A Lagrangian Approach to Optimal Lotteries in Non-Convex Economies*

Chengfeng Shen† Felix Kübler‡ Yucheng Yang§ Zhennan Zhou¶

June 30, 2025

Abstract

We develop a new method to efficiently solve for optimal lotteries in models with non-convexities. To employ a Lagrangian framework, we prove that the value of the saddle point that characterizes the optimal lottery is the same as the value of the dual of the deterministic problem. Our algorithm solves the dual of the deterministic problem via sub-gradient descent. We prove that the optimal lottery can be directly computed from the deterministic optima that are computed along the iterations. We analyze the computational complexity of our algorithm and show that the worst-case complexity is orders of magnitude better than the one arising from a linear programming approach. We apply the method to two canonical problems with private information. First, we solve a principal-agent moral-hazard problem, demonstrating that our approach delivers substantial improvements in speed and scalability over traditional linear programming methods. Second, we study an optimal taxation problem with multidimensional hidden types, which was previously considered computationally challenging, and examine under which conditions the optimal tax schedule will involve lotteries.

Keywords: Private Information, Adverse Selection, Moral Hazard, Non-Convexities, Lotteries, Lagrangian Iteration, Mirrleesian Optimal Taxation.

JEL classification: C61, C63, D61, D82.

*We thank Albert Marcet, Nicola Pavoni, Ned Prescott, Simon Scheidegger, Florian Scheuer, Liangjie Wu, and seminar participants at Zurich, the CRETA conference at Warwick, BSE Summer Forum, SED, and SFI Research Days for helpful discussions. Kubler and Yang thank the Swiss National Science Foundation (SNF) for financial support. Shen and Zhou thank the support by the National Key R&D Program of China with project number 2021YFA1001200, and the NSFC with grant number 12171013.

†Peking University. Email: shencf1999@pku.edu.cn

‡University of Zurich and SFI. Email: fkubler@gmail.com

§University of Zurich and SFI. Email: yucheng.yang@uzh.ch

¶Westlake University. Email: zhouzhennan@westlake.edu.cn

1 Introduction

In economic models of adverse selection, signaling, and moral hazard constrained optimal allocations are typically difficult to characterize because the optimization problems involve non-convex feasible sets, making it challenging to find global optima and often resulting in computational inefficiencies. Motivated by these challenges, there has been substantial interest in exploring lottery solutions to models with private information (Myerson, 1982; Prescott and Townsend, 1984b; Stiglitz, 1982).[1] The lottery approach transforms the original problem by allowing actions and allocations to be randomized, replacing deterministic assignments with probability distributions. This reformulation linearizes the constraints in the probability space, rendering the problem convex, and ensuring the existence of solutions under standard conditions. However, this convexification comes at the cost of dramatically increasing the dimensionality of the problem. As a result, conventional linear programming methods, while theoretically applicable, face severe computational challenges due to the scale of the probability space.

In this paper, we develop a new, efficient, and general method to solve for lottery solutions in a wide class of models with non-convexities. As an example of such a model, it is useful to consider a lottery framework for models with private information, where a social planner maximizes expected utility by choosing a probability distribution over agents' actions and consumption allocations, rather than assigning a specific action and consumption to each agent. The action set is assumed to be finite[2] while the consumption set can be discrete or continuous. The allocations must satisfy two key sets of constraints. First, feasibility constraints ensure that the expected allocation does not exceed resource and technological limits. Second, incentive compatibility constraints guarantee that agents, given their private information, have no incentive to deviate from the prescribed probabilistic allocations.

We develop an algorithm to solve the maximization problem in lotteries using the Lagrangian framework. The Karush-Kuhn-Tucker (KKT) conditions of the Lagrangian formulation include two parts: the optimization condition, which identifies optimal allocations for a given set of Lagrange multipliers, and the complementary slackness conditions. Our method is based on several key theoretical insights. First, we show that the lottery solution assigns positive probability only to the deterministic allocations that maximize the Lagrangian. Second, we show that this implies that the value of the saddle point of the La-

[1]The concept of "lottery solution" is sometimes referred to as randomization in the public finance literature.

[2]Starting from Mirrlees (1999) there is a large strand of literature that considers a principal agent problem where the action set is continuous (see Renner and Schmedders (2015) for the computational difficulties that arise). By focusing on finitely many actions, we avoid many technical problems.

grangian of the lottery system must be the same as the value of the dual of the deterministic problem. This will be the same as the value of the primal problem if and only if all optimal lotteries are degenerate. Third, we show that the frequency of different optimal points of the deterministic system along the iterations approximates the probability distribution of the optimal lottery solution. Our main theoretical results establish that the proposed Lagrangian iteration algorithm converges to a lottery solution.

Our Lagrangian iteration method offers significant computational advantages compared to traditional linear programming approaches. When both the consumption and action sets are finite, we prove that our algorithm's (worst case) complexity is better than $(\frac{1}{\epsilon})^{\frac{2}{1-\rho}}|\hat{C}||A|^{2+\frac{1}{1-\rho}}$, where $|A|$, $|\hat{C}|$ are the cardinality of the action and consumption sets, ϵ denotes the accuracy, and $\rho \to 0$ governs the convergence rate of the algorithm. For high-dimensional problems, the term $|\hat{C}||A|^{2+\frac{1}{1-\rho}}$ dominates due to the curse of dimensionality. In contrast, the best-known complexity bounds for the interior point method for linear programming results in an overall complexity of approximately $(|A||\hat{C}|+|A|^2)^{3.5}$ for the linear programming approach.[3] Our method avoids the need to store or solve for the full probability distribution over allocations, and instead only tracks Lagrange multipliers and solves a sequence of deterministic problems, yielding substantial savings in both computation time and memory usage.

Further computational gains arise in models with special structure. In linear programming methods, the consumption (or "compensation") set must be finely discretized to capture optimality, especially in moral hazard problems, often resulting in millions of variables and constraints (Su and Judd, 2007). Our approach can circumvent this by solving for consumption allocations via a sequence of box-constrained optimization problems. In many models, these subproblems are convex, allowing our method to exploit standard optimization routines for faster and more reliable solutions - an advantage not available to linear programming.

The framework allows us to analyze both models with adverse selection and with moral hazard. We first consider a standard moral-hazard problem in which agents take unobservable actions that influence outcomes, and the planner allocates consumption contracts based on these outcomes. If preferences are separable in actions and consumption, optimal lotteries over consumption allocations are often degenerate (see Arnott and Stiglitz (1988), Jerez (2003)). This allows us to work with a continuous consumption (compensation) set and find the optimal consumption allocation directly as a function of actions. Numerical experiments confirm that the solutions obtained with our Lagrangian iteration algorithm closely match those from linear programming when the model size is manageable. However, as the action

[3] Note that these are worst-case bounds - the linear programs resulting from the problems considered in this paper have a lot of additional structure that might improve on these bounds. However, to our knowledge, no better bounds are available for the linear programs considered.

space becomes more granular, conventional linear programming approaches encounter significant memory limitations. In contrast, our Lagrangian iteration algorithm handles these high-resolution action grids efficiently, maintaining computational feasibility with reduced memory usage.

Our second application studies optimal taxation with multi-dimensional hidden types. We consider a setting with standard preferences over consumption and leisure, allowing for heterogeneity in both productivity and labor supply elasticity. Using an example from Judd et al. (2017), we show that the welfare gains from lotteries depend crucially on details of the calibration of preferences, and in some cases can be substantial. In this context, we also clarify under which conditions optimal solutions are deterministic.

Related Literature Lotteries play a pivotal role in mechanism design under non-convexities induced by asymmetric information. Seminal work by Myerson (1982) established that allowing for lotteries can both improve welfare and make the model more tractable. Prescott and Townsend (Prescott and Townsend (1984a), Prescott and Townsend (1984b)) first laid out the question of whether constrained optimal allocations could be decentralized through linear prices, explicitly linking the problem of mechanism design to the description of a competitive market. Although sufficient conditions for the optimality of deterministic contracts have been established (e.g., Strausz (2006); Hellwig (2007)), they are typically restrictive. Gauthier and Laroque (2014) provide simple sufficient conditions for lotteries to be part of an optimal contract. Since their analysis is local, these conditions are not necessary. As a by-product of our analysis, we can give a general necessary and sufficient condition for the use of lotteries in an optimal contract. More generally, lotteries may arise in constrained optimal allocations whenever non-convexities are present, as shown in applications ranging from stochastic tax schedules to labor market welfare programs (e.g., Weiss (1976); Brito et al. (1995); Gauthier and Laroque (2014); Arnott and Stiglitz (1988); Pavoni and Violante (2007); Armstrong et al. (2010); Pavoni et al. (2016)). Despite the theoretical justification, the practical use of lottery solutions has been limited by computational complexity. By introducing an efficient and scalable solution method, our work enables quantitative analysis of models previously considered infeasible and opens the door to studying optimal policy designs that incorporate lotteries.[4]

We also contribute to the literature on computational methods for solving lottery prob-

[4]One may argue that most contracts used in practice do not involve lotteries. However, as Prescott (1999) argues, there are no economic arguments to exclude contracts that involve a lottery. Then it is a quantitative question to explore under which conditions lotteries can lead to large welfare gains. To address this question, one needs to be able to compute optimal contracts with lotteries. This is precisely where an efficient and scalable method like ours can play an important role.

lems in models with private information. Most of the existing approaches focus on improving linear programming techniques. For example, Prescott (2004) and Prescott and Townsend (2006) exploit the block-triangular structure of the constraint matrix and apply Dantzig-Wolfe decomposition to reduce memory and computational cost, while Doepke and Townsend (2006) propose additional simplifications tailored to specific environments. An alternative approach by Su and Judd (2007), based on mathematical programming with equilibrium constraints (MPEC), improves accuracy in small-scale moral hazard problems. Our method advances the literature by introducing a novel solution method that transcends the linear programming framework, enabling the analysis of more complex and computationally demanding models that were previously infeasible.

Our second application contributes to the literature on optimal tax design under private information. Previous work has shown that lotteries can arise in optimal tax schedules with one-dimensional heterogeneity (e.g. Weiss (1976), Brito et al. (1995), Gauthier and Laroque (2014)). More recent studies, such as Judd et al. (2017), Moser and Olea de Souza e Silva (2019), and Boerma et al. (2022), analyze optimal deterministic tax systems with multidimensional types. Our framework allows for a fully nonlinear analysis of optimal taxation with multidimensional heterogeneity and reveals that lotteries naturally arise for a range of parameter values.

Lastly, our method builds on the subgradient descent literature for nondifferentiable convex optimization (Shor, 2012; Nedic and Bertsekas, 2001), and combines it with a key theoretical insight: the value of the saddle point characterizing the optimal lottery equals the value of the dual of the deterministic problem. This connection allows us to construct the optimal lottery from the sequence of deterministic solutions generated along the Lagrangian iterations. The approach is related in spirit to algorithms used to compute mixed-strategy equilibria in mean-field games (e.g., Shen et al., 2023), but is tailored to the structure of private information problems.

The remainder of the paper is organized as follows. In Section 2 we describe the setup, introduce the algorithm, and state our theorems. In Sections 3 and 4, we apply the method to two canonical problems with private information. Section 5 concludes. The proofs are collected in the appendix.

2 Lagrangian Iteration

2.1 General Setup

In this section, we present a general constrained maximization problem which subsumes many planning problems under asymmetric information considered in economics.

Let A denote a finite set of actions, and C be a compact set of consumption vectors in some Euclidean space $\mathbb{R}^n$. The payoff function $f : A \times C \to \mathbb{R}$ is continuous. In the deterministic problem, the social planner allocates action and consumption to solve the following optimization problem:

$$\begin{aligned} &\max_{a \in A,\, c \in C} f(a, c), \\ \textbf{s.t.} \quad & g_i(a, c) \le 0 \quad i \in \{1, \dots, m\}, \\ & h_j(a, c) \le 0 \quad j \in \{1, \dots, \ell\}, \end{aligned} \tag{2.1}$$

where the functions $g_i(.), i = 1, \dots, m$, and $h_j(.), j = 1, \dots, \ell$ are assumed to be continuous in c.[5] In the rest of the paper, we refer to (2.1) as the *deterministic problem*. A central challenge in such problems arises from the fact that the set of allocations that satisfy the constraints is generally non-convex. This non-convexity not only complicates the existence of a solution but also makes it computationally challenging to solve the problem, even when a solution exists.

Motivated by these difficulties, economists have explored the concept of *lottery solutions*, where the planner chooses a probability distribution over actions and consumption rather than a deterministic allocation (Prescott and Townsend, 1984a; Myerson, 1982; Arnott and Stiglitz, 1988). This approach offers a way to relax the non-convexities and can improve the value of the objective function. In this setting, instead of selecting a deterministic action-consumption pair from $A \times C$ as in (2.1), the social planner chooses a probability distribution over the set of possible outcomes, $A \times C$.

Let $\mathcal{P}(A \times C)$ denote the set of Borel probability measures over $A \times C$. An element of $\mathcal{P}(A \times C)$ is denoted by $x(a, dc)$, which specifies the probability of choosing action $a \in A$ and allocating consumption $c \in C$. Here, we assume that the variable a is discrete and the variable c is continuous. The measure x satisfies $x \ge 0$ and $\sum_{a \in A} \int_{c \in C} x(a, dc) = 1$.

The objective of the social planner in the lottery problem is to maximize the expected

[5] The reason why we distinguish between g-constraints and h-constraints and why A is assumed to be finite will become clear below when we describe the lottery problem.

payoff, which can be written as:

$$\sum_{a\in A}\int_{c\in C} f(a,c)x(a,dc).$$

The g-constraints from the original maximization problem is now also assumed to hold in expectation.

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x(a,dc) \leq 0 \quad \forall i\in\{1,\ldots,m\}.$$

In economic models the typical justification for this assumption is that there is a large number of identical agents and the $g-$constraints describe resource constraint. In contrast, we assume that the h-constraints have to be satisfied for each action a that has positive probability, i.e. for each action, a, in the support of x, the h-constraints must be satisfied in expectation over consumptions and the incentive compatibility constraints become:

$$\int_{c\in C} h_j(a,c)x(a,dc) \leq 0 \quad \forall j\in\{1,\ldots,\ell\}.$$

Note that allowing for continous actions $a\in A$ would lead to measure-theoretic problems since classic measure theory defines marginal integration and addresses problems in terms of almost every $a\in A$ when A is continuous, rather than considering every $a\in A$, which is required in our problem.

Thus, the lottery optimization problem for the social planner can be written as follows.

$$\begin{aligned}
&\max_{x\in\mathcal{P}(A\times C)} \sum_{a\in A}\int_{c\in C} f(a,c)x(a,dc),\\
\text{s.t. } &\sum_{a\in A}\int_{c\in C} g_i(a,c)x(a,dc) \leq 0 \quad \forall i\in\{1,\ldots,m\},\\
&\int_{c\in C} h_j(a,c)x(a,dc) \leq 0 \quad \forall a\in A,\ j\in\{1,\ldots,\ell\}.
\end{aligned} \tag{2.2}$$

Note that in this formulation, we do not require that across agents the distributions are independent. This requirement is of course an important part of the definition of a mixed strategy Nash equilibrium, where allowing for correlation significantly alters the set of equilibria; see Aumann (1974). In a social planning framework, it seems that not allowing for correlation would artificially restrict the set of possible contracts (see also Myerson (1982) for a discussion in the context of the principal-agent problem). In the examples below, correlation is never used in the optimal contract, but this is a result and not an assumption.

The introduction of lotteries transforms the original problem by extending the decision space from individual action-consumption pairs to probability distributions over these pairs. The set of distributions subject to the constraints is now convex, as the constraints are linear in the space of probability measures. This convexity ensures that the problem is well-defined and solvable, allowing for welfare improvements by permitting randomized allocations that are infeasible in deterministic settings. However, this formulation introduces a significant computational challenge, as optimizing over probability distributions inherently involves higher dimensionality. Linear programming methods are typically employed to solve such problems, but they face difficulties due to the high dimensionality of the probability space. In the next section, we propose a new solution method for the lottery problem based on Lagrangian iteration.

2.2 A Lagrange Approach

The essence of our method is to establish a connection between the deterministic system (2.1) and the lottery system (2.2), and to approximate a solution of (2.2) through iterations within the space $A \times C$. The crucial link between these two systems lies in the Lagrange multipliers. In this section, we first establish that the Lagrange multiplier method can be applied to solve the lottery problem (2.2), and then establish a link between the Lagrangian of the lottery problem and the Lagrangian of the deterministic problem.

2.2.1 Lagrangian for the lottery problem

We first demonstrate the validity of the Lagrange Multiplier Method for the lottery system (2.2). This is a linear programming problem in $\mathcal{P}(A \times C)$ and hence a convex optimization problem in $\mathcal{P}(A \times C)$. According to the Karush-Kuhn-Tucker Theorem, a convex optimization problem that satisfies Slater's conditions can be solved by the Lagrange Multiplier Method (see Theorem 1 in Section 8.3 in Luenberger (1997) for details). The definition of the condition is as follows.

Definition 2.1 *(Slater's condition) We say that Slater's condition holds for the lottery system (2.2), if the feasible set to (2.2) includes one inner point, i.e., there exists $x \in \mathcal{P}(A \times C)$ such that*

$$\sum_{a \in A} \int_{c \in C} g_i(a,c)x(a,dc) < 0, \quad \forall i \in \{1, \cdots, m\},$$

and

$$\int_{c \in C} h_j(a,c)x(a,dc) < 0, \quad \forall j \in \{1, \cdots, \ell\},\ a \in A.$$

Note here both inequalities must hold strictly.

We then have the following result.

Theorem 2.1 *Assume that the Slater's condition holds for the lottery system (2.2). The following properties hold:*

1. *The solution of problem (2.2) $x^* \in \mathcal{P}(A \times C)$ exists;*

2. *There exist Lagrangian multipliers $\lambda_i^*, \gamma_{j,a}^* \geq 0$ ($i \in \{1, \cdots, m\}$, $j \in \{1, \cdots, l\}$, $a \in A$) corresponding to x^*, i.e. there exists $\lambda^* \geq 0$, $\gamma^* \geq 0$ satisfying:*

   (a) *x^* is the maximizer of*

$$\begin{aligned} L(x; \lambda^*, \gamma^*) = & \sum_{a \in A} \int_{c \in C} f(a, c) x(a, dc) \\ & - \sum_{i=1}^{n} \lambda_i^* \sum_{a \in A} \int_{c \in C} g_i(a, c) x(a, dc) - \sum_{j=1}^{\ell} \sum_{a \in A} \gamma_{j,a}^* \int_{c \in C} h_j(a, c) x(a, dc). \end{aligned}$$

   (b) *λ^* and γ^* satisfy complementary conditions*

$$\lambda_i^* \sum_{a \in A} \int_{c \in C} g_i(a, c) x^*(a, dc) = 0, \ \forall i \in \{1, \cdots, m\}$$

   and

$$\gamma_{j,a}^* \int_{c \in C} h_j(a, c) x^*(a, dc) = 0, \ \forall j \in \{1, \cdots, \ell\}, \ a \in A.$$

Proof The result follows from Theorem 1 in section 8.3 in Luenberger (1997).[6] □

Slater's condition is sufficient but not necessary; hence, the Lagrangian multipliers may exist even when Slater's condition is violated. Generally, it is difficult to verify whether the Slater's condition holds in (2.2). However, if we relax the constraints in (2.2) to

$$\sum_{a \in A} \int_{c \in C} g_i(a, c) x(a, dc) \leq \epsilon, \quad \forall i \in \{1, \cdots, m\},$$

and

$$\int_{c \in C} h_j(a, c) x(a, dc) \leq \epsilon, \quad \forall j \in \{1, \cdots, \ell\}, \ a \in A,$$

[6] Indeed we need to verify the existence of the maximizer in 2(a) for applying the Theorem in Luenberger (1997). This follows directly from the fact that $\mathcal{P}(A \times C)$ is compact with respect to the *-weak topology. The existence of the maximizer can then be deduced by the standard compactness argument. The details are similar as in the proof of Theorem 2.5 below and we omit them here for simplification.

for some $\epsilon > 0$, then any feasible measure $x \in \mathcal{P}(A \times C)$ for the original lottery system (2.2) should satisfy these two relaxed constraints strictly.

We therefore consider the following relaxed system.

$$\begin{aligned} &\max_{x \in \mathcal{P}(A\times C)} \sum_{a\in A} \int_{c\in C} f(a,c)x(a,dc), \\ \text{s.t. } &\sum_{a\in A} \int_{c\in C} g_i(a,c)x(a,dc) \le \epsilon \quad \forall i \in \{1,\dots,m\}, \\ &\int_{c\in C} h_j(a,c)x(a,dc) \le \epsilon \quad \forall a \in A,\ j \in \{1,\dots,\ell\}. \end{aligned} \tag{2.3}$$

The Slater's condition for (2.3) is then satisfied when there is at least one feasible point for the original lottery system (2.2). We denote x^ϵ as the optimal solution to (2.3). The following theorem shows that x^ϵ will converge to x^*, which is the optimal solution to the original lottery system (2.2).

Theorem 2.2 *Assume that there exists at least one feasible point for the original lottery problem* (2.2)*, hence for any $\epsilon > 0$, the solution to the relaxed lottery problem* (2.3) *exists, denoted as x^ϵ. Then we can choose a sequence $\{\epsilon_n\} \to 0$, s.t. x^{ϵ_n} *-weak converges in the finite Borel measures space on $A \times C$, denoted as $\mathcal{M}(A \times C)$, to some $x^* \in \mathcal{P}(A \times C) \subset \mathcal{M}(A \times C)$ as $n \to \infty$, i.e. for any $\varphi \in C^0(A \times C)$, we have*

$$\sum_{a\in A} \int_{c\in C} \varphi(a,c)x^{\epsilon_n}(a,dc) \to \sum_{a\in A} \int_{c\in C} \varphi(a,c)x^*(a,dc), \quad \textit{as } n \to \infty.$$

Furthermore, x^* is an optimal solution to (2.2)*.*

Proof See Online Appendix A.3. □

According to theorem 2.2, we can apply the Lagrange Multiplier Method to the relaxed system (2.3) with a small relaxation parameter $\epsilon > 0$ to obtain its optimal solution x^ϵ, which can be considered as an approximate solution to the original lottery system (2.2). This theorem ensures that the Lagrange Multiplier Method is a valid numerical approach to solve the system (2.2). In the remainder of this paper, we will always assume that the Lagrangian multipliers to system (2.2) exist.

