WORKING PAPERS

N° 1537

MAY 2024

"A General Solution to the Quasi Linear Screening Problem"

Guillaume Carlier, Xavier Dupuis,
Jean-Charles Rochet, John Thanassoulis

Toulouse School of Economics

A General Solution to the Quasi Linear Screening Problem

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

May 20, 2024

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.

Keywords: Multidimensional screening, algorithm, numerical methods, price discrimination, optimal tax

JEL classification: C02, H21, D42

1 Introduction

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

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

*Ceremade, Université Paris-Dauphine, PSL Research University.

†Institut de Mathématiques de Bourgogne, UMR 5584 CNRS, Université de Bourgogne, F-21000 Dijon, France. The IMB to which belongs X. Dupuis receives support from the EIPHI Graduate School (contract ANR-17-EURE-0002).

‡Toulouse School of Economics. Rochet acknowledges the support of the European Research Council (ERC grant n° 101055239, DIPAMUTA). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

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

under the constraints that for all (i, j)

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

and for all i

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

In (1) y_i is a vector capturing the allocation to person i, $S_i(y_i)$ is the surplus created by this allocation, U_i is the utility received by person i and f_i is the frequency of this type of person in the population. y and U without subscripts denote the set of allocations and utilities.

In (2), the $\Lambda_{ij}(.)$ are the mimicking functions which will be specific to each application. These functions capture the utility gross of transfers available if a person j receives the allocation which was targeted at type i. Finally, the constraints (3) are individual rationality conditions.

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 analytical 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 incentive compatibility constraints are not *ex ante* identifiable which 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 multidimensional 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. This new formulation enables us to use the powerful algorithm proposed by Chambolle and Pock (2011) for solving such max-min problems. Since each step of the Chambolle-Pock algorithm itself requires an optimization problem to be solved, it is not obvious a priori that it can be used in practice for screening problems. Our first computational contribution is to

[1]Rochet (2024) surveys the literature on multidimensional screening that has followed that article.

show that these steps are indeed numerically feasible and we offer a method to solve them. A second advantage of our approach is that it can handle nonlinear parameterizations of utilities, which give rise to non-convex programs. We demonstrate that an extension of the Chambolle-Pock algorithm by Valkonen (2014) can be applied to such non-convex settings. Our approach is therefore applicable to all quasi-linear multidimensional screening problems.[2]

2 Motivations

Although there are many economic problems to which our algorithm can be applied, we will focus on just two particular applications: the multidimensional version of the multiproduct monopolist problem, 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 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 the 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 with several specifications represented by a vector of characteristics $y_i \in \mathbb{R}^d$, which denotes the set of characteristics targeted at consumers of type i . The dimension d represents the number of different features upon which the product can be differentiated,

[2] We do not claim however that the algorithm we propose is more efficient than existing ones developed for particular situations. For instance, the algorithm of Ekeland and Moreno-Bromberg (2010) handles very efficiently the case of a quadratic surplus and linear utilities. The more recent adaptative method of Mirebeau (2016) is certainly a more refined and efficient one to handle convexity constraints, but it requires types to be on a regular two-dimensional grid. Our approach is more flexible, as it does not share this limitation.

and each individual component in the vector y_i 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_i is given by a convex function $C(\cdot)$. The utility of buying this good for agent $i = 1, .., N$ is

$$U_i = \theta_i \cdot y_i - p_i, \tag{4}$$

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

$$\sum_i f_i[p_i - C(y_i)] = \sum_i f_i \left[\theta_i \cdot y_i - C(y_i) - U_i\right], \tag{5}$$

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 $U_i = \theta_i \cdot y_i - p_i$:

$$U_i \geq \theta_i \cdot y_j - p_j = U_j + (\theta_i - \theta_j) \cdot y_j, \tag{6}$$

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

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

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 that the distribution of types satisfy a monotone likelihood property. 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 several types 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.

There is a second canonical version of the multiproduct monopolist problem. In this version the seller's goods are indivisible and consumers have differing valuations for each of the N products on offer. Examples include a supermarket of groceries, a theatre or cinema selling multiple shows, or a bank offering its clients a number of different financial services and account types. Each consumer chooses a subset of the goods to buy, which is known as their '*bundle*'. This formulation of the problem again displays multiple type dimensions as consumers can have valuations over each of the N goods available. The initial insight that *mixed bundling*, that is offering prices for bundles of goods which differs from the sum of the component prices, can be part of the optimal pricing strategy is often credited to Adams and Yellen (1976). However optimal pricing policies were not established in that work. In a widely cited contribution Schmalensee (1984) considered the case of consumers' valuations being distributed according to a joint-normal distribution across two products. Once again optimal pricing was not achievable at that time, though Schmalensee (1984) proved that some bundle reduction was always more profitable than none if there was any negative correlation in valuations. McAfee et al. (1989) then demonstrated that mixed bundling was optimal if consumers' distributions were uncorrelated across producers.

The difficulty of working with multidimensional screening then forced the literature to adopt either the restriction that all consumers had to buy the whole bundle whilst keeping general distributions (e.g. Armstrong and Vickers (2010)), or that distributions were uniform on a square which allowed competition and other economic questions to be explored (e.g. Thanassoulis (2007)). The final approach is to determine the optimal prices numerically. With a small number of products standard optimisation routines can work (Chu et al. (2011)), though without convergence theorems such approaches can be impractical at scale.[3] We note below some of the limitations of off-the-shelf approaches and offer an algorithm which is flexible enough to solve such problems and for which convergence is guaranteed.

2.2 Joint taxation of saving and labour incomes

The question of the optimal mix between labour 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 greatly the employment and consumption choices of individuals and the investment decisions of firms.

