A General Solution to the Quasi Linear Screening Problem

Guillaume Carlier[*] Xavier Dupuis[†] Jean-Charles Rochet[‡] John Thanassoulis[§]

September 29, 2023

Abstract

We provide an algorithm for solving multidimensional screening problems which are intractable analytically. The algorithm is a primal-dual algorithm which alternates between optimising the primal problem of the surplus extracted by the principal and the dual problem of the optimal assignment to deliver to the agents for a given surplus. We illustrate the algorithm by solving (i) the generic monopolist price discrimination problem and (ii) an optimal tax problem covering income and savings taxes when citizens differ in multiple dimensions.

1 Introduction

We provide an algorithm that solves any principal agent problem of the following form:

$$\max_{y,U} \sum_{i=1}^{N} f_i[S_i(y_i) - \lambda U_i] \tag{1}$$

subject to incentive compatibility conditions for all (i, j)

$$U_i - U_j \geq \Lambda_{ij}(y_j) := b_i(y_j) - b_j(y_j) \tag{2}$$

and individual rationality conditions for all i

$$U_i \geq 0. \tag{3}$$

[*]Ceremade, Université Paris-Dauphine, PSL Research University

[†]Institut de Mathématiques de Bourgogne, UMR 5584 CNRS, Université de Bourgogne, F-21000 Dijon, France

[‡]Toulouse School of Economics

[§]Warwick Business School, University of Warwick, CEPR, and UK Competition and Markets Authority

The interpretation of this problem is the following. A principal wants to assign different goods, or bundles of goods, to a population of agents who can be of different types $i = 1, .., N$. Types are not publicly observable. The mass of agents of type i is denoted $f_i > 0$. The function $S_i(y_i)$ measures the surplus generated for the principal by agent i when she receives assignment y_i, U_i is agent i's utility, and $\lambda > 0$ is a weight applied by the principal to the agents' utility. We give below two examples that will be used for illustrating how our algorithm works.

This is a standard screening problem, but a general solution does not exist in large part due to the fact that in a multidimensional context there is no natural ranking of the agents. This means the binding constraints are not *ex ante* identifiable and frustrates local analyses. This contrasts with the case of one dimensional types when the Single Crossing Condition holds. In this one dimensional setting the local downward incentive compatibility constraints are binding for all types, allowing a global problem to be converted into a series of local optimisations which can be solved easily (Mussa and Rosen (1978)). In the multi-dimensional setting the direction of the binding incentive compatibility constraints can be subject to many variations which in turn can lead to rich new features in the optimal solution to the canonical principal agent problem.

We therefore move beyond the one-dimensional analysis which has dominated research thus far. Thanks to the characterization of implementability that was given in Rochet (1987)[1], we can reformulate our screening problem into a max min problem. Our algorithm is an extension of the powerful algorithm proposed by Chambolle and Pock (2011) for solving such max min problems.

2 Motivations

Although there are many economic problems to which our algorithm can be applied, we will focus on two particular applications: the multidimensional version of the multiproduct monopolist problem studied by Rochet and Choné (1998), and the joint taxation of labour and savings income.

2.1 Multiproduct Monopolist

Rochet and Choné (1998) have studied the multidimensional extension of the multiproduct monopolist problem of Mussa and Rosen (1978). They established that pooling, i.e. different types receiving the same assignment (this is also called bunching), is a general feature of optimal screening in multiple dimensions. This has important consequences, and makes analytical solutions hard, except in special cases. When the distribution of types is discrete, the informational rent of agents (see below for a formal definition) is a

[1]Rochet (2023) surveys the literature on multidimensional screening that has followed that article.

non-differentiable function of assignments in the pooling region. When the distribution of types is continuous, the Euler-Lagrange equation that characterizes solutions of control problems is not satisfied in the pooling region. Our algorithm overcomes these difficulties by using two ingredients: the use of proximal functions for avoiding non-differentiability problems, and a primal-dual approach to take into account the different expressions taken by the informational rent when different incentive compatibility constraints are binding. Our algorithm determines these endogenous pooling groups and binding incentive compatibility constraints as it proceeds through the optimisation process. Using our algorithm we are able to solve the discrete version of this monopolist pricing problem for an arbitrary number of types.

As a specific example, consider a monopolist selling a durable good (say a car) that can be designed in several specifications represented by a vector of characteristics $y \in \mathbb{R}^d$. The dimension d represents the number of different features upon which the product can be differentiated, and each individual component (y_i) in the vector y can be thought of as the quality level of that feature offered. The cost of producing one unit of the good with overall characteristics y is a convex function $C(y)$. The utility of buying this good for agent $i = 1, .., N$ is quasi-linear:

$$U_i = \theta_i . y_i - p_i,$$

where the vector $\theta_i \in \mathbb{R}^d$ represents the willingness to pay for a unit of quality across all the available dimensions of the good and (y_i, p_i) is the combination of characteristics and price that is designed for agents of type i. The firm wants to select the menu $(y_i, p_i), i = 1, ..., N$ of characteristics and prices that maximizes its profit

$$\sum_i f_i[p_i - C(y_i)] = \sum_i f_i \left[\theta_i . y_i - C(y_i) - U_i\right], \tag{4}$$

under the constraint that, for all i, agents of type i buy the product y_i at price p_i. This constraint can be decomposed into two conditions on individual utilities $U_i = \theta_i \cdot y_i - p_i$:

$$U_i \geq \theta_i \cdot y_j - p_j = U_j + (\theta_i - \theta_j) \cdot y_j,$$

for all i, j, expressing that agent i prefers the combination (y_i, p_i) that was designed for him to any combination (y_j, p_j) designed for another agent j, and

$$U_i \geq 0,$$

expressing the participation constraint of agent i. This is a particular case of our general problem if we take

$$S_i(y_i) = \theta_i.y_i - C(y_i), b_i(y_i) = \theta_i \cdot y_i, \ \lambda = 1.$$

Rochet and Choné (1998) consider a continuous version of this model and show that the solution necessarily involves some degree of pooling. Of course, pooling may already appear in dimension 1, but it can be ruled out by assuming a monotone likelihood property of the distribution of types. Armstrong (1996) shows that a simple form of pooling is generic in multidimensional screening problems: a positive measure of consumers is typically excluded. Rochet and Choné (1998) extend this result by showing that a second form of pooling is also typical of multidimensional problems: low type consumers are often offered a reduced set of products. Only high types are offered a wide set of products that are tailored to their taste differences. By contrast, low quality products are less differentiated and each of them is bought by a positive measure of consumers. Our algorithm allows us to characterise precisely the product range a monopolist would choose to have so as to optimally manage the allocation of its products to its clients.

2.2 Joint taxation of saving and labour incomes

The question of the optimal mix between labor and capital taxes is very old. However, influential books by Piketty (2014) or Saez and Zucman (2019) have recently restarted the debate. These books recommend a more comprehensive taxation of inheritance and savings. Such taxes, it is argued, would reduce inequality and would provide additional fiscal resources without distorting too much the employment and consumption choices of individuals and the investment decisions of firms.

However, most of the academic literature on optimal taxation, starting from the influential papers of Chamley (1986) and Judd (1985), argue on the contrary that capital (and by extension all financial activities) should not be taxed in the steady state of a standard economy when optimal income taxation is possible. The modern approach to optimal taxation, initiated by Mirrlees (1971), also recommends that capital should not be taxed at all: see in particular Atkinson and Stiglitz (1976) and Diamond and Mirrlees (1971).

But these results are not valid when heterogeneity between individuals is multidimensional. Labour income is one tool which can be used to screen the population, but with multiple dimensions more tools can be valuable. For example, Saez (2002) shows that taxing capital income is optimal when more productive people have a higher propensity to save – the tax on capital alongside labour allowing for better screening outcomes. Mirrlees (1976) himself was well aware of the fact that most of his results relied on the restrictive assumption that labor productivity is the only source of unobservable heterogeneity among individuals, an assumption that he adopted for pure tractability reasons.

Many papers have tried to extend the Mirrlees (1971) model to several dimensions of heterogeneity, but technical difficulties have hindered progress: we know very little about multidimensional screening problems in general. Explicit results have been obtained for particular distributions of types: Wilson (1993) for nonlinear pricing, Armstrong (1996) for multi-good monopoly pricing, Rochet and Choné (1998) for the hedonic version of the same model, Rochet and Thanassoulis (2019) for dynamic versions of the screening problem, and Rochet (2009) for the regulation of firms with different marginal and fixed costs. Their results show that the solution pattern may differ markedly from that of the unidimensional case. However, these results are only valid for very peculiar parametrizations of heterogeneity.

The recent literature on multidimensional screening has explored no less than five different approaches to overcome these difficulties.