2.2.2 The Relation between the Lottery Problem and the Deterministic Problem

Given Lagrange parameters λ, γ, we define the Lagrangian function in the probability space $\mathcal{P}(A \times C)$ as

$$L(x; \lambda, \gamma) :=$$
$$\sum_{a \in A} \int_{c \in C} f(a,c)x(a,dc) - \sum_{i=1}^{n} \lambda_i \sum_{a \in A} \int_{c \in C} g_i(a,c)x(a,dc) - \sum_{j=1}^{\ell} \sum_{a \in A} \gamma_{j,a} \int_{c \in C} h_j(a,c)x(a,dc), \tag{2.4}$$

and the Lagrangian function in the pure strategy space $A \times C$ as

$$\mathcal{L}(a, c; \lambda, \gamma) := f(a,c) - \sum_{i=1}^{n} \lambda_i g_i(a,c) - \sum_{j=1}^{\ell} \gamma_{j,a} h_j(a,c). \tag{2.5}$$

In this subsection we show that under the assumption that Slater's conditions holds, we have the following equivalence

$$\max_{x \in \mathcal{P}(A \times C)} \min_{(\lambda,\gamma) \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}} L(x; \lambda, \gamma) = \min_{(\lambda,\gamma) \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}} \max_{a \in A,\, c \in C} \mathcal{L}(a, c; \lambda, \gamma),$$

bridging the gap between the lottery problem (the left-hand side) and the dual of the deterministic problem (the right-hand side).

In order to understand why this is correct, the following theorem first establishes that for given λ, γ the optimal solutions to the maximization of $L(.; \lambda, \gamma)$ and $\mathcal{L}(.; \lambda, \gamma)$ have the same value.

Theorem 2.3 *Given $\lambda \geq 0$, $\gamma \geq 0$. Let L and $\mathcal{L}$ be defined as in (2.4) and (2.5) respectively. Then we have*

$$\max_{x \in \mathcal{P}(A \times C)} L(x; \lambda, \gamma) = \max_{a \in A, c \in C} \mathcal{L}(a, c; \lambda, \gamma). \tag{2.6}$$

Furthermore, if we define $Z = \arg\max_{a \in A, c \in C} \mathcal{L}(a, c; \lambda, \gamma)$, then

$$x^* \in \arg \max_{x \in \mathcal{P}(A \times C)} L(x; \lambda, \gamma)$$

if and only if the measure of Z^c with respect to x^* is zero, i.e.

$$x^*(Z^c) = 0. \tag{2.7}$$

Proof We first prove (2.6). By direct computation, we have

$$
\begin{aligned}
& L(x;\lambda,\gamma) \\
=& \sum_{a\in A}\int_{c\in C} f(a,c)x(a,dc) - \sum_{i=1}^{n}\lambda_i \sum_{a\in A}\int_{c\in C} g_i(a,c)x(a,dc) - \sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}\int_{c\in C} h_j(a,c)x(a,dc) \\
=& \sum_{a\in A}\int_{c\in C}\left(f(a,c) - \sum_{i=1}^{n}\lambda_i g_i(a,c) - \sum_{j=1}^{\ell}\gamma_{j,a}h_j(a,c)\right)x(a,dc) \\
:=& \sum_{a\in A}\int_{c\in C}\mathcal{L}(a,c;\lambda,\gamma)x(a,dc) \\
\leq& \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma).
\end{aligned}
\tag{2.8}
$$

Hence

$$
\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) \leq \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma).
$$

On the other hand, if $(a^*,c^*)\in\arg\max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma)$, then

$$
L(\delta_{(a^*,c^*)};\lambda,\gamma) = \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma),
$$

implying that

$$
\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) \geq \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma).
$$

Hence we can conclude that (2.6) holds.

Now we prove (2.7). According to (2.8), we know that

$$
L(x^*;\lambda,\gamma) = \sum_{a\in A}\int_{c\in C}\mathcal{L}(a,c;\lambda,\gamma)x^*(a,dc),
$$

and hence

$$
\begin{aligned}
0 &\leq L(x^*;\lambda,\gamma) - \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) \\
&= L(x^*;\lambda,\gamma) - \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma) \\
&= \sum_{a\in A}\int_{c\in C}\mathcal{L}(a,c;\lambda,\gamma)x^*(a,dc) - \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma) \\
&= \sum_{a\in A}\int_{c\in C}(\mathcal{L}(a,c;\lambda,\gamma) - \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma))x^*(a,dc).
\end{aligned}
$$

Therefore, $x^*\in\arg\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma)$ if and only if $\mathcal{L}(a,c;\lambda,\gamma) = \max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma)$ a.s. with respect to probability x^*, which is equivalent to $(a,c)\in Z$ a.s. with respect to

x^*. ☐

Even more surprisingly, it turns out that the lottery problem and the dual problem of the deterministic problem are essentially the same problem. The following theorem shows that these two problems share the same Lagrangian multipliers and the same optimal value.

Definition 2.2 *We define the dual problem of the deterministic problem (2.1) as*

$$\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{a\in A,\, c\in C} \mathcal{L}(a,c;\lambda,\gamma), \tag{2.9}$$

where $\mathcal{L}$ is defined in (2.5). If the infimum is attained at (λ^*,γ^*), then we call (λ^*,γ^*) a solution to the dual problem (2.9).

Theorem 2.4 *We assume that x^* is the solution to system (2.2), and the Lagrangian multipliers corresponding to x^* exist, denoted as $(\lambda^*,\gamma^*)\in\mathbb{R}^m\times\mathbb{R}^{\ell|A|}$. Then (λ^*,γ^*) is the solution to the dual problem of the deterministic problem, i.e.*

$$\max_{a\in A,\, c\in C} \mathcal{L}(a,c;\lambda^*,\gamma^*) = \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{a\in A,\, c\in C} \mathcal{L}(a,c;\lambda,\gamma), \tag{2.10}$$

where $\mathcal{L}$ is defined in (2.5). Furthermore, the optimal objective value of the dual problem (2.9) is the same as the optimal objective value of the lottery problem (2.2), i,e.

$$\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{a\in A,\, c\in C} \mathcal{L}(a,c;\lambda,\gamma) = \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc). \tag{2.11}$$

Proof We note that the equality (2.6) in Theorem 2.3 holds for any $(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}$. Therefore, (2.6) in Theorem 2.3 further yields that

$$\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) = \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{a\in A, c\in C} \mathcal{L}(a,c;\lambda,\gamma), \tag{2.12}$$

where L is defined in (2.4). According to (2.6) and (2.12), $\mathcal{L}(a,c;\lambda,\gamma)$ in (2.10) and (2.11) can be replaced to $L(x;\lambda,\gamma)$, i.e. to prove Theorem 2.4 it suffices to show that

$$\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda^*,\gamma^*) = \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma), \tag{2.13}$$

and

$$\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) = \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc). \tag{2.14}$$

We first prove (2.14). Since x^* satisfies all the constraints in (2.2), for any $\lambda \geq 0$, $\gamma \geq 0$, we know that

$$\sum_{i=1}^{n} \lambda_i \sum_{a\in A} \int_{c\in C} g_i(a,c)x^*(a,dc) + \sum_{j=1}^{\ell} \sum_{a\in A} \gamma_{j,a} \int_{c\in C} h_j(a,c)x^*(a,dc) \leq 0,$$

and hence

$$\begin{aligned}
&L(x^*;\lambda,\gamma)\\
=&\sum_{a\in A} \int_{c\in C} f(a,c)x^*(a,dc) - \sum_{i=1}^{n} \lambda_i \sum_{a\in A} \int_{c\in C} g_i(a,c)x^*(a,dc) - \sum_{j=1}^{\ell} \sum_{a\in A} \gamma_{j,a} \int_{c\in C} h_j(a,c)x^*(a,dc)\\
\geq&\sum_{a\in A} \int_{c\in C} f(a,c)x^*(a,dc).
\end{aligned} \tag{2.15}$$

Since $(\lambda,\, \gamma)$ is arbitrary chosen in $\mathbb{R}_+^m \times \mathbb{R}^{\ell|A|}$, the inequality (2.15) yields

$$\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} L(x^*;\lambda,\gamma) \geq \sum_{a\in A} \int_{c\in C} f(a,c)x^*(a,dc),$$

and further

$$\max_{x\in\mathcal{P}(A\times C)} \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} L(x;\lambda,\gamma) \geq \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} L(x^*;\lambda,\gamma) \geq \sum_{a\in A} \int_{c\in C} f(a,c)x^*(a,dc). \tag{2.16}$$

Due to the general inf-sup inequality, (2.16) implies that

$$\begin{aligned}
&\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma)\\
\text{(Inf-sup Inequality)} \geq &\max_{x\in\mathcal{P}(A\times C)} \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} L(x;\lambda,\gamma)\\
(2.16) \geq &\sum_{a\in A} \int_{c\in C} f(a,c)x^*(a,dc).
\end{aligned} \tag{2.17}$$

On the other hand, since (λ^*,γ^*) is the Lagrangian multipliers corresponding to x^*, we have

$$L(x^*;\lambda^*,\gamma^*) = \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda^*,\gamma^*), \tag{2.18}$$

implying that

$$
\begin{aligned}
&\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) \\
\leq &\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda^*,\gamma^*) \\
= &L(x^*;\lambda^*,\gamma^*).
\end{aligned}
\tag{2.19}
$$

Due to the complementary conditions satisfied by the Lagrangian multipliers, we have

$$
\begin{aligned}
&L(x^*;\lambda^*,\gamma^*) \\
=&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc) - \sum_{i=1}^{n}\lambda_i^*\sum_{a\in A}\int_{c\in C} g_i(a,c)x^*(a,dc) - \sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}^*\int_{c\in C} h_j(a,c)x^*(a,dc) \\
=&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc).
\end{aligned}
\tag{2.20}
$$

We combine (2.19) and (2.20) to obtain

$$
\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) \leq \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc). \tag{2.21}
$$

The equality (2.14) is then deduced from (2.17) and (2.21). Furthermore, the equality (2.13) can be implied directly from (2.14), (2.18), and (2.20). We can then combine (2.13) and (2.14) to finish the proof. □

Remark 2.1 *The theorems bridge the gap between the dual determinstic problem and the lottery problem, which will be utilized for the theoretical analysis of the algorithm in the next section.*

The theorems imply the following corollary that gives general conditions for nondegenerate lotteries to be optimal.

Corollary 2.1 *The Lagrangian function in $\mathcal{P}(A\times C)$, L, admits a nondegenerate lottery maximizer if and only if one of the following holds.*

1. *The Lagrangian function in $A\times C$, $\mathcal{L}$, has at least two different maximal points at the optimal λ^*,γ^*.*

2. *The deterministic saddle point problem does not satisfy strong duality, i.e.*

$$
\sup_{a\in A,\, c\in C} \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \mathcal{L}(a,c;\lambda,\gamma) \neq \inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} \sup_{a\in A,\, c\in C} \mathcal{L}(a,c;\lambda,\gamma).
$$

We use this result in Section 4 below to derive necessary conditions for lotteries to be optimal in a setting of Mirrleesian taxation with multi-dimensional types.

It is established in convex analysis that the Lagrangian dual problem corresponds to the closed convexification of the perturbed primal problem (see Theorem 2.2 and Lemma 2.1 in Bi (2020), and Chapter XII in Urruty and Lemaréchal (1993)). In this section, we demonstrate that the lottery relaxation provides a means of achieving this convexification.

2.3 An Algorithm to solve the Lagrangian optimization problem

We utilize the above results to construct an algorithm for solving the lottery problem. Theorem 2.3 implies that for given (λ, γ) the Lagrangian optimization problem in the probability space $\mathcal{P}(A \times C)$

$$\max_{x \in \mathcal{P}(A \times C)} L(x; \lambda, \gamma)$$

can be simplified to the Lagrangian optimization problem in the pure strategy space $A \times C$

$$\max_{a \in A, c \in C} \mathcal{L}(a, c; \lambda, \gamma).$$

Theorem 2.4 indicates how the optimal λ^*, γ^* can be obtained from the deterministic problem. We define

$$V(\lambda, \gamma) := \max_{a \in A,\, c \in C} \mathcal{L}(a, c; \lambda, \gamma), \tag{2.22}$$

and minimize $V(\cdot)$ via sub-gradient descent.

Definition 2.3 *We consider a convex function $F : \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|} \to \mathbb{R}$. For any $(\lambda, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}$, the sub-gradient of F at the point (λ, γ) is defined as*

$$\partial F(\lambda, \gamma) = \{d \in \mathbb{R}^m \times \mathbb{R}^{\ell|A|} |\, F(\lambda', \gamma') \geq F(\lambda, \gamma) + (\lambda' - \lambda, \gamma' - \gamma) \cdot d,\ \forall (\lambda', \gamma') \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}\}.$$

Lemma 2.1 *The dual function V defined in (2.22) is a convex function. Furthermore, for any $(\lambda, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}$ and for any (a_λ, c_λ) that maximize $\mathcal{L}(a, c; \lambda, \gamma)$, i.e.*

$$(a_\lambda, c_\lambda) \in \arg \max_{a \in A,\, c \in C} \mathcal{L}(a, c; \lambda, \gamma),$$

the following is a negative sub-gradient of V at (λ, γ).

$$(\Delta\lambda,\, \Delta\gamma) := \begin{cases} \Delta\lambda_i = g_i(a_\lambda, c_\lambda), & i \in \{1, \cdots, m\}; \\ \Delta\gamma_{j,a} = h_j(a_\lambda, c_\lambda), & j \in \{1, \cdots, l\},\ a = a_\lambda; \\ \Delta\gamma_{j,a} = 0, & j \in \{1, \cdots, l\},\ a \neq a_\lambda. \end{cases} \tag{2.23}$$

Proof See Online Appendix A.1. $\square$

Assuming that the Lagrangian multipliers at the k-th iteration are denoted as λ^k and γ^k, we first solve the unconstrained optimization problem in $A \times C$

$$(a^k, c^k) \in \arg\max_{a\in A, c\in C} \mathcal{L}(a, c; \lambda, \gamma).$$

We then update the multipliers using the projected sub-gradient descent method:

$$\lambda_i^{k+1} = \max\{\lambda_i^k + \mu^k g_i(a^k, c^k), 0\},\ \forall i \in \{1, \cdots, n\};$$

$$\gamma_{j,a^k}^{k+1} = \max\{\gamma_{j,a^k}^k + \mu^k h_j(a^k, c^k), 0\},\ \forall j \in \{1, \cdots, l\};$$

and

$$\gamma_{j,a}^{k+1} = \gamma_{j,a}^k,\ \forall j \in \{1, \cdots, l\},\ a \neq a^k.$$

These updates follow the negative subgradient of the dual function, with a projection onto the nonnegative orthant to enforce the non-negativity of multipliers as required by the KKT conditions. Intuitively, when a constraint is violated, its associated multiplier increases, tightening the penalty; when satisfied, the multiplier relaxes.

Finally, we collect all these pure strategies (a^k, c^k) during the iterations to construct an approximated lottery solution as

$$x^N := \frac{1}{\sum_{k=1}^N \mu^k} \sum_{k=1}^N \mu^k \delta_{(a^k, c^k)}. \tag{2.24}$$

It seems surprising at first that the empirical frequencies of different actions along the iterations converge to the optimal probabilities. The key to understanding this construction lies in the proof of Proposition 2.1, where it is shown that for every constraint i and iteration n

$$\lambda_i^{n+1} \geq \lambda_i^1 + \sum_{k=1}^n \mu_k g_i(a^k, c^k) = \lambda_i^1 + \left(\sum_{k=1}^n \mu^k\right) \sum_{a\in A} \int_{c\in C} g_i(a, c) x^n(a, dc),$$

which directly implies that

$$\sum_{a\in A} \int_{c\in C} g_i(a, c) x^n(a, dc) \leq \frac{\lambda_i^{n+1} - \lambda_i^1}{\sum_{k=1}^n \mu^k}.$$

The updating rule for the Lagrangian multipliers can be regarded as estimating the contribution to the constraints for accumulating (a^k, c^k) as a new point in the support of the optimal lottery with the weight μ^k.

We formally present the full algorithm in Algorithm 1.

Algorithm 1: Lagrangian Iteration Algorithm

Given $\lambda_i^1 (i \in \{1, \cdots, m\})$, $\gamma_{j,a}^1 (a \in A,\ j \in \{1, \cdots, \ell\})$, $\mu^1, \mu^2, ... \in \mathbb{R}_+$, $N \in \mathbb{N}_+$.

For $k = 1 : N$

Step 1. Solve the Lagrangian problem.

$$(a^k, c^k) \in \arg \max_{a \in A, c \in C} \mathcal{L}(a, c; \lambda^k, \gamma^k).$$

Step 2. Update the Lagrangian multipliers.

$$\lambda_i^{k+1} = \max\{\lambda_i^k + \mu^k g_i(a^k, c^k), 0\},\ \forall i \in \{1, \cdots, n\}.$$

$$\gamma_{j,a^k}^{k+1} = \max\{\gamma_{j,a^k}^k + \mu^k h_j(a^k, c^k), 0\},\ \forall j \in \{1, \cdots, \ell\}.$$

$$\gamma_{j,a}^{k+1} = \gamma_{j,a}^k,\ \forall j \in \{1, \cdots, \ell\}, a \neq a^k.$$

End

Step 3. Construct the lottery solution.

$$x^N := \frac{1}{\sum_{k=1}^N \mu^k} \sum_{k=1}^N \mu^k \delta_{(a^k, c^k)},$$

where $\delta_{(a^k,c^k)}$ is the $\delta-$measure at the point (a^k, c^k).

It is important to note that there are $\ell|A|$ Lagrange multipliers associated with the constraints related to unobserved actions, as the system (2.2) requires the incentive constraints to be satisfied for all $a \in A$. However, in each iteration, we update only ℓ of these multipliers, specifically those corresponding to the form $\gamma_{\cdot,a^k}$.

A key part of the algorithm consists of the maximization problem

$$(a^k, c^k) \in \arg \max_{a \in A, c \in C} \mathcal{L}(a, c; \lambda^k, \gamma^k).$$

Since it is assumed that the set A is finite, the feasibility of the algorithm is based on solving $\max_{c \in C} \mathcal{L}(\bar{a}, c; \lambda^k, \gamma^k)$ for a given action $\bar{a} \in A$. While this is generally a nonconvex problem (in which case one needs to discretize the set C and search for the best element), there are several important economic applications where it is, in fact, a convex problem (see the moral hazard example below). If it is not convex, one might be able to formulate it as a simple polynomial optimization problem and use methods as in Lasserre (2001). In the worst-case scenario, one needs to discretize the set C.

2.4 Theorems

In this section, we establish that, under appropriate conditions on the learning rates μ^k, the lottery solution x^N generated by Algorithm 1 provides an approximation to the optimal lottery solution, with a precise definition of "approximation" to follow. From our results above and general convergence results for subgradient algorithms (see, e.g. Nedic and Bertsekas (2001)), we can expect our algorithm to have the following convergence properties. The Lagrangian multiplier (λ^k, γ^k) during iterations converges to the optimal Lagrange multiplier, (λ^*, γ^*); the value of the objective function for the constructed lottery converges to the optimal value of the objective function; and the constructed lotteries x^n converge to the optimal lottery. These results will be established in this section.

For any $\epsilon > 0$, we define the ϵ-optimal solution to the system (2.2), to describe a probability measure on $A \times C$, that approximately satisfies all the constraints, and approximately attains the maximal objective function value with error ϵ.

Definition 2.4 *(ϵ-optimal solution) We denote by x^* the optimal solution to system (2.2). We call $\tilde{x}^\epsilon \in \mathcal{P}(A \times C)$ an ϵ-optimal solution to system (2.2), if it satisfies*

1.
$$\sum_{a \in A} \int_{c \in C} g_i(a, c) \tilde{x}^\epsilon(a, dc) \leq \epsilon, \quad \forall i \in \{1, ..., m\};$$

2.
$$\int_{c \in C} h_j(a, c) \tilde{x}^\epsilon(a, dc) \leq \epsilon, \quad \forall j \in \{1, \cdots, l\},\, a \in A;$$

3.
$$\sum_{a \in A} \int_{c \in C} f(a, c) \tilde{x}^\epsilon(a, dc) - \sum_{a \in A} \int_{c \in C} f(a, c) x^*(a, dc) \geq -\epsilon.$$

Remark 2.2 *The first two properties in the definition imply that $\tilde{x}^\epsilon$ is a feasible probability measure for the relaxed system (2.3). However, $\tilde{x}^\epsilon$ is different from the solution to the relaxed system (2.3), since we do not require $\tilde{x}^\epsilon$ to exactly attain the maximal objective function value among all feasible probability measures for (2.3). However, property 3 in this definition actually implies that $\tilde{x}^\epsilon$ can be regarded as an approximate optimal solution to (2.3).*

The following theorem shows that a sequence of ϵ-optimal solutions, denoted as $\{\tilde{x}^\epsilon\}_{\epsilon > 0}$, will *-weak converge to the optimal solution of the lottery system (2.2) up to a subsequence as $\epsilon \to 0$. The proof for this theorem is similar to that for Theorem 2.2. This theorem shows that an ϵ-optimal solution can be considered as a "close" approximate solution to the original lottery system (2.2) when ϵ is sufficiently small.

