arXiv:2304.01871v1 [math.OC] 4 Apr 2023

一种更快的指标算法及带切换成本的多臂老虎机计算研究

José Niño-Mora 统计学系 马德里卡洛斯三世大学
28903赫塔费（马德里），西班牙jose.nino@uc3m.es,
http://alum.mit.edu/www/jnimoraORCID：
0000-0002-2172-3983

发表于*INFORMS Journal on Computing*，第20卷，第255–269页，2008年
DOI: 10.1287/ijoc.1070.0238

摘要

我们研究带切换成本的难解多臂老虎机问题。Asawa和Teneketzis针对该问题提出了一种指标，该指标部分刻画了最优策略：为每个项目状态赋予一个“延续指数”（即其吉廷斯指数）和一个“切换指数”。他们建议将两者作为一个具有 $2n$ 个状态的项目的吉廷斯指数联合计算，而原项目有 n 个状态；这使得所需算术运算量相较于计算延续指数增加了八倍，即达到 $O(n^3)$。我们提出一种更快的解耦计算方法：第一阶段计算延续指数，第二阶段计算切换指数，所需算术运算量最多为 $n^2+O(n)$，速度提高一个数量级；总体算术运算量减少四倍，内存操作也大幅减少。分析利用了这样一个事实：Asawa和Teneketzis提出的指标，是项目在其休闲赌博机重新表述中的边际生产率指标；相关方法由作者提出。文中报告了大量计算实验，展示了新算法带来的显著运行时间加速，以及所得指数策略近似最优，并表明与基准吉廷斯指标策略相比，该策略在广泛随机生成的双项目和三项目实例中取得了显著优势。

关键词： 动态规划，马尔可夫，有限状态；老虎机，休闲型；切换成本；指标策略；边际生产率指标；算法分析；计算复杂度**MSC（2020）：** 90B36, 90C39, 90C40

1. 引言

吉廷斯指数为 吉廷斯指数 提供了一种可处理的 多臂老虎机问题 （MABP）求解方案。该问题要求为有限个随机项目寻找最优的序贯努力分配策略：在无限时域内的每个离散时段，都必须选择一个项目参与。项目被建模为 老虎机，即二元行动（1：主动；0：被动）的 马尔可夫决策过程（MDP），且只有在主动时状态才会改变。最优策略是在每个时段选择指标最大的项目，其中指标针对每个项目分别定义为其状态的函数。参见 Gittins 和 Jones （1974）以及Gittins （1979）。

这类结果背后的一个关键假设是，切换项目不产生任何成本。然而，正如 Banks和 Sundaram(1994年所指出的：“很难想象存在这样一种相关的经济决策问题，决策者可以在不同选项之间毫无成本地切换。”将此类成本纳入考量后，便得到带切换成本的多臂老虎机问题(MABPSC)。例如，参见 Agrawal 等人(1988年)、Van Oyen 和 Teneketzis (1994年)，以及 Jun(2004年的综述。

鉴于 Gittins 和 Jones 对 MABP 的研究结果，为 MABPSC 设计良好的指数策略颇具吸引力。此类策略会为每个项目 m 分配一个指标 $\nu_m(a_m^-,\ i_m)$，该指标是其先前行动 a_m^- 和当前状态 i_m 的函数，因此可分解为“继续指数” $\nu_m(1,\ i_m)$ 和“切换指数” $\nu_m(0,\ i_m)$。指数策略规定，在每个时点选择当前指标最大的项目。

尽管指标策略通常不是 MABPSC 的最优策略（参见 Banks和Sundaram（1994年）），Asawa 和 Teneketzis （1996年）提出了一个能够部分刻画最优策略的指标。他们的延续指数是吉廷斯指数，而他们的切换指数则是：通过使一个最初处于非激活状态的项目参与其中的停止规则，在单位期望折现时间内可实现的期望折现奖励减去切换成本的最大比率。

Asawa 和 Teneketzis （1996年）提出通过以下方式联合计算这两个指标：(i) 构造一个经过修改的项目，不含切换成本，但具有两倍数量的状态——即（$a_m^-,\ i_m$）；以及 (ii) 计算后者的吉廷斯指数。虽然已有多种算法可用于计算吉廷斯指数，但目前对于一般的 n状态项目，最佳复杂度由 快速枢轴算法实现，该算法由 Niño-Mora （2007a）提出，执行（2/3）n^3+O（n^2）次算术运算。然而，利用该算法通过他们的方案计算 Asawa 和 Teneketzis（AT）指标时，运算次数达到（2/3）（$2n$）^3+O（n^2）：相较于单独计算延续指数所需的工作量增加了八倍，这严重阻碍了其在大规模模型中的应用。因此，有必要开发一种显著更快的计算方法，这正是本文的首要目标。

我们通过一条看似迂回的途径实现这一目标：将经典的带切换成本的多臂老虎机重新表述为一种无休止多臂赌博机——即在被动时也能改变状态的多臂老虎机——且不含切换成本。这样，我们便可以采用 Whittle （1988年）提出、并由作者进一步发展的无休止多臂赌博机指标化方法（参见 Niño-Mora （2007b）），该方法为一类称为可指标化的受限多臂老虎机定义了边际生产率指标（MPI）。这构成了针对难以求解的多臂无休止赌博机问题（MARBP）的启发式指标策略的基础。

我们证明并利用了这样一个事实：对于带切换成本的多臂老虎机，其MPI与AT指标相同。根据 Niño-Mora（2002、2006年）对MPI的经济学刻画，即将其视为项目上工作边际生产率的状态依赖度量，这为将此类指标策略用作MABPSC的启发式方法提供了具有经济学意义的依据。基于MPI的优先级策略会动态地将精力分配到当前看起来生产率更高的项目上，并将各项目的MPI作为边际生产率的代理指标——因为它忽略了相互作用。

虽然 Asawa 和 Teneketzis （1996年）研究的是启动成本恒定的项目，但我们允许启动成本和停机成本依赖于状态，前提是两者之和在每个状态下均非负。

这确保延续指数至少与切换指数一样大，这与最优策略中直观的 滞后性质一致；这一性质由 Banks 和 Sundaram （1994年）指出："显而易见，在比较两个其他方面完全相同的臂时，如果其中一个在前一时期被使用过，那么曾在使用中的那个必然比闲置的那个更具吸引力。"另请参见 Dusonchet 和 Hongler （2003）。

我们采用 Niño-Mora (2001、2002、2006年) 提出的方法，证明我们关注的无休止老虎机是 *PCL* 可指数化的——它们满足部分守恒定律(PCL)，因此我们可以使用相关著作中提出的自适应贪心算法计算 MPI。我们进一步将该算法分解为一种更快的两阶段方法：第一阶段计算延续指数及其他量；第二阶段则利用第一阶段的输出计算切换指数。

要实现这一方案，第一阶段可以采用 Niño-Mora(2007a) 中的若干算法，例如 扩展输出快速选主元算法 FP(1)，其执行 $4/3$ 次n^3 +$O(n^2)$ 算术运算。对于第二阶段，我们提出一种切换指数算法，执行至多 $n^2 + O(n)$ 次算术运算，使计算切换指数所需的额外工作量微不足道。与 AT 方案相比，我们的两阶段方法使总体算术运算量减少了四倍，同时大幅减少了内存操作。该算法是本文的主要贡献。

一项计算研究表明，这种复杂度的改进带来了显著的运行时间加速。研究还通过一系列实验加以补充，结果显示，在大量双项目和三项目实例中，指数策略近乎最优，且相较于基准 Gittins 指数策略具有显著优势。

第2 节介绍模型及其无休止多臂赌博机重构。第3 节回顾无休止多臂赌博机指标化，给出基于动态规划 (DP) 的可指标化证明，说明 MPI 正是 AT指标，并概述后续采用的方法。第4节对经无休止重构的含切换成本项目进行 PCL-可索引性分析。第5 节基于这项分析开发新的解耦 MPI 算法。第7 节报告计算研究的结果。第8 节作出总结。

2. MABPSC与无休止多臂赌博机重构

2.1 模型构建

接下来，我们将介绍第1节概述的 MABPSC。考虑由 M 个有限状态项目构成的集合，在无限时域内的每个离散时段 $t = 0$、1、2、……，必须选择其中一个项目进行操作（主动），其余项目则处于休息状态（被动）。当项目 m 处于状态 i_m （属于其状态空间 N_m ）并被操作时，会产生主动奖励 $R^1_m(i_m) = R_m(i_m)$，其状态以概率 $p_m(i_m, j_m)$ 转移至 j_m 。如果项目处于休息状态，则会产生为零的被动奖励 $R^0_m(i_m) = 0$ ，且其状态保持不变。

切换项目会产生费用。当项目 m 处于状态 i_m 并被新启动（或处于休息状态）时，将产生启动成本 $c_m(i_m)$（或停机成本 $d_m(i_m)$。我们假设 $c_m(i_m) + d_m(i_m) \geq 0$。奖励和成本均按因子 $0 < \beta < 1$进行时间折现。

通过采用从调度策略 π中选择的策略来决定行动，该策略取自可行策略类 $\mathbf{\Pi}$ ，相对于状态和行动的历史而言是非预见性的，并且每次只启动一个项目。MABPSC 的目标是找到一种可行策略，使所得奖励减去所产生切换成本后的预期折现总值最大化。

记 $X_m(t) \in N_m$ 和 $a_m(t) \in \{0, 1\}$ 分别表示项目 m 在时刻的状态和动作 t,

。我们将使用以下记号

$$a_m^-(t) \triangleq a_m(t-1), \quad \bar{a}_m(t) \triangleq 1 - a_m(t), \quad \text{and} \quad \bar{a}_m^-(t) \triangleq \bar{a}_m(t-1). \tag{1}$$

由于初始条件必须指明每个项目 m 是否已处于初始设置状态，我们用 $a_m^-(0)$ 表示该状态。进一步定义项目的 增广状态 $\hat{}\ X_m(t) \triangleq (a_m^-(t)\ X_m(t))$，其在 增广状态空间 $\hat{}\ N_m$, $\{0,1\} \times N_m$上转移。因此，系统的联合增广状态 为 $\hat{}\mathbf{X}(t) \triangleq (\hat{}\ X_m(t))_{m=1}^M$ ，其联合动作过程 为 $\mathbf{a}(t) \triangleq (a_m(t))_{m=1}^M$。

给定联合初始状态 $\hat{}\mathbf{X}(0) = \hat{}\boldsymbol{\imath} = ((a_m^-, i_m))_{m=1}^M$ ，我们可以将 MABPSC 表述为

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \mathsf{E}_{\hat{\boldsymbol{\imath}}}^{\boldsymbol{\pi}} \left[\sum_{m=1}^{M} \sum_{t=0}^{\infty} \left\{ R_m^{a_m(t)} \left(X_m(t) \right) - c_m \left(X_m(t) \right) \bar{a}_m^-(t) a_m(t) - d_m \left(X_m(t) \right) a_m^-(t) \bar{a}_m(t) \right\} \beta^t \right], \tag{2}$$

其中 $\mathsf{E}_{\hat{\boldsymbol{\imath}}}^{\boldsymbol{\pi}}[\cdot\ ]$ 表示在初始联合状态等于$\hat{}\boldsymbol{\imath}$的条件下， $\boldsymbol{\pi}$ 下的期望。

2.2 无休止多臂赌博机重构

将 $\hat{}\ X_m(t)$ 视为项目 m 的状态，可将 (2) 重构为不含切换成本的 MARBP（参见第1节）。该重构中的奖励和转移概率如下。若项目处于增广状态 (a_m^-, i_m) 并被激活，则获得主动奖励 $\hat{}R_m^1(a_m^-, i_m)$ $R_m^1(i_m) - c_m(i_m)\bar{a}_m^-$ ，其增广状态以主动转移概率 $\hat{}p_m^1((a_m^-, i_m), (1, j_m))$, $p_m(i_m, j_m)$ 转移至 $(1, j_m)$。若项目处于休息状态，则获得单期被动奖励 $\hat{}R_m^0(a_m^-, i_m)$ $R_m^0(i_m) - d_m(i_m)a_m^-$ ，其增广状态以单位被动转移概率转移至 $(0, i_m)$： $\hat{}p_m^0((a_m^-, i_m),(0, i_m)) \equiv 1$。

因此，我们可以将（2）重新表述为 MARBP

$$\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \mathsf{E}_{\hat{\boldsymbol{\imath}}}^{\boldsymbol{\pi}} \left[\sum_{m=1}^{M} \sum_{t=0}^{\infty} \hat{R}_m^{a_m(t)} \left(\hat{X}_m(t) \right) \beta^t \right]. \tag{3}$$

这使我们能够采用 Whittle (1988) 提出的无休止多臂赌博机指标化方法，并由作者在近期研究中进一步发展（参见 Niño-Mora (2007b)）。不过必须强调，由此得到的指标策略通常并非最优。

3. 无休止多臂赌博机指标化

本节回顾无休止多臂赌博机指标化，给出基于动态规划的可指标化性证明，说明 MPI 即为 AT指标，并概述将采用的方法。关于 MPI 的定义、分析和计算，我们参考 Whittle (1988) 和 Niño-Mora (2001、2002、2006年)，其适用于上述单个项目的休闲赌博机重新表述。因此，以下省略项目标签 m，于是例如，$\hat{}N$、$\{0,1\} \times N$表示该项目的增广状态空间。我们用 π 和 Π 表示运行该项目的策略和策略空间；此处的记号用来区分它们与上文在多项目情形中使用的粗体记号 $\boldsymbol{\pi}$ 和 $\boldsymbol{\Pi}$ 。

此后，我们假设切换成本满足以下关键条件 .

