A Lagrangian Approach to Optimal Randomization*

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

First version: April 22, 2025
This version: March 26, 2026

Abstract

We develop a new method to efficiently solve a class of non-convex constrained optimization problems that frequently arise in economics. The optimal solution to these problems often involves randomization. To employ a Lagrangian framework, we prove that the value of the saddle point that characterizes the optimal random solution 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 random solution can be directly obtained from the deterministic optima that are computed along the iterations. We argue that for many non-convex economic problems, the performance of our method is orders of magnitude faster than existing methods, in particular than linear programming. To illustrate our method, we solve for optimal income taxes in the Mirrleesian model with multi-dimensional types. We show that heterogeneity in productivity and Frisch elasticity can make randomization welfare improving relative to the optimal deterministic tax schedule.

Keywords: Non-Convexities, Randomization, Lagrangian Iteration, Mirrleesian Optimal Taxation, Optimal Policy, Moral Hazard.

JEL classification: C61, C63, D61, D82.

*An earlier version of this work, which this paper supersedes, circulated under the title "A Lagrangian Approach to Optimal Lotteries in Non-Convex Economies". We thank Matias Bayas-Erazo, 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 (#10003091) 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. Email: fkubler@gmail.com
§University of Zurich and SFI. Email: yucheng.yang@uzh.ch
¶Westlake University. Email: zhouzhennan@westlake.edu.cn

1 Introduction

Many interesting economic questions can be formulated as non-convex optimization problems – prominent examples include contracting problems under moral hazard (Mirrlees (1999)), problems of optimal taxation (Mirrlees (1971), Cremer et al. (2001)), models of optimal fiscal policy (Aiyagari et al. (2002)), or models of macro-prudential regulation (Dávila and Korinek (2018)). In these environments, standard first-order conditions generally fail to characterize globally optimal solutions. Moreover, in the presence of non-convexities, random policies often outperform deterministic solutions (see, e.g. Myerson (1982); Prescott and Townsend (1984b); Stiglitz (1982); Citanna and Tvede (2024)).

In this paper, we develop a new, efficient, and general method to solve for optimal lottery solutions[1] in a wide class of models with non-convexities. 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 Lagrangian 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 no lottery solution yields a higher value than the deterministic solution. Using the Lagrangian dual to solve constrained optimization problems is a very common approach in applied mathematics (see, e.g., Boyd and Vandenberghe (2004)). By reformulating the primal optimization problem into the Lagrangian dual problem, a difficult non-convex problem can oftentimes be turned into a simpler problem that can be efficiently solved (see, e.g. Lemaréchal (2001)). For economic applications, the key problem is to formalize under which conditions the solution to the dual problem can be interpreted as a lottery solution to the original problem. We use sub-gradient descent to solve the dual problem iteratively. Our third theoretical insight is that the frequency of different optimal points along the iterations approximates the probability distribution of the optimal lottery solution. As in fictitious play for normal form games (see, e.g., Fudenberg and Levine (1998)), the (weighted) empirical distribution of past actions converges to the optimal distribution in the lottery solution.

In the economic literature, Prescott and Townsend (1984b) already emphasizes that the lottery approach transforms the original problem into a convex problem by replacing deterministic assignments with probability distributions. This reformulation linearizes the constraints in the probability space and, after discretizing all possible actions, linear programming can be employed to find optimal solutions. However, this convexification comes

[1]Throughout the paper, we use the terms "lottery solution" and "randomized solution" interchangeably to refer to the optimal, possibly random, solution of a non-linear programming problem.

at the cost of dramatically increasing the dimensionality of the problem. As a result, linear programming methods, while theoretically applicable, face a severe intractability in the dimensionality of the choice variables. Although it is possible to construct examples where our Lagrangian method faces the same curse of dimensionality, we argue that economic problems typically impose additional structure on the mathematical problem that allows us to partially overcome the curse.

To illustrate our method and its advantages over existing methods, we first consider a simple textbook principal agent problem (see, e.g. Salanié (2005, Chapter 5)). We use this simple example to illustrate the advantages of our methods over linear programming and to give an intuitive explanation for our method.

Our main 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. Since the Lagrangian decomposes across types, the complexity of our algorithm increases linearly in the number of types. This is in stark contrast to global constrained optimization methods for non-convex problems that are generally known to be NP-hard (e.g. Sahni (1974)). The lottery relaxation together with the *decomposability* property make the problem feasible. Using an example from Judd et al. (2017), we illustrate the efficiency of our method and clarify the roles that lotteries play in the optimal solution. We also show that the welfare gains from lotteries depend crucially on details of the calibration of preferences and in some cases can be substantial. We then consider a more complicated example where the production function is strictly concave over different types of labor and where there are skill complementarities. In this example, the problem only becomes decomposable after some variables are eliminated using first order conditions. This is impossible when using linear programming methods, where the number of variables increases exponentially with the number of sectors.

The remainder of the paper is organized as follows. After reviewing the existing literature in Section 2, we describe the setup, introduce the algorithm, and state our theorems in Section 3. In Section 4 we discuss the computational performance of our algorithm and compare it with existing methods. In Section 5, we illustrate the algorithm using a textbook principal-agent problem. In Sections 6, we apply the method to a standard optimal tax problem. Section 7 concludes. The proofs are collected in the Appendix A. In Online Appendix B, we provide worst-case complexity bounds for our method and compare them to bounds from linear programming. In Online Appendix C, we demonstrate how to handle a version of the tax problem with complementarities in production.

2 Related Literature

The seminal work by Myerson (1982) established that allowing lotteries can improve both 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) provides simple sufficient conditions for lotteries to be part of an optimal contract. Since their analysis is local, these conditions are not necessary.

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); Chang and Wildasin (1986); 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.[2] Most of the existing approaches for solving lottery problems in models with private information focus on improving linear programming techniques. For example, Prescott (2004) exploits the block-triangular structure of the constraint matrix and applies Dantzig-Wolfe decomposition to reduce memory and computational cost, while Doepke and Townsend (2006) proposes 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 Lagrangian iteration method offers significant computational advantages compared to traditional linear programming approaches.

Our application contributes to the literature on optimal tax design under private infor-

[2]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.

mation. 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.

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 paper, we demonstrate that the lottery relaxation provides a means of achieving this convexification and therefore provide economic foundations for convexification.

Lastly, our method builds on the subgradient descent literature for nondifferentiable convex optimization (Bertsekas, 2009; 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 constrained planning problems.

3 Lagrangian Iteration

This section introduces a general class of non-convex planning problems and presents an algorithm for computing the optimal lottery. We begin by stating the environment and constraints in Section 3.1. We then establish a key connection between the lottery problem and the deterministic Lagrangian in Section 3.2, which motivates our approach and provides the basis for the algorithm. Next, Section 3.3 presents the algorithm and Section 3.4 proves its convergence properties. Readers primarily interested in the method and its implementation may focus on Sections 3.1 and 3.3.

3.1 General Setup

In this section, we present a general constrained maximization problem that subsumes many planning problems considered in economics. Let A denote a finite set of actions, and let C be a n-dimensional box 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:

$$\begin{aligned} &\max_{a \in A,\, c \in C} f(a, c), \\ \textbf{s.t. } &g_i(a, c) \leq 0 \quad i \in \{1, \ldots, m\}, \\ &h_j(a, c) \leq 0 \quad j \in \{1, \ldots, \ell\}, \end{aligned} \tag{3.1}$$

where the functions $g_i(.), i = 1, \ldots, m$, and $h_j(.), j = 1, \ldots, \ell$ are assumed to be continuous in c.[3] and in the rest of the paper we refer to (3.1) 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 makes it computationally challenging to solve the problem. In examples where the set C can also be taken to be approximated by a finite set, an optimal solution can obviously be found by searching over all elements in $A \times C$ but the computational costs of this method increase exponentially in the dimension of $A \times C$.

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 (3.1), the social planner chooses a probability distribution over the set of possible outcomes, $A \times C$. The economic model then dictates which constraints can be relaxed to having to hold only in expectations and which constraints have to hold for each realization of the lottery. In our abstract formulation, assume that the $g-$constraints can be relaxed while the $h-$constraints have to hold for each realization of $a \in A$.

Formally, 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 an 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 objective of the social planner in the lottery problem is to maximize the expected 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 assumed to hold in expec-

[3]The reason why we distinguish between g-constraints and h-constraints and why A is assumed to be finite will become clear in the following when we describe the lottery problem,

tations, i.e.

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

In economic models, one 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.

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

Note that allowing for continuous 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.

$$\begin{aligned} &\max_{x\in\mathcal{P}(A\times C)} \sum_{a\in A}\int_{c\in C} f(a,c)x(a,dc), \\ \text{s.t. } &\sum_{a\in A}\int_{c\in C} g_i(a,c)x(a,dc) \leq 0 \quad \forall i \in \{1,\ldots,m\}, \\ &\int_{c\in C} h_j(a,c)x(a,dc) \leq 0 \quad \forall a \in A,\ j \in \{1,\ldots,\ell\}. \end{aligned} \tag{3.2}$$

Note that in this formulation, we do not require that the marginal distributions of $x \in \mathcal{P}(A\times C)$ across the components A and C to be 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 solution, but this is a result and not an assumption.

3.2 A Lagrange Approach

The essence of our method is to establish a connection between the deterministic system (3.1) and the lottery system (3.2). Given Lagrange parameters λ, γ, we define the Lagrangian

function in the probability space $\mathcal{P}(A \times C)$ as

$$L(x; \lambda, \gamma) :=$$
$$\sum_{a \in A} \int_{c \in C} f(a,c) x(a, dc) - \sum_{i=1}^{m} \lambda_i \sum_{a \in A} \int_{c \in C} g_i(a,c) x(a, dc) - \sum_{j=1}^{\ell} \sum_{a \in A} \gamma_{j,a} \int_{c \in C} h_j(a,c) x(a, dc), \tag{3.3}$$

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

$$\mathcal{L}(a, c; \lambda, \gamma) := f(a,c) - \sum_{i=1}^{m} \lambda_i g_i(a,c) - \sum_{j=1}^{\ell} \gamma_{j,a} h_j(a,c). \tag{3.4}$$

The main result of this subsection is that under the assumption that Slater's conditions holds, we have the following equivalence

$$\max_{x \in \mathcal{P}(A \times C)} \min_{(\lambda, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^{\ell |A|}_+} L(x; \lambda, \gamma) = \min_{(\lambda, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^{\ell |A|}_+} \max_{a \in A,\, c \in C} \mathcal{L}(a, c; \lambda, \gamma), \tag{3.5}$$

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

3.2.1 Prerequisites for a Lagrange approach

One well-known problem in solving the deterministic problem (3.1) lies in the fact that standard constraint qualifications, which are necessary for applying optimization methods from numerical analysis, often fail to hold at the solution (see, e.g., Su and Judd (2007) or Judd et al. (2017)). We therefore first demonstrate the validity of the Lagrange Multiplier Method for the lottery system (3.2). 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).[4] The definition of the condition is as follows.

Definition 3.1 *(Slater's condition) We say that Slater's condition holds for the lottery system (3.2), if the feasible set to (3.2) includes one inner point, i.e., there exists $x \in \mathcal{P}(A \times C)$ such that*

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

[4]Because the decision variable is a probability measure, (3.2) is an infinite-dimensional linear program. In finite dimensions, strong duality for linear programs can be established without Slater-type conditions. In infinite dimensions, a constraint qualification is typically invoked to ensure existence of multipliers.

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.

It is standard to show (see, e.g. Theorem 1 in section 8.3 in Luenberger (1997)[5]) that Slater's condition implies that the solution of problem (3.2) $x^* \in \mathcal{P}(A\times C)$ exists, and that there exist Lagrangian multipliers $\lambda_i^*, \gamma_{j,a}^* \geq 0$ ($i \in \{1,\cdots,m\}$, $j \in \{1,\cdots,l\}$, $a \in A$) such that x^* is the maximizer of $L(x;\lambda^*,\gamma^*)$. Generally, it is difficult to verify whether Slater's condition holds in (3.2). However, if we relax the constraints in (3.2) to

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x(a,dc) \leq \epsilon, \quad \forall i \in \{1,\cdots,m\}, \tag{3.6}$$

and

$$\int_{c\in C} h_j(a,c)x(a,dc) \leq \epsilon, \quad \forall j \in \{1,\cdots,\ell\},\ a \in A, \tag{3.7}$$

for some $\epsilon > 0$, then any feasible measure $x \in \mathcal{P}(A\times C)$ for the original lottery system (3.2) will strictly satisfy these two relaxed constraints. We therefore consider the following relaxed system.

$$\max_{x\in\mathcal{P}(A\times C)} \sum_{a\in A}\int_{c\in C} f(a,c)x(a,dc) \text{ s.t. (3.6) and (3.7)} \tag{3.8}$$

The Slater's condition for (3.8) is then satisfied when there is at least one feasible point for the original lottery system (3.2). We denote x^ϵ as the optimal solution to (3.8). The following theorem shows that x^ϵ will converge to x^*, which is the optimal solution to the original lottery system (3.2).

Theorem 3.1 *Assume that there exists at least one feasible point for the original lottery problem (3.2), hence for any $\epsilon > 0$, the solution to the relaxed lottery problem (3.8) 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*

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

[5]We need to verify the existence of the maximizer of the Lagrangian in order to apply 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.

Furthermore, x^* is an optimal solution to (3.2).

Theorem 3.1 ensures that the Lagrange Multiplier Method is a valid numerical approach to solve the system (3.2). In the remainder of this paper, we will always assume that the Lagrangian multipliers to system (3.2) exist. It is straightforward to modify our methods to compute a solution to the ϵ-relaxed problem for cases where this assumption fails.

3.2.2 The Relation between the Lottery Problem and the Deterministic Problem

In order to understand why (3.5) holds, we first establish that for given λ, γ the optimal solutions to the maximization of $L(.; \lambda, \gamma)$ and $\mathcal{L}(.; \lambda, \gamma)$ have the same value in the following theorem.

Theorem 3.2 *Given $\lambda \geq 0$, $\gamma \geq 0$. Let L and $\mathcal{L}$ be defined as in (3.3) and (3.4) respectively. Then we have*

$$\max_{x \in \mathcal{P}(A \times C)} L(x; \lambda, \gamma) = \max_{a \in A, c \in C} \mathcal{L}(a, c; \lambda, \gamma). \tag{3.9}$$

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.

$$x^*(Z^c) = 0. \tag{3.10}$$

As in mixed strategy Nash equilibrium, an optimal solution only put weights on different actions that yield the same value of the objective function. But this implies directly that

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

But since the lottery problem is convex and strong duality holds, this implies that the lottery problem and the dual problem of the deterministic problem are essentially the same problem. To formalize this, it is useful to explicitly define what we mean by the solution to the dual problem.

Definition 3.2 *We define the dual problem of the deterministic problem (3.1) as*

$$\inf_{(\lambda, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^{\ell |A|}_+} \max_{a \in A,\, c \in C} \mathcal{L}(a, c; \lambda, \gamma), \tag{3.11}$$

where $\mathcal{L}$ is defined in (3.4). If the infimum is attained at (λ^*, γ^*), then we call (λ^*, γ^*) a solution to the dual problem (3.11).

The following theorem shows that these two problems share the same Lagrangian multipliers and the same optimal value.

Theorem 3.3 *We assume that x^* is the solution to system (3.2), 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.*

$$\max_{a\in A,\, c\in C} \mathcal{L}(a, c; \lambda^*, \gamma^*) = \inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+} \max_{a\in A,\, c\in C} \mathcal{L}(a, c; \lambda, \gamma), \tag{3.12}$$

where $\mathcal{L}$ is defined in (3.4). Furthermore, the optimal objective value of the dual problem (3.11) is the same as the optimal objective value of the lottery problem (3.2), i,e.

$$\inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+} \max_{a\in A,\, c\in C} \mathcal{L}(a, c; \lambda, \gamma) = \sum_{a\in A} \int_{c\in C} f(a, c) x^*(a, dc). \tag{3.13}$$

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

Corollary 3.1 *The Lagrangian function in $\mathcal{P}(A \times C)$, L, admits a nondegenerate lottery maximizer and no deterministic 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^*, and none of these points are feasible.*

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

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

3.3 The Lagrangian iteration algorithm

We utilize the results above to construct an algorithm for solving the lottery problem. Theorem 3.2 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 $A \times C$ and Theorem 3.3 shows how the optimal λ^*, γ^* can be obtained from the deterministic problem: We define

$$V(\lambda, \gamma) := \max_{a\in A,\, c\in C} \mathcal{L}(a, c; \lambda, \gamma), \tag{3.14}$$

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

Definition 3.3 (Sub-gradient) *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 3.1 *The dual function V defined in (3.14) 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).

$$(\Delta\lambda, \, \Delta\gamma) := \begin{cases} \Delta\lambda_i = g_i(a_\lambda, c_\lambda), & i \in \{1, \cdots, m\}; \\ \Delta\gamma_{j,a} = h_j(a_\lambda, c_\lambda), & j \in \{1, \cdots, \ell\}, \, a = a_\lambda; \\ \Delta\gamma_{j,a} = 0, & j \in \{1, \cdots, \ell\}, \, a \neq a_\lambda. \end{cases} \tag{3.15}$$

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:

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

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

and

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

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

$$x^N := \frac{1}{\sum_{k=1}^N \mu^k} \sum_{k=1}^N \mu^k \delta_{(a^k,c^k)}. \tag{3.16}$$

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 A.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.

