



计算捆绑定价问题选择概率的梯度下降算法

献给韩继业教授 90 寿辰

张立平*, 赵逸潭

清华大学数学科学系, 北京 100084

E-mail: lipingzhang@tsinghua.edu.cn, 1176740923@qq.com

收稿日期: 2024-02-28; 接受日期: 2024-09-22; 网络出版日期: 2024-12-02; * 通信作者
国家自然科学基金 (批准号: 12171271) 资助项目

摘要 捆绑定价问题在服务业、电视和电影、媒体等行业有重要的应用, 得到了广泛研究. 本文证明捆绑定价问题交叉矩模型的目标函数在约束域内是强凸的且梯度是 Lipschitz 连续的. 基于此性质, 本文提出求解该模型的梯度下降算法并计算捆绑定价问题选择概率. 本文还证明了算法是线性收敛的. 数值实验验证了算法的有效性和交叉矩模型对于选择概率估计的准确性.

关键词 捆绑定价 交叉矩模型 梯度下降算法 选择概率

MSC (2020) 主题分类 90C30, 90C90, 90C25

1 引言

捆绑定价问题是指一个风险中性的公司将不同的产品打包成套餐出售, 所考虑的定价机制是规模定价, 即套餐的价格仅取决于套餐内商品的数量. 消费者对于商品的预期价格服从某种分布, 他们选择商品来最大化消费者剩余, 而公司的目标是最大化预期收益. 捆绑定价在电视、运输、媒体和服务等行业中被普遍采用. 例如, 有线电视公司允许客户在规定的范围内选择频道组合 (参见文献 [11]). 剧院公司和运动队在一个赛季中向客户销售批量的不同演出和比赛的门票, 每个套餐的价格通常只取决于套餐的大小 (参见文献 [6]).

尽管捆绑定价机制在实践中很有吸引力, 但要确定最优的捆绑价格是非常困难的, 其关键是确定给定价格下消费者选择不同套餐的分布情形, 这在数学上是一个离散选择问题, 涉及次序统计量的最大部分和的分布特征 (参见文献 [7, 8]).

由于捆绑定价问题考虑消费者对商品的最高预期价格, 所以计算选择概率需要考虑次序统计量的分布, 这在计算上非常复杂 (参见文献 [2, 12]). 2012 年, Mishra 等 [8] 首次提出了容易计算捆绑定价选择概率的交叉矩模型, 引起了学者们的关注和研究 [3, 4]. 在此基础上, Li 等 [7] 提出了在离散选择模

英文引用格式: Zhang L P, Zhao Y T. A gradient descent algorithm for computing choice probability in the bundle size pricing problem (in Chinese). Sci Sin Math, 2025, 55: 385–396, doi: 10.1360/SSM-2024-0046

型下, 仅使用客户估值的一阶和二阶矩来解决捆绑定价问题的改进方法, 从理论上证明了交叉矩模型下的捆绑定价问题是凸的. 改进的交叉矩模型从一组具有固定均值向量和协方差矩阵的分布中寻找极值分布, 以实现消费者效用的最大化, 套餐间的相关性可以通过协方差矩阵来获得, 从而可以通过数值模拟来计算各种捆绑的选择概率. 但是, 文献 [7] 没有提供用来求解交叉矩模型的算法, 仅是通过 MATLAB 工具箱里的约束优化求解器 “fmincon” 来求解.

鉴于此, 本文对文献 [7] 给出的交叉矩模型进行详细的分析, 并证明其目标函数在约束域上梯度是 Lipschitz 连续的. 根据这一性质, 本文提出一种梯度下降算法求解交叉矩模型以此计算捆绑定价的选择概率, 并证明该算法是线性收敛的. 数值实验验证了算法的有效性和交叉矩模型对于选择概率估计的准确性.

本文所用记号说明如下. 黑斜体小写字母表示向量, 如 $\mathbf{x}$, $\|\mathbf{x}\|$ 表示向量 $\mathbf{x}$ 的 Eucli 范数, $\|\mathbf{x}\|_1$ 表示向量 $\mathbf{x}$ 的 ℓ_1 范数. $\mathbf{e}$ 表示全 1 向量. $\mathbb{R}_+^n$ ($\mathbb{R}_{++}^n$) 表示 n 维非负 (正) 向量的集合. $\text{Diag}(\mathbf{x})$ 表示对角线元素为 $\mathbf{x}$ 的对角矩阵. $\text{diag}(X)$ 表示矩阵 X 的对角线元素组成的列向量. tr 表示矩阵的迹. $\text{int}(S)$ 表示集合 S 的内部, ∂S 表示集合 S 的边界.

本文余下内容的结构如下. 第 2 节介绍捆绑定价问题和交叉矩模型. 第 3 节提出计算选择概率的梯度下降算法并进行算法的收敛性分析. 第 4 节给出数值实验. 第 5 节给出本文结论.

2 捆绑定价与交叉矩模型

本节主要介绍文献 [7] 中给出的捆绑定价问题和交叉矩模型的数学形式.

考虑一个公司出售 n 个不同种类的商品给同质客户群体, 客户对商品的预期价格由一个随机向量 $\tilde{\mathbf{u}} = (\tilde{u}_1, \dots, \tilde{u}_n)^T$ 来表示. 假定 $\tilde{\mathbf{u}}$ 服从给定的有着有限一阶和二阶矩的分布 F , 令 $\tilde{u}_0 = 0$ 为选择之外的预期价格. 用 $\mathcal{S}$ 表示公司提供的捆绑销售的所有商品套餐, 则 $\mathcal{S} \subset [n] = \{1, \dots, n\}$ 且不妨设 $|\mathcal{S}| = m$. 令 c_s 为套餐 s 的成本. 令 $\mathbf{p} \in \mathbb{R}^m$ 为所有套餐的价格向量. 对于 $s \in \mathcal{S}$, 公司设定价格 p_s 来最大化收益, 消费者选择合适的套餐来最大化消费者剩余, 即最高预期价格减去实际支付的价格 p_s . 用 $\tilde{\mathbf{w}}$ 表示消费者对每个套餐的最高预期价格向量, 而 $\tilde{w}_s$ 显然是套餐中商品个数 s 以及预期价格 $\tilde{\mathbf{u}}$ 的函数. 定义 $\tilde{\mathbf{u}}$ 的第 i 个次序统计量 (即 $\tilde{\mathbf{u}}$ 中第 i 小的元素) 为 $\tilde{u}_{(i)}$, 则 $\tilde{w}_s(\tilde{\mathbf{u}})$ 可以写成

$$\tilde{w}_s(\tilde{\mathbf{u}}) = \sum_{k=0}^{s-1} \tilde{u}_{(n-k)}, \quad \forall s \in \mathcal{S}. \quad (2.1)$$