假设 3.1 $c_i + d_i \geq 0$, 其中 $i \in N$。

3.1 可指数化性、MPI 与滞后现象

我们评估在策略 $\pi \in \Pi$ 下，从 $(a_0^-, i_0) \in \hat{N}$ 开始，在项目上获得的净奖励（即计入切换成本），其奖励测度

$$f^{\pi}_{(a_0^-,i_0)} \triangleq \mathsf{E}^{\pi}_{(a_0^-,i_0)}\left[\sum_{t=0}^{\infty} \hat{R}^{a(t)}(\hat{X}(t))\beta^t\right],$$

并进一步 评估工作量度量所对应的工作量

$$g^{\pi}_{(a_0^-,i_0)} \triangleq \mathsf{E}^{\pi}_{(a_0^-,i_0)}\left[\sum_{t=0}^{\infty} a(t)\beta^t\right].$$

设想项目工作按每完成单位工作支付工资率 ν，这促使我们考虑 ν工资问题

$$\max_{\pi\in\Pi} f^{\pi}_{(a_0^-,i_0)} - \nu g^{\pi}_{(a_0^-,i_0)}, \tag{4}$$

即寻找一个可行策略，使其达到最大值 $v^*{}_{(a_0^-,\ i_0)}(\nu)$，即所获净奖励减去所付劳动力成本。我们将使用 (4) 来 标定每个状态下的 工作边际价值，方法是分析工资 ν 变化时其最优策略的结构。

由于 (4) 是一个有限状态、有限动作的 MDP，我们知道，对于任意工资 $\nu \in \mathbb{R}$，都存在一个平稳、确定且独立于初始状态的最优策略。每个此类策略均由其活动状态集（即活动集）唯一刻画；该集合是增广状态的子集，包含策略规定启动项目的状态。我们将活动集表示为

$$S_0 \oplus S_1 \triangleq \{0\} \times S_0 \cup \{1\} \times S_1, \quad S_0, S_1 \subseteq N.$$

因此，我们用 $S_0 \oplus S_1$ 表示的策略是：若项目之前处于休止状态（分别地，处于激活状态），且其原始状态 $X(t)$ 属于 S_0（分别地，属于 S_1），则激活该项目。

因此，对于任意 ν，存在唯一最优活动集 $S_0^*(\nu) \oplus S_1^*(\nu)$，它由所有激活项目是唯一最优行动的增广状态组成。此类集合对 $\subseteq$ 的依赖关系接下来用于定义可指数化性这一性质。

定义 3.2（可指数化性；MPI） 如果存在一个指标$\nu^*_{(a^-,i)}$，使得对于（a^-，i）$\in \hat{N}$，则称该项目可指数化：

$$S_0^*(\nu) = \{(0,i)\colon \nu^*_{(0,i)} > \nu\} \quad \text{and} \quad S_1^*(\nu) = \{(1,i)\colon \nu^*_{(1,i)} > \nu\}, \quad \nu \in \mathbb{R}. \tag{5}$$

我们称$\nu^*_{(a^-,i)}$ 为该项目的边际生产率指标（MPI），并将 $\nu^*_{(1,i)}$ 称为继续 *MPI*，将$\nu^*_{(0,i)}$ 称为切换 *MPI*。

因此，若在项目处于 $\nu^*_{(a^-,i)}$ 时选择投入（相应地，休息）该项目是最优的(a^-, i) 当且仅当 $\nu^*_{(a^-,i)} \geq \nu$（相应地，$\nu^*_{(a^-,i)} \leq \nu$），则该项目具有 MPI 可指数化性。 Whittle (1988年）在平均准则下为无休止多臂赌博机引入了可指数化性的概念。我们在 Niño-Mora (2006年）中引入了 MPI 一词，其动机是将指标刻画为项目在特定状态下工作所产生的边际生产率度量。

为了证明可指数化性并计算 MPI，Niño-Mora (2001, 2002, 2006年) 提出了一种通过猜测最优活动集结构来构造方法的途径，即构造一个激活集族 $\hat{\mathscr{F}} \subseteq 2^{\hat{N}}$，其中包含集合 $S_0^*(\nu) \oplus S_1^*(\nu)$，当 ν变化时。根据直觉，在假设 3.1下，最优

策略应具有滞后性质（参见第 1 节）：如果项目先前处于休息状态时，启动该项目是最优的，那么当该项目先前处于ˆ活动状态时，启动该项目也应是最优的。这使我们推断，$\mathscr{F}$ 的正确取值应为

$$\hat{\mathscr{F}} \triangleq \{S_0 \oplus S_1 \colon S_0 \subseteq S_1 \subseteq N\}. \tag{6}$$

ˆ ˆ**定义3.3**（$\mathscr{F}$**-可指数化性**）如果满足以下条件，我们称项目是 $\mathscr{F}$-可指数化的：**(i)** 它是可指数化的；ˆ 且 **(ii)** $S_0^*(\nu) \oplus S_1^*(\nu) \in \mathscr{F}$ 对于 $\nu \in \mathbb{R}$。

ˆ ˆ注意，$\mathscr{F}$ 表示一族策略。我们将在定理3.7 中证明，当 $\mathscr{F}$ˆ按式（6）定义时，本文关注的无休止项目确实是 $\mathscr{F}$-可指数化的。

3.2 化归为标准化的零停机成本情形

本节表明，只需考虑停机成本为零的情形，这是对Banks和Sundaram （1994年，第3节）关于恒定切换成本情形的结果的推广 $c_i \equiv c$，$d_i \equiv d$。假设在某一时刻（记为 $t = 0$），项目刚刚启动，将持续随机时长，该时长由停时规则 τ给定。记$\mathbf{R} = (R_j)$ 、$\mathbf{c} = (c_j)$ 和$\mathbf{d} = (d_j)$ 分别为其依赖状态的主动奖励、启动成本和停机成本向量，则从 $X(0) = i$ 开始，在此时间段内的预期折现净收益可写为

$$f_i^\tau(\mathbf{R}, \mathbf{c}, \mathbf{d}) \triangleq \mathsf{E}_i^\tau \left[-c_i + \sum_{t=0}^{\tau-1} R_{X(t)} \beta^t - d_{X(\tau)} \beta^\tau \right]. \tag{7}$$

我们有如下结果，其中$\mathbf{I}$ 是按状态空间 N 编号的单位矩阵，$\mathbf{P} = (p_{ij})_{i,\, j \in N}$ 是转移概率矩阵，而$\mathbf{0}$ 是一个零向量。

引理 3.4 $f_i^\tau(\mathbf{R}, \mathbf{c}, \mathbf{d}) = f_i^\tau(\mathbf{R} + (\mathbf{I} - \beta\mathbf{P})\mathbf{d}, \mathbf{c} + \mathbf{d}\ \mathbf{0})$。

证明。利用初等恒等式

$$d_{X(\tau)} \beta^\tau = d_i - \sum_{t=0}^{\tau-1} \{ d_{X(t)} - \beta d_{X(t+1)} \} \beta^t$$

得到

$$\begin{aligned} f_i^\tau(\mathbf{R}, \mathbf{c}, \mathbf{d}) &\triangleq -c_i + \mathsf{E}_i^\tau \left[\sum_{t=0}^{\tau-1} R_{X(t)} \beta^t - d_{X(\tau)} \beta^\tau \right] \\ &= -c_i - d_i + \mathsf{E}_i^\tau \left[\sum_{t=0}^{\tau-1} \{ R_{X(t)} + d_{X(t)} - \beta d_{X(t+1)} \} \beta^t \right] = f_i^\tau(\mathbf{R} + (\mathbf{I} - \beta\mathbf{P})\mathbf{d}, \mathbf{c} + \mathbf{d}, \mathbf{0}). \end{aligned}$$

□

引理 3.4 说明了如何消除停机成本：只需通过以下变换，将其并入修改后的启动成本和主动奖励中

$$\tilde{\mathbf{c}} \triangleq \mathbf{c} + \mathbf{d} \quad \text{and} \quad \tilde{\mathbf{R}} \triangleq \mathbf{R} + (\mathbf{I} - \beta\mathbf{P})\mathbf{d}. \tag{8}$$

请注意，在 $c_j \equiv c$ 和 $d_j \equiv d$ 的情形中，Banks和Sundaram (1994年) 中讨论的此类变换˜可化简为 $\tilde{c}_j \equiv c + d$ 和 $R_j = R_j + (1 - \beta)d$，与他们的结果一致。

为简化起见，以下讨论将集中于归一化情形 $c_i \geq 0$，$d_i \equiv 0$假设在需要时已完成（8）中的变换。

3.3 AT指标、MPI与可指标化证明

接下来，我们将给出归一化情形下继续型和切换型 AT 指标的表示，提供一个基于动态规划的 $\mathscr{F}$-可指数化性证明，并说明 AT 指标与对应的 MPI 指标一致。这些表示将通过工作测度和奖励测度 g_i^S及 f_i^S来给出，针对的是从 $i \in N$开始、在策略下不含切换成本的底层项目；当其原始状态 $X(t)$ 属于 $S \subseteq N$时，该策略会启动该项目。

定义继续和切换 AT 指标如下

$$\nu^{\mathrm{AT}}_{(a^-,i)} \triangleq \max_{\tau>0} \frac{\mathsf{E}_i^\tau\left[-(1-a^-)c_i + \sum_{t=0}^{\tau-1} R_{X(t)}\beta^t\right]}{\mathsf{E}_i^\tau\left[\sum_{t=0}^{\tau-1}\beta^t\right]} = \max_{S\subseteq N:\, i\in S} \frac{f_i^S - (1-a^-)c_i}{g_i^S}, \tag{9}$$

其中 τ 表示一个停时规则。因此， $\nu^{\mathrm{AT}}_{(1,i)}$ 恰好是该项目的吉廷斯指数。注意，式（9）中最右侧的等式成立，是因为只需考虑平稳确定性停时规则，而这些规则可由其继续集合表示。

下一个结果表明，这些指标与式（$\hat{}\mathscr{F}$ （6）中的策略族相一致。

引理 3.5 $\nu^{\mathrm{AT}}_{(1,i)} \geq \nu^{\mathrm{AT}}_{(0,i)}$，对于 $i \in N$。

证明。固定 $i \in N$。对于每个集合 $S \subseteq N$，假设3.1，即 $c_i \geq 0$，意味着

$$\frac{f_i^S}{g_i^S} \geq \frac{f_i^S - c_i}{g_i^S}.$$

在上述不等式两边分别对 $S \subseteq N$ 取最大值，可知该不等式仍成立；再利用 $\nu^{\mathrm{AT}}_{(1,i)}$ 和 $\nu^{\mathrm{AT}}_{(0,i)}$ 在（9）中的定义，即得结论。□

接下来，我们将利用动态规划证明，式（9）中定义的指标确实是定义 3.2 中项目的继续指标和切换MPI。我们将使用最优值函数 $v^*_{(a^-,i)}$ （ν）的贝尔曼方程 ν-工资问题（4）：对于 $i \in N$，

$$v^*_{(a^-,i)}(\nu) = \max\left\{\beta v^*_{(0,i)}(\nu), R_i - (1-a^-)c_i - \nu + \beta\sum_{j\in N} p_{ij} v^*_{(1,j)}(\nu)\right\}. \tag{10}$$

$\hat{}$ 我们首先证明，只需将注意力限制在激活集族 $\mathscr{F}$ 上即可。

引理 3.6 如果在 $(1,i)$ 中休息该项目是最优的，那么在 $(0,i)$ 中休息该项目也是最优的。

证明。结论可由以下蕴含关系得出：

$$\beta v^*_{(0,i)}(\nu) \geq R_i - \nu + \beta\sum_{j\in N} p_{ij} v^*_{(1,j)}(\nu) \Longrightarrow \beta v^*_{(0,i)}(\nu) \geq -c_i + R_i - \nu + \beta\sum_{j\in N} p_{ij} v^*_{(1,j)}(\nu),$$

