通过块坐标更新实现快速且可解释的费希尔市场动态

Tianlong Nan、Yuan Gao、Christian Kroer

哥伦比亚大学，美国纽约 {tianlong.nan,
gao.yuan, christian.kroer}@columbia.edu

arXiv:2303.00506v1 [cs.GT] 1 Mar 2023

摘要

我们研究如何利用可扩展的一阶优化方法计算大规模 Fisher 市场均衡。众所周知，市场均衡可以通过具有特定结构的凸规划来刻画，例如 Eisenberg-Gale 凸规划和 Shmyrev 凸规划。针对这些规划，已有高性能的确定性全梯度一阶方法。本文提出新的块坐标一阶方法来计算 Fisher 市场均衡，并证明这些方法可以解释为 tâtonnement 式或比例响应式动态，其中买方或物品逐一出现。我们对这些凸规划进行重新表述，并使用近端块坐标下降方法求解。这类方法每次迭代只更新决策变量中的少数几个坐标。借助这些方法收敛性分析的最新进展，以及均衡捕获凸规划的结构，我们证明了这些方法具有快速收敛率。

1 引言

在市场均衡（ME）中，通过为物品设定一组价格，并将物品分配给买方，使得每位买方都能最优地使用其预算，且所有物品均被完全分配。由于具有丰富的结构性质，并能提供强有力的公平性和效率保证，ME 长期以来一直用于开发公平分配和在线资源分配机制（Gao 和 Kroer 2021；Aziz 和 Ye 2014；Barman、Krishnamurthy 和 Vaish 2018；Arnsperger 1994）。市场模型及相应的均衡计算算法一直是市场设计以及经济学、计算机科学和运筹学相关领域的核心研究课题，并产生了实际影响（Scarf 等人 1967；Kantorovich 1975；Othman、Sandholm 和 Budish 2010；Daskalakis、Goldberg 和 Papadimitriou 2009；Cole 等人 2017；Kroer 等人 2019）。更具体地说，市场设计中的许多研究都依赖于这样一种假设：对于大型市场实例，能够高效地计算 ME。例如，著名的无货币公平分配机制等收入竞争均衡（CEEI）要求在统一买方预算下计算 Fisher市场的 ME（Varian 等人 1974）。近期研究还揭示了市场均衡与大型互联网市场背景下的重要解概念之间的紧密联系，例如重复拍卖中的节奏控制均衡（Conitzer 等人 2018, 2019；Kroer 和 Stier-Moses 2022）。受到上述经典及新兴应用的启发，我们希望为大型市场实例开发高效的均衡计算算法。一般而言，计算 ME 是一个困难问题（Chen 和 Teng 2009；Vazirani 和 Yannakakis 2011；Othman、Papadimitriou 和 Rubinstein 2016）。然而，对于 Fisher市场及某些类别的效用函数，已有高效算法（Devanur 等人 2008；Zhang 2011；Gao 和 Kroer（2020））；这些算法通常通过优化算法求解特定的凸规划，而该凸规划的解与 ME 互为对应。本文聚焦于两种著名的此类凸规划，即 Eisenberg-Gale（EG）凸规划（Eisenberg and Gale 1959；Eisenberg 1961）和 Shmyrev 凸规划（Shmyrev 2009；Cole 等人 2017）。

现有的大多数均衡计算文献研究的是静态市场，即在均衡计算过程中，每个时间步都有所有买方和物品参与。与此不同，我们研究的是每个时间步仅有随机子集的买方-物品对出现的情形。从计算角度看，这种设定有充分的动机，因为随机方法通常更适用于超大规模问题。其次，我们的模型能够刻画新类型的市场动态。我们利用随机一阶优化领域的最新进展，更具体地说，采用分块坐标类方法，为这一设定设计新的均衡计算算法。由此得到的均衡计算算法具有强有力的收敛保证，并且在数值实验中持续优于确定性的全信息算法。此外，许多优化步骤不仅能为市场迭代量提供高效的更新公式，还能转化为可解释的市场动态。

贡献概述。 我们提出了两种用于计算大规模市场均衡（ME）的随机算法：基于 *EG* 的（近端）分块坐标下降（BCDEG）和分块坐标比例响应 （BCPR）。这些算法通过在重新表述的均衡捕获凸规划上应用随机分块坐标类算法推导而来。具体而言，BCDEG 基于（近端）随机分块坐标下降（BCD），而 BCPR 基于 BCD 的非欧几里得（Bregman）版本。我们证明，这些

算法具有良好的理论收敛保证，并讨论了在实践中实现高效运行的重要细节。此外，我们证明，BCDEG 中的欧几里得投影到单纯形（以及其他已有的投影梯度型方法中的相应投影）具有试探式解释。随后，我们通过在合成与真实市场实例上开展广泛的数值实验，展示了我们算法的实际效率；结果发现，与比例响应动态等现有最先进的方法相比，我们的算法显著更快。

预备知识与符号约定。 除非另有说明，我们考虑一个线性费希尔市场，其中有 n 名买方 和 m 件物品。我们用 $i \in [n]$ 表示买方，用 $j \in [m]$ 表示物品。每位买方 i 的预算 为 $B_i > 0$，每件物品的供给均为一。买方 i 的分配 （或物品组合）是一个向量 $x_i \in \mathbb{R}_+^m$，用于指定买方 i 获得每件物品 j 的数量。给定 x_i，买方 i 获得的效用为 $\langle v_i, x_i\rangle = \sum_j v_{ij}x_{ij}$，其中 $v_i \in \mathbb{R}_+^m$ 是其估值向量。给定所有物品的价格 $p \in \mathbb{R}_+^m$（即物品 j 的价格为 p_j），买方 i 为 x_i支付 $\langle p, x_i\rangle = \sum_j p_j x_{ij}$。给定价格 p 和预算 B_i，若物品组合 x_i 满足 $\langle p, x_i\rangle \le B_i$，则其对买方 i 而言是预算可行的。我们还用 $x_{\cdot j} \in \mathbb{R}_+^n$ 表示一个向量，其中包含分配给所有买方的物品 j数量。买方 i 的需求集 是预算可行且效用最大化的分配所构成的集合：

$$\mathcal{D}_i(p) = \arg\max_{x_i}\{u_i = \langle v_i, x_i\rangle : \langle p, x_i\rangle \le B_i\}. \qquad \text{(D)}$$

市场均衡（ME）是一个分配-价格对（x^*，p^*），满足对于所有 i，$x_i^* \in \mathcal{D}_i(p^*)$；并且对于所有 j，$\sum_i x_{ij}^* \le 1$，若 $p_j^* > 0$ 则取等号。

2 相关工作

自 Eisenberg 和 Gale 的开创性研究（Eisenberg and Gale 1959；Eisenberg 1961）以来，关于费希尔市场均衡计算的研究已十分丰富，其中许多工作基于市场均衡的凸优化刻画（Devanur 等人 2008；Zhang 2011；Birnbaum、Devanur 和 Xiao 2011；Cole 等人 2017；Gao 和 Kroer（2020）；Garg 和 Kapoor2006）。针对更一般市场模型的均衡计算算法也得到了广泛研究，这些模型包含额外约束，例如不可分物品以及对每位买方可选物品组合的限制（Othman、Sandholm 和 Budish 2010；Budish 等人 2016）；此类算法通常基于近似算法、混合整数规划建模以及局部搜索启发式方法。

在 Gao 和 Kroer（2020）中，作者考虑了三种（确定性）一阶优化方法，即投影梯度法（PG）、Frank-Wolfe（FW）和镜像下降法（MD），用于求解刻画市场均衡的凸规划。据我们所知，目前尚无关于市场均衡块坐标方法的研究结果。在优化文献中，关于新型块坐标算法及其分析的研究已经十分丰富，并且仍在持续发展（例如，Tseng（2001）；Wright（2015）；Beck 和 Tetruashvili（2013）；Hanzely 和 Richtárik（2019、2021）；Liu 和 Wright（2015）；Nesterov（2012）；Richtárik 和 Taká˘c（2014）；Gao 等人（2020）；Attouch、Bolte 和 Svaiter（2013）；Zhang（2020）；Reddi 等人（2016））。如前所述，我们的 BCDEG 算法基于应用于 EG 的近端块坐标下降（PBCD）。在不同的误差界条件下，已经证明，对于非光滑有限和复合优化问题（其目标函数形式为 $F(x) = \sum_i f_i(x) + \psi(x)$），PBCD 最后一次迭代的均方误差呈线性收敛（Richtárik 和 Taká˘c 2014；Karimi、Nutini 和 Schmidt 2016；Reddi 等人 2016）。对于 BCPR，我们采用近期提出的非欧几里得（Bregman）PBCD 的分析方法（Gao 等人 2020）；该方法又借鉴了 Bauschke、Bolte 和 Teboulle（2017）发展的收敛理论。

3 EG 程序的块坐标下降算法

（近端）块坐标下降（BCD）方法常用于求解目标函数由光滑部分和（可能的）非光滑部分构成的问题。后者通常按块坐标可分。BCD 算法在每次迭代中只更新一个较小的坐标块，因此每次迭代的成本远低于确定性方法，从而能够扩展到规模非常大的实例。刻画市场均衡的 EG 凸规划可以写成适用于近端块坐标下降方法的形式。对于效用函数为线性的买方，EG 凸规划为：

$$\begin{aligned} \max_{x} \quad & \sum\nolimits_i B_i \log u_i \\ \text{s.t.} \quad & u_i \le \langle v_i, x_i\rangle \quad \forall\, i \\ & \sum\nolimits_i x_{ij} \le 1 \quad \forall\, j \\ & x \ge 0. \end{aligned} \qquad \text{(EG)}$$

(EG) 的任意一个最优解 x^*，以及与约束$\sum_i x_{ij} \le 1$、$j \in [m]$ 相关联的（唯一）最优拉格朗日乘子 $p^* \in \mathbb{R}_+^m$，构成一个市场均衡。事实上，当效用函数为凹、连续、非负且齐次度为 1 时，这一结论更一般地成立。

要应用块坐标下降，请注意，(EG) 具有以下形式：

$$\min_{x\in\mathbb{R}^{m\times n}} F(x) := f(x) + \psi(x) = f(x) + \sum\nolimits_j \psi_j(x_{\cdot j}) \quad (1)$$

其中 $f(x) = -\sum_i B_i \log \langle v_i, x_i\rangle$，并且 $\psi_j(x_{\cdot j}) = 0$ 当 $\sum_i x_{ij} \le 1$，$x_{\cdot j} \ge 0$ 否则。

因此，非光滑项 $\psi(x)$ 可按物品 j 分解，因此我们可以将 x 视为一组块：每个物品 j 对应一个分配变量块，表示该物品 j 分配给各买方的数量。我们用$x_{\cdot j} = (x_{1j}, \ldots, x_{nj})$ 表示第 j 个块。已知 f 在 x 处的全梯度为 $\nabla f(x)$，我们还需要第 j 个块 $x_{\cdot j}$ 对应的偏梯度，记为 $\nabla_{\cdot j} f(x)$，即 $\nabla_{\cdot j} f(x) = (\nabla_{1j} f(x) \ldots, \nabla_{nj} f(x))$。

在块坐标下降法的每次迭代中，我们首先随机选择一个索引，并确定相应的步长。然后，若满足条件，则通过相应的近端映射生成下一个迭代点；否则，按另一种方式生成，其中该映射等于

$$\arg\min_{y_{\cdot j}} \langle \nabla_{\cdot j} f(x), y_{\cdot j} - x_{\cdot j}\rangle + \frac{1}{\eta_j}\|y_{\cdot j} - x_{\cdot j}\|^2 + \psi_j(y_{\cdot j}). \quad (2)$$

上述近端映射等价于

$$T_j(x) = \text{Proj}_{\Delta_n}\Big(x_{\cdot j} - \eta_{j'}\nabla_{\cdot j} f(x)\Big), \tag{3}$$

因此，我们可以通过对 n 维单纯形进行一次投影来生成 $T_j(x) \in \mathbb{R}^n$。由于投影是计算成本最高的部分，我们的方法最终比完整的投影梯度下降便宜 m 倍。

为确保 $\nabla_{\cdot j} f(x)$ 存在，我们需要在每次迭代中将买家效用保持在远离零的位置。为此，令 $u_i = \langle v_i, (B_i / \sum_{l=1}^{n} B_l)\vec{1}\rangle$ 表示比例分配的效用。然后，我们执行"二次外推"，将目标函数 $f(x)$ 替换为函数 $\tilde{f}(x) = \sum_i f_i(x_i) = \sum_i \tilde{g}_i(u_i)$，其中 $\tilde{g}_i(u_i) = -B_i \log u_i$（如果 $u_i \geq \underline{u}_i$），而 $\log \underline{u}_i - \frac{B_i}{\underline{u}_i}(u_i - \underline{u}_i) + \frac{B_i}{2\underline{u}_i^2}(u_i - \underline{u}_i)^2$ 其他情况。对于（EG），将 f 替换为 $\tilde{f}$ 不会影响任何最优解（Gao 和 Kroer（2020），引理 1）。

我们还需要以下 Lipschitz 界，它确保迭代点执行的是下降步，即只要步长不过大，期望目标值就不会增加。对于特定块，确保迭代点下降的允许步长上界由关于该块坐标的 Lipschitz 常数决定。更多细节和证明请参见附录 A。

引理 1。对于任意 $j \in [m]$，令

$$L_j = \max_{i \in [n]} \frac{B_i v_{ij}^2}{\underline{u}_i^2}. \tag{4}$$

那么，对于所有满足 x、且仅在第 j 个块上有所差异的 x、$y \in \mathcal{X}$，我们有

$$\tilde{f}(y) \leq \tilde{f}(x) + \langle \nabla_{\cdot j}\tilde{f}(x), y - x\rangle + \frac{L_j}{2}\|y - x\|^2. \tag{5}$$

令 L 为决定（全）梯度下降步长的"全局" Lipschitz 常数。令 $L_{\max} = \max_j L_j$。通常，很容易看出 $1 \leq L/L_{\max} \leq n$，但利用梯度 $(\nabla f)_{ij} = B_i v_{ij}/u_i$，我们可以得到比这更强的界。当块 j 中仅有变量 $x_{\cdot j}$ 发生变化时，L_j 受限于第 j 个块（子）海森矩阵的最大对角值。相比之下，对于 L，它取决于所有买方 i 中第 i 个块（子）海森矩阵的最大迹。从市场的角度来看，这可以解释如下：当我们只调整一个块时，每个买方的效用会根据该物品发生波动。相反，对于全梯度方法，每件物品都会对 u_i 的变化产生贡献。这意味着 L_j/L 的比值大致为 $\max_i v_{ij}^2 / \max_i \|v_i\|^2$。

算法1给出了 EG 凸规划的（近端）块坐标下降算法。每次迭代只需在一个 n 维单纯形上进行一次投影，而完整的投影梯度方法则需要进行 m 次投影。此外，我们在附录A中证明，对于线性效用，当估值矩阵 v_{ij} 稀疏时，还可以进一步降低算法1每次迭代的计算成本。

算法1： BCDEG（近端）块坐标 EG 程序的下降法

输入： 初始 x^0，步长 $\eta_1^k, \eta_2^k, \cdots, \eta_m^k, \forall k \in \mathbb{N}$

对于 $k \leftarrow 1, 2, \cdots$ **执行**

- 以概率 $1/m$ 选择 $j_k \in [m]$；
- $g^{k-1} \leftarrow \nabla_{\cdot j_k}\tilde{f}(x^{k-1})$；
- $x_{\cdot j_k}^k \leftarrow$ 近端算子$_{\Delta_n}\left(x_{\cdot j_k}^{k-1} - \eta_{j_k}^k g^{k-1}\right)$；
- $x_{\cdot j}^k \leftarrow x_{\cdot j}^{k-1}$，$\forall j \neq j_k$；

线搜索策略。正如 Richtárik 和 Takáč（2014）等人在先前工作中所提到的，线搜索通常对块坐标下降非常有帮助。如果能够以较低成本执行线搜索，就可以大幅加快数值收敛。我们将在后文证明，这一点在我们的设置中确实成立。

我们在算法1中引入线搜索，并采用（逐坐标的）l_2 光滑性条件。线搜索对算法1的修改如下：计算出 $x_{\cdot j_k}^k$ 后，我们检查是否

$$\eta_{j_k}\|\nabla_{\cdot j_k}\tilde{f}(x^k) - \nabla_{\cdot j_k}\tilde{f}(x^{k-1})\| \leq \|x_{\cdot j_k}^k - x_{\cdot j_k}^{k-1}\|.$$

如果此检验通过，我们就将步长乘以一个较小的因子，然后进入下一次迭代。如果检验未通过，则将步长乘以一个较小的因子使其减小，并重新计算 $x_{\cdot j_k}^k$。该线搜索算法每次迭代的实现成本为 $O(n)$，而全梯度法则需要 $O(nm)$ 的时间。附录 A 给出了带线搜索的 BCDEG（BCDEG-LS）的完整说明。

收敛性分析。接下来，我们将证明，在步长足够大且固定的情况下，算法1具有线性收敛性，并证明线搜索变体也具有线性收敛性。

沿袭关于满足"松弛强凸性"条件的结构化凸优化问题中一阶方法线性收敛性的既有研究，我们证明 BCDEG 生成的迭代点的期望目标值线性收敛。关于确保一阶方法线性收敛的这些松弛充分条件的更多详情，请参见 Karimi、Nutini 和 Schmidt（2016）及其中的参考文献。

Gao 和 Kroer（2020）证明了，(EG) 及其他均衡捕获凸规划可以重新表述为满足这些条件的形式。因此，在这些凸规划上运行一阶方法，可以得到线性收敛的均衡计算算法。类似于 Gao 和 Kroer（2020，定理2）以及 Karimi、Nutini 和 Schmidt（2016，(39)）的证明，我们首先建立一个近端算子-*PŁ*不等式。

引理 2。对于任意可行的 x 和任意 $L > 0$，定义 $\mathcal{D}_\psi(x, L)$ 为

$$-2L\min_y \langle \nabla\tilde{f}(x), y - x\rangle + \frac{L}{2}\|y - x\|^2 + \psi(y) - \psi(x)$$

于是我们得到如下不等式

$$\frac{1}{2}\mathcal{D}_\psi(x, L) \geq \min\left\{\frac{\mu}{\theta^2(A, C)}, L\right\}(F(x) - F^*) \tag{6}$$

其中， $\theta(A, C)$ 是(EG)的最优解多面体集合的霍夫曼常数，该集合由矩阵 A 和 C 刻画，其中 C 是一个用于刻画最优性条件的矩阵，而 F^* 是最优值。

在前述不等式中，他们实质上证明了，对于任意 $L \geq \lambda = \mu/\theta^2(A, C) > 0$， $\frac{1}{2}\mathcal{D}_\psi(x, L) \geq \lambda(F(x) - F^*)$。然而，我们在此推导出一个近端 PŁ 不等式，为任意 $L > 0$ 给出 $\frac{1}{2}\mathcal{D}_\psi(x, L)$ 的下界。

接着，结合引理 2 与 Richtárik 和 Taká˘c（2014，引理 2 和 3），我们可以证明 BCDEG 的主要收敛定理。

定理 1. 给定初始迭代点 x^0 和步长 $\eta_j^0 = 1/L_j$，其中 $\forall\, j$ 满足 (4)，令 x^k 为算法*1*生成的随机迭代点。那么，

$$\mathbf{E}\left[F(x^{k+1})\right] - F^* \leq (1-\rho)^k \left(F(x^0) - F^*\right), \quad (7)$$

where $\rho = \min\left\{\frac{\mu}{mL_{\max}\theta^2(A,C)}, \frac{1}{m}\right\}$ *and* $L_{\max} = \max_j L_j$.