(2.1) 表示对于大小为 s 的套餐, 消费者选取他们认为最值钱的 s 个商品.

我们需要如下假设^[1, 7]. 假设所有商品的价格敏感性都相同且等于 1, 商品的估值是累加的, 购买的套餐中每个商品最多包含一件; 假设商品之间的边际成本是相同的, 即 c_s 是一个确定的量. 在这样的假设下, 每个套餐的需求即为消费者选取这个套餐的概率. 记 $\tilde{\mathbf{w}}$ 的联合分布函数为 G , 则由 (2.1) 及 $\tilde{\mathbf{u}}$ 服从给定的有着有限一阶和二阶矩的分布 F 可知, G 是关于 F 的函数, 从而所谓的捆绑定价问题可以表示为如下的数学形式:

$$\begin{aligned} \max_{\mathbf{p} \geq 0} \quad & \sum_{s \in \mathcal{S}} (p_s - c_s) q_s^*(\mathbf{p}) \\ \text{s.t.} \quad & q_s^*(\mathbf{p}) = \mathbb{P}_{\tilde{\mathbf{w}} \sim G} \left(s = \arg \max_{i \in \{0\} \cup \mathcal{S}} \{ \tilde{w}_i(\tilde{\mathbf{u}}) - p_i \} \right), \quad \forall s \in \mathcal{S}, \end{aligned} \quad (2.2)$$

其中 $q^*(p)$ 表示由给定价格 p 诱导的套餐 s 的预期需求. 捆绑定价问题 (2.2) 是一个标准的离散选择框架下的利润最大化问题, 其难点在于如何描述预期需求 $q^*(p)$. 2022 年, Li 等^[7] 利用交叉矩模型^[8,9] 得到了 (2.2) 的凸优化模型.

虽然刻画分布函数 G 很困难, 但 $\tilde{w}(\tilde{u})$ 的一阶矩和二阶矩可以通过 F 很容易地计算出来, 而且这些经验估计都是无偏的, 并记 $\omega = \mathbb{E}_G[\tilde{w}(\tilde{u})]$ 和 $\Sigma = \text{Cov}_G(\tilde{w}(\tilde{u}))$. 用 Θ 表示 $\tilde{w}$ 的期望为 ω 和方差为 Σ 的分布集合, 即

$$\Theta = \{\theta \mid \mathbb{E}_\theta[\tilde{w}] = \omega, \text{Cov}_\theta(\tilde{w}) = \Sigma\}.$$

显然, G 为 Θ 中的一个元素. 对于给定的价格 p , 交叉矩模型^[8] 给出了如下的消费者预期效用剩余的上界:

$$Z^* = \sup_{\theta \in \Theta} \mathbb{E}_\theta \left(\max_{s \in \{0\} \cup \mathcal{S}} \{\tilde{w}_s(\tilde{u}) - p_s\} \right). \quad (2.3)$$

于是, Li 等^[7] 给出了如下的交叉矩模型下的捆绑定价问题来近似问题 (2.2):

$$\begin{aligned} \max_{p \geq 0} \quad & \sum_{s \in \mathcal{S}} (p_s - c_s) q_s(p) \\ \text{s.t.} \quad & q_s(p) = \mathbb{P}_{\tilde{w} \sim \theta^*(p)} \left(s = \arg \max_{i \in \{0\} \cup \mathcal{S}} \{\tilde{w}_i(\tilde{u}) - p_i\} \right), \quad \forall s \in \mathcal{S}, \end{aligned} \quad (2.4)$$

其中 $\theta^*(p)$ 是由 (2.3) 确定的最优分布. 显然, 对于价格 p 的不同取值和固定的 Θ , $\theta^*(p)$ 也可能不同. 虽然 $\theta^*(p)$ 与 G 不同, 但它们有着相同的期望和方差, 从而 $q_s(p)$ 是 $q_s^*(p)$ 的一个很好的近似. 因此, Li 等^[7] 在假设方差 Σ 是正定矩阵的条件下证明了捆绑定价问题的交叉矩模型 (2.4) 可以转化为如下的凸优化 (参见文献 [7, 定理 2]):

$$\max_{q \in \Delta_m} \pi(q) = q^T(\omega - c) + \frac{1}{2} q^T \text{diag}(\Sigma^{1/2}(T(q))^{-1/2} \Sigma^{1/2}) - q^T \Sigma^{1/2}(T(q))^{-1/2} \Sigma^{1/2} q, \quad (2.5)$$

其中 c 是套餐的预期成本, $\pi(q)$ 是预期利润, $\Delta_m = \{q \in \mathbb{R}_+^m \mid e^T q \leq 1\}$ 且

$$T(q) = \Sigma^{1/2} S(q) \Sigma^{1/2}, \quad S(q) = \text{Diag}(q) - qq^T. \quad (2.6)$$

虽然 (2.5) 是一个凸规划, 但并不容易求解. 根据文献 [7, 定理 1], 在已经给定套餐价格 p 并且客户对套餐的预期价格的均值向量 ω 和方差矩阵 Σ 已知的情形下, 可以通过求解如下的优化问题来计算捆绑定价 (2.4) 的选择概率:

$$q(p) = \arg \max_{x \in \Delta_m} \{(\omega - p)^T x + \text{tr}((T(x))^{1/2})\}, \quad (2.7)$$

其中 $T(x)$ 的定义由 (2.6) 给出. 文献 [7] 没有给出计算选择概率的算法, 因此在下一节将基于问题的特有结构提出一个新的梯度下降算法求解 (2.7) 以解决交叉矩模型下捆绑定价的选择概率计算问题.