However most of the academic literature on optimal taxation, starting with 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

[3]For example Chu et al. (2011) note (footnote 27) that they have manually checked different start points to draw comfort that the prices they determine are indeed optimal.

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, and there is an asymmetry of information between the tax payer and the government. 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 labour 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. In fact we know very little about multidimensional screening problems in general. Explicit results have been obtained for particular distributions of types in a variety of contexts. For example Wilson (1993) in his study of nonlinear pricing, and as noted above Armstrong (1996) and Rochet and Choné (1998) in the setting of multi-good monopoly pricing model. Rochet and Thanassoulis (2019) establish conditions such that dynamic screening, in which the menu offered evolves across time, can be optimal, indicating a new frontier across which the principal can optimise.[4] This literature highlights that the solution pattern in the multidimensional problem may differ markedly from that of the unidimensional case.

In the case of optimal taxation, the recent literature on the multidimensional problem 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 the approach adopted by 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,

[4]See also Rochet (2009) for the regulation of firms with different marginal and fixed costs.

Jacquet et al. (2013) study taxation 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.[5] 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.

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 or 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 will use the algorithm we offer to solve 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 $t = 1, 2$.

[5] 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.

Similarly to Diamond (1998) our agents have quasi linear preferences:[6]

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

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 of which are exogenous and uniform across agents. The principal seeks the tax system that maximizes a weighted sum of a Rawlsian objective and utilitarian welfare:

$$W = \lambda \min_j V_j + (1 - \lambda) \sum_i f_i V_i, \tag{7}$$

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. This allows us to rewrite the objective function (7) as $W = \min_j V_j + (1-\lambda) \sum_i f_i U_i$. Moreover, if we define the total surplus from citizens i as the sum of their utility and tax payment, $S_i = V_i + T_i$, and using that $\sum_i f_i T_i = G$, the objective of the principal can be rewritten as:[7]

$$W = \sum_i f_i [S_i - \lambda U_i]. \tag{8}$$

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

$$S_i(y_i) = u(e_i - s_i) + Rs_i + (w - x_i)l_i, \tag{9}$$

and

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

An interesting economic question, which a solution to (7) 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

[6] Diamond (1998) and Saez (2001) have shown that quasi-linearity allows a deeper and more intuitive comprehension of optimal tax systems. The absence of wealth effects greatly simplifies the analysis.

[7] To derive (8) note we can write

$$W = \min_j V_j + \sum_i f_i (S_i - T_i - \min_j V_j) - \lambda \sum_i f_i U_i,$$

and this differs from (8) by the constant G.

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 disultility 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, 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

3.1 Additional Notation

For expositional simplicity, we restrict our discussion to the case where the assignment y can be any vector[8] in $\mathbb{R}^d$. 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: $\mathbb{R}^N \to \mathbb{R}^{N\times N}$ by

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

Note that $D(\mathbb{R}^N) \subset H_N$ where H_N denotes the ($N \times (N-1)$ dimensional) space of $N \times N$ matrices with zero entries on the diagonal.

The inner product[9] of the matrix Du with matrix $v \in \mathbb{R}^{N\times N}$ is defined as:

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

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

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

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

[8]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.

[9]This is sometimes referred to as the Frobenius inner product or the scalar product for matrices. It should not be confused with matrix multiplication.

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).

3.2 Existence and First Order conditions

Existence of a solution and first-order optimality conditions are given by:

Proposition 1 *Assuming that for every i*

$$S_i(y_i) \to -\infty \text{ as } |y_i| \to \infty, \tag{10}$$

then (1)-(2)-(3) admits 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

$$\hat{u}_i - \hat{u}_j > \nabla\Lambda_{ij}(\overline{y}_j)\hat{y}_j,\ \forall(i, j) \in A, \tag{11}$$

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:

$$\lambda f = \overline{\mu} + D^*\overline{v},\ f_j\nabla S_j(\overline{y}_j) = \sum_i \overline{v}_{ij}\nabla\Lambda_{ij}(\overline{y}_j),\ \forall j \tag{12}$$

together with the complementary slackness conditions:

$$\overline{\mu}_i\overline{U}_i = 0,\ \overline{v}_{ij}(D\overline{U} - \Lambda(\overline{y}))_{ij} = 0. \tag{13}$$

Proof. Condition (10) 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 $\min_i U_i = 0$. 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 (12)-(13) 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 (11). ■

Let us briefly comment on the assumptions in the previous proposition. First observe that (10) 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 (11) is easily obtained. Note also that when Λ is linear and the S_i's are concave, the first-order conditions (12)-(13) are sufficient conditions.

Of course (10) is a technical assumption which is not needed for the existence of an optimal solution if y is constrained to remain in a compact set. This holds in the particular case of a multiproduct monopolist facing linear costs and considering allocations which can be randomisations over bundles – as initially explored by McAfee and McMillan (1988).[10] In this case, S and Λ are linear and there are additional linear constraints: for every j, $y_j \in \Gamma$ where Γ is a convex compact set defined by finitely many linear inequalities. This makes (1)-(2)-(3) a linear programming problem with a compact constrained set and the existence of a solution is straightforward. Characterization of optimal solutions can be obtained by similar conditions as (12)-(13) with additional KKT multiplier terms corresponding to the constraints $y_j \in \Gamma$.

3.3 Alternative Numerical Methods

We develop below a primal-dual method that we propose for solving this problem. It is easily implementable, flexible enough to address the multidimensional screening problems discussed in the paper, and convergence is guaranteed.