Theorem 2.5 *We denote $\tilde{x}^{\epsilon}$ as an ϵ-optimal solution to the system (2.2). Then we can choose a sequence $\{\epsilon_n\} \to 0$, s.t. $\tilde{x}^{\epsilon_n}$ *-weak converges in the finite Borel measures space on $A \times C$, denoted as $\mathcal{M}(A \times C)$, to some $\tilde{x}^* \in \mathcal{P}(A \times C) \subset \mathcal{M}(A \times C)$, as $n \to \infty$, i.e. for any $\varphi \in C^0(A \times C)$, we have*

$$\sum_{a\in A}\int_{c\in C}\varphi(a,c)\tilde{x}^{\epsilon^n}(a,dc) \to \sum_{a\in A}\int_{c\in C}\varphi(a,c)\tilde{x}^*(a,dc),\ as\ n\to\infty.$$

Furthermore, $\tilde{x}^*$ is an optimal solution to (2.2).

Proof See Online Appendix A.4. □

We now examine the property of the lottery solution x^N generated by Algorithm 1. The first result shows that, if the Lagrangian multipliers λ^k, γ^k in Algorithm 1 converges, then, under some assumptions on the updating rate μ^k, the solution x^N is an $\epsilon-$optimal solution for arbitrary small $\epsilon > 0$ when N is sufficiently large.

Proposition 2.1 *We assume that the sequence $(\mu^k)_{k=1}^{\infty}$ satisfies*

$$\sum_{k=1}^{\infty}\mu^k = \infty.$$

We assume that (λ^k, γ^k) converges to some (λ^*, γ^*) when $k \to \infty$ and x^* is the solution to system (2.2). Then for any $\epsilon > 0$, there exists $N \in \mathbb{N}_+$, such that when $n > N$, x^n is an ϵ-optimal solution to system (2.2).

Proof See Online Appendix A.5. □

The following proposition addresses the behavior of the Lagrangian multiplier during iterations and gives sufficient conditions on the learning rates $(\mu^k)_{k=1}^{\infty}$, to guarantee that the Lagrangian multipliers in Algorithm 1 indeed converge to the minimizer of the dual problem.

Proposition 2.2 *We assume that the sequence $(\mu^k)_{k=1}^{\infty}$ satisfies*

$$\sum_{k=1}^{\infty}\mu^k = \infty, \quad and \quad \sum_{k=1}^{\infty}(\mu^k)^2 < \infty.$$

Let x^* be the solution to system (2.2), and corresponding Lagrangian multipliers to x^* exist. Then (λ^k, γ^k) generated by Algorithm 1 converge to some (λ^*, γ^*), where (λ^*, γ^*) is a minimizer of the dual problem

$$\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} V(\lambda,\gamma) := \max_{a\in A,\, c\in C}\mathcal{L}(a,c;\lambda,\gamma),$$

where $\mathcal{L}$ is defined in (2.5).

Proof See Online Appendix A.6. $\square$

We then obtain the following property of x^N as in Algorithm 1:

Theorem 2.6 *We assume that the sequence $(\mu^k)_{k=1}^{\infty}$ satisfies*

$$\sum_{k=1}^{\infty} \mu^k = \infty, \quad \text{and} \quad \sum_{k=1}^{\infty} (\mu^k)^2 < \infty.$$

Let x^* be the solution to system (2.2), and the corresponding Lagrangian multipliers to x^* exist. Then for any $\epsilon > 0$, there exists $N \in \mathbb{N}_+$, such that when $n > N$, x^n obtained in Algorithm 1 is an $\epsilon-$optimal solution to system (2.2).

Proof The theorem is proven directly by combining Proposition 2.1 and Proposition 2.2. $\square$

Theorem 2.6 demonstrates that (under the appropriate assumptions on the learning rates μ^k), the lottery we constructed by the algorithm is an approximation of the optimal lottery.

2.5 Computational Complexity and Comparison to Existing Methods

In this section, we analyze the computational complexity of Algorithm 1, and compare it to the linear programming approach widely used in the literature (Prescott and Townsend, 1984a; Prescott, 2004).

2.5.1 Computational Complexity

General Estimate We denote that $\lambda = (\lambda_i)_{i \in \{1,\cdots,m\}}$ and $\gamma = (\gamma_{j,a})_{j \in \{1,\cdots,\ell\}, a \in A}$. We define the function

$$\Lambda(\lambda, \gamma) = \sum_{i=1}^{m} \lambda_i^2 + \sum_{j=1}^{\ell} \sum_{a \in A} \gamma_{j,a}^2.$$

Suppose that all the assumptions in Theorem 2.6 hold, then (λ^k, γ^k) converges and hence remains bounded during the iterations in Algorithm 1. Additionally, we assume there exist two constants $M \geq 0, \bar{\Lambda} \geq 0$ such that $|g_i(a,c)| \leq M, |h_j(a,c)| \leq M$ and $\|(\lambda^k, \gamma^k)\|_\infty + \Lambda(\lambda^1, \gamma^1) \leq \bar{\Lambda}$ throughout the iterations.

First, we analyze the number of iterations required to obtain an ϵ-optimal lottery solution. We use the symbol $\gtrsim$ to denote greater than or equal to a constant multiple. The following proposition provides a rough estimate of the number of iterations required.

Proposition 2.3 *We take all the assumptions in Theorem 2.6. Additionally, we assume that there exist two constants $M \geq 0, \bar{\Lambda} \geq 0$ such that $\|g_i\|_\infty \leq M\,(i = 1, \cdots, m)$, $\|h_j\|_\infty \leq M\,(j = 1, \cdots, \ell)$, and $\|(\lambda^k, \gamma^k)\|_\infty + \Lambda(\lambda^1, \gamma^1) \leq \bar{\Lambda}$ throughout the iterations. We take $\mu^k \sim k^{-\frac{1}{2}(1+\rho)}$ for $0 < \rho \leq 1$. For $\epsilon > 0$, when*

$$n \gtrsim \begin{cases} \left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}} (m+\ell)^{\frac{1}{1-\rho}}, & \rho < 1 \\ e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{m+\ell}}{\epsilon}}, & \rho = 1, \end{cases} \tag{2.25}$$

the x^n obtained from Algorithm 1 is an ϵ-optimal solution.

Proof See Online Appendix A.7. □

Remark 2.3 *The estimate is constructed based on three fundamental inequalities: (A.10), (A.11), and (A.42) (see Online Appendix A.7 for details). Notably, each of these inequalities contains a term in the numerator that represents the difference between the Lagrangian multipliers at two different iterations. For further analysis, we have only utilized the boundedness of these multipliers. Theorem 2.2 establishes that the Lagrangian multipliers converge to the solution of the dual problem associated with the deterministic system. Furthermore, by Proposition 2.8 in Nedic and Bertsekas (2001), if we additionally assume that the dual problem is strongly convex in a neighborhood of the optimal Lagrangian multipliers, i.e.,*

$$V(\lambda, \gamma) - V(\lambda^*, \gamma^*) \geq K\, dist^2\left((\lambda, \gamma), \Lambda^*\right), \quad \forall (\lambda, \gamma) \text{ s.t. } dist\left((\lambda, \gamma), \Lambda^*\right) < \tau, \tag{2.26}$$

for some given $\tau > 0$, where V and Λ^* are defined in (2.22) and (A.25), respectively, then for any $\epsilon > 0$, we can select a learning rate satisfying $\mu^k \sim O(1/k)$ such that when

$$n \gtrsim \frac{M^2(m+\ell)}{\epsilon^2}, \tag{2.27}$$

it follows that $dist((\lambda^n, \gamma^n), \Lambda^*) < \epsilon$. This implies that if the number of iterations is chosen as in (2.25), then in the final iterations, the Lagrangian multipliers will be very close to Λ^*, thereby allowing for an improved estimate. This observation also suggests a practical algorithm for constructing lotteries: we can first run the iterative process for a certain number of rounds and then use the results from the later iterations to construct the lottery.

Next, we consider the computational complexity of each iteration. We discretize the set C into the same finite grid $\hat{C}$ as in the linear programming approach. For each iteration, we first solve the problem

$$(a^k, c^k) \in \arg\max_{a\in A, c\in \hat{C}} \left[f(a,c) - \sum_{i=1}^{m} \lambda_i^k g_i(a,c) - \sum_{j=1}^{\ell} \gamma_{j,a} h_j(a,c) \right]. \tag{2.28}$$

The computation of this optimization problem involves evaluating the Lagrangian function at $|A||\hat{C}|$ points, with each evaluation requiring $O(m+\ell)$ computations. Thus, the complexity of finding the maximum point of the Lagrangian function is $O\left(|A||\hat{C}|(m+\ell)\right)$. The complexity of updating the Lagrange multipliers is $O\left(m+\ell\right)$. Therefore, the complexity for each iteration is:

$$O\left(|A||\hat{C}|(m+\ell)\right). \tag{2.29}$$

Combining (2.25) and (2.29), we have the theorem for the overall computational complexity for finding an ϵ-optimal lottery solution as follows:

Theorem 2.7 *We take all the assumptions in Theorem 2.6. Additionally, we assume there exist two constants $M \geq 0, \bar{\Lambda} \geq 0$ such that $\|g_i\|_\infty \leq M\,(i = 1,\cdots,m)$, $\|h_j\|_\infty \leq M\,(j = 1,\cdots,\ell)$, and $\|(\lambda^k,\gamma^k)\|_\infty + \Lambda(\lambda^1,\gamma^1) \leq \bar{\Lambda}$ throughout the iterations. We take $\mu^k \sim k^{-\frac{1}{2}(1+\rho)}$ for $0 < \rho \leq 1$. Then for $\epsilon > 0$, the overall computational complexity for finding an ϵ-optimal lottery solution is*

$$\begin{cases} O\left(\left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}} (|A||\hat{C}|(m+\ell)^{1+\frac{1}{1-\rho}})\right), & \rho < 1; \\ O\left(e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{m+\ell}}{\epsilon}} (|A||\hat{C}|(m+\ell))\right), & \rho = 1. \end{cases} \tag{2.30}$$

Proof The overall complexity (2.30) can be directly obtained by multiplying (2.25) and (2.29). $\square$

Estimate for Partial First-order Approach As indicated above, a crucial part of our algorithm consists of solving the maximization problem in Step 1. So far, our complexity analysis assumed that this is done by discretizing C and simple grid-search. However, in many economic applications, the subproblem (2.28) exhibits specific structural properties that allow for more efficient solution methods. For example, in the moral hazard problem we considered in the next section with a utility function that is separable in action a and consumption c, the subproblem(2.28) satisfies the following properties

$$\begin{aligned} f(a,c) &= u_0(a) + \sum_{r=1}^{d} u_r(a) w_r(c_r), \quad \forall a \in A,\, c = (c_1, \cdots, c_d) \in C \subset \mathbb{R}^d \\ h_j(a,c) &= v_{j,0}(a) + \sum_{r=1}^{d} v_{j,r}(a) w_r(c_r), \quad \forall j \in \{1, \cdots, \ell\},\, a \in A,\, c = (c_1, \cdots, c_d) \in C \subset \mathbb{R}^d, \end{aligned} \tag{2.31}$$

where $C = \times_{r=1}^{d} [c_{r,\min}, c_{r,\max}]$, for some strictly increasing and strictly concave functions $w_r(c_r)$[7]. The following Lemma implies that, with the separable property (2.31), and the assumption that the resource constraints are strictly increasing and weakly convex with respective to each c_r, the first-order approach for c can be implemented without the need to grid the set C for solving the subproblem (2.28), further enhancing the algorithm's efficiency.

Lemma 2.2 *Given $\lambda > 0\,(\lambda \geq 0$ and there exists $i \in \{1, \cdots, m\}$ such that $\lambda_i > 0), \gamma \geq 0$. We assume that the separable property (2.31) holds and that the resource constraints $g_i(a,c)$ are strictly increasing and weakly convex with respective to each c_r. For any $a \in A$, we consider the following system:*

$$\begin{cases} \mathcal{A}(a,\gamma,r) \frac{\partial}{\partial c_r} w_r(c_r) = \sum_{i=1}^{m} \lambda_i \frac{\partial}{\partial c_r} g_i(a,c), & \mathcal{A}(a,r,\gamma) > 0; \\ c_r = c_{r,min}, & \mathcal{A}(a,r,\gamma) \leq 0, \end{cases} \tag{2.32}$$

where

$$\mathcal{A}(a,\gamma,r) = u_r(a) - \sum_{j=1}^{\ell} \gamma_{j,a} v_{j,r}(a).$$

If (2.32) admits a solution $c(a) \in C$ for any $a \in A$, then the solution (a^*, c^*) to the subproblem (2.28) should satisfy $c^* = c(a^*)$.

Proof See Online Appendix A.2. □

If all assumptions in Lemma 2.2 hold, then (2.32) can be applied for solving the subproblem (2.28). To be precise, the subproblem (2.28) can be solved by the following three-step procedure:

1. Computing $\mathcal{A}(a,\gamma,r)$ for all $a \in A$ and $r \in \{1, \cdots, d\}$;

[7] This property holds because each consumption c_r appears in f and h in a single, specifically concave form. See Section 3 for more details. The case there is more specific, as $w_1 = \cdots = w_{|Q|}$.

2. Determining $c(a)$ for all $a \in A$ by

$$\begin{cases} \mathcal{A}(a,\gamma,r)\frac{\partial}{\partial c_r}w_r(c_r(a)) = \sum_{i=1}^m \lambda_i \frac{\partial}{\partial c_r} g_i(a,c(a)), & \mathcal{A}(a,r,\gamma) > 0; \\ c_r = c_{r,\min}, & \mathcal{A}(a,r,\gamma) \le 0. \end{cases} \tag{2.33}$$

3. Determining optimal $a \in A$ for the subproblem (2.28) with $c = c(a)$.

If additionally each g_i is linear with respective to c[8], then $\sum_{i=1}^m \lambda_i \frac{\partial}{\partial c_r} g_i(a,c(a)) = \mathcal{B}(a,r,\lambda)$ is independent on c and (2.33) can be solved by

$$\begin{cases} c_r = \left(\frac{\partial w_r}{\partial c_r}\right)^{-1}\left(\frac{\mathcal{B}(a,r,\lambda)}{\mathcal{A}(a,r,\gamma)}\right), & \mathcal{A}(a,r,\gamma) > 0; \\ c_r = c_{r,\min}, & \mathcal{A}(a,r,\gamma) \le 0. \end{cases}$$

It is directly to check that the complexity for this three-step procedure is

$$O\left(\underbrace{d\ell|A|}_{\text{STEP1}} + \underbrace{dm|A|}_{\text{STEP2}} + \underbrace{d(m+\ell)|A|}_{\text{STEP3}}\right) \sim O\left(d(m+\ell)|A|\right), \tag{2.34}$$

where d is the dimension of the consumption space C.

We combine (2.34) and proposition 2.3 to obtain the complexity estimate for the first-order approach as following:

Theorem 2.8 *We take all the assumptions in Theorem 2.6. Additionally, we assume there exist two constants $M \ge 0, \bar{\Lambda} \ge 0$ such that $\|g_i\|_\infty \le M$ $(i = 1, \cdots, m)$, $\|h_j\|_\infty \le M$ $(j = 1, \cdots, \ell)$, and $\|(\lambda^k, \gamma^k)\|_\infty + \Lambda(\lambda^1, \gamma^1) \le \bar{\Lambda}$ throughout the iterations. Also, we assume that each g_i is linear with respective to c. We take $\mu^k \sim k^{-\frac{1}{2}(1+\rho)}$ for $0 < \rho \le 1$. Then for $\epsilon > 0$, the overall computational complexity for finding an ϵ-optimal lottery solution is*

$$\begin{cases} O\left(\left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}} d|A|(m+\ell)^{1+\frac{1}{1-\rho}}\right), & \rho < 1; \\ O\left(e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{m+\ell}}{\epsilon}} d|A|(m+\ell)\right), & \rho = 1. \end{cases} \tag{2.35}$$

where d is dimension of the consumption space C.

[8]This assumption holds for a broad class of problems, including the moral hazard problem discussed in the next section, where the resource constraint is imposed on the total consumption amount, see Section 3 for details. In general, equation (2.33) can be solved using standard methods, such as Newton's method. In this context, the complexity of solving the nonlinear equation (2.33) for a given $a \in A$ is polynomial in d and m, which is still significantly more efficient than discretizing the consumption set. The resulting complexity analysis would follow a structure similar to Theorem 2.8, where $|\hat{C}|$ is replaced by an expression in terms of d. For simplicity, we omit a rigorous analysis of these cases.

Proof The overall complexity (2.35) can be directly obtained by multiplying (2.25) and (2.34). □

Remark 2.4 *By using the first-order approach, we can replace $|\hat{C}|$ in Theorem 2.7 by d in the complexity estimate.*

2.5.2 Comparison to Existing Methods

In the literature, the most commonly used method for solving the lottery problem (2.2) is to formulate it as linear program and to use standard algorithms for linear programming (Prescott and Townsend, 1984a; Prescott, 2004). To use linear programming, we first need to discretize the set C into a finite grid $\hat{C}$. This transforms (2.2) into a finite-dimensional problem:

$$\begin{aligned} &\max_{\hat{x}\in\mathcal{P}(A\times\hat{C})} \sum_{a\in A}\sum_{c\in\hat{C}} f(a,c)\hat{x}(a,c), \\ \textbf{s.t.}\ &\sum_{a\in A}\sum_{c\in\hat{C}} g_i(a,c)\hat{x}(a,c) \le 0\ (i\in\{1,\cdots,m\}), \\ &\sum_{c\in\hat{C}} h_j(a,c)\hat{x}(a,c) \le 0\ (j\in\{1,\cdots,l\},\ a\in A). \end{aligned} \tag{2.36}$$

This formulation constitutes a standard linear programming problem, where $\hat{x}$ is a vector of dimension $|A||\hat{C}|$, and the number of inequality constraints is $\ell|A|+m$. According to classical theory on interior-point algorithms for linear programming (Nocedal and Wright, 2006), the computational complexity of solving this problem is $O\left(\log(1/\delta)\cdot(|A||\hat{C}|+\ell|A|+m)^{3.5}\right)$, where δ denotes the error for complementary conditions in interior-point algorithms.[9]

Comparing the complexity of our approach in Theorem 2.7 to that of the linear programming method, we see that in the case that $|\hat{C}|\sim|A|\sim\ell\gg m$,

$$|A||\hat{C}|(m+\ell)^{1+\frac{1}{1-\rho}} \sim |A|^{3+\frac{1}{1-\rho}} \ll |A|^7 \sim (|A||\hat{C}|+\ell|A|+m)^{3.5},$$

for ρ close to 0. Considering the fact that Algorithm 1 actually needs less iterations than the estimation in Theorem 2.3, the Lagrangian iteration algorithm can perform significantly better than using linear programming directly when facing problems with large $|A|$ and $|\hat{C}|$ and the required accuracy ϵ is not very stringent.

[9]In Prescott and Townsend (1984a); Prescott (2004), the Dantzig-Wolfe decomposition is employed to enhance the linear programming algorithm, addressing both efficiency and memory concerns. However, this algorithm requires the computation of extreme points of the linear constraints and the construction of iterations based on these extreme points. This suggests that the method effectively relies on an extended version of the simplex algorithm, which lacks well-established theoretical complexity results. Consequently, to provide a theoretical comparison of computational complexity, we focus on traditional interior-point methods.

We note that for a problem discretized from a continuous problem, the scales $|A|$ and $|\hat{C}|$ are chosen in relation to ϵ, such that the exact solution to the discretized problem is an ϵ-optimal solution to the continuous problem. Therefore, the complexity term involving ϵ, specifically $\left(\frac{\frac{M}{\rho}+\bar{\Lambda}+1}{\epsilon}\right)^{\frac{2}{1-\rho}}$, cannot be neglected when analyzing the overall complexity of our methods. However, for problems with high-dimensional action and consumption sets, this complexity term is small compared to the term we analyzed earlier, namely

$$|A||\hat{C}|(m+\ell)^{1+\frac{1}{1-\rho}}.$$

For instance, consider a problem with a d-dimensional action set and a d-dimensional consumption set, i.e., $A = \times_{i=1}^{d} A_i$ and $C = \times_{i=1}^{d} C_i$. A typical choice would be $|A| = \prod_{i=1}^{d} |A_i| \sim \left(\frac{1}{\epsilon}\right)^{\frac{d}{2}}$ and $|\hat{C}| = \prod_{i=1}^{d} |\hat{C}_i| \sim \left(\frac{1}{\epsilon}\right)^{\frac{d}{2}}$. [10] Therefore, when d is large,

$$\left(\frac{\frac{M}{\rho}+\bar{\Lambda}+1}{\epsilon}\right)^{\frac{2}{1-\rho}} \ll \left(\frac{1}{\epsilon}\right)^{\frac{d}{2}} = |A|,$$

and the complexity for our methods is smaller than the complexity for the linear programming method. Furthermore, for problems satisfying assumptions in Theorem (2.8), such as the moral hazard problems, the first-order approach for c can be implemented without the need to grid the set C, further improving the algorithm's efficiency by a factor of approximately $\frac{|\hat{C}|}{d}$, where d is the dimension of C.

Memory Advantage In practical computation, another advantage of the Lagrangian iteration algorithm is that the memory needed is significantly reduced. Using linear programming, we need memory for an element $x \in \mathcal{P}(A \times C)$ together with a series of constraints concerning this element, while in the Lagrangian iteration algorithm, we only need the memory for saving $m + \ell|A|$ Lagrange multipliers λ^k, γ^k, and the memory for solving a deterministic optimization sub-problem in each iteration.