3 计算选择概率的梯度下降算法

本节将提出求解 (2.7) 的梯度下降算法并通过分析模型 (2.7) 的重要性质建立算法的线性收敛性.

3.1 梯度下降算法

令 $\mathbf{x} \in \mathbb{R}^m$, 定义

$$f(\mathbf{x}) = -(\boldsymbol{\omega} - \mathbf{p})^T \mathbf{x} - \text{tr}((\boldsymbol{\Sigma}^{1/2} S(\mathbf{x}) \boldsymbol{\Sigma}^{1/2})^{1/2}), \quad (3.1)$$

则 (2.7) 可以改写为

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } \mathbf{x} \in \Delta_m := \{\mathbf{x} \in \mathbb{R}_+^m \mid \mathbf{e}^T \mathbf{x} \leq 1\}. \end{aligned} \quad (3.2)$$

通过下面的引理证明了 $f(\mathbf{x})$ 的定义域就是 Δ_m , 因此上述约束优化问题可以转化为无约束优化问题来求解. 下面的引理在文献 [3] 中已有描述, 便于阅读, 下面给出详细证明.

引理 3.1 $S(\mathbf{x}) = \text{Diag}(\mathbf{x}) - \mathbf{x}\mathbf{x}^T$ 为半正定矩阵的充要条件是 $\mathbf{x} \in \Delta_m$.

证明 充分性 对任意给定的 $\mathbf{x} \in \Delta_m$, 若 $\mathbf{x} = \mathbf{0}$, 则 $S(\mathbf{x})$ 为零矩阵. 显然是半正定的. 设 $\mathbf{x} \neq \mathbf{0}$, 任取 $\mathbf{d} \in \mathbb{R}^m$ 且 $\mathbf{d} \neq \mathbf{0}$, 则根据 Jensen 不等式可得

$$\frac{1}{\mathbf{e}^T \mathbf{x}} \left(\sum_{i=1}^m d_i x_i \right)^2 = \mathbf{e}^T \mathbf{x} \left(\sum_{i=1}^m d_i \frac{x_i}{\mathbf{e}^T \mathbf{x}} \right)^2 \leq \mathbf{e}^T \mathbf{x} \sum_{i=1}^m d_i^2 \frac{x_i}{\mathbf{e}^T \mathbf{x}} = \sum_{i=1}^m d_i^2 x_i.$$

由上式可推导出

$$\begin{aligned} \mathbf{d}^T S(\mathbf{x}) \mathbf{d} &= \mathbf{d}^T \text{Diag}(\mathbf{x}) \mathbf{d} - (\mathbf{d}^T \mathbf{x})^2 \\ &= \sum_{i=1}^m d_i^2 x_i - \left(\sum_{i=1}^m d_i x_i \right)^2 \\ &\geq \frac{1}{\mathbf{e}^T \mathbf{x}} \left(\sum_{i=1}^m d_i x_i \right)^2 - \left(\sum_{i=1}^m d_i x_i \right)^2 \\ &\geq 0. \end{aligned}$$

因此, $S(\mathbf{x})$ 在 Δ_m 上是半正定的.

必要性 对于任意 $i \in [m]$, 取 $\mathbb{R}^m$ 的自然基 $\mathbf{e}_i$, 由 $S(\mathbf{x})$ 半正定知 $\mathbf{e}_i^T S(\mathbf{x}) \mathbf{e}_i \geq 0$. 又因为

$$\mathbf{e}_i^T S(\mathbf{x}) \mathbf{e}_i = x_i - x_i^2,$$

故有 $0 \leq x_i \leq 1$.

同理可得

$$0 \leq \mathbf{e}^T S(\mathbf{x}) \mathbf{e} = \mathbf{e}^T \mathbf{x} - (\mathbf{e}^T \mathbf{x})^2 \Rightarrow 0 \leq \mathbf{e}^T \mathbf{x} \leq 1.$$

因此, $\mathbf{x} \in \Delta_m$. □

根据引理 3.1 的证明, 立即可证得如下结论.

引理 3.2 $S(\mathbf{x})$ 为正定矩阵的充要条件是 $\mathbf{x} \in \text{int}(\Delta_m) := \{\mathbf{x} \in \mathbb{R}_{++}^m \mid \mathbf{e}^T \mathbf{x} < 1\}$.

引理 3.1 说明 $f(\mathbf{x})$ 的定义域就是约束优化问题 (2.7) 的可行域, 因此不能通过罚函数类算法来求解 (2.7). 在已知方差矩阵 $\boldsymbol{\Sigma}$ 正定的假设条件下, Ahipasaoglu 等 [3] 证明了

$$h(\mathbf{x}) = \text{tr}((\boldsymbol{\Sigma}^{1/2} S(\mathbf{x}) \boldsymbol{\Sigma}^{1/2})^{1/2})$$

在 Δ_m 上是强凹函数, 在 $\text{int}(\Delta_m)$ 上是可微的且

$$\nabla h(\mathbf{x}) = \frac{1}{2} \text{diag}(\Sigma^{1/2}(T(\mathbf{x}))^{-1/2} \Sigma^{1/2}) - \Sigma^{1/2}(T(\mathbf{x}))^{-1/2} \Sigma^{1/2} \mathbf{x},$$

其中 $T(\mathbf{x}) = \Sigma^{1/2} S(\mathbf{x}) \Sigma^{1/2}$. 于是由 (3.1) 和文献 [7, 定理 2] 立即可知以下结论成立.