It is important to note that there are $\ell|A|$ Lagrange multipliers associated with the constraints related to unobserved actions, as the system (3.2) 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 $\bar{a} \in A$. This is generally a nonconvex problem, and in the worst-case scenario, one needs to discretize the set C and search over all $A \times C$. In this case, the algorithm's computational efficiency is comparable to linear programming (see Appendix B). However, a central claim of this paper is that in typical economic applications this maximization problem can be solved relatively efficiently.

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.

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

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

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

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).

3.4 Theoretical Properties of the Algorithm

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 formally in this section.

For any $\epsilon > 0$, we first define the ϵ-optimal solution to the system (3.2), 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 3.4 *(ϵ-optimal solution) We denote by x^* the optimal solution to system (3.2). We call $\tilde{x}^\epsilon \in \mathcal{P}(A \times C)$ an ϵ-optimal solution to system (3.2), if it satisfies the relaxed*

constraints (3.6) and (3.7) and if

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

Note that $\tilde{x}^{\epsilon}$ is different from the solution to the relaxed system (3.8), since we do not require $\tilde{x}^{\epsilon}$ to exactly attain the maximal objective function value among all feasible probability measures for (3.8). However, property 3 in this definition actually implies that $\tilde{x}^{\epsilon}$ can be regarded as an approximate optimal solution to (3.8) and we have the following analog to Theorem 3.1.

Theorem 3.4 *We denote $\tilde{x}^{\epsilon}$ as an ϵ-optimal solution to the system (3.2). 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*

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

Furthermore, $\tilde{x}^{*}$ is an optimal solution to (3.2).

We now examine the property of the lottery solution x^{N} generated by Algorithm 1 and have the following result.

Theorem 3.5 *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 (3.2), 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 (3.2).

Theorem 3.5 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.

The final theoretical result derives a bound on the number of iterations needed to obtain a given accuracy of the solution. We let $\lambda = (\lambda_i)_{i\in\{1,\cdots,m\}}$ 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 3.5 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. 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 to obtain an ϵ-solution.

Proposition 3.1 *We take all the assumptions in Theorem 3.5. 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*

$$n \gtrsim \begin{cases} \left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}} (m+\ell)^{\frac{1}{1-\rho}}, & \rho < 1 \\ e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{m+\ell}}{\epsilon}}, & \rho = 1, \end{cases} \tag{3.17}$$

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

From Equation (3.17) it appears as if the learning rate parameter ρ should be set relatively close to zero. However, 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, then for any $\epsilon > 0$, we can select a learning rate satisfying $\mu^k \sim O(1/k)$ (i.e. with ρ close to 1) such that when

$$n \gtrsim \frac{M^2(m+\ell)}{\epsilon^2}, \tag{3.18}$$

it follows that $\text{dist}((\lambda^n, \gamma^n), \Lambda^*) < \epsilon$.

In general, it is important to point out that sub-gradient descent is generally much less efficient than gradient descent applied to convex and smooth problems. In the latter case, the algorithm can be refined by using line-search or other methods and, in general, the number of iterations needed to achieve precision ϵ is of the order $\frac{1}{\epsilon}$ or even $\frac{1}{\sqrt{\epsilon}}$ (see Shor (2012)). If the optimal solution is deterministic, our problem becomes smooth (at least in a neighborhood of the optimal solution) and there are more efficient ways to find the optimal solution. In this paper, we focus mainly on the more difficult problem, where the optimal solution involves lotteries.

4 Computational Performance

4.1 Computational Performance without Additional Structure

It is well established since at least Sahni (1974) that without any further assumptions, the maximization problem (3.1) to solve for a deterministic solution is not computationally tractable. Linear programming methods are typically used to solve for a lottery solution to (3.1), but little is gained in terms of computational efficiency without any additional structure. To be precise, to use linear programming, we first need to discretize the set C into a finite grid $\hat{C}$. This transforms (3.2) into a finite-dimensional problem:

$$\begin{aligned} &\max_{\hat{x}\in\mathcal{P}(A\times\hat{C})} \sum_{a\in A}\sum_{c\in\hat{C}} f(a,c)\hat{x}(a,c), \\ \textbf{s.t.}\ &\sum_{a\in A}\sum_{c\in\hat{C}} g_i(a,c)\hat{x}(a,c) \leq 0 \ (i\in\{1,\cdots,m\}), \\ &\sum_{c\in\hat{C}} h_j(a,c)\hat{x}(a,c) \leq 0 \ (j\in\{1,\cdots,\ell\},\ a\in A). \end{aligned} \tag{4.1}$$

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 $\ell|A|+m$. The deterministic problem can be approximately solved by exhaustive search over $A\times\hat{C}$, while using linear programming to obtain the lottery solution means to solve a linear program with $|A||\hat{C}|$ variables – this is generally obviously more computationally intensive than grid-search. In applications where a very fine discretization of C is needed to obtain an economically meaningful solution, both methods are not tractable. Furthermore, if C is a high dimensional set, and one wants to maintain the same number of grid points along each dimension of C, both methods' complexity is obviously exponential in the dimension of C. This is true even if the complexity of linear programming is polynomial in the number of variables.

For the abstract mathematical problem, our Lagrangian algorithm performs better in terms of the worst case complexity compared to linear programming. The number of iterations needed to obtain a given accuracy ϵ is bounded in Proposition 3.1 and does not depend on the number of variables, but increases polynomially in the number of constraints. The overall complexity of our algorithm can then be obtained by combining the required number of iterations (Proposition 3.1), and a straightforward estimate of the complexity of the grid search in Step 1 – see Theorem B.1 and Section B.2 (for comparison to linear programming) in Online Appendix B for details. However, without further structure, our algorithm does not perform significantly better in practical computations. For finite $\hat{C}$, the grid search in Step 1 of Algorithm 1 suffers from the same curse of dimensionality as linear programming

and exhaustive search for a deterministic solution.

One of the key insights of this paper is that typical economic problems contain additional structure which stems from the standard assumption that the utility- and production-functions are concave and that helps overcome the curse of dimensionality for the Step 1 of our algorithm.

4.2 Computational Performance with Additional Structure

Two key structures that arise in many economic problems and can significantly improve the efficiency of our algorithm are *decomposability* and *partial concavity*. Both will be formally defined later in this section, but we provide some intuition here.

Decomposability allows a high-dimensional consumption set C to be decomposed into lower-dimensional components $C = \times_{k=1}^{d} C_k$, with all related functions that are additively separable across these components. This structure enables us to drastically reduce the scale of grid search in Step 1 of our algorithm. While linear programming can also exploit decomposability to reduce dimensionality, the mechanism is less direct—an illustration is provided in the moral hazard example of Section 5. Importantly, decomposability does not improve efficiency if one attempts to apply grid-search to (3.1) to solve for a deterministic solution.

Partially concavity refers to the property that the Lagrangian of problem (3.1) is pseudo-concave in some dimensions of C. This permits a first-order approach for those dimensions in Step 1, avoiding the need for discretization. Unlike grid search or linear programming—which require full discretization—partial concavity can be leveraged only within our algorithm, allowing some dimensions to be handled in continuous form.

The applicability of different methods to the two structures can be illustrated by the following table:

	Lagrangian Iteration	**LP**	**Deterministic Solution**
Decomposibility	✓	✓	×
Partial Concavity	✓	×	×

Table 1: Applicability of different methods to the two structures

4.2.1 Decomposability

To further discuss decomposability, we first give the formal definition.

Definition 4.1 *We call an economic decomposable, if there exists* $C_1, \cdots, C_d$, *s.t.* $C = \times_{k=1}^{d} C_k$, *and for any* $c = (c_1, \cdots, c_d)$, *we have*

$$f(a,c) = \sum_{k=1}^{d} f_k(a,c_k),\ g_i(a,c) = \sum_{k=1}^{d} g_{i,k}(a,c_k)(\forall 1 \le i \le m),\ h_j(a,c) = \sum_{k=1}^{d} h_{j,k}(a,c_k)(\forall 1 \le j \le \ell).$$

Decomposability is a common structure in multi-agent models; we illustrate it with specific examples from principal-agent problems in Section 5 and optimal taxation problems in Section 6. If an economic problem is decomposable, then given the Lagrange multipliers $\lambda \ge 0$, $\gamma \ge 0$, the Lagrangian function defined in (3.4) can be further simplified as

$$\begin{aligned}
\mathcal{L}(a,c;\lambda,\gamma) &= f(a,c) - \sum_{i=1}^{m} \lambda_i g_i(a,c) - \sum_{j=1}^{\ell} \gamma_{j,a} h_j(a,c) \\
&= \sum_{k=1}^{d} f_k(a,c_k) - \sum_{i=1}^{m} \lambda_i \sum_{k=1}^{d} g_{i,k}(a,c_k) - \sum_{j=1}^{\ell} \gamma_{j,a} \sum_{k=1}^{d} h_{j,k}(a,c_k) \\
&= \sum_{k=1}^{d} \left[f_k(a,c_k) - \sum_{i=1}^{m} \lambda_i g_{i,k}(a,c_k) - \sum_{j=1}^{\ell} \gamma_{j,a} h_{j,k}(a,c_k) \right] \\
&:= \sum_{k=1}^{d} \mathcal{L}_k(a,c_k;\lambda,\gamma).
\end{aligned}$$

In Step 1 of each iteration of our algorithm, we need to solve the problem

$$(a^*, c^*) = \arg \max_{a \in A,\, c \in C} \mathcal{L}(a,c;\lambda,\gamma), \tag{4.2}$$

for given $\lambda \ge 0$, $\gamma \ge 0$. Without decomposability (or any other simplifying structure), this would require a grid search of size $|A|\,|\hat{C}| = |A| \left(\prod_{k=1}^{d} |\hat{C}_k| \right)$, where $\hat{C}_k$ is a discretization of C_k and $\hat{C}$ is the corresponding discretization of $C = \times_{k=1}^{d} C_k$.

With decomposability, however, the maximization problem (4.2) simplifies dramatically. Using the decomposition $\mathcal{L}(a,c;\lambda,\gamma) = \sum_{k=1}^{d} \mathcal{L}_k(a,c_k;\lambda,\gamma)$, we obtain

$$\begin{aligned}
(a^*, c^*) &= \arg \max_{a \in A,\, c \in C} \sum_{k=1}^{d} \mathcal{L}_k(a,c_k;\lambda,\gamma) \\
&= \arg \max_{a \in A} \sum_{k=1}^{d} \max_{c_k \in C_k} \mathcal{L}_k(a,c_k;\lambda,\gamma),
\end{aligned} \tag{4.3}$$

which requires only a grid search of scale $|A| \left(\sum_{k=1}^{d} |\hat{C}_k| \right)$ – a substantial reduction when

$d > 1$.

We do not know any way to take advantage of decomposability in finding deterministic solution. However, similar complexity reduction can be achieved for linear programming, though in a less direct manner. We discuss this in detail in Appendix B.2.1.

4.2.2 Partially Concavity

Partially concavity is a stronger structure than decomposability. It allows us to handle Step 1 of our algorithm without discretizing the consumption set C, by instead employing a first-order approach for certain dimensions. To lay the groundwork for further discussion, we first provide a formal definition of partially concavity.

Definition 4.2 *We call an economic partially concave, if there exists C_0, C_1, s.t. $C = C_0 \times C_1$, and for any $\lambda \geq 0$, $\gamma \geq 0$, $a \in A$, $c_0 \in C_0$, the Lagrangian $\mathcal{L}(a, (c_0, c_1); \lambda, \gamma)$ is continuously differentiable and pseudo-concave[6] in c_1.*

A sufficient condition for this property is obviously that for all $a \in A$, $c_0 \in C_0$ the objective function f and the constraints $-g, -h$ of the original problem (3.1) are concave in c_1. As we will illustrate in a simple example in Section 5 this is not a necessary condition. In Online Appendix B.1.1 we give a more general sufficient condition for a partially concave problem, and derive the worst case complexity of our method for this case. Intuitively, partial concavity is a natural feature of many economic problems since many economic problems are likely to be convex in some variables. More specific examples will be discussed in Section 5 and Section 6.

If the economic problem is partially concave, the computational burden can be reduced by using methods from convex optimization to solve for c_1 in Step 1 of every iteration of our Lagrangian algorithm. Specifically, for a partially concave problem, Step 1 can be written as

$$\begin{aligned}(a^*, c^*) &= \arg \max_{a \in A,\, c \in C} \mathcal{L}(a, c; \lambda, \gamma) \\ &= \arg \max_{a \in A,\, c_0 \in C_0} \max_{c_1 \in C_1} \mathcal{L}(a, (c_0, c_1); \lambda, \gamma),\end{aligned} \tag{4.4}$$

for given $\lambda \geq 0$, $\gamma \geq 0$. The maximization problem (4.4) thus decomposes into two subproblems:

[6] Following Mangasarian (1965) we say a differentiable function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is pseudo-concave if

$$x, y \in X : \quad D_x f(x) \cdot (x - y) \geq 0 \Rightarrow f(x) \geq f(y).$$

1. For each $a \in A$ and $c_0 \in C_0$, solve

$$c_1^*(a, c_0) = \arg\max_{c_1 \in C_1} \mathcal{L}(a, (c_0, c_1); \lambda, \gamma). \tag{4.5}$$

2. Solve

$$(a^*, c_0^*) = \arg\max_{a \in A,\, c_0 \in C_0} \mathcal{L}\big(a, (c_0, c_1^*(a, c_0)); \lambda, \gamma\big). \tag{4.6}$$

Because $\mathcal{L}$ is pseudo-concave in c_1, abstracting from boundary solutions, the subproblem (4.5) can be solved by the first-order condition:

$$\nabla_{c_1} \mathcal{L}\big(a, (c_0, c_1^*(a, c_0)); \lambda, \gamma\big) = 0. \tag{4.7}$$

The first-order condition (4.7) is typically analytically solvable or easily solved numerically, eliminating the need to discretize C_1 and perform a grid search. This not only improves efficiency, but also reduces memory usage. Depending on the dimension of c^1 this reduction can be substantial.

Partial concavity is an analytical property of the continuous economic problem. It cannot be effectively exploited by linear programming or grid-search methods for finding a deterministic solution, as these approaches require discretizing the problem first and thus abandon its analytical structure. In the remainder of the paper, we then consider economic applications and illustrate how these two properties play key roles for the tractability of our method.

5 Application to a Textbook Principal-Agent Problem

In this section, we illustrate our method as well as its advantages over the conventional linear programming approach using a textbook principal agent problem with moral hazard (e.g., in Salanié (2005, Section 5.2)). The agent can take unobserved actions $a \in A$, which influence the probability distribution of their output $q \in Q$ through conditional probabilities $p(q|a)$. 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 a closed interval. For a given output q and a given consumption contract $c(q)$, the principal has utility $v(q - c(q))$. In the deterministic solution, the principal's goal is to allocate an action and a consumption contract $\{a, c(q)\}$ for different states of output to each agent to

maximize her expected utility:

$$\max_{a,c(q)} \sum_q p(q|a)v(q-c(q)), \tag{5.1}$$

subject to the participation constraint and incentive compatibility constraint:

$$\begin{aligned} &\sum_q p(q|a)u(c(q),a) \geq \underline{U}; \\ &\sum_q p(q|a)u(c(q),a) \geq \sum_q p(q|\hat{a})u(c(q),\hat{a}),\ \forall \hat{a} \in A. \end{aligned} \tag{5.2}$$

Comparing this problem to the general form (3.1), the function $f(a,c)$ is defined as $f(a,c) = \sum_q v(q-c(q))p(q|a)$. The participation constraint function is $g_1(a,c) = \sum_q u(c(q),a)p(q|a) - \underline{U}$. The incentive constraint functions are $h_{\hat{a}}(a,c) = \sum_q p(q|\hat{a})u(c(q),\hat{a}) - \sum_q p(q|a)u(c(q),a)$, $\forall \hat{a} \in A$.

Lottery Solutions. We now consider the lottery solution for the moral hazard problem (see, e.g. Arnott and Stiglitz (1988)). Instead of choosing a deterministic action and outcome-contingent consumption contract $\{a, c(q)\}$, the planner chooses a probabilistic allocation over actions and consumption contracts. Mathematically, the problem is formulated 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 (5.1) becomes

$$\sum_{a \in A} \int_{c \in C^{|Q|}} x(a, dc) \sum_{q \in Q} p(q|a)v(q-c(q)). \tag{5.3}$$

The participation constraint in (5.2) is assumed to only hold in expectations,

$$\sum_{a \in A} \int_{c \in C^{|Q|}} x(a, dc) \sum_{q \in Q} p(q|a)u(c(q),a) \geq \underline{U}, \tag{5.4}$$

and the incentive constraints in (5.2) become

$$\int_{c \in C^{|Q|}} x(a, dc) \sum_q p(q|a)u(c(q),a) \geq \int_{c \in C^{|Q|}} x(a, dc) \sum_q p(q|\hat{a})u(c(q),\hat{a}),\ \forall (a,\hat{a}) \in A \times A. \tag{5.5}$$

The interpretation is that the principal randomizes over actions and compensation schedules, commits to a possibly random compensation schedule to each agent and announces a recommended action to the agent. The agent undertakes the recommended action if the

expected payoff is weakly larger than for all other actions, taking the random compensation schedule as given. The agent commits to participate before he observes the realization of the lottery over recommended actions. The incentive constraint has to hold for each recommended action.