3 Application to Moral Hazard Problems

In this section, we apply our method to a principal agent problem with moral hazard. We illustrate the method as well as the advantages of our algorithm over linear programming

[10] Heuristically, we define the function $f(x_1, \ldots, x_d) = -\frac{1}{d}\sum_{k=1}^{d} x_k^2$ over the domain $[-1,1]^d$. The maximum value of f should be 0. If we choose $h = \frac{1}{2N+1}$ and evaluate f on the discrete set $\{(2k+1)h : -N-1 \leq k \leq N\}^d$, the maximum value of f should be $-h^2$. Thus, to achieve an ϵ-optimal solution, we select $h \sim \sqrt{\epsilon}$, leading to the number of points $|\{(2k+1)h : -N-1 \leq k \leq N\}^d| \sim \left(\frac{1}{\epsilon}\right)^{\frac{d}{2}}$.

The agents can take unobserved actions $a \in A$, which influence the probability distribution of their output $q \in Q$ through conditional probability $p(q|a)$. The action set A and the output set Q are finite sets. The total amount of consumption goods available to the social planner is the sum of the outputs of all agents. An agent has utility $u(c, a)$ from action a and consumption $c \in C$, where the consumption set C is compact. In the deterministic solution, the social planner's goal is to allocate an action and a consumption contract $\{a, c(q)\}$ for different states of output to each agent to maximize the total utility:

$$\max_{a,c(q)} \sum_q p(q|a)u(c(q), a), \tag{3.1}$$

subject to the resource constraint and incentive compatibility constraint:

$$\begin{aligned} &\sum_q p(q|a)(c(q) - q) \leq 0; \\ &\sum_q p(q|a)u(c(q), a) \geq \sum_q p(q|\hat{a})u(c(q), \hat{a}),\ \forall \hat{a} \in A \end{aligned} \tag{3.2}$$

Comparing this problem to the general form (2.1), the function $f(a, c)$ is defined as $f(a, c) = \sum_q u(a, c(q))p(q|a)$. The resource constraint function is $g_1(a, c) = \sum_q (c(q) - q)p(q|a)$. The incentive constraint functions are $h_{\hat{a}}(a, c) = \sum_q p(q|\hat{a})u(c(q), \hat{a}) - \sum_q p(q|a)u(a, c(q))$, $\forall \hat{a} \in A$.

Lottery Solutions. We now consider the lottery solution for the static moral hazard problem. We now consider the problem in the space $\mathcal{P}(A \times C^{|Q|})$. An element $x \in \mathcal{P}(A \times C^{|Q|})$ can be expressed as $x = x(a, c) = x(a, c(q_1), \cdots, c(q_{|Q|}))$. The objective function (3.1) becomes

$$\sum_{a \in A} \int_{c \in C^{|Q|}} x(a, dc) \sum_{q \in Q} p(q|a)u(c(q), a). \tag{3.3}$$

The resource constraint in (3.2) is changed by

$$\sum_{a \in A} \int_{c \in C^{|Q|}} x(a, dc) \sum_{q \in Q} p(q|a)(c(q) - q) \leq 0 \tag{3.4}$$

and the incentive constraints in (3.2) become

$$\int_{c \in C^{|Q|}} x(a, dc) \sum_q p(q|a)u(c(q), a) \geq \int_{c \in C^{|Q|}} x(a, dc) \sum_q p(q|\hat{a})u(c(q), \hat{a}),\ \forall (a, \hat{a}) \in A \times A. \tag{3.5}$$

We define $\pi \in \mathcal{P}(C \times Q \times A)$ such that $\pi(dc, q, a) = x(a, dc)p(q|a)$, then (3.3) can be simplified as

$$\sum_{q,a} \int_c \pi(dc, q, a)u(c, a), \tag{3.6}$$

while the resource constraint (3.4) and (3.5) become

$$\sum_{q,a} \int_c \pi(dc, q, a)(c - q) \leq 0, \tag{3.7}$$

$$\text{and} \ \sum_{q} \int_c \pi(dc, q, a)u(c, a) \geq \sum_{q} \int_c \pi(dc, q, a)\frac{p(q|\hat{a})}{p(q|a)}u(c, \hat{a}), \quad \forall (a,\ \hat{a}) \in A \times A. \tag{3.8}$$

respectively. Additionally, since $\pi(dc, q, a) = x(a, dc)p(q|a)$, it must satisfy the coinciding condition

$$\int_c \pi(dc, \bar{q}, \bar{a}) = x(\bar{a}, dc)p(\bar{q}|\bar{a}) = p(\bar{q}|\bar{a}) \sum_{q} \int_c \pi(dc, q, \bar{a}), \quad \forall \bar{q} \in Q, \bar{a} \in A. \tag{3.9}$$

We combine (3.6),(3.7),(3.8) and (3.9) to formulate the lottery problem for the static moral hazard problem as follows:

$$\begin{aligned}
&\max_{\pi \in \mathcal{P}(C \times Q \times A)} \sum_{q,a} \int_c \pi(dc, q, a)u(c, a), \\
\textbf{s.t.} \ &\sum_{q,a} \int_c \pi(dc, q, a)(c - q) \leq 0; \\
&\int_c \pi(dc, \bar{q}, \bar{a}) = p(\bar{q}|\bar{a}) \sum_{q} \int_c \pi(dc, q, \bar{a}), \quad \forall \bar{q} \in Q, \bar{a} \in A; \\
&\sum_{q} \int_c \pi(dc, q, a)u(c, a) \geq \sum_{q} \int_c \pi(dc, q, a)\frac{p(q|\hat{a})}{p(q|a)}u(c, \hat{a}), \quad \forall (a,\ \hat{a}) \in A \times A;
\end{aligned} \tag{3.10}$$

In particular, if C is a finite set, we have the lottery problem for the static moral hazard problem as in Prescott (1998):

$$
\begin{aligned}
& \max_{\pi} \sum_{c,q,a} \pi(c,q,a) u(c,a), \\
\textbf{s.t.} \quad & \sum_{c,q,a} \pi(c,q,a)(c-q) \leq 0; \\
& \sum_{c} \pi(c,\bar{q},\bar{a}) = p(\bar{q}|\bar{a}) \sum_{c,q} \pi(c,q,\bar{a}), \quad \forall \bar{q}, \bar{a}; \\
& \sum_{c,q} \pi(c,q,a) u(c,a) \geq \sum_{c,q} \pi(c,q,a) \frac{p(q|\hat{a})}{p(q|a)} u(c,\hat{a}), \quad \forall a,\ (\hat{a} \neq a) \in A \times A; \\
& \sum_{c,q,a} \pi(c,q,a) = 1; \\
& \pi(c,q,a) \geq 0, \quad \forall c,q,a.
\end{aligned}
\tag{3.11}
$$

We now consider the Lagrangian Iteration algorithm for (3.10). Assume a separable utility function $u(c,a)$ of the form $u(c,a) = v(c) + w(a)$, where v is increasing and concave with $v'(c_{min}) = \infty$, and w is decreasing. Therefore, according to Lemma 2.2, the first-order approach can be applied to the Lagrangian sub-problem. To be precise, after straightforward computations, the first-order derivative of the Lagrangian function $\mathcal{L}$ with respect to $c(q)$ is given by:

$$
\begin{aligned}
\frac{\partial \mathcal{L}(a,c;\lambda^k,\gamma^k)}{\partial c(q)} &= \left[p(q|a) - \sum_{\hat{a} \in A} \gamma^k_{\hat{a},a} \left(p(q|\hat{a}) - p(q|a) \right) \right] v'(c) - \lambda^k p(q|a) \\
&:= \mathcal{A}(a,q,\gamma^k) v'(c) - \mathcal{B}(a,q,\lambda^k).
\end{aligned}
\tag{3.12}
$$

Here, $\mathcal{B}(a,q,\lambda^k) \geq 0$. When $\mathcal{A}(a,q,\gamma^k) > 0$, the right-hand side of (3.12) decreases with respect to c and $v'(c_{min}) = \infty$. Therefore, the optimal consumption $c^k(a,q,\lambda^k,\gamma^k)$ is determined by the first-order condition (FOC):

$$
c^k(a,q,\lambda^k,\gamma^k) = \max \left\{ c_{max}, (v')^{-1} \left(\frac{\mathcal{B}(a,q,\lambda^k)}{\mathcal{A}(a,q,\gamma^k)} \right) \right\}.
$$

If $\mathcal{A}(a,q,\gamma^k) \leq 0$, the right-hand side of (3.12) is non-positive, and thus the optimal consumption is:

$$
c^k(a,q,\lambda^k,\gamma^k) = c_{min}.
$$

Thus, we can analytically determine the optimal consumption for each iteration using the FOC.

3.1 Simplified algorithm

We present the algorithm for the static moral hazard problem using the FOC approach in Algorithm 2.[11]

Algorithm 2: Algorithm for the Static Moral Hazard Problem: FOC Approach

Let $a_1, a_2 \in A$. Given $\lambda^1, \gamma^1_{a_1,a_2}, \mu^1, \mu^2, \cdots \in \mathbb{R}_+$.

For $k = 1 : \infty$

1. Updating $\mathcal{A}$ and $\mathcal{B}$.

$$\mathcal{A}(a,q,\gamma^k) = p(q|a) - \sum_{\hat{a}\in A} \gamma^k_{\hat{a},a}\left(p(q|\hat{a}) - p(q|a)\right),\ \mathcal{B}(a,q,\lambda^k) = \lambda^k p(q|a)$$

2. Updating $c^k(a,q,\lambda^k,\gamma^k)$ by first-order conditions for each $a \in A$.

$$c^k(a,q,\lambda^k,\gamma^k) = \begin{cases} \max\{c_{max}, (v')^{-1}\frac{\mathcal{B}(a,q,\lambda^k)}{\mathcal{A}(a,q,\gamma^k)}\}, & \mathcal{A}(a,q,\gamma^k) > 0; \\ c_{min}, & \mathcal{A}(a,q,\gamma^k) \leq 0. \end{cases}$$

3. Updating $c^k(q)$ and a^k by finding the optimal a to maximize $\mathcal{L}(a, c^k(a,q,\lambda^k,\gamma^k), \lambda^k, \gamma^k)$.

$$a^k = \arg\max_{a\in A} \mathcal{L}(a, c^k(a,q,\lambda^k,\gamma^k), \lambda^k, \gamma^k),\ c^k(q) = c^k(a^k, q, \lambda^k, \gamma^k).$$

4. Update the Lagrangian multipliers λ^k and γ^k.

$$\lambda^{k+1} = \max\{\lambda^k + \mu^k \sum_q p(q|a^k)(c^k(q) - q), 0\}.$$

For $\hat{a} = 1 : |A|$

$$\gamma_{\hat{a},a^k} = \max\{\gamma_{\hat{a},a^k} + \mu^k \left[\sum_q p(q|\hat{a})u(c^k(q),\hat{a}) - p(q|a^k)u(c^k(q),a^k)\right], 0\}$$

End

End

$$x^N = \frac{1}{\sum_{k=1}^N \mu^k} \sum_{k=1}^N \mu^k \delta_{(a^k,c^k)}.$$

[11] Note that the incentive constraints associated with nearby actions $a \in A$ and $\hat{a} \in A$ may be weak for all consumption contracts. As a result, they contribute little to the Lagrangian and can be difficult to detect numerically. To better capture the effect of all incentive constraints in practical computation, it is useful to rescale them. In our numerical implementation below, we rescale each constraint $h_{\hat{a}}(a,c) = \sum_q p(q|\hat{a})u(c(q),\hat{a}) - \sum_q p(q|a)u(c(q),a)$ by dividing it by $|a - \hat{a}|^2$.

According to Theorem 2.8, the computational complexity for obtaining an ϵ-optimal lottery solution by Algorithm 2 is:

$$\begin{cases} O\left(\left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}}(|A|^{2+\frac{1}{1-\rho}}|Q|)\right), & \rho < 1; \\ O\left(e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{|A|}}{\epsilon}}(|A|^2|Q|)\right), & \rho = 1. \end{cases}$$

In comparison, the complexity of solving the linear programming in formulation of (3.10) by interior algorithm is

$$O\left(\log(1/\delta)\cdot(|A||Q||C|+|A|^2)^{3.5}\right),$$

where δ denotes the error tolerance for the complementary conditions in interior-point algorithms.

3.2 Solution Results and Comparison to Linear Programming

We consider Example 1 from Prescott (1998). In this example, the consumption set is $C=[0,2]$, the output set is $Q=\{0.5,1.5\}$, and the action set is $A=0.05:\Delta a:1.95$, where $\Delta a=0.025$. The utility function is defined as $u(a,c)=\sqrt{c}+0.8\sqrt{2-a}$. The relationship between output and action is given by:

$$p(q=1.5 \mid a)=\begin{cases} \frac{1-(1-a)^{0.2}}{2}, & \text{if } a<1, \\ \frac{1+(a-1)^{0.2}}{2}, & \text{if } a\geq 1. \end{cases}$$

The solution obtained using linear programming in Prescott (1998), where the set C is discretized to $\hat{C}=0:0.01:2$, is:

$$\pi(a=0.050)=0.0924,\ \pi(a=1.075)=0.9076$$

$$\pi(c=1.20 \mid q=0.5, a=0.05)=1,\ \pi(c=1.19 \mid q=1.5, a=0.05)=1.$$

$$\pi(c=0.54 \mid q=0.5, a=1.075)=0.5311,\ \pi(c=0.55 \mid q=0.5, a=1.075)=0.4689.$$

$$\pi(c=1.40 \mid q=1.5, a=1.075)=1.$$

We implement Algorithm 2 for this problem. We choose the initial Lagrangian multipliers as $\lambda^1 = 0.5$ and $\mu^1 = 0$. The learning rate is chosen as

$$\mu^k = \frac{1}{\left(k + \frac{1}{|\Delta a|^2}\right)^{0.8}}.$$

We iterate for $100/\Delta a = 4000$ rounds[12] , and the trajectories of λ^k, a^k, c^k, and $V(\lambda^k, \gamma^k)$ are shown in Figure 1.

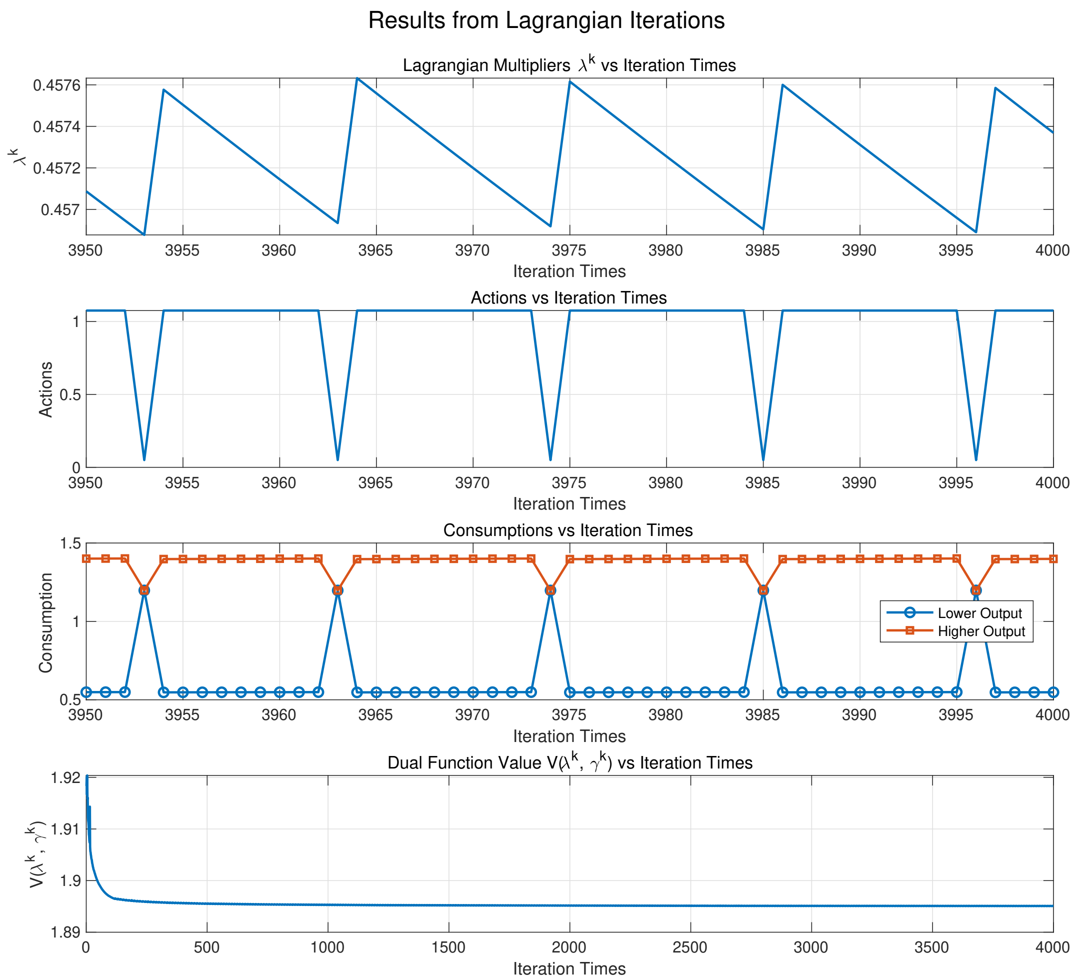

Figure 1: Relation between Lagrangian multipliers λ^k, optimal actions a^k, optimal consumptions c^k, value of the dual function $V(\lambda^k, \gamma^k)$ and Iteration Rounds when $\Delta a = 0.025$

As observed, a^k oscillates between 0.050 and 1.075 during iterations. Notably, there exists

[12]The total number of iterations, N, is chosen empirically, following the principle $N \propto \ell \sim \frac{1}{\Delta a}$, as derived from the condition (2.27).

a relationship between λ^k and a^k: when λ^k exceeds a certain threshold, indicating that the principal places greater emphasis on satisfying the resource constraints, the resulting action a^k stabilizes around 1.075. Conversely, when λ^k falls below this threshold, reflecting a reduced focus on the resource constraints, a^k stabilizes around 0.050. We use the results from rounds 3800 to 4000 to construct a lottery $\hat{\pi}$. The resulting lottery is:

$$\hat{\pi}(a = 0.050) = \frac{\sum_{k=3800}^{4000} \mu^k \mathbf{1}_{a^k=0.050}}{\sum_{k=3800}^{4000} \mu^k} \approx 0.0945.$$

This value is close to the solution obtained from linear programming, where $\pi(a = 0.050) = 0.0924$. When $a^k = 0.050$, the consumption contract approximates $c^k(q = 0.5) = c^k(q = 1.5) = 1.20$; and when $a^k = 1.075$, the consumption contract approximates $c^k(q = 0.5) = 0.55$ and $c^k(q = 1.5) = 1.40$. These results are consistent with those obtained by linear programming.

In terms of efficiency, the linear programming method takes 1.26 seconds using the optimization toolbox in MATLAB, whereas our method takes only 0.05 seconds to run 4000 Lagrangian iterations to obtain the approximate lottery solution. As shown in the last subfigure of Figure 1, the dual function $V(\lambda^k, \gamma^k)$ (defined in (2.22)) rapidly converges to the minimum $V(\lambda^*, \gamma^*)$ within the first few hundreds of iterations. The subsequent iterations primarily serve to construct lotteries and do not require large learning rates. This observation could inspire the careful design of learning rates to improve computational efficiency in practical applications.

We also perform $100/\Delta a$ iterations with varying Δa values. The computational time for each choice of Δa is presented in Table 1. In addition, the table provides the sizes of the corresponding linear programming problems, highlighting the direct implications on memory requirements.

Although the linear programming solver can also handle the problem with $\Delta a = 0.0125$ within 6.84 s, it cannot solve the finer grid case of $\Delta a = 0.00625$ on a regular laptop due to exceeding memory limits. In contrast, our Lagrangian iteration method remains computationally efficient and requires significantly less memory. Furthermore, Figure 2 below illustrates the trajectories of λ^k, a^k, c^k, and $V(\lambda^k, \gamma^k)$ for $\Delta a = 0.00625$. In this case, a^k oscillates between 1.0625 and 0.05, which differs from the case when $\Delta a = 0.025$, where a^k oscillates between 1.075 and 0.05. This observation suggests that finer discretization captures additional solutions, and, along with Figure 2, it indicates that 16,000 iterations are sufficient to approximate the lottery solution for this problem.

Δa	Iterations	CPU time	Size of LP		
			#Vars	#Equality Cons	#Inequality Cons
0.2	500	0.006	4020	21	91
0.1	1000	0.01	8040	41	381
0.05	2000	0.02	15678	79	1483
0.025	4000	0.05	30954	155	5853
0.0125	8000	0.16	61506	307	23257
0.00625	16000	0.98	122610	611	92721

Table 1: Computational Performance for Different Δa. "# Vars" refers to the number of variables in the LP method, "# Equality Cons" refers to the number of equality constraints, and "# Inequality Cons" refers to the number of inequality constraints, excluding the non-negativity constraints $\pi(c, q, a) \geq 0$.