定理 3.1 设 Σ 正定, 则 $f(\mathbf{x})$ 是 $\text{int}(\Delta_m)$ 上的强凸函数, 即存在常数 $\mu > 0$ 使得

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|^2, \quad \forall \mathbf{x}, \mathbf{y} \in \text{int}(\Delta_m), \quad (3.3)$$

其中 $\nabla f(\mathbf{x}) = \mathbf{p} - \boldsymbol{\omega} - \nabla h(\mathbf{x})$.

下面的定理已由文献 [7, 引理 1] 给出.

定理 3.2 设 Σ 正定, 则当 $\mathbf{x}$ 趋于 $\partial\Delta_m$ 时, $\|\nabla f(\mathbf{x})\| \rightarrow \infty$.

定理 3.2 说明, 当 $\mathbf{x}$ 趋于 $\partial\Delta_m$ 时, 问题 (3.2) 的目标函数的方向导数趋于 $+\infty$. 结合定理 3.1 及凸函数的性质可知, 问题 (3.2) 的目标函数在 $\text{int}(\Delta_m)$ 上必存在极小值, 进而存在梯度为 0 的点 [3, 7]. 因此, 问题 (3.2) 的一阶最优性条件为

$$\nabla f(\mathbf{x}) = 0, \quad \mathbf{x} \in \text{int}(\Delta_m). \quad (3.4)$$

对于任意给定的套餐价格 $\mathbf{p}$, (3.4) 可以给出唯一的最优选择概率 $\mathbf{q}(\mathbf{p})$, 它就是 $f(\mathbf{x})$ 在 $\text{int}(\Delta_m)$ 上的全局极小值点, 记作 $\mathbf{x}^*$. 根据定理 3.1 和引理 3.3, 目标函数 $f(\mathbf{x})$ 的强凸性和其梯度函数的 Lipschitz 连续性都是受限制的, 因此已有的算法并不能直接用于求解 $\mathbf{x}^*$. 于是, 本文提出如下的梯度算法 (算法 1) 求解这个全局极小值点 $\mathbf{x}^*$.

算法 1 计算选择概率的梯度下降法

Input: $\boldsymbol{\omega}, \Sigma, \mathbf{p}$, 初始迭代点 $\mathbf{x}_0 \in \text{int}(\Delta_m)$, 初始迭代步长 $\alpha_0 \in (0, 1)$, 参数 $\beta, \gamma \in (0, 1)$, 终止精度 $\varepsilon > 0$.

Iteration:

```

1:  $k \leftarrow 0$ ,
2:  $\mathbf{g}_0 \leftarrow \nabla f(\mathbf{x}_0)$ ,
3: while  $\|\mathbf{g}_k\| > \varepsilon$  do
4:   while  $\mathbf{x}_k - \alpha_k \mathbf{g}_k \notin \text{int}(\Delta_m)$  or  $f(\mathbf{x}_k - \alpha_k \mathbf{g}_k) > f(\mathbf{x}_k) - \gamma \alpha_k \|\mathbf{g}_k\|^2$  do
5:      $\alpha_k \leftarrow \beta \alpha_k$ ,
6:   end while
7:    $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \alpha_k \mathbf{g}_k$ ,
8:    $\mathbf{g}_{k+1} \leftarrow \nabla f(\mathbf{x}_{k+1})$ ,
9:    $\mathbf{y}_k \leftarrow \mathbf{g}_{k+1} - \mathbf{g}_k$ ,
10:   $\alpha_{k+1} \leftarrow -\frac{\alpha_k \mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T \mathbf{y}_k}$ ,
11:   $k \leftarrow k + 1$ ,
12: end while

```

Output: 最优解 $\mathbf{x}_k$.

关于算法 1, 我们有以下说明:

(i) 步骤 10 给出的步长是 Barzilai-Borwein (BB) 型步长 [5, 10]. 这是因为令 $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, 则步骤 10 可以等价地写成

$$\alpha_{k+1} \leftarrow -\frac{\alpha_k \mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T \mathbf{y}_k} = \frac{\mathbf{s}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{y}_k}.$$

根据定理 3.1, f 强凸, 故有 $\mathbf{s}_k^T \mathbf{y}_k > 0$, 上式是有意义的.

(ii) 步骤 4–6 不会陷入死循环. 因为 $\mathbf{x}_0 \in \text{int}(\Delta_m)$ 且 Δ_m 是凸集, 所以存在充分小的 $\delta > 0$ 使得迭代点列 $\{\mathbf{x}_k\}$ 满足 $\mathbf{x}_k \in \text{int}(\Delta_m)$ 且

$$f(\mathbf{x}_k - \alpha_k \mathbf{g}_k) \leq f(\mathbf{x}_k) - \gamma \alpha_k \|\mathbf{g}_k\|^2.$$

因此, 算法 1 是有定义的, 可将其看作是一种 BB 梯度下降算法.

3.2 收敛性分析

本小节证明算法 1 是线性收敛的. 在此之前, 首先需要给出下面的引理.

引理 3.3 取 $\mathbf{x}_0$ 满足 $f(\mathbf{x}_0) < \inf_{\mathbf{x} \in \partial \Delta_m} f(\mathbf{x})$, 令

$$\Omega = \{\mathbf{x} \in \text{int}(\Delta_m) \mid f(\mathbf{x}) \leq f(\mathbf{x}_0)\}, \quad (3.5)$$

则下列结论成立.

- (i) Ω 为紧凸集且存在一个开集 $O \supset \Omega$ 使得 $O \subset \Delta_m$.
- (ii) $\nabla f(\mathbf{x})$ 在 Ω 上是 Lipschitz 连续的, 即存在常数 $L > 0$ 使得

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \Omega;$$

而且对任意的 $\mathbf{x}, \mathbf{y} \in \Omega$, 有

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2. \quad (3.6)$$

证明 (i) 根据定理 3.1, 可知 Ω 为紧凸集. 由 $\mathbf{x}_0$ 的定义可知 $\Omega \cap \partial \Delta_m = \emptyset$. 又因为 Ω 是紧集, 所以存在开集 $O \supset \Omega$ 使得 $O \subset \Delta_m$.