The first approach is to make assumptions on preferences and technology such that the multidimensional problem reduces to a one-dimensional screening problem. This is what Kleven et al. (2009) have done in their analysis of the optimal taxation of couples. Similarly, Choné and Laroque (2010) consider an optimal taxation problem with two dimensions of heterogeneity (labour productivity and the opportunity cost of labour) but they simplify the incentive problem by assuming that individual labour supply only depends on a unidimensional combination of the two parameters. Beaudry et al. (2009) use similar simplifications in their analysis of employment subsidies.

A second approach is to assume that the government only has one instrument, e.g. taxing total income, independently of its composition. Rothschild and Scheuer (2013, 2016) study the general equilibrium impact of taxation in a multisector economy where agents have different (unobservable) productivities in the different sectors. Similarly, by adapting the techniques introduced by Rochet and Stole (2002) for non-linear pricing, Jacquet et al. (2013) study the taxation of labor income when individuals differ in two dimensions: skill and cost of participating in the labour market, whilst the government can only tax labour income.

A third approach is the variational approach of Golosov et al. (2014) for continuous distributions of types. Roughly speaking, the idea is to compute the (Gateaux) differential of social welfare with respect to the different policy instruments available to the government (here the different taxes). This allows one to analyze the impact of (infinitesimal) tax reforms.[2] This amounts to a calculus of variations problem constrained by a partial differential equation. The problem is that the approach is only valid when there is no bunching, i.e. different types always get different allocations. However, bunching is very frequent in multidimensional screening problems.

[2]Similarly, Renes and Zoutman (2017) adopt a mechanism design approach and solve the relaxed problem (first order approach) where the second order conditions of individuals' optimization programs are neglected.

A fourth approach is purely numerical. Tarkiainen and Tuomala (1999, 2007) consider an income tax model where individuals differ by their productivity and their cost of labor participation. They develop numerical methods that allow them to solve this problem for particular specifications of preferences and type distributions. Similarly, Judd et al. (2017) use a non-standard optimization algorithm to solve particular specifications of highly complex taxation problems with 5 dimensions of heterogeneity. However, none of these papers provide a convergence theorem. As acknowledged by Tarkianen and Tuomala, these numerical approaches seem to work for special parametrizations but there is no guarantee that the algorithms would also converge for other specifications. A more promising approach is developed in Boerma et al.(2022), who use Legendre transforms to transform the screening problem into a linear program. They are able to numerically solve a large scale multidimensional tax problem that is calibrated to the US economy.

Finally, the fifth approach is only illustrative: it focuses on 2×2 models with two dimensions of heterogeneity and two possible values for each parameter. Using the methodology introduced by Armstrong and Rochet (1999), such models are fully solvable. For example, Cremer et al. (2001, 2003) show that taxing capital and luxury goods can be optimal in a 2×2 model where individuals differ in their initial endowments as well as their labour productivities. Similarly, Boadway et al. (2002) show that negative marginal tax rates can be optimal in a 2×2 model where individuals differ by their preferences for leisure as well as their labor productivity. The problem is that these models are purely illustrative: the need to restrict to 2×2 types means they cannot be calibrated to real data.

We examine a simple extension of the Mirrlees optimal tax problem to the case where agents have two dimensions of heterogeneity: their initial endowments e_i and their disutilities of working x_i. Agents consume at two dates $t = 1, 2$ and have quasi linear preferences:

$$V_i = u(C_i^1) + C_i^2 = u(e_i - s_i) + Rs_i + (w - x_i)l_i - T_i,$$

where $y_i = (s_i, l_i)$ denote the (observable) decisions of agent i : savings s_i and labor supply $0 \leq l_i \leq 1$. T_i denotes the total tax paid by agent i. R is the return on savings and w the unit wage. Both are exogenous and uniform across agents. The principal seeks the tax system that maximizes a weighted sum of a Rawlsian objective and utilitarian welfare:

$$W = \lambda \min_j V_j + (1 - \lambda) \sum_i f_i V_i, \tag{5}$$

with $0 \leq \lambda \leq 1$, under the constraint that tax revenue is sufficient to finance public expenditures of G, which is taken as exogenously given. Note that no participation constraints are required in this context of obligatory tax. However, the problem can be put into our general form by defining incremental utilities by $U_i = V_i - \min_j(V_j)$, which implies by definition that $U_i \geq 0$ for all i. Moreover, if we set $S_i(y_i) = u(e_i - s_i) + Rs_i + (w - x_i)l_i$, and using that $\sum_i f_i T_i = G$, the objective of the principal can be rewritten, up to a constant:

$$W = \sum_i f_i[S_i - \lambda U_i]. \tag{6}$$

It is easy to see that this program is a particular case of the general problem when

$$S_i(y_i) = u(e_i - s_i) + Rs_i + (w - x_i)l_i, \tag{7}$$

and

$$b_i(y) = u(e_i - s) - lx_i.$$

An interesting economic question, which a solution to (5) would allow us to address, is whether the taxation of savings should be independent of labour income. In particular, should savings be taxed more heavily for employed or unemployed people? The involved trade-off can be understood by looking at a model with two dimensional types and two possible values for each dimension: taxpayers may have a low or high cost of participating in the labour force, and a low or high initial endowment. Given the preferences, a separable tax schedule would imply that savings and labour supply decisions are independent: labour supply only depends on the first dimension of heterogeneity (the personal disutility of working and so participating in the labour force), while savings only depend on the second dimension (initial endowments). We will see in Section 7 that in general, for some parameter values, greater societal welfare is generated if the planner conditions the tax on savings on the citizen's workforce status. This can be seen by direct computations in the 2×2 model, but it very hard to assess in a calibrated model that reproduces data more accurately. Our algorithm allows us to solve such models without having to assume unrealistic distributions of types.

3 Problem Preliminaries

For the sake of simplicity, we only discuss the case where the assignment y can be any vector in $\mathbb{R}^d$.[3] We assume that the functions b_i and S_i are smooth for all i. There are N types, with weights in the population $f_i > 0$. For an easy representation of the constraints, we define the linear operator D: $\mathbb{R}^N \to \mathbb{R}^{N\times(N-1)}$ by

$$(Du)_{ij} := u_i - u_j,$$

where $\mathbb{R}^{N\times(N-1)}$ denotes the set of $N \times N$ matrices with zero entries on the diagonal.

[3]The extension to the case where the y_i's are constrained to lie in a certain box of $\mathbb{R}^d$, possibly dependent on i, is straightforward.

The scalar product[4] of Du with a vector v in $\mathbb{R}^{N\times(N-1)}$ is thus

$$(Du)\cdot v = \sum_{i,j}(u_i - u_j)v_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N}(v_{ij} - v_{ji})u_i.$$

The adjoint D^* of this operator is the linear mapping from $\mathbb{R}^{N\times(N-1)}$ to $\mathbb{R}^N$ defined by

$$(D^*v)\cdot u = (Du)\cdot v,\ \forall (u,v) \in \mathbb{R}^N \times \mathbb{R}^{N\times(N-1)}.$$

Hence, it is given, for all i, by:

$$(D^*v)_i := \sum_{j=1}^{N}(v_{ij} - v_{ji}).$$

We shall also use the more concise notation Λ for the map appearing in the right-hand side of (2). For $y = (y_1, ..., y_N) \in \mathbb{R}^{d\times N}$ and all (i,j):

$$\Lambda_{ij}(y) = b_i(y_j) - b_j(y_j).$$

Setting $S(y) := \sum_i f_i S_i(y_i)$, the screening problem (1)-(2)-(3) can be rewritten as

$$\max_{y,U}\{(S(y) - \lambda f\cdot U)\ : DU \geq \Lambda(y),\, U \geq 0\}$$

where the notation $A \geq B$ for matrices (respectively vectors) A and B means that $A - B$ has all nonnegative entries (respectively coordinates). Existence of a solution and first-order optimality conditions are given by:

Proposition 1 *Assuming that for every i*

$$S_i(y_i) \to -\infty \text{ as } |y_i| \to \infty, \tag{8}$$

then (1)-(2)-(3) admits at least a solution. Let $(\overline{y}, \overline{U})$ be such a solution, and let A be the set of binding IC constraints, i.e. the set of (i,j)'s for which $\overline{U}_i - \overline{U}_j = \Lambda_{ij}(\overline{y}_j)$. If either Λ is linear or the IC constraints are qualified at $(\overline{y}, \overline{U})$ i.e. there exist $\hat{y}, \hat{u}$ such that

$$\hat{u}_i - \hat{u}_j > \nabla\Lambda_{ij}(\overline{y}_j)\hat{y}_j,\ \forall (i,j) \in A. \tag{9}$$

then there exist multipliers $\overline{\mu}_i \geq 0$ (for the IR constraints (3)), multipliers $\overline{v}_{ij} \geq 0$ (for the IC constraints (2)) such that:

$$\lambda f = \overline{\mu} + D^*\overline{v},\ f_j \nabla S_j(\overline{y}_j) = \sum_i \overline{v}_{ij}\nabla\Lambda_{ij}(\overline{y}_j),\ \forall j \tag{10}$$