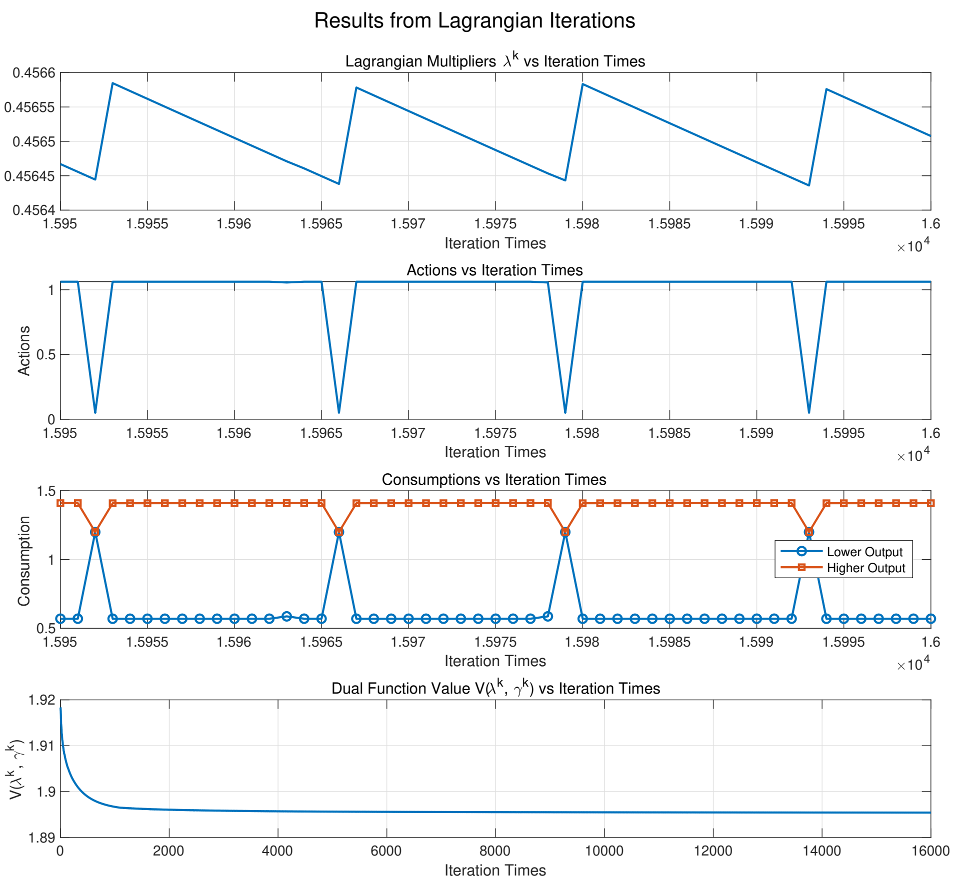

Figure 2: Relation between Lagrangian multipliers λ^k, optimal actions a^k, optimal consumptions c^k, value of the dual function $V(\lambda^k, \gamma^k)$ and Iteration Rounds when $\Delta a = 0.00625$.

4 Mirrleesian optimal taxation with multidimensional types

Following Mirrlees (1971) and Saez (2001), there is now a large literature on optimal income taxation. An important question is how their classical results change if one considers economies where tax-payers differ across several dimensions (see, e.g. Judd et al. (2017) Bergstrom and Dodds (2021), Boerma et al. (2022)). Solving these models is computationally challenging since the so-called first order approach is often not valid. As Judd et al. (2017) points out, there is no justification for this approach in the multidimensional case.

We make the multi-dimensional problem tractable by allowing for randomization. The planner chooses a lottery over consumption and output allocations for each type subject to incentive constraints that ensure that workers accurately report their types. We therefore deviate from the classical setting in Saez (2001) in two important dimensions. First, we assume that the planner can offer different tax-schedules to different types (this assumption is not uncommon in the literature, see, for example, Boerma et al. (2022)). Second, as in Weiss (1976), Brito et al. (1995), Hellwig (2007), we examine optimal taxation in an economy with a continuum of identical agents within each type where the planner can offer random tax-schedules.

In our setup, there are finitely many types of agents $h \in H$, a finite set, and a continuum of identical agents within each type. Each agent of type h has preferences over consumption and labor supply, which we represent by a continuous utility function $\bar{u}_h(c_h, \ell_h)$. Each agent has type-specific productivity w_h, $h \in H$, and produces $w_h \ell_h$ units of output when working ℓ_h hours. It is useful to write the indirect utility function over consumption and output as

$$u_h(c, y) = \bar{u}_h(c, \frac{y}{w_h})$$

The planner sets taxes as a function of output. An allocation is therefore $(c_h, y_h)_{h \in H} \in C \times Y$, where $C \subset \mathbb{R}^{|H|}$ and $Y \subset \mathbb{R}^{|H|}$ are assumed to be compact sets. A utilitarian social planner solves

$$\max_{x \in \mathcal{P}(C \times Y)} \int_{(c,y) \in C \times Y} \left(\sum_{h \in H} u_h(c_h, y_h) \right) x(dc, dy), \tag{4.1}$$

$$\text{s.t.} \quad \int_{(c,y) \in C \times Y} \left(\sum_h c_h - \sum_h y_h \right) x(dc, dy) \leq 0, \tag{4.2}$$

$$\int_{(c,y) \in C \times Y} \left(u_h(c_h, y_h) - u_h(c_{h'}, y_{h'}) \right) x(dc, dy) \geq 0 \text{ for all } h, h' \in H. \tag{4.3}$$

This problem is equivalent to the optimal taxation problem of maximizing (4.1) by choosing an income tax schedule $T_\omega(y)$ for each type and each possible realization of a public randomization device, ω, and letting each agent report their type, observe a public random signal ω and then choose output y and consumption $c = y - T_\omega(y)$ (see, e.g., Brito et al. (1995) or Hellwig (2007))

It is easy to verify that this problem fits directly into our general setup. Moreover, our algorithm can be efficiently applied even when the number of agents becomes large. The reason for this is that the Lagrangian to the deterministic problem can be written as

$$\begin{aligned}\mathcal{L}(c, y, \lambda, \gamma) &= \sum_{h \in H} u_h(c_h, y_h) + \sum_{h,h' \in H} \lambda_{h,h'}(u_h(c_h, y_h) - u_h(c_{h'}, y_{h'})) - \gamma(\sum_{h \in H} c_h - \sum_{h \in H} y_h) \\ &= \sum_{h \in H} \mathcal{L}_h(c_h, y_h, \lambda, \gamma),\end{aligned}$$

where $\lambda = (\lambda_{h,h'})_{h,h' \in H}$ and

$$\mathcal{L}_h(c_h, y_h, \lambda, \gamma) = (1 + \sum_{h' \in H, h' \neq h} \lambda_{h,h'}) u_h(c_h, y_h) - \sum_{h' \in H, h' \neq h} \lambda_{h',h} u_{h'}(c_h, y_h) - \gamma(c_h - y_h).$$

In Step 1 of our computational method, one therefore has to solve H maximization problems in two variables. The computational costs increase linearly in the number of agents.

To investigate optimal lotteries in this setting, we follow Judd et al. (2017) and assume that for each type $h \in H$, the utility function is given by

$$\bar{u}_h(c, y) = \log(c) - \psi_h \frac{\left(\frac{y}{w_h}\right)^{\frac{1}{\eta_h}+1}}{\frac{1}{\eta_h} + 1}. \tag{4.4}$$

We allow for heterogeneity in the Frisch elasticity η, productivity w and preference for leisure ψ. In general, one could also allow for heterogeneity in the utility for consumption, and Cole (1989) and Kehoe et al. (2002) examined lotteries in this setting. They point out that the common assumption of decreasing absolute risk aversion rules out lotteries. In contrast, there is typically no strong link between Frisch elasticity and productivity (see, e.g., Chetty et al. (2013)).

But can lotteries improve the objective of the planner's problem if elasticities are heterogeneous? Corollary 2.1 allows us to derive simple necessary conditions for this.

4.1 Necessary conditions for lotteries

Under Slater's condition, there exists a $p \in \mathcal{P}(C \times Y)$ satisfying the resource constraint and all incentive constraints strictly, if agents' utilities only differ via the parameters ψ_h and w_h, it is never optimal to employ lotteries. This is formalized in the following proposition.

Proposition 4.1 *Suppose that for all h, $\eta_h = \eta$. Given any $\epsilon > 0$, $c_{\max} > \epsilon$, $\ell_{\max} > 0$, we consider the planner problem (4.1) subject to (4.2) and (4.3), with the consumption set $C = [\epsilon, c_{\max}]^{|H|}$, the output set $Y = \times_{h\in H}[0, \ell_{max} w_h]$, and the utility function defined by (4.4). If Slater's condition holds, then the optimal solution to the planner problem does not involve lotteries.*

Proof See Online Appendix A.8. □

This proposition implies that the heterogeneity in Frisch elasticities is a necessary condition for lotteries to be optimal. To understand in which scenarios it is also sufficient, it is useful to consider a relaxed problem with partial incentive constraints as follows:

$$\int_{(c,y)\in C\times Y} \left(u_h(c_h, y_h) - u_h(c_{h'}, y_{h'})\right) x(dc, dy) \geq 0 \text{ for all } h, h' \text{ with } \eta_h \geq \eta_{h'}. \tag{4.5}$$

This constraint ensures that agents with lower Frisch elasticities will never pretend to be those with higher elasticities. Without an upper bound on the labor supply, the optimal value for the planner problem (4.1) subject to (4.2) and (4.5) can approximate that of the original problem arbitrarily well if we allow for lotteries. The key intuition is simple. The planner can use lotteries that assign an extremely high labor supply with extremely low probability to deter agents with low Frisch elasticity from imitating high elasticity agents. This is formalized in the following lemma.

Lemma 4.1 *The maximal objective value for the planner problem (4.1) subject to (4.2) and (4.3), and the maximal objective value for the planner problem (4.1) subject to (4.2) and (4.5) have the same limit as $\ell_{\max} \to \infty$.*

Proof See Online Appendix A.9. □

As the upper bound on maximal labor supply becomes very large, the lotteries needed to prevent low elasticity agents from reporting high elasticities become more and more degenerate. The following proposition proves that there will never be other lotteries.

Proposition 4.2 *We consider the planner problem (4.1) subject to (4.2) and (4.5) with the consumption set $C = \times_{h\in H}[\epsilon, c_{\max}]$, the output set $Y = \times_{h\in H}[0, \ell_{max} w_h]$, and the utility* *function defined by* (4.4)*. Assume that the Slater's condition holds, then the optimal solution to the planner problem does not involve lotteries.*

Proof See Online Appendix A.10. □

Putting together Lemma 4.1 and Proposition 4.2, the optimality of lotteries depends crucially on the upper bound on labor supply $\ell_{\max}$.[13]

4.2 Lottery solution and welfare gain in a calibrated example

We consider an example of two-dimensional heterogeneity where agents differ in productivity w_h as well as in labor supply elasticity η_h. Following Section 5.1 in Judd et al. (2017), we choose five values of $w_h \in \{1, 2, 3, 4, 5\}$ and five values of $\eta_h \in \{\frac{1}{8}, \frac{1}{5}, \frac{1}{3}, \frac{1}{2}, 1\}$, set $\psi_h = 1$, yielding 25 distinct types indexed by (w_h, η_h) and 600 incentive compatibility constraints. We take the optimal deterministic allocations from Judd et al. (2017) as a benchmark and compare them to our optimal lottery solutions in Table 2. We choose $\ell_{\max} = 1.2$ to nest Judd et al. (2017)'s optimal deterministic labor supply.

In order to make welfare comparisons economically meaningful, we denote welfare differences in terms of the Hicksian "compensating variation in resources" relative to the full-information (first-best) problem. Formally, let $\Delta^W(m)$ denote the highest level of social welfare that can be achieved in the full information setup when the economy has m fewer units of total resources, defined by:

$$\Delta^W(m) = \max_{x \in \mathcal{P}(C \times Y)} \int_{(c,y) \in C \times Y} \left(\sum_{h \in H} u_h(c_h, y_h) \right) x(dc, dy), \tag{4.6}$$

$$\text{s.t.} \int_{(c,\ell) \in C \times L} \left(\sum_h c_h - \sum_h y_h \right) x(dc, dy) \leq -m. \tag{4.7}$$

The function $\Delta^W(m)$ is monotonic and decreasing in m. Let u^L denote the optimal welfare of the lottery problem, and let u^D denote the optimal welfare of the deterministic problem. For each of $i \in \{L, D\}$, define m^i as the loss of resources in the full information problem that produces the same welfare level as u^i; that is, $\Delta^W(m^i) = u^i$. We then report welfare losses, WL^i, $i = L, D$, as the relative compensating variation in resources.

[13]Of course, this upper bound is not independent of the choice of ψ - we discuss this in more detail at the end of the next section.

Type (w, η)	Deterministic solution (c, y)	Lottery solution over c	Lottery solution over y	Full information solution (c, y)
1, 1	1.68, 0.42	1.697	0.43	3.17, 0.32
1, $\frac{1}{2}$	1.77, 0.62	1.789	0.63	3.17, 0.56
1, $\frac{1}{3}$	1.79, 0.65	1.803	0.65	3.17, 0.68
1, $\frac{1}{5}$	1.83, 0.77	1.842	0.77	3.17, 0.79
1, $\frac{1}{8}$	1.86, 0.86	1.876	0.86	3.17, 0.87
2, 1	1.86, 0.86	1.876	0.86	3.17, 1.26
2, $\frac{1}{2}$	2.03, 1.39	2.049	(1.38, 98.91%) (2.4, 1.09%)	3.17, 1.59)
2, $\frac{1}{3}$	2.07, 1.50	2.080	1.49	3.17, 1.72)
2, $\frac{1}{5}$	2.16, 1.74	2.179	1.75	3.17, 1.82
2, $\frac{1}{8}$	2.2, 1.83	2.220	1.84	3.17, 1.89
3, 1	2.2, 1.83	2.456	(1.74 78.08%) (3.6, 21.92%)	3.17, 2.84
3, $\frac{1}{2}$	2.47, 2.49	2.611	(2.62, 95.21%) (3.6, 4.79%)	3.17, 2.92
3, $\frac{1}{3}$	2.47, 2.49	2.622	(2.66, 96.26%) (3.6, 3.74%)	3.17, 2.95
3, $\frac{1}{5}$	2.55, 2.68	2.690	2.85	3.17, 2.97
3, $\frac{1}{8}$	2.62, 2.85	2.707	2.88	3.17, 2.98
4, 1	3.36, 4.00	3.176	(2.82, 63.74%) (4.8, 36.26%)	3.17, 5.06
4, $\frac{1}{2}$	3.36, 4.00	3.357	(4.02, 99.60%) (4.8, 0.40%)	3.17, 4.50
4, $\frac{1}{3}$	3.36, 4.00	3.361	4.03	3.17, 4.32
4, $\frac{1}{5}$	3.36, 4.00	3.361	4.03	3.17, 4.19
4, $\frac{1}{8}$	3.36, 4.00	3.361	4.03	3.17, 4.12
5, 1	4.87, 5.87	4.816	5.85	3.17, 7.90
5, $\frac{1}{2}$	4.49, 5.56	4.475	5.57	3.17, 6.28
5, $\frac{1}{3}$	4.34, 5.43	4.323	5.44	3.17, 5.82
5, $\frac{1}{5}$	4.11, 5.24	4.100	5.25	3.17, 5.48
5, $\frac{1}{8}$	4.00, 5.14	3.979	5.14	3.17, 5.29

Table 2: Optimal deterministic solution versus optimal lottery solution. Deterministic solutions are from Judd et al. (2017). For the lottery solutions, the deterministic allocations are shown directly; lotteries are listed with each value followed by the corresponding probabilities π_i.

In this example, the deterministic mechanism yields a welfare loss of 7.53% relative to the first best full-information benchmark. Allowing for lotteries reduces this loss to 7.27%, implying a relative welfare gain of approximately 3.46%. These gains stem from randomization over labor supply. This is consistent with convex preferences and linear technologies because of the non-convexity implied by the IC constraints. A high elasticity of labor supply implies a low effective risk aversion to lotteries over labor supply. In our example, agents with high labor elasticity and intermediate productivities (such as types

$(w, \eta) = (3, 1), (3, 1/2), (4, 1), (4, 1/2))$ are most likely to receive lottery allocations.

In the example, the welfare gains from randomization are rather small. If one assumes that it is costly to use lotteries, this might be consistent with the fact that one does not observe them being used in actual tax systems. However, as we increase $\ell_{\max}$, the welfare gains from the lottery tax scheme become more pronounced. In the limit $\ell_{\max} \to \infty$, the welfare losses from asymmetric information shrink to 3.33% under the optimal lottery mechanism. This implies that 55.8% of the welfare loss from asymmetric information can be eliminated through lotteries and that the welfare gains from lotteries can be very large.

Of course, for empirically grounded specifications of preferences, the value of $\ell_{\max}$ is typically normalized with corresponding adjustments to the labor disutility parameter ψ. Our calibration of $\psi = 1$ is relatively low compared to values used in the literature - for example, Kaplan et al. (2018) adopts $\psi = 2.2$ with $\ell_{\max} = 1$, and Heathcote et al. (2017) estimate a substantial dispersion of ψ in the data.

5 Conclusion

In this paper, we develop a new method to compute lottery solutions in non-convex economies. Using a Lagrangian framework and iteratively decoupling optimization and complementary slackness conditions, our method provides a computationally efficient and memory-scalable approach to these high-dimensional problems. This allows us to analyze models of private information that were previously intractable due to the limitations of traditional linear programming methods.

References

Armstrong, C. S., D. F. Larcker, and C.-L. Su (2010): "Endogenous selection and moral hazard in compensation contracts," Operations Research, 58, 1090–1106.

Arnott, R. and J. Stiglitz (1988): "Randomization with Asymmetric Information," RAND Journal of Economics, 19, 344–362.

Aumann, R. J. (1974): "Subjectivity and correlation in randomized strategies," Journal of mathematical Economics, 1, 67–96.

Bergstrom, K. and W. Dodds (2021): "Optimal taxation with multiple dimensions of heterogeneity," Journal of Public Economics, 200, 104442.

Bi, Y. (2020): Analysis of Convex Relaxations for Nonconvex Optimization, Cornell University.

Boerma, J., A. Tsyvinski, and A. P. Zimin (2022): "Bunching and taxing multidimensional skills," Tech. rep., National Bureau of Economic Research.

Brito, D. L., J. H. Hamilton, S. M. Slutsky, and J. E. Stiglitz (1995): "Randomization in optimal income tax schedules," Journal of Public Economics, 56, 189–223.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2013): "Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities," NBER macroeconomics Annual, 27, 1–56.

Cole, H. L. (1989): "Comment: general competitive analysis in an economy with private information," International Economic Review, 249–252.

Doepke, M. and R. M. Townsend (2006): "Dynamic mechanism design with hidden income and hidden actions," Journal of Economic Theory, 126, 235–285.

Gauthier, S. and G. Laroque (2014): "On the value of randomization," Journal of Economic Theory, 151, 493–507.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): "Optimal tax progressivity: An analytical framework," The Quarterly Journal of Economics, 132, 1693–1754.

Hellwig, M. F. (2007): "The undesirability of randomized income taxation under decreasing risk aversion," Journal of Public Economics, 91, 791–816.

Jerez, B. (2003): "A dual characterization of incentive efficiency," Journal of Economic Theory, 112, 1–34.

Judd, K. L., D. Ma, M. A. Saunders, and C.-L. Su (2017): "Optimal Income Taxation with Multidimensional Taxpayer Types," .

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary policy according to HANK," American Economic Review, 108, 697–743.

Kehoe, T. J., D. K. Levine, and E. C. Prescott (2002): "Lotteries, sunspots, and incentive constraints," Journal of Economic Theory, 107, 39–69.

Lasserre, J. B. (2001): "Global optimization with polynomials and the problem of moments," SIAM Journal on optimization, 11, 796–817.

Luenberger, D. G. (1997): Optimization by vector space methods, John Wiley & Sons.

Mirrlees, J. A. (1971): "An exploration in the theory of optimum income taxation," The review of economic studies, 38, 175–208.

——— (1999): "The theory of moral hazard and unobservable behaviour: Part I," The Review of Economic Studies, 66, 3–21.

Moser, C. and P. Olea de Souza e Silva (2019): "Optimal paternalistic savings policies," .

Myerson, R. B. (1982): "Optimal coordination mechanisms in generalized principal–agent problems," Journal of mathematical economics, 10, 67–81.

Nedic, A. and D. Bertsekas (2001): "Convergence rate of incremental subgradient algorithms," Applied Optimization, 54, 223–264.

Nocedal, J. and S. Wright (2006): Numerical Optimization, Springer Science & Business Media.

Pavoni, N., O. Setty, and G. L. Violante (2016): "The design of 'soft'welfare-to-work programs," Review of Economic Dynamics, 20, 160–180.

Pavoni, N. and G. L. Violante (2007): "Optimal welfare-to-work programs," The Review of Economic Studies, 74, 283–318.

Prescott, E. C. and R. M. Townsend (1984a): "General competitive analysis in an economy with private information," International Economic Review, 1–20.

——— (1984b): "Pareto optima and competitive equilibria with adverse selection and moral hazard," Econometrica: journal of the econometric society, 21–45.

Prescott, E. S. (1998): "Computing Moral Hazard Problems Using the Dantzig-Wolfe Decomposition Algorithm," .

——— (1999): "A primer on moral-hazard models," FRB Richmond Economic Quarterly, 85, 47–77.

——— (2004): "Computing solutions to moral-hazard programs using the Dantzig–Wolfe decomposition algorithm," Journal of Economic Dynamics and Control, 28, 777–800.

Prescott, E. S. and R. M. Townsend (2006): "Firms as clubs in Walrasian markets with private information," Journal of Political Economy, 114, 644–671.

Renner, P. and K. Schmedders (2015): "A polynomial optimization approach to principal–agent problems," Econometrica, 83, 729–769.

Saez, E. (2001): "Using elasticities to derive optimal income tax rates," The review of economic studies, 68, 205–229.

Shen, C., Y. Luo, and Z. Zhou (2023): "Fictitious Play via Finite Differences for Mean Field Games with Optimal Stopping," arXiv preprint arXiv:2310.05741.

Shor, N. Z. (2012): Minimization methods for non-differentiable functions, vol. 3, Springer Science & Business Media.

Stiglitz, J. E. (1982): "Self-selection and Pareto efficient taxation," Journal of public economics, 17, 213–240.

Strausz, R. (2006): "Deterministic versus stochastic mechanisms in principal–agent models," Journal of economic theory, 128, 306–314.

Su, C.-L. and K. L. Judd (2007): "Computation of moral-hazard problems," Tech. rep., working paper.

Urruty, J.-B. H. and C. Lemaréchal (1993): Convex analysis and minimization algorithms II, Springer-Verlag.

Weiss, L. (1976): "The desirability of cheating incentives and randomnness in the optimal income tax," Journal of Political Economy, 84, 1343–1352.