(ii) 由定理 3.1 及 $\nabla f(\mathbf{x})$ 的定义知, $\nabla f(\mathbf{x})$ 在 Ω 上连续可微. 因为 Ω 为紧凸集, 所以 $\nabla f(\mathbf{x})$ 在 Ω 上是 Lipschitz 连续的, 其 Lipschitz 常数记作 L . 因此, $f(\mathbf{x})$ 在 Ω 上有二次上界, 即对任意的 $\mathbf{x}, \mathbf{y} \in \Omega$, (3.6) 成立. $\square$

下面证明算法 1 的收敛性, 可归结为以下定理.

定理 3.3 设 $\{\mathbf{x}_k\}$ 是由算法 1 产生的迭代点列, $\{\mathbf{g}_k\}$ 是相应的 f 的梯度序列. 设初始迭代点 $\mathbf{x}_0$ 满足 $f(\mathbf{x}_0) < \inf_{\mathbf{x} \in \partial \Delta_m} f(\mathbf{x})$, 则要么算法 1 有限步终止, 要么成立

$$\lim_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0, \quad \lim_{k \rightarrow \infty} \mathbf{x}_k = \mathbf{x}^*. \quad (3.7)$$

证明 设算法 1 不是有限步终止的, 则需要证明 (3.7) 成立. 记 $\{\alpha_k\}$ 为算法 1 产生的迭代步长序列, 则 α_k 由以下方式给出:

$$\hat{\alpha}_k = \frac{\mathbf{s}_{k-1}^T \mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}, \quad \alpha_k = \beta^{n_k} \hat{\alpha}_k,$$

其中当 $k = 0$ 时 $\hat{\alpha}_k = \alpha_0$, 正整数 n_k 定义如下:

$$n_k := \min\{n \in \mathbb{N} \mid \mathbf{x}_k - \beta^n \hat{\alpha}_k \mathbf{g}_k \in \text{int}(\Delta_m), f(\mathbf{x}_k - \beta^n \hat{\alpha}_k \mathbf{g}_k) \leq f(\mathbf{x}_k) - \gamma \beta^n \hat{\alpha}_k \|\mathbf{g}_k\|^2\}.$$

下面分 3 步来证明 (3.7) 成立.

第 1 步 存在常数 $M_1, M_2 > 0$ 使得对任意的 k 有 $M_1 < \hat{\alpha}_k < M_2$.

我们来证明第 1 中的结论. 根据定理 3.1, 在 (3.3) 中分别取 $\mathbf{x} = \mathbf{x}_k, \mathbf{y} = \mathbf{x}_{k+1}$ 和 $\mathbf{x} = \mathbf{x}_{k+1}, \mathbf{y} = \mathbf{x}_k$, 可得

$$\begin{aligned} \mathbf{s}_k^T \nabla f(\mathbf{x}_{k+1}) &\geq f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) + \frac{\mu}{2} \|\mathbf{s}_k\|^2, \\ -\mathbf{s}_k^T \nabla f(\mathbf{x}_k) &\geq f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) + \frac{\mu}{2} \|\mathbf{s}_k\|^2. \end{aligned}$$

以上两式相加可得

$$\hat{\alpha}_{k+1} = \frac{\mathbf{s}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{y}_k} \leq \frac{1}{\mu}.$$

令

$$M_2 = \max \left\{ \frac{1}{\mu}, \alpha_0 \right\},$$

则对任意的 k 有 $\hat{\alpha}_k \leq M_2$ 成立.

因 f 强凸, 故有 $\mathbf{s}_k^T \mathbf{y}_k > 0$, 从而 α_k 均为正值. 又因

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \gamma \alpha_k \|\mathbf{g}_k\|^2,$$

故 $\{f(\mathbf{x}_k)\}$ 单调下降. 令

$$\Omega := \{\mathbf{x} \in \text{int}(\Delta_m) \mid f(\mathbf{x}) \leq f(\mathbf{x}_0)\},$$

则有 $\{\mathbf{x}_k\} \subset \Omega$. 根据引理 3.3(ii), 可得 $\|\mathbf{y}_k\| \leq L \|\mathbf{s}_k\|$. 因此, 根据 Cauchy 不等式, 可得

$$\hat{\alpha}_{k+1} = \frac{\mathbf{s}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{y}_k} \geq \frac{\|\mathbf{s}_k\|}{\|\mathbf{y}_k\|} \geq \frac{1}{L}.$$

令

$$M_1 = \min \left\{ \frac{1}{L}, \alpha_0 \right\},$$

则对任意的 k , $\hat{\alpha}_k \geq M_1$ 成立. 综上, 第 1 步得证.

第 2 步 存在 $M > 0$, 使得当 $0 \leq \alpha \leq \frac{M}{\beta}$ 时, 对任意的 k 有

$$\mathbf{x}_k - \alpha \mathbf{g}_k \in \text{int}(\Delta_m), \quad f(\mathbf{x}_k - \alpha \mathbf{g}_k) \leq f(\mathbf{x}_k) - \gamma \alpha \|\mathbf{g}_k\|^2,$$

进而有

$$\min\{M, M_1\} \leq \alpha_k \leq M_2.$$

我们来证明第 2 步中的结论. 由引理 3.3(i) 可知

$$d_1 := \text{dist}(\partial\Omega, \partial\Delta_m) > 0 \quad \text{且} \quad d_2 := \sup_{\mathbf{x} \in \Omega} \|\nabla f(\mathbf{x})\| < \infty,$$

则当 $0 \leq \alpha \leq \frac{d_1}{2d_2}$ 时有

$$\mathbf{x}_k - \alpha \mathbf{g}_k \in \text{int}(\Delta_m).$$

另外, 由引理 3.3(ii) 可知

$$f(\mathbf{x}_k - \alpha \mathbf{g}_k) \leq f(\mathbf{x}_k) - \alpha \|\mathbf{g}_k\|^2 + \frac{\alpha^2 L}{2} \|\mathbf{g}_k\|^2.$$

所以, 当 $0 \leq \alpha \leq \frac{2(1-\gamma)}{L}$ 时, 有

$$f(\mathbf{x}_k - \alpha \mathbf{g}_k) \leq f(\mathbf{x}_k) - \gamma \alpha \|\mathbf{g}_k\|^2.$$

因此, 取

$$M = \beta \min \left\{ \frac{d_1}{2d_2}, \frac{2(1-\gamma)}{L} \right\},$$

则第 2 步得证.