[4]This is the standard scalar product for vectors, and not the matrix product.

together with the complementary slackness conditions:

$$\overline{\mu}_i \overline{U}_i = 0, \ \overline{v}_{ij}(D\overline{U} - \Lambda(\overline{y}))_{ij} = 0. \tag{11}$$

Proof. Condition (8) and the constraint $U \geq 0$, guarantee that one can reduce the maximization problem to a compact set for the y_i's so that $\Lambda_{ij}(y_j)$ can be bounded a priori. One can also choose U such that $\min_i U_i = 0$, but the IC constraint imposes that $U_j \leq \min_i U_i + \max_{kl} -\Lambda_{kl}(y_l)$ so that the U_i's can also be chosen to remain in a bounded set, we are therefore left to maximizing a continuous function over a compact set and the existence claim follows. The necessity of the first-order optimality conditions (10)-(11) for nonnegative multipliers $\overline{\mu}$ and $\overline{v}$ follows from the Karush-Kuhn-Tucker Theorem: see e.g. Carlier (2022) Proposition 4.9 for the case of affine constraints and Theorem 4.5 for nonlinear constraints satisfying the qualification condition (9). ■

Let us briefly comment the assumptions in the previous proposition. First observe that (8) is automatically satisfied in the Multiproduct Monopolist problem as soon as the cost C is superlinear i.e. $C(y)/|y| \to +\infty$ as $|y| \to +\infty$. As for the qualification condition, Lemma 1 in the Appendix gives a simple case where the condition (9) is easily obtained. Note also that when Λ is linear and the S_i's are concave, the first-order conditions (10)-(11) are sufficient conditions.

4 Feasibility, informational rent and duality

The aim of this section is to reformulate the generic principal-agent model (1)-(2)-(3) in terms of the assignment vector y only.

4.1 Feasibility

Let us first introduce a definition:

Definition 1 *Let $\Lambda \in \mathbb{R}^{N\times(N-1)}$ be a matrix with zero diagonal entries. We will say that Λ is feasible whenever there exists $U \in \mathbb{R}^N$ such that $DU \geq \Lambda$.*

Since DU is unchanged when adding a constant to U, one sees that feasibility of Λ is the same as the existence of a $U \in \mathbb{R}^N$ such that $U \geq 0$ and $DU \geq \Lambda$. Since $\Lambda_{ii} = 0$ the feasibility condition $DU \geq \Lambda$ can be rewritten as requiring the existence of U such that $U = T_\Lambda(U)$ where T_Λ is the self-map of $\mathbb{R}^N$ given by

$$T_\Lambda(U)_i := \max_j \{U_j + \Lambda_{ij}\}. \tag{12}$$

This characterisation of feasibility, results from the application of Theorem 1 in Rochet (1987) to our context. Formally we have

Proposition 2 *Let $\Lambda \in \mathbb{R}^{N\times(N-1)}$ (with $\Lambda_{ii} = 0$). The following are equivalent:*

1. *Λ is feasible,*

2. *Whenever $i_0, \dots i_L, i_{L+1} = i_0$ is a cycle in the set of indices in $\{1, \dots, N\}$, one has*

$$\sum_{k=0}^{L} \Lambda_{i_k i_{k+1}} \leq 0. \tag{13}$$

3. *Defining T_Λ by (12), the sequence starting from $u^0 = 0$ and inductively defined by $u^{n+1} = T_\Lambda(u^n)$ for $n \geq 1$ converges (monotonically and in at most $N-1$ steps) to the smallest nonnegative fixed point of T_Λ.*

Proof. Suppose Λ is feasible, let U be such that $U_i - U_j \geq \Lambda_{ij}$ for every i, j. If $i_0, \dots i_L, i_{L+1} = i_0$ is a cycle, then

$$\sum_{k=0}^{L} \Lambda_{i_k i_{k+1}} \leq \sum_{k=0}^{L} (U_{i_k} - U_{i_{k+1}}) = 0$$

so that 1. $\Rightarrow$ 2.

Assume that Λ satisfies (13) and define u^n by $u^0 = 0$ and $u^{n+1} = T_\Lambda(u^n)$ for $n \geq 1$. We will show that $u^N = u^{N-1}$. Since $\Lambda_{ii} = 0$, we have $0 \leq u^n \leq u^{n+1}$ in particular $u^N \geq u^{N-1}$. One easily checks inductively that

$$u_i^n = \max\Big\{ \sum_{k=0}^{n-1} \Lambda_{i_k i_{k+1}} \ : \ i_0 = i, \ i_1, \dots, i_n \in \{1, \dots, N\}^n \Big\}. \tag{14}$$

Therefore $u_i^N = \sum_{k=0}^{N-1} \Lambda_{i_k i_{k+1}}$ for some $i_1, \dots, i_N \in \{1, \dots, N\}^N$ and $i_0 = i$. Necessarily $i_k = i_{l+1}$ for some pair of indices k and l such that $0 \leq k \leq l \leq N-1$. Hence, thanks to (13) we have

$$\sum_{j=k}^{l} \Lambda_{i_j i_{j+1}} \leq 0$$

so that

$$u_i^N \leq \sum_{j \in \{0,..,N-1\} \setminus \{k,\dots,l\}} \Lambda_{i_j i_{j+1}} \leq u_i^{N-1}$$

where the last inequality follows from (14) and the fact that $i_k = i_{l+1}$. This shows that $u^n = u^{N-1}$ for $n \geq N-1$ so that u^n converges to a nonnegative fixed point of T_Λ in at most $N-1$ steps. If u is a nonnegative fixed point of T_Λ, monotonicity of T_Λ and an obvious induction argument show $u^n \leq u$ which implies that u^n converges to the smallest nonnegative fixed point of T_Λ. So we have 2. $\Rightarrow$ 3.

If 3. holds, there exists $u \geq 0$ such that $u = T_\Lambda(u)$, hence $u_i - u_j \geq \Lambda_{ij}$ i.e. Λ is feasible, and so 3. $\Rightarrow$ 1. ■

Note that part 3 of Proposition 2 gives a constructive way to solve $DU \geq \Lambda(y)$, $U \geq 0$ when $\Lambda(y)$ is feasible. The minimality of the fixed point $T_\Lambda^{N-1}(0)$ also implies:

Corollary 1 *If Λ is feasible, the least nonnegative fixed-point of T_Λ, $u = T_\Lambda^{N-1}(0)$ is the unique solution of*

$$\min \Big\{ \sum_i f_i U_i \ : \ DU \geq \Lambda, \, U \geq 0 \Big\}$$

for every collection of positive weights $f_i > 0$.

A dual characterization of feasibility (upon which our algorithm will in part rely) is the following:

Lemma 1 *Let $\Lambda \in \mathbb{R}^{N \times N}$ (with $\Lambda_{ii} = 0$). Then Λ is feasible if and only if for every $v \in \mathbb{R}^{N \times N}$,*

$$\Big(v \geq 0 \text{ and } D^* v = 0 \Big) \Rightarrow v \cdot \Lambda \leq 0.$$

Proof. If Λ is feasible, there exists U such that $\Lambda \leq DU$. Hence if $v \geq 0$ and $D^* v = 0$ we have $v \cdot DU = 0 \geq v \cdot \Lambda$. Conversely, suppose that Λ is not feasible: it is impossible to find U and a matrix $M \geq 0$ such that $-\Lambda = D(-U) + M$. Geometrically this means that $-\Lambda \notin D(\mathbb{R}^N) + \mathbb{R}_+^{N \times N}$. The set $D(\mathbb{R}^N) + \mathbb{R}_+^{N \times N}$ is clearly convex, we claim that it is also closed. To show this, take a sequence $\mu^n \in \mathbb{R}_+^{N \times N}$, another sequence $u^n \in \mathbb{R}^N$ and assume that $\mu^n + Du^n$ converges. Since the sum of the entries of Du^n vanishes and μ^n is nonnegative, the convergence of the sum of the entries of μ^n implies that μ^n is bounded. Hence it has a convergent subsequence, which implies that Du^n also has a convergent subsequence. Since $D(\mathbb{R}^N)$ is closed, the limit of this subsequence of $\mu^n + Du^n$ belongs to $D(\mathbb{R}^N) + \mathbb{R}_+^{N \times N}$. We can therefore strictly separate $-\Lambda$ from $D(\mathbb{R}^N) + \mathbb{R}_+^{N \times N}$ i.e. find $v \in \mathbb{R}^{N \times N}$ and $\varepsilon > 0$ such that

$$-v \cdot \Lambda \leq -\varepsilon + v \cdot \mu + v \cdot Du, \ \forall (\mu, u) \in \mathbb{R}_+^{N \times N} \times \mathbb{R}^N$$