Karimi、Nutini 和 Schmidt（2016）也提出了一种应用于 EG 的块坐标方法。与他们的结果不同，我们的方法允许采用更大的步长，且步长可以因块而异；此外，还支持线搜索策略。正如我们在第 8 节所示，这有助于提升实际性能。

4 投影梯度步的经济学解释

正如 Goktas、Viqueira 和 Greenwald（2021）所指出的，通过 BCDEG 等投影梯度法计算市场均衡的一个缺点是，这些方法无法自然地解释为市场动态。为弥补这一不足，本节将展示：算法1，以及更一般的投影梯度下降，可以解释为平衡供给与需求的分布式定价动态。

对于单个买方i 和选定的物品 j，算法1中的投影步骤如下（其中我们用 j 表示 j_k，并为简洁起见省略时间索引）：

$$x_{\cdot j}^k \leftarrow \text{Proj}_{\Delta^n}\left(x_{\cdot j}^{k-1} - \eta_j g^{k-1}\right). \quad (8)$$

众所周知，将向量 $y \in \mathbb{R}^n$ 投影到单纯形 $\Delta^n = \left\{x \in \mathbb{R}_+^n : \sum_i x_i = 1\right\}$，可以使用 O（n对数 n）算法求得（我们所知最早的发现来自 Held、Wolfe 和 Crowder（1974）；关于单纯形投影的较新研究，参见附录 B）。关键步骤是找到满足以下条件的唯一数 t：

$$\sum\nolimits_i (y_i - t)^+ = 1 \quad (9)$$

并按分量计算解 $x = (y - t \cdot \mathbf{1})^+$。如果 y_i 已排序，就可以用一个简单的单遍算法完成这一计算。事实上，数值 t 对应于 KKT 条件中约束$\sum_i x_i = 1$ 的（唯一）最优拉格朗日乘子。

在算法1的投影步骤中，(9) 的形式为

$$\sum\nolimits_i (x_{ij}^{k-1} - \eta_j g_i^{k-1} - t)^+ = 1. \quad (10)$$

回顾一下，我们有 $g^{k-1} < 0$。注意，(10) 的左边关于 t 非增，并且在解附近严格递减（因为左边必须有一些项为正，才能使总和为 1）。此外，令 $t = 0$ 可得下界 $\sum_i (x_{ij}^{k-1} - \eta_j g_i^{k-1} - t)^+ > \sum_i x_{ij}^{k-1} = 1$。因此，唯一解 t^* 必为正。现在，我们将 t 改写为$\eta_j p_j$，其中 p_j为某个“价格”。于是，(10) 可写为

$$\sum\nolimits_i D_i^k(p_j) = 1,\ D_i^k(p_j) = (x_{ij}^{k-1} - \eta_j g_i^{k-1} - \eta_j p_j)^+. \quad (11)$$

换言之，投影步骤等价于求解满足式 (11) 的p_j^k。这里，D_i^k可以看作买方 i 在时刻 k 给定先前分配x_{ij}^{k-1} 时的线性需求函数。解 p_j^k可以视为市场出清价格，因为 $\sum_i D_i^k(p_j^k)$恰好等于物品 j的单位供给。设定价格后，就可以像单纯形投影算法中那样轻松计算更新后的分配，即对所有 i， $x_{ij}^k = \left(x_{ij}^{k-1} - \eta_j g^{k-1} - \eta_j p_j^k\right)^+$。等价地，新的分配由当前线性需求函数给出： $x_{ij}^k = D_i^k(p_j^k)$。

综上所述，我们可以将算法1重述为以下动态定价步骤。在每个时刻 k，都会发生以下事件。

- 均匀随机地抽取一个物品 j。
- 对于每位买方 i，其需求函数变为$D_i^k(p_j) \leftarrow (x_{ij}^{k-1} - \eta_j g^{k-1} - \eta_j p_j)^+$。
- 求出满足 $\sum_i D_i^k(p_j^k) = 1$的唯一价格 p_j^k。
- 每位买方通过 $x_{ij}^k = D_i^k(p_j^k)$ 选择其对物品 j 的新分配。

为了直观理解线性需求函数，请注意，当 $u_i \geq u_i$时，我们有 $g_i^{k-1} = = B_i v_{ij}/u_i^{k-1}$，因此 $D_i^k\left(B_i v_{ij}/u_i^{k-1}\right) = x_{ij}^{k-1}$ 成立。换言之，如果物品 j 的价格为 $(B_i/u_i^{k-1})v_{ij}$，那么买方 i 当前的需求恰好等于上一轮的分配 x_{ij}^{k-1}。这可以用求解 EG 时熟悉的方式来解释：设 $\beta_i^{k-1} = B_i/u_i^{k-1}$ 为时刻 $k-1$买方 i的效用价格，那么在看到价格 p_j^k后，买方 i 会增加对价格高于其当前效用价格的商品的分配，并减少对价格低于其当前效用价格的商品的分配。步长 η_j 表示买方 i对价格变化的响应程度。

价格的设定方式使供给与需求相等，这令人联想到试探式价格设定。不同之处在于，买方会逐渐适应变化的环境，而价格的设定则旨在根据当前的买方需求函数实现精确的市场出清。

5 比例响应动态的相对块坐标下降算法

Birnbaum、Devanur 和 Xiao（2011）指出，针对线性买方效用的比例响应动态（PR），可以通过对以买方出价表述的 Shmyrev 凸规划应用带 KL 散度的镜像下降（MD）推导出来

（Shmyrev 2009；Cole 等人 2017）。作者利用 Shmyrev 凸规划的“相对光滑性”条件，推导出了 PR（MD）的 $O(1/k)$ 末次迭代收敛保证。此后，这一结果得到进一步推广，并促成了适用于更一般的相对光滑问题、速度更快的镜像下降类算法（Hanzely 和 Richtárik 2021；Lu、Freund 和 Nesterov 2018；Gao 等人 2020）。本节提出 PR 动态的一种随机化扩展，并称之为分块坐标比例响应（BCPR）。BCPR 基于近期提出的一种随机镜像下降算法（Gao 等人 2020）。我们给出步长规则，并证明每次迭代只涉及单个买方，且可在 $O(m)$ 时间内完成。

令 $b \in \mathbb{R}^{n\times m}$ 表示所有买方对所有物品的出价矩阵，$p_j(b) := \sum_i b_{ij}$ 表示给定出价 b 时物品 j 的价格。若 $v_{ij} > 0$，则记 $a_{ij} = \log v_{ij}$；否则记为 0。Shmyrev 凸规划为

$$\begin{aligned} \max_{b} \quad & \sum\nolimits_{i,j} a_{ij} b_{ij} - \sum\nolimits_j p_j(b) \log p_j(b) \\ \text{s.t.} \quad & \sum\nolimits_j b_{ij} = B_i \quad \forall i \\ & b \geq 0. \end{aligned} \tag{S}$$

已知，(S) 的一个最优解 $(b^*,\ p^*)$ 给出均衡价格 p^*_j。对于所有 $i,\ j$，可通过 $x^*_{ij} = b^*_{ij}/p^*_j$ 构造相应的均衡分配。(S) 可改写为一个光滑有限和凸函数与一个可能非光滑的凸可分正则项之和的最小化问题：

$$\min_{b\in\mathbb{R}^{m\times n}} \Phi(b) := \varphi(b) + r(b) = \varphi(b) + \sum\nolimits_i r_i(b_i) \tag{12}$$

其中 $\varphi(b) = -\sum_{i,\ j} b_{ij} \log\left(\frac{v_{ij}}{p_j(b)}\right)$，且 $r_i(b_i) = 0$，当 $\sum_j b_{ij} = B_i,\quad b_i \geq 0$ 时；否则为 $+\infty$。

现在，我们介绍用于（12）的相对随机分块坐标下降（RBCD）方法。在对 b 进行近端更新时，我们使用 KL 散度作为 Bregman 距离。令 $D_{\mathrm{KL}}(q^1, q^2) = \sum_i q^1_i \log(q^1_i/q^2_i)$ 表示 q^1 与 q^2 之间的 KL 散度（假设 $\sum_i q^1_i = \sum_i q^2_i$ 和 $q^1_i, q^2_i > 0,\ \forall i$）。在每次迭代中，给定当前的 b，我们均匀随机地选择 $i \in [n]$，并且仅更新 i-th 块坐标 b_i。下一个迭代点 b^+ 对于 i 为 $b^+_i = T_i(b_i)$，对于其余块则为 $b^+ = b$。这里，$T_i(b_i)$ 等于

$$\arg\min_a \langle \nabla_i \varphi(b), a - b_i\rangle + \frac{1}{\alpha_i} D_{\mathrm{KL}}(a, b_i) + r_i(a) \tag{13}$$

其中，$\alpha_i > 0$ 为步长。众所周知，（13）等价于以下简单的显式更新公式：

$$b^+_{ij} = \frac{1}{Z_i} b_{ij} \left(\frac{v_{ij}}{p_j}\right)^{\alpha_i} \quad \forall j = 1, 2, \ldots, m \tag{14}$$

其中 Z_i 是一个归一化常数，使得 $\sum_j b^+_{ij} = B_i$。

算法2给出了完整的分块坐标比例响应动态。

收敛性分析。 (S) 的目标函数相对于 $L = 1$ 是光滑的（Birnbaum、Devanur 和 Xiao 2011，引理7）。这意味着 $\alpha_i = 1$ 是安全的

算法2： 分块坐标比例响应（BCPR）

输入： 初始 b^0, p^0，步长 $\alpha^k_1, \ldots, \alpha^k_n,\ \forall k \in \mathbb{N}$

对于 $k \leftarrow 1, 2, \ldots$ **执行**

- 以概率 $1/n$ 选取 $i_k \in [n]$；
- 根据式（14）以步长 $\alpha^k_{i_k}$ 计算 $b^+_{i_k}$；
- $b^k_{i_k} \leftarrow b^+_{i_k}$ 和 $b^k_{i'} \leftarrow b^{k-1}_{i'},\quad \forall i' \neq i_k$;
- $p^k_j \leftarrow \sum_i b^k_{ij},\quad \forall j$;

步长下界。对于算法2，Gao 等人（2020）给出了最后迭代次线性收敛速率，其中 θ_i 是 Bauschke、Bolte 和 Teboulle（2017）引入的 Bregman对称性度量。对于 KL 散度 $\theta_i = 0$。他们的证明对于 $\alpha_i = 1/L_i$ 仍然成立，由此得到以下结果。

定理2。 设 b^k 为算法2生成的随机迭代点，其中对所有 k 均有 $\alpha^k_i = 1/L_i$，则

$$\mathbf{E}\big[\Phi(b^k)\big] - \Phi^* \leq \frac{n}{n+k}\Big(\Phi^* - \Phi(b^0) + D(b^*, b^0)\Big) \tag{15}$$

其中 Φ^* 为最优目标值。

线搜索。 为加快 BCPR，我们引入线搜索策略和自适应步长策略。带线搜索的BCPR可以通过比较 $D_{\mathrm{KL}}(p^+,\ p)$ 和 $D_{\mathrm{KL}}(b^+_i,\ b_i)$ 来实现，耗时 $O(m)$，远低于计算 (S) 的完整目标函数值所需的时间。此外，通过存储 p，我们还避免了在每次迭代中访问所有变量。因此，每次迭代所需访问的数据量和计算量为 $O(m)$（而全梯度方法需要 $O(nm)$）。在自适应策略中，我们利用 Lipschitz估计和闭式公式计算更大的步长。带线搜索的BCPR（BCPR-LS）和自适应 BCPR（A-BCPR）在附录C中正式给出。

与 BCDEG 一样，我们的实验表明，较大的步长可以加速算法2。当我们考虑由线搜索或自适应策略生成的一系列步长 $\{\alpha^k_i\}_{k\in\mathbb{N}}$ 时，可以证明以下不等式

$$\begin{aligned} &\mathbf{E}\big[\Phi(b^k) - \Phi(b^*)\big] \\ &\leq \frac{n}{n+k}\left(\Phi(b^*) - \Phi(b^0) + \sum_i \frac{1}{\alpha^0_i} D(b^*_i, b^0)\right) \\ &+ \frac{1}{k}\sum_{l=1}^{k} \mathbf{E}\left[\sum_i \frac{1}{\alpha^l_i} D(b^*_i, b^l) - \sum_i \frac{1}{\alpha^{l-1}_i} D(b^*_i, b^l)\right]. \end{aligned} \tag{16}$$

然而，我们无法保证收敛，因为无法确保上述最后一项收敛。

6 带线搜索的比例响应

在本节中，我们通过开发带线搜索的镜像下降（MDLS）算法，将原始PR动力学扩展为带线搜索的比例响应动态（PRLS）。直观地说，线搜索策略基于反复增大步长并

检查相对光滑条件；当条件不满足时则减小步长。这与 Gao 和 Kroer （2020，A.5）中的带线搜索的投影梯度法（PGLS）类似，但用 Bregman 散度替代了 ℓ_2 范数。在保证相同的次线性末次迭代收敛的同时，该方法允许使用更大的步长。一般的 MDLS 算法见附录 D。

收敛速度。 Birnbaum、Devanur 和 Xiao（2011，定理 3）证明，采用固定步长 $\alpha^k = 1/L$ 可确保以 $1/k$ 速度次线性收敛。建立该收敛速度的关键步骤之一是下降引理，该引理在线搜索情形下同样成立：

引理 3 b^k $D\ b^*\ b^{k+1}$ $\leq$。设为 $D\ b^*\ b^k$ k *MDLS* 迭代点。那么，对于所有，(，)(，)。

我们有如下定理。

定理 3。 设 b^k 为从任意可行的初始解 b^0出发，由 *PRLS* 生成的迭代点，则有

$$\varphi(b^k) - \varphi^* \leq \frac{1}{\rho^-} \cdot \frac{D(b^*, b^0)}{k} \tag{17}$$

其中 ρ^- 为步长的收缩因子。

与 BCPR-LS 中的结果不同，确定性算法在线搜索下仍能保持收敛保证。块坐标情形的主要问题在于 $\mathbf{E}[D_i(b^*, b^k)]$, 缺乏单调性，而确定性算法避免了这一问题。在证明中，我们还给出了一个更紧的、依赖于路径的界。

7 CES 效用函数的块坐标下降算法

本节将说明，对于买方具有常替代弹性（CES）效用函数的情形（对于 $\rho \in (0,1)$），块坐标下降算法同样适用。

形式上，对于参数为 $v_i \in \mathbb{R}_+^m$ 和 $\rho \in (0,1)$ 的买方 i，其 CES 效用函数为 $u_i(x_i) = (\sum_j v_{ij} x_{ij}^\rho)^{\frac{1}{\rho}}$，其中 v_{ij} 表示买方i对物品 j 的单位估值。CES 效用函数是凸的、连续的、非负且齐次的，因此，所得市场均衡仍可由 EG 凸规划刻画。