第 3 步 $\lim_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$ 且 $\lim_{k \rightarrow \infty} \mathbf{x}_k = \mathbf{x}^*$.

我们来证明第 3 步中的结论. 函数值序列 $\{f(\mathbf{x}_k)\}$ 单调下降且有界, 故而收敛. 又因为

$$\gamma \alpha_k \|\mathbf{g}_k\|^2 \leq f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}),$$

所以,

$$\lim_{k \rightarrow \infty} \alpha_k \|\mathbf{g}_k\|^2 = 0.$$

由第 2 步知, $\{\alpha_k\}$ 满足 $\min\{M, M_1\} \leq \alpha_k \leq M_2$. 因此有

$$\lim_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0. \quad (3.8)$$

下面用反证法证明 $\lim_{k \rightarrow \infty} \mathbf{x}_k = \mathbf{x}^*$. 若不然, 则存在 $\varepsilon_0 > 0$ 以及子列 $\{\mathbf{x}_{k_j}\}$ 使得 $\|\mathbf{x}_{k_j} - \mathbf{x}^*\| > \varepsilon_0$. 令 $K = \max_{\mathbf{x} \in \Delta_m} \|\mathbf{x}\|$, 则根据 $\{f(\mathbf{x}_k)\}$ 单调下降、定理 3.1 及 (3.3), 可得

$$\begin{aligned} \|\mathbf{g}_{k_j}\| &\geq \frac{1}{2K} \|\mathbf{x}_{k_j} - \mathbf{x}^*\| \|\mathbf{g}_{k_j}\| \geq \frac{1}{2K} (\mathbf{x}_{k_j} - \mathbf{x}^*)^T \mathbf{g}_{k_j} \\ &\geq \frac{1}{2K} (f(\mathbf{x}_{k_j}) - f(\mathbf{x}^*)) + \frac{\mu}{4K} \|\mathbf{x}_{k_j} - \mathbf{x}^*\|^2 \\ &\geq \frac{\mu \varepsilon_0^2}{4K}. \end{aligned}$$

这与 (3.8) 矛盾. 因此, 第 3 步得证, 即 (3.7) 成立. 定理证毕. $\square$

我们可以进一步证明当 $\gamma \in [0.5, 1)$ 时, 算法 1 是 Q- 线性收敛的.

定理 3.4 设 $\{\mathbf{x}_k\}$ 是由算法 1 产生的迭代点列. 令 $\delta_k = \|\mathbf{x}_k - \mathbf{x}^*\|$, 则当 $\gamma \in [0.5, 1)$ 时, 存在 $q \in (0, 1)$ 使得当 $\delta_{k+1} \leq q \delta_k$.

证明 由 δ_k 的定义及 $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{g}_k$ 知

$$\delta_{k+1}^2 = \|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 = \delta_k^2 + \alpha_k^2 \|\mathbf{g}_k\|^2 + 2\alpha_k \mathbf{g}_k^T (\mathbf{x}^* - \mathbf{x}_k). \quad (3.9)$$

令 $\epsilon_k = f(\mathbf{x}^*) - f(\mathbf{x}_k)$, 则 $\epsilon_k \leq 0$ 且 $\epsilon_{k+1} \geq \epsilon_k$. 由定理 3.1 知

$$\mathbf{g}_k^T (\mathbf{x}^* - \mathbf{x}_k) \leq \epsilon_k - \frac{\mu}{2} \delta_k^2.$$

结合 (3.9), 有

$$\delta_{k+1}^2 \leq (1 - \mu\alpha_k)\delta_k^2 + \alpha_k(2\epsilon_k + \alpha_k\|\mathbf{g}_k\|^2). \quad (3.10)$$

另外, 根据算法 1 的迭代过程 $\gamma \in [0.5, 1)$, 可得

$$\alpha_k\|\mathbf{g}_k\|^2 \leq \frac{1}{\gamma}(f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})) \leq 2(\epsilon_{k+1} - \epsilon_k).$$

从而, 有

$$\alpha_k(2\epsilon_k + \alpha_k\|\mathbf{g}_k\|^2) \leq 2\alpha_k\epsilon_{k+1} \leq 0. \quad (3.11)$$

根据定理 3.3 中的第 1 步可知, 存在常数 $C > 0$ 使得 $C < \mu\alpha_k < 1$. 因此, 结合 (3.10) 和 (3.11), 可得

$$\delta_{k+1}^2 \leq (1 - C)\delta_k^2. \quad (3.12)$$

取 $q = 1 - C$, 则 $q \in (0, 1)$, 且 (3.12) 表明定理结论成立. $\square$

定理 3.3 和 3.4 说明算法 1 要么有限步终止, 输出基于交叉矩模型 (3.2) 的捆绑定价的唯一最优的选择概率 $\mathbf{x}^*$; 要么产生迭代点列线性收敛于其唯一最优的选择概率 $\mathbf{x}^*$. 特别地, 从定理 3.4 可以看出, 对于给定精度 $\varepsilon > 0$ (如 $\varepsilon = 10^{-8}$), 由

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq q\|\mathbf{x}_{k-1} - \mathbf{x}^*\| \leq \cdots \leq q^k\|\mathbf{x}_0 - \mathbf{x}^*\|$$

可知, 算法 1 至多 $K = \lceil \log(C\varepsilon)/\log q \rceil$ 步迭代可保证 $\|\mathbf{x}_k - \mathbf{x}^*\| < \varepsilon$, 其中 $C = 1/\|\mathbf{x}_0 - \mathbf{x}^*\|$. 从这个角度讲, 算法 1 的收敛速度在一定条件下是优于 Q- 线性收敛的.