(Online) Appendix

A Proofs

A.1 Proof of Lemma 2.1

Given $(\lambda, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}$. For any $(\lambda', \gamma') \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}$, we have

$$\begin{aligned}
V(\lambda', \gamma') &= \max_{a\in A,\, c\in C} \mathcal{L}(a, c; \lambda', \gamma') \\
&= \max_{a\in A,\, c\in C} f(a, c) - \sum_{i=1}^{m} \lambda'_i g_i(a, c) - \sum_{j=1}^{\ell} \gamma'_{j,a} h_j(a, c) \\
&\geq f(a_\lambda, c_\lambda) - \sum_{i=1}^{m} \lambda'_i g_i(a_\lambda, c_\lambda) - \sum_{j=1}^{\ell} \gamma'_{j,a} h_j(a_\lambda, c_\lambda) \\
&= f(a_\lambda, c_\lambda) - \sum_{i=1}^{m} \lambda_i g_i(a_\lambda, c_\lambda) - \sum_{j=1}^{\ell} \gamma_{j,a} h_j(a_\lambda, c_\lambda) + (\lambda' - \lambda, \gamma' - \gamma) \cdot (-(\Delta\lambda, \Delta\gamma)) \\
&= V(\lambda, \gamma) + (\lambda' - \lambda, \gamma' - \gamma) \cdot (-(\Delta\lambda, \Delta\gamma)) .
\end{aligned} \tag{A.1}$$

For any (λ'_1, γ'_1), $(\lambda'_2, \gamma'_2) \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}$ and $0 \leq \mu \leq 1$, such that

$$\mu(\lambda'_1, \gamma'_1) + (1 - \mu)(\lambda'_2, \gamma'_2) = (\lambda, \mu),$$

(A.1) implies that

$$\begin{aligned}
&\mu V(\lambda'_1, \gamma'_1) + (1 - \mu) V(\lambda'_2, \gamma'_2) \\
\geq &\mu \left(V(\lambda, \gamma) + ((\lambda'_1, \gamma'_1) - (\lambda, \gamma)) \cdot (-(\Delta\lambda, \Delta\gamma))\right) \\
&+ (1 - \mu) \left(V(\lambda, \gamma) + ((\lambda'_2, \gamma'_2) - (\lambda, \gamma)) \cdot (-(\Delta\lambda, \Delta\gamma))\right) \\
= &V(\lambda, \gamma),
\end{aligned}$$

implying that V is convex. Furthermore, by the definition of a sub-gradient we know from (A.1) that $(\Delta\lambda,\, \Delta\gamma)$ is a negative sub-gradient of V at (λ, γ).

A.2 Proof of Lemma 2.2

With the property (2.31), the subproblem (2.28) can be written as

$$(a^*, c^*) \in \arg \max_{a \in A, c \in \hat{C}} \mathcal{L}(a, c; \lambda, \gamma) := \mathcal{A}_0(a, \gamma, 0) + \sum_{r=1}^{d} \mathcal{A}(a, \gamma, r) w_r(c_r) - \sum_{i=1}^{m} \lambda_i g_i(a, c). \quad \text{(A.2)}$$

We define the function on c as

$$\mathcal{F}(c; a^*, \lambda, \gamma) := \mathcal{L}(a, c^*; \lambda, \gamma).$$

Then (A.2) yields that c^* is the maximizer of $\mathcal{F}$, given a^*, λ, and γ. It is direct to show that

$$\frac{\partial \mathcal{F}(c; a^*, \lambda, \gamma)}{\partial c_r} = \frac{\partial \mathcal{L}(a^*, c; \lambda, \gamma)}{\partial c_r} = \mathcal{A}(a^*, \gamma, r) \frac{\partial}{\partial c_r} w_r(c_r) - \sum_{i=1}^{m} \lambda_i \frac{\partial}{\partial c_r} g_i(a^*, c). \quad \text{(A.3)}$$

When $\mathcal{A}(a^*, \gamma, r) \leq 0$, the right-hand side of (A.3) is negative, implying that $\mathcal{F}$ is strictly decreasing with respect to c_r, and hence $c_r^* = c_r(a^*) = c_{r,\min}$.

We then define $\hat{c} := (c_r)_{r \in \{A(a^*,\gamma,r)>0\}}$ and $\tilde{c} := (c_r)_{r \in \{A(a^*,\gamma,r)\leq 0\}}$, and hence $c = (\hat{c}, \tilde{c})$. According to the discussion above, we know that $\tilde{c}^* = \tilde{c}(a^*) = (c_{r,\min})_{r \in \{A(a^*,\gamma,r)\leq 0\}}$. By the fact that c^* maximizes $\mathcal{F}(c; a^*, \lambda, \gamma)$, it is then straightforward to show that $\hat{c}^*$ should maximize $\mathcal{F}(\hat{c}, \tilde{c}^*; a^*, \lambda, \gamma)$, or equivalently, $\hat{c}^*$ should maximize

$$\sum_{r \in \{A(a^*,\gamma,r)>0\}} \mathcal{A}(a^*, \gamma, r) w_r(c_r) - \sum_{i=1}^{m} \lambda_i g_i(a^*, \hat{c}, \tilde{c}^*), \quad \text{(A.4)}$$

which is a strictly concave function in $\hat{c}$. Therefore, the following first order condition to (A.4),

$$\mathcal{A}(a^*, \gamma, r) \frac{\partial}{\partial c_r} w_r(c_r) = \sum_{i=1}^{m} \lambda_i \frac{\partial}{\partial c_r} g_i(a^*, \hat{c}, \tilde{c}^*), \quad r \in \{A(a^*, \gamma, r) > 0\},$$

admits at most one solution $\hat{c}(a^*)$. From the assumption we know that $(\hat{c}(a^*), \tilde{c}^*) \in C$, then we have $c^* = (\hat{c}(a^*), \tilde{c}^*)$. Hence, we have shown that (a^*, c^*) satisfies the system (A.3).

A.3 Proof of Theorem 2.2

For any $\epsilon > 0$, we know that $x^\epsilon \in \mathcal{P}(A \times C) \subset \mathcal{M}(A \times C)$, and hence

$$\sum_{a \in A} \int_{c \in C} |x^\epsilon(a, dc)| = \sum_{a \in A} \int_{c \in C} x^\epsilon(a, dc) = 1,$$

which is uniformly bounded with respective to ϵ. According to Alaoglu's theorem, we can choose a sequence $\{\epsilon_n\} \to 0$, such that x^{ϵ_n} *-weak converges to some $x^* \in \mathcal{M}(A \times C)$, i.e. for any $\varphi(a,c) \in C^0(A \times C)$, we have

$$\sum_{a\in A} \int_{c\in C} \varphi(a,c) x^{\epsilon_n}(a,dc) \to \sum_{a\in A} \int_{c\in C} \varphi(a,c) x^*(a,dc). \tag{A.5}$$

Next, we prove that x^* is the optimal solution to system (2.2).

We take $\varphi_1 \equiv 1$ in (A.5), and obtain

$$\sum_{a\in A} \int_{c\in C} x^*(a,dc) = 1,$$

implying that $x^* \in \mathcal{P}(A \times C)$.

We take $\varphi_2 \equiv g_i$ for some $i \in \{1, \cdots, m\}$ in (A.5), and obtain

$$\begin{aligned} &\sum_{a\in A} \int_{c\in C} g_i(a,c) x^*(a,dc) \\ &= \lim_{\epsilon_n \to 0} \sum_{a\in A} \int_{c\in C} g_i(a,c) x^{\epsilon_n}(a,dc) \\ &\le \lim_{\epsilon_n \to 0} \epsilon_n = 0. \end{aligned}$$

Similarly, for $j \in \{1, \cdots, \ell\}, a' \in A$, we take

$$\varphi(a,c) = \begin{cases} h_j(a',c), & a = a'; \\ 0, & otherwise. \end{cases}$$

we can prove that

$$\int_{c\in C} h_j(a',c) x^*(a',dc) \le 0.$$

Therefore x^* satisfy all the constraints in (2.2) and is a feasible probability measure to (2.2).

To show that x^* is an optimal solution to system (2.2), it then suffices to show that for any $x \in \mathcal{P}(A \times C)$, which is a feasible measure to (2.2), we have

$$\sum_{a\in A} \int_{c\in C} f(a,c) x^*(a,dc) \ge \sum_{a\in A} \int_{c\in C} f(a,c) x(a,dc). \tag{A.6}$$

We prove (A.6) by contradiction. If (A.6) does not hold, then we can find $\tilde{x} \in \mathcal{P}(A \times C)$, such that $\tilde{x}$ satisfy all the constraints in (2.2) and

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc) < \sum_{a\in A}\int_{c\in C} f(a,c)\tilde{x}(a,dc). \tag{A.7}$$

By the continuity of f and the *-weak convergence of x^{ϵ_n}, we know that

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc) = \lim_{\epsilon_n\to 0}\sum_{a\in A}\int_{c\in C} f(a,c)x^{\epsilon_n}(a,dc),$$

and hence we can find $\epsilon_{n_0} > 0$, such that

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^{\epsilon_{n_0}}(a,dc) < \sum_{a\in A}\int_{c\in C} f(a,c)\tilde{x}(a,dc)$$

according to (A.7). However, we know that $\tilde{x}$ is a feasible probability measure for (2.2) and hence a feasible probability measure for (2.3) with $\epsilon = \epsilon_n$. This contradicts the fact that x^{ϵ_n} is the optimal solution to (2.3) with $\epsilon = \epsilon_n$. Therefore, we finish the proof.

A.4 Proof of Theorem 2.5

Following the same steps in the proof for Theorem 2.2, we can choose a sequence $\{\epsilon_n\}_{n=1}^{\infty} \to 0$, such that $\tilde{x}^{\epsilon_n}$ *-weak converges to some $\tilde{x}^* \in \mathcal{M}(A\times C)$, and $\tilde{x}^*$ satisfy all the constraints in the lottery system (2.2). We suppose that x^* is an optimal solution to (2.2), it suffices to show that

$$\sum_{a\in A}\int_{c\in C} f(a,c)\tilde{x}^*(a,dc) \geq \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc). \tag{A.8}$$

Indeed, by property 3 in the definition of an $\epsilon-$optimal solution, we know that

$$\sum_{a\in A}\int_{c\in C} f(a,c)\tilde{x}^{\epsilon_n}(a,dc) \geq \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc) - \epsilon_n. \tag{A.9}$$

The inequality (A.8) can then be directly obtained by taking limits on both sides of (A.9) .

A.5 Proof of Proposition 2.1

We first prove that property 1 of definition 2.4 holds, i.e.

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x^{\epsilon}(a,dc) \leq \epsilon, \quad \text{for } i \in \{1,\cdots,m\}.$$

For $i \in \{1, ..., m\}$, we have

$$\sum_{a \in A} \int_{c \in C} g_i(a,c) x^n(a,dc) = \frac{1}{\sum_{k=1}^n \mu_k} \sum_{k=1}^n \mu_k g_i(a^k, c^k).$$

by the definition of x^n that

$$x^n = \sum_{k=1}^n \delta_{(a^k,c^k)}.$$

By the updating rule for λ_i, written

$$\lambda_i^{k+1} = \max\{\lambda_i^k + \mu^k g_i(a^k,c^k), 0\} \geq \lambda_i^k + \mu^k g_i(a^k,c^k), \quad k = 1, \cdots, n,$$

we know that

$$\sum_{k=1}^n \lambda_i^{k+1} \geq \sum_{k=1}^n \left[\lambda_i^k + \mu^k g_i(a^k,c^k)\right],$$

which can be simplified as

$$\lambda_i^{n+1} \geq \lambda_i^1 + \sum_{k=1}^n \mu_k g_i(a^k,c^k) = \lambda_i^1 + \left(\sum_{k=1}^n \mu^k\right) \sum_{a \in A} \int_{c \in C} g_i(a,c) x^n(a,dc).$$

Hence

$$\sum_{a \in A} \int_{c \in C} g_i(a,c) x^n(a,dc) \leq \frac{\lambda_i^{n+1} - \lambda_i^1}{\sum_{k=1}^n \mu^k}. \tag{A.10}$$

Property 1 is then implied by the fact that $\lambda_i^{n+1} \to \lambda_i^*$ (which leads to the fact that the numerator $\lambda_i^{n+1} - \lambda_i^1$ is bounded) and $\sum_{k=1}^n \mu^k \to \infty$.

We then prove that property 2 of definition 2.4 holds. The spirit of the proof is the same as that for property 1. For $j \in \{1, \cdots, l\}, a \in A$, we have

$$\int_{c \in C} h_j(a,c) x^n(a,dc) = \frac{1}{\sum_{k=1}^n \mu^k} \sum_{\{k : a^k = a\}} \mu^k h_j(a, c^k),$$

by the definition of x^n. By the updating rule of $\gamma_{j,\hat{a},a}$, we have

$$\gamma_{j,a}^{n+1} \geq \gamma_{j,a}^1 + \sum_{\{k : a^k = a\}} \mu^k h_j(a,c^k) = \gamma_{j,a}^1 + \left(\sum_{k=1}^n \mu^k\right) \int_{c \in C} h_j(a,c) x^n(a,dc).$$

Hence

$$\int_{c \in C} h_j(a,c) x^n(a,dc) \leq \frac{\gamma_{j,a}^{n+1} - \gamma_{j,a}^1}{\sum_{k=1}^n \mu^k}. \tag{A.11}$$

Property 2 holds by the fact that $\gamma_{j,a}^n \to \gamma_{j,a}^*$ and $\sum_{k=1}^n \mu^k \to \infty$.

Now we prove property 3 of definition 2.4 holds, i.e.

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,dc) \geq \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc) - \epsilon, \tag{A.12}$$

where x^* is the optimal solution of (2.2). The proof for property 3 can be divided into two steps.

Step 1. We define the sets

$$\Lambda_1^g := \{i \in \{1, 2..., m\}, \lambda_i^* > 0\},\ \Lambda_1^h := \{(j,a) \in \{1,\cdots,l\} \times A, \gamma_{j,a}^* > 0\}.$$

$$\Lambda_2^g := \{i \in \{1, 2..., m\}, \lambda_i^* = 0\},\ \Lambda_2^h := \{(j,a) \in \{1,\cdots,l\} \times A, \gamma_{j,a}^* = 0\}.$$

In this step, we prove that when $i \in \Lambda_1^g$, we have

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x^n(a,dc) \to 0, \tag{A.13}$$

and when $(j,a) \in \Lambda_1^h$, we have

$$\int_{c\in C} h_j(a,c)x^n(a,dc) \to 0. \tag{A.14}$$

Before we prove (A.13) and (A.14), we first interpret these two formulas. Since we have already proven the property 1 and property 2 of definition 2.4 hold, we now have

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x^n(a,dc) \leq \epsilon, \quad \forall i \in \{1,\cdots,m\},$$

and

$$\int_{c\in C} h_j(a,c)x^n(a,dc) \leq \epsilon, \quad \forall (j,a) \in \{1,\cdots,\ell\} \times A.$$

These two formulas (A.13) and (A.14) further state that the complementary conditions

$$\lambda_i^* \sum_{a\in A}\int_{c\in C} g_i(a,c)x^n(a,dc) = 0, \quad \forall i \in \{1,\cdots,m\},$$

and

$$\gamma_{j,a}^* \int_{c\in C} h_j(a,c)x^n(a,dc) = 0, \quad \forall (j,a) \in \{1,\cdots,\ell\} \times A,$$

are approximately satisfied when n is sufficiently large, which in a way implies that when

$i \in \Lambda_1^g$ or $(j, a) \in \Lambda_1^h$, the corresponding constraints are active.

Now we prove (A.13). For every $i \in \Lambda_1^g$, according to the fact that $\lambda_i^n \to \lambda_i^* > 0$, we know that there exists $N_i \in \mathbb{N}_+$, such that when $n > N_i$, we have $\lambda_i^n > 0$. Then by the updating rule of λ_i^n, when $n \geq N_i$ we have

$$0 < \lambda_i^{k+1} = \max\{\lambda_i^k + \mu^k g_i(a^k, c^k), 0\},$$

and thus

$$\lambda_i^{n+1} = \lambda_i^n + \mu^k g_i(a^k, c^k), \quad \forall n \geq N_i.$$

Therefore

$$\begin{aligned}
\lambda_i^{N_i+n+1} &= \lambda_i^{N_i} + \sum_{k=1}^{n} \mu^{N_i+k} g(a^{N_i+k}, c^{N_i+k}) \\
&= \lambda_i^{N_i} + \sum_{k=1}^{N_i+n} \mu^k g(a^k, c^k) - \sum_{k=1}^{N_i} \mu^k g(a^k, c^k) \\
&= \lambda_i^{N_1} + \left(\sum_{k=1}^{N_i+n} \mu^k\right) \sum_{a\in A} \int_{c\in C} g_i(a, c) x^{N_i+n}(a, dc) - \left(\sum_{k=1}^{N_i} \mu^k\right) \sum_{a\in A} \int_{c\in C} g_i(a, c) x^{N_i}(a, dc), \quad \forall n \geq 0.
\end{aligned}$$

Hence

$$\sum_{a\in A} \int_{c\in C} g_i(a, c) x^{N_i+n}(a, dc) = \frac{\lambda_i^{N_i+n+1} - \lambda_i^{N_i}}{\sum_{k=1}^{N_i+n} \mu^k} + \frac{\sum_{k=1}^{N_i} \mu^k}{\sum_{k=1}^{N_i+n} \mu^k} \sum_{a\in A} \int_{c\in C} g_i(a, c) x^{N_i}(a, dc). \quad \text{(A.15)}$$

Since $\lambda_i^n \to \lambda_i^*$ (hence $\lambda_i^{N_i+n+1} - \lambda_i^{N_i}$ is bounded) and $\sum_{k=1}^{N_i+n} \mu^k \to \infty$, the right hand side of (A.15) converges to 0 as $n \to \infty$, and we have finished the claim (A.13). The proof for (A.14) is the same as the one for (A.13) so we omit the details here.

Step 2. We then utilize these claims (A.13) and (A.14) to prove (A.12). By the definition of x^n, we have

$$\sum_{a\in A} \int_{c\in C} f(a, c) x^n(a, dc) = \frac{1}{\sum_{k=1}^{n} \mu^k} \sum_{k=1}^{n} \mu^k f(a^k, c^k) \quad \text{(A.16)}$$

By the fact that

$$(a^k, c^k) \in \arg \max_{a\in A, c\in C} f(a, c) - \sum_{i=1}^{m} \lambda_i^k g_i(a, c) - \sum_{j=1}^{\ell} \sum_{\hat{a}\in A} \gamma_{j,a}^k h_j(a, c),$$

we have

$$
\begin{aligned}
&f(a^k, c^k) - \sum_{i=1}^{m} \lambda_i^k g_i(a^k, c^k) - \sum_{j=1}^{\ell} \gamma_{j,a^k}^k h_j(a^k, c^k) \\
\geq & f(a, c) - \sum_{i=1}^{m} \lambda_i^k g_i(a, c) - \sum_{j=1}^{\ell} \gamma_{j,a}^k h_j(a, c), \text{ for all } a \in A, c \in C.
\end{aligned}
\tag{A.17}
$$

The key of the proof is to utilize the inequality (A.17) for each $1 \leq k \leq n$ and give a lower bound for the right hand side of (A.16). To be precise, for any $a \in A$, $c \in C$, we have

$$
\begin{aligned}
&\frac{1}{\sum_{k=1}^n \mu^k} \sum_{k=1}^{n} \mu^k f(a^k, c^k) \\
=&\frac{1}{\sum_{k=1}^n \mu^k} \sum_{k=1}^{n} \mu^k \Bigg(f(a^k, c^k) - \sum_{i=1}^{m} \lambda_i^k g_i(a^k, c^k) - \sum_{j=1}^{\ell} \gamma_{j,a^k}^k h_j(a^k, c^k) \\
&+ \sum_{i=1}^{m} \lambda_i^k g_i(a^k, c^k) + \sum_{j=1}^{\ell} \gamma_{j,a^k}^k h_j(a^k, c^k) \Bigg) \\
\geq&\frac{1}{\sum_{k=1}^n \mu^k} \sum_{k=1}^{n} \mu^k \Bigg(f(a, c) - \sum_{i=1}^{m} \lambda_i^k g_i(a, c) - \sum_{j=1}^{\ell} \gamma_{j,a}^k h_j(a, c) \\
&+ \sum_{i=1}^{m} \lambda_i^k g_i(a^k, c^k) + \sum_{j=1}^{\ell} \gamma_{j,a^k}^k h_j(a^k, c^k) \Bigg)
\end{aligned}
\tag{A.18}
$$

We utilize the relations $\lambda_i^k = (\lambda_i^k - \lambda_i^*) + \lambda_i^*$ and $\gamma_{j,a}^k = (\gamma_{j,a}^k - \gamma_{j,a}^*) + \gamma_{j,a}^*$ to simplify the terms $\sum_i^m \lambda_i^k g_i(a^k, c^k)$ and $\sum_{j=1}^{\ell} \gamma_{j,a^k}^k h_j(a^k, c^k)$ in the brackets on the right hand side of (A.18), and have

$$
\begin{aligned}
&\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k f(a^k,c^k)\\
\geq&\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left(f(a,c)-\sum_{i=1}^{m}\lambda_i^k g_i(a,c)-\sum_{j=1}^{\ell}\gamma_{j,a}^k h_j(a,c)\right.\\
&+\sum_{i=1}^{m}\lambda_i^* g_i(a^k,c^k)+\sum_{j=1}^{\ell}\gamma_{j,a^k}^* h_j(a^k,c^k)\\
&\left.+\sum_{i=1}^{m}(\lambda_i^k-\lambda_i^*)g_i(a^k,c^k)+\sum_{j=1}^{\ell}(\gamma_{j,a^k}^k-\gamma_{j,a^k}^*)h_j(a^k,c^k)\right)\\
=&f(a,c)-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{i=1}^{m}\lambda_i^k g_i(a,c)-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{j=1}^{\ell}\gamma_{j,a}^k h_j(a,c)\\
&+\sum_{i=1}^{m}\lambda_i^*\sum_{a\in A}\int_{c\in C}g_i(a,c)x^n(a,dc)+\sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}^*\int_{c\in C}h_j(a,c)x^n(a,dc)\\
&+\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left(\sum_{i=1}^{m}(\lambda_i^k-\lambda_i^*)g_i(a^k,c^k)\right)+\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left(\sum_{j=1}^{\ell}(\gamma_{j,a^k}^k-\gamma_{j,a^k}^*)h_j(a^k,c^k)\right).
\end{aligned}
\tag{A.19}
$$