CES 效用函数的 BCDEG。所得 EG 程序与式 (1) 形式相似，其中 $f(x) = -\sum_i \frac{B_i}{\rho} \log \langle v_i, x_i^\rho \rangle$，且 $x_i^\rho = (x_{i1}^\rho, \ldots, x_{im}^\rho)$ 与式 (1) 中相同，均为可分的 $\psi(x)$。因此，与线性 Fisher 市场一样，我们可以应用分块坐标下降。

由于 u和 x 可能在某些迭代点处达到 0，我们需要采用平滑外推技术来确保梯度存在。首先，与 Zhang（2011，引理 8）类似，我们对所有 i，j 给出 x^* 的下界：$v_{ij} > 0$，这确保当 $\min_{i,\ j} v_{ij} > 0$时，外推不会影响均衡。我们的界比 Zhang（2011）的更紧。

引理 4. 对于具有 *CES* 效用函数、 $\rho \in (0,1)$ 且市场均衡分配为 x^* 的市场，对于任意满足 $v_{ij} > 0$ 的 i、j，有 $x_{ij}^* \geq x_{ij}^* := \omega_1(i)^{\frac{1}{1-\rho}} \omega_2(i)^{\frac{\rho(\rho+1)}{1-\rho}}$，其中 $\omega_1(i) = \frac{B_i}{m \sum_l B_l}$，$\omega_2(i) =$最小值$_{j:v_{ij}>0}\, v_{ij}$/ 最大值$_j\, v_{ij}$。

在附录E中，我们展示如何利用这一界进行安全外推。由此得到以下关于将 BCDEG 应用于 CES 效用的定理。由于它与线性情形相似，我们在附录中给出完整算法。该定理是引理 4 和 Richtárik 和 Takáˇc（2014，定理7）的直接推论。

定理4。 设 x^k 为 *BCDEG* 针对 *CES* 效用函数（$\rho \in (0,1)$）生成的随机迭代点，步长为 $\eta_j^k = 1/L_j\ \forall\, k$，其中

$$L_j = \max_{i \in [n]} \frac{B_i v_{ij} \underline{x}_{ij}^{\rho-2}}{\underline{u}_i(\rho)} \quad \textit{for all } j \in [m] \tag{18}$$

且 $\underline{u}_i(\rho) = \sum_j v_{ij} \underline{x}_{ij}^{*\rho}$（定义见引理 *4*）。那么，

$$\mathbf{E}\big[F(x^k)\big] - F^* \leq \left(1 - \frac{\mu(L)}{m}\right)^k \big(F(x^0) - F^*\big) \tag{19}$$

其中 $\mu(L)$ 是关于加权范数 $\sum_j L_j \|\cdot\|_{\cdot j}^2$的强凸模量。

CES 效用函数下的 BCPR。与线性效用不同，CES 效用的 EG 程序无法转化为简单的对偶问题。因此，我们无法将 $\rho \in (0,\ 1)$ 下的 PR 视为镜像下降算法，也无法用典型的相对光滑性技术对其进行分析。然而，Zhang（2011）仍证明了 $\rho \in (0,\ 1)$ 下 PR 的收敛性。我们证明，仍可将他们的证明推广，从而证明 CES 效用下块坐标 PR 的收敛性。

定理5。设 b^k 为 *CES* 效用函数 *(*$\rho \in (0,\ 1)$*)* 下由 *BCPR* 生成的随机迭代点。对于任意 $\epsilon > 0$，当

$$k \geq \frac{2 \log \frac{\sqrt{8D(b^*, b^0)} W^{\frac{1}{1-\rho}}}{\epsilon}}{\log \frac{n}{n-1+\rho}}, W = \frac{n}{\min_{v_{ij}>0} v_{ij} \cdot \min_i B_i}, \tag{20}$$

we have $\mathbf{E}\left[\frac{|b_{ij} - b_{ij}^*|}{b_{ij}^*}\right] \leq \epsilon$ *for all* i, j *such that* $v_{ij} > 0$.

8 数值实验

我们基于模拟数据（线性效用和 CES ($\rho \in (0,\ 1)$) 效用）以及真实数据进行了数值实验，以测试我们算法的可扩展性。

为衡量所执行的工作量，我们统计对估值矩阵中各单元格的访问次数。对于确定性算法，每次迭代的成本为 $n \times m$；对于我们的 BCDEG 算法，每次迭代的成本为 n；对于我们的 BCPR 算法，每次迭代的成本为 m。对于采用线搜索的算法，我们还统计执行线搜索所需的估值访问次数。

为衡量解的准确性，我们使用 u 与 u^*之间的对偶间隙和平均相对差异。实例规模足够小，因此我们可以使用 Mosek（2010）计算均衡效用 u^*。

模拟低秩实例。为了模拟市场实例，我们生成一组估值，以近似模拟低秩估值。这类估值在现实市场中普遍存在，且 Kroer 等人（2019）曾在市场均衡的背景下对其进行研究。物品 j 的估值以及

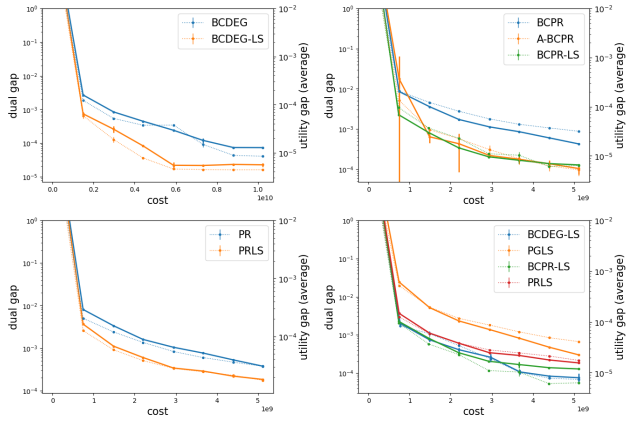

图1：模拟低秩实例上的性能。随机算法使用种子 $0-9$ 实现。我们还绘制了垂直误差线，表示不同种子下的标准差。左侧 y 轴显示以对偶间隙衡量的性能（每种算法用实线表示），右侧 y 轴显示以效用衡量的性能（每种算法用虚线表示）。x 轴显示执行的工作量。

买方 i 的估值生成方式如下：$v_{ij} = v_i v_j + \epsilon_{ij}$，其中 $v_i \sim \mathcal{N}(1, 1)$，$v_j \sim \mathcal{N}(1, 1)$，且 $\epsilon_{ij} \sim$ 服从均匀（0，1）分布。这里，买方 i 对物品 j 的估值由三部分组成：物品 j 本身的价值（v_j）、买方 i 的平均估值（v_i）以及随机项 ϵ_{ij}。我们考虑满足 $n = m = 400$ 的市场。所有预算和供给均等于 1。

电影评分实例。我们使用从 Twitter 收集的一个名为 *Movietweetings* 的电影评分数据集，生成一个市场实例（Dooms、De Pessemier 和 Martens（2013））。在此，用户被视为买方，电影被视为物品，评分被视为估值。每位买方的预算设为一个单位，每件物品的供给设为一个单位。我们使用“snapshots 200K”数据集，并移除记录过少的用户和电影。我们使用矩阵补全软件 *fancyimpute*（Rubinsteyn 和 Feldman（2016））估算缺失的估值。最终得到的实例包含 $n = 691$ 个买方和 $m = 632$ 件物品。

首先，我们比较各个新算法在不同步长策略下的表现：BCDEG 与 BCDEG-LS、BCPR 与 A-BCPR 与 BCPR-LS，以及 PR 与 PRLS。结果显示在图1和图2的左上、右上和左下角。在所有情况下，我们都看到新的线搜索变体表现最佳。

其次，我们将块坐标算法与最先进的确定性市场均衡算法进行比较：PGLS（Gao 和 Kroer（2020））和 PRLS。结果显示在图 1 和图 2 右下角。对于所有市场，在各项指标上表现最佳的都是 BCDEG 或 BCPR-LS，其次依次是 PRLS 和 PGLS。总体而言，我们发现随机分块坐标算法在各方面都比对应的确定性算法收敛更快，即使确定性方法已迭代数千次也是如此。因此，即使在高精度要求下，块坐标方法似乎也更优；同时，由于执行的更新次数多得多，其早期表现也更好

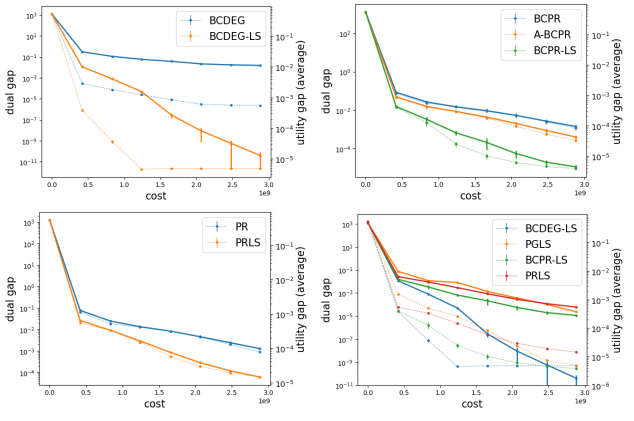

图2：在电影评分实例上的表现。图的设置与图1相同。

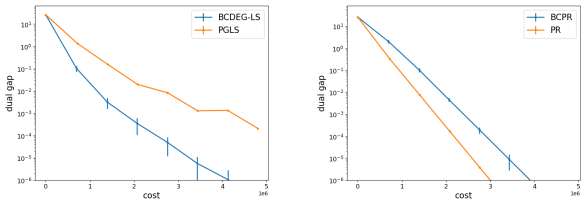

图3：在 CES 效用函数的模拟实例上的表现，其中 $\rho \in (0,1)$。设置与图1相同。

单位工作量下。

CES 效用。与线性效用类似，我们生成一个 $n = m = 200$ 规模市场实例，其中 CES 效用参数生成方式为 $v_{ij} = v_i v_j + \epsilon_{ij}$，其中 $v_i \sim \mathcal{N}(1, 0.2^2)$，$v_j \sim \mathcal{N}(1, 0.2^2)$，且 $\epsilon_{ij} \sim$ 均匀分布$(0, 0.2)$。对于 CES 实例，我们无法从现有锥优化求解器获得高精度解，因此仅根据对偶间隙衡量性能。有些出乎意料的是，我们发现对于 CES 效用，普通 PR 收敛非常快；比 BCPR 及所有版本的 BCDEG 都快。此外，BCDEG 的步长极小，因此我们必须使用 BCDEG-LS。在此处，我们为模拟效用中的分布采用了特定的小参数。关于 CES 效用的更多讨论和实验见附录E。

9 结论与未来工作

我们提出了两种用于计算大规模市场均衡的随机分块坐标算法：针对 EG 的（近端）块坐标下降（BCDEG）和分块坐标比例响应（BCPR）。我们为每种算法提供了理论收敛保证，并通过数值实验表明，它们的表现优于现有最先进算法。我们还为 BCDEG 中使用的投影梯度更新提出了一种新的经济学解释。未来，我们希望推导出采用自适应步长的 BCPR 的次线性收敛率，扩展该算法以利用分布式并行计算能力，并使其适用于更一般的动态场景和其他买家效用模型。

参考文献

Arnsperger, C. 1994. 无嫉妒性与分配正义。《经济调查杂志》, 8(2): 155–186.

Attouch, H.; Bolte, J.; and Svaiter, B. F. 2013. 半代数问题与温和问题中下降方法的收敛性：近端算法、前向–后向分裂及正则化 Gauss–Seidel 方法。《数学规划》，137(1): 91–129. Aziz, H.; and Ye, C. 2014. 分段常数和分段均匀估值的切蛋糕算法。载于《国际网络与互联网经济学会议》，1–14. 施普林格。Barman, S.; Krishnamurthy, S. K.; and Vaish, R. 2018. 寻找公平且有效的分配。载于《2018年ACM经济学与计算会议论文集》，557–574. Bauschke, H. H.; Bolte, J.; and Teboulle, M. 2017. 超越利普希茨梯度连续性的下降引理：重新审视一阶方法及其应用。《运筹学数学》，42(2): 330–348. Beck, A.; and Tetruashvili, L. 2013. 关于块坐标下降类方法的收敛性。*SIAM*优化学报, 23(4): 2037–2060. Birnbaum, B.; Devanur, N. R.; and Xiao, L. 2011. 通过梯度下降实现费希尔市场的分布式算法。载于《第12届ACM电子商务会议论文集》，127–136. ACM. Budish, E.; Cachon, G. P.; Kessler, J. B.; and Othman, A. 2016. 课程匹配：组合分配中等收入竞争均衡近似的大规模实现。《运筹学》，65(2): 314–336. Chen, G.; and Teboulle, M. 1993. 使用布雷格曼函数的类近端最小化算法的收敛性分析。*SIAM*优化学报，3(3): 538–543. Chen, X.; and Teng, S.-H. 2009. 支出并不比交易容易：费希尔均衡与阿罗–德布鲁均衡的计算等价性。载于《算法与计算国际研讨会》，647–656. 施普林格。Chen, Y.; and Ye, X. 2011. 投影到单纯形。*arXiv*预印本 *arXiv:1101.6081*。Cole, R.; Devanur, N. R.; Gkatzelis, V.; Jain, K.; Mai, T.; Vazirani, V. V.; and Yazdanbod, S. 2017. 凸规划对偶性、费希尔市场与纳什社会福利。载于《第18届ACM经济学与计算会议（EC 2017）》。计算机协会公司。Condat, L. 2016. 快速投影到单纯形及 ℓ_1球。《数学规划》，158(1): 575–585. Conitzer, V.; Kroer, C.; Panigrahi, D.; Schrijvers, O.; Sodomka, E.; Stier-Moses, N. E.; and Wilkens, C. 2019. 第一价格拍卖市场中的竞价节奏均衡。载于《2019年ACM经济学与计算会议论文集》。ACM。Conitzer, V.; Kroer, C.; Sodomka, E.; and Stier-Moses, N. E. 2018. 拍卖市场中的乘法竞价节奏均衡。载于《国际网络与互联网经济学会议》。

Daskalakis, C.; Goldberg, P. W.; and Papadimitriou, C. H. 2009. 纳什均衡计算的复杂性。*SIAM*计算杂志，39(1)：195–259。Devanur, N. R.; Papadimitriou, C. H.; Saberi, A.; and Vazirani, V. V. 2008. 通过凸规划的原始-对偶算法求解市场均衡。*ACM*学报（*JACM*），55(5)：1–18。Dooms, S.; De Pessemier, T.; and Martens, L. 2013. Movietweetings：从Twitter收集的电影评分数据集。Eisenberg, E. 1961. 效用函数聚合。管理科学，7(4)：337–350。Eisenberg, E.; and Gale, D. 1959. 主观概率的一致性：彩池法。数学统计年鉴，30(1)：165–168。Gao, T.; Lu, S.; Liu, J.; and Chu, C. 2020. 用于非Lipschitz优化的随机布雷格曼坐标下降法。*arXiv*预印本 *arXiv:2001.05202*。Gao, Y.; and Kroer, C. 2020. 大规模市场均衡计算的一阶方法。载于神经信息处理系统*2020*，*NeurIPS 2020*。Gao, Y.; and Kroer, C. 2021. 无限维费雪市场：均衡、对偶性与优化。载于*AAAI*人工智能会议论文集。Garg, R.; and Kapoor, S. 2006. 市场均衡的拍卖算法。《运筹学数学》，31(4)：714–729。Goktas, D.; Viqueira, E. A.; and Greenwald, A. 2021. 费雪市场均衡的消费者理论刻画。载于国际网络与互联网经济学会议，334–351。施普普林格。Hanzely, F.; and Richtárik, P. 2019. 任意采样下的加速坐标下降与小批量处理的最优速率。载于第*22*届国际人工智能与统计会议，304–312。PMLR。Hanzely, F.; and Richtárik, P. 2021. 随机镜像下降法的最快速率。《计算优化与应用》，1–50。Held, M.; Wolfe, P.; and Crowder, H. P. 1974. 次梯度优化的验证。《数学规划》，6(1)：62–88。Hoffman, A. J. 2003. 关于线性不等式组的近似解。载于*Alan J. Hoffman*论文选集：附评论，174–176。世界科技出版公司。Kantorovich, L. 1975. 经济学中的数学：成就、困难与展望。技术报告，诺贝尔奖委员会。Karimi, H.; Nutini, J.; and Schmidt, M. 2016. Polyak–Łojasiewicz条件下梯度法和近端梯度法的线性收敛。载于机器学习与知识发现联合欧洲会议，795–811。施普林格。Kroer, C.; Peysakhovich, A.; Sodomka, E.; and Stier-Moses, N. E. 2019. 利用抽象计算大型市场均衡。载于*2019*年*ACM*经济学与计算会议论文集，745–746。