Suppose now that $v \cdot Du < 0$ for some $u \in \mathbb{R}^N$. Multiplying this vector u by a large positive constant gives a contradiction. Similarly if $v \cdot Du > 0$, multiplying u by a large negative constant gives a contradiction. Thus it must be that $v \cdot Du = 0$ for all $u \in \mathbb{R}^N$ i.e. $D^* v = 0$. By a similar reasoning on μ, the above condition implies $v \geq 0$ and also $v \cdot \Lambda \geq \varepsilon > 0$ which is the desired conclusion. ■

4.2 Informational rent and duality

For fixed assignment vector y, the informational rent $R(y)$ left to the agents is defined by the value of the sub-problem:

$$R(y) := \inf\Big\{ \sum_i f_i U_i \ : \ U_i \geq 0, \ U_i - U_j \geq \Lambda_{ij}(y_j) \Big\}. \tag{15}$$

The interpretation is that $R(y)$ is the minimum expected pay-off that must be left to the agents in order to implement the assignment y. We adopt the convention that $\inf \emptyset = +\infty$ so that $R(y) = +\infty$ whenever $\Lambda(y)$ is not feasible. The next proposition gives a dual expression for the informational rent, for which it is convenient to introduce the closed and convex (but unbounded) set

$$K := \{v \in \mathbb{R}^{N \times (N-1)} \ : \ v \geq 0, \ D^* v \leq \lambda f\}, \tag{16}$$

as well as its support function:

$$\sigma_K(\Lambda) := \sup\{v \cdot \Lambda \ : \ v \in K\}, \ \forall \Lambda \in \mathbb{R}^{N \times (N-1)}.$$

Proposition 3 *The informational rent $R(y)$ is the value of the dual problem:*

$$R(y) = \sup\Big\{ v \cdot \Lambda(y) = \sum_{i,j} v_{ij} \Lambda_{ij}(y_j) \ : \ \sum_j (v_{ij} - v_{ji}) \leq f_i, \ v_{ij} \geq 0 \Big\} \tag{17}$$

We thus have

$$\lambda R(y) = \sigma_K(\Lambda(y)). \tag{18}$$

Moreover whenever $\Lambda(y)$ is feasible, there exists $v \in K$ such that $\lambda R(y) = v \cdot \Lambda(y)$.

Proof. If $\Lambda(y)$ is not feasible, then $R(y) = +\infty$ and it follows from Lemma 1 that there is some $v_0 \geq 0$ such that $D^* v_0 = 0$ and $v_0 \cdot \Lambda(y) > 0$. Since for $t > 0$, $tv_0 \geq 0$ and $D^*(tv_0) = 0 \leq f$, we have

$$\sup\{v \cdot \Lambda(y), \ v \geq 0, D^* v \leq f\} \geq \sup_{t>0} tv_0 \cdot \Lambda(y) = +\infty = R(y).$$

Assume now that $\Lambda(y)$ is feasible, then the admissible set in the right-hand side of (15) is nonempty. We claim that the infimum in (15) is a minimum: if U^n is a minimizing sequence, it is nonnegative and $f \cdot U^n$ is bounded from above and since $f > 0$ this implies that U^n is bounded, and hence has a subsequence which converges to a solution of the minimization problem in (15). Now we can invoke the duality Theorem for linear programming (see e.g. Theorem 6.5 in Carlier (2022)): if the linear minimization problem in (15) admits a solution, so does its dual problem which is exactly the linear maximization problem in (17) and the values of both problems agree. ■

The informational rent is thus the composition of the support function σ_K of the feasible set K of the dual problem by the “mimicking” functions Λ_{ij} which represent the gain of agent i when he mimicks agent j. Note that σ_K only depends on the distribution of the agent types $\{f_i\}$, not on the economic fundamentals of the problem. Moreover $\sigma_K(\Lambda)$ is infinite iff there is a cycle on which the sum of the Λ is positive.

When the Single Crossing Condition holds, the binding IC constraints are always the local downward constraints (independently of the assignment) and the support function has a simple linear expression:

$$\sigma_K(\Lambda) = \lambda \sum_i (1 - F_i)\Lambda_{i+1,i},$$

where $F_i = \sum_{j<i} f_j$. However in the multidimensional case, the sup in the definition of the support function is not always attained for the same vector v when we consider different assignments $y = (y_1, ..., y_N) \in \mathbb{R}^{d\times N}$. For these assignments, the rent $R(y)$ must be written as the sup of two or more affine mappings and is therefore not differentiable.

4.3 Maxmin reformulation, optimality conditions

Using (18) and Proposition 3, we will establish that the initial screening problem (1)-(2)-(3) is equivalent to

$$\max_{y\in\mathbb{R}^{d\times N}} S(y) - \sigma_K(\Lambda(y)). \tag{19}$$

By definition of σ_K, this rewrites in maxmin form

$$\max_y \min_{v\in K} S(y) - v \cdot \Lambda(y). \tag{20}$$

Formally we have:

Proposition 4 *$(\overline{y}, \overline{U})$ solves (1)-(2)-(3) if and only if $\overline{y}$ solves (19) and*

$$\sigma_K(\Lambda(\overline{y})) = \lambda \sum_i f_i \overline{U}_i,\ D\overline{U} \geq \Lambda(\overline{y}),\ \overline{U} \geq 0.$$

Note also that one can recover the optimal $\overline{U}$ from an optimal $\overline{y}$ using Proposition 2. Indeed, if $\overline{y}$ solves (19) (so that $\Lambda(\overline{y})$ is feasible) and $\overline{U}$ is the smallest nonnegative fixed point of $T_{\Lambda(\overline{y})}$ (obtained as in Proposition 2) then $(\overline{y}, \overline{U})$ solves (1)-(2)-(3).

Now observe that the KKT conditions (10)-(11) imply that

$$D^*\overline{v} \leq \lambda f \text{and } \overline{v} \geq 0 \text{ i.e. } \overline{v} \in K$$

and

$$\lambda R(\overline{y}) \leq \lambda f \cdot \overline{U} = D^*\overline{v} \cdot \overline{U} = \overline{v} \cdot D\overline{U} = \overline{v} \cdot \Lambda(\overline{y}) \leq \sigma_K(\Lambda(\overline{y})) = \lambda R(\overline{y})$$

which, thanks to Proposition 3, yields

$$\sigma_K(\Lambda(\overline{y})) = \overline{v} \cdot \Lambda(\overline{y}).$$

We can therefore reformulate the necessary conditions (10)-(11) for the initial formulation (1)-(2)-(3) in terms of conditions in the variables y and v (multipliers for the IC constraints) instead of y and U:

Proposition 5 *Assume that $(\overline{y}, \overline{U})$ solves (1)-(2)-(3) and the IC constraints are qualified (see (9)) at $(\overline{y}, \overline{U})$. Then, there exists $\overline{v} \in K$ such that*

$$\sigma_K(\Lambda(\overline{y})) = \overline{v} \cdot \Lambda(\overline{y}), \tag{21}$$

and

$$f_j \nabla S_j(\overline{y}_j) = \sum_i \overline{v}_{ij} \nabla \Lambda_{ij}(\overline{y}_j). \tag{22}$$

In terms of sufficient conditions, we have:

Proposition 6 *Assume $(\overline{y}, \overline{v}) \in \mathbb{R}^{d\times N} \times K$ satisfy conditions (21)-(22) of Proposition 5 and that $\overline{U}$ is the smallest nonnegative fixed point of $T_{\Lambda(\overline{y})}$ (see Proposition 2). Then,*

1. *if $\overline{y}$ is a local (resp. global) maximizer of $y \mapsto S(y) - \overline{v} \cdot \Lambda(y)$, it is a local (resp. global) solution of (19) that is $(\overline{y}, \overline{U})$ is a local (global) solution of (1)-(2)-(3),*

2. *if*

$$\sum_j (f_j D^2 S(\overline{y}_j) - \sum_{ij} \overline{v}_{ij} D^2 \Lambda_{ij}(\overline{y}_j))(h_j, h_j) < 0$$

for every nonzero $h \in \mathbb{R}^{d\times N}$ such that there exist u_i such that

$$\nabla \Lambda_{ij}(\overline{y}_j) \cdot h_j = u_i - u_j \text{ when } (i,j) \in A \text{ and } \overline{v}_{ij} > 0$$

and

$$\nabla \Lambda_{ij}(\overline{y}_j) \cdot h_j \leq u_i - u_j \text{ when } (i,j) \in A \text{ and } \overline{v}_{ij} = 0$$

where A is the set of binding incentive compatibility constraints at $(\overline{y}, \overline{U})$, then $\overline{y}$ is a local solution of (19).