However one might wonder if other currently standard numerical methods could also work, such as explicit gradient methods, linearization of the constraints, and Newton methods. Newton methods are well-known to converge only very close to the solution for unconstrained problems and are not very well-suited to handle constraints. For optimization problems with a large number of constraints, such as the screening problems we consider in this paper, the standard projected gradient method on the initial formulation of the problem is not really an option since each step is as costly as the intial problem. Moreover, when incentive compatibility constraints are nonlinear, linearizing them may lead to infeasible points, and to stability and convergence issues. We believe that this is where our duality (max-min formulation) arguments are particularly appealing: they enable us to satisfy the constraints by updating the multipliers at each step in a suitable way. Of course, it comes with a cost: we have to project onto a set K in the space of matrices at each step, but as shown in the appendix this is tractable. As for feasibility, it is worth mentioning that our algorithm finds at each step a feasible point (see footnote

[10]Randomisations were subsequently shown to be optimal in some cases (Thanassoulis, 2004).

15). More standard explicit methods based on gradients and linearizations techniques, would not.

Linear programming methods can also be used such as the celebrated Frank-Wolfe algorithm which consists in solving a linearized problem at each step. In theory these linearized problems can be solved exactly (by the simplex algorithm), or approximately (by interior point methods). However such methods are well-known to have a complexity that scales badly with the number of constraints. For instance, in the worst cases, the simplex algorithm may have to explore all the vertices of the constrained set. The number of steps grows exponentially with the size of the problem. This makes LP prohibitively costly, except for special constrained sets like the simplex or the box, on which optimizing a linear function is easy. However, the geometry of the incentive compatibility constraints does not fall in this category. To illustrate the limitations of LP solvers, it might be worth recalling that the well-known optimal transportation problem is an LP problem but except for small size instances, efficient methods are not by linear programming because they have a cubic complexity. In our screening framework, unless the initial problem is linear, the situation is even worse if we have to use LP at each step of the algorithm. So one reason for not using linear programming is its computational cost. There are also stability issues due to the fact that solutions of LP (which may be non-unique) are discontinuous with respect to the objective and the constraints. This is a second reason why we did not adopt LP like strategies.

Finally, for the quadratic version of the Rochet Choné problem that we solve in section 6, other methods are more efficient. For example, Mirebeau's adaptative algorithm is perhaps the current state of the art for handling convexity constraints when types are on a regular two-dimensional grid. Similarly, Ekeland and Moreno-Bromberg (2010) use a fast linear quadratic solver which achieves great efficiency, but only works with a quadratic cost. The algorithm we propose is different because it also involves dual variables. It is more widely applicable because it can handle non-regular grids, non-uniform distributions and more general surpluses as well as nonlinear utilities.

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. This will allow us to divide the optimisation problem into sub-problems which we will then be able to show can be brought to existing optimisation algorithms.

4.1 Feasibility

Let us first introduce a definition:

Definition 1 *Let $\Lambda \in H_N$ (i.e. Λ is an $N \times N$ 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 $DU \geq \Lambda$. Further, using $\Lambda_{ii} = 0$ we can show that 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

$$T_\Lambda(U)_i := \max_j \{U_j + \Lambda_{ij}\}. \tag{14}$$

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

Proposition 2 *Let $\Lambda \in H_N$. 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*
$$\sum_{k=0}^{L} \Lambda_{i_k i_{k+1}} \leq 0. \tag{15}$$
3. *Defining T_Λ by (14), 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 non-negative 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 (15) 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

$$u_i^n = \max \Big\{ \sum_{k=0}^{n-1} \Lambda_{i_k i_{k+1}} \ : \ i_0 = i, \ i_1, \dots, i_n \in \{1, \dots, N\}^n \Big\}. \tag{16}$$

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

(15) we have

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

so that

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

where the last inequality follows from (16) 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. ⇒ 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. ⇒ 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 any collection of positive weights $f_i > 0$.

Intuitively the utility assignment is determined by allowing each possible agent type just enough utility that it outweighs the benefits of mimicking another type. The proof of Proposition 2 works by assessing all possible chains of mimickry, e.g. type i pretending to be type j who in turn pretends to be type k and so on. The fixed point finds the most tempting mimicking type at every stage in this chain. This process does not rely on the number of types of i or j in the population, that is the frequency f_i has no role. All that is required is that a type i or j could exist. This is enough as it permits each type the possibility of pretending to be that type. This is why Corollary 1 applies for all distributions f_i, given $f_i > 0$ for all i.

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

Lemma 1 *Let $\Lambda \in H_N$. 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 $DU \geq \Lambda$. 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 given by the value of the sub-problem:

$$R(y) := \inf \Big\{ \sum_i f_i U_i \ : \ U_i \geq 0, \ U_i - U_j \geq \Lambda_{ij}(y_j) \Big\}. \tag{17}$$

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

$$K := \{v \in \mathbb{R}^{N\times N} \ : \ v \geq 0, \ D^* v \leq \lambda f\}, \tag{18}$$

as well as its support function:

$$\sigma_K(\Lambda) := \sup\{v \cdot \Lambda \ : \ v \in K\},$$

defined for all $\Lambda \in H_N$.[11]

[11]The support function of a set is an important tool in Convex Geometry. It can be thought of as defining the set of hyperplanes which enclose the set (Hug et al., 2020).