We note that the economy is decomposable as in Definition 4.1; hence we can define $\pi \in \mathcal{P}(C \times Q \times A)$ such that $\pi(dc, q, a) = x(a, dc)p(q|a)$, and then the planner's objective function (5.3) can be written as

$$\sum_{q,a} \int_c \pi(dc, q, a)v(q - c), \tag{5.6}$$

while the participation constraint (5.4) and the incentive constraint (5.5) become

$$\sum_{q,a} \int_c \pi(dc, q, a)u(c, a) \geq \underline{U}, \tag{5.7}$$

$$\text{and } \sum_q \int_c \pi(dc, q, a)u(c, a) \geq \sum_q \int_c \pi(dc, q, a)\frac{p(q|\hat{a})}{p(q|a)}u(c, \hat{a}), \quad \forall(a, \hat{a}) \in A \times A. \tag{5.8}$$

respectively.

To apply the Lagrangian iteration algorithm, note that the deterministic Lagrangian can be written as

$$\mathcal{L}(a, c; \lambda, \gamma) = \sum_q p(q|a)v(q - c(q)) - \sum_{\hat{a} \in A} \gamma_{a,\hat{a}} \sum_q (p(q|\hat{a})u(c(q), \hat{a}) - p(q|a)u(c(q), a)) \tag{5.9}$$

$$- \lambda(\underline{U} - \sum_q p(q|a)u(c(q), a)). \tag{5.10}$$

It is easy to see that the problem is decomposable in outcomes $q \in Q$ and that we can write

$$\mathcal{L}(a, c; \lambda, \gamma) = \sum_{q \in Q} \mathcal{L}_q(a, c(q); \lambda, \gamma).$$

For the discretized version of the problem with finite C, decomposability implies that for a given a, λ, γ one needs to consider $|Q||C|$ possible c instead of $|C|^{|Q|}$ points. However, under mild assumptions on fundamentals, the problem becomes partially convex and one does not need to discretize the set C at all. The following lemma provides sufficient conditions for partially concavity in this setting.

Lemma 5.1 *Suppose $v(.)$ is C^2, concave and increasing and for each $a \in A$, $u(., a)$ is C^2*

concave and increasing. Also assume that for each $a \in A$, $\frac{\partial u(c,a)}{\partial c} = C(a)\tilde{u}'(c)$ *for a decreasing function* $\tilde{u}'(.)$. *Then the economy is partially concave, i.e. the Lagrangian* $\mathcal{L}(a, c; \lambda, \gamma)$ *is continuously differentiable and pseudo-concave in* c.

5.1 A numerical example

We adopt the parameterization following Example 1 in Prescott (1998).[7] 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 agents' utility function is defined as $u(a, c) = \sqrt{c} + 0.8\sqrt{2 - a}$ and the principal is risk neutral. The relationship between output and action is given by:

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

Clearly, the assumptions in Lemma 5.1 are satisfied, and the problem is partially concave. Therefore, in each Step 1 of our algorithm, the optimal consumption $c^k(a, q, \lambda^k, \gamma^k)$ is determined by the first-order condition (FOC):

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

where

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

when $\mathcal{A}(a, q, \gamma^k) > 0$; and the optimal consumption is determined by

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

when $\mathcal{A}(a, q, \gamma^k) \leq 0$. Thus, we can analytically determine the optimal consumption for each iteration using the FOC.

To solve this problem with linear programming, Prescott (1998) discretizes the set C to $\hat{C} = 0 : 0.01 : 2$, and obtains the following solution:

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

[7] The formulation of the principal agent problem in Prescott (1998) is slightly non-standard in that he maximizes the agent's expected utility subject to the resource constraint. However, it is easy to see that if we set the outside option $\underline{U}$ to the maximizing value in Prescott (1998), his problem is mathematically equivalent to our framework.

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

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

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

In this solution, the expected utility of the agent is approximately 1.8950. We use our algorithm with $\underline{U} = 1.8950$ and continuous c. We choose the initial Lagrangian multipliers as $\lambda^1 = 0.5$ and $\gamma^1 = 0$. The learning rate is chosen as

$$\mu^k = \frac{1}{\left(k + \frac{1}{|\Delta a|^2}\right)^{0.7}}.$$

The total number of iterations, N, is chosen from the relation $N \propto \ell \sim \frac{1}{\Delta a}$, as derived from the condition (3.18) and we take $100/\Delta a = 4000$ iterations. The trajectories of λ^k, a^k, c^k, and $V(\lambda^k, \gamma^k)$ are shown in Figure 1. The action a^k oscillates between 0.050 and 1.075 during iterations. Notably, there exists a relationship between λ^k and a^k: when λ^k is below a certain threshold, indicating that the principal places smaller emphasis on satisfying the participation constraint, the resulting action a^k stabilizes around 1.075. Conversely, when λ^k exceeds this threshold, reflecting an increased focus on the participation constraint, a^k stabilizes around 0.050. We use the results from rounds 3750 to 4000 to construct a lottery $\hat{\pi}$. The resulting lottery is:

$$\hat{\pi}(a = 0.050) = \frac{\sum_{k=3750}^{4000} \mu^k \mathbf{1}_{a^k = 0.050}}{\sum_{k=3750}^{4000} \mu^k} \approx 0.0917.$$

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 highly consistent with those obtained with linear programming.

5.2 Efficiency and comparison with linear programming

In this textbook example, the Lagrangian iteration method exhibits highly efficient computational performance not only by exploiting the low dimensionality but also the decomposability and partial concavity of the problem.[8] In our small numerical example, our method

[8] As Prescott (1998) and Su and Judd (2007) point out, the decomposability in $c(q)$ also simplifies the linear programming approach and the number of variables can be taken to be roughly $|Q||C||A|$ (instead of $|C|^{|Q|}|A|$ in the general problem). However, the linear programming approach cannot exploit partial

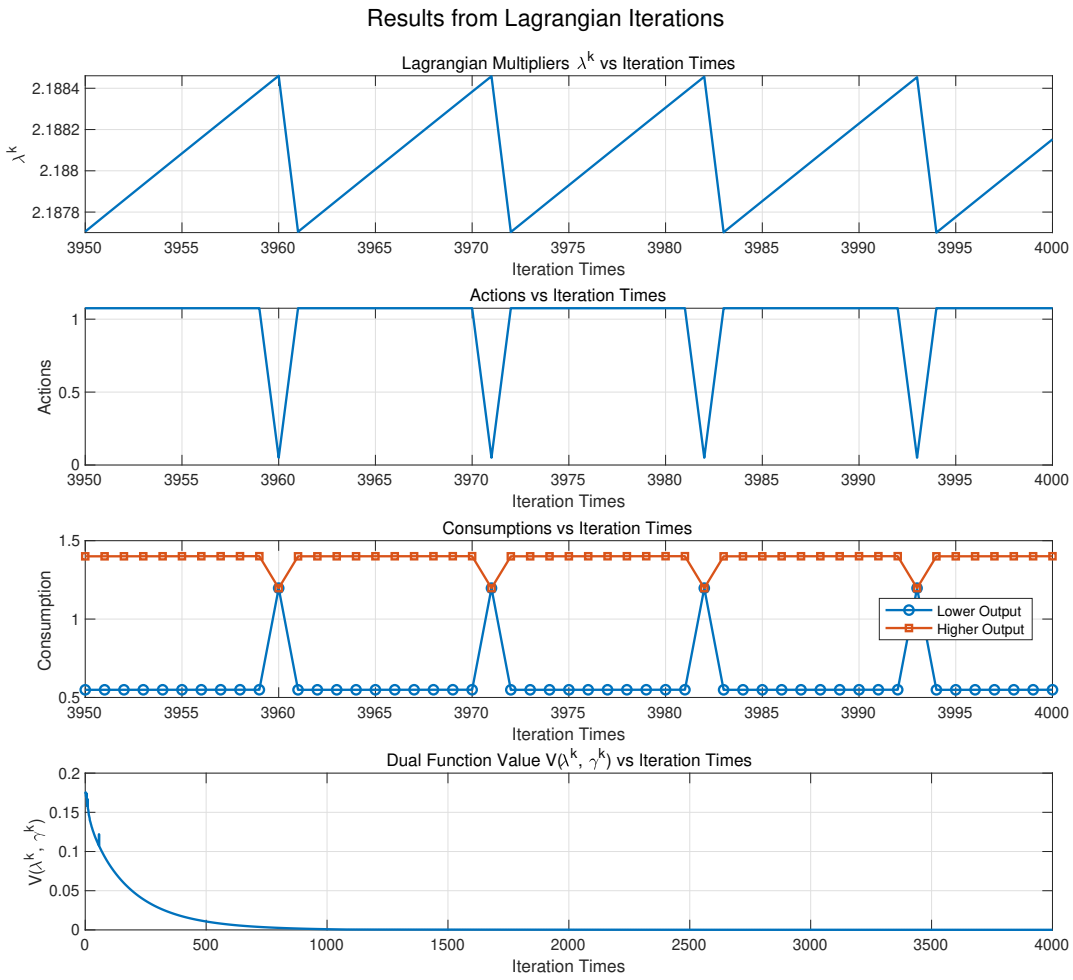


Figure 1: Relation between Lagrangian multipliers λ^k, optimal actions a^k, optimal consumptions c^k, value of the dual function $V(\lambda^k, \gamma^k)$ and Iteration Rounds when $\Delta a = 0.025$

takes only 0.05 seconds to run 4000 Lagrangian iterations to obtain the approximate lottery solution, while the linear programming method takes 1.69 seconds[9] using the optimization toolbox in MATLAB. As shown in the last sub-figure of Figure 1, the dual function $V(\lambda^k, \gamma^k)$ (defined in (3.14)) rapidly converges to the minimum $V(\lambda^*, \gamma^*) \approx 0$ within the first few hundreds of iterations. The subsequent iterations serve primarily to build lotteries and do not require large learning rates. A careful design of learning rates can therefore improve computational efficiency in practical applications.

It is clear that a finer discretization of C would dramatically increase the running time of

concavity and the size of the problem still grows dramatically in $|C|$, $|Q|$ and $|A|$.

[9]This CPU time accounts solely for the LP solver runtime, excluding the matrix construction time.

linear programming, but this does not affect our method since we do not need to discretize C. Slightly more surprising is the fact that using a larger set of possible actions slows down the linear programming approach dramatically while having only modest effects on the run-time of our algorithm. To demonstrate this, we vary Δa and perform $100/\Delta a$ iterations for each specification. The computational time for each choice of Δa is presented in Table 2. In addition, the table provides the sizes of the corresponding linear programming problems, highlighting the direct implications on memory requirements.

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 2: 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.

Although the linear programming solver can also handle the problem with $\Delta a = 0.0125$ within 12.93 seconds, 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. The discretization of the action set has a significant effect on the optimal solution.

6 Mirrleesian optimal taxation with multidimensional types

Following Mirrlees (1971), there is now a large literature on optimal income taxation. An important question is how the classical results change if one considers economies where taxpayers differ across several dimensions (see, e.g. Judd et al. (2017), Bergstrom and Dodds (2021)). Solving these models is computationally challenging since there is no justification for the so-called first order approach in the multidimensional case (see Judd et al. (2017)

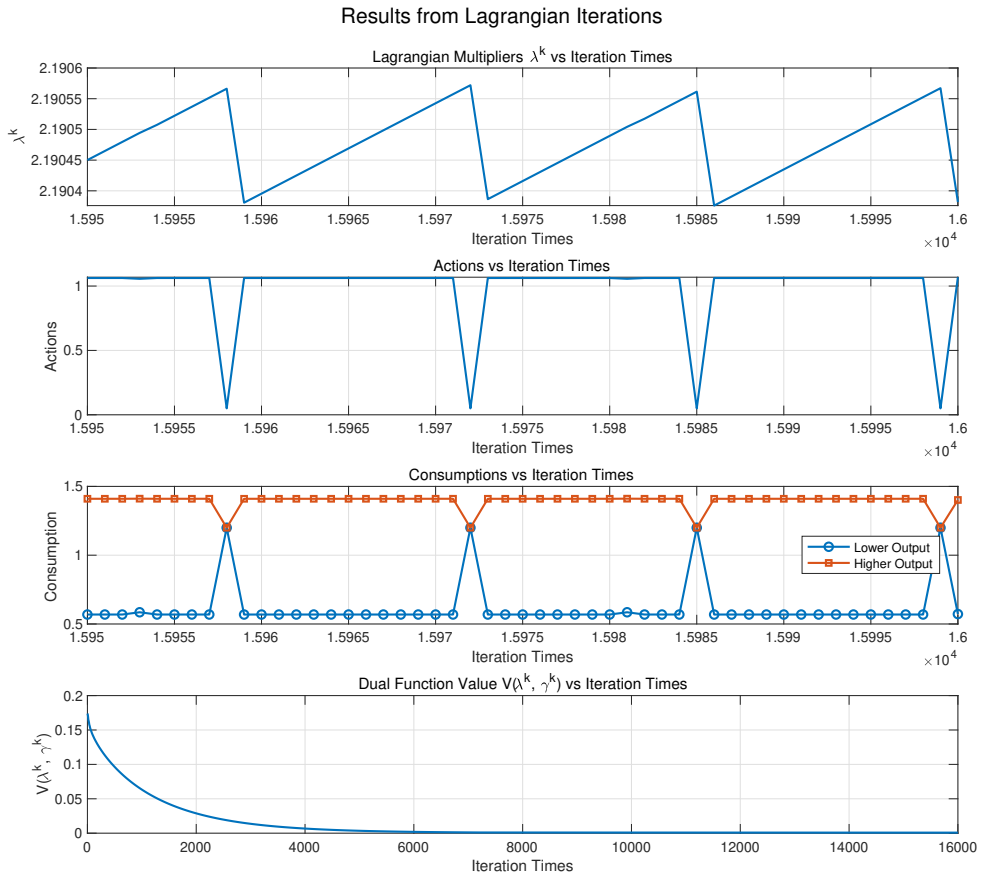


Figure 2: Relation between Lagrangian multipliers

and Boerma et al. (2022)). We show that our Lagrangian approach often makes the multi-dimensional problem tractable. Following Weiss (1976), Brito et al. (1995) or Hellwig (2007), we deviate from the classical setting in Saez (2001) and examine optimal taxation in an economy with a continuum of identical agents within each type where the planner can offer random tax-schedules that an agent must commit to prior to observing the randomized outcome and deciding how much labor to supply.

There are finitely many types of agents $\theta \in \Theta$, a finite set, and a continuum of identical agents within each type. A type is characterized by a preference parameter, η and a productivity parameters ω. Each agent of type $\theta = (\eta, \omega)$ has preferences over consumption (c) and effort (a), which we represent by a continuous utility function $\bar{u}_\eta(c, a)$. Each agent has type-specific productivity ω and, given effort a, an agent of type $\theta = (\eta, \omega)$ receives a wage ω and has income $y = \omega a$. While income is observable, the agent's effort and preferences are not. We assume that each $a \in [0, a_{\max}]$ - unlike above we want to think of a to be a

continuous variable. It is useful to write the indirect utility function over consumption and income as

$$u_\theta(c, y) = \bar{u}_\theta(c, \frac{y}{w_\theta}). \tag{6.1}$$

A deterministic allocation is therefore $(c_\theta, y_\theta)_{\theta\in\Theta} \in C \times Y$, where $C \subset \mathbb{R}^{|\Theta|}$ is compact and $Y = \times_{\theta\in\Theta}[0, w_\theta a_{\max}] \subset \mathbb{R}^{|\Theta|}$ is a $|\Theta|$-dimensional box.