Proof. 1. Follows from $S(y) - \sigma_K(\Lambda(y)) \leq S(y) - \overline{v} \cdot \Lambda(y)$ with an equality for $y = \overline{y}$. 2. Is a (local) sufficient second-order condition which can be found in Chapter 3 ((Proposition 3.3.2 and its refined version in Exercise 3.3.7) of Bertsekas (2009)). ■

We conclude this section with the following remark:

Remark 2 *If S is concave differentiable and Λ is affine, $S - \sigma_K \circ \Lambda$ is concave so conditions (21)-(22) are in fact necessary and sufficient (global) optimality conditions for problem (19).*

5 The Algorithm

We now describe a proximal primal-dual algorithm to find a pair $(\overline{y}, \overline{v}) \in \mathbb{R}^{d\times N} \times K$ which solves the optimality conditions (21)-(22). We assume that S is concave and differentiable, and that Λ is smooth. We start with the case in which Λ is linear. Then the algorithm coincides with that proposed by Chambolle and Pock (2011).[5]

5.1 On proximal methods

Before describing the algorithm, let us recall some concepts from convex analysis, with the aim of giving some insights on proximal methods to the unfamiliar reader. Let $\varphi : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a convex, lower semi continuous function which is not identically $+\infty$. Given $x \in \mathbb{R}^m$, the subdifferential of φ at x, $\partial\varphi(x)$ is defined by

$$\partial\varphi(x) := \{p \in \mathbb{R}^m \ : \ \varphi(z) - \varphi(x) \geq p \cdot (z - x), \ \forall z \in \mathbb{R}^m\},$$

hence $\overline{x}$ minimizes φ if and only if $0 \in \partial\varphi(\overline{x})$ (which in the event φ is differentiable at $\overline{x}$ reduces to the standard first-order condition $0 = \nabla\varphi(\overline{x})$). The proximal operator of φ, was introduced in Moreau (1965) and is given by

$$\mathrm{prox}_\varphi(x) := \underset{z\in\mathbb{R}^m}{\mathrm{argmin}}\{\frac{1}{2}|z-x|^2 + \varphi(z)\}, \ \forall x \in \mathbb{R}^m.$$

The map $x \in \mathbb{R}^m \mapsto \mathrm{prox}_\varphi(x)$ is single-valued and one-Lipschitz (see Moreau (1965)) and

$$z = \mathrm{prox}_\varphi(x) \iff x \in z + \partial\varphi(z).$$

In particular

$$\overline{x} \text{ minimizes } \varphi \iff \overline{x} \in \overline{x} + \partial\varphi(\overline{x}) \iff \overline{x} = \mathrm{prox}_\varphi(\overline{x}).$$

So minimizing φ is equivalent to finding a fixed point of prox_φ (or $\mathrm{prox}_{\tau\varphi}$ with $\tau > 0$). This is the basic idea behind the proximal point algorithm

$$x_{k+1} = \mathrm{prox}_\varphi(x_k)$$

[5]It is worth recalling, especially in this special issue, that the algorithm of Chambolle and Pock (2011) is itself an extension of the classical Arrow-Hurwicz algorithm (Arrow et al. (1958)).

introduced in Martinet (1972). This algorithm is well-known to converge to a minimizer provided such a minimizer exists, see Rockafellar (1976). The proximal point algorithm has several appealing properties, not only because it allows for nonsmooth objective φ but also because it satisfies by construction the inequality $\varphi(x_{k+1}) + \frac{1}{2}|x_{k+1} - x_k|^2 \leq \varphi(x_k)$ which ensures φ decreases along its iterates. Of course, to use proximal methods in practice, one should be able to compute prox_φ efficiently. We end this paragraph by a simple example (which will be useful in our setting). If C is a nonempty closed and convex subset of $\mathbb{R}^m$, its characteristic function χ_C:

$$\chi_C(x) := \begin{cases} 0 \text{ if } x \in C \\ +\infty \text{ otherwise} \end{cases}$$

is lower semi continuous and convex. Its proximal operator prox_{χ_C} coincides with the projection (closest point map) proj_C onto C.

5.2 The linear case

If the utilities b_i are linear i.e of the form

$$b_i(y) = \theta_i \cdot y$$

where $\theta_i \in \mathbb{R}^d$ is the constant marginal utility of agent i (e.g. their willingness to pay for quality), Λ is the linear map

$$\Lambda_{ij}(y) := (\theta_i - \theta_j) \cdot y,$$

defined in (2).[6] The problem (19) is a concave maximization problem equivalent to finding $(\overline{y}, \overline{v}) \in \mathbb{R}^{d\times N} \times K$ which solve for the optimality conditions (21)-(22). For given step sizes $\tau > 0$ and $\sigma > 0$, the Chambolle-Pock algorithm proposes the following iterations:

$$y_{k+1} = \mathrm{prox}_{-\tau S}(y_k - \tau\Lambda^*(v_k)), \tag{23}$$

$$\widetilde{y}_{k+1} = 2y_{k+1} - y_k, \tag{24}$$

$$v_{k+1} = \mathrm{proj}_K(v_k + \sigma\Lambda(\tilde{y}_{k+1})). \tag{25}$$

Theorem 1 from Chambolle and Pock (2011) (also see He and Yuan (2012) for another proof) guarantees that the iterates above converge to a solution of the system (21)-(22) if $\tau > 0$, $\sigma > 0$ satisfy $\tau\sigma\|\Lambda\|^2 < 1$ where

$$\|\Lambda\|^2 := \sup_{y\neq 0} \frac{\|\Lambda(y)\|^2}{\|y\|^2} \leq \max_i \sum_j |\theta_i - \theta_j|^2.$$

[6]We denote by Λ^* its adjoint. This is defined so that if v is a $N \times N$ matrix, $\Lambda^* v \in \mathbb{R}^N$ is given by $(\Lambda^* v)_j = \sum_i (\theta_i - \theta_j) v_{ij}$.

We can therefore use this algorithm to solve the linear version of the general multidimensional principal-agent problem.

5.3 The General Case

When Λ is nonlinear, it is possible to use the linearization of primal updates[7] which leads to the algorithm proposed and analyzed by Valkonen (2014):

$$y_{k+1} = \text{prox}_{-\tau S}(y_k - \tau(\Lambda'(y_k))^* v_k), \tag{26}$$

$$\widetilde{y}_{k+1} = 2y_{k+1} - y_k, \tag{27}$$

$$v_{k+1} = \text{proj}_K(v_k + \sigma\Lambda(\tilde{y}_{k+1})). \tag{28}$$

Note that if these iterates converge to some $\overline{y}$, $\overline{v}$ one should have

$$\overline{y} = \text{prox}_{-\tau S}(\overline{y} - \tau(\Lambda'(\overline{y}))^* \overline{v}) \text{ i.e. } \nabla S(\overline{y}) = \Lambda'(\overline{y})^* \overline{v}$$

and

$$\overline{v} = \text{proj}_K(\overline{v} + \sigma\Lambda(\overline{y})) \text{ i.e. } \overline{v} \in K, \text{ and } \sigma_K(\Lambda(\overline{y})) = \overline{v} \cdot \Lambda(\overline{y})).$$

In other words, the pair $(\overline{y}, \overline{v})$ satisfies the first-order conditions (21)-(22) from Proposition 5. The (local) convergence analysis of the above algorithm to a solution of (21)-(22) is rather involved and can be found under various technical assumptions[8] in Valkonen (2014), see also the very recent references Valkonen (2023) and Gao and Zhang (2023) (where a shorter local convergence proof can be found, for a slightly different algorithm where the linearization of the nonlinear map Λ is used at the level of the updates (28) for v instead of the updates (26) for y).

Once again we have converted the multidimensional screening problem into a form which can be tackled using a recently developed algorithm. Of course, the steps of the above algorithm should be tractable enough to make the algorithm effective. We explain in the Appendix how the proximal steps (23) (or (26)) and (24) (or (27)) can be handled in practice.

6 Illustration 1: the Multiproduct Monopolist

In this section we apply our algorithm to the multiproduct monopolist problem described in Section 2.1. We suppose that the monopolist produces a product whose quality or type can be described by two characteristics $y \in \mathbb{R}^2$ with components y_1, y_2 capturing the

[7] $\Lambda'(y)$ denotes the derivative of Λ at y and $\Lambda'(y)^*$ denotes its adjoint.