We further define that

$$
I_1=f(a,c)-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{i=1}^{m}\lambda_i^k g_i(a,c)-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{j=1}^{\ell}\gamma_{j,a}^k h_j(a,c),
$$

$$
I_2=\sum_{i=1}^{m}\lambda_i^*\sum_{a\in A}\int_{c\in C}g_i(a,c)x^n(a,dc)+\sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}^*\int_{c\in C}h_j(a,c)x^n(a,dc),
$$

and

$$
I_3=\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left(\sum_{i=1}^{m}(\lambda_i^k-\lambda_i^*)g_i(a^k,c^k)\right)+\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left((\gamma_{j,a^k}^k-\gamma_{j,a^k}^*)h_j(a^k,c^k)\right).
$$

The inequality (A.19) can then be written as

$$
\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k f(a^k,c^k)\geq I_1+I_2+I_3. \tag{A.20}
$$

Next we give lower bound estimation of I_3, I_2 and I_1 respectively. For I_3, By the facts that

$\lambda_i^k \to \lambda_i^*$ and $\gamma_{j,a}^k \to \gamma_{j,a}^*$, together with the boundedness of g_i and h_j, we know that

$$\sum_{i=1}^{m}(\lambda_i^k - \lambda_i^*)g_i(a^k, c^k) + \sum_{j=1}^{\ell}(\gamma_{j,a^k}^k - \gamma_{j,a^k}^*)h_j(a^k, c^k) \to 0. \tag{A.21}$$

Since $\sum_{k=1}^{\infty} \mu^k = \infty$, (A.21) yields

$$I_3 = \frac{1}{\sum_{k=1}^{n} \mu^k} \sum_{k=1}^{n} \mu^k \left(\sum_{i=1}^{m} (\lambda_i^k - \lambda_i^*) g_i(a^k, c^k) \right) + \frac{1}{\sum_{k=1}^{n} \mu^k} \sum_{k=1}^{n} \mu^k \left(\sum_{\hat{a} \in A} (\gamma_{j,\hat{a},a^k}^k - \gamma_{j,\hat{a},a^k}^*) h_j(a^k, c^k, \hat{a}) \right) \to 0.$$

Hence, for any $\epsilon > 0$, there exists $N_1 \in \mathbb{N}_+$, such that when $n > N_1$,

$$I_3 > -\frac{\epsilon}{2}. \tag{A.22}$$

Next, we consider the lower bound of I_2. By the formulas (A.13) and (A.14) proven in step 1, we have

$$\begin{aligned}
I_2 &= \sum_{i=1}^{m} \lambda_i^* \sum_{a \in A} \int_{c \in C} g_i(a, c) x^n(a, dc) + \sum_{j=1}^{\ell} \sum_{a \in A} \gamma_{j,a}^* \int_{c \in C} h_j(a, c) x^n(a, dc) \\
&= \sum_{i \in \Lambda_1^g} \lambda_i^* \sum_{a \in A} \int_{c \in C} g_i(a, c) x^n(a, dc) + \sum_{(j,a) \in \Lambda_1^h} \gamma_{j,a}^* \int_{c \in C} h_j(a, c) x^n(a, dc) \\
&\to 0.
\end{aligned}$$

Therefore there exists $N_2 \in \mathbb{N}_+$, such that when $n > N_2$, we have

$$I_2 > -\frac{\epsilon}{2}. \tag{A.23}$$

We combine (A.20), (A.22) and (A.23) and see that when $n > \max\{N_1, N_2\}$, we have

$$\begin{aligned}
&\frac{1}{\sum_{k=1}^{n} \mu^k} \sum_{k=1}^{n} \mu^k f(a^k, c^k) \\
\geq & I_1 - \epsilon \\
= & f(a, c) - \frac{1}{\sum_{k=1}^{n} \mu^k} \sum_{k=1}^{n} \mu^k \sum_{i=1}^{m} \lambda_i^k g_i(a, c) - \frac{1}{\sum_{k=1}^{n} \mu^k} \sum_{k=1}^{n} \mu^k \sum_{j=1}^{\ell} \gamma_{j,a}^k h_j(a, c) - \epsilon
\end{aligned} \tag{A.24}$$

We multiply $x^*(a, dc)$ to both sides of (A.24) and integral on $A \times C$. The inequality (A.24) then implies

$$
\begin{aligned}
&\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k f(a^k,c^k)\\
\geq &\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc)-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{i=1}^{m}\lambda_i^k\sum_{a\in A}\int_{c\in C} g_i(a,c)x^*(a,dc)\\
&-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}^k\int_{c\in C} h_j(a,c)x^*(a,dc)-\epsilon\\
\geq &\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc)-\epsilon,
\end{aligned}
$$

where the last inequality holds by the fact that x^* satisfy all the constraints in (2.2). Hence we prove the property 3 and finish the proof for Proposition 2.1.

A.6 Proof of Proposition 2.2

We define the set of minimizers to the dual problem as

$$
\Lambda^* := \left\{(\lambda^*,\gamma^*)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|},\ V(\lambda^*,\gamma^*)=\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}} V(\lambda,\gamma)\right\}. \tag{A.25}
$$

(2.10) in Theorem 2.4 implies that Λ^* is a non-empty set. Furthermore, for any $(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}$, according to Lemma 2.1, $(\Delta\lambda,\Delta\gamma)$ defined in (2.23) is a negative sub-gradient of $V(\lambda,\gamma)$, and

$$
\|(\Delta\lambda,\Delta\gamma)\|_2^2=\sum_{i=1}^{m}|g_i(a_\lambda,c_\lambda)|^2+\sum_{j=1}^{\ell}|h_j(a_\lambda,c_\lambda)|^2
$$

is uniformly bounded due to the continuity of the constraints g_i, h_j, and the compactness of the consumption set C and the action set A. The theorem therefore follows from Proposition 2.7 in Nedic and Bertsekas (2001).

A.7 Proof of Proposition 2.3

The proof can be divided into two steps. At the first step, we show the properties of the lottery solution x^n generated from Algorithm 1 as follows:

$$
\sum_{a\in A}\int_{c\in C} g_i(a,c)x^n(a,\mathrm{d}c)\leq\frac{\bar{\Lambda}}{\sum_{k=1}^{n}\mu^k}; \tag{A.26}
$$

$$\int_{c\in C} h_j(a,c)x^n(a,\mathrm{d}c) \le \frac{\bar{\Lambda}}{\sum_{k=1}^n \mu^k}; \tag{A.27}$$

and

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c) \ge \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c) - \frac{M(m+\ell)\sum_{k=1}^n(\mu^k)^2+\bar{\Lambda}}{\sum_{k=1}^n \mu^k}. \tag{A.28}$$

At the second step, we show that we can choose μ^k satisfying the condition $\mu^k \sim k^{-\frac{1}{2}(1+\rho)}$, such that when (2.25) holds, x^n is an ϵ-optimal solution.

Step 1. According to (A.10) and (A.11) in the proof of Proposition 2.1, we have:

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x^n(a,\mathrm{d}c) \le \frac{\lambda_i^{n+1}-\lambda_i^1}{\sum_{k=1}^n \mu^k} \le \frac{\lambda_i^{n+1}}{\sum_{k=1}^n \mu^k} \le \frac{\bar{\Lambda}}{\sum_{k=1}^n \mu^k};$$

and

$$\int_{c\in C} h_j(a,c)x^n(a,\mathrm{d}c) \le \frac{\gamma_{a,j}^{n+1}-\gamma_{a,j}^1}{\sum_{k=1}^n \mu^k} \le \frac{\gamma_{a,j}^{n+1}}{\sum_{k=1}^n \mu^k} \le \frac{\bar{\Lambda}}{\sum_{k=1}^n \mu^k}.$$

Therefore, (A.26) and (A.27) hold.

It then remains to show that (A.28) holds in this step. The proof for (A.28) can be further divided into three sub-steps. First we show that

$$\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c) \\
\ge&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c) \\
&+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k) + \frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^{\ell}\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k).
\end{aligned} \tag{A.29}$$

Then, we give an estimate of the term

$$\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k) + \frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^{\ell}\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k) \tag{A.30}$$

in the right hand side of (A.29). Finally, we conclide (A.28).

1-1. Show that (A.29) holds. According to (A.16) and (A.18) in the proof of Proposition 2.1, for every $a \in A$, $c \in C$, we have[14]:

$$
\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c)\\
\geq&\frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k\left(f(a,c)-\sum_{i=1}^m \lambda_i^k g_i(a,c)-\sum_{j=1}^{\ell}\gamma_{j,a}^k h_j(a,c)+\sum_{i=1}^m \lambda_i^k g_i(c^k)+\sum_{j=1}^{\ell}\gamma_{j,a^k}^k h_j(a^k,c^k)\right)\\
=&f(a,c)-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a,c)-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^{\ell}\sum_{k=1}^n \mu^k\gamma_{j,a}^k h_j(a,c)\\
&+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^{\ell}\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k).
\end{aligned}
\tag{A.31}
$$

We multiply $x^*(a,dc)$ to both sides in (A.31) and integral on $A\times C$ and have

$$
\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c)\\
\geq&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c)-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k\sum_{i=1}^m \lambda_i^k\sum_{a\in A}\int_{c\in C} g_i(a,c)x^*(a,dc)\\
&-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k\sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}^k\int_{c\in C} h_j(a,c)x^*(a,dc)\\
&+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^{\ell}\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k).
\end{aligned}
\tag{A.32}
$$

Since $x^*(a,dc)$ satisfies all the constraints in (2.2), i.e.

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x^*(a,dc)\leq 0,\qquad \int_{c\in C} h_j(a,c)x^*(a,dc)\leq 0,$$

we thus obtain the estimate (A.29) from (A.32).

1-2. Estimate the term (A.30). Recall that (A.30) is

$$\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^{\ell}\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k).$$

[14]There will be a slightly abuse of notations here. The (a,c) in the first line just means the variable of integration; and the (a,c) in the rest refers to the arbitrary chosen $a\in A$, $c\in C$.

By direct computation, we have

$$
\begin{aligned}
&\Lambda(\lambda^{k+1},\gamma^{k+1}) - \Lambda(\lambda^k,\gamma^k)\\
=&\left(\sum_{i=1}^m 2\lambda_i^k\left(\lambda_i^{k+1}-\lambda_i^k\right) + \sum_{j=1}^{\ell} 2\gamma_{j,a^k}^k\left(\gamma_{j,a^k}^{k+1}-\gamma_{j,a^k}^k\right)\right) + \left(\sum_{i=1}^m \left(\lambda_i^{k+1}-\lambda_i^k\right)^2 + \sum_{j=1}^{\ell}\left(\gamma_{j,a^k}^{k+1}-\gamma_{j,a^k}^k\right)^2\right).
\end{aligned}
\tag{A.33}
$$

We recall that the updating rules for λ_i, written

$$
\lambda_i^{k+1} = \max\{\lambda_i^k + \mu^k g_i(a^k,c^k), 0\},
$$

can be rewritten as

$$
\lambda_i^{k+1} - \lambda_i^k = \mu^k\tau_i^k g_i(a^k,c^k), \tag{A.34}
$$

for some $0 \le \tau_i^k \le 1$, where $\tau_i^k < 1$ if and only if $g_i(a^k,c^k) < 0$ and $\lambda_i^{k+1} = 0$. Similarly, there exists $0 \le \tau_{j,a^k}^k \le 1$, such that

$$
\gamma_{j,a^k}^{k+1} - \gamma_{j,a^k}^k = \mu^k\tau_{j,,a^k}^k h_j(a^k,c^k), \tag{A.35}
$$

where $\tau_{j,a^k}^k < 1$ if and only if $h_j(a^k,c^k) < 0$ and $\gamma_{j,a^k}^{k+1} = 0$. Therefore we can utilize (A.34) and (A.35) to simplify the right hand side of (A.33) and obtain

$$
\begin{aligned}
&\Lambda(\lambda^{k+1},\gamma^{k+1}) - \Lambda(\lambda^k,\gamma^k)\\
=&\left(\sum_{i=1}^m 2\mu^k\tau_i^k g_i(a^k,c^k)\lambda_i^k + \sum_{j=1}^{\ell} 2\mu^k\tau_{j,a^k}^k h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
&+\left(\sum_{i=1}^m (\mu^k)^2(\tau_i^k)^2 g_i^2(a^k,c^k) + \sum_{j=1}^{\ell}(\mu^k)^2(\tau_{j,a^k}^k)^2 h_j^2(a^k,c^k)\right)\\
=&2\mu^k\left(\sum_{i=1}^m g_i(a^k,c^k)\lambda_i^k + \sum_{j=1}^{\ell} h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
&-2\mu^k\left(\sum_{i=1}^m (1-\tau_i^k)g_i(a^k,c^k)\lambda_i^k + \sum_{j=1}^{\ell}(1-\tau_{j,a^k}^k)h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
&+\left(\sum_{i=1}^m (\mu^k)^2(\tau_i^k)^2 g_i^2(a^k,c^k) + \sum_{j=1}^{\ell}(\mu^k)^2(\tau_{j,a^k}^k)^2 h_j^2(a^k,c^k)\right).
\end{aligned}
\tag{A.36}
$$

Furthermore, when $1-\tau_i^k > 0$, we know that $\lambda_i^{k+1} = 0$, and according to (A.34), we have

$$
\lambda_i^k = -\mu^k\tau_i^k g_i(a^k,c^k). \tag{A.37}
$$

Similarly, when $1-\tau_{j,a^k}^k>0$, we have the following equality according to (A.35)

$$\gamma_{j,a^k}^k=-\mu^k\tau_{j,a^k}^k h_j(a^k,c^k). \tag{A.38}$$

We utilize the equality (A.37) and the equality (A.38) to obtain

$$\begin{aligned} &-2\mu^k\left(\sum_{i=1}^m(1-\tau_i^k)g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}(1-\tau_{j,a^k}^k)h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\ =&2(\mu^k)^2\left(\sum_{i=1}^m(1-\tau_i^k)\tau_i^k g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(1-\tau_{j,^k}^k)\tau_{j,a^k}^k h_j^2(a^k,c^k)\right). \end{aligned} \tag{A.39}$$

We combine (A.36) and (A.39), then we have

$$\begin{aligned} &\Lambda(\lambda^{k+1},\gamma^{k+1})-\Lambda(\lambda^k,\gamma^k)\\ =&2\mu^k\left(\sum_{i=1}^m g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\ &-2\mu^k\left(\sum_{i=1}^m(1-\tau_i^k)g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}(1-\tau_{j,a^k}^k)h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\ &+\left(\sum_{i=1}^m(\mu^k)^2(\tau_i^k)^2g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(\mu^k)^2(\tau_{j,a^k}^k)^2h_j^2(a^k,c^k)\right)\\ =&2\mu^k\left(\sum_{i=1}^m g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\ &2(\mu^k)^2\left(\sum_{i=1}^m(1-\tau_i^k)\tau_i^k g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(1-\tau_{j,^k}^k)\tau_{j,a^k}^k h_j^2(a^k,c^k)\right)\\ &+\left(\sum_{i=1}^m(\mu^k)^2(\tau_i^k)^2g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(\mu^k)^2(\tau_{j,a^k}^k)^2h_j^2(a^k,c^k)\right).\\ =&2\mu^k\left(\sum_{i=1}^m g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\ &+(\mu^k)^2\left(\sum_{i=1}^m(2-\tau_i^k)\tau_i^k g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(2-\tau_{j,a^k}^k)\tau_{j,a^k}^k h_j^2(a^k,c^k)\right), \end{aligned}$$

which implies that

$$\begin{aligned}
&2\mu^k\left(\sum_{i=1}^{m} g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell} h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
&=\left(\Lambda(\lambda^{k+1},\gamma^{k+1})-\Lambda(\lambda^k,\gamma^k)\right)\\
&\quad-(\mu^k)^2\left(\sum_{i=1}^{m}(2-\tau_i^k)\tau_i^k g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(2-\tau_{j,a^k}^k)\tau_{j,a^k}^k h_j^2(a^k,c^k)\right).
\end{aligned}\tag{A.40}$$

We add up equality (A.40) for $k=1,\cdots,n$. After some direct computation, we have an estimate to (A.30) as

$$\begin{aligned}
&\frac{1}{\sum_{k=1}^n\mu^k}\sum_{i=1}^{m}\sum_{k=1}^{n}\mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n\mu^k}\sum_{j=1}^{\ell}\sum_{k=1}^{n}\mu^k\gamma_{j,a^k}^k h_j(a^k,c^k)\\
=&\frac{\Lambda(\lambda^{n+1},\gamma^{n+1})-\Lambda(\lambda^1,\gamma^1)}{2\sum_{k=1}^n\mu^k}\\
&-\frac{\sum_{k=1}^n(\mu^k)^2\left(\sum_{i=1}^m(2-\tau_i^k)\tau_i^k g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(2-\tau_{j,a^k}^k)\tau_{j,a^k}^k h_j^2(a^k,c^k)\right)}{2\sum_{k=1}^n\mu^k}.
\end{aligned}\tag{A.41}$$

1-3. We then utilize the estimate (A.41) to derive (A.28). We combine (A.29) and (A.41) and obtain

$$\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c)\\
\geq&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c)+\frac{\Lambda(\lambda^{n+1},\gamma^{n+1})-\Lambda(\lambda^1,\gamma^1)}{2\sum_{k=1}^n\mu^k}\\
&-\frac{\sum_{k=1}^n(\mu^k)^2\left(\sum_{i=1}^m(2-\tau_i^k)\tau_i^k g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(2-\tau_{j,a^k}^k)\tau_{j,a^k}^k h_j^2(a^k,c^k)\right)}{2\sum_{k=1}^n\mu^k}.
\end{aligned}\tag{A.42}$$

By the fact that $|g_i|\leq M$, $|h_j|\leq M$, the estimate (A.42) further implies

$$\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c)\\
\geq&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c)-\frac{M^2(m+\ell)\sum_{k=1}^n(\mu^k)^2+\Lambda(\lambda^1,\gamma^1)}{2\sum_{k=1}^n\mu^k}\\
\geq&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c)-\frac{M^2(m+\ell)\sum_{k=1}^n(\mu^k)^2+\bar{\Lambda}}{\sum_{k=1}^n\mu^k}.
\end{aligned}$$

Therefore, we have shown that (A.28) holds.

Step 2. Choosing $\mu^k = \frac{1}{M\sqrt{m+\ell}k^{\frac{1}{2}(1+\rho)}}$, where $0 < \rho \leq 1$, gives:

$$\sum_{k=1}^{n} \mu^k \sim \begin{cases} \frac{n^{\frac{1}{2}(1-\rho)}}{M\sqrt{m+\ell}}, & \rho < 1 \\ \frac{\log n}{M\sqrt{m+\ell}}, & \rho = 1, \end{cases} \tag{A.43}$$

and

$$\sum_{k=1}^{n} (\mu^k)^2 \sim \frac{1}{M^2 \rho (m+\ell)}. \tag{A.44}$$

Thus, for any $\epsilon > 0$, when (2.25) holds, i.e.

$$n \gtrsim \begin{cases} \left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}} (m+\ell)^{\frac{1}{1-\rho}}, & \rho < 1 \\ e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{m+\ell}}{\epsilon}}, & \rho = 1, \end{cases}$$

we can implement (A.43) and (A.44) to (A.26),(A.27) and (A.28), and hence obtain:

$$\sum_{a \in A} \int_{c \in C} g_i(a,c) x^n(a, \mathrm{d}c) \leq \epsilon;$$

$$\int_{c \in C} h_j(a,c) x^n(a, \mathrm{d}c) \leq \epsilon;$$

and

$$\begin{aligned} &\sum_{a \in A} \int_{c \in C} f(a,c) x^n(a, \mathrm{d}c) \\ \geq &\sum_{a \in A} \int_{c \in C} f(a,c) x^*(a, \mathrm{d}c) - \epsilon. \end{aligned}$$

after some straightforward computation.