Kroer, C.; and Stier-Moses, N. E. 2022. 大规模互联网市场中的市场均衡模型。金融与运营交界处的创新技术。*Springer*供应链管理系列。*Springer Nature*。Liu, J.; and Wright, S. J. 2015. 异步随机坐标下降：并行性与收敛性质。*SIAM*优化学报，25(1)：351–376。Lu, H.; Freund, R. M.; and Nesterov, Y. 2018. 通过一阶方法进行相对光滑凸优化及其应用。*SIAM*优化学报，28(1)：333–354。Mosek, A. 2010. MOSEK优化软件。在线地址：*http://www. mosek. com*，54(2-1)：5。Nesterov, Y. 2012. 超大规模优化问题中坐标下降方法的效率。*SIAM*优化学报，22(2)：341–362。Othman, A.; Papadimitriou, C.; and Rubinstein, A. 2016. 通过均衡实现公平的复杂性。*ACM*经济学与计算学报（*TEAC*），4(4)：1–19。Othman, A.; Sandholm, T.; and Budish, E. 2010. 求解近似竞争均衡：高效且公平的课程分配。载于*AAMAS*，第10卷，873–880。Perez, G.; Barlaud, M.; Fillatre, L.; and Régin, J.-C. 2020. 用于投影到单纯形和 ℓ_1 球的过滤桶聚类方法。《数学规划》，182(1)：445–464。Reddi, S. J.; Sra, S.; Póczos, B.; and Smola, A. 2016. 非光滑非凸优化的快速随机方法。*arXiv*预印本 *arXiv:1605.06900*。Richtárik, P.; and Taká ̆c, M. 2014. 用于复合函数最小化的随机分块坐标下降方法的迭代复杂度。《数学规划》，144(1)：1–38。Rubinsteyn, A.; and Feldman, S. 2016. fancyimpute：Python插补库。Scarf, H.; et al. 1967. 均衡价格的计算。耶鲁大学考尔斯经济学研究基金会，纽黑文，康涅狄格州。Shmyrev, V. I. 2009. 用于在固定预算下求解线性交换模型均衡的算法。《应用与工业数学杂志》，3(4)：505。Tseng, P. 2001. 不可微最小化的块坐标下降方法的收敛性。《优化理论与应用杂志》，109(3)：475–494。Varian, H. R.; et al. 1974. 公平、嫉妒与效率。《经济理论杂志》，9(1)：63–91。Vazirani, V. V.; and Yannakakis, M. 2011. 可分、分段线性凹效用下的市场均衡。*ACM*学报（*JACM*），58(3)：1–25。Wright, S. J. 2015. 坐标下降算法。《数学规划》，151(1)：3–34。Zhang, H. 2020. 通过统一误差界条件对梯度型方法线性收敛性的新分析。《数学规划》，180(1)：371–416。Zhang, L. 2011. 费雪市场中的比例响应动态。理论计算机科学，412(24)：2691–2698。

EG 程序的块坐标下降算法

BCDEG 的高效实现

为了高效实现算法1，我们会在内存中暂存一个向量 $u \in \mathbb{R}^n$，它表示所有买方当前的效用。然后，在每次迭代 k中，我们只需在更新某个特定块 j_k时替换 $x_{\cdot j_k}$，即可动态更新 u，耗时为 $O(n)$。这样便无需执行形如 $u_i = \langle v_i, x_i \rangle$的计算；若对每个买方都执行此类计算，则需要 $O(nm)$ 时间。详见算法3。

算法3： BCDEG 的高效实现

输入： 初始迭代点 x^0 $\eta_1^k\ \eta_2^k\ \dots\ \eta_m^k\ \ \forall\, k \in \mathbb{N}$，步长 ,,,, $u_i^0 \leftarrow \langle v_i, x_i^0 \rangle$ i 对所有 ;**对于** $k \leftarrow 1, 2, \dots$ **执行**以概率 $1/m$选取 $j_k \in [m]$；**对于** $i \leftarrow 1, 2, \cdots, n$ **执行若** $u_i^{k-1} \geq u_i$ **则** $g_i^{k-1} \leftarrow -B_i \log u_i^{k-1}$ **否则** $g_i^{k-1} \leftarrow -B_i \log u_i - \frac{B_i}{u_i}\left(u_i^{k-1} - u_i\right) + \frac{B_i}{2u_i^2}\left(u_i^{k-1} - u_i\right)^2$ $x_{\cdot j_k}^k \leftarrow$ 近端算子$(x_{\cdot j_k}^{k-1} - \eta_{j_k}^k g^{k-1})$;$x_{\cdot j}^k \leftarrow x_{\cdot j}^{k-1} \quad \forall j \neq j_k$;$u^k \leftarrow u^{k-1} + v_{j_k}($ $x_{\cdot j_k}^k - x_{\cdot j_k}^{k-1}$);

带线搜索的 EG 程序（近端）块坐标下降

我们在此正式陈述 BCDEG-LS。在我们的数值实验结果中，该算法通常优于其他算法。它与 BCDEG 的区别在于：执行试探性近端步后，我们会检查下降条件（算法4中的 if 条件）。如果未通过下降条件检查（if-else 语句中为 true 的情况），则减小步长并再次执行试探步。如果通过检查，则接受该试探性近端步。

算法4： BCDEG-LS：带线搜索的 EG 程序（近端）块坐标下降

输入： 初始迭代点 x^0，初始步长 $\eta_1^0,\ \eta_2^0, \dots, \eta_m^0$, $\rho_-,\ \ \rho_+ \in \mathbb{R}^+$ **对于** $k \leftarrow 1,\ 2, \dots$ **执行**以概率 $1/m$选取 $j_k \in [m]$；$g^{k-1} \leftarrow \nabla_{\cdot j_k} \tilde{f}(x^{k-1})$; $(*)$ $x_{\cdot j_k}^+ \leftarrow$ 近端算子$_{\Delta_n}\left(x_{\cdot j_k}^{k-1} - \eta_{j_k} g^{k-1}\right)$;~ $g^+ \leftarrow \nabla_{\cdot j_k} f(x^+)$;**如果** $\eta_{j_k}\|g^+ - g^{k-1}\| > \|x_{\cdot j_k}^+ - x_{\cdot j_k}^{k-1}\|$ **则** $\eta_{j_k} \leftarrow$ 最大化 $\{\rho_- \eta_{j_k}, 1/L_j\}$，其中 L_j 由 (4) 计算；返回 $(*)$ 并重新计算 $x_{\cdot j_k}^+$ 和 g^+；**否则**$\eta_{j_k} \leftarrow \rho_+ \eta_{j_k}$;$x_{\cdot j_k}^k \leftarrow x_{\cdot j_k}^+$ 和 $x_{\cdot j}^{(k)} \leftarrow x_{\cdot j}^{(k-1)} \quad \forall\, j \neq j_k$;

引理 1 的证明

~~证明。~~ 通过简单计算可以验证，$f_i(x)$ 和 $\nabla f_i(x)$ 在其定义域上都是连续的。~Hessian 矩阵 f 的每个元素都具有以下形式：

$$\nabla_{ij}^2 \tilde{f}(x) = \begin{cases} B_i v_{ij}^2 / u_i^2(x) & u_i(x) \geq \underline{u}_i \\ B_i v_{ij}^2 / \underline{u}_i^2 & u_i(x) < \underline{u}_i. \end{cases}$$

设 x 和 y 是仅在一个坐标 j上不同的两个向量。那么，

$$\left|\nabla_{ij}\tilde{f}(y)-\nabla_{ij}\tilde{f}(x)\right|=\left|\nabla_{ij}\tilde{f}(y_i)-\nabla_{ij}\tilde{f}(x_i)\right|\leq\left|\frac{B_i v_{ij}^2}{\underline{u}_i^2}\right||y_{ij}-x_{ij}|=L_j\|y_i-x_i\|, \tag{21}$$

其中，不等式是由

$$\nabla_{ij}^2\tilde{f}(x)\leq\frac{B_i v_{ij}^2}{\underline{u}_i^2}\leq\max_{i\in[n]}\frac{B_i v_{ij}^2}{\underline{u}_i^2}=L_j.$$

因此，

$$\left\|\nabla_{\cdot j}\tilde{f}(y)-\nabla_{\cdot j}\tilde{f}(x)\right\|^2=\sum_i\left|\nabla_{ij}\tilde{f}(y)-\nabla_{ij}\tilde{f}(x)\right|^2\leq L_j^2\sum_i\|y_i-x_i\|^2=L_j^2\|y-x\|^2=L_j^2\|y_{\cdot j}-x_{\cdot j}\|^2, \tag{22}$$

也就是说，

$$\left\|\nabla_{\cdot j}\tilde{f}(y)-\nabla_{\cdot j}\tilde{f}(x)\right\|\leq L_j\|y_{\cdot j}-x_{\cdot j}\|. \tag{23}$$

为推导式 (5)，我们可以应用积分：

$$\begin{aligned}\tilde{f}(y)-\tilde{f}(x)-\langle\nabla_{\cdot j}\tilde{f},y-x\rangle&=\int_0^1\langle\nabla_{\cdot j}\tilde{f}(x+\tau(y-x))-\nabla_{\cdot j}\tilde{f}(x),y-x\rangle\,d\tau\\&\leq\int_0^1\|\nabla_{\cdot j}\tilde{f}(x+\tau(y-x))-\nabla_{\cdot j}\tilde{f}(x)\|\|y-x\|\,d\tau\\&\leq\int_0^1\tau L_j\|y-x\|^2\,d\tau=\frac{L_j}{2}\|y-x\|^2.\end{aligned} \tag{24}$$

□

注。 作为对比，我们给出标准（全局）Lipschitz 常数 L 的证明如下。与上面的证明对比可知，当变量 $x_{\cdot j}$ 中只有一个块发生变化时，逐坐标 Lipschitz 常数 L_j 由海森矩阵所有 i对角线元素中的最大值界定。相比之下，对于 L，它取决于海森矩阵所有 i子迹中的最大值。

引理 5。 对于任意 $j\in[m]$，令

$$L=\max_{i\in[n]}\frac{B_i\|v_i\|^2}{\underline{u}_i^2}. \tag{25}$$

那么，对于所有的 x、$y\in\mathcal{X}$，我们都有

$$\tilde{f}(y)\leq\tilde{f}(x)+\langle\nabla\tilde{f}(x),y-x\rangle+\frac{L}{2}\|y-x\|^2. \tag{26}$$

˜证明。 令 x 和 y 为定义域 f 中的两个向量，则

$$\left|\nabla_{ij}\tilde{f}(y)-\nabla_{ij}\tilde{f}(x)\right|=\left|\nabla_{ij}\tilde{f}(y_i)-\nabla_{ij}\tilde{f}(x_i)\right|\leq\sum_{j'=0}^{n-1}\left|\nabla_{ij}\tilde{f}(y_i^{j'})-\nabla_{ij}\tilde{f}(x_i^{j'})\right|\leq\sum_{j'=0}^{n-1}\left|\frac{\partial^2}{\partial x_{ij}\partial x_{ij'}}\tilde{f}(x)\right|\left|y_i^{j'}-x_i^{j'}\right|$$

其中，$y_i^{j'}$ 是一个向量，其前（$j'+1$）个分量是 y_i 的前（$j'+1$）个分量的副本，其余分量是 x_i 的后（$m-j'-1$）个分量的副本；$x_i^{j'}$ 是一个向量，其前 j' 个分量是 y_i 的前 j' 个分量的副本，其余分量是 x_i 的后（$m-j'$）个分量的副本。也就是说，每一对 $y_i^{j'}$ 和 $x_i^{j'}$ 仅在第（$j'+1$）个分量上有所不同。此外，对于所有 i，j'，我们都有 $|y_i^{j'}-x_i^{j'}|\leq\|y_i-x_i\|$。

$$\sum_j\left|\nabla_{ij}\tilde{f}(y)-\nabla_{ij}\tilde{f}(x)\right|^2\leq\left(\sum_j\left|\nabla_{ij}\tilde{f}(y)-\nabla_{ij}\tilde{f}(x)\right|\right)^2\leq\left(\sum_j\sum_{j'}\frac{B_i v_{ij}v_{ij'}}{\underline{u}_i^2}\|y_i-x_i\|\right)^2\leq L^2\|y_i-x_i\|^2$$

由于

$$\frac{\partial^2}{\partial x_{ij}\partial x_{ij'}}\tilde{f}(x)\leq\frac{B_i v_{ij}v_{ij'}}{\underline{u}_i^2}\quad\text{and}\quad\sum_j\sum_{j'}\frac{B_i v_{ij}v_{ij'}}{\underline{u}_i^2}=\frac{B_i\|v_i\|^2}{\underline{u}_i^2}\leq\max_{i\in[n]}\frac{B_i\|v_i\|^2}{\underline{u}_i^2}=L.$$

因此，

$$\left\|\nabla\tilde{f}(y)-\nabla\tilde{f}(x)\right\|^2=\sum_i\sum_j\left|\nabla_{ij}\tilde{f}(y)-\nabla_{ij}\tilde{f}(x)\right|^2\le L^2\sum_i\|y_i-x_i\|^2=L^2\|y-x\|^2. \tag{27}$$

□

注意，由于 $\max_{i\in[n]}\frac{B_i\|v_i\|^2}{u_i^2}\le\max_{i\in[n]}\frac{B_i}{u_i^2}\max_{i\in[n]}\|v_i\|^2$，这是一个比 Gao 和 Kroer (2020) 更紧的界。

引理 2 的证明

证明。 由于 $g(u)$ 是强凸的，因此存在唯一的 u^*，使得

$$g(u^*)=g^*\qquad Ax^*=u^*\quad\forall x^*\in\mathcal{X}^*.$$

因此，对于某个 C，最优解集合 $\mathcal{X}^*$ 可由以下多面体集合描述:

$$\mathcal{X}^*=\{x^*:Ax^*=u^*,Cx^*\le d\}.$$

假设最优多面体集合 $\mathcal{X}^*$ 非空，则 Hoffman 不等式（Hoffman 2003）表明存在某个正数，使得

$$\|x-x^*\|\le\theta(A,C)\left\|\begin{bmatrix}Ax-u^*\\(Cx-d)^+\end{bmatrix}\right\|\quad\forall\,x. \tag{28}$$

令 $\mathcal{X}$ 为可行分配集合。对于任意满足 $Cx\le d$ 的 $x\in\mathcal{X}$，我们有

$$\|x-x^*\|\le\theta(A,C)\|Ax-Ax^*\|. \tag{29}$$

由 $\tilde{g}$ 的强凸性和式 (29)，我们有

$$\begin{aligned}F^*=F(x^*)&\ge F(x)+\langle\nabla\tilde{f}(x),x^*-x\rangle+\frac{\mu}{2}\|Ax^*-Ax\|^2+\psi(x^*)-\psi(x)\\&\ge F(x)+\langle\nabla\tilde{f}(x),x^*-x\rangle+\frac{\mu}{2\theta^2(A,C)}\|x^*-x\|^2+\psi(x^*)-\psi(x).\end{aligned} \tag{30}$$

如果 $L\ge\frac{\mu}{\theta^2(A,C)}$，则

$$\begin{aligned}F^*&\ge F(x)+\langle\nabla\tilde{f}(x),x^*-x\rangle+\frac{\mu}{2\theta^2(A,C)}\|x^*-x\|^2+\psi(x^*)-\psi(x)\\&\ge F(x)+\min_{y\in\mathbb{R}^{nm}}\left\{\langle\nabla\tilde{f}(x),y-x\rangle+\frac{\mu}{2\theta^2(A,C)}\|y-x\|^2+\psi(y)-\psi(x)\right\}\\&=F(x)-\frac{\theta^2(A,C)}{2\mu}\mathcal{D}_\psi(x,\frac{\mu}{\theta^2(A,C)}).\end{aligned} \tag{31}$$

因此，

$$\frac{1}{2}\mathcal{D}_\psi(x,\frac{\mu}{\theta^2(A,C)})\ge\frac{\mu}{\theta^2(A,C)}(F(x)-F^*). \tag{32}$$

由于 $\mathcal{D}(x,\cdot)$ 是非递减的（Karimi、Nutini 和 Schmidt，2016，引理 1），我们有

$$\frac{1}{2}\mathcal{D}_\psi(x,L)\ge\frac{\mu}{\theta^2(A,C)}(F(x)-F^*). \tag{33}$$

如果 $L<\frac{\mu}{\theta^2(A,C)}$，则

$$\begin{aligned}F^*&\ge F(x)+\langle\nabla\tilde{f}(x),x^*-x\rangle+\frac{\mu}{2\theta^2(A,C)}\|x^*-x\|^2+\psi(x^*)-\psi(x)\\&\ge F(x)+\langle\nabla\tilde{f}(x),x^*-x\rangle+\frac{L}{2}\|x^*-x\|^2+\psi(x^*)-\psi(x)\\&\ge F(x)+\min_{y\in\mathbb{R}^{nm}}\left\{\langle\nabla\tilde{f}(x),y-x\rangle+\frac{L}{2}\|y-x\|^2+\psi(y)-\psi(x)\right\}\\&=F(x)-\frac{1}{2L}\mathcal{D}_\psi(x,L).\end{aligned} \tag{34}$$

因此，

$$\frac{1}{2}\mathcal{D}_\psi(x,L)\ge L(F(x)-F^*). \tag{35}$$

综合这两种情况，我们可以得出式（6）。 □

定理 1 的证明

下面的收敛性分析是针对均匀（块）坐标下降算法的典型分析。利用式（6），很容易推导出 BCDEG 和 BCDEG-LS 的线性收敛率。Karimi、Nutini 和 Schmidt（2016，附录 H）给出了这一证明方案；这里我们则给出一个变体，其中不同块和不同迭代采用不同的步长。

证明。 为提醒读者注意记号，我们重新定义

$$T_j(x) = \arg\min_{y_{\cdot j}} \langle \nabla_{\cdot j} f(x), y_{\cdot j} - x_{\cdot j} \rangle + L_j \|y_{\cdot j} - x_{\cdot j}\|^2 + \psi_j(y_{\cdot j}) \tag{36}$$