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

$$R(y) = \sup\Big\{v \cdot \Lambda(y) = \sum_{i,j} v_{ij}\Lambda_{ij}(y_j) \ : \ \sum_j (v_{ij} - v_{ji}) \leq f_i, \ v_{ij} \geq 0\Big\}. \tag{19}$$

We thus have

$$\lambda R(y) = \sigma_K(\Lambda(y)). \tag{20}$$

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 (17) is nonempty. We claim that the infimum in (17) 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 (17). 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 (17) admits a solution, so does its dual problem which is exactly the linear maximization problem in (19) 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 Max-min reformulation, optimality conditions

Using (20) and Proposition 3, we will establish that the initial screening problem (1)-(2)-(3) is equivalent to

$$\max_{y\in\mathbb{R}^{d\times N}} S(y) - \sigma_K(\Lambda(y)). \tag{21}$$

By definition of σ_K, this can be rewritten in max-min form

$$\max_{y} \min_{v\in K} S(y) - v \cdot \Lambda(y). \tag{22}$$

Formally we have:

Proposition 4 *$(\overline{y}, \overline{U})$ solves (1)-(2)-(3) if and only if $\overline{y}$ solves (21) 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 (21) (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 (12)-(13) imply that

$$D^*\overline{v} \leq \lambda f \text{ and } \overline{v} \geq 0 \text{ i.e. } \overline{v} \in K,$$
$$\text{and} \quad \lambda R(\overline{y}) \leq \lambda f \cdot \overline{U} = D^*\overline{v} \cdot \overline{U} = \overline{v} \cdot D\overline{U} = \overline{v} \cdot \Lambda(\overline{y}) \leq \sigma_K(\Lambda(\overline{y})) = \lambda R(\overline{y})$$

which, thanks to Proposition 3, yields

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

We can therefore reformulate the necessary conditions (12)-(13) 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 (11)) at $(\overline{y}, \overline{U})$. Then there exists $\overline{v} \in K$ such that*

$$\sigma_K(\Lambda(\overline{y})) = \overline{v} \cdot \Lambda(\overline{y}), \tag{23}$$

and

$$f_j \nabla S_j(\overline{y}_j) = \sum_i \overline{v}_{ij} \nabla \Lambda_{ij}(\overline{y}_j). \tag{24}$$

In terms of sufficient conditions, we have:

Proposition 6 *Assume* $(\overline{y}, \overline{v}) \in \mathbb{R}^{d\times N} \times K$ *satisfy conditions (23)-(24) 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 (21) that is* $(\overline{y}, \overline{U})$ *is a local (global) solution of (1)-(2)-(3),*

2. *if*

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

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 (21).*

Proof. 1. Follows from $S(y) - \sigma_K(\Lambda(y)) \leq S(y) - \overline{v} \cdot \Lambda(y)$ with 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 linear or affine,* $S - \sigma_K \circ \Lambda$ *is concave so conditions (23)-(24) are in fact necessary and sufficient (global) optimality conditions for problem (21).*

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 (23)-(24). We assume that S is concave and differentiable, and that Λ is smooth. We start with the case in which Λ is linear. In this case the algorithm we propose coincides with that developed by Chambolle and Pock (2011).[12]

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 φ :

[12] 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)).

$\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous 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

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

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

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

In particular

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

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

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

introduced in Martinet (1972). This algorithm is known to converge to a minimizer provided such a minimizer exists, see Rockafellar (1976). The proximal point algorithm has several appealing properties. These include that it allows for a nonsmooth objective φ and also that the sequence $\varphi(x_k)$ is decreasing. The latter follows as by construction the proximal operator satisfies the inequality $\varphi(x_{k+1}) + \frac{1}{2}|x_{k+1} - x_k|^2 \leq \varphi(x_k)$. Of course, to use proximal methods in practice, one should be able to compute $\operatorname{prox}_\varphi$ 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 $\operatorname{prox}_{\chi_C}$ coincides with the projection (closest point map) proj_C onto C.

5.2 Proximal primal-dual algorithm in the linear case

Suppose 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), and so Λ is the linear map

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

defined in (2).[13] The problem (21) 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 (23)-(24). For given step sizes $\tau > 0$ and $\sigma > 0$, the Chambolle-Pock algorithm consists of the following iterations:

$$y_{k+1} = \text{prox}_{-\tau S}(y_k - \tau\Lambda^*(v_k)), \tag{25}$$

$$\widetilde{y}_{k+1} = 2y_{k+1} - y_k, \tag{26}$$

$$v_{k+1} = \text{proj}_K(v_k + \sigma\Lambda(\tilde{y}_{k+1})). \tag{27}$$

Theorem 1 from Chambolle and Pock (2011) (also see He and Yuan (2012) for a simpler proof) guarantees that the iterates above converge to a solution of the system (23)-(24) if $\tau > 0$, $\sigma > 0$ satisfy $\tau\sigma\|\Lambda\|^2 < 1$ where

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

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

5.3 Proximal primal-dual algorithm in the General Case

When Λ is nonlinear, it is possible to use the linearization of primal updates which leads to the algorithm proposed and analyzed by Valkonen (2014):[14]

$$y_{k+1} = \text{prox}_{-\tau S}(y_k - \tau(\Lambda'(y_k))^* v_k), \tag{28}$$

$$\widetilde{y}_{k+1} = 2y_{k+1} - y_k, \tag{29}$$

$$v_{k+1} = \text{proj}_K(v_k + \sigma\Lambda(\tilde{y}_{k+1})). \tag{30}$$

[13] 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 $(\Lambda^* v)_j = \sum_i (\theta_i - \theta_j) v_{ij}$.

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

Note that if these iterates converge to some $\overline{y}$, $\overline{v}$ one will 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 (23)-(24) from Proposition 5 and in particular $\Lambda(\overline{y})$ is feasible[15] in the sense of definition 1. The (local) convergence analysis of the above algorithm to a solution of (23)-(24) is rather involved and can be found under various technical assumptions[16] in Valkonen (2014), and more recently in Valkonen (2023). Theorem 1 in Gao and Zhang (2023) gives a shorter proof of local convergence (for a slightly different algorithm where the linearization of the nonlinear map Λ is used at the level of the updates (30) for v instead of the updates (28) for y).