该关系成立，因为根据假设 $c_i \geq 0$，由 3.1 可知（回想一下， $d_i \equiv 0$）。 □

接下来给出本节的主要结果（其证明使用了第 3.5 节的结果）。

$\hat{}$**定理 3.7** 重新表述后的休闲项目是 $\mathscr{F}$-可指标化的，其 *MPI* 由 (9)。

证明。我们首先证明，当且仅当 $\nu \geq \nu^{\mathrm{AT}}_{(1,i)}$ 时，在 (1， i) 处休息该项目是最优的。我们有 e:

$$\begin{aligned}
\text{it is optimal to rest at } (1,i) &\iff 0 \geq \max_{S_0 \subseteq S_1 \subseteq N:\, i \in S_1} f^{S_0 \oplus S_1}_{(1,i)} - \nu g^{S_0 \oplus S_1}_{(1,i)} \\
&\iff \nu \geq \max_{S_0 \subseteq S_1 \subseteq N:\, i \in S_1} \frac{f^{S_0 \oplus S_1}_{(1,i)}}{g^{S_0 \oplus S_1}_{(1,i)}} \\
&\iff \nu \geq \max_{S_1 \subseteq N:\, i \in S_1} \frac{f^{S_1}_i}{g^{S_1}_i} = \nu^{\mathrm{AT}}_{(1,i)},
\end{aligned}$$

其中我们用到了引理3.6 以及下文引理4.2(b) 和4.6(b) 中的以下关系：$g^{S_0 \oplus S_1}{}_{(1,i)} = g^{S_1}_i$以及 $f^{S_0 \oplus S_1}{}_{(1,i)} = f^{S_1}_i$，分别对应 $S_0 \subseteq S_1 \subseteq N$ 和 $i \in S_1$。

还需证明，当且仅当 $\nu \geq \nu^{\mathrm{AT}}_{(0,i)}$时，在 (0, i) 处让项目休息是最优的。我们有：

$$\begin{aligned}
\text{it is optimal to rest at } (0,i) &\iff 0 \geq \max_{S_0 \subseteq S_1 \subseteq N:\, i \in S_0} f^{S_0 \oplus S_1}_{(0,i)} - \nu g^{S_0 \oplus S_1}_{(0,i)} \\
&\iff \nu \geq \max_{S_0 \subseteq S_1 \subseteq N:\, i \in S_0} \frac{f^{S_0 \oplus S_1}_{(0,i)}}{g^{S_0 \oplus S_1}_{(0,i)}} \\
&\iff \nu \geq \max_{S_1 \subseteq N:\, i \in S_0} \frac{f^{S_1}_i - c_i}{g^{S_1}_i} = \nu^{\mathrm{AT}}_{(0,i)},
\end{aligned}$$

其中我们用到了引理3.6 以及下文引理4.2(c) 和4.6(c) 中的以下关系：$g^{S_0 \oplus S_1}{}_{(0,i)} = g^{S_1}_i$以及 $f^{S_0 \oplus S_1}{}_{(0,i)} = f^{S_1}_i - c_i$，对应 $i \in S_0 \subseteq S_1 \subseteq N$。

上述结果表明，该项目具有指标性，其继续指标和切换MPI由 (9) 给出。由引理^$\mathscr{F}$-可指标化的结论则可由引理3.5推出。□

3.4 关于最优策略的一个性质

主要结果由 Asawa 和 Teneketzis (1996年) 得出，表明他们的延续指数和切换指数在一定程度上刻画了 MABPSC 的最优策略。

定理 3.8（Asawa 和 Teneketzis (1996年)） *MABPSC* 的一个最优策略具有如下性质：只有在达到相应（延续或切换）指数的停止时刻，才需要决定参与哪个项目。

在 Asawa 和 Teneketzis (1996年，第 3.C 节）中，说明了如何构造达到定理 3.8 中停止时刻的延续（停止）集合。下面的结果结合上述关于式（9）中各指数的 MPI 刻画，重新讨论这一问题。该结果可由上一节的结果直接推出，因此省略证明。

命题 3.9

(a) 达到延续指数 $\nu^*_{(1,i)}$ 的最小延续集合（见 (9)为

$$S^*_1\left(\nu^*_{(1,i)}\right) = \{j \in N : \nu^*_{(1,j)} > \nu^*_{(1,i)}\}.$$

（b）达到切换指数的最小延续集合$\nu^*_{(0,i)}$ 在（9）是

$$S^*_1\left(\nu^*_{(0,i)}\right) = \{j \in N : \nu^*_{(1,j)} > \nu^*_{(0,i)}\}.$$

注意，命题 3.9 表明，尽管基于 MPI 的 MABPSC 策略通常并非最优，但它满足定理 3.8中的最优策略性质。

3.5 PCL-可索引性与 MPI算法

接下来，我们概述将采用的方法，以计算 MPI。该方法基于自适应贪心MPI 算法，该算法由 Niño-Mora (2001, 2002) 提出。此算法可计算满足一种称为 *PCL*-可索引性的强形式可索引性的无休止多臂赌博机的 MPI。

给定 $a \in \{0, 1\}$ 和 $S_0 \oplus S_1 \in \hat{\mathscr{F}}$，用 h$a$， $S_0 \oplus S_1$i 表示这样一种策略：在初始阶段采取行动 a，此后采用 $S_0 \oplus S_1$-激活策略 。对于每个增广状态 (a^-, i) 和 $\hat{}$ 活动集 $S_0 \oplus S_1 \in \mathscr{F}$，定义边际工作量度量

$$w_{(a^-,i)}^{S_0 \oplus S_1} \triangleq g_{(a^-,i)}^{\langle 1, S_0 \oplus S_1 \rangle} - g_{(a^-,i)}^{\langle 0, S_0 \oplus S_1 \rangle}, \tag{11}$$

以及边际奖励测度

$$r_{(a^-,i)}^{S_0 \oplus S_1} \triangleq f_{(a^-,i)}^{\langle 1, S_0 \oplus S_1 \rangle} - f_{(a^-,i)}^{\langle 0, S_0 \oplus S_1 \rangle} \tag{12}$$

以及边际生产率测度

$$\nu_{(a^-,i)}^{S_0 \oplus S_1} \triangleq \frac{r_{(a^-,i)}^{S_0 \oplus S_1}}{w_{(a^-,i)}^{S_0 \oplus S_1}}. \tag{13}$$

稍后我们将看到（参见命题 4.4）， $w^{S_0 \oplus S_1}(a^-, i) > 0$，从而确保 $\nu^{S_0 \oplus S_1}(a^-, i)$ 定义良好。

我们将使用表 1 中的自适应贪心MPI算法 AG $_{\hat{}\mathscr{F}}$，其中 n，$|N|$ 是原始表述中的项目状态数。算法的输入包括项目的参数（未显式列出）以及给定的激活集族 $\mathscr{F}$。算法输出一个由不同增广状态组成的序列 $\{(a_k^-, i_k)\}_{k=1}^{2n}$，这些状态覆盖 $\hat{}N$，并满足 $\hat{}S^k$、$\{(a_1^-, i_1)\ldots\ldots, (a_k^-, i_k)\} \in \hat{}\mathscr{F}$（对于 $1 \le k \le 2n$），同时给出相应的指标值 $\{\nu^*(a_k^-, i_k)\}_{k=1}^{2n}$。注意，每个活动集 $\hat{}S^k$ 都具有 $S_0 \oplus S_1$ 的形式，其中 $S_0 \subseteq S_1 \subseteq N$ 为某个值。为方便记号，我们在该算法中使用紧凑形式 $\hat{}S^k$ 。

表1：MPI算法AG的版本 1 $_{\hat{}\mathscr{F}}$。

AG算法 $\hat{}$ $\mathscr{F}$:
输出： $\{(a_k^-, i_k)\ \nu^*\ {}_{(a_k^-\ i_k)}\}_{k=1}^{2n}$
$S^0 := \emptyset$
对于 $k := 1$ **到** $2n$，**执行**
  选择 $(a_k^-, i_k) \in$ 使 $\{\nu \hat{} S^{k-1}(a^-,i)$ 达到最大值：$(a^-, i) \in \hat{}N \setminus \hat{}S^{k-1}, \hat{}S^{k-1} \cup \{(a^-, i)\} \in \hat{}\mathscr{F})\}$
  $\nu^*_{(a_k^-\ i_k)} := \nu_{(a_k^-\ i_k)}^{\hat{}S^{k-1}}$; $\hat{}S^k := \hat{}S^{k-1} \cup \{(a_k^-, i_k)\}$
结束 { for }

定义 3.10 (PCL($\hat{}\mathscr{F}$)-可指数化性) 我们称该项目是 $PCL(\hat{}\mathscr{F})$-可指数化的，如果：

(i) 正边际工作负载： $w^{S_0 \oplus S_1}(a^-, i) > 0$ 对于 $(a^-, i) \in \hat{}N$ 和 $S_0 \oplus S_1 \in \hat{}\mathscr{F}$。

(ii) 单调非递增指标：AG算法 $_{\hat{}\mathscr{F}}$’的输出满足

$$\nu^*_{(a_1^-, i_1)} \ge \nu^*_{(a_2^-, i_2)} \ge \cdots \ge \nu^*_{(a_{2n}^-, i_{2n})}.$$

表2：MPI算法AG的版本 2 $\hat{}_{\mathscr{F}}$。

AG算法AG $\hat{}_{\mathscr{F}}$:
输出： $\{(0,i_0^{k_0})\ \nu^*_{(0,i_0^{k_0})}\}_{k_0=1}^n$, $\{(1,i_1^{k_1})\ \nu^*_{(1,i_1^{k_1})}\}_{k_1=1}^n$

$S_0^0:=\emptyset$; $S_1^0:=\emptyset$; $k_0:=1$; $k_1:=1$
当 $k_0+k_1\le 2n+1$ **时执行**
  若 $k_1\le n$ **选取** $j_1^{\max}\in$ 使 $\{\nu_1^{(k_0-1,k_1-1}\arg\max)_{(,j)}:\ j\in N\setminus S_1^{k_1-1}\}$
  若 $k_0<k_1$ **选取** $j_0^{\max}\in$ 使 $\{\nu_0^{(k_0-1,k_1-1}\arg\max)_{(,j)}:\ j\in S_1^{k_1-1}\setminus S_0^{k_0-1}\}$
  若 $k_1=n+1$ **或** $\{k_0<k_1\le n$ **且** $\nu^{(k_0-1,k_1-1)}{}_{(1,j_1^{\max})}<\nu^{(k_0-1,k_1-1)}{}_{(0,j_0^{\max})}\}$

$i_0^{k_0}:=j_0^{\max}$; $\nu^*_{(0,i_1^{k_0})}:=\nu^{(k_0-1,k_1-1}_{(0,i_1^{k_0})}$; $S_0^{k_0}:=S_0^{k_0-1}\cup\{i_0^{k_0}\}$; $k_0:=k_0+1$

  else

$i_1^{k_1}:=j_1^{\max}$; $\nu^*_{(1,i_1^{k_1})}:=\nu^{(k_0-1,k_1-1}_{(1,i_1^{k_1})}$; $S_1^{k_1}:=S_1^{k_1-1}\cup\{i_1^{k_1}\}$; $k_1:=k_1+1$

  结束 { if }
结束 { while }

以下结果见 Niño-Mora (2001, 推论 2)，而在本文采用其表述的更一般设定下，见 Niño-Mora (2002, 定理 6.3)。

定理 3.11 一个 *PCL*($\hat{}_{\mathscr{F}}$)-可指数化的休闲项目是$\hat{}_{\mathscr{F}}$-可指数化的，而指标 $\nu^*_{(a^-,i)}$ 由算法 AG $\hat{}_{\mathscr{F}}$ 计算得出，是其 *MPI*。

$\hat{}$虽然在上文的定理 3.7 中已经证明了 $\mathscr{F}$-可指数化性，但我们将在第3.11 节中援引定理4.4 ，以确保算法 AG $\hat{}_{\mathscr{F}}$ 计算 MPI 的有效性。请注意，在算法 AG $\hat{}_{\mathscr{F}}$ 中，选择 (a_k^-,i_k) 时遇到的并列情况可以任意打破。

3.6 MPI 算法的版本 2

表 1 中的 MPI 算法（版本 1）很容易改写为表 2中更明确的版本 2，该版本使用了（6）中 $\hat{}_{\mathscr{F}}$ 的定义。在这一版本及后续版本中，我们采用更方便的类算法记法，例如将 $\nu^{S_0^{k_0-1}\oplus S_1^{k_1-1}}{}_{(0,j)}$ 写作 $\nu^{(k_0-1,k_1-1}{}_{(0,j)}$。请注意，两个版本中构造的活动集及其对应计数器 k 和 k_0,k_1，通过 $\hat{}$ $S^{k-1}=S_0^{k_0-1}\oplus S_1^{k_1-1}$、$k=k_0+k_1-1$ 和 $k_0\le k_1$ 相互关联。