BCDEG 为

$$T_j^k(x) = \arg\min_{y_{\cdot j}} \langle \nabla_{\cdot j} f(x), y_{\cdot j} - x_{\cdot j} \rangle + \frac{1}{\eta_j^k} \|y_{\cdot j} - x_{\cdot j}\|^2 + \psi_j(y_{\cdot j}) \tag{37}$$

BCDEG-LS 为。

于是，我们有（以 BCDEG 为例）

$$\begin{aligned}
\mathbf{E}[F(x^{k+1}) \mid x^k] &\le F(x^k) + \frac{1}{m} \sum_j \left\{ \langle \nabla_{\cdot j} \tilde{f}(x^k), T_j(x) - x_{\cdot j}^k \rangle + \frac{L_j}{2} \|T_j(x) - x_{\cdot j}^k\|^2 + \psi_j(T_j(x)) \right\} \qquad (38) \\
&= \tilde{f}(x^k) + \frac{1}{m} \sum_j \min_{y_{\cdot j}} \left\{ \langle \nabla_{\cdot j} \tilde{f}(x^k), y_{\cdot j} - x_{\cdot j}^k \rangle + \frac{L_j}{2} \|y_{\cdot j} - x_{\cdot j}^k\|^2 + \psi_j(y_{\cdot j}) \right\} \\
&= \tilde{f}(x^k) + \frac{1}{m} \min_y \left\{ \langle \nabla \tilde{f}(x^k), y - x^k \rangle + \frac{1}{2} \|y - x^k\|_L^2 + \psi(y) \right\} \\
&= F(x^k) + \frac{1}{m} \min_y \left\{ \langle \nabla \tilde{f}(x^k), y - x^k \rangle + \frac{1}{2} \|y - x^k\|_L^2 + \psi(y) - \psi(x^k) \right\} \\
&\le F(x^k) + \frac{1}{m} \min_y \left\{ \langle \nabla \tilde{f}(x^k), y - x^k \rangle + \frac{L_{\max}}{2} \|y - x^k\|^2 + \psi(y) - \psi(x^k) \right\} \\
&= F(x^k) - \frac{1}{2mL_{\max}} \mathcal{D}_\psi(x^k, L_{\max}), \qquad (39)
\end{aligned}$$

˜其中，我们使用了 f 的 Lipschitz 光滑性（对于 BCDEG-LS，则使用线搜索条件）、 $T_j(x)$ 的定义，以及每个块 j 上的可分性。

对于 BCDEG-LS，将（38）替换为以下不等式:

$$\mathbf{E}[F(x^{k+1}) \mid x^k] \le F(x^k) + \frac{1}{m} \sum_j \left\{ \langle \nabla_{\cdot j} \tilde{f}(x^k), T_j(x) - x_{\cdot j}^k \rangle + \frac{1}{2\eta_j^k} \|T_j(x) - x_{\cdot j}^k\|^2 + \psi_j(T_j(x)) \right\}, \tag{40}$$

并根据 $1/\eta_j^k \le L_j \le L_{\max}$ （用于最后一个不等式），我们得到（39）。

根据引理 2，我们有

$$\mathbf{E}[F(x^{k+1}) \mid x^k] - F(x^k) \le -\frac{1}{2mL_{\max}} \mathcal{D}_\psi(x^k, L_{\max}) \le -\min\left\{ \frac{\mu}{mL_{\max}\theta^2(A,C)}, \frac{1}{m} \right\} (F(x^k) - F^*). \tag{41}$$

在两边减去 F^*，重新整理上述不等式，并对 j_1、.....、 j_k 取期望，通过归纳法可得（7）。

□

B 单纯形投影的算法细节

我们给出单纯形投影算法的一个关键定理，以及一个形式化表述的算法，作为对不熟悉该算法的读者的背景介绍。

定理 6. （*Chen* 和 *Ye 2011*，定理 *2.2*。）对于任意向量 $y \in \mathbb{R}^n$， y 在 Δ^n 上的投影由 $y - t$ 的正ˆ部分得到:

$$x = (y - \hat{t})^+, \tag{42}$$

where ˆt 是 $\{t_i : i = 0, \ldots, n-1\}$ 中唯一落入相应区间的值，如下所示 *ws,*

$$t_i := \frac{\sum_{j=i+1}^n y_{(j)} - 1}{n-i}, i = 0, \ldots, n-1, \quad \textit{where } t_1 \le y_{(1)} \textit{ and } y_{(i)} \le t_i \le y_{(i+1)}, i = 1, \ldots, n-1. \tag{43}$$

因此，要找到 x，我们只需找到 (43) 中落入相应区间的 t_i，将其认定为最优的 $\hat{t}$，然后根据 (42) 计算 x_i。上述定理导出了以下算法。

算法 5： 单纯形投影

输入： $y = (y_1, y_2, \ldots, y_n)^\top \in \mathbb{R}^n$ **输出：** $x = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$ 作为 y 在 Δ^n 上的投影。按升序对 y 排序，得到 $y_{(1)} \le \cdots \le y_{(n)}$；$\hat{t} \leftarrow \frac{\sum_{j=1}^{n} y_{(j)} - 1}{n}$；**对于** $i \leftarrow n-1, \ldots, 1$ **执行** $t_i \leftarrow \frac{\sum_{j=i+1}^{n} y_{(j)} - 1}{n-i}$；**如果** $t_i \ge y_{(i)}$ **则** $\hat{t} \leftarrow t_i$ 并跳出循环；$x \leftarrow (y - \hat{t})^+$；

关于单纯形投影替代算法的讨论

本文假设上述算法是一种将数组投影到单纯形上的高效方法，并对基于该算法的投影型 ME 求解算法进行了解释。这类投影算法有多个变体，其中一些能够改善复杂度结果或实际性能。Condat（2016）总结了其中大多数算法；这些算法采用不同方法来寻找正确的阈值。一些算法利用问题结构，获得$O(n)$的期望或实际复杂度。然而，在最坏情况下，仍没有比 $O(n\log(n))$ 更好的结果。Perez 等人（2020）提出了一种具有最坏情况线性时间复杂度的算法。他们的结果类似于桶排序，假设可能遇到的取值集合是常数。

C 分块坐标比例响应

带线搜索的分块坐标比例响应

我们正式表述带线搜索的分块坐标比例响应（BCPR-LS）算法如下。请注意，$\delta \in [0, 1)$ 是一个“保守”因子——我们可以设置 $\delta > 0$，采用比标准线搜索（$\delta = 0$）更保守的步长策略。在 $\delta > 0$ 情况下，我们可以保证 BCPR-LS 具有某种（较弱的）收敛性质。

算法 6： 带线搜索的分块坐标比例响应（BCPR）

输入： $b^0 \in \mathbb{R}^{n \times m}, p^0 \in \mathbb{R}^m, \alpha_1, \ldots, \alpha_n \in \mathbb{R}^+, \rho_-, \rho_+ \in \mathbb{R}^+$

对于 $k \leftarrow 1, 2, \ldots$ **执行**

- 以概率 $1/n$ 选择 $i_k \in [n]$；
- （*）根据式（14）以步长（$1 - \delta$）计算 $b_{i_k}^+ \alpha_{i_k}$；
- $p^+ \leftarrow p + b_{i_k}^+ - b_{i_k}$；
- **如果** $\alpha_{i_k} D_{\mathrm{KL}}(p^+, p) > D_{\mathrm{KL}}(b_{i_k}^+, b_{i_k})$ **则**
  - $\alpha_{i_k} \leftarrow \rho_- \alpha_{i_k}$；
  - 回到（*）并重新计算 b_{ik}^+ 和 p^+；
- **else**
  - $\alpha_{i_k} \leftarrow \rho_+ \alpha_{i_k}$；
  - $b_{i_k}^k \leftarrow b_{i_k}^+$ 以及 $b_{i'}^k \leftarrow b_{i'}^{k-1} \quad \forall\, i' \ne i_k$；

式（14）的证明

证明。原始最小化问题可以通过引入一个拉格朗日乘子 λ，采用拉格朗日方法求解：

$$\min \mathcal{L}(a_1, a_2, \ldots, a_m, \lambda) = \langle \nabla_i \varphi(b), a - b_i \rangle + \frac{1}{\alpha} D_{\mathrm{KL}}(a, b_i) + \lambda \Big(B_i - \sum_j a_j \Big). \tag{44}$$

令 $\nabla_{a_1, a_2, \ldots, a_m, \lambda} \mathcal{L} = 0$，我们有

$$\frac{\partial}{\partial a_j} \left(\langle \nabla_i \varphi(b), a - b_i \rangle + \frac{1}{\alpha} D_{\mathrm{KL}}(a, b_i) \right) = \lambda \quad \forall j = 1, 2, \ldots, m \quad \text{and} \quad \sum_j a_j = B_i. \tag{45}$$

偏导数等于

$$1-\log\left(\frac{v_{ij}}{p_j}\right)+\frac{1}{\alpha}\left(1+\log\frac{a_j}{b_{ij}}\right)\quad\forall j=1,2,\dots,m.$$

设 $a_j=\frac{1}{Z}b_{ij}\left(\frac{v_{ij}}{p_j}\right)^{\alpha}$, $\lambda=1+\frac{1}{\alpha}-\log Z$, 其中 $Z=\frac{1}{B_i}\sum_j b_{ij}\left(\frac{v_{ij}}{p_j}\right)^{\alpha}$。我们可以验证

$$1+\frac{1}{\alpha}-\log Z=\frac{\partial}{\partial a_j}\left(\langle\nabla_i\varphi(b),a-b_i\rangle+\frac{1}{\alpha}D_{\mathrm{KL}}(a,b_i)\right)=\lambda\quad\text{and}\quad\sum_j a_j=\frac{1}{Z}\sum_j b_{ij}\left(\frac{v_{ij}}{p_j}\right)^{\alpha}=B_i. \tag{46}$$

因此，$(a_1,\ a_2,\ \dots,\ a_m,\ \lambda)$ 是拉格朗日函数的一个驻点。由于该问题具有凸性，$a=(a_1,\ a_2,\ \dots,\ a_m)$ 是原始最小化问题的一个解。 □

坐标相对光滑性（期望意义下）

在本节余下部分中，除 A-BCPR 外，我们使用一般的 Bregman 距离 $D(\sum_i L_i D_i\ a_i\ b_i$ $L\in\mathbb{R}^n$ $D_i\ a_i\ b_i$ a $T^k=\sum_i T_i^k$ a $=$ ，$b)$ 来度量，$b\in\mathbb{R}^n$ 与 $D_L(,\ b)(,\ )$ 之间的距离，其中 $(,\ )$ 是坐标 Bregman 距离，且 。我们还如下引入（类似于 l_2 范数块坐标下降）：

$$T^k=\sum_i T_i^k=\operatorname*{arg\,min}_{a\in\mathbb{R}^{n\times m}}\left\{\langle\nabla\varphi(b),a-b\rangle+\sum_i\frac{1}{\alpha_i^k}D_i(a_i,b_i)+r(a)\right\}. \tag{47}$$

因此，以下引理是标准结论。

引理 6. 若 b^k 是由算法 *6* 生成的随机迭代点，则

$$\begin{aligned}\mathbf{E}\left[\varphi(b^{k+1})\,\middle|\,b^k\right]\leq&\varphi(b^k)+\frac{1}{n}\langle\nabla\varphi(b^k),T^k-b^k\rangle+\frac{1}{n}D_{1/\alpha^k}(T^k,b^k) &(48)\\ =&\varphi(b^k)+\mathbf{E}\left[\langle\nabla\varphi(b^k),b^{k+1}-b^k\rangle\,\middle|\,b^k\right]+\mathbf{E}\left[D_{1/\alpha^k}(b^{k+1},b^k)\,\middle|\,b^k\right]. &(49)\end{aligned}$$

注意，(48) 是 h-ESO 假设的一个特例（采样基数等于 1，即均匀采样），该假设由 Hanzely 和 Richtárik（2021）提出。从本质上说，这就是期望意义下的坐标相对光滑性。

证明。 给定任意坐标 i，我们有（令 b^+ 和 p^+ 分别为 b 和 p的下一个迭代点）

$$\begin{aligned}\varphi(b^+)-\varphi(b)-\langle\nabla_i\varphi(b),b_i^+-b_i\rangle=&\varphi(b^+)-\varphi(b)-\langle\nabla\varphi(b),b^+-b\rangle\\ =&-\sum_{i,j}b_{ij}^+\log\frac{v_{ij}}{p_j^+}+\sum_{i,j}b_{ij}\log\frac{v_{ij}}{p_j}-\sum_{i,j}\left(1-\log\frac{v_{ij}}{p_j}\right)(b_{ij}^+-b_{ij})\\ =&\sum_{i,j}b_{ij}^+\log\frac{p_j^+}{p_j}=\sum_j p_j^+\log\frac{p_j^+}{p_j},\end{aligned} \tag{50}$$

并且 $D_{\mathrm{KL}}(b_i^+,\ b_i)=\sum_j b_{ij}^+\log\frac{b_{ij}^+}{b_{ij}}$。线搜索条件保证

$$\varphi(b^+)\leq\varphi(b)+\langle\nabla_i\varphi(b),b_i^+-b_i\rangle+\frac{1}{\alpha_i}D_{\mathrm{KL}}(b_i^+,b_i). \tag{51}$$

注意，若令 $a\in\mathbb{R}$ 和 $f(x)=x\log\frac{x}{x-a}$，则 $f'(x)=\log\left(1+\frac{a}{x-a}\right)-\frac{a}{x-a}\leq\frac{a}{x-a}-\frac{a}{x-a}=0$。由于 $p_j^+-p_j=b_{ij}^+-b_{ij}$，令 $a=b_{ij}^+-b_{ij}$，我们有

$$\sum_j p_j^+\log\frac{p_j^+}{p_j}=\sum_j f(p_j^+)\leq\sum_j f(b_{ij}^+)=\sum_j b_{ij}^+\log\frac{b_{ij}^+}{b_{ij}}, \tag{52}$$

其中我们使用了 $p_j^+\geq b_{ij}^+$ 和 $f(x)$ 的单调性。注意，当对所有 i 都有 $\alpha_i=1$ 时，式 (51) 始终成立。因此，由于每个坐标块上的可分性，我们有

$$\begin{aligned}\mathbf{E}\left[\varphi(b^{k+1})\,\middle|\,b^k\right]\leq&\sum_i\frac{1}{n}\left\{\varphi(b^k)+\langle\nabla_i\varphi(b),T_i^k-b_i^k\rangle+\frac{1}{\alpha_i^k}D_{\mathrm{KL}}(T_i^k,b_i^k)\right\}\\ =&\varphi(b^k)+\frac{1}{n}\langle\nabla\varphi(b^k),T^k-b^k\rangle+\frac{1}{n}D_{1/\alpha^k}(T^k,b^k).\end{aligned} \tag{53}$$

众所周知

$$\mathbf{E}\left[\langle \nabla\varphi(b^k), b^{k+1}-b^k\rangle \mid b^k\right] = \sum_i \frac{1}{n}\langle \nabla_i\varphi(b^k), T_i^k - b_i^k\rangle = \frac{1}{n}\langle \nabla\varphi(b^k), T^k - b^k\rangle \tag{54}$$

and

$$\mathbf{E}\left[D_{1/\alpha^k}(b^{k+1}, b^k) \mid b^k\right] = \sum_i \frac{1}{n}\frac{1}{\alpha_i^k} D_i(T_i^k, b_i^k) = \frac{1}{n} D_{1/\alpha^k}(T^k, b^k), \tag{55}$$

(48) 和 (49) 等价。

□

一个有用的下降引理

接下来，我们将考虑带线搜索的 BCPR（BCPR-LS）。我们可以直接看出，BCPR（采用固定步长）的次线性收敛率是我们收敛性质的一个特例。

下面的引理可由引理 6 和三点性质推出（例如，Hanzely 和 Richtárik（2021，引理 3.5））。除了已有结果之外，我们还证明了一个适用于线搜索和 $\alpha_i^k = (1-\delta)/L_i^k$ 步长策略的版本。

引理7. *若 b^k 是由算法 6 生成的随机迭代点，则*

$$\begin{aligned}\mathbf{E}\left[\Phi(b^{k+1}) \mid b^k\right] &\le \frac{n-1}{n}\Phi(b^k) + \frac{1}{n}\left\{\Phi(u) + D_{\alpha^k}(u, b^k) - D_{\alpha^k}(u, T^k) - \delta D_{\alpha^k}(T^k, b^k)\right\} \\ &= \frac{n-1}{n}\Phi(b^k) + \frac{1}{n}\Phi(u) + \frac{1}{n}D_{\frac{1}{\alpha^k}}(u, b^k) - \frac{1}{n}\mathbf{E}\left[D_{\frac{1}{\alpha^k}}(u, b^{k+1}) \,\middle|\, b^k\right] - \frac{\delta}{n}\mathbf{E}\left[D_{\frac{1}{\alpha^k}}(b^{k+1}, b^k) \,\middle|\, b^k\right], \end{aligned} \tag{56}$$