Thus we have converted the multidimensional screening problem into a form which can be tackled using recent algorithms. We explain in the Appendix how the proximal steps (25) (or (28)) and (26) (or (29)) 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 quality of each characteristic. The marginal cost of producing a product with quality y is assumed to be given by a quadratic function

$$C(y) := \frac{1}{2}\left(y_1^2 + y_2^2\right). \tag{31}$$

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 $y = \theta$.

To illustrate our algorithm let us suppose that consumers are uniformly distributed on an $N \times N$ grid supported on the square $[1, 2]^2$. The use of the uniform distribution

[15] Let us emphasize that in fact, at each step of our algorithm, we find a feasible point, indeed it directly follows from (30) that $\Lambda(\tilde{y}_{k+1}) + \frac{v_k - v_{k+1}}{\sigma}$ is feasible.

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

of types and quadratic cost function is purely for expositional simplicity and to reflect the textbook cases studied in the literature. The algorithm we propose permits arbitrary distributions of types with any surplus functions which are smooth as long as the problem has an interior solution such that the qualification condition of Proposition 1 is satisfied.

In Figure 1 we solve two versions of this monopolist 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 in which we study precisely which incentive compatibility constraints are binding.

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 characteristics a type θ_i receives equal to her type: $y_i = \theta_i$. 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 assignments 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 assigned 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.

Our finding that all clients who have a high valuation on at least one dimension receive close to first-best, while there is significant distortion only for clients who have a lower type on all dimensions extends the general intuition as to how sellers should optimally offer volume discounts. It is known that in one dimension volume discounts apply only to the highest volumes so that the high value clients purchase (close to) the first-best quantity (Maskin and Riley, 1984). We show that volumes (or quality) are most distorted downwards and bunched for those at the bottom for the value distribution in all dimensions. Whereas those nearer the top in any dimension escape bunching and see less by way of distortion.

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 pattern of binding IC constraints forms a distinctive '*tree pattern*' in which the branches of the tree are the binding IC constraints. One-dimensional problems have a trivial such tree pattern; whereas the complexity of the tree in the multidimensional case is closely tied to the difficulty in solving these problems, a problem which our algorithm overcomes. The green spots in Panel (b2) further show that the rents of the bottom six client types are fully extracted.

We offer a second illustration to demonstrate the versatility of our algorithm. We solve the problem posed by Schmalensee (1984). Suppose that consumers are distributed over an $N \times N$ grid with the coordinates of the grid points capturing the value consumers have for each one of two possible products. Further assume that the weight at each grid point is adjusted so as to approximate a population with valuations drawn from a joint normal distribution. Valuations are not therefore independent and neither are they uniform. Suppose that the monopolist sells these two products and consumers' demand at most one unit of each good at her valuation. The monopolist wishes to choose the optimal price profile $\{p_1, p_2, p_b\}$ for the two component goods 1 and 2 and the bundle of both goods. If marginal costs are set to zero then the monopolist's problem is:

$$\max_{\{p_1,p_2,p_b\}} p_b \cdot Q_b(p_1,p_2,p_b) + p_1 \cdot Q_1(p_1,p_2,p_b) + p_2 \cdot Q_2(p_1,p_2,p_b)$$
$$\text{subject to } \{p_1,p_2\} > 0,\ 0 < p_b \leq p_1 + p_2$$

where Q_1 is the demand for good one only, and so on for the other functions. This program can be rewritten as a screening problem identically to (4)–(6) with $C(y_i) = 0$ and with the additional constraint that

$$y_i \in [0,1]^2.$$

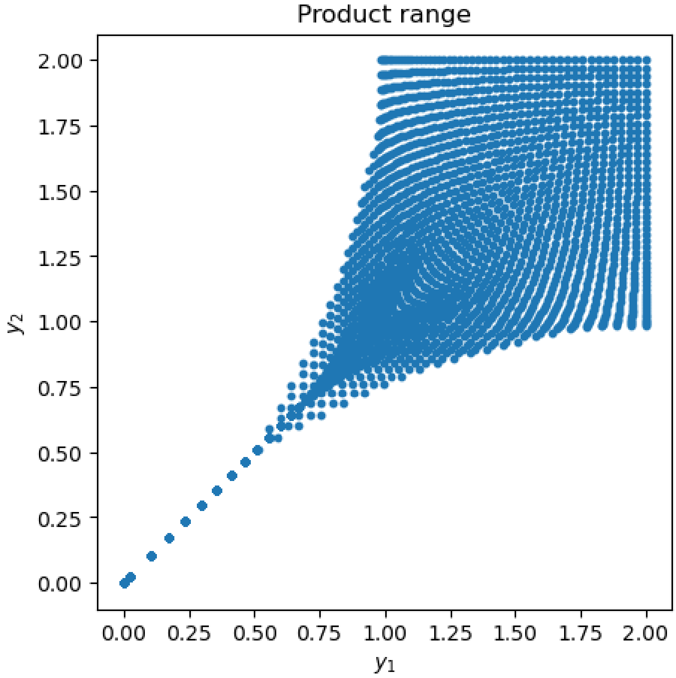
That is, we restrict clients to get no more than one unit of a good.[17]

In Figure 2 we solve the Schmalensee (1984) problem at 5000 distinct pairs of parameter values for mean product valuation μ and correlation coefficient ρ. For a given (μ, ρ) pair we model consumers as having valuations drawn from a normal distribution approximated by a 10×10 grid on $[0, 10]^2$ with variance $\sigma^2 = 10$. Figure 2 shows the richness of the pattern of optimal prices. We see that the optimal component good prices are sensitive to both the mean valuation in the population and the correlation the consumers have between their component good valuations. The optimal bundle prices remain sensitive to the the average valuation of each product (μ), and so to the average valuation clients have for the bundle, while bundle prices are not highly sensitive to the correlation between the product valuations (ρ). The spirit of this result – that mixed bundling allows less volatility in bundle sales than in individual good sales – is captured in the literature only when the law of large numbers can be invoked by assuming an infinity of products (Bakos and Brynjolfsson (1999)). The individual product prices are sensitive to the correlation