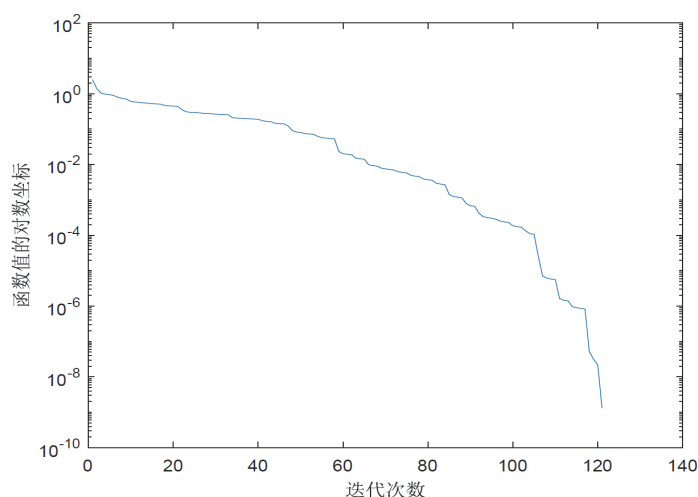
4 数值实验

本节通过数值模拟来验证算法 1 的实际效果, 分别从参数和初值两方面来展示其性能. 最后利用算法 1 计算由交叉矩模型 (3.2) 给出的选择概率并与实际的选择概率相比较, 验证交叉矩模型 (3.2) 和算法 1 的有效性.

首先通过数值模拟得到输入值 $\boldsymbol{\omega}$ 和 $\boldsymbol{\Sigma}$, 并确定合适的价格 $\mathbf{p}$. 设定商品的个数 n 并假定公司提供全部大小的套餐, 即 $\mathcal{S} = [n]$. 假设顾客对商品的预期价格服从多元正态分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0)$, 取 $\boldsymbol{\mu} = 3\mathbf{e}$, $\boldsymbol{\Sigma}_0$ 的特征值在 $[0, 1]$ 上均匀抽样得到, 对应的特征向量由每个元素均为 $[-1, 1]$ 上均匀抽样的随机矩阵正交化得到. 由此, 得到 $N = 10,000$ 个顾客的预期价格向量的样本 $\mathbf{u}^1, \dots, \mathbf{u}^N$, 再根据 (2.1) 可以得到顾客对于套餐的预期价格向量 $\mathbf{w}^1, \dots, \mathbf{w}^N$. 计算它们的均值向量和协方差矩阵分别为 $\boldsymbol{\omega}$ 和 $\boldsymbol{\Sigma}$. 我们随机生成一个可行的价格 $\mathbf{p}$: 在 (2.5, 3.5) 上均匀抽取 n 个数并降序排列, 记为 $p_1 \geq p_2 \geq \cdots \geq p_n$. 令 $\mathbf{p} = (sp_s)_{1 \leq s \leq n}$, 保证了套餐中的商品越多, 每个商品的平均价格就越低. 这样的设定符合现实中的常见情形.

设置参数 $n = 10$, $\gamma = 0.5$, $\varepsilon = 10^{-3}$, 取初值 $\mathbf{x}_0 = \frac{1}{2n}\mathbf{e}$. 利用算法 1 得到迭代序列 $\{\mathbf{x}_k\}$ 和近似最优解 $\mathbf{x}^*$, 将 $f(\mathbf{x}_k) - f(\mathbf{x}^*)$ 绘制于对数坐标轴, 如图 1 所示. 从图 1 可以看出算法 1 在 120 次迭代收敛.

下面考察参数 γ 对算法 1 收敛速度的影响. 固定 $\varepsilon = 10^{-3}$, 取 $n = 10, 20, \dots, 50$ 和 $\gamma = 0.1, 0.2, \dots, 0.9$, 记录算法 1 收敛所需的迭代次数. 对于同一组 n 值、不同 γ 值的实验采用相同的 $\boldsymbol{\omega}$ 和 $\boldsymbol{\Sigma}$, 实

图 1 (网络版彩图) 函数值 $f(x_k) - f(x^*)$ 随迭代次数的变化数值结果

验结果见表 1. 从表中数据可以看出, 在相同 n 值下, 收敛所需要的迭代次数随 γ 的增大先减少后增大, 大概在 $\gamma = 0.4, 0.5, 0.6$ 时收敛最快.

最后, 通过算法 1 计算捆绑定价问题下使用交叉矩模型 (3.2) 估计的选择概率, 并与真实的选择概率作比较, 验证交叉矩模型 (3.2) 对选择概率估计的准确性. 对于 $n = 10, 20, 30, 40, 50$, 利用算法 1 求解 (3.2) 得到选择概率 $\mathbf{q}_{\text{CMM}}$. 同时利用生成的 10,000 组顾客对于套餐的最高预期价格向量, 根据

$$\mathbf{q}_s^*(\mathbf{p}) = \mathbb{P}_{\mathbf{w} \sim G} \left(s = \arg \max_{i \in \{0\} \cup \mathcal{S}} \{w_i(\mathbf{u}) - p_i\} \right)$$

得到顾客真实的选择概率 $\mathbf{q}^*$. 我们利用 $\delta_{\mathbf{q}} = \frac{1}{2} \|\mathbf{q}_{\text{CMM}} - \mathbf{q}^*\|_1$ 来衡量交叉矩模型 (3.2) 对选择概率的预测值和真实值的误差.