其中 T^k 定义于 (47)，且 $u \in \operatorname{dom}\varphi \cup \operatorname{dom} h$。

证明。给定任意 b^k（第 k次迭代时的迭代点），我们有

$$\begin{aligned}\mathbf{E}\left[\Phi(b^{k+1}) \mid b^k\right] \le& \varphi(b^k) + \mathbf{E}\left[\langle \nabla\varphi(b^k), b^{k+1}-b^k\rangle \mid b^k\right] + \mathbf{E}\left[D_{L^k}(b^{k+1}, b^k) \mid b^k\right] + \mathbf{E}\left[r(b^{k+1}) \mid b^k\right] \\ =& \varphi(b^k) + \frac{1}{n}\langle \nabla\varphi(b^k), T^k - b^k\rangle + \frac{1}{n}D_{L^k}(T^k, b^k) + \frac{1}{n}\sum_i r_i\left(T_i^k\right) + \frac{1}{n}\sum_i\sum_{i'\neq i} r_{i'}\left(b_{i'}^k\right) \\ =& \varphi(b^k) + \frac{1}{n}\langle \nabla\varphi(b^k), T^k - b^k\rangle + \frac{1}{n}D_{L^k}(T^k, b^k) + \frac{1}{n}r\left(T^k\right) + \frac{n-1}{n}r\left(b^k\right) \\ =& \frac{n-1}{n}\Phi(b^k) + \frac{1}{n}\left\{\varphi(b^k) + \langle \nabla\varphi(b^k), T^k - b^k\rangle + D_{L^k}(T^k, b^k) + r\left(T^k\right)\right\}, \end{aligned} \tag{57}$$

这由引理 6 以及 r的可分性得出。

由（47）的最优性可得

$$v^k = -\nabla\varphi(b^k) - \sum_i \frac{1}{\alpha_i^k}\nabla h_i(T_i^k) + \sum_i \frac{1}{\alpha_i^k}\nabla h_i(b_i^k) \tag{58}$$

对于某个 $v^k \in \partial r(T^k)$。

由 φ 和 r的凸性，对于任意 $u \in \operatorname{dom}\varphi \cup \operatorname{dom} h$,

$$r(T^k) \le r(u) - \langle v^k, u - T^k\rangle, \tag{59}$$

$$\varphi(b^k) \le \varphi(u) - \langle \nabla\varphi(b^k), u - b^k\rangle. \tag{60}$$

根据 Bregman 距离的定义，我们有所谓的三点恒等式：

$$\langle \nabla h_i(T_i^k) - \nabla h_i(b_i^k), u_i - T_i^k\rangle = D_i(u_i, b_i^k) - D_i(u_i, T_i^k) - D_i(T_i^k, b_i^k) \tag{61}$$

对于所有 $u_i \in \operatorname{dom} h_i$ 以及所有 $i \in [n]$。

然后，结合以上所有不等式和等式，我们得到

$$\begin{aligned}\mathbf{E}\left[\Phi(b^{k+1}) \mid b^k\right] \leq & \frac{n-1}{n}\Phi(b^k)+\frac{1}{n}\left\{\varphi(b^k)+\langle\nabla\varphi(b^k), T^k-b^k\rangle+D_{L^k}(T^k,b^k)+r\left(T^k\right)\right\} \\ \leq & \frac{n-1}{n}\Phi(b^k)+\frac{1}{n}\left\{\varphi(b^k)+\langle\nabla\varphi(b^k), u-b^k\rangle\right\} \\ & +\frac{1}{n}\left\{D_{L^k}(T^k,b^k)+r\left(u\right)+\sum_i \frac{1}{\alpha_i^k}\langle\left(\nabla h_i\left(T_i^k\right)-\nabla h_i\left(b_i^k\right)\right), u-T^k\rangle\right\} \\ \leq & \frac{n-1}{n}\Phi(b^k)+\frac{1}{n}\varphi(u)+\frac{1}{n}\left\{D_{L^k}(T^k,b^k)+r\left(u\right)+\sum_i \frac{1}{\alpha_i^k}\langle\left(\nabla h_i\left(T_i^k\right)-\nabla h_i\left(b_i^k\right)\right), u-T^k\rangle\right\} \\ = & \frac{n-1}{n}\Phi(b^k)+\frac{1}{n}\left\{\Phi(u)+D_{\alpha^k}(u,b^k)-D_{\alpha^k}(u,T^k)-\delta D_{\alpha^k}(T^k,b^k)\right\} \end{aligned} \tag{62}$$

对于任意 u。这里，我们使用 (57)、(59) 以及 (58)、(60)、(61) 和 $\alpha_i^k=(1-\delta)/L_i^k$（对于 $\delta\in[0,1)$）。(56) 可由与 (54) 和 (55) 类似的恒等关系推出。 □

以下推论给出了 BCPR-LS 的一些（较弱的）收敛性质。

推论 1. 若 b^k 是由算法 *6* 生成的随机迭代点，则

(i) $E[\Phi(b^k)]$ 单调不增；*(ii)* $\sum_{k=0}^{\infty}E[D_{1/\alpha^k}(b^{k+1},b^k)]<\infty$，$E[D_i(b_i^{k+1},b_i^k)]\to 0$ 当 $k\to\infty$ 时趋于 i；*(iii)* $\sum_{k=0}^{\infty}E[D_{1/\alpha^k}(b^k,b^{k+1})]<\infty$，$E[D_i(b_i^k,b_i^{k+1})]\to 0$ 当 $k\to\infty$ 时趋于 i；*(iv)* $E[D_{1/\alpha^k}(b^*,b^k)]\leq E[D_{1/\alpha^k}(b^*,b^{k+1})]$ 当 $k\to\infty$时趋于。

证明。
(i) 令 $u=b^k$ 属于 (56)，我们有

$$\mathbf{E}\left[\Phi(b^{k+1}) \mid b^k\right] \leq \Phi(b^k) \tag{63}$$

利用 Bregman 距离的非负性，可得期望意义下的目标下降。(ii) 此外，我们有

$$\mathbf{E}[\Phi(b^{k+1})|b^k] \leq \Phi(b^k)+\frac{1}{n}\left\{-D_{\alpha^k}(b^k,T^k)-\delta D_{\alpha^k}(T^k,b^k)\right\} \leq \Phi(b^k)-\frac{\delta}{n}D_{\alpha^k}(T^k,b^k). \tag{64}$$

众所周知，$\mathbf{E}[D_{\alpha^k}(b^{k+1},b^k)|b^k]=\frac{1}{n}D_{\alpha^k}(T^k,b^k)$，(64) 等价于

$$\delta\mathbf{E}[D_{\alpha^k}(b^{k+1},b^k)|b^k] \leq \Phi(b^k)-\mathbf{E}[\Phi(b^{k+1})|b^k]. \tag{65}$$

将 (65) 对 $0,1,\ldots,k-1$ 求和，并对 $i_0,i_1,\ldots,i_{k-1}$ 取期望，我们有

$$\delta\sum_{l=0}^{k}\mathbf{E}[D_{\alpha^l}(b^{l+1},b^l)] \leq \Phi(b^0)-\mathbf{E}[\Phi(b^k)]<\infty. \tag{66}$$

令 $k\to\infty$，我们有 $\sum_{l=0}^{\infty}\mathbf{E}[D_{\alpha^l}(b^{l+1},b^l)]<\infty$，这意味着 $\mathbf{E}[D_{\alpha^k}(b^{k+1},b^k)]\to 0$，当 $k\to\infty$ 时。(iii) 类似于 (64)，我们有

$$\mathbf{E}[\Phi(b^{k+1})|b^k] \leq \Phi(b^k)+\frac{1}{n}\left\{-D_{\alpha^k}(b^k,T^k)-\delta D_{\alpha^k}(T^k,b^k)\right\} \leq \Phi(b^k)-\frac{1}{n}D_{\alpha^k}(b^k,T^k). \tag{67}$$

然后，(iii) 可由与 (ii) 证明相同的论证得出。
(iv) 由于 $\mathbf{E}[\Phi(b^k)]$ 是非增的，且 $\mathbf{E}[\Phi(b^k)]\geq\Phi^*>-\infty$，因此 $\mathbf{E}[\Phi(b^{k+1})-\Phi(b^k)]\to 0$。
令 $u=b^*$ 属于 (62)，我们有

$$\mathbf{E}\left[\Phi(b^{k+1}) \mid b^k\right] \leq \frac{n-1}{n}\Phi(b^k)+\frac{1}{n}\left\{\Phi(b^*)+D_{\alpha^k}(b^*,b^k)-D_{\alpha^k}(b^*,T^k)\right\} \tag{68}$$

这是由于 Bregman 距离非负。将 (68) 重新整理后，对 $0,\ldots,k$ 求和，并利用塔式性质，可得

$$(n-1)\mathbf{E}[\Phi(b^{k+1})-\Phi(b^k)]+\left(\mathbf{E}[\Phi(b^{k+1})]-\Phi(u)\right) \leq \mathbf{E}[D_{\alpha^k}(u,b^k)]-\mathbf{E}[D_{\alpha^k}(u,b^{k+1})]. \tag{69}$$

当 $k\to\infty$时，(69) 的左侧变为非负，这意味着 (iv)。 □

定理2的证明

证明。对于任意 $u \in \operatorname{dom} h$，我们有

$$\begin{aligned}\mathbf{E}\left[D_{1/\alpha^k}(u, b^{k+1}) \,\middle|\, b^k\right] &= \sum_i \frac{1}{n}\left\{\frac{1}{\alpha_i^k} D_i(u_i, T_i^k) + \sum_{i' \neq i} \frac{1}{\alpha_{i'}^k} D_i(u_{i'}, b_{i'}^k)\right\} \\ &= \frac{1}{n} D_{1/\alpha^k}(u, T^k) + \frac{n-1}{n} D_{1/\alpha^k}(u, b^k). \end{aligned} \tag{70}$$

根据引理7和（70），令 $u = b^*$，则得到

$$\mathbf{E}\left[\Phi(b^{k+1}) + D_{1/\alpha^k}(b^*, b^{k+1}) \,\middle|\, b^k\right] \leq \frac{n-1}{n}\Phi(b^k) + \frac{1}{n}\Phi(b^*) + D_{1/\alpha^k}(b^*, b^k). \tag{71}$$

重排（71），得到

$$\mathbf{E}\left[\Phi(b^{k+1}) - \Phi(b^*) + D_{1/\alpha^k}(b^*, b^{k+1}) \,\middle|\, b^k\right] \leq \frac{n-1}{n}\left(\Phi(b^k) - \Phi(b^*)\right) + D_{1/\alpha^k}(b^*, b^k). \tag{72}$$

对 $\{i_0, i_1, \ldots, i_k\}$ 取（72）的期望，并将该不等式对 $l = 0$、.....、 k 求和，得到

$$\begin{aligned}&\mathbf{E}\left[\Phi(b^{k+1}) - \Phi(b^*)\right] + \mathbf{E}\left[D_{1/\alpha^k}(b^*, b^{k+1})\right] \\ &\leq \left(\Phi(b^0) - \Phi(b^*)\right) - \frac{1}{n}\sum_{l=0}^{k} \mathbf{E}\left[\Phi(b^l) - \Phi(b^*)\right] + D_{1/\alpha^0}(b^*, b^0) + \sum_{l=1}^{k} \mathbf{E}\left[D_{1/\alpha^l}(b^*, b^l) - D_{1/\alpha^{l-1}}(b^*, b^l)\right]. \end{aligned} \tag{73}$$

容易看出，当 $\alpha_i^k = 1/L_i$时，（73）中的最后一项为零，重排后即可得到次线性收敛率。对于线搜索情形，我们可以推导出（16）。

备注。 尽管（73）中最后一项有许多等价表达式（如下文示例所示），但我们无法保证其收敛。

$$\begin{aligned}&\sum_{l=1}^{k} \mathbf{E}\left[D_{1/\alpha^l}(b^*, b^l) - D_{1/\alpha^{l-1}}(b^*, b^l)\right] \\ =&\sum_{l=1}^{k} \mathbf{E}\left[\sum_i \left(\frac{1}{\alpha_i^l} - \frac{1}{\alpha_i^{l-1}}\right) D_i\left(b_i^*, b_i^l\right)\right] \\ =&\sum_i \mathbf{E}\left[\sum_{l=1}^{k} \left(\frac{1}{\alpha_i^l} - \frac{1}{\alpha_i^{l-1}}\right) D_i\left(b_i^*, b_i^l\right)\right] \\ =&\sum_i \mathbf{E}\left[\sum_{l=1}^{k} \frac{1}{\alpha_i^{l-1}}\left(D_i(b_i^*, b_i^{l-1}) - D_i(b_i^*, b_i^l)\right)\right] - \mathbf{E}\left[\frac{1}{\alpha_i^0} D_i(b_i^*, b_i^0)\right] + \mathbf{E}\left[\frac{1}{\alpha_i^k} D_i(b_i^*, b_i^k)\right]. \end{aligned} \tag{74}$$

□

用于PR动力学的自适应块坐标下降算法（A-BCPR）

我们还给出了一种自适应步长策略。为此，我们找到一系列 $\{L_i^k\}_{k\in\mathbb{N}}$，使得在第 k次迭代时，每个 $\{L_i^k\}_{k\in\mathbb{N}}$ 都满足与 $L_i^k < 1$ 的坐标相对光滑性。我们陈述以下引理来支撑该算法。

引理 8. 令 b_{ij}^{k+1} 为 *BCPR* 中在 $1 \leq \alpha_i^k \leq \bar{\alpha}_i^k$、 $\forall\, i$、 k 条件下由 b_{ij}^k 得到的下一个迭代点，其中 1 和 $\bar{\alpha}_i^k$ 分别为步长的下界和上界。令

$$\beta_i^k = \left(\max_{j: b_{ij}^* > 0} \frac{v_{ij}}{p_j^k} \Big/ \min_{j: b_{ij}^* > 0} \frac{v_{ij}}{p_j^k}\right)^{\bar{\alpha}_i} \tag{75}$$

and

$$\theta_i^k = \max_j \frac{b_{ij}^k}{p_j^k}. \tag{76}$$

当 $1 \le \beta_i^k \le \sqrt{2}$ 时，我们有

$$\varphi(b^{k+1}) \le \varphi(b^k) + \langle \nabla_i \varphi(b^k), b_i^{k+1} - b_i^k \rangle + L_i^k D_{\mathrm{KL}}(b_i^{k+1}, b_i^k) \tag{77}$$

其中

$$L_i^k = \frac{3}{4-\beta_i^k}\left(\theta_i^k + \frac{2\beta_i^k - 1}{6\beta_i^k}{\theta_i^k}^2\right). \tag{78}$$

证明。令 $c_j = p_j^+ - b_{ij}^+ = p_j - b_{ij}$。 $D_{\mathrm{KL}}(c_j + b_{ij}^+, c_j + b_{ij})$ 的泰勒级数，其余项采用拉格朗日形式 is

$$D_{\mathrm{KL}}(c_j + b_{ij}^+, c_j + b_{ij}) = (c_j + b_{ij}) \log \frac{c_j + b_{ij}^+}{c_j + b_{ij}} - b_{ij}^+ + b_{ij} = \frac{\left(b_{ij}^+ - b_{ij}\right)^2}{2\left(c_j + b_{ij}\right)} - \frac{\left(b_{ij}^+ - b_{ij}\right)^3}{6\left(c_j + b_{ij}\right)^2} + \frac{\left(b_{ij}^+ - b_{ij}\right)^4}{12\left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^3} \tag{79}$$

对于某个 $\lambda \in [0,1]$。

我们假设预算均为单位值，即 $\sum_j b_{ij} = 1, \forall i$。由于 $1 \le \alpha_i \le \bar{\alpha}_i$，我们可以推导出

$$\frac{b_{ij}^+}{b_{ij}} = \frac{\left(\frac{v_{ij}}{p_j}\right)^{\alpha_i}}{\sum_l b_{il}\left(\frac{v_{il}}{p_l}\right)^{\alpha_i}} \le \frac{\max_j \left(\frac{v_{ij}}{p_j}\right)^{\alpha_i}}{\min_j \left(\frac{v_{ij}}{p_j}\right)^{\alpha_i}} = \left(\frac{\max_j \frac{v_{ij}}{p_j}}{\min_j \frac{v_{ij}}{p_j}}\right)^{\alpha_i} \le \left(\frac{\max_j \frac{v_{ij}}{p_j}}{\min_j \frac{v_{ij}}{p_j}}\right)^{\bar{\alpha}_i} = \beta_i. \tag{80}$$

类似地，我们有

$$\frac{b_{ij}^+}{b_{ij}} = \frac{\left(\frac{v_{ij}}{p_j}\right)^{\alpha_i}}{\sum_l b_{il}\left(\frac{v_{il}}{p_l}\right)^{\alpha_i}} \ge \frac{\min_j \left(\frac{v_{ij}}{p_j}\right)^{\alpha_i}}{\max_j \left(\frac{v_{ij}}{p_j}\right)^{\alpha_i}} = \left(\frac{\min_j \frac{v_{ij}}{p_j}}{\max_j \frac{v_{ij}}{p_j}}\right)^{\alpha_i} \ge \left(\frac{\min_j \frac{v_{ij}}{p_j}}{\max_j \frac{v_{ij}}{p_j}}\right)^{\bar{\alpha}_i} = \frac{1}{\beta_i}. \tag{81}$$