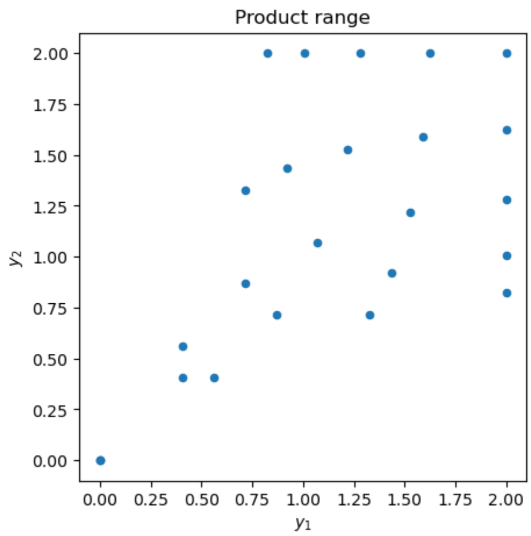
[17] As marginal costs are linear the algorithm will pick out $y_i \in \{0, 1\}$ as optimal.

of values and adjust so that the bundle discount can be set to optimally encourage bundle sales amongst those with the highest valuations. Figure 2 demonstrates visually that our algorithm allows this formulation of the general monopolist's problem to be solved and so showcases its flexibility even when consumers' valuations are not uniformly distributed. The proportional bundle discount $\frac{2p-p_b}{2p}$ is plotted in the third figure of Figure 2. The beauty of the optimal pattern yields some insight into why analytical solutions are so difficult to establish in this area. We see that bundle reductions become less generous as the correlation coefficient between the goods valuations becomes more positive – but the effect is not uniform and depends also on the average good valuations μ.

Panel (a)



Panel (b1)



Panel (b2)

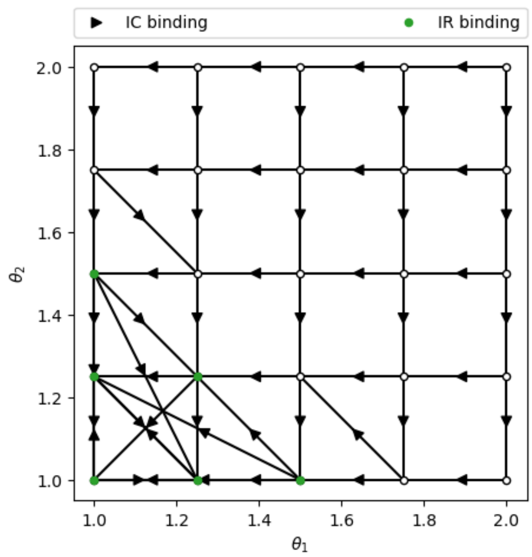


Figure 1: Solution to the Multidimensional Monopolist Problem

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 *tree structure* of binding incentive compatibility (IC) and individual rationality (IR) constraints for the 5×5 case at the optimal assignment of products to clients. All types are connected to the lowest type $(1, 1)$ by sequences of arrows (the 'branches' of the tree). Assignments are distorted in order to decrease the rents of all types located higher up the tree. Clients are supported on $[1, 2]^2$ and production costs are given in (31). Note that the green spots in panel b2 depict that rents are fully extracted by the monopolist – the individual rationality constraints are binding. Code for the simulation algorithm is available at `https://github.com/x-dupuis/screening-algo`.