The planner chooses a lottery over $(c_\theta, y_\theta)_{\theta\in\Theta}$ to solve

$$\max_{x\in\mathcal{P}(C\times Y)} \int_{(c,y)\in C\times Y} \left(\sum_{\theta\in\Theta} u_\theta(c_\theta, y_\theta)\right) x(dc, dy), \tag{6.2}$$

$$s.t. \int_{(c,y)\in C\times Y} (u_\theta(c_\theta, y_\theta) - u_\theta(c_{\theta'}, y_{\theta'})))x(dc, dy) \geq 0 \text{ for all } \theta, \theta' \in \Theta. \tag{6.3}$$

$$\int_{(c,y)\in C\times Y} \left(\sum_{\theta\in\Theta} c_\theta - \sum_{\theta\in\Theta} y_\theta\right) x(dc, dy) \leq 0. \tag{6.4}$$

The assumption of expected utility implies that the IC-constraint (6.3) only has to hold in expectations. Since we assume a continuum of agents within each type the resource constraints for consumption and for labor-supply only have to hold in expectations and they are combined into a single constraint (6.4). This problem is equivalent to the optimal taxation problem where the planner chooses a possibly stochastic income tax schedule $T_\theta(y)$ for each type θ. After each agent reports her type $\hat{\theta}$, the planner uses a public randomization device to observe signal s which then implies a tax schedule $T_{s,\hat{\theta}}(.)$ for reported type $\hat{\theta}$. For simplicity and without loss of generality, we think of the tax schedule as being characterized by an income level $\bar{y}(s, \hat{\theta})$ and being of the form $T_{s\hat{\theta}}(y) = y$ if $y \neq \bar{y}(s, \hat{\theta})$ and $T_{s\hat{\theta}}(y) = \bar{T}_{s\hat{\theta}}$ if $y = \bar{y}(s, \hat{\theta})$ (see, e.g., Brito et al. (1995) or Hellwig (2007)). This ensures that by reporting type $\hat{\theta}$ an agent commits to earning an income $\bar{y}(s, \hat{\theta})$ and obtaining consumption $c(s, \hat{\theta}) = \bar{y}(s, \hat{\theta}) - \bar{T}_{s,\hat{\theta}}$ whenever the public signal is s.

It is easy to verify that this problem fits directly into our general setup. Moreover, our algorithm can be applied even when the number of agents becomes large. The reason for this is that the problem is decomposable as described in Section 4 and the Lagrangian to the deterministic problem can be written as

$$\begin{aligned}\mathcal{L}(c, y, \lambda, \gamma) &= \sum_{\theta\in\Theta} u_\theta(c_\theta, y_\theta) + \sum_{\theta,\theta'\in\Theta} \lambda_{\theta,\theta'}(u_\theta(c_\theta, y_\theta) - u_\theta(c_{\theta'}, y_{\theta'})) - \gamma(\sum_{\theta\in\Theta} c_\theta - \sum_{\theta\in\Theta} y_\theta) \\ &= \sum_{\theta\in\Theta} \mathcal{L}_\theta(c_\theta, y_\theta, \lambda, \gamma),\end{aligned}$$

where $\lambda = (\lambda_{\theta,\theta'})_{\theta,\theta' \in \Theta}$ and

$$\mathcal{L}_\theta(c_\theta, y_\theta, \lambda, \gamma) = (1 + \sum_{\theta' \in \Theta, \theta' \neq \theta} \lambda_{\theta,\theta'}) u_\theta(c_\theta, y_\theta) - \sum_{\theta' \in \Theta, \theta' \neq \theta} \lambda_{\theta',\theta} u_{\theta'}(c_\theta, y_\theta) - \gamma(c_\theta - y_\theta).$$

In Step 1 of Algorithm 1, one therefore has to solve $|\Theta|$ maximization problems in two variables. The computational costs in Step 1 increase linearly in the number of types. The number of constraints increases polynomially, and therefore Proposition 3.1 implies that the number of iterations needed increases polynomially.

Note that the fact that the problem is decomposable can also be exploited when using linear programming to solve the problem. The probability that one needs to assign to (c, y) can be decomposed into independent probabilities for each (c_θ, y_θ). Suppose that for each θ, the set of feasible (c_θ, y_θ) is discretized into M points. Instead of solving a linear program with $M^{|\Theta|}$ variables that would become very quickly infeasible, one can solve a linear program with $M|\Theta|$ variables that remains feasible for moderate M. In contrast, a simple exhaustive search for the optimal deterministic solution is clearly infeasible, since one needs to search in $M^{|\Theta|}$ dimensional space.

6.1 Lottery solution in a calibrated example

To investigate optimal lotteries in this setting, we follow Judd et al. (2017)[10] and assume that for each type $\theta = (\eta, \omega) \in \Theta$, the utility function is given by

$$u_{(\eta,\omega)}(c, y) = \log(c) - \frac{\left(\frac{y}{\omega}\right)^{\frac{1}{\eta}+1}}{\frac{1}{\eta}+1}. \tag{6.5}$$

To understand how lotteries can improve upon this deterministic solution for this specific utility function, it is useful to consider a relaxed problem with partial incentive constraints as follows:

$$\int_{(c,y) \in C \times Y} (u_\theta(c_\theta, y_\theta) - u_\theta(c_{\theta'}, y_{\theta'}))\, x(dc, dy) \geq 0 \text{ for all } \theta, \theta' \text{ with } \eta_\theta \geq \eta_{\theta'}. \tag{6.6}$$

This constraint ensures that agents with lower Frisch elasticities will never pretend to be those with higher elasticities. If $a_{\max}$ is sufficiently large, the planner can always circumvent all other IC-constraints with lotteries. With very low probability, a given type is required to

[10] To the best of our knowledge, this is the only paper that solved the deterministic Mirrleesian problem for multi-dimensional types without additional simplifying assumptions. We use their deterministic optimal solutions to show the possible welfare gains from lotteries.

earn a large income. A type with the same productivity but lower η will never want to imitate the higher η type because the higher curvature in his utility over labor implies much higher disutility from the lottery. Without an upper bound on labor supply, the optimal value for the planner problem (6.2) subject to (6.4) and (6.6) can approximate arbitrarily well that of the original problem if we allow for lotteries. The following proposition formalizes this.

Proposition 6.1 *The maximum objective value for the planner problem* (6.2) *subject to* (6.4) *and* (6.3)*, and the maximum objective value for the planner problem* (6.2) *subject to* (6.4) *and* (6.6) *have the same limit as* $a_{\max} \to \infty$*.*

Following Section 5.1 in Judd et al. (2017), we choose five values of $\omega_\theta \in \{1, 2, 3, 4, 5\}$ and five values of $\eta_\theta \in \{\frac{1}{8}, \frac{1}{5}, \frac{1}{3}, \frac{1}{2}, 1\}$, resulting in 25 distinct types indexed by (w_θ, η_θ) and 600 incentive compatibility constraints. It is useful to first consider the "first best "solution where the planner faces no incentive constraints. In this case, consumption is constant across all types and income increases with productivity ω. For effort-levels above one, it is optimal for the planner to require less effort for lower η-types while for effort-levels below one, the planner requires more effort for lower η types. Columns 2 and 3 of Table 3 show the optimal deterministic solution compared to the first best solution. It is interesting to note that in the deterministic solution there is a large variation in consumption and a lot of "bunching"where different types receive the same allocation.

To understand how lotteries can improve upon this deterministic solution, it is useful to note the following.

Proposition 6.2 *There exists* $\bar{a}$ *such that for all* $a_{\max} \geq \bar{a}$ *the optimal lottery solution requires lotteries for all agents* (ω, η) *with* $\eta \neq \frac{1}{8}$*. In these solutions, lotteries always require these agents to earn income* $a_{\max}\omega$ *with some positive probability* π_θ*.*

The fourth column of Table 3 shows the optimal allocation in the limit. Compared to the deterministic solution, we observe that there is no bunching and that consumption across types is notably smoother. For types with productivity $\omega \in \{1, 2, 3\}$ the average required income is quite similar to the first best solution. In that solution, lower η types have to work more, and lotteries can be used to prevent them from imitating higher η types. For high productivity types, this is no longer the case; income is decreasing in η for $\omega = 4$ and almost constant for $\omega = 5$ while in the first-best it is now increasing in η[11]. The high η high ω types have an incentive to mimic the high η low ω types that have lower income (hence lower a). Although lotteries discourage low η types from doing so, the high η types' IC constraint can only be satisfied if they are not required to earn too high an income.

[11]This is due to the fact, that high productivity agents work more than one unit and the objective function is maximized if high η (low curvurture) agents work more. For $a < 1$ it is the other way around.

Type (ω, η)	Deterministic solution (c, y)	First Best (c, y)	$a_{\max} = \infty$ (c, Ey)	$a_{\max} = 1.2$ (c, Ey)	$a_{\max} = 0.8$ (c, Ey)
1, 1	1.68, 0.42	3.17, 0.32	1.93, 0.38 (L)	1.70, 0.43	1.52, 0.52
1, $\frac{1}{2}$	1.77, 0.62	3.17, 0.56	2.00, 0.63 (L)	1.79, 0.62	1.62, 0.69
1, $\frac{1}{3}$	1.79, 0.65	3.17, 0.68	2.03, 0.72 (L)	1.80, 0.65	1.63, 0.71 (L)
1, $\frac{1}{5}$	1.83, 0.77	3.17, 0.79	2.04, 0.82 (L)	1.84, 0.77	1.66, 0.80
1, $\frac{1}{8}$	1.86, 0.86	3.17, 0.87	2.05, 0.88	1.88, 0.86	1.66, 0.80
2, 1	1.86, 0.86	3.17, 1.26	2.21, 1.10 (L)	1.88, 0.86	1.78, 1.08
2, $\frac{1}{2}$	2.03, 1.39	3.17, 1.59	2.28, 1.48 (L)	2.05, 1.39 (L)	2.00, 1.60
2, $\frac{1}{3}$	2.07, 1.50	3.17, 1.72	2.29, 1.67 (L)	2.08, 1.49	2.00, 1.60
2, $\frac{1}{5}$	2.16, 1.74	3.17, 1.82	2.30, 1.87 (L)	2.18, 1.75	2.00, 1.60
2, $\frac{1}{8}$	2.2, 1.83	3.17, 1.89	2.30, 1.92	2.22, 1.84	2.00, 1.60
3, 1	2.2, 1.83	3.17, 2.84	2.77, 2.29 (L)	2.46, 2.50 (L)	2.39, 2.40
3, $\frac{1}{2}$	2.47, 2.49	3.17, 2.92	2.74, 2.62 (L)	2.61, 2.66 (L)	2.39, 2.40
3, $\frac{1}{3}$	2.47, 2.49	3.17, 2.95	2.68, 2.77 (L)	2.62, 2.69 (L)	2.39, 2.40
3, $\frac{1}{5}$	2.55, 2.68	3.17, 2.97	2.61, 2.91 (L)	2.69, 2.85	2.39, 2.40
3, $\frac{1}{8}$	2.62, 2.85	3.17, 2.98	2.56, 2.99	2.71, 2.88	2.39, 2.40
4, 1	3.36, 4.00	3.17, 5.06	3.62, 3.72 (L)	3.18, 3.52 (L)	2.75, 3.20
4, $\frac{1}{2}$	3.36, 4.00	3.17, 4.5	3.42, 3.92 (L)	3.36, 4.02 (L)	2.75, 3.20
4, $\frac{1}{3}$	3.36, 4.00	3.17, 4.32	3.26, 4.01 (L)	3.36, 4.03	2.75, 3.20
4, $\frac{1}{5}$	3.36, 4.00	3.17, 4.19	3.06, 4.06 (L)	3.36, 4.03	2.75, 3.20
4, $\frac{1}{8}$	3.36, 4.00	3.17, 4.12	2.91, 4.09	3.36, 4.03	2.75, 3.20
5, 1	4.87, 5.87	3.17, 7.9	4.76, 5.25 (L)	4.81, 5.85	2.91, 2.52 (L)
5, $\frac{1}{2}$	4.49, 5.57	3.17, 6.28	4.38, 5.34 (L)	4.47, 5.57	3.03, 4.00
5, $\frac{1}{3}$	4.34, 5.44	3.17, 5.82	4.08, 5.35 (L)	4.32, 5.44	3.03, 4.00
5, $\frac{1}{5}$	4.11, 5.25	3.17, 5.48	3.71, 5.31 (L)	4.10, 5.25	3.03, 4.00
5, $\frac{1}{8}$	4.00, 5.14	3.17, 5.29	3.39, 5.25	3.98, 5.14	3.03, 4.00

Table 3: Optimal deterministic solution versus first best versus optimal lottery solutions for different values of $a_{\max}$. (L) indicate lotteries. Deterministic solutions are from Judd et al. (2017).

With smaller values of a_{max} things become much more complicated, and statements can only be made for specific examples. We use our algorithm to compute the optimal solutions for $a_{\max} = 1.2$. Since we assume that utility for consumption is identical across all agents, each Lagrangian $\mathcal{L}_\theta(., y_\theta, \lambda, \gamma)$ is pseudo-concave in consumption, and this can be determined via the first order conditions. Step 1 of the Algorithm 1 therefore reduces to $|\Theta|$ one-dimensional grid searches, and the example can be solved in about 3 minutes using 2×10^6 Lagrangian iterations[12].

As a simple check, we also compute (the almost trivial) solution for the very tight constraint $a_{\max} = 0.8$. With the additional constraint $a \leq 1.2$ the optimal first-best solution would also change and with the constraint $a \leq 0.8$ both the first-best and the (second-best) deterministic solution change. Here, we want to focus on comparing the lottery solutions for the three cases. The last two columns in Table 3 show that there is no clear pattern on who receives lotteries and who does not. For $a_{\max} = 1.2$ no type with $\omega = 1, 5$ faces a lottery. For $\omega = 3$ we still observe three lotteries relaxing the IC constraints across different η. For $\omega = 4$, similar to the deterministic solution we obtain bunching. Many lotteries "disappear" as $a_{\max}$ decreases. For $a_{\max} = 0.8$ the upper bound is almost always binding in which case lotteries obviously become irrelevant. Interestingly, two types who did not face lotteries in the $a_{\max} = 1.2$ case now face lotteries. Agent (5,1) needs to be offered to work less in order not to imitate lower ω types. This can only be achieved efficiently if he does so randomly, deterring all agents with $\omega = 5$ and $\eta < 1$ to imitate him. For the $a_{max} = 1.2$ case this problem did not occur because the planner wanted him to work more (since effort was above one!) and compensated him with more consumption. Agent (1,1/3) faces a lottery that prevents him from being bundled with agents $(1, 1/5)$ and $(1, 1/8)$.

6.2 Welfare gain from randomization

We now examine under which conditions the use of lotteries can lead to significant welfare gains. 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

[12]We take the optimal deterministic allocations from Judd et al. (2017) as a benchmark and compare them to our optimal lottery solutions, choosing the initial Lagrangian multipliers as $\gamma^1 = 0.5$ and $\lambda^1 = 0$. The learning rate is chosen as

$$\mu^k = \begin{cases} \frac{1}{(k+1000)^{0.6}}, & 1 \leq k < 10^6; \\ \frac{1}{(k+1000)^{0.8}}, & 10^6 \leq k \leq 2 \times 10^6, \end{cases}$$

and we use the last 10^5 iteration rounds to construct the lottery solution.

welfare that can be achieved in the full information setup when the economy has m fewer units of total resources, defined by:

$$\Delta^W(m) = \max_{x \in \mathcal{P}(C \times Y)} \int_{(c,y) \in C \times Y} \left(\sum_{\theta \in \Theta} u_\theta(c_\theta, y_\theta) \right) x(dc, dy), \tag{6.7}$$

$$\text{s.t.} \int_{(c,y) \in C \times Y} \left(\sum_{\theta} c_\theta - \sum_{\theta} y_\theta \right) x(dc, dy) \leq -m. \tag{6.8}$$

The function $\Delta^W(m)$ is monotonic and decreases 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.

In our example, the deterministic mechanism yields a welfare loss of 7.16% relative to the first best full-information benchmark. As can be expected, potential welfare gains from lotteries depend crucially on $a_{\max}$. For $a_{\max} = 1.2$, allowing for lotteries reduces this loss to 6.91%[13], implying a relative welfare gain of approximately 3.49%. In the limit $a_{\max} \to \infty$, the welfare losses from asymmetric information shrink to 5.21% under the optimal lottery mechanism. This implies that 27.23% of the welfare loss from asymmetric information can be eliminated through lotteries and that the welfare gains from lotteries can be very large.

7 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 high-dimensional problems.

Our method has broad potential applications, since non-convexities are pervasive throughout economic theory. For example, Shen et al. (2025) extends our approach to recursive contract and proposes a general method for dynamic problems that directly addresses the long-standing critique that recursive contracts can mishandle non-convexities.

[13] We use the smallest dual value observed during the iterations for this estimation. The lottery constructed above yields a slightly lower loss of 6.88%, as some incentive constraints are slightly violated, with a maximum error of 7.6×10^{-4}.

References

Aiyagari, S. R., A. Marcet, T. J. Sargent, and J. Seppälä (2002): "Optimal taxation without state-contingent debt," Journal of Political Economy, 110, 1220–1254.

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.

Bertsekas, D. (2009): Convex optimization theory, vol. 1, Athena Scientific.

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.

Boyd, S. and L. Vandenberghe (2004): Convex optimization, Cambridge university press.

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.

Chang, F.-R. and D. Wildasin (1986): "Randomization of commodity taxes: an expenditure minimization approach," Journal of Public Economics, 31, 329–345.

Citanna, A. and M. Tvede (2024): "Do taxspots matter?: A study of optimal tax uncertainty," Journal of Economic Theory, 220, 105884.

Cremer, H., P. Pestieau, and J.-C. Rochet (2001): "Direct versus indirect taxation: the design of the tax structure revisited," International Economic Review, 42, 781–800.

Dávila, E. and A. Korinek (2018): "Pecuniary externalities in economies with financial frictions," The Review of Economic Studies, 85, 352–395.