于是，我们得到

$$\begin{aligned}\frac{\left(b_{ij}^+ - b_{ij}\right)^2}{2\left(c_j + b_{ij}\right)} - \frac{\left(b_{ij}^+ - b_{ij}\right)^3}{6\left(c_j + b_{ij}\right)^2} + \frac{\left(b_{ij}^+ - b_{ij}\right)^4}{12\left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^3} &\ge \frac{\left(b_{ij}^+ - b_{ij}\right)^2}{2\left(c_j + b_{ij}\right)} - \frac{\left(b_{ij}^+ - b_{ij}\right)^3}{6\left(c_j + b_{ij}\right)^2} \\ &\ge \left(\frac{1}{2(c_j + b_{ij})} - \frac{(\beta_i - 1) b_{ij}}{6(c_j + b_{ij})^2}\right)(b_{ij}^+ - b_{ij})^2. \end{aligned} \tag{82}$$

另一方面，当$1 < \beta_i \le \sqrt{2}$时，我们有，

$$\frac{\left(b_{ij}^+ - b_{ij}\right)^2}{2\left(c_j + b_{ij}\right)} - \frac{\left(b_{ij}^+ - b_{ij}\right)^3}{6\left(c_j + b_{ij}\right)^2} + \frac{\left(b_{ij}^+ - b_{ij}\right)^4}{12\left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^3} \le \left(\frac{1}{2(c_j + b_{ij})} + \frac{(1 - \frac{1}{2\beta_i}) b_{ij}}{6(c_j + b_{ij})^2}\right)(b_{ij}^+ - b_{ij})^2. \tag{83}$$

为验证 (83)，我们可以重新整理该式，并检验以下不等式是否与之等价：

$$\left(b_{ij}^+ - b_{ij}\right)^2 (c_j + b_{ij})^2 \le 2\left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^3 \left(b_{ij}^+ - \frac{1}{2\beta_i} b_{ij}\right). \tag{84}$$

通过考虑两种情况，可以验证上述不等式。若 $b_{ij}^+ \ge b_{ij}$,

$$\begin{aligned}\left(b_{ij}^+ - b_{ij}\right)^2 (c_j + b_{ij})^2 &\le \left(b_{ij}^+ - b_{ij}\right)^2 \left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^2 \\ &= \left(b_{ij}^+ - b_{ij}\right)\left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^2 \left(b_{ij}^+ - b_{ij}\right) \\ &\le 2\left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^3 \left(b_{ij}^+ - \frac{1}{2\beta_i} b_{ij}\right) \end{aligned} \tag{85}$$

其中，当 $1 < \beta_i \le 3$ 时，我们使用了 $b_{ij}^+ - b_{ij} \le 2\left(c_j + b_{ij} + \lambda(b_{ij}^+ - b_{ij})\right)$ 和 $b_{ij}^+ - b_{ij} \le b_{ij}^+ - \frac{1}{2\beta_i} b_{ij}$。

若$b_{ij}^+ < b_{ij}$,

$$\begin{aligned}\left(b_{ij}^+ - b_{ij}\right)^2 (c_j + b_{ij})^2 \le b_{ij}^+ (c_j + b_{ij})^2 \left(b_{ij} - b_{ij}^+\right) &\le \left(c_j + b_{ij}^+\right)\left(c_j + \beta_i b_{ij}^+\right)^2 \left(b_{ij} - b_{ij}^+\right) \\ &\le \beta_i^2 \left(c_j + b_{ij}^+\right)^3 \frac{2}{\beta_i^2}\left(b_{ij}^+ - \frac{1}{2\beta_i} b_{ij}\right) \\ &\le 2\left(c_j + \lambda b_{ij}^+ + (1-\lambda) b_{ij}\right)^3 \left(b_{ij}^+ - \frac{1}{2\beta_i} b_{ij}\right) \end{aligned} \tag{86}$$

where we use $b_{ij}-b_{ij}^{+}\le b_{ij}^{+}$ holds when $1<\beta_i\le 2$, and $b_{ij}-b_{ij}^{+}\le \frac{2}{\beta_i^2}\left(b_{ij}^{+}-\frac{1}{2\beta_i}b_{ij}\right)$ holds when $1<\beta_i\le\sqrt{2}$.

结合 (79)、(82) 和 (83)，我们得到

$$\begin{aligned}\frac{\varphi(b^{+})-\varphi(b)-\langle\nabla_i\varphi(b),b_i^{+}-b_i\rangle}{D_{\mathrm{KL}}(b_i^{+},b_i)}=&\sum_j D_{\mathrm{KL}}(c_j+b_{ij}^{+},c_j+b_{ij})\Big/\sum_j D_{\mathrm{KL}}(b_{ij}^{+},b_{ij})\\ \le&\sum_j\left(\frac{1}{2(c_j+b_{ij})}+\frac{(1-\frac{1}{2\beta_i})b_{ij}}{6(c_j+b_{ij})^2}\right)(b_{ij}^{+}-b_{ij})^2\Big/\sum_j\left(\frac{1}{2b_{ij}}-\frac{(\beta_i-1)}{6b_{ij}}\right)(b_{ij}^{+}-b_{ij})^2\\ \le&\max_j\left(\frac{1}{2(c_j+b_{ij})}+\frac{(1-\frac{1}{2\beta_i})b_{ij}}{6(c_j+b_{ij})^2}\right)\Big/\left(\frac{1}{2b_{ij}}-\frac{(\beta_i-1)}{6b_{ij}}\right)\\ =&\max_j\frac{3}{4-\beta_i}\left(\frac{b_{ij}}{p_j}+\frac{2\beta_i-1}{6\beta_i}\frac{b_{ij}^2}{p_j^2}\right)\end{aligned}\tag{87}$$

其中，第二个不等式成立是因为两个求和中的所有项均为非负数。由于（87）式右侧随 $\frac{b_{ij}}{p_j}$ 增大，引理 8 得证。

— □

注意，根据买方最优性条件，当 $b_{ij}^{*}>0, v_{ij}/p_j^{*}=\max_j v_{ij}/p_j^{*}$ 时，可得 $\beta_i^{*}=1$。因此，引理 8 有意义。

我们将比例响应动态的自适应块坐标下降算法表述如下。

算法 7： 自适应块坐标比例响应（A-BCPR）

输入： $b^0\in\mathbb{R}^{n\times m},\bar{\alpha}_1,\dots,\bar{\alpha}_n\in\mathbb{R}^{+}$

对于 $k\leftarrow 1,2,\cdots$ **执行**

- 以概率 $1/n$选择 $i_k\in[n]$；
- 根据 (75)、(76) 和 (78) 计算 L_i^k；
- $\alpha_{i_k}\leftarrow$最小{最大$\{1/L_i^k,1\}$，$\bar{\alpha}_i\}$；
- 使用步长 α_{i_k}根据 (14) 计算 $b_{i_k}^{+}$；
- $p^k\leftarrow p^{k-1}-b_{ik}+b_{ik}^{+}$；
- $b_{i_k}^{(k)}\leftarrow b_{i_k}^{+}$ and $b_{i'}^{(k)}\leftarrow b_{i'}^{(k-1)}\,\forall\, i'\neq i_k$；

D 采用线搜索的比例响应

我们在此给出一般的带线搜索的镜像下降（MD-LS）。

算法 8：线搜索 $(b^{k+1},\ \alpha_k,\ p_k) \leftarrow \mathcal{LS}_{\rho^+,\ \rho^-,\ \alpha_{\max}}(b,\ \alpha,\ p_{\text{prev}})$ 输入：当前迭代点 b，初始步长 α，上一次迭代中的线搜索步数 p_{prev}。输出：当前步长 α_k，下一迭代点 b^{k+1}，当前线搜索步数 p_k。如果 $p_{\text{prev}} = 0$（上一次迭代中未进行回溯搜索），则设置 $\alpha^{(0)} = \min\{\rho^+\alpha, \alpha_{\max}\}$（增加步长）；否则设置 $\alpha^{(0)} = \alpha$；对于 $p \leftarrow 0$，1，2，.....，执行以下操作：1. 计算 $b^{(p)} = \arg\min_{b' \in \mathcal{X}} \left\{ \langle \nabla\varphi(b)\ b' - b\rangle + \frac{1}{\alpha^{(p)}} D(b',\ b) \right\}$。2. 如果 $\varphi(b^{(p)}) \leq \varphi(b) + \langle \nabla\varphi(b)\ b^{(p)} - b\rangle + \frac{1}{\alpha^{(p)}} D(b^{(p)},\ b)$，则退出。3. 设置 $\alpha^{(p+1)} = \rho^-\alpha^{(p)}$ 并继续。

返回 $b^{k+1} = b^{(p)}, \alpha_k = \alpha^{(p)}, p_k = p$。

算法 9： 带线搜索的镜像下降（MD-LS）

输入： 初始迭代点 $b^0 \in \text{relint}(\mathcal{X})$，初始步长 $\alpha^{-1} \leq \alpha_{\max}, p_{-1} = 0$**对于** $k \leftarrow 0$, 1, 2, ... **执行**使用算法 8 计算 $(b^{k+1}, \alpha_k, p_k) = \mathcal{LS}_{\rho^+,\rho^-,\alpha_{\max}}(b^k, \alpha_{k-1}, p_{k-1})$。

定理 3 的证明

证明。 我们证明次线性收 敛性适用于一般的带线搜索的镜像下降，这立即推出定理 3 as w我们知道 w $L = 1$对于 (S).

与 Birnbaum、Devanur 和 Xiao（2011，引理 5）中的情况相同，不难看出，算法 9 仍然是一个下降目标算法，即

$$\varphi(b^{k+1}) \leq \varphi(b^k),\ k \in \mathbb{N}. \tag{88}$$

由 f 的凸性以及算法 8 中的中止条件可知

$$\varphi(b^k) + \langle \nabla\varphi(b^k), b^{k+1} - b^k\rangle \leq \varphi(b^{k+1}) \leq \varphi(b^k) + \langle \nabla\varphi(b^k), b^{k+1} - b^k\rangle + \frac{1}{\alpha^k} D(b^{k+1}, b^k). \tag{89}$$

利用 Birnbaum、Devanur、 和 Xiao（2011，引理 9）（最初见于 Chen 和 Teboulle（1993）），对任意 $u \in \mathcal{X}$，有 ，

$$\alpha^k \left(\varphi(b^k) + \langle \nabla\varphi(b^k), b^{k+1} - b^k\rangle\right) + D(b^{k+1}, b^k) + D(u, b^k) \leq \alpha^k \left(\varphi(b^k) + \langle \nabla\varphi(b^k), u - b^k\rangle\right) + D(u, b^k) \tag{90}$$

然后

$$\langle \nabla\varphi(b^k), b^{k+1} - b^k\rangle + \frac{1}{\alpha^k} D(b^{k+1}, b^k) \leq \langle \nabla\varphi(b^k), b - b^k\rangle + \frac{1}{\alpha^k} \left(D(u, b^k) - D(u, b^k)\right). \tag{91}$$

上述结果与 Birnbaum、Devanur 和 Xiao（2011，推论 10）类似。设 b^* 是一个最优解。然后，类似于 Birnbaum、Devanur 和 Xiao（2011，引理 8），我们有

$$\begin{aligned}
\varphi(b^{k+1}) &\leq \varphi(b^k) + \langle \nabla\varphi(b^k), b^{k+1} - b^k\rangle + \frac{1}{\alpha^k} D(b^{k+1}, b^k) \\
&\leq \varphi(b^k) + \langle \nabla\varphi(b^k), b^* - b^k\rangle + \frac{1}{\alpha^k} \left(D(b^*, b^k) - D(b^*, b^{k+1})\right) \\
&\leq \varphi(b^*) + \frac{1}{\alpha^k} \left(D(b^*, b^k) - D(b^*, b^{k+1})\right).
\end{aligned} \tag{92}$$

这由此推出引理 3，因为

$$D(b^*, b^k) - D(b^*, b^{k+1}) \geq \alpha^k \left(\varphi(b^{k+1}) - \varphi(b^*)\right) \geq 0. \tag{93}$$

将 (92) 对 $k = 0$，1、.....求和，得到

$$k \left(\varphi(b^k) - \varphi^*\right) \leq \sum_{l=0}^{k-1} (\varphi(b^{l+1}) - \varphi^*) \leq \sum_{l=0}^{k-1} \frac{D(b^*, b^l) - D(b^*, b^{l+1})}{\alpha^l} \tag{94}$$

其中，第一个不等式由 (88) 得出。重新整理上式可得以下路径依赖界：

$$\varphi(b^k)-\varphi^* \le \frac{1}{k}\left(\frac{1}{\alpha_0}D(b^*,b^0)-\frac{1}{\alpha_k}D(b^*,b^k)+\sum\nolimits_{l=1}^{k-1}\left(\frac{1}{\alpha_l}-\frac{1}{\alpha_{l-1}}\right)D(b^*,b^l)\right). \tag{95}$$

根据引理 3 和 $\alpha^l \ge \frac{\rho_-}{L}$，可得（17）：

$$k\left(\varphi(b^k)-\varphi^*\right) \le \sum_{l=0}^{k-1}\frac{D(b^*,b^l)-D(b^*,b^{l+1})}{\alpha^l} \le \frac{L}{\rho_-}\sum_{l=0}^{k-1}\left(D(b^*,b^l)-D(b^*,b^{l+1})\right) \le \frac{L}{\rho_-}D(b^*,b^0). \tag{96}$$

□

E CES 效用函数的块坐标下降算法：CES 效用函数二次外推的详细说明

在本节中，我们用 u_i^ρ 替代 u_i（ρ），以避免混淆，因为 u_i（ρ）的决策变量是 x_i。将 CES 效用函数的 EG 凸规划的目标函数记为

$$f(x)=\sum_i f_i(x_i)=\sum_i g_i(u_i^\rho(x_i)) \quad \text{and} \quad u_i^\rho(x_i)=\sum_j u_{ij}^\rho(x_{ij}) \tag{97}$$

其中

$$g_i(u_i^\rho)=-\frac{B_i}{\rho}\log u_i^\rho \quad \text{and} \quad u_{ij}^\rho(x_{ij})=v_{ij}x_{ij}^\rho, \quad \rho\in(0,1). \tag{98}$$

然后，我们分别用 $\tilde{g}$ 和 $\tilde{u}^\rho$（定义如下）替换 g 和 u^ρ。与线性情形类似，

$$\tilde{g}_i(u_i^\rho):=\begin{cases}-\frac{B_i}{\rho}\log u_i^\rho & u_i^\rho \ge \underline{u}_i^\rho \\ -\frac{B_i}{\rho}\log \underline{u}_i^\rho-\frac{B_i}{\rho \underline{u}_i^\rho}\left(u_i^\rho-\underline{u}_i^\rho\right)+\frac{B_i}{2\rho \underline{u}_i^{\rho 2}}\left(u_i^\rho-\underline{u}_i^\rho\right)^2 & \text{otherwise}\end{cases}, \tag{99}$$

and

$$\tilde{u}_{ij}^\rho(x_{ij}):=\begin{cases}v_{ij}x_{ij}^\rho & x_{ij}\ge \underline{x}_{ij} \\ v_{ij}\underline{x}_{ij}^\rho+\rho v_{ij}\underline{x}_{ij}^{\rho-1}\left(x_{ij}-\underline{x}_{ij}\right)+\frac{1}{2}\rho(\rho-1)v_{ij}\underline{x}_{ij}^{\rho-2}\left(x_{ij}-\underline{x}_{ij}\right)^2 & \text{otherwise}\end{cases} \tag{100}$$

其中，$\underline{u}_i$ 和 $\underline{x}_{ij}$ 是 u_i^* 和 x_{ij}^* 的特定下界，适用于某个 i，j。根据 Gao 和 Kroer（2020，引理 1）中的类似讨论，我们知道对于所有 i，u_i 都存在。根据引理 4，我们知道对于所有 i，j，x_{ij} 都存在。

引理 4 的证明

这里，我们给出引理 4 的证明。该证明利用了均衡处的不动点性质，类似于 Zhang（2011，引理 8），但得到了更紧的界。

证明。考虑一个买方 i 和一个物品 j，满足 $v_{ij}>0$。如果对于买方 i，所有其他 $k\ne j$ 都与 $v_{ik}=0$相关联，那么显然有 $b_{ij}^*=B_i$ 和$x_{ij}^*=b_{ij}^*/p_j^*\ge B_i/\sum_l B_l$。

否则，如果对于某个 $k\ne j$，存在 $v_{ik}>0$，由于市场均衡是比例响应动态的不动点（自然也可推广到块坐标情形），我们有

$$\left(\frac{v_{ij}}{p_j^*}\right)^\rho {b_{ij}^*}^{\rho-1}=\left(\frac{v_{ik}}{p_k^*}\right)^\rho {b_{ik}^*}^{\rho-1}. \tag{101}$$

重排（101）后，我们得到

$$b_{ij}^*=\left(\frac{v_{ij}}{v_{ik}}\right)^{\frac{\rho}{1-\rho}} b_{ik}^*\left(\frac{p_j^*}{p_k^*}\right)^{\frac{\rho}{\rho-1}} \tag{102}$$

and

$$x_{ij}^*=\frac{b_{ij}^*}{p_j^*}=\left(\frac{v_{ij}}{v_{ik}}\right)^{\frac{\rho}{1-\rho}} b_{ik}^* {p_j^*}^{\frac{1}{\rho-1}} {p_k^*}^{\frac{\rho}{1-\rho}}. \tag{103}$$