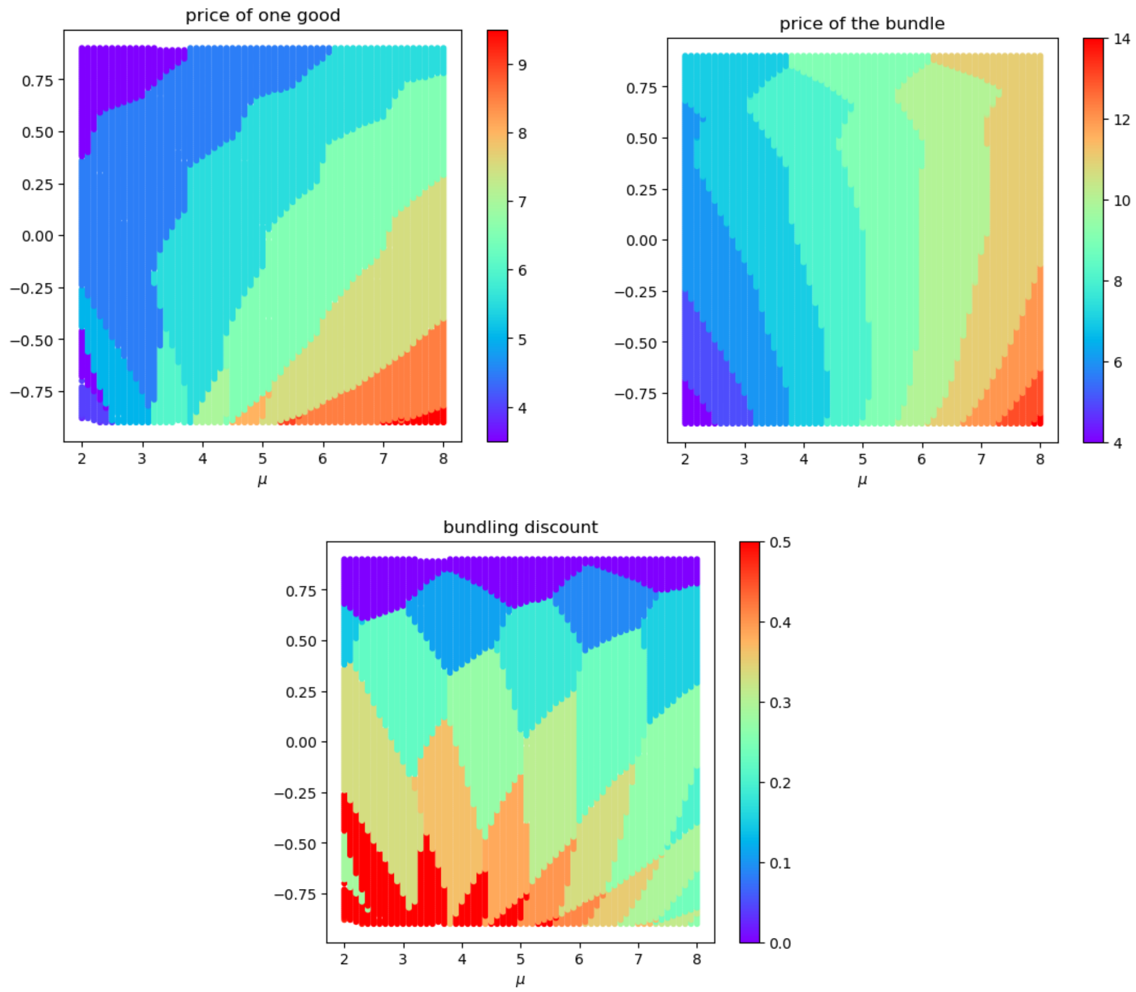


Figure 2: Solution to the Schmalensee (1984) problem

Notes: Consumers are distributed according to a joint-normal, and not a uniform, distribution. The vertical axis gives the correlation coefficient ρ in all three graphs. The richness of the pricing solution is apparent. For every pair of parameters (μ, ρ), optimal prices are calculated using our algorithm for a set of consumers distributed according to a joint-normal distribution approximated by the values it takes on a 10×10 grid covering $[0, 10]^2$. The variance of the joint-normal is set to $\sigma^2 = 10$. Non-deterministic prices are not, in general, optimal under this discrete distribution. The small number of such points have been dropped from the figures. The bundling discount is calculated as $1 - \frac{p_b}{2p}$. A proportionate bundle reduction of 0.5 indicates parameters for which only the bundle is sold ($p_b = p$). 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 (8) and the surplus available to citizens was given in (9). In this formulation we allow citizens to be differentiated on two dimensions; the type of individual i is a couple (e_i, x_i). In this illustration suppose that endowments e_i are distributed uniformly on a regular grid on $[1, e_{max}]$. Labour disutilities x_i are distributed uniformly on a regular grid on $[0, 1]$.[18] 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 assumed to be

$$u(C_1) := \frac{1}{\eta}\left(1 - e^{-\eta C_1}\right), \tag{32}$$

with $\eta = 1$. Period one consumption is $e - s$ while period two consumption is the sum of investment returns Rs and labour income, minus taxes. The return on savings is R and the social planner's Rawlsian weight is λ. Using the exponential utility as depicted in (32), we can establish in closed form the first best allocation when types are publicly observable. It is characterized by labour supply $l = 1$ 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 labour disutility x is the same for all agents, so that the agents only differ in their wealth endowment 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

$$T'(s(e)) = R - u'(e - s(e)), \tag{33}$$

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

$$U^{*\prime}(e) = u'(e - s(e)) > 0. \tag{34}$$

[18] The use of the uniform distribution here is for expositional convenience – the algorithm applies to general distributions.

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

$$x^*(e) = w + U_1^*(e) - U_0^*(e). \tag{35}$$

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 (33) and (34) 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.

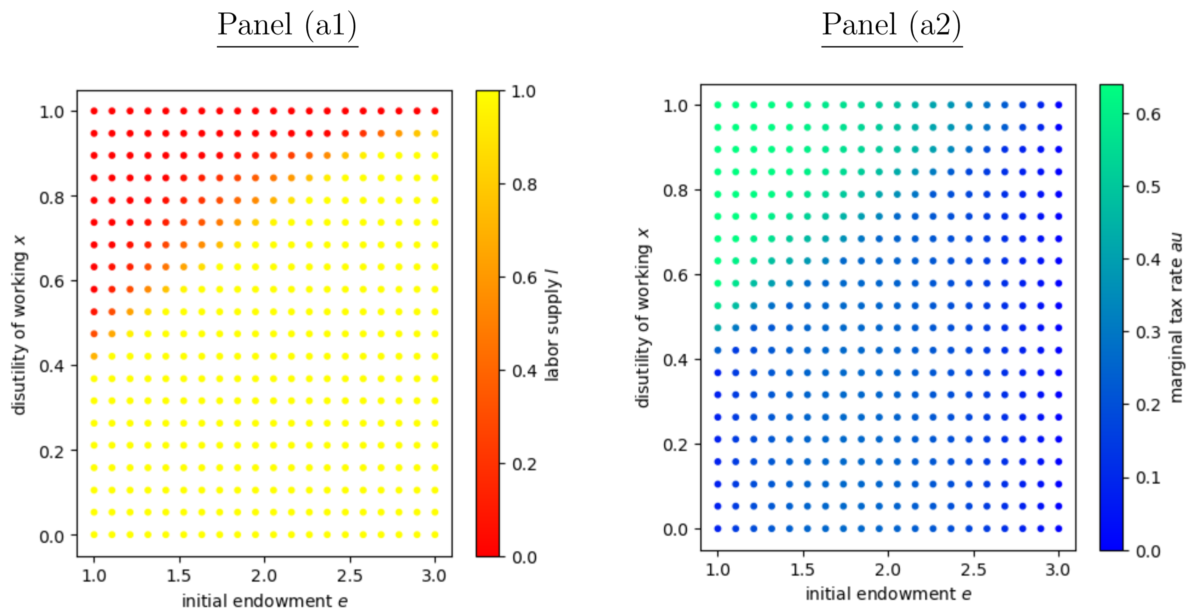


Figure 3: Solution to the multidimensional tax problem
Panel (a1) depicts the work decision citizens make in response to the optimal tax scheme. Panel (a2) depicts the marginal tax rate on savings given in (33). The simulation sets $\lambda = 1/2, R = 1, \eta = 1$, and citizens are modelled as taking 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 3. Panel (a1) of Figure 3 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 3. Panel (a2) plots the marginal tax rate with respect to saving which is given in (33). 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.