A.8 Proof of Proposition 4.1

The deterministic problem for (4.1) subject to (4.2) and (4.3) is written

$$
\begin{aligned}
&\max_{c\in C,\, y\in Y} \sum_{h\in H} \omega_h \left(\log(c_h) - \psi_h \frac{(\frac{y_h}{w_h})^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1} \right), \\
&\text{s.t.} \log(c_h) - \psi_h \frac{(\frac{y_h}{w_h})^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1} \geq \log(c_{h'}) - \psi_h \frac{(\frac{y_{h'}}{w_h})^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1}, \quad \forall h,\, h' \in H, \\
&\sum_{h\in H} \omega_h y_h \geq \sum_{h\in H} \omega_h c_h.
\end{aligned}
\tag{A.45}
$$

We define

$$
\tilde{\psi}_h = \frac{\psi_h}{w_h^{\frac{1}{\eta}+1}\left(\frac{1}{\eta}+1\right)}, \quad \forall h \in H.
$$

Then the deterministic problem (A.45) can be simplified as

$$
\begin{aligned}
&\max_{c\in C,\, y\in Y} \sum_{h\in H} \omega_h \left(\log(c_h) - \tilde{\psi}_h \frac{y_h^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1} \right), \\
&\text{s.t.} \log(c_h) - \tilde{\psi}_h \frac{(y_h)^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1} \geq \log(c_{h'}) - \tilde{\psi}_h \frac{(y_{h'})^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1}, \quad \forall h,\, h' \in H, \\
&\sum_{h\in H} \omega_h y_h \geq \sum_{h\in H} \omega_h c_h.
\end{aligned}
\tag{A.46}
$$

We introduce Lagrangian multipliers $\lambda_{h,h'} \geq 0$, $\forall h,\, h \in H$ for the incentive constraints and $\mu \geq 0$ for the resource constraint. The Lagrangian function of problem (A.45) is written:

$$
\begin{aligned}
\mathcal{L}(c, y; \lambda, \mu) =& \sum_{h\in H} \left[\left(\omega_h + \sum_{h'\in H} \lambda_{h,h'} - \sum_{h'\in H} \lambda_{h',h} \right) \log(c_h) - \mu\omega_h c_h \right] + \\
& \left[\left(-\omega_h - \sum_{h'\in H} \tilde{\psi}_h \lambda_{h,h'} + \sum_{h'\in H} \tilde{\psi}_{h'} \lambda_{h',h} \right) \frac{y_h^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1} + \mu\omega_h y_h \right] \\
:=& \sum_{h\in H} \left[\mathcal{A}_h(\lambda) \log(c_h) - \mu\omega_h c_h \right] + \left[-\mathcal{B}_h(\lambda) \frac{y_h^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1} + \mu\omega_h y_h \right].
\end{aligned}
\tag{A.47}
$$

Since the Slater's condition holds, according to Theorem 2.4, there exists (λ^*, μ^*) minimizing the dual function

$$
V(\lambda, \mu) := \max_{c\in C,\, y\in Y} \mathcal{L}(c, y; \lambda, \mu),
$$

and $V(\lambda^*, \mu^*)$ equals to the optimal objective value of the lottery problem (4.1) subject to (4.2) and (4.3). It is then straightforward to verify that $\mu^* > 0$ (when $\mu = 0$, we can take $\bar{c}_h \equiv c_{\max}$ and $\underline{y}_h \equiv 0$ to attain $\mathcal{L}(\bar{c}, \underline{y}; \lambda, 0)$ higher than the objective value of the optimal lottery solution).

Therefore, for any $h \in H$, $A_h(\lambda^*) \log(c_h) - \mu^* \omega_h c_h$ is whether strictly decreasing($A_h(\lambda^*) \leq 0$) or strictly concave($A_h(\lambda^*) > 0$) with respect to c_h. Hence the maximum c_h^* is unique and can be calculated by the first order condition when $A_h(\lambda^*) > 0$. To be precise, we have

$$c_h^* \in \arg\max \left(\mathcal{A}_h(\lambda^*) \log(c_h) - \mu^* \omega_h c_h\right) = \begin{cases} \{\min(\max(\frac{\mathcal{A}_h(\lambda^*)}{\mu^* \omega_h}, c_{\max}), \epsilon)\}, & \mathcal{A}_h(\lambda^*) > 0, \\ \{\epsilon\}, & \mathcal{A}_h(\lambda^*) \leq 0. \end{cases}$$

Similarly, the maximum y_h^* is also unique and can be calculated by

$$y_h^* \in \arg\max \left(-\mathcal{B}_h(\lambda^*) \frac{y_h^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1} + \mu^* \omega_h y_h\right) = \begin{cases} \{\min(\max(\left(\frac{\mu^* \omega_h}{\mathcal{B}_h(\lambda^*)}\right)^\eta, \ell_{\max} w_h), 0)\}, & \mathcal{B}_h(\lambda^*) > 0, \\ \{\ell_{\max} w_h\}, & \mathcal{B}_h(\lambda^*) \leq 0. \end{cases}$$

According to Result 1 of Corollary 2.1, the optimal solution to the planner problem (4.1) subject to (4.2) and (4.3) does not involve lotteries.

A.9 Proof of Lemma 4.1

We first introduce a Lemma for the proof.

Lemma A.1 *There exists $\bar{\ell} > 0$, s.t. when $\ell_{\max} \geq \bar{\ell}$, the solution to the planner problem (4.1) subject to (4.2) and (4.5) remains invariant with respect to $\ell_{\max}$.*

We denote $v_1(\ell_{\max})$ the maximal objective value of the planner problem (4.1) subject to (4.2) and (4.3); and $v_2(\ell_{\max})$ the maximal objective value of the planner problem (4.1) subject to (4.2) and (4.5). Since (4.5) involves less constraints than (4.3), we know $v_1(\ell_{\max}) \leq v_2(\ell_{\max})$ for any $\ell_{\max} > 0$. According to Lemma A.1, we have $v_2(\ell_{\max}) = v_2(\bar{\ell})$ for any $\ell_{\max} \geq \bar{\ell}$. Hence, to prove this lemma, it then suffices to show that

$$\begin{aligned} &\lim_{\ell_{\max} \to +\infty} v_2(\ell_{\max}) - v_1(\ell_{\max}) \\ =& \lim_{\ell_{\max} \to +\infty} (v_2(\ell_{\max}) - v_1(\ell_{\max}))_+ \\ =& \lim_{\ell_{\max} \to +\infty} (v_2(\bar{\ell}) - v_1(\ell_{\max}))_+ = 0. \end{aligned}$$

Step 1. The Slater's condition implies that, there exists $M > 0$, s.t. for any $\ell_{\max} > \bar{\ell}$, the Lagrangian multipliers for the planner problem (4.1) subject to (4.2) and (4.3) satisfies $\max\{\max_{h, h' \in H} \lambda^*_{h,h'}, \mu^*\} < M$.

To be precise, for $\ell_{\max} = \bar{\ell}$, according to the Slater's condition, there exists a probability distribution $\bar{p} \in \mathcal{P}(C \times Y)$, s.t. $\bar{p}$ satisfies all the constraints in (4.2) and (4.3) strictly. Hence, there exists $M > 0$, such that when $\max\{\max_{h,h' \in H} \lambda_{h,h'}, \mu\} \geq M$, we have $L(\bar{p}; \lambda, \mu) > v_2(\bar{\ell})$, where L denotes the Lagrangian function of the planner problem (4.1) subject to (4.2) and (4.3). Since $\max_p L(p; \lambda, v) \geq L(\bar{p}; \lambda, v) > v_2(\bar{\ell}) \geq v_1(\ell_{\max}) = \max_p L(p; \lambda^*, \mu^*)$, we know that $(\lambda, \mu) \neq (\lambda^*, \mu^*)$. Therefore, $\max\{\max_{h,\, h' \in H} \lambda^*_{h,h'},\, \mu^*\} < M$.

Step 2. We show that, for any $\epsilon > 0$, there exists $\ell(\epsilon) > \bar{\ell}$, such that when $\ell_{\max} > \ell(\epsilon)$, the Lagrangian multipliers of the planner problem (4.1) subject to (4.2) and (4.3) satisfies $\lambda^*_{h,h'} < \epsilon$, $\forall h, h' \in H$, s.t. $\eta_h < \eta_{h'}$.

Suppose $H = \cup_{i=1}^N \{h = (i, h_{-\eta}) | h_{-\eta} \in H_i\}$, and for any $h = (i, h_{-\eta}) \in H$, $\eta_h = \eta_i$. Suppose that $0 < \eta_N < \cdots < \eta_1$. We first define the set

$$\mathcal{G} = \{((h, h') \in H^2,\ \text{s.t. } \eta_h < \eta_{h'}\} = \{(h = (i, h_{-\eta}), h' = (j, h'_{-\eta})) \in H^2,\ \text{s.t. } i > j\}.$$

Assume that for a planner problem (4.1) subject to (4.2) and (4.3) with $\ell_{\max} > 0$, there exists some Lagrangian multipliers $\lambda^*_{h,h'} \geq \epsilon$, $(h = (i, h_{-\eta}), h' = (j, h'_{-\eta})) \in \mathcal{G}$. We consider the Lagrangian function of the deterministic problem for (4.1) subject to (4.2) and (4.3),

$$\begin{aligned}
\mathcal{L}(c, y; \lambda^*, \mu^*) &= \Bigg[-\left(\omega_h + \sum_{h' \in H} \tilde{\psi}_h \lambda^*_{h,h'} - \sum_{k=i} \sum_{h'_{-\eta} \in H_k} \tilde{\psi}_{h'} \lambda^*_{h',h} \right) y_h^{\frac{1}{\eta_i}+1} + \\
&\qquad \sum_{k \neq i} \left(\sum_{h'_{-\eta} \in H_k} \tilde{\psi}_{h'} \lambda^*_{h',h} \right) y_h^{\frac{1}{\eta_k}+1} + \mu^* \omega_h y_h \Bigg] + \mathcal{L}^{c, y_{-h}}(c, y_{-h}; \lambda^*, \mu^*) \\
&:= \left[\sum_{k=1}^N \mathcal{B}_{h,k}(\lambda^*) y_h^{\frac{1}{\eta_k}+1} + \mu^* \omega_h y_h \right] + \mathcal{L}^{c, y_{-h}}(c, y_{-h}; \lambda^*, \mu^*),
\end{aligned}$$

where $h' = (k, h'_{-\eta})$, $\mathcal{L}^{c,y_{-h}}(c, y_{-h}; \lambda^*, \mu^*)$ is independent on y_h, $\mathcal{B}_{h,j}(\lambda^*) \geq \epsilon \tilde{\psi}_{h'}$, and $\mathcal{B}_{h,k}(\lambda^*) \geq 0$, $\forall k \neq i$. According to Step 1, the Lagrangian multipliers λ^*, μ^* are bounded, hence there exists $M > 0$ that is independent on λ^*, μ^*, s.t. $|\mathcal{B}_{h,k}(\lambda^*)| \leq M$, $\omega_h \mu^* \leq M$ and $\mathcal{L}^{c,y_{-h}}(c, y_{-h}; \lambda^*, \mu^*)|_{c \equiv c_{\max}, y_{-h} \equiv 0} \geq -M$. Therefore

$$\begin{aligned}
&\max_{c \in C, y \in Y} \mathcal{L}(c, y; \lambda^*, \mu^*) \\
\geq &\max_{y_h \in [0, w_h \ell_{\max}]} \left[\sum_{k=1}^N \mathcal{B}_{h,k}(\lambda^*) y_h^{\frac{1}{\eta_k}+1} + \mu^* \omega_h y_h \right] - M \\
\geq &\max_{y_h \in [0, w_h \ell_{\max}]} \left[\epsilon \tilde{\psi}_{h'} y_h^{\frac{1}{\eta_j}+1} - M \sum_{k=1}^i y_h^{\frac{1}{\eta_k}+1} \right] - M \to \infty \text{ as } y_h \to \infty.
\end{aligned} \tag{A.48}$$

On the other hand,

$$\max_{c\in C, y\in Y} \mathcal{L}(c,y;\lambda^*,\mu^*) = v_1(\ell_{\max}) \le v_2(\bar{\ell}). \tag{A.49}$$

Hence there exists some $\bar{\ell}(M,\epsilon,h,h') > 0$, s.t. $\ell_{\max} < \ell_{\epsilon,h,h'}$. Since the set $\mathcal{G}$ is finite, we can define

$$\bar{\ell}(\epsilon) = \max_{(h,h')\in\mathcal{G}} \bar{\ell}(M,\epsilon,h,h').$$

Therefore, when $\ell_{\max} > \bar{\ell}(\epsilon)$, the Lagrangian multipliers should satisfy $\lambda^*_{h,h'} < \epsilon$, $\forall (h,h') \in \mathcal{G}$.
Step 3. We denote $(\bar{c},\bar{y})$ the solution to the planner problem (4.1) subject to (4.2) and (4.5) with $\ell_{\max} \ge \bar{\ell}$. It is straightforward to check that, for any $\epsilon > 0$, there exists $\delta > 0$, such that when $\max_{(h,h')\in\mathcal{G}} \lambda^*_{h,h'} \le \delta$,

$$\mathcal{L}(\bar{c},\bar{y};\lambda^*,\gamma^*) \ge v_2(\bar{\ell}) - \epsilon.$$

According to Step 2, there exists $\bar{\ell}(\delta) > 0$, such that when $\ell_{\max} \ge \bar{\ell}(\delta)$, the Lagrangian multipliers of the planner problem (4.1) subject to (4.2) and (4.3) satisfy $\max_{(h,h')\in\mathcal{G}} \lambda^*_{h,h'} \le \delta$. Hence

$$v_1(\ell_{\max}) = \max_{c\in C, y\in Y} \mathcal{L}(c,y;\lambda^*,\gamma^*) \ge \mathcal{L}(\bar{c},\bar{y};\lambda^*,\gamma^*) \ge v_2(\bar{\ell}) - \epsilon,$$

implying that

$$(v_2(\bar{\ell}) - v_1(\ell_{\max}))_+ \le \epsilon.$$

Since ϵ is arbitrary, we can conclude that $\lim_{\ell_{\max}\to\infty}(v_2(\bar{\ell}) - v_1(\ell_{\max}))_+ = 0$ and finish the proof.

Proof (*Proof of Lemma* (A.1)) According to Lemma 4.2, for any $\ell_{\max} > 0$, the solution to the planner problem (4.1) subject to (4.2) and (4.3) does no involve lottery in labor supply. We arbitrarily take a $\underline{\ell} > 0$. Since

$$u_h(c_h,y_h) \le u_h(c_{\max},0), \quad \forall h \in H, c_h \in [\epsilon, c_{\max}],\, y_h \in [0, w_h\ell_{\max}],$$

and

$$\lim_{y_h\to\infty} \omega_h u_h(c_h,y_h) \to -\infty, \quad \forall h \in H, c_h \in [\epsilon, c_{\max}],\, y_h \in [0, w_h\ell_{\max}],$$

there exists $\bar{\ell}(h) > 0$, s.t. when $y_h > \bar{\ell}(h)w_h$,

$$\omega_h u_h(c_h,y_h) + \sum_{h'\ne h} \omega_{h'} u_{h'}(c_{h'},y_{h'}) < \left(v_2(\underline{l}) - \sum_{h'\ne h} \omega_{h'} u_{h'}(c_{\max},0)\right) + \sum_{h'\ne h} \omega_{h'} u_{h'}(c_{\max},0) = v_2(\underline{l}).$$

Therefore, for any $\ell_{\max} > \underline{\ell}$, the solution $(c(\ell_{\max}), y(\ell_{\max}))$ to the planner problem (4.1) subject to (4.2) and (4.5) satisfies $y_h(\ell_{\max}) \le \bar{\ell}(h)w_h$. We take $\bar{\ell} = \max_{h\in H}\bar{\ell}(h)$. Then for any $\ell_{\max} > \bar{\ell}$, the solution to the planner problem (4.1) subject to (4.2) and (4.3), remains invariant to the solution to the problem with $\ell_{\max} = \bar{\ell}$. Hence we finish the proof. $\square$

A.10 Proof of Proposition 4.2

Suppose $H = \cup_{i=1}^{N}\{h = (i, h_{-\eta}) | h_{-\eta} \in H_i\}$, and for any $h = (i, h_{-\eta}) \in H$, $\eta_h = \eta_i$. Suppose that $0 < \eta_N < \cdots < \eta_1$. Similarly to the proof for Lemma 4.1, we define

$$\tilde{\psi}_h = \frac{\psi_h}{w_h^{\frac{1}{\eta_h}+1}\left(\frac{1}{\eta_h}+1\right)}, \quad \forall h \in H.$$

Then the deterministic problem for (4.1) subject to (4.2) and (4.5) can be written

$$\begin{aligned}
&\max_{c\in C,\, y\in Y} \sum_{h\in H} \omega_h \left(\log(c_h) - \tilde{\psi}_h \frac{y_h^{\frac{1}{\eta_h}+1}}{\frac{1}{\eta_h}+1}\right),\\
&\text{s.t.} \log(c_h) - \tilde{\psi}_h \frac{(y_h)^{\frac{1}{\eta_h}+1}}{\frac{1}{\eta_h}+1} \ge \log(c_{h'}) - \tilde{\psi}_h \frac{(y_{h'})^{\frac{1}{\eta_h}+1}}{\frac{1}{\eta_h}+1}, \quad \forall h = (i, h_{-\eta}), h' = (j, h'_{-\eta}) \in H, \text{ s.t. } i \le j,\\
&\sum_{h\in H}\omega_h y_h \ge \sum_{h\in H}\omega_h c_h;
\end{aligned} \tag{A.50}$$

and the Lagrangian function of problem (A.50) has the structure

$$\begin{aligned}
\mathcal{L}(c, y; \lambda, \mu) = &\sum_{(i,h_{-\eta})\in H}\left[-\left(\omega_h + \sum_{j\ge i}\sum_{h'_{-\eta}\in H_j}\tilde{\psi}_h\lambda_{h,h'} - \sum_{j=i}\sum_{h'_{-\eta}\in H_j}\tilde{\psi}_{h'}\lambda_{h',h}\right) y_h^{\frac{1}{\eta_i}+1} + \right.\\
&\left.\sum_{j=1}^{i-1}\left(\sum_{h'_{-\eta}\in H_j}\tilde{\psi}_{h'}\lambda_{h',h}\right) y_h^{\frac{1}{\eta_j}+1} + \mu\omega_h y_h\right] + \mathcal{L}^c(c; \lambda, \mu)\\
:= &\sum_{h\in H}\left[-\mathcal{B}_{h,i}(\lambda) y_h^{\frac{1}{\eta_i}+1} + \sum_{j=1}^{i-1}\mathcal{B}_{h,j}(\lambda) y_h^{\frac{1}{\eta_j}+1} + \mu\omega_h y_h\right] + \mathcal{L}^c(c; \lambda, \mu),
\end{aligned} \tag{A.51}$$

where $h' = (j, h'_{-\eta})$, $\mathcal{B}_{h,j}(\lambda) \ge 0$, $\forall 1 \le j \le i-1$, and $\mathcal{L}^c(c; \lambda, \mu)$ is independent on $y \in Y$. Similar to the proof for Lemma 4.1, the minimizer of

$$V(\lambda, \mu) := \max_{c\in C, y\in C}\mathcal{L}(c, y; \lambda, \mu)$$

satisfies $\mu^* > 0$. We introduce a technical Lemma here to progress the proof:

Lemma A.2 *Given $i \in \{1, \cdots, N\}$, and $1 < p_1 < \cdots < p_N$. We consider the function*

$$f_i(y) = -a_i y^{p_i} + \sum_{j=1}^{i-1} a_j y^{p_j} + a_0 y,$$

where $a_j \geq 0$, $\forall 1 \leq j \leq i-1$. and $a_0 > 0$. Then f_i admits a unique maximal point in the interval $[0, w_i \ell_{\max}]$.

For any $h \in H$, according to Lemma A.2 with $p_i = \frac{1}{\eta_i}+1$, $p_j = \frac{1}{\eta_j}+1 (\forall 1 \leq j \leq i-1)$, $a_i = \mathcal{B}_{h,i}(\lambda^*)$, $a_j = \mathcal{B}_{h,j}(\lambda^*)(\forall 1 \leq j \leq i-1)$, and $a_0 = \mu^* \omega_h$, we know that the maximal c_h^* is unique. Hence according to Result 1 of Corollary 2.1, the optimal solution to the planner problem (4.1) subject to (4.2) and (4.5) does not involve lotteries in labor.

Proof (*Proof of Lemma* (A.2)) If $a_i \leq 0$, then $f_i(y)$ is strictly increasing in the interval $[0, w_h \ell_{\max}]$ and has a unique maximal point $y^* = w_h \ell_{\max}$.

If $a_i > 0$, it is straightforward to have

$$f_i'(y) = -a_i p_i y^{p_i - 1} + \sum_{j=1}^{i-1} a_j p_j y^{p_j - 1} + a_0,$$

and

$$\begin{aligned}
f_i''(y) =& - a_i p_i (p_i - 1) y^{p_i - 2} + \sum_{j=1}^{i-1} a_j p_j (p_j - 1) y^{p_j - 2} \\
=& \frac{p_i - 1}{y} (f_i'(y) - \sum_{j=1}^{i-1} a_j p_j y^{p_j - 1} - a_0) + \sum_{j=1}^{i-1} a_j p_j (p_j - 1) y^{p_j - 2} \\
=& \frac{(p_i - 1) f_i'(y)}{y} - \frac{(p_i - 1) a_0}{y} - \sum_{j=1}^{i-1} a_j p_j (p_i - p_j) y^{p_j - 2} \\
\leq& \frac{(p_i - 1) f_i'(y)}{y} - \frac{(p_i - 1) a_0}{y}.
\end{aligned}$$

Therefore,

$$[\frac{f_i'(y)}{y^{p_i - 1}}]' = \frac{f_i''(y)}{y^{p_i - 1}} - \frac{(p_i - 1) f_i'(y)}{y^{p_i}} \leq -\frac{(p_i - 1) a_0}{y^{p_i}} < 0, \forall y > 0.$$

implying that $\frac{f_i'(y)}{y^{p_i - 1}}$ is strictly decreasing when $y > 0$. We take $y = 0$ in $f_i'(y)$ and obtain $f_i'(0) = a_0 > 0$. Therefore, by the continuity of $f_i'(y)$, there exists $\delta > 0$, such that when $y \in [0, \delta]$, $f'(y) > 0$. If for any $y \in [\delta, w_h \ell_{\max}]$, $\frac{f_i'(y)}{y^{p_i - 1}} > 0$, then $f_i'(y) > 0$, $f_i(y)$ is strictly increasing and has a unique maximal point $y^* = w_h \ell_{\max}$; if there exists $y_0 \in [\delta, w_h \ell_{\max}]$, s.t. $\frac{f_i'(y_0)}{y_0^{p_i - 1}} = 0$, then $f_i'(y_0) = 0$, $f_i'(y_0) > 0$ for $y < y_0$, and $f_i'(y_0) < 0$ for $y > y_0$, hence $f_i(y)$ has a unique maximal point $y^* = y_0$. Therefore we finish the proof. $\square$