这样的版本 2 将其输出拆分为两个增广状态字符串 $\{(0,i_0^{k_0})\}_{k_0=1}^n$ 和 $\{(1,i_1^{k_1})\}_{k_1=1}^n$，其中 $N=\{i_0^1,\ldots,i_0^n\}=\{i_1^1,\ldots,i_1^n\}$，并附带相应的切换指标值和延续指标值。算法中的活动集 $S_0^{k_0}$ 和 $S_1^{k_1}$ 分别为 $S_0^{k_0}=\{i_0^1,\ldots,i_0^{k_0}\}$ 和 $S_1^{k_1}=\{i_1^1,\ldots,i_1^{k_1}\}$，并满足 $S_0^{k_0}\subset S_1^{k_1}$（对于 $1\le k_0<k_1\le n$）。

我们将在后文展示，版本 2 仍可大幅简化，将继续MPI和切换MPI的计算拆分为两个算法。不过，要实现这样的简化，我们需要在下一节进行PCL-可索引性分析。

4. PCL-可索引性分析

本节旨在对上述具有切换成本的单个项目进行PCL-可索引性分析，并采用其休闲赌博机重新表述。

4.1 工作量与边际工作量度量

我们首先讨论如何计算（ $g^{S_0\oplus S_1}(a^-,i)$ 和 $w^{S_0\oplus S_1}(a^-,i)$ ）的工作量与边际工作量度量，其中 $(a^-,i) \in \hat{} N$和 $S_0 \oplus S_1 \in \hat{} \mathscr{F}$。我们将证明，它们与底层非休眠项目的相应度量 g_i^S 和 w_i^S 密切相关；对于该项目，平稳确定性策略由其活动集 $S \subseteq N$ 表示。

F对于每个 $S \subseteq N$，工作量度量 g_i^S 由以下求值方程刻画

$$g_i^S = \begin{cases} 1 + \beta \sum_{j\in S} p_{ij} g_j^S & \text{if } i \in S \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

注意，(14) 可化为一个系数矩阵为 $\mathbf{I}_S - \beta \mathbf{P}_{SS}$ 的线性方程组，其中 $\mathbf{I}_S$ 是以 S 为索引的单位矩阵，且 $\mathbf{P}_{SS}$ ，$(p_{ij})_{i,j\in S}$。该方程组的解唯一，因为矩阵理论保证，若 $\mathbf{P}_{SS}$ 是次随机矩阵（即其元素非负，且各行元素之和至多为 1），并且 $0 < \beta < 1$，则形如 $\mathbf{I}_S - \beta \mathbf{P}_{SS}$ 的矩阵可逆。

此外，边际工作量度量 w_i^S由以下方式求得

$$w_i^S \triangleq g_i^{\langle 1,S\rangle} - g_i^{\langle 0,S\rangle} = 1 + \beta \sum_{j\in N} p_{ij} g_j^S - \beta g_i^S = \begin{cases} (1-\beta) g_i^S & \text{if } i \in S \\ 1 + \beta \sum_{j\in S} p_{ij} g_j^S & \text{otherwise.} \end{cases} \tag{15}$$

注意，(15) 蕴含

$$w_i^S > 0, \quad i \in N. \tag{16}$$

现在我们回到项目的休闲赌博机重新表述。以下结果给出了给定活动集 $S_0 \oplus S_1 \in \hat{} \mathscr{F}$ 时工作量度量 $g^{S_0\oplus S_1}(a^-,\ i)$ 的求值方程。

引理4.1

$$g_{(a^-,i)}^{S_0\oplus S_1} = \begin{cases} 1 + \beta \sum_{j\in N} p_{ij} g_{(1,j)}^{S_0\oplus S_1} & \text{if } i \in S_{a^-} \\ \beta g_{(0,i)}^{S_0\oplus S_1} & \text{otherwise.} \end{cases}$$

下一个结果用 g_i^S表示工作量度量 $g^{S_0\oplus S_1}(a^-,\ i)$ 。

ˆ引理 4.2 对于 $S_0 \oplus S_1 \in \mathscr{F}$:

(a) $g_{(a^-,i)}^{S_0\oplus S_1} = g_i^{S_1} = 0$, *for* $a^- \in \{0,1\}, i \in N \setminus S_1$.

(b) $g_{(1,i)}^{S_0\oplus S_1} = g_i^{S_1}$, *for* $i \in S_1$.

(c) $g_{(0,i)}^{S_0\oplus S_1} = g_i^{S_1}$, *for* $i \in S_0$.

(d) $g_{(0,i)}^{S_0\oplus S_1} = 0$, *for* $i \in S_1 \setminus S_0$.

证明。（a）这部分可直接由策略 $S_0 \oplus S_1$ 的定义得出。

(b) 对于 $i \in S_1$，我们可以写成

$$g_{(1,i)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in S_1} p_{ij} g_{(1,j)}^{S_0\oplus S_1} + \beta \sum_{j\in N\setminus S_1} p_{ij} g_{(1,j)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in S_1} p_{ij} g_{(1,j)}^{S_0\oplus S_1},$$

这里用到了引理 4.1 和 (a) 部分。因此，$g^{S_0\oplus S_1}{}_{(1,i)}$ 满足 (14) 中针对 $g_i^{S_1}$ 的求值方程，其中 $i \in S_1$，由此得到所需结果。

(c) 对于 $i \in S_0$，我们有

$$g_{(0,i)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in S_1} p(i,j) g_{(1,j)}^{S_0\oplus S_1} + \beta \sum_{j\in N\setminus S_1} p_{ij} g_{(1,j)}^{S_0\oplus S_1} = g_{(1,i)}^{S_0\oplus S_1} = g_i^{S_1},$$

这里用到了引理 4.1、(b) 部分以及关系式 $S_0 \subseteq S_1$。

(d) 这部分可直接由策略 $S_0 \oplus S_1$ 的定义得到。 □

关于 $w^{S_0\oplus S_1}{}_{(a^-,\ i)}$，利用 (11) 和引理 4.1 可得

$$w_{(0,i)}^{S_0\oplus S_1} = w_{(1,i)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in N} p_{ij} g_{(1,j)}^{S_0\oplus S_1} - \beta g_{(0,i)}^{S_0\oplus S_1}. \tag{17}$$

下面的结果用 $w^{S_0\oplus S_1}{}_{(a^-,\ i)}$ 表示为 w_i^S 在 (15) 中。

引理 4.3 对于 $a^- \in \{0,\ 1\}$ $S_0 \oplus S_1 \in \mathscr{F}$：

(a) $w^{S_0\oplus S_1}{}_{(a^-,\ i)} = w_i^{S_1}$，对于 $i \in S_0 \cup N \setminus S_1$。(b) $w^{S_0\oplus S_1}{}_{(a^-,\ i)} = w_i^{S_1}/(1-\beta)$，对于 $i \in S_1 \setminus S_0$。

证明。（a）对于 $i \in S_0 \cup N \setminus S_1$，我们可以写成：

$$w_{(a^-,i)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in N} p_{ij} g^{S_0\oplus S_1}(1,j) - \beta g_{(0,i)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in S_1} p_{ij} g_j^{S_1} - \beta g_i^{S_1} = w_i^{S_1},$$

其中我们使用了 (17)、引理 4.2(a, b, c) 以及 (15)。

（b）对于 $i \in S_1 \setminus S_0$，我们可以写成：

$$w_{(a^-,i)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in N} p_{ij} g^{S_0\oplus S_1}(1,j) - \beta g_{(0,i)}^{S_0\oplus S_1} = 1 + \beta \sum_{j\in S_1} p_{ij} g_j^{S_1} = w_i^{S_1} + \beta g_i^{S_1} = w_i^{S_1}/(1-\beta),$$

其中我们使用了（17）、引理 4.2（a、b、d）以及（15）。 □

由此可得以下结果（参见定义 3.10（i））。

命题 4.4 $w^{S_0\oplus S_1}{}_{(a^-,\ i)} > 0$，对于 $(a^-,\ i) \in \hat{}N$，$S_0 \oplus S_1 \in \hat{}\mathscr{F}$。

证明。该结果可由引理 16) 直接推出，引理 4.3。 □

4.2 奖励与边际奖励测度

接下来，我们讨论如何计算度量 $f^{S_0\oplus S_1}(a^-,\ i)$ 和 $r^{S_0\oplus S_1}(a^-,\ i)$，并将它们与底层无启动成本的非休息项目的对应度量$f_i^S$和 r_i^S联系起来。

For 对于每个活动集 $S\subseteq N$，奖励测度 f_i^S由以下方程刻画 s

$$f_i^S=\begin{cases}R_i+\beta\sum_{j\in S}p_{ij}f_j^S & \text{if } i\in S\\ 0 & \text{otherwise,}\end{cases}\tag{18}$$

其解唯一，而边际奖励测度 r_i^S由下式给出

$$r_i^S\triangleq f_i^{\langle 1,S\rangle}-f_i^{\langle 0,S\rangle}=R_i+\beta\sum_{j\in S}p_{ij}f_j^S-\beta f_i^S=\begin{cases}(1-\beta)f_i^S & \text{if } i\in S\\ R_i+\beta\sum_{j\in S}p_{ij}f_j^S & \text{otherwise.}\end{cases}\tag{19}$$

回到休息型表述，下一个结果给出了奖励测度的求值方程 $f^{S_0\oplus S_1}(a^-,\ i)$，对于给定的活动集 $S_0\oplus S_1\in\hat{}\mathscr{F}$。回顾式（1）。

引理 4.5

$$f_{(a^-,i)}^{S_0\oplus S_1}=\begin{cases}R_i-(1-a^-)c_i+\beta\sum_{j\in N}p_{ij}f_{(1,j)}^{S_0\oplus S_1} & \text{if } i\in S_{a^-}\\ \beta f_{(0,i)}^{S_0\oplus S_1} & \text{otherwise.}\end{cases}$$

下一个结果用 f_i^S 表示奖励测度 $f^{S_0\oplus S_1}(a^-,\ i)$。

$\hat{}$**引理 4.6** 对于 $S_0\oplus S_1\in\mathscr{F}$:

(a) $f_{(a^-,i)}^{S_0\oplus S_1}=0=f_i^{S_1}$, *for* $a^-\in\{0,1\},i\in N\setminus S_1$.

(b) $f_{(1,i)}^{S_0\oplus S_1}=f_i^{S_1}$, *for* $i\in S_1$.

(c) $f_{(0,i)}^{S_0\oplus S_1}=f_i^{S_1}-c_i$, *for* $i\in S_0$.

(d) $f_{(0,i)}^{S_0\oplus S_1}=0=f_i^{S_0}$, *for* $i\in S_1\setminus S_0$.

证明。(a) 这部分由策略 $S_0\oplus S_1$ 的定义立即得出。(b) 对于 $i\in S_1$，我们可以写成:

$$f_{(1,i)}^{S_0\oplus S_1}=R_i+\beta\sum_{j\in S_1}p_{ij}f_{(1,j)}^{S_0\oplus S_1}+\beta\sum_{j\in N\setminus S_1}p_{ij}f_{(1,j)}^{S_0\oplus S_1}=R_i+\beta\sum_{j\in S_1}p_{ij}f_{(1,j)}^{S_0\oplus S_1},$$

这里我们使用了引理 4.5 以及 (a) 部分。因此，对于 $f^{S_0\oplus S_1}(1,\ i)$，当 $i\in S_1$时，满足式 (18) 中相应项 $f_i^{S_1}$的求值方程，由此得到结论。

(c) 对于 $i\in S_0$，我们可以写成:

$$f_{(0,i)}^{S_0\oplus S_1}=R_i-c_i+\beta\sum_{j\in S_1}p_{ij}f_{(1,j)}^{S_0\oplus S_1}+\beta\sum_{j\in N\setminus S_1}p_{ij}f^{S_0\oplus S_1}(1,j)=f_{(1,i)}^{S_0\oplus S_1}-c_i=f_i^{S_1}-c_i,$$

这里我们结合 (a, b) 部分使用了 $S_0\subseteq S_1$。

(d) 这部分直接由策略 $S_0\oplus S_1$ 的定义得出。 □

关于边际奖励测度 $r^{S_0\oplus S_1}{}_{(a^-,\ i)}$，我们利用（12）和引理4.5 得出

$$r^{S_0\oplus S_1}_{(a^-,i)} = R_i - (1-a^-)c_i + \beta\sum_{j\in N} p_{ij} f^{S_0\oplus S_1}_{(1,j)} - \beta f^{S_0\oplus S_1}_{(0,i)}. \tag{20}$$

以下结果用 $r^{S_0\oplus S_1}{}_{(a^-,\ i)}$ 表示为 r^S_i。

引理 4.7 对于 $S_0 \oplus S_1 \in \mathscr{F}$:

(a) $r^{S_0\oplus S_1}_{(0,i)} = r^{S_0\oplus S_1}_{(1,i)} - c_i$, *for* $i \in N$.