The dependence of the optimal labour taxes on the level of capital tax being paid is likely to be an important feature of the structure of optimal tax in richer, multidimensional settings. Simple discrete $n \times n$ examples for small n make it hard to discern the nature of this dependence. Numerical approaches, such as that championed here, potentially allow us to gain new insights on this topic.

Our initial question was whether the savings of employed people should be taxed differently to those of the unemployed. From Panel (a1) of Figure 3 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 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.

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 (11) 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 (25) (or (28)) and (26) (or (29)) can be handled in practice.

9.2.1 Updates for the primal variables y

The proximal steps (25) (or (28)) 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

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

Note that $S(y) = \sum_i f_i S_i(y_i)$ is a separable function so that these 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 completed by standard methods such as gradient ascent. Note that if S_i is quadratic (as in our Multiproduct 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,

$$K := \{v \in \mathbb{R}^{N\times N},\ v \geq 0,\ D^* v \leq \eta\} \tag{36}$$

where $\eta := \lambda f$. The projection onto K 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 wish to solve

$$\inf_{v\in K} |v - w|^2 = \sum_{1\leq i,j\leq N} (v_{ij} - w_{ij})^2. \tag{37}$$

This is a quadratic problem with finitely many linear and inequality constraints. The unique solution v of (37) 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

$$v - w = \mu - D\beta,\ \mu \geq 0,\ \mu \cdot v = 0, \tag{38}$$

as well as

$$\beta \geq 0,\ D^*v \leq \eta,\ \beta \cdot (D^*v - \eta) = 0, \tag{39}$$

One can simply eliminate μ and rewrite (38) as $v = v(\beta)$ depending only on β (we insist here that β only has dimension N) with

$$v = (w - D\beta)_+ \text{ i.e. } v_{ij}(\beta) := \max(w_{ij} - (\beta_i - \beta_j), 0). \tag{40}$$

We are left to find β in such a way that $v(\beta)$ fulfills (39). 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 (37) with K given by (36) if only if $v = v(\beta)$ and β solves

$$\inf_{\lambda\in\mathbb{R}_+^N} \Phi(\beta) + \eta\cdot\beta \tag{41}$$

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 (41) if only if

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

which is (39). ■

So projecting onto K consists in minimizing a smooth and convex function on $\mathbb{R}^N$ with only nonnegativity constraints (in (41)). One can therefore use a (projected) gradient method and, we propose, use Nesterov's acceleration as follows:

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

is M-Lipschitz with $M := \|D^*\|_2$, the 2-operator norm[19] of D^*. The standard projected gradient method for (41), 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_+ 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 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

$$\beta_{k+1} = \Pi_+\Big(\beta_k - \frac{1}{M}(\eta - D^* v(\overline{\beta}_k))\Big) \tag{42}$$

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)).

These steps have been implemented in our code[20] which we used to solve the illustrative examples found in §6 and §7.

[19] 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^*.

[20] Available at `https://github.com/x-dupuis/screening-algo`.

References

Adams, W. J. and J. L. Yellen (1976). Commodity bundling and the burden of monopoly. *The quarterly journal of economics 90*(3), 475–498.

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.

Armstrong, M. and J. Vickers (2010). Competitive non-linear pricing and bundling. *The Review of Economic Studies 77*(1), 30–60.

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.

Bakos, Y. and E. Brynjolfsson (1999). Bundling information goods: Pricing, profits, and efficiency. *Management science 45*(12), 1613–1630.

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*, 607–622.

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

Chu, C. S., P. Leslie, and A. Sorensen (2011). Bundle-size pricing as an approximation to mixed bundling. *The American Economic Review*, 263–303.

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. (1998). Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, 83–95.

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

Ekeland, I. and S. Moreno-Bromberg (2010). An algorithm for computing solutions of variational problems with global convexity constraints. *Numer. Math. 115*(1), 45–69.

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.

Hug, D., W. Weil, et al. (2020). *Lectures on convex geometry*, Volume 286. Springer.

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.

Maskin, E. and J. Riley (1984). Monopoly with incomplete information. *The RAND Journal of Economics 15*(2), 171–196.

McAfee, R. P. and J. McMillan (1988). Multidimensional incentive compatibility and mechanism design. *Journal of Economic theory 46*(2), 335–354.

McAfee, R. P., J. McMillan, and M. D. Whinston (1989). Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics 104*(2), 371–383.

Mirebeau, J.-M. (2016). Adaptive, anisotropic and hierarchical cones of discrete convex functions. *Numer. Math. 132*(4), 807–853.

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. (2024). multidimensional screening after 37 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. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68 (1)*, 205–229.

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.

Schmalensee, R. (1984). Gaussian demand and commodity bundling. *Journal of Business*, S211–S230.

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.

Thanassoulis, J. (2004). Haggling over substitutes. *Journal of Economic theory 117*(2), 217–245.

Thanassoulis, J. (2007). Competitive mixed bundling and consumer surplus. *Journal of Economics & Management Strategy 16*(2), 437–467.

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.