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

Fudenberg, D. and D. K. Levine (1998): The theory of learning in games, MIT press.

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

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

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

Lemaréchal, C. (2001): "Lagrangian relaxation," in Computational combinatorial optimization: optimal or provably near-optimal solutions, Springer, 112–156.

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

Mangasarian, O. (1965): "Pseudo-Convex Functions," SIAM Journal on Control and Optimization, 3, 281.

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.

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

Sahni, S. (1974): "Computationally related problems," SIAM Journal on computing, 3, 262–279.

Salanié, B. (2005): The economics of contracts: a primer, MIT press.

Shen, C., F. Kübler, and Z. Zhou (2025): "Recursive contracts in non-convex environments," arXiv preprint arXiv:2511.20303.

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.

Appendices

A Proofs

A.1 Proof of Lemma 3.1

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

$$\mu(\lambda_1', \gamma_1') + (1 - \mu)(\lambda_2', \gamma_2') = (\lambda, \gamma),$$

(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 Theorem 3.1

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

$$\sum_{a\in A} \int_{c\in C} \varphi(a,c) x^{\epsilon_n}(a,dc) \to \sum_{a\in A} \int_{c\in C} \varphi(a,c) x^*(a,dc). \tag{A.2}$$

Next, we prove that x^* is the optimal solution to system (3.2).

We take $\varphi_1 \equiv 1$ in (A.2), 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.2), and obtain

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

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

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

Therefore x^* satisfy all the constraints in (3.2) and is a feasible probability measure to (3.2).

To show that x^* is an optimal solution to system (3.2), it then suffices to show that for any $x \in \mathcal{P}(A \times C)$, which is a feasible measure to (3.2), we have

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

We prove (A.3) by contradiction. If (A.3) does not hold, then we can find $\tilde{x} \in \mathcal{P}(A \times C)$,

such that $\tilde{x}$ satisfy all the constraints in (3.2) and

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc) < \sum_{a\in A}\int_{c\in C} f(a,c)\tilde{x}(a,dc). \tag{A.4}$$

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.4). However, we know that $\tilde{x}$ is a feasible probability measure for (3.2) and hence a feasible probability measure for (3.8) with $\epsilon = \epsilon_n$. This contradicts the fact that x^{ϵ_n} is the optimal solution to (3.8) with $\epsilon = \epsilon_n$. Therefore, we finish the proof.

A.3 Proof of Theorem 3.2

We first prove (3.9). By direct computation, we have

$$\begin{aligned}
&L(x;\lambda,\gamma)\\
=&\sum_{a\in A}\int_{c\in C} f(a,c)x(a,dc) - \sum_{i=1}^{m}\lambda_i\sum_{a\in A}\int_{c\in C} g_i(a,c)x(a,dc) - \sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}\int_{c\in C} h_j(a,c)x(a,dc)\\
=&\sum_{a\in A}\int_{c\in C}\left(f(a,c) - \sum_{i=1}^{m}\lambda_i g_i(a,c) - \sum_{j=1}^{\ell}\gamma_{j,a}h_j(a,c)\right)x(a,dc)\\
:=&\sum_{a\in A}\int_{c\in C}\mathcal{L}(a,c;\lambda,\gamma)x(a,dc)\\
\le&\max_{a\in A,c\in C}\mathcal{L}(a,c;\lambda,\gamma).
\end{aligned} \tag{A.5}$$

Hence

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

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 (3.9) holds.

Now we prove (3.10). According to (A.5), 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^*.

A.4 Proof of Theorem 3.3

We note that the equality (3.9) in Theorem 3.2 holds for any $(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+$. Therefore, (3.9) in Theorem 3.2 further yields that

$$\inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) = \inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+} \max_{a\in A, c\in C} \mathcal{L}(a,c;\lambda,\gamma), \tag{A.6}$$

where L is defined in (3.3). According to (3.9) and (A.6), $\mathcal{L}(a,c;\lambda,\gamma)$ in (3.12) and (3.13) can be replaced to $L(x;\lambda,\gamma)$, i.e. to prove Theorem 3.3 it suffices to show that

$$\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda^*,\gamma^*) = \inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma), \tag{A.7}$$

and

$$\inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+} \max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma) = \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc). \tag{A.8}$$

We first prove (A.8). Since x^* satisfies all the constraints in (3.2), for any $\lambda\geq 0$, $\gamma\geq 0$,

we know that

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

and hence

$$\begin{aligned}
&L(x^*;\lambda,\gamma)\\
=&\sum_{a\in A}\int_{c\in C}f(a,c)x^*(a,dc)-\sum_{i=1}^{m}\lambda_i\sum_{a\in A}\int_{c\in C}g_i(a,c)x^*(a,dc)-\sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}\int_{c\in C}h_j(a,c)x^*(a,dc)\\
\ge&\sum_{a\in A}\int_{c\in C}f(a,c)x^*(a,dc).
\end{aligned}\tag{A.9}$$

Since $(\lambda,\ \gamma)$ is arbitrary chosen in $\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}$, the inequality (A.9) yields

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

and further

$$\max_{x\in\mathcal{P}(A\times C)}\inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+}L(x;\lambda,\gamma)\ge\inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+}L(x^*;\lambda,\gamma)\ge\sum_{a\in A}\int_{c\in C}f(a,c)x^*(a,dc).\tag{A.10}$$

Due to the general inf-sup inequality, (A.10) implies that

$$\begin{aligned}
&\inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+}\max_{x\in\mathcal{P}(A\times C)}L(x;\lambda,\gamma)\\
(\text{Inf-sup Inequality})\ge&\max_{x\in\mathcal{P}(A\times C)}\inf_{(\lambda,\gamma)\in\mathbb{R}^m_+\times\mathbb{R}^{\ell|A|}_+}L(x;\lambda,\gamma)\\
(\text{A.10})\ge&\sum_{a\in A}\int_{c\in C}f(a,c)x^*(a,dc).
\end{aligned}\tag{A.11}$$

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

$$L(x^*;\lambda^*,\gamma^*)=\max_{x\in\mathcal{P}(A\times C)}L(x;\lambda^*,\gamma^*),\tag{A.12}$$

implying that

$$\begin{aligned}&\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}}\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma)\\ \leq&\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda^*,\gamma^*)\\ =&L(x^*;\lambda^*,\gamma^*).\end{aligned}\tag{A.13}$$

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

$$\begin{aligned}&L(x^*;\lambda^*,\gamma^*)\\ =&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc)-\sum_{i=1}^m\lambda_i^*\sum_{a\in A}\int_{c\in C} g_i(a,c)x^*(a,dc)-\sum_{j=1}^\ell\sum_{a\in A}\gamma_{j,a}^*\int_{c\in C} h_j(a,c)x^*(a,dc)\\ =&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc).\end{aligned}\tag{A.14}$$

We combine (A.13) and (A.14) to obtain

$$\inf_{(\lambda,\gamma)\in\mathbb{R}_+^m\times\mathbb{R}_+^{\ell|A|}}\max_{x\in\mathcal{P}(A\times C)} L(x;\lambda,\gamma)\leq\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc).\tag{A.15}$$

The equality (A.8) is then deduced from (A.11) and (A.15). Furthermore, the equality (A.7) can be implied directly from (A.8), (A.12), and (A.14). We can then combine (A.7) and (A.8) to finish the proof.

A.5 Proof of Theorem 3.4

Following the same steps in the proof for Theorem 3.1, 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 (3.2). We suppose that x^* is an optimal solution to (3.2), it suffices to show that

$$\sum_{a\in A}\int_{c\in C} f(a,c)\tilde{x}^*(a,dc)\geq\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc).\tag{A.16}$$

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

$$\sum_{a\in A}\int_{c\in C} f(a,c)\tilde{x}^{\epsilon_n}(a,dc)\geq\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,dc)-\epsilon_n.\tag{A.17}$$

The inequality (A.16) can then be directly obtained by taking limits on both sides of (A.17)

.

A.6 Proof of Theorem 3.5

The Theorem follows from the following two propositions

Proposition A.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 (3.2). 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 (3.2).*

Proof We first prove that property 1 of definition 3.4 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

$$\sum_{a\in A} \int_{c\in C} g_i(a,c) x^{n}(a, dc) \leq \frac{\lambda_i^{n+1} - \lambda_i^1}{\sum_{k=1}^{n} \mu^k}. \tag{A.18}$$

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 3.4 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

$$\int_{c \in C} h_j(a,c) x^n(a, dc) \leq \frac{\gamma_{j,a}^{n+1} - \gamma_{j,a}^1}{\sum_{k=1}^n \mu^k}. \tag{A.19}$$

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 3.4 holds, i.e.

$$\sum_{a \in A} \int_{c \in C} f(a,c) x^n(a, dc) \geq \sum_{a \in A} \int_{c \in C} f(a,c) x^*(a, dc) - \epsilon, \tag{A.20}$$

where x^* is the optimal solution of (3.2). 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

$$\sum_{a \in A} \int_{c \in C} g_i(a,c) x^n(a, dc) \to 0, \tag{A.21}$$

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

$$\int_{c \in C} h_j(a,c) x^n(a, dc) \to 0. \tag{A.22}$$

Before we prove (A.21) and (A.22), we first interpret these two formulas. Since we have

already proven the property 1 and property 2 of definition 3.4 hold, we now have

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

and

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

These two formulas (A.21) and (A.22) 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.21). 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 $n\ge N_i$ we have

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

and thus

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

Therefore

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

Hence

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x^{N_i+n}(a,dc)=\frac{\lambda_i^{N_i+n+1}-\lambda_i^{N_i}}{\sum_{k=1}^{N_i+n}\mu^k}+\frac{\sum_{k=1}^{N_i}\mu^k}{\sum_{k=1}^{N_i+n}\mu^k}\sum_{a\in A}\int_{c\in C} g_i(a,c)x^{N_i}(a,dc).\quad\text{(A.23)}$$

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.23) converges to 0 as $n \to \infty$, and we have finished the claim (A.21). The proof for (A.22) is the same as the one for (A.21) so we omit the details here.

Step 2. We then utilize these claims (A.21) and (A.22) to prove (A.20). By the definition of x^n, we have

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,dc) = \frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k f(a^k,c^k) \tag{A.24}$$

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

$$\begin{aligned} & f(a^k,c^k) - \sum_{i=1}^m \lambda_i^k g_i(a^k,c^k) - \sum_{j=1}^\ell \gamma_{j,a^k}^k h_j(a^k,c^k) \\ \geq & f(a,c) - \sum_{i=1}^m \lambda_i^k g_i(a,c) - \sum_{j=1}^\ell \gamma_{j,a}^k h_j(a,c), \text{ for all } a\in A, c\in C. \end{aligned} \tag{A.25}$$

The key of the proof is to utilize the inequality (A.25) for each $1 \leq k \leq n$ and give a lower bound for the right hand side of (A.24). To be precise, for any $a \in A$, $c \in C$, we have

$$\begin{aligned} & \frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k f(a^k,c^k) \\ = & \frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k \left(f(a^k,c^k) - \sum_{i=1}^m \lambda_i^k g_i(a^k,c^k) - \sum_{j=1}^\ell \gamma_{j,a^k}^k h_j(a^k,c^k) \right. \\ & \left. + \sum_{i=1}^m \lambda_i^k g_i(a^k,c^k) + \sum_{j=1}^\ell \gamma_{j,a^k}^k h_j(a^k,c^k) \right) \\ \geq & \frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k \left(f(a,c) - \sum_{i=1}^m \lambda_i^k g_i(a,c) - \sum_{j=1}^\ell \gamma_{j,a}^k h_j(a,c) \right. \\ & \left. + \sum_{i=1}^m \lambda_i^k g_i(a^k,c^k) + \sum_{j=1}^\ell \gamma_{j,a^k}^k h_j(a^k,c^k) \right) \end{aligned} \tag{A.26}$$

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.26),

and have

$$
\begin{aligned}
&\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k f(a^k,c^k)\\
\geq&\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left(f(a,c)-\sum_{i=1}^{m}\lambda_i^k g_i(a,c)-\sum_{j=1}^{\ell}\gamma_{j,a}^k h_j(a,c)\right.\\
&+\sum_{i=1}^{m}\lambda_i^* g_i(a^k,c^k)+\sum_{j=1}^{\ell}\gamma_{j,a^k}^* h_j(a^k,c^k)\\
&\left.+\sum_{i=1}^{m}(\lambda_i^k-\lambda_i^*)g_i(a^k,c^k)+\sum_{j=1}^{\ell}(\gamma_{j,a^k}^k-\gamma_{j,a^k}^*)h_j(a^k,c^k)\right)\\
=&f(a,c)-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{i=1}^{m}\lambda_i^k g_i(a,c)-\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{j=1}^{\ell}\gamma_{j,a}^k h_j(a,c)\\
&+\sum_{i=1}^{m}\lambda_i^*\sum_{a\in A}\int_{c\in C}g_i(a,c)x^n(a,dc)+\sum_{j=1}^{\ell}\sum_{a\in A}\gamma_{j,a}^*\int_{c\in C}h_j(a,c)x^n(a,dc)\\
&+\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left(\sum_{i=1}^{m}(\lambda_i^k-\lambda_i^*)g_i(a^k,c^k)\right)+\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\left(\sum_{j=1}^{\ell}(\gamma_{j,a^k}^k-\gamma_{j,a^k}^*)h_j(a^k,c^k)\right).
\end{aligned}
\tag{A.27}
$$

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.27) can then be written as

$$
\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k f(a^k,c^k)\geq I_1+I_2+I_3. \tag{A.28}
$$

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

$$\sum_{i=1}^{m}(\lambda_i^k - \lambda_i^*)g_i(a^k, c^k) + \sum_{j=1}^{\ell}(\gamma_{j,a^k}^k - \gamma_{j,a^k}^*)h_j(a^k, c^k) \to 0. \tag{A.29}$$

Since $\sum_{k=1}^{\infty}\mu^k = \infty$, (A.29) 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 $n > N_1$,

$$I_3 > -\frac{\epsilon}{2}. \tag{A.30}$$

Next, we consider the lower bound of I_2. By the formulas (A.21) and (A.22) 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

$$I_2 > -\frac{\epsilon}{2}. \tag{A.31}$$

We combine (A.28), (A.30) and (A.31) and see that when $n > \max\{N_1, N_2\}$, we have

$$\begin{aligned} &\frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k f(a^k, c^k) \\ \geq& I_1 - \epsilon \\ =& f(a,c) - \frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{i=1}^{m}\lambda_i^k g_i(a,c) - \frac{1}{\sum_{k=1}^{n}\mu^k}\sum_{k=1}^{n}\mu^k\sum_{j=1}^{\ell}\gamma_{j,a}^k h_j(a,c) - \epsilon \end{aligned} \tag{A.32}$$

We multiply $x^*(a, dc)$ to both sides of (A.32) and integral on $A \times C$. The inequality (A.32)

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 (3.2). Hence we prove the property 3 and finish the proof for Proposition A.1. □

Proposition A.2 *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 (3.2), 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 (3.4).

Proof We define the set of minimizers to the dual problem as

$$
\Lambda^* := \left\{ (\lambda^*, \gamma^*) \in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|},\, V(\lambda^*, \gamma^*) = \inf_{(\lambda,\gamma)\in \mathbb{R}_+^m \times \mathbb{R}_+^{\ell|A|}} V(\lambda, \gamma) \right\}. \tag{A.33}
$$

(3.12) in Theorem 3.3 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 3.1, $(\Delta\lambda, \Delta\gamma)$ defined in (3.15) 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 3.1

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:

$$\sum_{a\in A}\int_{c\in C} g_i(a,c)x^n(a,\mathrm{d}c) \le \frac{\bar{\Lambda}}{\sum_{k=1}^n \mu^k}; \tag{A.34}$$

$$\int_{c\in C} h_j(a,c)x^n(a,\mathrm{d}c) \le \frac{\bar{\Lambda}}{\sum_{k=1}^n \mu^k}; \tag{A.35}$$

and

$$\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c) \ge \sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c) - \frac{M(m+\ell)\sum_{k=1}^n(\mu^k)^2+\bar{\Lambda}}{\sum_{k=1}^n \mu^k}. \tag{A.36}$$

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 (3.17) holds, x^n is an ϵ-optimal solution.

Step 1. According to (A.18) and (A.19) in the proof of Proposition A.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.34) and (A.35) hold.

It then remains to show that (A.36) holds in this step. The proof for (A.36) can be further divided into three sub-steps. First we show that

$$\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c)\\
\ge&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c)\\
&+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^\ell\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k).
\end{aligned} \tag{A.37}$$

Then, we give an estimate of the term

$$\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^\ell\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k) \tag{A.38}$$

in the right hand side of (A.37). Finally, we conclude (A.36).