[8] Among these assumptions is the requirement-as for the original Chambolle-Pock algorithm-that the steps τ and σ are small enough so that $\tau\sigma M^2 < 1$ with M a bound on the Lipschitz constant of $D\Lambda$.

quality of each characteristic. The marginal cost of producing a product with quality y is given by a quadratic function

$$C(y) := \frac{1}{2}\left(y_1^2 + y_2^2\right). \tag{29}$$

The monopolist serves a population of consumers who are characterised by a type vector $\theta \in \mathbb{R}^2$. The components of the type vector capture the willingness to pay for each characteristic. The total willingness to pay of type θ for a good with characteristics y is therefore the scalar product $\theta \cdot y$. The monopolist's first best would choose an assignment of a given product type for each consumer which maximised the surplus created for each client and extracted that surplus in the price charged. Hence the first best would be for a client of type θ to receive a product with characteristics $y = \theta$.

In our algorithm we suppose that consumers are uniformly distributed on an $N \times N$ grid supported on the square $[1, 2]^2$. In Figure 1 we solve two versions of this problem, a large version in which $N^2 = 2,500$ individual consumer types are modelled, and a smaller version with $N^2 = 25$ individual types which allows the binding incentive compatibility constraints to be studied.

Panel (a) of Figure 1 demonstrates how the monopolist optimally distorts her product range so as to maximise her profit. Recall that the first best has the product a type θ receives equal to her type. Under the asymmetric information constraint the '*no distortion at the top*' result which holds in the one-dimensional case is almost true for types who have the highest willingness to pay for at least one of the product characteristics, and holds exactly for the clients with the highest willingness to pay on both dimensions. It is the South West tail of the clients who find themselves with the most distorted assignments. These clients form a *Stingray's tail* which is a typical shape in these problems. The client of type $(1, 1)$ who has the lowest willingness to pay is optimally not served at all, and clients with low willingness to pay have their product significantly distorted towards lower quality on both dimensions. There is also bunching so that multiple types of low-valuation clients are served with the same product.

These two features – the Stingray's tail and bunching – are more clearly seen in the (less busy) 5×5 example in panels (b1) and (b2) of Figure 1. The Stingray's tail is displayed in Panel (b1) where the South West clients with the lowest valuations have their products distorted downwards. The bunching can be seen from Panel (b2) which depicts that in the South West corner of the support of client types the binding incentive compatibility constraints are not just in the local downward and leftward direction. Instead three, or sometimes four IC constraints become simultaneously binding (as evidenced by the multiple arrows depicting binding IC constraints). This is because these types are optimally served the same product characteristics, or are indifferent between two distorted products. The green spots in Panel (b2) further shows that the rents of the bottom six client types are fully extracted.

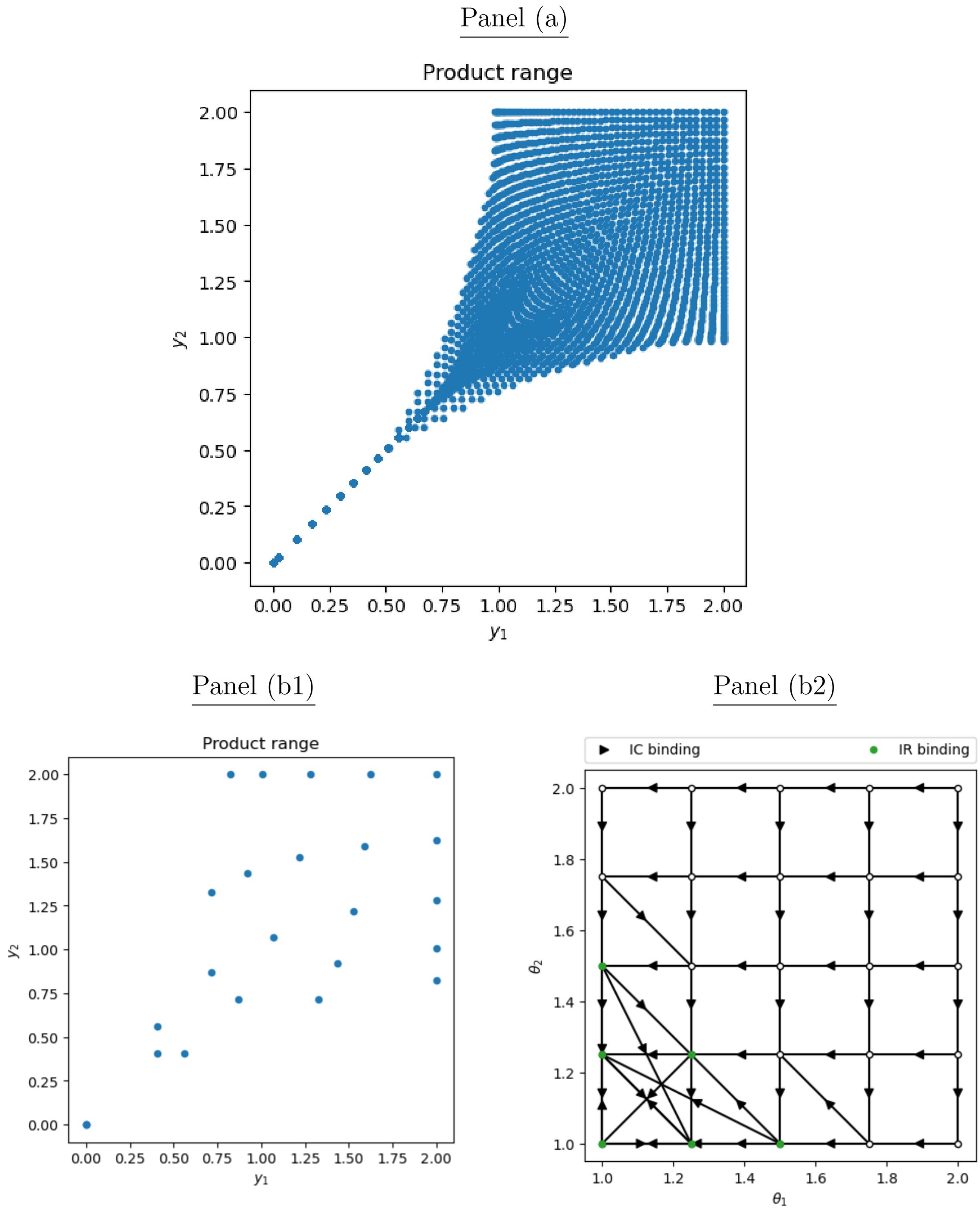

Figure 1: Solution to the Multidimensional Monopolist Problem
Notes: Panel (a) depicts the optimal product mix for a monopolist serving $2,500 = 50 \times 50$ consumers – the *Stingray's tail* is evident. Panel (b1) depicts the optimal product mix with $25 = 5 \times 5$ consumers. Panel (b2) depicts the binding incentive compatibility (IC) and individual rationality (IR) constraints for the 5×5 case at the optimal assignment of products to clients. Clients are supported on $[1, 2]^2$ and production costs are given in (29). Code for the simulation algorithm is available at `https://github.com/x-dupuis/screening-algo`.

7 Illustration 2: Joint Taxation of Labour and Savings Incomes

In this section we apply our algorithm to a second canonical problem in economics: optimal taxation when citizens differ in more than one dimension. This problem was introduced in section 2.2. To study this tax setting the social planner's objective function was defined in (6) and the surplus available to citizens was given in (7). In this formulation we allow citizens to be differentiated on two dimensions; the type of individual i is a couple (e_i, x_i). Endowments e_i are assumed to be distributed uniformly on a regular grid on $[1, e_{max}]$. Labour disutilities x_i are distributed uniformly on a regular grid on $[0, 1]$. We set the competitive wage at $w = 1$ so that the utility of full-time work is $1 - x_i > 0$. The utility of consumption at date 1 is

$$u(C_1) := \frac{1}{\eta}\left(1 - e^{-\eta C_1}\right),$$

with $\eta = 1$. Period one consumption is $e - s$ while period two consumption is the sum of investment returns Rs and labour income. The return on savings is R and the Rawlsian weight is λ. With these specifications, the first best allocation, obtained when types are publicly observable, is characterized by $l = 1$ for all (everyone participates in the labor force) as labour contributes positively to each citizen's surplus. Further, the first best allocation would result in identical consumption at date 1 such that : $C_1 = \ln\frac{1}{R}$ for all. Such an allocation, in the presence of full information, would be implemented by personalized lump-sum taxes that do not depend on the labour or savings decisions of the agents.