(b) $r^{S_0\oplus S_1}_{(1,i)} = r^{S_1}_i$, *for* $i \in N \setminus S_1$.

(c) $r^{S_0\oplus S_1}_{(1,i)} = r^{S_1}_i + \beta c_i$, *for* $i \in S_0$.

(d) $r^{S_0\oplus S_1}_{(1,i)} = r^{S_1}_i/(1-\beta)$, *for* $i \in S_1 \setminus S_0$.

证明。（a）这一部分由（20）立即得出。

（b）对于 $i \in N\setminus S_1$，可写为:

$$r^{S_0\oplus S_1}_{(1,i)} = R_i + \beta\sum_{j\in N} p_{ij} f^{S_0\oplus S_1}_{(1,j)} - f^{S_0\oplus S_1}_{(1,i)} = R_i + \beta\sum_{j\in S_1} p_{ij} f^{S_1}_j - f^{S_1}_i = r^{S_1}_i,$$

这里用到了（20）、引理 4.6（a，b）以及（19）。

（c）对于 $i \in S_0$，可写为:

$$r^{S_0\oplus S_1}_{(1,i)} = R_i + \beta\sum_{j\in N} p_{ij} f^{S_0\oplus S_1}_{(1,j)} - \beta f^{S_0\oplus S_1}_{(0,i)} = R_i + \beta\sum_{j\in S_1} p_{ij} f^{S_1}_j - \beta(f^{S_1}_i - c_i) = r^{S_1}_i + \beta c_i,$$

其中我们使用了 $S_0 \subseteq S_1$，（20），引理4.6（a、b、c）和（19）。

（d）对于 $i \in S_1\setminus S_0$，我们可以写为:

$$r^{S_0\oplus S_1}_{(1,i)} = R_i + \beta\sum_{j\in N} p_{ij} f^{S_0\oplus S_1}_{(1,j)} - \beta f^{S_0\oplus S_1}_{(0,i)} = f^{S_1}_i = \frac{r^{S_1}_i}{1-\beta},$$

其中我们使用了（20），引理4.6（a、b、d）、（18）和（19）。证明完毕。 □

4.3 边际生产率测度

我们继续讨论边际生产率测度的计算 $\nu^{S_0\oplus S_1}{}_{(a^-,\ i)}$ 在（13）中的计算，并将其与给定的底层非休息项目对应测度联系起来:

$$\nu^S_i \triangleq \frac{r^S_i}{w^S_i}, \quad i \in N, S \subseteq N. \tag{21}$$

下一个结果将所需的 $\nu^{S_0\oplus S_1}{}_{(a^-,\ i)}$ 表示为 ν^S_i。

引理 4.8 对于 $S_0 \oplus S_1 \in \mathscr{F}$:

(a) $\nu^{S_0\oplus S_1}_{(0,i)} = \nu^{S_0\oplus S_1}_{(1,i)} - c_i/w^{S_0\oplus S_1}_{(1,i)}$, *for* $i \in N$.

(b) $\nu^{S_0\oplus S_1}_{(1,i)} = \nu^{S_1}_i = \nu^{\emptyset\oplus S_1}_{(1,i)}$, *for* $i \in N \setminus S_0$.

(c) $\nu^{S_0\oplus S_1}_{(1,i)} = \nu^{S_1}_i + \beta c_i / w^{S_1}_i$, *for* $i \in S_0$.

(d) $\nu^{S_0\oplus S_1}_{(0,i)} = \nu^{S_1}_i - (1-\beta) c_i / w^{S_1}_i = \nu^{\emptyset\oplus S_1}_{(0,i)}$, $i \in S_1 \setminus S_0$.

证明。所有部分都可由式（13）、（21）、引理 4.3 和引理 4.7。 □

上述结果使我们能够将 MPI 算法的版本 2 改写为表 3 中所示、更加明确的版本 3。与前面一样，例如，我们将 $\nu^{S^0_{k_0}\oplus S^1_{k_1}}$(0,$j$) 写作 $\nu^{(k_0,k_1)}$ (0,j)，并将 $\nu^{S_{k_1}}_j$写作$_{\nu^{(k_1)}\,j}$。请注意，在版本 3 中，我们依据引理 4.8(d)，使用 $\nu^{(0,k_1-1)}$ (0,j)（表示 $\nu^{S_0\oplus S_{k_1-1}}$ (0,j)）代替 $\nu^{(k_0-1,k_1-1)}$ (0,j)。这样做是出于计算方面的考虑，因为存储量 $\nu^{(0,k_1-1)}$ (0,j) 所需的维度比存储 $\nu^{(k_0-1,k_1-1)}$ (0,j) 少一个。

表 3：AG 算法的版本 3 $\hat{}_{\mathscr{F}}$。

算法AG $\hat{}_{\mathscr{F}}$:
输出： $\{(0,i^{k_0}_0)\ \nu^*_{(0,i^{k_0}_0)}\}^n_{k_0=1}$, $\{i^{k_1}_1, \nu^*_{i^{k_1}_1}\}^n_{k_1=1}$

$S^0_0 := \emptyset$; $S^0_1 := \emptyset$; $k_0 := 1$; $k_1 := 1$
当 $k_0 + k_1 \le 2n+1$ **时执行**
  如果 $k_1 \le n$ **选择** $j^{\max}_1 \in \arg\max\{\nu^{(k_1-1)\,j} : j \in N \setminus S^{k_1-1}_1\}$
  $\nu^{(0,k_1-1)\,(0,j)} := \nu^{(k_1-1)\,j} - (1-\beta) c_j / w^{(k_1-1)\,j}$, $j \in S^{k_1-1}_1 \setminus S^{k_0-1}_0$
  如果 $k_0 < k_1$ **选择** $j^{\max}_0 \in \arg\max\{\nu^{(0,k_1-1)\,(,j)}_0 : j \in S^{k_1-1}_1 \setminus S^{k_0-1}_0\}$
  如果 $k_1 = n+1$ **或** $\{k_0 < k_1 \le n$ **且** $\nu^{(k_1-1}_{j^{\max}_1} < \nu^{(0,k_1-1)\,)\,(,)}_{0\,j^{\max}_0}\}$
    $i^{k_0}_0 := j^{\max}_0$; $\nu^*_{(0,i^{k_0}_0)} := \nu^{(0,k_1-1)}_{(0,i^{k_0}_0)}$; $S^{k_0}_0 := S^{k_0-1}_0 \cup \{i^{k_0}_0\}$; $k_0 := k_0 + 1$
  else
    $i^{k_1}_1 := j^{\max}_1$; $\nu^*_{i^{k_1}_1} := \nu^{(k_1-1)}_{i^{k_1}_1}$; $S^{k_1}_1 := S^{k_1-1}_1 \cup \{i^{k_1}_1\}$; $k_1 := k_1 + 1$
  结束 { if }
结束 { while }

$\hat{}$4.4 证明 $PCL(\mathscr{F})$-可指数化性

$\hat{}$现在我们可以着手证明 $PCL(\mathscr{F})$-可指数化性（参见定理 3.11）。

定理 4.9 重新表述的休闲项目是 *PCL*($\hat{}_{\mathscr{F}}$)-可指标化的。

证明。定义 3.10(i) 已在命题 4.4 中得到证明。接下来需要证明条件 (ii)，即由算法 AG $\hat{}_{\mathscr{F}}$ 计算出的连续指标值单调不增。参照算法的版本 1，我们将证明第 k 个和第 $(k+1)$ 个计算出的指标值

满足 $\nu^*_{(a_k^-,\ i_k)} \geq \nu^*_{(a_{k+1}^-,\ i_{k+1})}$ 对 $1 \leq k < 2n$。现在，令 $\hat{S}^{k-1}$ 和 $\hat{S}^k = \hat{S}^{k-1} \cup \{(a_k^-,\ i_k)\}$ 如表 1 所示，我们利用 Niño-Mora (2002，命题 6.4(b, c)) 得到

$$r^{\hat{S}^k}_{(a^-,i)} - r^{\hat{S}^{k-1}}_{(a^-,i)} = \frac{r^{\hat{S}^{k-1}}_{(a_k^-,i_k)}}{w^{\hat{S}^{k-1}}_{(a_k^-,i_k)}}\left(w^{\hat{S}^k}_{(a^-,i)} - w^{\hat{S}^{k-1}}_{(a^-,i)}\right), \quad (a^-,i) \in \hat{N},$$

这可立即利用（13）以及 $\nu^*_{(a_k^-,i_k)} = \nu^{\hat{S}^{k-1}}_{(a_k^-,i_k)}$ 重新表述为

$$\nu^{\hat{S}^k}_{(a^-,i)} = \nu^*_{(a_k^-,i_k)} - \frac{w^{\hat{S}^{k-1}}_{(a^-,i)}}{w^{\hat{S}^k}_{(a^-,i)}}\left(\nu^*_{(a_k^-,i_k)} - \nu^{\hat{S}^{k-1}}_{(a^-,i)}\right), \quad (a^-,i) \in \hat{N}.$$

现在，取 $(a^-,i) = (a_{k+1}^-, i_{k+1})$ 在后一个恒等式中，并利用命题 4.4 可得

$$\nu^*_{(a_k^-,i_k)} \geq \nu^*_{(a_{k+1}^-,i_{k+1})} \iff \nu^*_{(a_k^-,i_k)} \geq \nu^{\hat{S}^{k-1}}_{(a_{k+1}^-,i_{k+1})},$$

这表明，只需证明 $\nu^*_{(a_k^-,i_k)} \geq \nu^{\hat{S}^{k-1}}_{(a_{k+1}^-,i_{k+1})}$。

对于 此目的，回想一下，该算法选取 (a_k^-, i_k) 和 (a_{k+1}^-, i_{k+1})，使得

$$(a_k^-, i_k) \in \arg\max\{\nu^{\hat{S}^{k-1}}_{(a^-,i)} \colon (a^-,i) \in N \setminus \hat{S}^{k-1}, \hat{S}^{k-1} \cup \{(a^-,i)\} \in \hat{\mathscr{F}})\},$$
$$(a_{k+1}^-, i_{k+1}) \in \arg\max\{\nu^{\hat{S}^k}_{(a^-,i)} \colon (a^-,i) \in N \setminus \hat{S}^k, \hat{S}^k \cup \{(a^-,i)\} \in \hat{\mathscr{F}})\},$$

因此需要考虑两种情况。若 $\hat{S}^{k-1} \cup \{(a_{k+1}^-, i_{k+1})\} \in \hat{\mathscr{F}}$，则上述关系推出$\nu^*_{(a_k^-,i_k)} = \nu_{\hat{S}^{k-1}}(a_k^-,i_k) \geq \nu_{\hat{S}^{k-1}}(a_{k+1}^-,i_{k+1})$，如所需。若 $\hat{S}^{k-1} \cup \{(a_{k+1}^-, i_{k+1})\} \notin \hat{\mathscr{F}}$，由于 $\hat{S}^k \cup \{(a_{k+1}^-, i_{k+1})\} \in \hat{\mathscr{F}}$，根据式 (6) 中定义的 $\mathscr{F}$ 的结构可知，它必为 $a_k^- = 1$，且 $(a_{k+1}^-, i_{k+1}) = (0, i_k)$。现在我们结合引理 4.8(a)和后一个恒等式，写出

$$\nu^*_{(a_k^-,i_k)} = \nu^*_{(1,i_k)} = \nu^{\hat{S}^{k-1}}_{(1,i_k)} = \nu^{\hat{S}^{k-1}}_{(0,i_k)} + c_i / w^{\hat{S}^{k-1}}_{(1,i_k)} \geq \nu^{\hat{S}^{k-1}}_{(0,i_k)} = \nu^{\hat{S}^{k-1}}_{(a_{k+1}^-,i_{k+1})},$$

其中该不等式由假设3.1 和命题 4.4得出。 □

5. MPI 的解耦计算

在本节中，我们旨在通过在两阶段方案中将继续 MPI 和切换 MPI 的计算解耦，进一步简化版本 3 的 MPI 算法 AG $_{\hat{\mathscr{F}}}$。