1-1. Show that (A.37) holds. According to (A.24) and (A.26) in the proof of Proposition A.1, for every $a\in A$, $c\in C$, we have[14]:

$$\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c)\\
\geq&\frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k\left(f(a,c)-\sum_{i=1}^m\lambda_i^k g_i(a,c)-\sum_{j=1}^\ell\gamma_{j,a}^k h_j(a,c)+\sum_{i=1}^m\lambda_i^k g_i(c^k)+\sum_{j=1}^\ell\gamma_{j,a^k}^k h_j(a^k,c^k)\right)\\
=&f(a,c)-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a,c)-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^\ell\sum_{k=1}^n \mu^k\gamma_{j,a}^k h_j(a,c)\\
&+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^\ell\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k).
\end{aligned}\tag{A.39}$$

We multiply $x^*(a,dc)$ to both sides in (A.39) and integral on $A\times C$ and have

$$\begin{aligned}
&\sum_{a\in A}\int_{c\in C} f(a,c)x^n(a,\mathrm{d}c)\\
\geq&\sum_{a\in A}\int_{c\in C} f(a,c)x^*(a,\mathrm{d}c)-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k\sum_{i=1}^m\lambda_i^k\sum_{a\in A}\int_{c\in C} g_i(a,c)x^*(a,dc)\\
&-\frac{1}{\sum_{k=1}^n \mu^k}\sum_{k=1}^n \mu^k\sum_{j=1}^\ell\sum_{a\in A}\gamma_{j,a}^k\int_{c\in C} h_j(a,c)x^*(a,dc)\\
&+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{i=1}^m\sum_{k=1}^n \mu^k\lambda_i^k g_i(a^k,c^k)+\frac{1}{\sum_{k=1}^n \mu^k}\sum_{j=1}^\ell\sum_{k=1}^n \mu^k\gamma_{j,a^k}^k h_j(a^k,c^k).
\end{aligned}\tag{A.40}$$

Since $x^*(a,dc)$ satisfies all the constraints in (3.2), i.e.

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

[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 $a\in A$, $c\in C$.

we thus obtain the estimate (A.37) from (A.40).

1-2. Estimate the term (A.38). Recall that (A.38) is

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

By direct computation, we have

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

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

$$\lambda_i^{k+1}-\lambda_i^k=\mu^k\tau_i^k g_i(a^k,c^k),\tag{A.42}$$

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

$$\gamma_{j,a^k}^{k+1}-\gamma_{j,a^k}^k=\mu^k\tau_{j,,a^k}^k h_j(a^k,c^k),\tag{A.43}$$

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.42) and (A.43) to simplify the right hand side of (A.41) and obtain

$$
\begin{aligned}
&\Lambda(\lambda^{k+1},\gamma^{k+1})-\Lambda(\lambda^k,\gamma^k)\\
=&\left(\sum_{i=1}^{m}2\mu^k\tau_i^k g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}2\mu^k\tau_{j,a^k}^k h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
&+\left(\sum_{i=1}^{m}(\mu^k)^2(\tau_i^k)^2 g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(\mu^k)^2(\tau_{j,a^k}^k)^2 h_j^2(a^k,c^k)\right)\\
=&2\mu^k\left(\sum_{i=1}^{m}g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
&-2\mu^k\left(\sum_{i=1}^{m}(1-\tau_i^k)g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}(1-\tau_{j,a^k}^k)h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
&+\left(\sum_{i=1}^{m}(\mu^k)^2(\tau_i^k)^2 g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(\mu^k)^2(\tau_{j,a^k}^k)^2 h_j^2(a^k,c^k)\right).
\end{aligned}
\tag{A.44}
$$

Furthermore, when $1-\tau_i^k>0$, we know that $\lambda_i^{k+1}=0$, and according to (A.42), we have

$$
\lambda_i^k=-\mu^k\tau_i^k g_i(a^k,c^k). \tag{A.45}
$$

Similarly, when $1-\tau_{j,a^k}^k>0$, we have the following equality according to (A.43)

$$
\gamma_{j,a^k}^k=-\mu^k\tau_{j,a^k}^k h_j(a^k,c^k). \tag{A.46}
$$

We utilize the equality (A.45) and the equality (A.46) to obtain

$$
\begin{aligned}
&-2\mu^k\left(\sum_{i=1}^{m}(1-\tau_i^k)g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^{\ell}(1-\tau_{j,a^k}^k)h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
=&2(\mu^k)^2\left(\sum_{i=1}^{m}(1-\tau_i^k)\tau_i^k g_i^2(a^k,c^k)+\sum_{j=1}^{\ell}(1-\tau_{j,k}^k)\tau_{j,a^k}^k h_j^2(a^k,c^k)\right).
\end{aligned}
\tag{A.47}
$$

We combine (A.44) and (A.47), then we have

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

which implies that

$$
\begin{aligned}
&2\mu^k\left(\sum_{i=1}^m g_i(a^k,c^k)\lambda_i^k+\sum_{j=1}^\ell h_j(a^k,c^k)\gamma_{j,a^k}^k\right)\\
=&\big(\Lambda(\lambda^{k+1},\gamma^{k+1})-\Lambda(\lambda^k,\gamma^k)\big)\\
&-(\mu^k)^2\left(\sum_{i=1}^m(2-\tau_i^k)\tau_i^kg_i^2(a^k,c^k)+\sum_{j=1}^\ell(2-\tau_{j,a^k}^k)\tau_{j,a^k}^kh_j^2(a^k,c^k)\right).
\end{aligned}
\tag{A.48}
$$

We add up equality (A.48) for $k = 1, \cdots, n$. After some direct computation, we have an estimate to (A.38) as

$$\begin{aligned}
&\frac{1}{\sum_{k=1}^n \mu^k} \sum_{i=1}^m \sum_{k=1}^n \mu^k \lambda_i^k g_i(a^k, c^k) + \frac{1}{\sum_{k=1}^n \mu^k} \sum_{j=1}^\ell \sum_{k=1}^n \mu^k \gamma_{j,a^k}^k h_j(a^k, c^k) \\
=&\frac{\Lambda(\lambda^{n+1}, \gamma^{n+1}) - \Lambda(\lambda^1, \gamma^1)}{2\sum_{k=1}^n \mu^k} \\
&- \frac{\sum_{k=1}^n (\mu^k)^2 \left(\sum_{i=1}^m (2 - \tau_i^k) \tau_i^k g_i^2(a^k, c^k) + \sum_{j=1}^\ell (2 - \tau_{j,a^k}^k) \tau_{j,a^k}^k h_j^2(a^k, c^k) \right)}{2\sum_{k=1}^n \mu^k}.
\end{aligned} \tag{A.49}$$

1-3. We then utilize the estimate (A.49) to derive (A.36). We combine (A.37) and (A.49) and obtain

$$\begin{aligned}
&\sum_{a\in A} \int_{c\in C} f(a,c) x^n(a, \mathrm{d}c) \\
\geq &\sum_{a\in A} \int_{c\in C} f(a,c) x^*(a, \mathrm{d}c) + \frac{\Lambda(\lambda^{n+1}, \gamma^{n+1}) - \Lambda(\lambda^1, \gamma^1)}{2\sum_{k=1}^n \mu^k} \\
&- \frac{\sum_{k=1}^n (\mu^k)^2 \left(\sum_{i=1}^m (2 - \tau_i^k) \tau_i^k g_i^2(a^k, c^k) + \sum_{j=1}^\ell (2 - \tau_{j,a^k}^k) \tau_{j,a^k}^k h_j^2(a^k, c^k) \right)}{2\sum_{k=1}^n \mu^k}.
\end{aligned} \tag{A.50}$$

By the fact that $|g_i| \leq M$, $|h_j| \leq M$, the estimate (A.50) 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.36) holds.

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

$$\sum_{k=1}^n \mu^k \sim \begin{cases} \frac{n^{\frac{1}{2}(1-\rho)}}{M\sqrt{m+\ell}}, & \rho < 1 \\ \frac{\log n}{M\sqrt{m+\ell}}, & \rho = 1, \end{cases} \tag{A.51}$$

and

$$\sum_{k=1}^{n}(\mu^k)^2 \sim \frac{1}{M^2\rho(m+\ell)}. \tag{A.52}$$

Thus, for any $\epsilon > 0$, when (3.17) 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.51) and (A.52) to (A.34),(A.35) and (A.36), 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 Lemma 5.1

For all $q \in Q$ the partial derivative of the Lagrangian with respect to $c(q)$ can be written as

$$\frac{\partial\mathcal{L}(a,c;\lambda,\gamma)}{\partial c(q)} = -p(q|a)v'(q-c(q))-\tilde{u}'(c(q))\left(\sum_{\hat{a}\in A}\gamma_{a,\hat{a}}(p(q|\hat{a})C(\hat{a}) - p(q|a)C(a)) + \lambda p(q|a)C(a)\right).$$

If for a given $a \in A$ the term in the brackets is positive, the derivative is negative and the function is strictly differentiably monotone and hence (since one-dimensional) pseudo-concave. If the term in the brackets is negative the partial derivative is decreasing in $c(q)$ and hence the Lagrangian is concave (and therefore quasi-concave).

A.9 Proof of Proposition 6.1

To prove the proposition we first need the following results that is interesting in itself. It shows that if the original IC constraint 6.3 is replaced by 6.6, the optimal solution never

involves lotteries.

Proposition A.3 *We consider the planner problem (6.2) subject to (6.4) and (6.6) with the consumption set $C = \times_{h\in H}[\epsilon, c_{\max}]$, the output set $Y = \times_{h\in H}[0, a_{max}w_h]$, and the utility function defined by (6.5). Assume that the Slater's condition holds, then the optimal solution to the planner problem does not involve lotteries.*

To proceed in the proofs of the propositions, we will need the following technical lemma.

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

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

where $a_j \geq 0$, $\forall j \neq i$. and $a_0 > 0$. Then one of the following holds:

1. *f_i admits a unique maximal point in the interval $[0, w_h a_{\max}]$.*
2. *f_i admits two maximal points in the interval $[0, w_h a_{\max}]$, and one equals $w_h a_{\max}$.*

Furthermore, when $a_{i+1} = \cdots = a_N = 0$, then f admits a unique maximal point in the interval $[0, w_h a_{\max}]$.

Proof 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 $a_i > 0$, we define $p_0 = 1$, and it is straightforward to have

$$f_i'(y) = \sum_{j=i+1}^{N} a_j p_j y^{p_j - 1} - a_i p_i y^{p_i - 1} + \sum_{j=0}^{i-1} a_j p_j y^{p_j - 1},$$

and

$$\begin{aligned} f_i''(y) =& \sum_{j=i+1}^{N} a_j p_j (p_j - 1) y^{p_j - 2} - a_i p_i (p_i - 1) y^{p_i - 2} + \sum_{j=1}^{i-1} a_j p_j (p_j - 1) y^{p_j - 2} \\ =& \sum_{j=i+1}^{N} a_j p_j (p_j - 1) y^{p_j - 2} + \frac{p_i - 1}{y}(f_i'(y) - \sum_{j=0}^{i-1} a_j p_j y^{p_j - 1} - \sum_{j=i+1}^{N} a_j p_j y^{p_j - 1}) + \sum_{j=1}^{i-1} a_j p_j (p_j - 1) y^{p_j - 2} \\ =& \frac{(p_i - 1) f_i'(y)}{y} - \sum_{j=0}^{i-1} a_j p_j (p_i - p_j) y^{p_j - 2} + \sum_{j=i+1}^{N} a_j p_j (p_j - p_i) y^{p_j - 2}. \end{aligned}$$

Therefore,

$$\begin{aligned}\left[\frac{f_i'(y)}{y^{p_i-1}}\right]' &= \frac{f_i''(y)}{y^{p_i-1}} - \frac{(p_i-1)f_i'(y)}{y^{p_i}} \\ &= y^{-1}\left[\sum_{j=i+1}^{N} a_j p_j(p_j-p_i)y^{p_j-p_i} - \sum_{j=0}^{i-1} a_j p_j(p_i-p_j)y^{p_j-p_i}\right] \\ &:= y^{-1} g_i(y).\end{aligned} \tag{A.53}$$

Since $a_0 > 0$, $g_i(y)$ is strictly increasing in y, and $g_i(0) = -\infty$. Therefore, $f_i'(y)/y^{p_i-1}$ is whether

1. strictly decreasing, or
2. strictly decreasing in $[0, \bar{y}]$ for some $\bar{y} \in [0, w_h\ell_{\max}]$ and then strictly increasing in $[\bar{y}, w_h\ell_{\max}]$.

We take $y = 0$ in $f_i'(y)$ and obtain $f_i'(0) = a_0 > 0$. By the continuity of $f_i'(y)$, there exists $\delta > 0$, such that when $y \in [0, \delta]$, $f'(y) > 0$.

For case 1, 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 $y^* = y_0$.

For case 2, since $f_i'(y)/y^{p_i-1}$ first decreases and then increases, $f_i'(y)$ may have 0, 1 or 2 zeros in $[0, w_h\ell_{\max}]$. The schematic diagram of $f_i'(y)/y^{p_i-1}$ is shown in Figure 3. The case that $f_i'(y)$ has 0 or 1 zeros is discussed in case 1. When $f_i'(y)$ has two zeros $0 < y_1 < y_2 < w_h\ell_{\max}$, $f_i(y)$ is increasing in $[0, y_1]$, decreasing in $[y_1, y_2]$, and increasing again in $[y_2, w_h\ell_{\max}]$, and hence has two local maximum y_1 and $w_h\ell_{\max}$. When $f_i(y_1) = f_i(w_h\ell_{\max})$, f_i has two maximal point with one equals $w_h\ell_{\max}$; otherwise $f_i(y)$ has a unique maximal point.

When $a_{i+1} = \cdots = a_N = 0$, (A.53) yields that $f_i'(y)/y^{p_i-1}$ is strictly decreasing, and hence $f_i(y)$ has a unique maximal point. Therefore, we finish the proof. □

Proof (*Proof of Proposition A.3*) To simplify notation, we denote types slightly differently and let $\Theta = H = \cup_{i=1}^{N}\{h = (i, h_{-\eta}) | h_{-\eta} \in H_i\}$, and for any $h = (i, h_{-\eta}) \in \Theta$, $\eta_h = \eta_i$. Suppose that $0 < \eta_N < \cdots < \eta_1$. We define

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

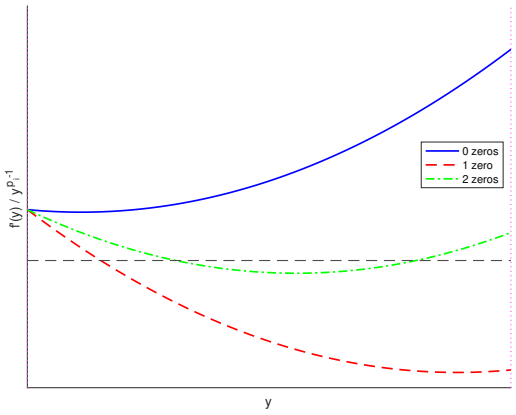


Figure 3: Schematic diagram of $\frac{f'(y)}{y^{p_i-1}}$.

Then the deterministic problem for (6.2) subject to (6.4) and (6.6) can be written

$$\begin{aligned}
&\max_{c\in C,\, y\in Y} \sum_{h\in H} \left(\log(c_h) - \tilde{\psi}_h y_h^{\frac{1}{\eta_h}+1}\right), \\
&\text{s.t.}\, \log(c_h) - \tilde{\psi}_h y_h^{\frac{1}{\eta_h}+1} \geq \log(c_{h'}) - \tilde{\psi}_h y_{h'}^{\frac{1}{\eta_h}+1}, \quad \forall h = (i, h_{-\eta}), h' = (j, h'_{-\eta}) \in H, \text{ s.t. } i \leq j, \\
&\sum_{h\in H} y_h \geq \sum_{h\in H} c_h;
\end{aligned} \tag{A.54}$$

and the Lagrangian function of problem (A.54) has the structure

$$\begin{aligned}
\mathcal{L}(c, y; \lambda, \mu) = &\sum_{(i,h_{-\eta})\in H} \left[-\left(\tilde{\psi}_h + \sum_{j\geq i} \sum_{h'_{-\eta}\in H_j} \tilde{\psi}_h \lambda_{h,h'} - \sum_{j=i} \sum_{h'_{-\eta}\in H_j} \tilde{\psi}_{h'} \lambda_{h',h}\right) y_h^{\frac{1}{\eta_i}+1} + \right. \\
&\left. \sum_{j=1}^{i-1} \left(\sum_{h'_{-\eta}\in H_j} \tilde{\psi}_{h'} \lambda_{h',h}\right) y_h^{\frac{1}{\eta_j}+1} + \mu\omega_h y_h \right] + \mathcal{L}^c(c; \lambda, \mu) \\
:= &\sum_{h\in H} \left[-\mathcal{B}_{h,i}(\lambda) y_h^{\frac{1}{\eta_i}+1} + \sum_{j=1}^{i-1} \mathcal{B}_{h,j}(\lambda) y_h^{\frac{1}{\eta_j}+1} + \mu\omega_h y_h \right] + \mathcal{L}^c(c; \lambda, \mu),
\end{aligned} \tag{A.55}$$

where $h' = (j, h'_{-\eta})$, $\mathcal{B}_{h,j}(\lambda) \geq 0$, $\forall 1 \leq j \leq i-1$, and $\mathcal{L}^c(c; \lambda, \mu)$ is independent on $y \in Y$. It is then straightforward to verify that the minimizer of

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

satisfies $\mu^* > 0$ (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). For any $h \in H$, according to Lemma A.1 with $p_i = \frac{1}{\eta_i} + 1$, $p_j = \frac{1}{\eta_j} + 1 (\forall 1 \leq j \leq i-1)$, $a_i = \mathcal{B}_{h,i}(\lambda^*)$, $a_j = \mathcal{B}_{h,j}(\lambda^*) (\forall 1 \leq$

$j \leq i-1), a_j = 0 (\forall i+1 \leq j \leq N)$ and $a_0 = \mu^* w_h$, we know that the maximal y_h^* is unique. The uniqueness of maximal c_h^* is straightforward since $\mathcal{L}^c(c; \lambda^*, \mu^*)$ has the form $\sum_h X_h \log c_h - Y_h c_h$, where $Y_h > 0$ for any h. Hence according to Result 1 of Corollary 3.1, the optimal solution to the planner problem (6.2) subject to (6.4) and (6.6) does not involve lotteries. □

We introduce an additional Lemma needed for the proof of Proposition 6.1.