When x is the same for all agents, which only differ in e, the second best allocation can be implemented by a savings tax $T(s)$. The indirect utility of an agent of type e is denoted by $U^*(e) + (w - x)l$ where

$$U^*(e) = \max_s u(e - s) + Rs - T(s).$$

Note that the marginal tax rates are such that

$$T'(s(e)) = R - u'(e - s(e)), \tag{30}$$

where $s(e)$ denotes the savings of agent e in the second best allocation. The envelope theorem then implies

$$U^{*\prime}(e) = u'(e - s(e)) > 0. \tag{31}$$

The economic question we want to investigate is whether it is optimal to tax the savings of employed people at a higher or lower rate than unemployed people in the general case where both e and x are heterogenous and privately observable. Assuming for simplicity that l can only take the values 0 or 1, the principal will offer a menu of tax schedules $(T_l(s), l = 0, 1)$, giving rise to an indirect utility function

$$\max(U_0^*(e), U_1^*(e) + w - x).$$

The critical value of x above which an agent of type (e, x) decides not to work is thus

$$x^*(e) = w + U_1^*(e) - U_0^*(e). \tag{32}$$

This critical value is increasing in the citizen's endowment e if and only if $U_1^{*\prime}(e) > U_0^{*\prime}(e)$, which arises if and only if the marginal tax rates are lower for employed rather than unemployed agents. Using (30) and (31) we see that we can establish whether the tax rates on savings is affected by employment status by comparing the marginal utility of consumption across differing endowments.

Panel (a1) Panel (a2)

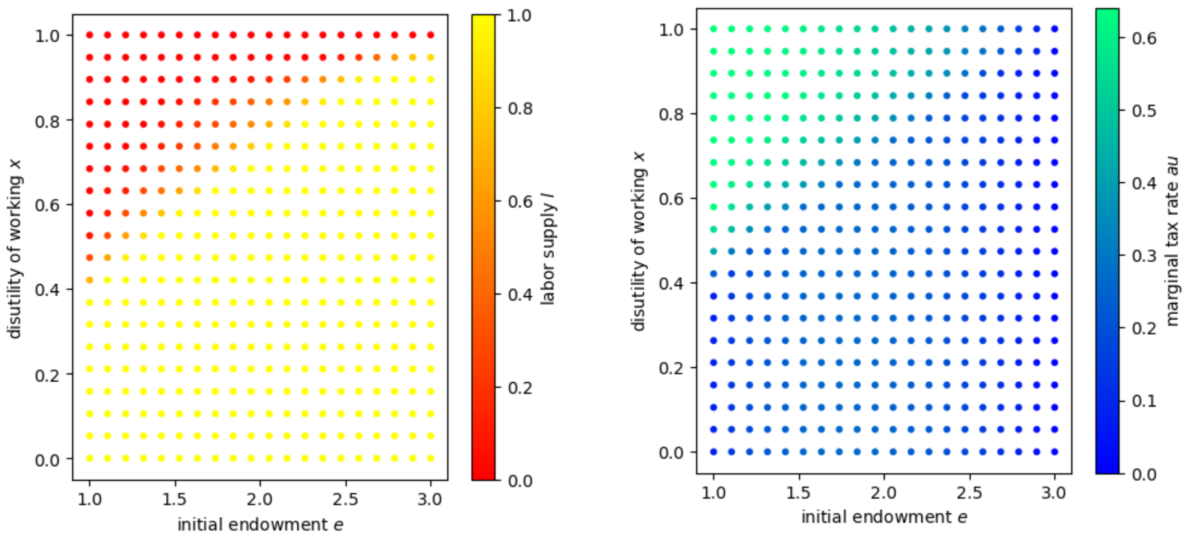

Figure 2: Solution to the multidimensional tax problem
Panel (a1) depicts the work decision citizens make in response to the optimal tax scheme. Panel (a2) depicts the marginal tax rate on savings given in (30). The simulation sets $\lambda = 1/2, R = 1, \eta = 1$, and citizens are modelled as taking one of $400 = 20 \times 20$ types. Code for the simulation algorithm is available at `https://github.com/x-dupuis/screening-algo`.

We solve for the optimal tax in a 20×20 example in Figure 2. Panel (a1) of Figure 2 depicts the citizen's labour decision when faced with the optimal tax scheme implemented by the social planner. We see that the labour force participation decision optimally depends on *both* the citizen's disutility of labour and their initial endowment. Panel (a1) shows us that the critical point at which citizens swap from not-working to working,

$x^*(e)$ is increasing in the citizen's endowment. This shows that optimality requires a tax on savings, and further that this tax depends upon the citizen's disutility of labour (and therefore on their labour force decision). This is confirmed in Panel (a2) of Figure 2. Panel (a2) plots the marginal tax rate with respect to saving which is given in (30). We see that the optimal savings tax depends upon the disutility of labour and so differs depending on the citizen's optimal labour force decision.

Our initial question was therefore whether the savings of employed people should be taxed differently to those of the unemployed. From Panel (a1) of Figure 2 the unemployed have the lowest endowments and the highest disutility of working. From Panel (a2) we see that the marginal tax rate is highest for these people. So in our simple formulation a form of no-distortion-at-the-top applies in which those with the largest endowment enjoy zero marginal savings taxes, but for those with smaller endowments savings taxes at the margin are larger, and the marginal tax rates are largest for those who do not supply labour so as to create the maximal incentive to work and not just consume from one's initial endowment.

We hope our algorithm will allow these results to be expanded and refined in much larger simulations making full use of the multidimensionality of the problem.

8 Conclusion

The objective of this paper is to make easily accessible to the research community an efficient algorithm which allows one to solve any discrete, quasi-linear screening problem of reasonable size. The two examples analyzed here are only illustrative and do not have any pretense to realism. However, our hope is that this algorithm will be used by specialists in the different topics that can be modelled as screening problems, including of course taxation and multiproduct design and pricing. The power of our algorithm makes it effective for large numbers of types, which allows one to approximate continuous distributions closely. We also hope to extend it, in subsequent research, to non quasi-linear environments.

Acknowledgements

J.-C. Rochet acknowledges financial support from ERC grant 101055239 DIPAMUTA. The IMB to which belongs X. Dupuis receives support from the EIPHI Graduate School (contract ANR-17-EURE-0002).

9 APPENDIX

9.1 A simple case where IC constraints are qualified

Lemma 2 *If y splits into $y = (x, z)$ and $b_i(y) = \varphi_i(x) + \theta_i z$ with $i \neq j \Rightarrow \theta_i \neq \theta_j$ then (9) holds at any admissible (y, U).*

Proof. Take $\hat{u}_i = \frac{1}{2}|\theta_i|^2$, $\hat{y}_i = (0, \theta_i)$ and note that $D\hat{u}_{ij} - \nabla\Lambda_{ij}(y_j)\hat{y}_j = \frac{1}{2}|\theta_i - \theta_j|^2$. ■

9.2 Feasibility of the proximal steps

Let us now explain how the proximal steps (23) (or (26)) and (24) (or (27)) can be handled in practice.

9.2.1 Updates for the primal variables y

The proximal steps (23) (or (26)) involve the proximal operator of the convex and smooth function $-\tau S$, that is given $y^0 \in \mathbb{R}^{d\times N}$ we have to solve

$$\sup_y S(y) - \frac{1}{2\tau}|y - y^0|^2$$

but $S(y) = \sum_i f_i S_i(y_i)$ is a separable function so that this proximal steps can be split into simple (strictly concave and smooth) optimization problems in dimension d only:

$$\sup_{y_i} f_i S_i(y_i) - \frac{1}{2\tau}(y_i - y_i^0)^2$$

which can be done by standard methods such as gradient ascent. Note that if S_i is quadratic (as in our Multiprodcut Monopolist illustration), this proximal step is in closed form.

9.2.2 Updates for the multipliers: projecting onto K

Recall that K is the closed convex set of $N \times N$ matrices,

$$K := \{v \in \mathbb{R}^{N\times N},\ v_{ii} = 0,\ v \geq 0,\ D^* v \leq \eta\} \tag{33}$$

where $\eta := \lambda f$. The projection onto K step could be a serious bottleneck for the algorithm if projecting onto K was costly. Our aim now is to explain how to project onto K efficiently. Given $w \in \mathbb{R}^{N\times N}$ we thus wish to solve

$$\inf_{v\in K} |v - w|^2 = \sum_{1\leq i,j\leq N} (v_{ij} - w_{ij})^2. \tag{34}$$

This is a quadratic problem with finitely many linear and inequality constraints. The unique solution v of (34) is characterized by the following KKT conditions: there exist $\mu \in \mathbb{R}_+^{N\times N}$ (multipliers for the nonnegativity constraints) and $\beta \in \mathbb{R}^N$ (multipliers for the constraints on D^*v) such that

$$v - w = \mu - D\beta,\ \mu \geq 0,\ \mu \cdot v = 0, \tag{35}$$

as well as

$$\beta \geq 0,\ D^*v \leq \eta,\ \beta \cdot (D^*v - \eta) = 0, \tag{36}$$

One can simply eliminate μ and simply rewrite (35) as $v = v(\beta)$ depending only on β (we insist here that β only has dimension N) with

$$v = (w - D\beta)_+ \text{ i.e. } v_{ij}(\beta) := \max(w_{ij} - (\beta_i - \beta_j), 0) \tag{37}$$

and we are left to find β in such a way that $v(\beta)$ fulfills (36). At this point it is useful to observe the following