5.1 第一阶段：计算延续指数

我们从延续指数 $\nu^*_{(1,i)}$开始，该指数（参见第 3.3节）是无切换成本项目的 Gittins 指数 ν^*_i。我们还需要一些额外量，以供第 5.2节讨论的第二阶段切换指数算法使用。

为计算吉廷斯指数及其他量，我们采用表 4[1] 中的算法方案 AG。这是 Varaiya 等人 （1985）算法的一个变体，按照 Niño-Mora（2007a）的方式重新表述。有关算法 FP(1) 等实现，请参见后者的论文；该算法执行（4/3）$n^3 + O(n^2)$ 次算术运算；或者也可使用 完全主元选取算法 CP——尽管其运算次数更多，但我们发现它在实践中速度更快

表 4：Gittins 指数算法方案 AG[1]。

ALGORITHM AG^1:
Output: $\{i_1^{k_1}, \nu^*_{i_1^{k_1}}, (w_j^{(k_1)}, \nu_j^{(k_1)}), j \in S_1^{k_1}\}_{k_1=1}^n$

set $S_1^0 := \emptyset$; **compute** $\{(w_i^{(0)}, \nu_i^{(0)}) \colon i \in N\}$
for $k_1 := 1$ **to** n **do**
  pick $i_1^{k_1} \in \arg\max \{\nu_i^{(k_1-1)} \colon i \in N \setminus S_1^{k_1-1}\}$
  $\nu^*_{i_1^{k_1}} := \nu^{(k_1-1)}_{i_1^{k_1}}$; $S_1^{k_1} := S_1^{k_1-1} \cup \{i_1^{k_1}\}$
  compute $\{(w_i^{(k_1)}, \nu_i^{(k_1)}) \colon i \in N\}$
end

5.2 第二阶段：计算切换指数

接下来我们讨论切换指数的计算。考虑表 5中的算法 AG^0，该算法以 AG的输出作为输入，并生成一个跨越 的状态序列，以及非递增的指标值 $(0,\ i_0^{k_0})$。现在我们可以给出本文的主要结果。

定理 5.1 算法 AG^0 计算切换指数 $\nu^*_{(0,i)}$。

证明。该结果可由以下观察得出：算法 AG^0 是通过将表 3 中的 MPI 算法 $\mathscr{F}$ AG ˆ $\nu^*_{(0,i)}$ 的计算与 ν_i^* 的计算解耦后，得到的 □版本 3。

接下来，我们评估切换指数算法的计算复杂度。

命题 5.2 算法 AG^0 至多执行 $n^2 + O(n)$ 次算术运算。

证明。操作次数主要由以下语句决定

$$\nu^{(0,k_1)}_{(0,j)} := \nu_j^{(k_1)} - \hat{c}_j / w_j^{(k_1)}, \quad j \in S_1^{k_1} \setminus S_0^{k_0},$$

至多执行 $2k_1$ 次操作。将 $k_1 = 1$、.....、 n 相加即可得出结果。 □

因此我们看到，切换指数的计算效率很高，比延续指数快一个数量级。还需注意，AG算法[0] 涉及处理大小为$n \times n$的矩阵，而 AT 方案所需处理的矩阵大小为 $2n \times 2n$，因此还能大幅减少开销较高的内存操作。

6. 指标对切换成本的依赖关系

接下来，我们讨论在 $c_i \equiv c$和 $d_i \equiv d$的情况下，指标对切换成本的依赖关系的一些性质。下文将 $\nu^*_{(1,i)}(d)$——由于它不依赖于 c ——记作 $\nu^*_{(0,i)}(c,\ d)$，并用 ν_i^* 表示无切换成本时底层项目的吉廷斯指数。

命题 6.1

表 5：切换指数算法 AG^0。

算法 AG^0:
输入： $\{i_1^{k_1}, \nu^*_{i_1^{k_1}}, \ (w^{(k_1)}{}_j, \ \nu^{(k_1)}{}_j) \ \ j \in S_1^{k_1}\}_{k_1=1}^{n}$
输出： $\{i_0^{k_0}, \nu^*{}^{1}_{(0,i_0^{k_0})}\}_{k_0=1}^{n}$

$\hat{c}_j := (1-\beta)c_j, \ j \in N; \quad S_0^0 := \emptyset; \quad S_1^0 := \emptyset; \quad k_0 := 0$
对 $k_1 := 1$ **至** n **执行**
  $S_1^{k_1} := S_1^{k_1-1} \cup \{i_1^{k_1}\}$; 增广$_1 :=$ `false`
  $\nu^{(0,k_1)}_{0,\ j} := \nu^{(k_1)}{}_j - \hat{c}_j / w_j^{(k_1)}, \ j \in S_1^{k_1} \setminus S_0^{k_0}$
  当 $k_0 < k_1$ **且不**(扩充$_1$) **时**
    选取 $j_0^{\max} \in \arg\max \{\nu_0^{(0,k_1)}{}_{(,j)} : \ j \in S_1^{k_1} \setminus S_0^{k_0}\}$
    若 $k_1 = n$ **or** $\nu^*_{i_1^{k_1}} < \nu^{(0,\ k_1)}_{(0,j_0^{\max})}$
      $i_0^{k_0+1} := j_0^{\max}$; $\nu^*{}^{(0,}_{i_0^{k_0+1}\ \ )}$ (style id='1'>$:= \nu^{(0,k_1)}_{(0,i_0^{k_0+1}\ \ )}$
      $S_0^{k_0+1} := S_0^{k_0} \cup \{i_0^{k_0+1}\}$; $k_0 := k_0 + 1$
    else
      增广$_1 :=$ `true`
    结束{ if }
  结束 { while }
结束 { for }

（a） $\nu^*_{(1,i)}(d) = \nu_i^* + (1-\beta)d$。（b）当 $c+d$ 足够大时，$\nu^*_{(0,i)}(c, \ d) = \nu_i^N - (1-\beta)c$。

(c) $\nu^*_{(0,i)}(c,d)$ 关于(c,d)是分段线性凸函数，关于 c 递减，并关于 d非递增。

˜证明。(a) 注意到 $\nu^*_{(1,i)}(d)$ 是一个具有修正奖励的项目的吉廷斯指数 $R_j = R_j + (1-\beta)d$（参见第 3.2节）。在奖励中加上一个常数项，会使吉廷斯指数增加同样的常数，因此可得所需结论。

(b) 利用第 3.2 节和 3.3 节，我们得到

$$\nu^*_{(0,i)}(c,d) = \max_{S \subseteq N:\ i \in S} \frac{f_i^S - c - \{1-(1-\beta)g_i^S\}d}{g_i^S} = (1-\beta)d + \max_{S \subseteq N:\ i \in S} \frac{f_i^S - (c+d)}{g_i^S}, \qquad (22)$$

其中 f_i^S是底层非休眠项目在奖励 R_j 下的奖励测度——请注意，对应的修正奖励 $\tilde{R}_j$下的奖励测度如上所述为 $f_i^S(d) = f_i^S + (1-\beta)dg_i^S$。式（22）中的第二个等式表明，当 $c+d$足够大时，项 $(c+d)/g_i^S$将占主导地位，因此给定表达式的最大值可通过使分母最大化来取得g_i^S。分母的最大值在 $S = N$处取得，此时 $g_i^N = 1/(1-\beta)$。由于 $\nu_i^N = r_i^N / w_i^N = f_i^N / g_i^N$，故得证。

(c) 式（22）中的第一个恒等式将 $\nu^*_{(0,i)}(c,d)$ 表示为关于 (c,d) 的线性函数的最大值，这些函数关于 c递减，关于 d单调不增，由此可得结论。□

注意，命题 6.1(a) 表明，继续留在一个活跃项目上的动机随其停机成本线性增加。此外，当 $\beta \nearrow 1$ 时，对切换成本的依赖关系消失，因为切换指数收敛于未折扣的吉廷斯指数。

接下来，我们给出两个例子来说明上述结果。第一个例子涉及一个状态空间为 $N = \{1,2,3\}$、启动成本为 c 的项目，

$$d = 0, \quad \beta = 0.95, \quad \mathbf{R}^0 = \mathbf{0}, \quad \mathbf{R}^1 = \begin{bmatrix} 0.7221 \\ 0.9685 \\ 0.1557 \end{bmatrix}, \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} 0.8061 & 0.1574 & 0.0365 \\ 0.1957 & 0.0067 & 0.7976 \\ 0.1378 & 0.5959 & 0.2663 \end{bmatrix}.$$

图 1 展示了每个状态的切换指数随 c变化的情况，与命题 6.1(b, c) 一致，表明由该指标诱导的状态排序会随 c 变化。

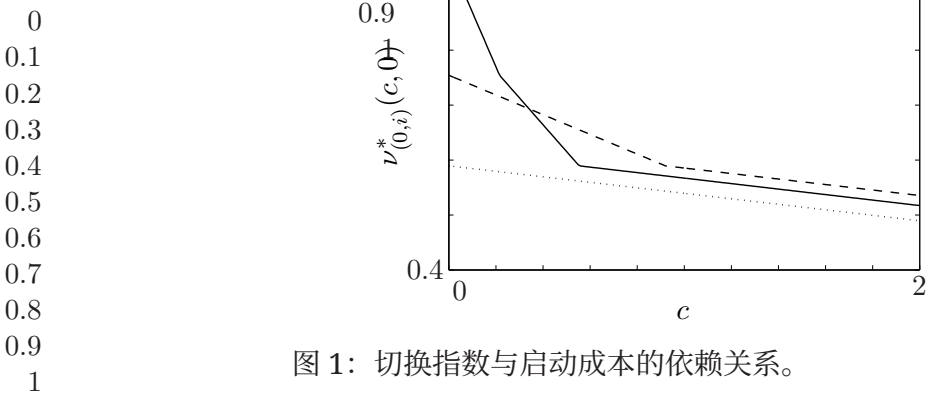
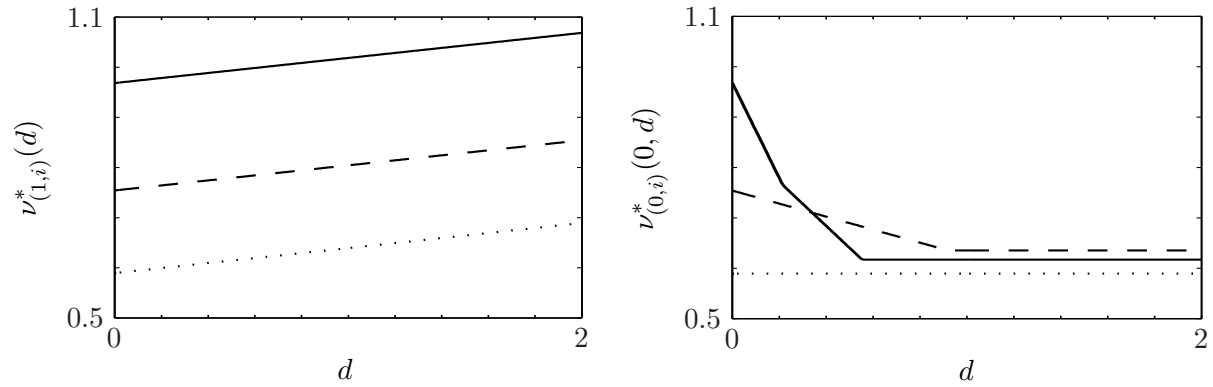


图 1：切换指数与启动成本的依赖关系。

在第二个例子中，我们令停机成本 d 变化，取 $c = 0$。图 2 展示了每个状态的指标随 d变化的情况。图形同样与命题 6.1一致。



Fi图 2：继续指数和切换指数对停机成本的依赖关系。

7. 计算实验

接下来，我们报告基于作者对本文结果的实现所开展的计算研究。实验在一台配备 4 GB 内存的 HP xw9300 AMD Opteron 254（2.8 GHz）工作站上运行 MATLAB R2006b，并使用 Windows XP x64 系统。

第一个实验考察了解耦指数计算方法的运行时间性能。我们让 MATLAB 为每种状态空间大小生成带有启动成本的随机项目实例 $n = 500, 1000, \ldots\ldots, 5000$。对于每个这样的 n，MATLAB 记录了以下计算所需的时间：(i) 使用 Niño-Mora(2007a中的 FP(1) 算法计算延续指数及其他量；(ii) 使用算法AG^{0}计算切换指数；以及 (iii) 通过 AT 方案计算两个指数——使用 Niño-Mora (2007a中的 FP(0) 算法。

结果显示于图 3。左侧面板绘出了计算两个指标的总运行时间（小时），并与 n进行比较，同时给出了三次最小二乘（LS）拟合曲线，这与理论上的$O(n^3)$ 复杂度一致。虚线对应 AT 方案，实线对应我们提出的两阶段方法。结果表明，后者始终比前者快约四倍。