Lemma A.2 *There exists $\bar{a} > 0$, s.t. when $a_{\max} \geq \bar{a}$, the solution to the planner problem (6.2) subject to (6.4) and (6.6) remains invariant with respect to $a_{\max}$.*

Proof We denote $v_2(a_{\max})$ as the maximal objective value of the planner problem (6.2) subject to (6.4) and (6.6). According to Proposition A.3, for any $a_{\max} > 0$, the solution to the planner problem (6.2) subject to (6.4) and (6.6) does no involve lottery in labor supply. We arbitrarily take a $\underline{a} > 0$. Since

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

and

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

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

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

Therefore, for any $a_{\max} > \underline{a}$, the solution $(c(a_{\max}), y(a_{\max}))$ to the planner problem (6.2) subject to (6.4) and (6.6) satisfies $y_h(a_{\max}) \leq \bar{a}(h) w_h$. We take $\bar{a} = \max_{h \in H} \bar{a}(h)$. Then for any $a_{\max} > \bar{a}$, the solution to the planner problem (6.2) subject to (6.4) and (6.3), remains invariant to the solution to the problem with $a_{\max} = \bar{a}$. Hence we finish the proof. □

Proof of Proposition 6.1

We denote $v_1(a_{\max})$ as the maximal objective value of the planner problem (6.2) subject to (6.4) and (6.3); and $v_2(a_{\max})$ as the maximal objective value of the planner problem (6.2) subject to (6.4) and (6.6). Since (6.6) involves less constraints than (6.3), we know $v_1(a_{\max}) \leq v_2(a_{\max})$ for any $a_{\max} > 0$. According to Lemma A.2, we have $v_2(a_{\max}) = v_2(\bar{a})$

for any $a_{\max} \geq \bar{a}$. Hence, to prove this lemma, it then suffices to show that

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

Step 1. The Slater's condition implies that, there exists $M > 0$, s.t. for any $a_{\max} > \bar{a}$, the Lagrangian multipliers for the planner problem (6.2) subject to (6.4) and (6.3) satisfies $\max\{\max_{h,\, h'\in H} \lambda^*_{h,h'},\, \mu^*\} < M$.

To be precise, for $a_{\max} = \bar{a}$, 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 (6.4) and (6.3) strictly. Hence, there exists $M > 0$, such that when $\max\{\max_{h,h'\in H} \lambda_{h,h'}, \mu\} \geq M$, we have $L(\bar{p}; \lambda, \mu) > v_2(\bar{a})$, where L denotes the Lagrangian function of the planner problem (6.2) subject to (6.4) and (6.3). Since $\max_p L(p; \lambda, v) \geq L(\bar{p}; \lambda, v) > v_2(\bar{a}) \geq v_1(a_{\max}) = \max_p L(p; \lambda^*, \mu^*)$, 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 $a(\epsilon) > \bar{a}$, such that when $a_{\max} > a(\epsilon)$, the Lagrangian multipliers of the planner problem (6.2) subject to (6.4) and (6.3) 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 (6.2) subject to (6.4) and (6.3) with $a_{\max} > 0$, 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 (6.2) subject to (6.4) and (6.3),

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

where $h' = (k, h'_{-\eta})$, $\mathcal{L}^{c,y_{-h}}(c, y_{-h}; \lambda^*, \mu^*)$ is independent on y_h, $\mathcal{B}_{h,j}(\lambda^*) \geq \epsilon \tilde{\psi}_{h'^*}$ for some

$h'^* = (j, h'^*_{-\eta})$, 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

$$\begin{aligned} &\max_{c \in C, y \in Y} \mathcal{L}(c, y; \lambda^*, \mu^*) \\ \geq &\max_{y_h \in [0, w_h a_{\max}]} \left[\sum_{k=1}^{N} \mathcal{B}_{h,k}(\lambda^*) y_h^{\frac{1}{\eta_k}+1} + \mu^* \omega_h y_h \right] - M \\ \geq &\max_{y_h \in [0, w_h a_{\max}]} \left[\epsilon \tilde{\psi}_{h'^*} y_h^{\frac{1}{\eta_j}+1} - M y_h^{\frac{1}{\eta_i}+1} \right] - M \to \infty \text{ as } y_h \to \infty. \end{aligned} \tag{A.56}$$

On the other hand,

$$\max_{c \in C, y \in Y} \mathcal{L}(c, y; \lambda^*, \mu^*) = v_1(a_{\max}) \leq v_2(\bar{a}). \tag{A.57}$$

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

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

Therefore, when $a_{\max} > \bar{a}(\epsilon)$, 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 (6.2) subject to (6.4) and (6.6) with $a_{\max} \geq \bar{a}$. It is straightforward to check that, for any $\epsilon > 0$, there exists $\delta > 0$, such that when $\max_{(h,h') \in \mathcal{G}} \lambda^*_{h,h'} \leq \delta$,

$$\mathcal{L}(\bar{c}, \bar{y}; \lambda^*, \mu^*) \geq L(\bar{c}, \bar{y}; \lambda^{**}, \mu^{**}) - \epsilon \geq v_2(\bar{a}) - \epsilon,$$

where

$$\lambda^{**}_{h,h'} = \begin{cases} \lambda^*_{h,h'}, & (h, h') \notin \mathcal{G}; \\ 0, & (h, h') \in \mathcal{G}. \end{cases}$$

According to Step 2, there exists $\bar{a}(\delta) > 0$, such that when $a_{\max} \geq \bar{a}(\delta)$, the Lagrangian multipliers of the planner problem (6.2) subject to (6.4) and (6.3) satisfy $\max_{(h,h') \in \mathcal{G}} \lambda^*_{h,h'} \leq \delta$. Hence

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

implying that

$$(v_2(\bar{a}) - v_1(a_{\max}))_+ \leq \epsilon.$$

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

A.10 Proof of Proposition 6.2

To prove the proposition we first need the following result. It shows that for every $a_{\max} > 0$, whenever the optimal solution of some agent h involves lottery, the lottery requires the agent to earn income $a_{\max} w_h$ with some positive probability π_h.

Proposition A.4 *We consider the planner problem (6.2) subject to (6.4) and (6.3) with the consumption set $C = \times_{h\in H}[\epsilon, c_{\max}]$, the output set $Y = \times_{h\in H}[0, a_{max} w_h]$, and the utility function defined by (6.5). Assume that the Slater's condition holds, then the optimal solution to the planner problem does not involve lotteries in consumption. If the optimal solution requires lottery for agent θ, then the lottery requires this agent to earn income $a_{\max} w_h$ with some positive probability π_h.*

Proof Similarly as in (A.55), we have the Lagrangian as

$$\mathcal{L}(c, y; \lambda^*, \mu^*) = \sum_{h\in H} \Bigg[\underbrace{\sum_{j=i+1}^{N} \mathcal{B}_{h,j}(\lambda^*) y_h^{\frac{1}{\eta_j}+1} - \mathcal{B}_{h,i}(\lambda^*) y_h^{\frac{1}{\eta_i}+1} + \sum_{j=1}^{i-1} \mathcal{B}_{h,j}(\lambda^*) y_h^{\frac{1}{\eta_j}+1} + \mu^* w_h y_h}_{I} \Bigg] + \mathcal{L}^c(c; \lambda^*, \mu^*),$$

where $\mathcal{B}_{h,j}(\lambda^*) \geq 0 (\forall j \neq i)$ and $a_0 = \mu^* w_h > 0$. According to Lemma A.1, if I has multiple maximizers $y_h^{*,1} < y_h^{*,2}$, then $y_h^{*,2} = a_{\max} w_h$. Hence according to Result 1 of Corollary 3.1, we can conclude this proposition. □

To prove the proposition, it then suffices to show that the optimal solution for agents with $\eta \neq 1/8$ requires lotteries. To show this, we will need the following lemma.

Lemma A.3 *Assume that $(\bar{c}, \bar{y})$ is the solution to the planner problem (6.2) subject to (6.4) and (6.6) as $a_{\max} \to \infty$, as defined in Lemma A.2.*[15] *When there exists h, h' satisfying $\eta_h < \eta_{h'}$ such that*

$$u_h(\bar{c}_h, \bar{y}_h) - u_h(\bar{c}_{h'}, \bar{y}_{h'}) < 0,$$

then there exists $\bar{a}$, s.t. when $a_{\max} \geq \bar{a}$, the optimal solution for agent h' in the planner problem (6.2) subject to (6.4) and (6.3) requires lottery.

Proof We prove by contradiction. If not, then there exists h, h' satisfying $\eta_h < \eta_{h'}$, and $a_{\max}^n \to \infty$, s.t.

$$u_h(\bar{c}_h, \bar{y}_h) - u_h(\bar{c}_{h'}, \bar{y}_{h'}) < 0,$$

[15] According to Proposition A.3, the optimal solution to the problem with partial incentive constraints is always deterministic, and hence unique, implying the uniqueness of $(\bar{c}, \bar{y})$.

and the optimal solution with $a_{\max} = a^n_{\max}$ for agent h' does not involve lottery. Assume that the Lagrangian multipliers for $a_{\max} = a^n_{\max}$ is (λ^n, μ^n). According to Step1 and Step2 of the proof for Proposition 6.1, (λ^n, μ^n) is bounded and $\lambda^n_{h,h'} \to 0$ for $h, h' \in H$ s.t. $\eta_h < \eta_{h'}$.

Therefore, up to a subsequence, $(\lambda^n, \mu^n) \to (\lambda^*, \mu^*)$ s.t. $\lambda^*_{h,h'} = 0$ for $h, h' \in H$ satisfying $\eta_h \leq \eta_{h'}$. Then for any $a_{\max} > 0$, we have

$$\begin{aligned}\max_{c\in C, y\in \times_{h\in H}[0, a_{\max} w_h]} \mathcal{L}(c, y; \lambda^*, \mu^*) &= \lim_{n\to\infty} \max_{c\in C, y\in \times_{h\in H}[0, a_{\max} w_h]} \mathcal{L}(c, y; \lambda^n, \mu^n) \\ &\leq \lim_{n\to\infty} \max_{c\in C, y\in \times_{h\in H}[0, a^n_{\max} w_h]} \mathcal{L}(c, y; \lambda^n, \mu^n) \\ &= \lim_{n\to\infty} v_2(a^n_{\max}),\end{aligned}$$

where $v_2(a^n_{\max})$ is defined as in Lemma A.2. On the other hand, we have

$$\begin{aligned}\max_{c\in C, y\in \times_{h\in H}[0, a_{\max} w_h]} \mathcal{L}(c, y; \lambda^*, \mu^*) \geq & \inf_{\lambda,\mu} \max_{c\in C, y\in \times_{h\in H}[0, a_{\max} w_h]} \mathcal{L}(c, y; \lambda^*, \mu^*) \\ = & v_2(a_{\max}).\end{aligned}$$

Therefore, according to Lemma A.2 there exists $\bar{a}_1 > 0$, s.t. when $a_{\max} > \bar{a}_1$,

$$\max_{c\in C, y\in \times_{h\in H}[0, a_{\max} w_h]} \mathcal{L}(c, y; \lambda^*, \mu^*) = v_2(\bar{a}_1) = \lim_{n\to\infty} v_2(a^n_{\max}).$$

Hence λ^*, μ^* is the Lagrangian multipliers for the planner problem (6.2) subject to (6.4) and (6.3), and

$$\arg \max_{c\in C, y\in \mathbb{R}^{|H|}_+} \mathcal{L}(c, y; \lambda^*, \mu^*) = (\bar{c}, \bar{y}).$$

We define $y^n_{h'}$ as the optimal income for agent h' with $a_{\max} = a^n_{\max}$. Since the total welfare is $v_2(\bar{a}_1)$ when n is sufficiently large, $y^n_{h'}$ is bounded by some $\bar{y}_1$. By continuity, $y^n_{h'} \to \bar{y}_{h'}$. Assume that c^n_h is the optimal consumption, y^n_h is in the support of the optimal solution, then by continuity we have $\lim_{n\to\infty} c^h_n = \bar{c}^h$, $\liminf_{n\to\infty} y^n_h \geq \bar{y}_h$. It is then directly to verify that the incentive constraint for h and h' does not hold when $a_{\max} = a^n_{\max}$ as $n \to \infty$. □

The proposition is finally obtained by the fact that the solution to the planner problem (6.2) subject to (6.4) and (6.3) violates all the constraints that are not included.

B Worst case complexity bounds (online only)

B.1 Lagrangian iteration

Proposition 3.1 provides bounds on the total number of iterations needed to obtain a desired accuracy. 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

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

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:

$$O\left(|A||\hat{C}|(m+\ell)\right). \tag{B.2}$$

Combining (3.17) and (B.2), we have the theorem for the overall computational complexity for finding an ϵ-optimal lottery solution as follows:

Theorem B.1 *We take all the assumptions in Theorem 3.5. 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*

$$\begin{cases} O\left(\left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}} (|A||\hat{C}|(m+\ell)^{1+\frac{1}{1-\rho}})\right), & \rho<1; \\ O\left(e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{m+\ell}}{\epsilon}} (|A||\hat{C}|(m+\ell))\right), & \rho=1. \end{cases} \tag{B.3}$$

Proof The overall complexity (B.3) can be directly obtained by multiplying (3.17) and (B.2). □

B.1.1 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, as explained in Section 4, in many economic applications, the subproblem (B.1) exhibits specific structural properties that allow for more efficient solution methods. For example, in the moral hazard problem with a utility function that is separable in action a and consumption c, the subproblem(B.1) satisfies the following properties

$$\begin{aligned} f(a,c) &= u_0(a) + \sum_{r=1}^{d} u_r(a) w_r(c_r), \quad \forall a \in A,\ c = (c_1, \cdots, c_d) \in C \subset \mathbb{R}^d \\ h_j(a,c) &= v_{j,0}(a) + \sum_{r=1}^{d} v_{j,r}(a) w_r(c_r), \quad \forall j \in \{1, \cdots, \ell\},\ a \in A,\ c = (c_1, \cdots, c_d) \in C \subset \mathbb{R}^d, \end{aligned} \tag{B.4}$$

where $C = \times_{r=1}^{d} [c_{r,\min}, c_{r,\max}]$, for some strictly increasing and strictly concave functions $w_r(c_r)$[16]. The following Lemma implies that, with the separable property (B.4), and the assumption that the resource constraints are strictly increasing and weakly convex with respect 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 (B.1), further enhancing the algorithm's efficiency.

Lemma B.1 *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 (B.4) holds and that the resource constraints $g_i(a,c)$ are strictly increasing and weakly convex with respect to each c_r. For any $a \in A$, we consider the following system:*

$$\begin{cases} \mathcal{A}(a,\gamma,r) \frac{\partial}{\partial c_r} w_r(c_r) = \sum_{i=1}^{m} \lambda_i \frac{\partial}{\partial c_r} g_i(a,c), & \mathcal{A}(a,r,\gamma) > 0; \\ c_r = c_{r,min}, & \mathcal{A}(a,r,\gamma) \leq 0, \end{cases} \tag{B.5}$$

where

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

If (B.5) admits a solution $c(a) \in C$ for any $a \in A$, then the solution (a^*, c^*) to the subproblem (B.1) should satisfy $c^* = c(a^*)$.

Proof With the property (B.4), the subproblem (B.1) can be written as

$$(a^*, c^*) \in \arg \max_{a \in A, c \in \hat{C}} \mathcal{L}(a, c; \lambda, \gamma) := \mathcal{A}_0(a,\gamma,0) + \sum_{r=1}^{d} \mathcal{A}(a,\gamma,r) w_r(c_r) - \sum_{i=1}^{m} \lambda_i g_i(a,c). \tag{B.6}$$

[16]This property holds because each consumption c_r appears in f and h in a single, specifically concave form. See Section 5 for more details. The case there is more specific, as $w_1 = \cdots = w_{|Q|}$.