具体而言，我们考虑 $b_{ij}^*<\frac{B_i}{m}$，因为否则下界可以直接得到。对于这些 i，j，我们总能选择另一个 k（如上所述，至少存在一个），使得 $b_{ik}^*\ge \frac{B_i}{m}$。因此，我们可以根据 $b_{ij}^*\le p_j^*$、(101) 和 $b_{ik}^*\le p_k^*$ 对 p_j^* 进行界定：

$$p_j^*={p_j^*}^{1-\rho}{p_j^*}^\rho \ge {b_{ij}^*}^{1-\rho}{p_j^*}^\rho=\left(\frac{v_{ij}}{v_{ik}}\right)^\rho {p_k^*}^\rho {b_{ik}^*}^{1-\rho}\ge\left(\frac{v_{ij}}{v_{ik}}\right)^\rho {b_{ik}^*}\ge \omega_2(i)^\rho\frac{B_i}{m} \tag{104}$$

其中 $\omega_2(i) = $最小值$_{j:v_{ij}>0}\, v_{ij}/$最大值$_j\, v_{ij}$。结合 (103) 和 (104)，我们得到

$$x_{ij}^* = \left(\frac{v_{ij}}{v_{ik}}\right)^{\frac{\rho}{1-\rho}} b_{ik}^* {p_j^*}^{\frac{1}{\rho-1}} {p_k^*}^{\frac{\rho}{1-\rho}} \geq \omega_2(i)^{\frac{\rho}{1-\rho}} \frac{B_i}{m} \left(\sum_l B_l\right)^{\frac{1}{\rho-1}} \left(\omega_2(i)^\rho \frac{B_i}{m}\right)^{\frac{\rho}{1-\rho}} = \omega_1(i)^{\frac{1}{1-\rho}} \omega_2(i)^{\frac{\rho(1+\rho)}{1-\rho}} \tag{105}$$

其中 $\omega_1(i) = \frac{B_i}{m\sum_l B_l}$。其他所有情况的下界都比这更大。 □

定理4的证明

我们使用以下引理，为 $f(x) = \sum_i \tilde{g}(\sum_j \tilde{u}_{ij}^\rho(x_{ij}))$ 建立光滑性保证。**引理 9**。 对于任意 $j \in [m]$，令

$$L_j = \max_{i\in[n]} \frac{B_i v_{ij} \underline{x}_{ij}^{\rho-2}}{\underline{u}_i(\rho)}. \tag{106}$$

那么，对于所有可行的 x、y，若 x 和 y 仅在第 j 个块上不同，则有

$$\tilde{f}(y) \leq \tilde{f}(x) + \langle \nabla_{\cdot j}\tilde{f}(x), y - x\rangle + \frac{L_j}{2}\|y-x\|^2. \tag{107}$$

证明。 为简洁起见，令 $\tilde{u}_i$ 表示 $\tilde{u}_i(\rho)$。我们可以得到

$$\begin{aligned}\nabla_{ij}\tilde{f} = \frac{\partial^2}{\partial x_{ij}^2}\tilde{f} = \frac{\partial^2 \tilde{g}_i}{\partial \tilde{u}_i^2}\left(\frac{\partial \tilde{u}_i}{\partial x_{ij}}\right)^2 + \frac{\partial \tilde{g}_i}{\partial \tilde{u}_i}\frac{\partial^2 \tilde{u}_i}{\partial x_{ij}^2} &\leq \frac{B_i}{\rho \underline{u}_i^2}\left(\rho v_{ij}\underline{x}_{ij}^{\rho-1}\right)^2 + \left(-\frac{B_i}{\rho \underline{u}_i}\right)\left(\rho(\rho-1)v_{ij}\underline{x}_{ij}^{\rho-2}\right) \\ &= \frac{B_i v_{ij}\underline{x}_{ij}^{\rho-2}}{\underline{u}_i}\left(\rho\frac{v_{ij}\underline{x}_{ij}^{\rho}}{\underline{u}_i} + (1-\rho)\right) \leq \frac{B_i v_{ij}\underline{x}_{ij}^{\rho-2}}{\underline{u}_i}.\end{aligned} \tag{108}$$

于是，(107) 可由与引理 1 相同的推导得出。 □

定理4可由 Richtárik 和 Taká˘c（2014，定理7）以及以下事实得到：当 $\tilde{f}$ 具有上述形式时，关于范数 $\sum_j L_j\|\cdot\|_{\cdot j}^2$， $\tilde{f}$ 存在强凸模量$\mu(L)$。

备注。 与引理 9 类似，我们可以证明其具有常数为 $L = \max_{i\in[n]} \frac{\rho B_i \|v_i\|^2}{u_i^2} x_i^{2\rho-2}$ 的标准 Lipschitz 光滑性。这可视为引理 5 的推广结果，因为我们有

$$\frac{\partial^2}{\partial x_{ij}\partial x_{ij'}}\tilde{f}(x) = \frac{\partial \tilde{u}_i}{\partial x_{ij}}\frac{\partial^2 \tilde{g}_i}{\partial \tilde{u}_i^2}\frac{\partial \tilde{u}_i}{\partial x_{ij'}} \leq \rho v_{ij}v_{ij'}\frac{B_i}{\underline{u}_i^2}\underline{x}_{ij}^{\rho-1}\underline{x}_{ij'}^{\rho-1}. \tag{109}$$

定理5的证明

我们使用 Zhang（2011）引入的一个辅助函数，他们利用该函数证明确定性 PR 的收敛性。证明。 在此证明中，$D_i(\cdot,\ \cdot)$ 表示决策变量第 i 个块的 KL 散度。定义

$$T_{ij}^k = B_i \frac{v_{ij}\left(\frac{b_{ij}^k}{p_j^k}\right)^\rho}{\sum_{j'} v_{ij'}\left(\frac{b_{ij'}^k}{p_{j'}^k}\right)^\rho}, \quad u_{ij}^k(\rho) = v_{ij}\left(\frac{b_{ij}^k}{p_j^k}\right)^\rho, \quad u_i^k(\rho) = \sum_j v_{ij}\left(\frac{b_{ij}^k}{p_j^k}\right)^\rho,$$

因此 $T_{ij}^k = B_i u_{ij}^k(\rho)/u_i^k(\rho)$。

设 $m(k,i) = D_i(b_i^*, b_i^k)$ 且 $m(k) = \sum_i m(k,i)$，则有

$$\mathbf{E}\left[m(k+1) \,\middle|\, b^k\right] = \sum_i \frac{1}{n}\left\{D_i(b_i^*, T_i^k) + \sum_{i'\neq i} m(k,i')\right\} = \frac{1}{n}\sum_i D_i(b_i^*, T_i^k) + \frac{n-1}{n}m(k). \tag{110}$$

与 Zhang（2011）类似，我们得到

$$\begin{aligned}D_i(b_i^*, T_i^k) = \sum_j b_{ij}^* \log\frac{b_{ij}^*}{T_{ij}^k} = \sum_j b_{ij}^* \log\frac{b_{ij}^* u_i^k(\rho)}{B_i u_{ij}^k(\rho)} &= \sum_j b_{ij}^* \log\frac{u_{ij}^*(\rho)u_i^k(\rho)}{u_i^*(\rho)u_{ij}^k(\rho)} \\ &= \rho\sum_j b_{ij}^* \log\frac{b_{ij}^*}{b_{ij}^k} - \rho\sum_j b_{ij}^* \log\frac{p_j^*}{p_j^k} - \sum_j b_{ij}^* \log\frac{u_i^*(\rho)}{u_i^k(\rho)}.\end{aligned} \tag{111}$$

因此，

$$\begin{aligned}\mathbf{E}\left[m(k+1)\mid b^k\right] &= \frac{1}{n}\sum_{i,j}\left\{\rho b_{ij}^*\log\frac{b_{ij}^*}{b_{ij}^k}-\rho b_{ij}^*\log\frac{p_j^*}{p_j^k}-b_{ij}^*\log\frac{u_i^*(\rho)}{u_i^k(\rho)}\right\}+\frac{n-1}{n}m(k)\\ &= \frac{\rho}{n}m(k)+\frac{n-1}{n}m(k)-\frac{\rho}{n}\sum_j p_j^*\log\frac{p_j^*}{p_j^k}-\frac{1}{n}\sum_i B_i\log\frac{u_i^*(\rho)}{u_i^k(\rho)}\\ &= \left(1-\frac{1-\rho}{n}\right)m(k)-\frac{\rho}{n}D_{\mathrm{KL}}(p^*,p^k)-\frac{1}{n}\sum_i B_i\log\frac{\langle v_i,x_i^{*\rho}\rangle}{\langle v_i,x_i^{k\rho}\rangle}. \end{aligned}\tag{112}$$

由于 $D_{\mathrm{KL}}(p^*,\ p^k)\ge 0$ 且 $\sum_i \frac{B_i}{\rho}\log\frac{\langle v_i,x_i^{*\rho}\rangle}{\langle v_i,x_i^{k\rho}\rangle}\ge 0$（由 EG 程序的最优性可知），我们有

$$\mathbf{E}\left[m(k+1)\mid b^k\right]\le \rho' m(k),\quad \rho'=1-\frac{1-\rho}{n}.\tag{113}$$

对 $i_1,\ i_2,\ \ldots\ldots,\ i_k$取期望后，得到 $\mathbf{E}[m(k)]\le(\rho')^k m(0)$。

注意，我们有以下引理。

引理 10。（*Zhang 2011*，引理 *9*）对于两个正序列 $\{b_{ij}^*\}_{i\in[n],\ j\in[m]}$ 和 $\{b_{ij}\}_{i\in[n],\ j\in[m]}$，满足 $\sum_{i,\ j} b_{ij}^* =\sum_{i,\ j} b_{ij}$，令 $\eta=\max_{i,\ j}\frac{|b_{ij}-b_{ij}^*|}{b_{ij}^*}$。则，

$$D_{\mathrm{KL}}(b^*,b)\ge\frac{1}{8}\min\{1,\eta\}\eta\min_{i,j}b_{ij}^*.\tag{114}$$

我们考虑 $\eta<1$ 的情形。根据 Zhang（2011，引理 8），我们知道，对于任意 $i,\ j$，只要 $v_{ij}>0$，就有 $W=\frac{n\min_{v_{ij}>0}v_{ij}\cdot\min_i B_i}{}$。然后，由（114）可知

$$m(k)=D_{\mathrm{KL}}(b^*,b)\ge\frac{1}{8}\eta^2\min_{i,j}b_{ij}^*\ge\frac{1}{8}\eta^2\left(\frac{1}{W^2}\right)^{\frac{1}{1-\rho}}=\frac{1}{8}\left(\frac{|b_{ij}-b_{ij}^*|}{b_{ij}^*}\right)^2\left(\frac{1}{W^2}\right)^{\frac{1}{1-\rho}}.\tag{115}$$

对上述不等式两边取期望，并应用詹森不等式，可得

$$\left(\mathbf{E}\left[\frac{|b_{ij}-b_{ij}^*|}{b_{ij}^*}\right]\right)^2\le\mathbf{E}\left(\frac{|b_{ij}-b_{ij}^*|}{b_{ij}^*}\right)^2\le 8W^{\frac{2}{1-\rho}}\mathbf{E}[m(k)].\tag{116}$$

因此，为了保证 $\mathbf{E}\frac{|b_{ij}-b_{ij}^*|}{b_{ij}^*}\le\epsilon<1$，我们需要 $8W^{\frac{2}{1-\rho}}\mathbf{E}[m(k)]\le\epsilon^2$。由于 $\mathbf{E}[m(k)]\le(\rho')^k m(0)$，只需满足

$$(\rho')^k m(0)\le\frac{1}{8}\left(\frac{1}{W^2}\right)^{\frac{1}{1-\rho}}\epsilon^2,\tag{117}$$

这等价于 $k\ge 2\log\frac{\sqrt{8m(0)}W^{\frac{1}{1-\rho}}}{\epsilon}\Big/\log\frac{n}{n-1+\rho}$。

□

关于 CES 效用的块坐标下降方法及更多实验结果的讨论

我们证明，对于线性效用，可以通过应用坐标下降算法高效求解 EG 程序。与线性效用不同，使用块坐标下降类方法计算 CES（$\rho\in(0,\ 1)$）效用对应的市场均衡时，会遇到一些困难。

关键问题在于，尽管我们可以得到 x_{ij}^*（对于那些 $i,\ j$：$v_{ij}>0$）的下界，但这个下界太小，无法产生合理的（逐坐标）Lipschitz 常数，因为该下界出现在 L_j 的界中（见 (106)）。

尽管我们的算法配备了线搜索策略，但我们发现它们仍难以找到高效的步长。我们甚至发现很难选择合适的初始步长。步长往往太小，无法达到均衡，尤其是在迭代次数不多时。因此，BCDEG-LS 和 PGLS 的表现不如 BCPR 和 PR，因为后两者无需选择步长；它们在步长为 1 时仍能（次线性）收敛。

当估值波动性较小时，这一困难可以克服——在这种情况下，方程 (106) 中 x 的下界更大。关于 CES 效用的更多实验结果，请参见图 4。我们使用了模型 $v_{ij}=v_i v_j+\epsilon$，其中 $v_i\sim\mathcal{N}(1,\ \hat{}v)$、$v_j\sim\mathcal{N}(1,\ \hat{}v)$ 和 $\epsilon\sim$ 均匀分布$(0,\ \hat{}v)$，其中 $\hat{}v$ 是用于控制波动性的参数。这里，我们设定 $n=m=200$。

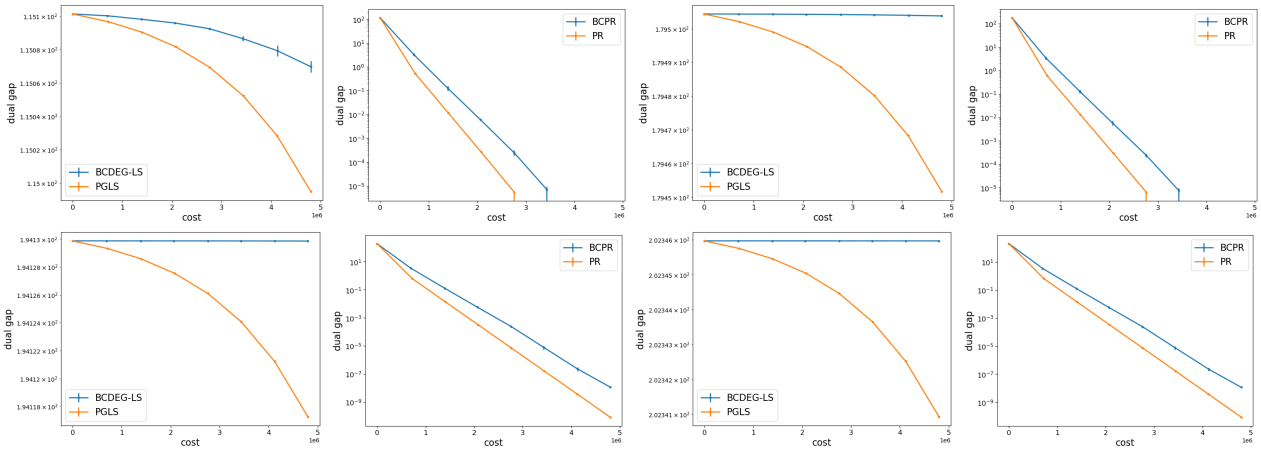

图 4：采用 CES 效用的更多模拟实例上的性能。（从上到下依次为 $\hat{}v = 0.4$，0。6，0。8，1。0）。图的设置与图 1 相同。

F 墙钟运行时间讨论

本文根据不同算法查询买家-物品估值的次数（即访问估值矩阵单元格的次数）对它们进行比较。从这个角度看，如果将一次查询计为一个单位成本，那么物品块算法（BCDEG、BCDEG-LS）在每次迭代中的成本为 n 个单位，买家块算法（BCPR、A-BCPR、BCPR-LS）在每次迭代中的成本为 m 个单位，而全梯度算法（PGLS、PR）在每次迭代中的成本为 nm 个单位。比较这些算法的另一种方式是比较墙钟运行时间。在本节中，我们说明为何选择以估值访问次数作为衡量标准。

对于投影梯度类方法（全确定性方法和 BCDEG 变体），每种方法每完成 n 单位的工作就执行一次 n 维投影，因此所有这些方法都可以直接根据我们的度量进行比较。同样，所有比例响应类算法每进行 m 次查询，完成的工作量都相同。因此，它们也可以直接根据我们的度量进行比较。将投影梯度类方法与 PR 类方法进行比较时，两者的区别在于前者需要进行投影。不过，实践中已知，最佳投影类算法的高效实现往往也能在线性时间内运行（Condat 2016）。因此，跨类别比较也应当是公平的。墙钟运行时间可能与我们的度量不同，主要是因为实现效率存在差异。我们的所有实验均使用 Python 实现。因此，就墙钟时间而言，确定性方法可能比块坐标方法更有优势，因为 NumPy 和 MKL 能够实现高度优化的矩阵/向量运算。换言之，确定性方法可以将更多时间花在“C 语言中”，而不是“Python 中”。如果所有算法的实现都没有采用利用多核 CPU 的高度优化矩阵运算，那么墙钟运行时间的实验结果将与我们基于查询的复杂度成本一致。同样，如果所有算法都经过充分优化，以充分利用硬件、并行化等，我们仍可以预期得到相似的结果。