对于每一组 n 值, 我们做 100 次实验, 得到 100 个 $\delta_{\mathbf{q}}$ 值, 通过绘制箱线图来刻画 $\delta_{\mathbf{q}}$ 的分布, 结果如图 2 所示. 图 2 显示交叉矩模型 (3.2) 对选择概率的预测值相比真实值的误差在绝大多数情形下会小于 0.1, 而且几乎不随 n 的改变而变化. 这说明用交叉矩模型 (3.2) 来近似捆绑定价问题下的选择概率是合适的, 同时也说明算法 1 是有效的.

表 1 迭代次数: 参数 γ 的选取对算法 1 收敛速度的影响

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
$\gamma = 0.1$	605	3,126	2,360	20,112	28,524
$\gamma = 0.2$	486	3,326	1,827	19,931	19,760
$\gamma = 0.3$	330	2,665	2,449	11,931	20,842
$\gamma = 0.4$	314	2,089	1,474	7,477	14,564
$\gamma = 0.5$	314	2,089	1,474	7,477	14,564
$\gamma = 0.6$	467	1,702	2,485	12,342	16,593
$\gamma = 0.7$	629	1,863	2,604	8,627	14,578
$\gamma = 0.8$	1,258	3,144	4,716	19,428	32,036
$\gamma = 0.9$	2,350	6,429	11,412	45,765	69,585

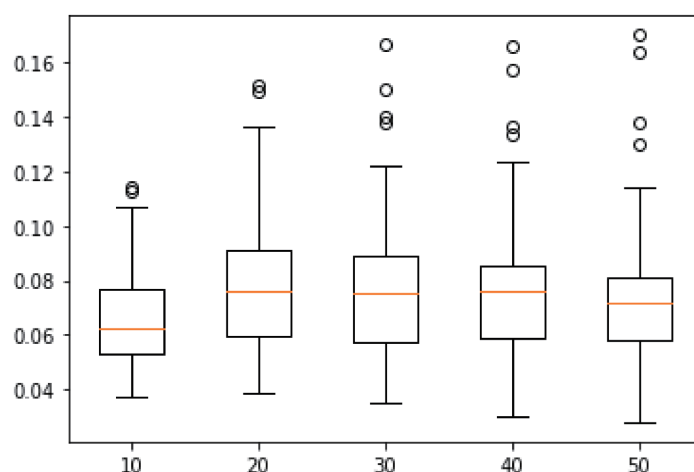


图 2 (网络版彩图) 不同 n 值下交叉矩模型 (3.2) 对选择概率的预测值相比真实值的误差

5 结论

本文考虑了捆绑定价的选择概率计算问题, 它是解决捆绑定价问题的关键. 本文建立了新形式的交叉矩模型 (3.2), 证明了其目标函数是强凸的并且梯度是 Lipschitz 连续的 (定理 3.1 和引理 3.3). 首次提出了 BB 梯度算法计算捆绑定价的选择概率 (算法 1), 证明了算法的收敛性 (定理 3.3), 并给出了线性收敛速度 (定理 3.4). 数值实验验证了算法的有效性和交叉矩模型 (3.2) 预测捆绑定价的选择概率的准确性. 在后面的工作中, 我们将研究直接求解基于一阶矩和二阶矩的捆绑定价的交叉矩模型 (2.5) 的有效算法.

参考文献

- 1 Abdallah T. On the benefit (or cost) of large-scale bundling. *Product Oper Manag*, 2019, 28: 955–969
- 2 Abdallah T, Asadpour A, Reed J. Large-scale bundle-size pricing: A theoretical analysis. *Oper Res*, 2021, 69: 1158–1185
- 3 Ahipasaoglu S D, Li X, Natarajan K. A convex optimization approach for computing correlated choice probabilities with many alternatives. *IEEE Trans Automat Control*, 2019, 64: 190–205
- 4 Ahipasaoglu S D, Meskarian R, Magnanti T L, et al. Beyond normality: A cross moment-stochastic user equilibrium model. *Trans Res Part B Methodol*, 2015, 81: 333–354
- 5 Barzilai J, Borwein J M. Two-point step size gradient methods. *IMA J Numer Anal*, 1988, 8: 141–148
- 6 Chu C S, Leslie P, Sorensen A. Bundle-size pricing as an approximation to mixed bundling. *Amer Econ Rev*, 2011, 101: 263–303
- 7 Li X, Sun H, Teo C P. Convex optimization for bundle size pricing problem. *Manage Sci*, 2022, 68: 1095–1106
- 8 Mishra V K, Natarajan K, Tao H, et al. Choice prediction with semidefinite optimization when utilities are correlated. *IEEE Trans Automat Control*, 2012, 57: 2450–2463
- 9 Natarajan K, Teo C P. On reduced semidefinite programs for second order moment bounds with applications. *Math Program*, 2017, 161: 487–518
- 10 Raydan M. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J Optim*, 1997, 7: 26–33
- 11 Wu J F, Tawarmalani M, Kannan K N. Cardinality bundling with Spence-Mirrlees reservation prices. *Manage Sci*, 2019, 65: 1891–1908
- 12 Wu S Y, Hitt L M, Chen P Y, et al. Customized bundle pricing for information goods: A nonlinear mixed-integer programming approach. *Manage Sci*, 2008, 54: 608–622

A gradient descent algorithm for computing choice probability in the bundle size pricing problem

Liping Zhang & Yitan Zhao

Abstract Bundle size pricing has important applications in the service industry, TV and film, media, and other industries and so has been widely studied. In this paper, we prove that the objective function of the cross-moment model of the bundle size pricing problem is strongly convex and the gradient is Lipschitz continuous in the constrained region. Based on this property, we propose a gradient descent algorithm to compute choice probability by solving this model. Moreover, we prove that the proposed algorithm is linearly convergent. Numerical experiments verify the effectiveness of the algorithm and the accuracy of the cross-moment model for the choice probability estimation.

Keywords bundle size pricing, cross-moment model, gradient descent algorithm, choice probability

MSC(2020) 90C30, 90C90, 90C25

doi: 10.1360/SSM-2024-0046