We define the function on c as

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

Then (B.6) yields that c^* is the maximizer of $\mathcal{F}$, given a^*, λ, and γ. It is direct to show that

$$\frac{\partial \mathcal{F}(c; a^*, \lambda, \gamma)}{\partial c_r} = \frac{\partial \mathcal{L}(a^*, c; \lambda, \gamma)}{\partial c_r} = \mathcal{A}(a^*, \gamma, r) \frac{\partial}{\partial c_r} w_r(c_r) - \sum_{i=1}^{m} \lambda_i \frac{\partial}{\partial c_r} g_i(a^*, c). \tag{B.7}$$

When $\mathcal{A}(a^*, \gamma, r) \leq 0$, the right-hand side of (B.7) 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

$$\sum_{r \in \{A(a^*,\gamma,r)>0\}} \mathcal{A}(a^*, \gamma, r) w_r(c_r) - \sum_{i=1}^{m} \lambda_i g_i(a^*, \hat{c}, \tilde{c}^*), \tag{B.8}$$

□

which is a strictly concave function in $\hat{c}$. Therefore, the following first order condition to (B.8),

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

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 (B.7).

If all assumptions in Lemma B.1 hold, then (B.5) can be applied for solving the subproblem (B.1). To be precise, the subproblem (B.1) can be solved by the following three-step procedure:

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

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

$$\begin{cases} \mathcal{A}(a, \gamma, r) \frac{\partial}{\partial c_r} w_r(c_r(a)) = \sum_{i=1}^{m} \lambda_i \frac{\partial}{\partial c_r} g_i(a, c(a)), & \mathcal{A}(a, r, \gamma) > 0; \\ c_r = c_{r,\min}, & \mathcal{A}(a, r, \gamma) \leq 0. \end{cases} \tag{B.9}$$

3. Determining optimal $a \in A$ for the subproblem (B.1) with $c = c(a)$.

If additionally each g_i is linear with respective to c[17], 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 (B.9) can be solved by

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

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

$$O\left(\underbrace{d\ell|A|}_{\text{STEP1}} + \underbrace{dm|A|}_{\text{STEP2}} + \underbrace{d(m+\ell)|A|}_{\text{STEP3}}\right) \sim O\left(d(m+\ell)|A|\right), \tag{B.10}$$

where d is the dimension of the consumption space C.

We combine (B.10) and proposition 3.1 to obtain the complexity estimate for the first-order approach as following:

Theorem B.2 *We take all the assumptions in Theorem 3.5. 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. 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 $0 < \rho \leq 1$. Then for $\epsilon > 0$, the overall computational complexity for finding an ϵ-optimal lottery solution is*

$$\begin{cases} O\left(\left(\frac{\frac{M}{\rho}+\bar{\Lambda}}{\epsilon}\right)^{\frac{2}{1-\rho}} d|A|(m+\ell)^{1+\frac{1}{1-\rho}}\right), & \rho < 1; \\ O\left(e^{\frac{(\frac{M}{\rho}+\bar{\Lambda})\sqrt{m+\ell}}{\epsilon}} d|A|(m+\ell)\right), & \rho = 1. \end{cases} \tag{B.11}$$

where d is dimension of the consumption space C.

Proof The overall complexity (B.11) can be directly obtained by multiplying (3.17) and (B.10). □

Remark B.1 *By using the first-order approach, we can replace $|\hat{C}|$ in Theorem B.1 by d in the complexity estimate.*

[17]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 5 for details. In general, equation (B.9) can be solved using standard methods, such as Newton's method. In this context, the complexity of solving the nonlinear equation (B.9) 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 B.2, where $|\hat{C}|$ is replaced by an expression in terms of d. For simplicity, we omit a rigorous analysis of these cases.

B.2 Comparison to Existing Methods

In the literature, the most commonly used method for solving the lottery problem (3.2) 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 (3.2) into a finite-dimensional problem:

$$\begin{aligned} &\max_{\hat{x}\in\mathcal{P}(A\times\hat{C})} \sum_{a\in A}\sum_{c\in\hat{C}} f(a,c)\hat{x}(a,c), \\ \textbf{s.t.}\ &\sum_{a\in A}\sum_{c\in\hat{C}} g_i(a,c)\hat{x}(a,c) \le 0 \ (i\in\{1,\cdots,m\}), \\ &\sum_{c\in\hat{C}} h_j(a,c)\hat{x}(a,c) \le 0 \ (j\in\{1,\cdots,l\},\ a\in A). \end{aligned} \tag{B.12}$$

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.[18]

Comparing the complexity of our approach in Theorem B.1 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 Proposition 3.1, 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.

We note that for a problem that is 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,

[18] 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.

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 $|\hat{C}| = \prod_{i=1}^{d} |\hat{C}_i| \sim \left(\frac{1}{\epsilon}\right)^{\frac{d}{2}}$. [19] 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 (B.2), 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.

[19] 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 $|\{(2k+1)h : -N-1 \leq k \leq N\}^d| \sim \left(\frac{1}{\epsilon}\right)^{\frac{d}{2}}$.

B.2.1 Decomposabiliy in linear programming

For a decomposable problem, the linear programming formulation (B.12) can be equivalently rewritten as

$$\begin{aligned}
&\max_{\left(\hat{x}_k \in \mathcal{P}(A\times \hat{C}_k)\right)_{1\le k\le d}} \sum_{a\in A}\sum_{k=1}^{d}\sum_{c_k\in\hat{C}_k} f_k(a,c_k)\hat{x}_k(a,c_k),\\
\textbf{s.t.}\ &\sum_{a\in A}\sum_{k=1}^{d}\sum_{c_k\in\hat{C}_k} g_{i,k}(a,c_k)\hat{x}_k(a,c_k)\le 0\ (i\in\{1,\cdots,m\}),\\
&\sum_{k=1}^{d}\sum_{c_k\in\hat{C}_k} h_{j,k}(a,c_k)\hat{x}_k(a,c_k)\le 0\ (j\in\{1,\cdots,\ell\},\ a\in A),\\
&\sum_{c_1\in\hat{C}_1}\hat{x}_1(a,c_1)=\cdots=\sum_{c_d\in\hat{C}_d}\hat{x}_d(a,c_d):=\pi(a),(a\in A).
\end{aligned} \tag{B.13}$$

where the joint distribution $\hat{x}(a,c)$ is recovered as $\hat{x}(a,c)=\pi(a)\prod_{k=1}^{d}\hat{x}_k(c_k|a)$. A concrete illustration of a similar reformulation is given in Section 5.

C Production Complementarities

We now introduce non-linear production and complementarities between different types of workers into our example from Section 6. Each productivity type $\omega\in\Omega$ now corresponds to a different occupation of the worker. There is a representative firm that takes the labor supply of different occupations $\bar{l}=(\bar{l}_\omega)_{\omega\in\Omega}$ as given and produces the consumption good according to a strictly concave production function $f(\bar{l})$. Wages differ across occupations $(w_\omega)_{\omega\in\Omega}$ and are assumed to be exogenously given (they might be disconnected from the marginal product of labor due to the imperfectly competitive labor market, or, if they equal to the marginal product, the planner does not take this dependence into account when setting taxes and hence determining labor supply in different sectors). It is assumed that they satisfy $\sum_{\omega\in\Omega} w_\omega\bar{l}_\omega\le f(\bar{l})$. An agent of type $\theta=(\eta,\omega)$ who supplies a_θ units of labor receives an income $y_\theta=a_\theta w_{\omega_\theta}$ that is observable and taxable. The type is not observable, and the planner cannot tax different occupations differently – the latter assumption is motivated by existing labor-income tax schemes in most OECD countries, Boerma et al. (2022) allow for occupation-specific taxation and discuss institutional details. For simplicity, we assume that the planner can tax profits 100 percent and ignore the important question of optimal capital taxation.

The planner chooses a lottery over $(c_\theta, y_\theta)_{\theta\in\Theta}$ to solve

$$\max_{x\in\mathcal{P}(C\times Y)} \int_{(c,y)\in C\times Y} \left(\sum_{\theta\in\Theta} u_\theta(c_\theta, y_\theta)\right) x(dc, dy), \tag{C.1}$$

$$s.t. \int_{(c,y)\in C\times Y} (u_\theta(c_\theta, y_\theta) - u_\theta(c_{\theta'}, y_{\theta'}))x(dc, dy) \geq 0 \text{ for all } \theta, \theta' \in \Theta. \tag{C.2}$$

$$\int_{(c,y,\bar{l})\in C\times Y\times \bar{L}} \left(\sum_{\theta} c_\theta - f(\bar{l})\right) x(dc, dy, d\bar{l}) \leq 0 \tag{C.3}$$

$$\int_{(c,y,\bar{l})\in C\times Y\times \bar{L}} \left(\bar{l}_{\bar{\omega}} - \sum_{\theta:\omega_\theta=\bar{\omega}} \frac{y_\theta}{w_\theta}\right) x(dc, dy, d\bar{l}) \leq 0 \text{ for all } \bar{\omega} \in \Omega, \tag{C.4}$$

where $\bar{L}$ is a compact subset of $\mathbb{R}^{|\Omega|}$. For non-concave production functions this formulation is obviously a bit strange since the lotteries cannot be interpreted in an economic meaningful way unless one also assumes that there is a continuum of identical producers. However, since we focus on a strictly concave production function it is never optimal to randomize over $\bar{l}$ and our algorithm can be applied. In fact, the property of partial concavity ensures that the problem remains decomposable. The reason for this is that the Lagrangian to the deterministic problem can be written as

$$\begin{aligned}\mathcal{L}(c, y, \bar{l}, \lambda, \gamma, \nu) \quad = \quad & \sum_{\theta\in\Theta} u_\theta(c_\theta, y_\theta) + \sum_{\theta,\theta'\in\Theta} \lambda_{\theta,\theta'}(u_\theta(c_\theta, y_\theta) - u_\theta(c_{\theta'}, y_{\theta'})) - \gamma(\sum_{\theta\in\Theta} c_\theta - f(\bar{l})) \\ & - \sum_{\omega\in\Omega} \nu_\omega(\bar{l}_\omega - \sum_{\theta:\omega_\theta=\omega} \frac{y_\theta}{w_\theta})\end{aligned}$$

Since f is concave the Lagrangian is concave in $\bar{l}$ and we must have that at the optimal solution

$$\gamma \frac{\partial f(\bar{l})}{\partial \bar{l}_\omega} = \nu_\omega \text{ for all } \omega \in \Omega$$

and hence we can write $\bar{l}(\nu, \gamma)$ and $f(\bar{l}) = \tilde{f}(\frac{\nu}{\gamma})$ for some function $\tilde{f}(.)$ that often can be solved for analytically.[20] We can then write

$$\mathcal{L}(c, y, \tilde{\ell}, \lambda, \gamma, \nu) = \sum_{\theta\in\Theta} \mathcal{L}_\theta(c_\theta, y_\theta, \lambda, \gamma, \nu),$$

[20] One typical example is that f is the Cobb-Douglas production function.

with

$$\mathcal{L}_\theta(c_\theta, y_\theta, \lambda, \gamma, \nu) = (1 + \sum_{\theta' \in \Theta, \theta' \neq \theta} \lambda_{\theta,\theta'}) u_\theta(c_\theta, y_\theta) -$$

$$\sum_{\theta' \in \Theta, \theta' \neq \theta} \lambda_{\theta',\theta} u_{\theta'}(c_\theta, y_\theta) - \gamma \left(c_\theta - \frac{\tilde{f}(\nu/\gamma)}{|\Theta|} \right) - \sum_{\omega \in \Omega} \nu_{\omega_\theta} \left(\frac{\bar{l}_{\omega_\theta}(\nu, \gamma)}{|\{\theta : \omega_\theta = \omega\}|} - \frac{y_\theta}{w_\theta} \right).$$

The problem becomes decomposable after taking advantage of partial concavity.

In linear programming, on the other hand, the problem can only be partially decomposed. While one can decompose the probabilities across agents, one cannot separate out the probabilities over different $\bar{l}_\omega$. The problem therefore quickly becomes intractable as $|\Omega|$ grows beyond 2.

A numerical example To illustrate our method, we take $|\Omega| = 5$ and vary the number of different η types from 2 to 7. Let $u_\theta(.)$ be defined as in (6.5) and let the production function be

$$f(\bar{l}) = \bar{Z} \Pi_{\omega \in \Omega} \left(\bar{l}_\omega \right)^\omega . \tag{C.5}$$

Suppose $\omega \in \{0.05, 0.1, 0.15, 0.2, 0.25\}$, and consider 4 different specifications for η_θ: $\eta^1 = (1, 1/7)$, $\eta^2 = (1, 1/4, /1/7)$, $\eta^3 = (1, 1/3, 1/5, 1/7)$ and $\eta^4 = (1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7)$. Without loss of generality, we choose $\bar{Z} = 3$. Each agent can work at most $a_{\max} = 1$. In this example, we specify the equilibrium wages $w^* = \{w^*_\omega\}_{\omega \in \Omega}$ to be at their competitive level but assume that the planner takes wages as exogenously given[21].

For the Lagrangian Iteration Algorithm, the initial Lagrangian multipliers are chosen by $\lambda = 0$, $\gamma = 1$, $\nu = 1$. For each optimization problem, the learning rate is

$$\mu^k = \begin{cases} \frac{2}{(1+|\Theta|^2)^{0.51}}, & 1 \leq k < \frac{K}{2}; \\ \frac{2}{(1+|\Theta|^2)^{0.8}}, & \frac{K}{2} \leq k \leq K. \end{cases}$$

Table 4 shows how running times increase in the number of types. Following our theoretical analysis above, we obviously need to increase the number of iterations as the number

[21] To compute the wages, we use a simple fixed point iteration algorithm as following for finding the equilibrium wages. We use Algorithm 1 with less iterations to solve the optimization problem (6.2) to (C.4), and obtain an optimal lottery $x^*(w^t) \in \mathcal{P}(C \times Y \times \tilde{L})$. Then we update w^t as

$$w^{t+1} = \frac{\partial f}{\partial \bar{l}}(\bar{l}^*(w^t)).$$

We iterate over t until $\|w^t - w^{t-1}\|$ lies below some $tol > 0$ In this example, we choose the initial $w^0 = \{0.3, 0.4, 0.5, 0.6, 0.7\}$, $tol = 10^{-4}$, and iteration rounds $K = 8 \times 10^4 \times (|\Theta|/5)^2$.

$\lvert\Theta\rvert$	Iterations ($\times 10^5$)	CPU time (s)	Max. constraint violation
10	6	14	1.8×10^{-5}
15	14.5	38	2.6×10^{-5}
20	28	93	3.3×10^{-4}
35	73.5	477	1.4×10^{-4}

Table 4: Computational Performance for Different number of types $|\Theta|$.

of constraints increases. For the case of 35 types (i.e. 7 different η-types and 5 occupations), to keep the maximal constraint violation relatively low we need to increase the number of iterations quadratically in the number of agents. The table confirms that in many cases the bottleneck of our algorithm is not only the global optimization problem in Step 1, but also the number of iterations needed in the sub-gradient descent method.

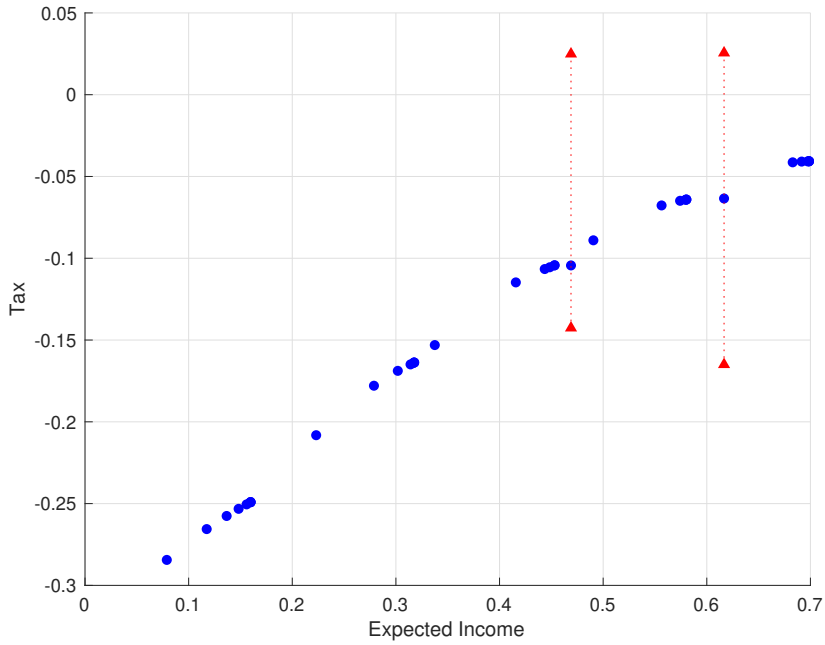


Figure 4: Relation between tax and expected income. The support of lotteries is indicated by red triangles.

For the case of five occupations and seven preference types, Figure 4 plots the relation between expected income and tax across all 35 types. Two types are offered lotteries, and their taxes are a bit lower on average. The tax schedule in this example is remarkably close to linear[22] but its slope is much smaller than one (as it would be in the first best).

[22]Note that taxes are always negative. Since capital is taxed fully, the revenue guarantees negative income taxes.