右侧面板绘制了切换指数算法与 n的运行时间（单位为秒），并给出了二次最小二乘拟合，这与理论上的 $O(n^2)$复杂度一致。时间尺度的变化展示了运行时间提升了一个数量级。

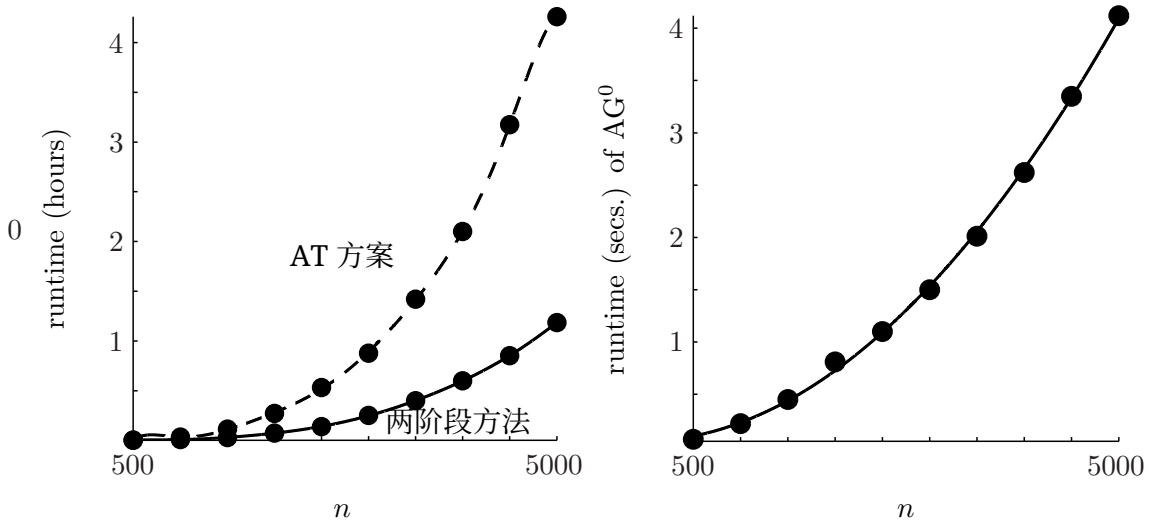


图3：实验1：指数算法的运行时间。

在以下实验中，通过使用经由 TOMLAB 与 MATLAB 接口连接的 CPLEXLP求解器求解动态规划方程的线性规划（LP）形式，为每个实例计算最优策略。MPI 和基准策略（Gittins 指数）则通过 MATLAB 求解其求值方程进行评估。

第二个实验评估了在双项目实例中，MPI 策略的表现如何取决于共同的固定启动成本 c 和折扣因子 β ——停机成本为零。我们随机生成了 100 个实例（每个项目的状态空间为 $N = \{1, \ldots, n\}$，其中 $n = 10$）。在每个实例中，每个项目的参数值均独立生成：转移概率（通过将一个元素服从 Uniform$[0, 1]$ 的矩阵按行缩放，使每行之和为 1 得到）以及主动奖励（Uniform$[0, 1]$）。被动奖励设为零。对于每个实例 $k = 1, \ldots, 100$ 以及每一对 $(c, \beta) \in [0, 1] \times [0.2, 0.9]$——采用 0.1 的网格步长——计算最优目标值 $v^{(k),\mathrm{opt}}$、MPI 策略 $(v^{(k),\mathrm{MPI}})$ 和基准策略 $(v^{(k),\mathrm{bench}})$ 的目标值，以及 MPI 策略对应的相对次优差距$\Delta^{(k),\mathrm{MPI}} = 100(v^{(k),\mathrm{opt}} - v^{(k),\mathrm{MPI}})/|v^{(k),\mathrm{opt}}|$和 MPI 策略相对于基准策略的次优差距比 $\rho^{(k),\mathrm{MPI},\mathrm{bench}} = 100(v^{(k),\mathrm{opt}} - v^{(k),\mathrm{MPI}})/(v^{(k),\mathrm{opt}} - v^{(k),\mathrm{bench}})$。随后，对每一对 (c, β)，将这些值在 100 个实例上取平均，得到平均值 Δ^{MPI} 和

$\rho^{\text{MPI, bench}}$。

目标值 $v^{(k),\text{opt}}$, $v^{(k),\text{MPI}}$ 和 $v^{(k),\text{bench}}$ 按以下方式计算。首先，如上所述，计算相应的价值函数 $v^{(k)}$, $_{\text{opt}}((a_1^-,i_1),(a_2^-,i_2))$、 $v^{(k)}$,MPI $((a_1^-,i_1),(a_2^-,i_2))$ 和 $v^{(k)}$,bench $((a_1^-,i_1),(a_2^-,i_2))$，并将其表示为以初始联合增广状态 $((a_1^-,i_1),(a_2^-,i_2))$ 为索引的向量。然后，假设两个项目最初都处于非活跃状态，计算目标值：

$$v^{(k),\pi} \triangleq \frac{1}{n^2} \sum_{i_1,i_2 \in N} v^{(k),\pi}_{((0,i_1),(0,i_2))}, \quad \pi \in \{\text{opt}, \text{MPI}, \text{bench}\}. \tag{23}$$

图 4 左侧面板绘制了 Δ^{MPI} 随启动成本 c 变化的曲线，针对多个 β，并使用 MATLAB 的三次插值。此类差距在 0 处 $c = 0$ 时从 —开始（此时可恢复最优策略），随后增至低于 *0*25% 的最大值，约在 $c \approx 0.3$处，然后逐渐减小，并在约 0 处再次降至 $c \approx 0.9$ ，且当 c取值更大时保持为零。这种行为符合预期：当 c 足够大时，最优策略只会选择一个项目并始终坚持选择它。右侧面板中的图显示， Δ^{MPI} 会随 β增大。

图5 展示了 MPI 相对于基准策略的次优差距比 $\rho^{\text{MPI, bench}}$。左侧图显示，MPI 策略的平均次优差距始终低于基准策略次优差距的 40%。此外，当 $c = 0$ 以及 c 足够大时，该比值均为 0，因为此时 MPI 策略是最优的。右侧图显示，该比值随 β增大。

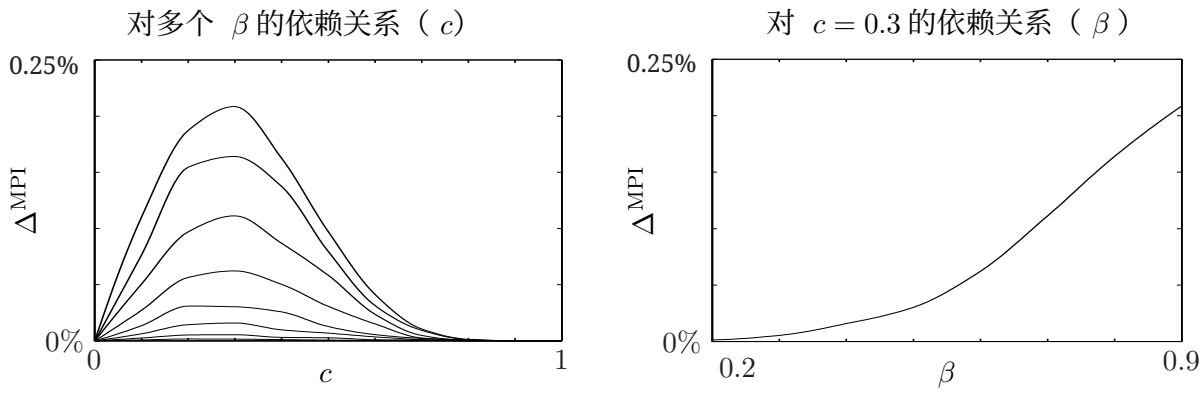
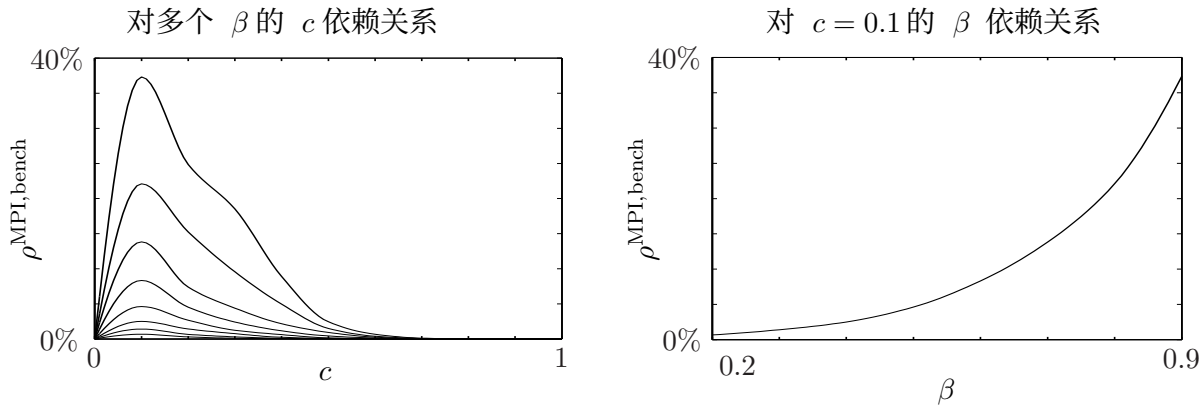


图 4：实验 2：MPI 策略的平均相对次优差距。



Fig图 5：实验 2：MPI 相对于基准策略的平均次优性差距比 y.

第三项实验考察了共同停机成本 d 对 MPI 策略相对性能的影响——启动成本设为零。其他方面与第二项实验相同。结果图见图 6 和 7，与上一项实验的图相似。因此，共同停机成本对相对性能的影响似乎与共同启动成本的影响相似。

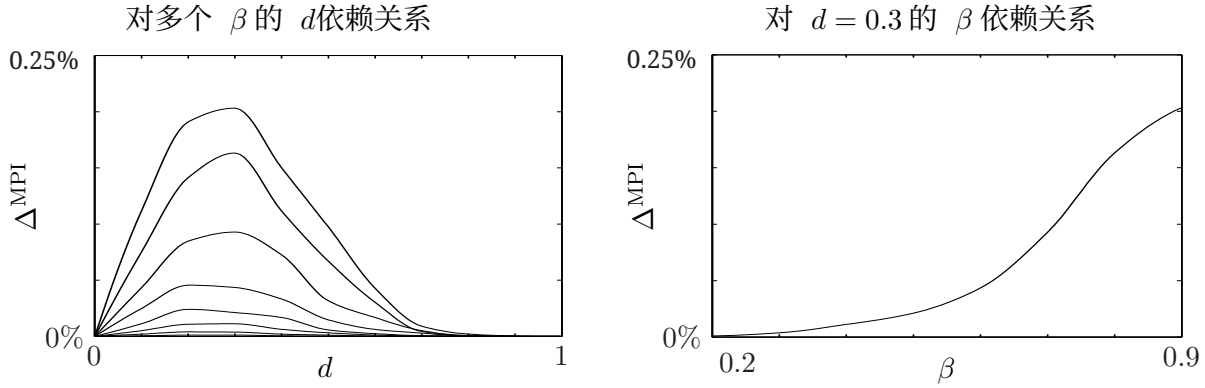


图6：实验3：MPI 策略的平均相对次优差距。

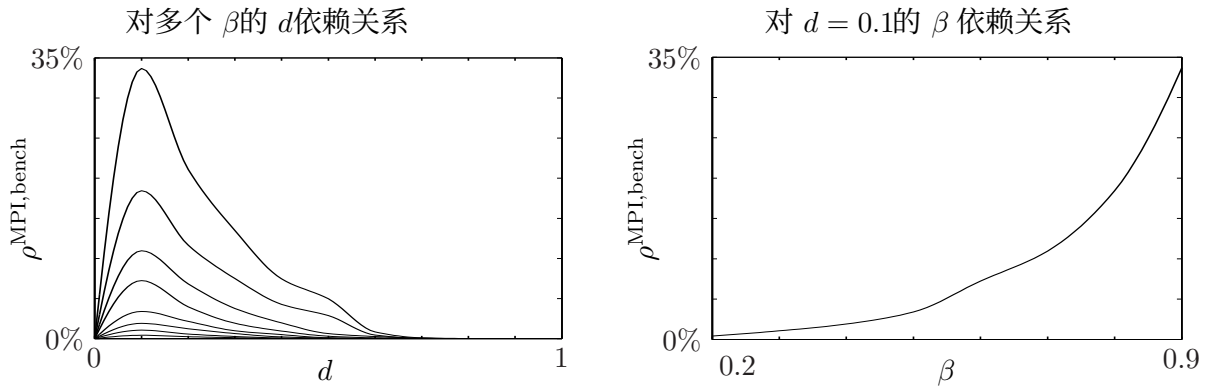


图7：实验3：MPI 相对于基准策略的平均次优性差距比。

第四个实验评估了非对称恒定启动成本 $(c_1、\ c_2) \in [0,1]^2$在无停机成本的双项目实例中以及 $\beta = 0.9$的影响。图 8 展示了 Δ^{MPI} 和 $\rho^{\mathrm{MPI}、\ \mathrm{bench}}$所得的等高线图。我们发现，$\Delta^{\mathrm{MPI}}$ 达到约 0.23% 的最大值；当两个启动成本趋近于零，或其中任一成本增大到足够大时，该值均趋于零。至于$\rho^{\mathrm{MPI}、\ \mathrm{bench}}$，其最大值约为 39.4%；当任一启动成本增大到足够大，或两个成本均趋近于零时，该值均趋于零。