Lemma 3 *Define for every $\beta \in \mathbb{R}^N$*

$$\Phi(\beta) := \frac{1}{2}|v(\beta)|^2 = \frac{1}{2}\sum_{1\leq i,j\leq N}(w_{ij} - \beta_i + \beta_j)_+^2$$

then v solves (34) with K given by (33) if only if $v = v(\beta)$ and β solves

$$\inf_{\lambda\in\mathbb{R}_+^N} \Phi(\beta) + \eta \cdot \beta \tag{38}$$

Proof. Observe that Φ is convex and differentiable (it is even $C^{1,1}$ i.e. has a Lipschitz gradient) and

$$\nabla\Phi(\beta) = -D^*v(\beta)$$

so β solves (38) if only if

$$D^*v(\beta) \leq \eta,\ (\eta - D^*v(\beta)) \cdot \beta = 0$$

which is (36). ■

So the good news is that projecting onto K consists in minimizing a smooth and convex function on $\mathbb{R}^N$ with only nonnegativity constraints (in (38)). One can therefore use a (projected) gradient method and, what is even more important, use Nesterov's acceleration as follows.

Since $\nabla\Phi(\beta) = -D^*v(\beta)$ and v is 1-Lipschitz (for the euclidean norm of $\mathbb{R}^N$), $\nabla\Phi$ is

M-Lipschitz with $M := \|D^*\|_2$ is the 2-operator norm[9] of D^*. The standard projected gradient method for (38), consists, given an initial guess β_0, in iteratively setting

$$\beta_{k+1} = \Pi_+\Big(\beta_k - \frac{1}{M}(\eta - D^* v(\beta_k))\Big)$$

where Π_+ just consists of taking componentwise the positive part. This is simple to implement but converges quite slowly that is the difference between the desired minimum and the function to be minimized evaluated at β_k is $O(1/k)$. Nesterov's acceleration (Nesterov (1983), Beck and Teboulle (2009)) enables one to reach an error $O(1/k^2)$ with the same computational cost just by choosing properly varying gradient steps t_k by starting with $t_0 = 0$ and the recursion

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}.$$

Given an initial guess $\beta_0 = \overline{\beta}_0$, Nesterov's iterates, are then given by

$$\beta_{k+1} = \Pi_+\Big(\beta_k - \frac{1}{M}(\eta - D^* v(\overline{\beta}_k))\Big) \tag{39}$$

and

$$\overline{\beta}_{k+1} = \beta_{k+1} + \left(\frac{t_k - 1}{t_{k+1}}\right)(\beta_{k+1} - \beta_k).$$

The error between the minimum and the cost computed at β_k is $O(1/k^2)$ (see Beck and Teboulle (2009)).

[9] i.e. $\|D^*\|_2$ is the square root of $\sup\{\sum_i (D^* v)_i^2 \ : \ \sum_{ij} v_{ij}^2 \leq 1\}$ which is also the largest eigenvalue of DD^*.

References

Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica 64*(1), 51–75.

Armstrong, M. and J.-C. Rochet (1999). Multi-dimensional screening:: A user's guide. *European Economic Review 43*(4-6), 959–979.

Arrow, K. J., L. Hurwicz, and H. Uzawa (1958). *Studies in linear and non-linear programming*, Volume II of *Stanford Mathematical Studies in the Social Sciences*. Stanford University Press, Stanford, CA. With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow.

Atkinson, A. B. and J. E. Stiglitz (1976). The design of tax structure: direct versus indirect taxation. *Journal of public Economics 6*(1-2), 55–75.

Beaudry, P., C. Blackorby, and D. Szalay (2009). Taxes and employment subsidies in optimal redistribution programs. *American Economic Review 99*(1), 216–242.

Beck, A. and M. Teboulle (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci. 2*(1), 183–202.

Bertsekas, D. (2009). *Convex optimization theory*, Volume 1. Athena Scientific.

Boadway, R., M. Marchand, P. Pestieau, and M. del Mar Racionero (2002). Optimal redistribution with heterogeneous preferences for leisure. *Journal of Public Economic Theory 4*(4), 475–498.

Carlier, G. (2022). *Classical and modern optimization.* Adv. Textb. Math. Hackensack, NJ: World Scientific.

Chambolle, A. and T. Pock (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision 40*(1), 120–145.

Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica: Journal of the Econometric Society*, 607–622.

Choné, P. and G. Laroque (2010). Negative marginal tax rates and heterogeneity. *American Economic Review 100*(5), 2532–2547.

Cremer, H., P. Pestieau, and J.-C. Rochet (2001). Direct versus indirect taxation: the design of the tax structure revisited. *International Economic Review 42*(3), 781–800.

Cremer, H., P. Pestieau, and J.-C. Rochet (2003). Capital income taxation when inherited wealth is not observable. *Journal of Public Economics 87*(11), 2475–2490.

Diamond, P. A. and J. A. Mirrlees (1971). Optimal taxation and public production i: Production efficiency. *The American economic review 61*(1), 8–27.

Gao, Y. and W. Zhang (2023). An alternative extrapolation scheme of PDHGM for saddle point problem with nonlinear function. *Comput. Optim. Appl. 85*(1), 263–291.

Golosov, M., J. Hassler, P. Krusell, and A. Tsyvinski (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica 82*(1), 41–88.

He, B. and X. Yuan (2012). Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. *SIAM J. Imaging Sci. 5*(1), 119–149.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013). Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory 148*(5), 1770–1805.

Judd, K. L. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of public Economics 28*(1), 59–83.

Judd, K. L., D. Ma, M. A. Saunders, and C.-L. Su (2017). Optimal income taxation with multidimensional taxpayer types.

Kleven, H. J., C. T. Kreiner, and E. Saez (2009). The optimal income taxation of couples. *Econometrica 77*(2), 537–560.

Martinet, B. (1972). Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox. *C. R. Acad. Sci. Paris Sér. A-B 274*, A163–A165.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies 38*(2), 175–208.

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of public Economics 6*(4), 327–358.

Moreau, J.-J. (1965). Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France 93*, 273–299.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic theory 18*(2), 301–317.

Nesterov, Y. E. (1983). A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR 269*(3), 543–547.

Piketty, T. (2014). *Capital in the twenty-first century.* Harvard University Press.

Renes, S. and F. Zoutman (2017). As easy as abc? multidimensional screening in public finance.

Rochet, J.-C. (1987). A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of mathematical Economics 16*(2), 191–200.

Rochet, J.-C. (2009). Monopoly regulation without the spence–mirrlees assumption. *Journal of Mathematical Economics 45*(9-10), 693–700.

Rochet, J.-C. (2023). multidimensional screening after 36 years. *TSE discussion paper*.

Rochet, J.-C. and P. Choné (1998). Ironing, sweeping, and multidimensional screening. *Econometrica*, 783–826.

Rochet, J.-C. and L. A. Stole (2002). Nonlinear pricing with random participation. *The Review of Economic Studies 69*(1), 277–311.

Rochet, J.-C. and J. Thanassoulis (2019). Intertemporal price discrimination with two products. *The RAND Journal of Economics 50*(4), 951–973.

Rockafellar, R. T. (1976). Monotone operators and the proximal point algorithm. *SIAM J. Control Optim. 14*(5), 877–898.

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the roy model. *The Quarterly Journal of Economics 128*(2), 623–668.

Rothschild, C. and F. Scheuer (2016). Optimal taxation with rent-seeking. *The Review of Economic Studies 83*(3), 1225–1262.

Saez, E. (2002). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics 83*(2), 217–230.

Saez, E. and G. Zucman (2019). *The triumph of injustice: How the rich dodge taxes and how to make them pay.* WW Norton & Company.

Tarkiainen, R. and M. Tuomala (1999). Optimal nonlinear income taxation with a two-dimensional population; a computational approach. *Computational Economics 13*, 1–16.

Tarkiainen, R. and M. Tuomala (2007). On optimal income taxation with heterogeneous work preferences. *International Journal of Economic Theory 3*(1), 35–46.

Valkonen, T. (2014). A primal-dual hybrid gradient method for nonlinear operators with applications to MRI. *Inverse Problems 30*(5), 055012, 45.

Valkonen, T. ([2023] ©2023). First-order primal-dual methods for nonsmooth non-convex optimization. In *Handbook of mathematical models and algorithms in computer vision and imaging—mathematical imaging and vision*, pp. 707–748. Springer, Cham.

Wilson, R. B. (1993). *Nonlinear pricing.* Oxford University Press, USA.