第五个实验评估了 β 变化时，状态依赖型启动成本对无停机成本双项目实例的影响。我们让 MATLAB 为每个实例生成相互独立的 Uniform[0,1] 状态依赖型启动成本。图 9'的左图绘制了 Δ^{MPI} 与 β的关系，显示前者随后者增加而增加，但始终远低于 0.14%。右图绘制了$\rho^{\mathrm{MPI}、\ \mathrm{bench}}$的关系，该值同样随 β增加而增加，并保持在 4% 以下。

第六个也是最后一个实验，是实验 2 的三项目对应实验，基于 100 个随机样本，每个样本由三个 8 状态项目组成。结果见图10 和11，它们分别对应实验 2 的图4 和5。比较图4 和10 可见，后者中 MPI 策略的性能略有下降，不过 Δ^{MPI}

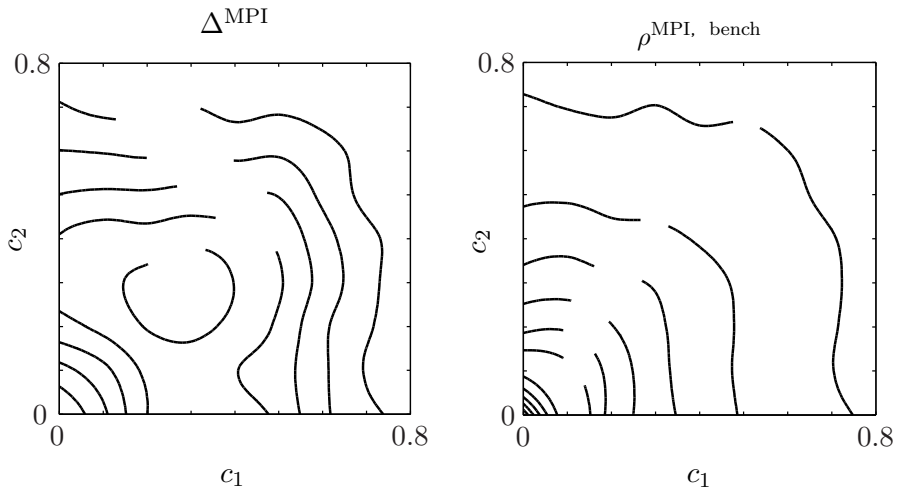


图 8：Ex　第 4 页：MPI 策略相对于（c_1，c_2）在 $\beta = 0.9$ 下的平均相对（%）性能。

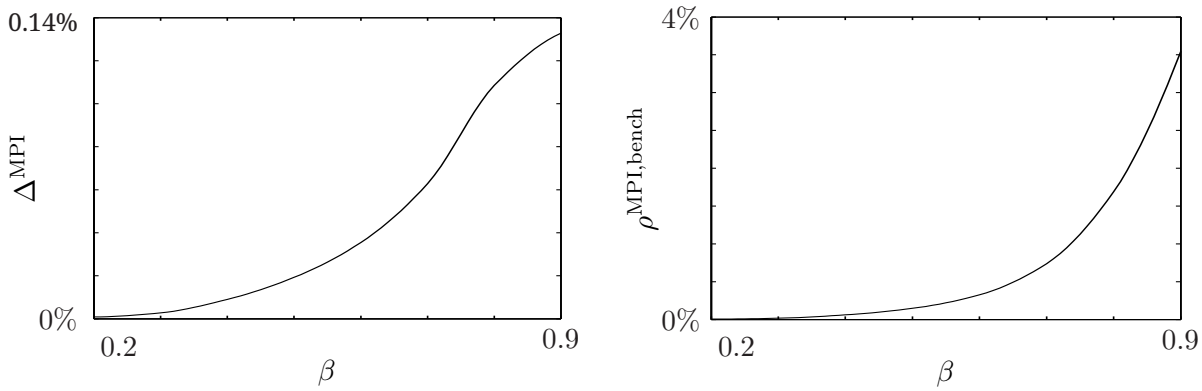


图 10：实验 5：MPI 策略在状态依赖型启动成本下的平均性能。

仍然相当小，不到0。3%。比较图5 和 11 可见，$\rho^{\mathrm{MPI\ 的值相近,\ bench}}$。

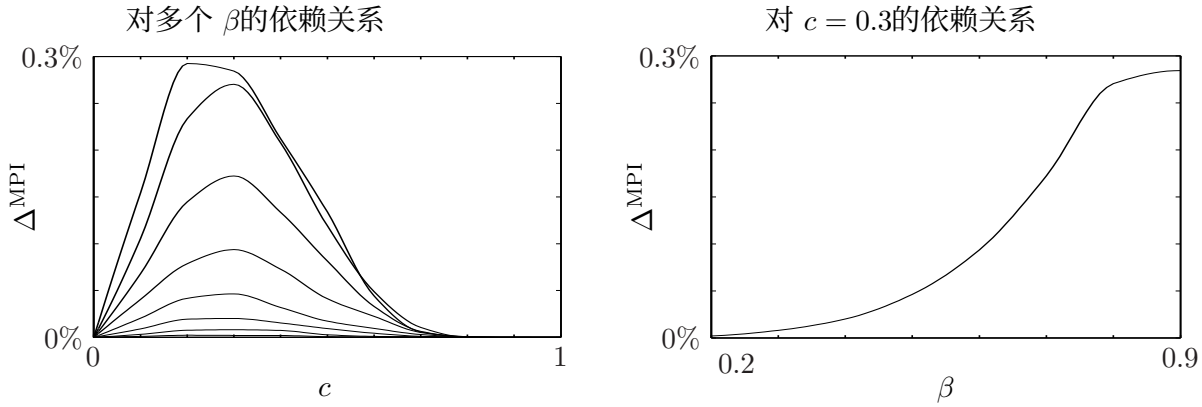


图 10：实验 6：三项目实例中图 4的对应结果。

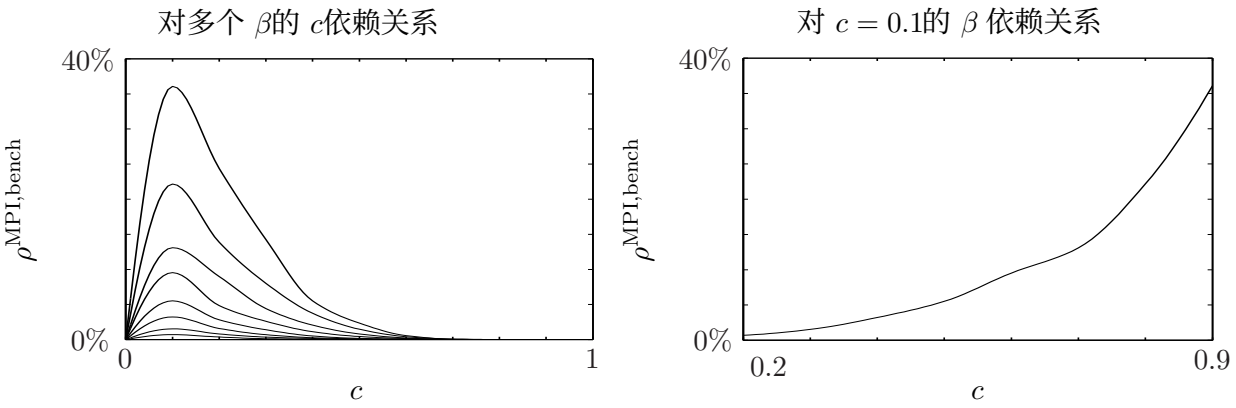


图 11：实验 6：图 5的三项目实例对应图。

8. 结论

本文证明了基于Asawa 和 Teneketzis (1996年) 提出的切换指数的带切换成本的多臂老虎机指数策略具有可处理性和实用价值。该方法基于Whittle(1988年) 提出的无休止赌博机指数化理论，并由作者进一步发展（参见Niño-Mora (2007b）。Niño-Mora (2007c) 报告了模型扩展方面的结果，该扩展还纳入了切换延迟。本文的分析仅部分适用于这种情形，因为此时无休止赌博机的重新表述会得到未必具有 PCL 可指数化性质的半马尔可夫项目。其他值得研究的扩展涉及无休止老虎机及其特定模型，例如Reiman 和 Wein (1998年)。

致谢

作者感谢副编辑和两位审稿人提出的建议，这些建议帮助改进了本文。这项工作始于作者在庞培法布拉大学（巴塞罗那）任教期间，获得了西班牙教育与科学部通过项目 BEC2000-1027 和 MTM2004-02334 提供的资助、拉蒙·卡哈尔研究员奖和 I3 资助、欧盟卓越网络 EuroNGI 和 EuroFGI 的资助，以及马德里自治区通过 UC3M-MTM-05-075 和 CCG06-UC3M/ESP-0767 提供的资助。作者曾在德国瓦德恩举行的 Schloss Dagstuhl「不完全信息下的优化算法」研讨会上介绍了这项工作的部分内容（2005年）。

参考文献

Agrawal, R., M. V. Hegde, D. Teneketzis. 1988年. 多臂老虎机问题中带切换成本的渐近高效自适应分配规则。《*IEEE*自动控制汇刊》**33**899–906.

Asawa, M., D. Teneketzis. 1996年. 带切换惩罚的多臂老虎机。《*IEEE*自动控制汇刊》**41**328–348.

Banks, J. S., R. K. Sundaram. 1994年. 切换成本与吉廷斯指数。《计量经济学》**62**687–694。

Dusonchet, F., M.-O. Hongler. 2003. 一类具有切换成本的确定性退化双臂老虎机问题的最优滞回。《自动化》**39** 1947–1955。Gittins, J. C. 1979. 多臂老虎机过程与动态分配指数（附讨论）。《皇家统计学会杂志*B*辑》**41**148–177。Gittins, J. C., D. M. Jones. 1974. 用于序贯实验设计的动态分配指数。J. Gani, K. Sarkadi, I. Vincze, 编。《统计学进展》（欧洲统计学家会议，布达佩斯，*1972*年）。北荷兰，阿姆斯特丹，荷兰。241–266。Jun, T. 2004. 带切换成本的老虎机问题综述。《经济学家》**152** 513–541。Niño-Mora, J. 2001. 无休止老虎机、部分守恒定律与可指数化性。《应用概率学进展》**33** 76–98。Niño-Mora, J. 2002. 无休止项目与排队准入控制的动态分配指数：一种多面体方法。《数学规划》**93**361–413。Niño-Mora, J. 2006. 无休止老虎机的边际生产率指数、收益递减，以及按订单生产／按库存生产队列的最优控制 $M/G/1$ 。《数学运筹学研究》**31** 50–84。Niño-Mora, J. 2007a. 一种用于吉廷斯指数和马尔可夫链最优停止的（$2/3$）n^3 快速主元算法。《*INFORMS*计算杂志》**19** 页码待定。Niño-Mora, J. 2007b. 通过无休止老虎机边际生产率指数实现动态优先级分配（附讨论）。《*Top*》**15** 页码待定。Niño-Mora, J. 2007c. 计算带切换惩罚的老虎机的指数策略。*SMCtools '07*：*2007*年结构化马尔可夫链求解工具研讨会论文集。ACM，纽约，NY。Reiman, M. I., L. M. Wein. 1998. 带设置操作的双类别队列动态调度。《运筹学研究》**46** 532–547。Van Oyen, M. P., D. Teneketzis. 1994年. 带切换惩罚的森林网络最优随机调度。《应用概率学进展》**26** 474–497。Varaiya, P. P., J. C. Walrand, C. Buyukkoc. 1985. 多臂老虎机问题的扩展：折扣情形。《*IEEE*自动控制汇刊》**30** 426–439。Whittle, P. 1988年. 无休止老虎机：变化世界中的活动分配。J. Gani, 编。《应用概率学庆典》，《应用概率学杂志》**25A**。应用概率信托基金，谢菲尔德，英国。287–298。