NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Sergey Kokovin, Babu Nahata,
and Evgeny Zhelobodko

DISTORTION IN SCREENING AND SPATIAL PREFERENCES

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: ECONOMICS
WP BRP 83/EC/2014

Sergey Kokovin,[1] *Babu Nahata,*[2] *and Evgeny Zhelobodko*[3]

DISTORTION IN SCREENING AND SPATIAL PREFERENCES

Abstract

We study a multidimensional screening problem with minimal restrictions on valuations. Our ε-relaxation of the constraints excludes bunching and cycles in the graph of active incentive-compatibility constraints. Therefore, the Lagrange multipliers do exist and enable us in characterizing distortion. In particular, under "spatial" preferences that include both the Hotelling and the Spence-Mirrlees cases, the solution has a simple planar graph. Consequently, the pattern of distortion is *centrifugal*, i.e., the points of service are biased towards the low-valuation market segments.

Keywords: incentive compatibility, multidimensional screening, second-degree price discrimination, non-linear pricing, product line, distortion, envy-graphs.

JEL Codes: D42, D82, L10, L12, L40.

1 Introduction

The modern theory of screening or non-linear pricing does consider multidimensional goods or services, or/and situations when consumers' valuations for the commodity are not strictly ordered either in a vertical or a horizontal sense (see reviews by Rochet and Stole (2003), Armstrong (2006), Stole (2007)). "Vertically ordered" valuations in our context mean that they satisfy the Spence-Mirrlees single-crossing condition (SCC).[4] That is, a higher-type agent values the commodity higher, the valuations cross only once at zero and the demands do not cross. Another popular simplifying assumption originating from Hotelling's linear-city model is the "horizontal" ordering: all agents are identical, except for the locations of their bliss points in some unidimensional space of quantity/quality. Thus, non-participation by the consumers does not imply a common outside option, but the level of reservation utility is the same for all

[1]National Research University Higher School of Economics, pr. Rimskogo-Korsakova, 47, St. Petersburg 190068, Russia: tel/fax 8-812- 6779384, and National Research University Novosibirsk State University, ul. Pirogova, 2, Novosibirsk 630090, Russia, skokov7@gmail.com

[2]Department of Economics, University of Louisville, Louisville, Kentucky 40292, USA. e-mail: nahata@louisville.edu

[3]*Evgeny Zhelobodko*: in memoriam 25.09.1973–27.03.2013. Kokovin gratefully acknowledges the support from the Economics Education and Research Consortium Inc (EERC)–grant 06-056; grants RFBR 12-06-00174a, and 11.G34.31.0059 from the Russian Government. Nahata acknowledges support from the University of Louisville. We thank R. Ericson, V. Polterovich and A. Savvateev for their valuable comments.

[4]When a consumer type i has some willingness-to-pay or monetary valuation $V_i(x)$ for quantity/quality $x \geq 0$, SCC is usually understood as $V'_{i+1}(x) > V'_i(x)\ \forall i, x,\ V_i(0) = 0$.

types. Under both these simplifying assumptions, the conclusions about the solution structure, distortion and informational rent are well known. Vertical ordering ensures efficiency at-the-top (the highest-demand type) and a downward distortion below with informational rent for all higher types. Horizontal ordering under monopoly results in overall efficiency without any informational rent (see Nahata et al., 2003, Andersson, 2008). Without these two traditional restrictions on the preference ordering, similar conclusions become more complicated, but we show that the topic is tractable.

In order to motivate our paper, consider the Hotelling linear city, but consisting of many blocks. Each block is inhabited by a block-specific mass of consumers with a block-specific reservation utility, some blocks may be empty. A monopolist designs a pizzeria chain, serving each block with its own pizzeria, or may leave some blocks unserved. The four questions we address are: How to find the location/pricing solution? Will the solution be socially efficient or distorted? Will the location pattern be grouped towards the consumers with higher willingness to pay, or dispersed? Who gets the informational rent?

Though formally we focus on monopoly, our intention in the design of the model is also to include indirectly oligopolistic markets with free entry, for example fast-food chain stores. In such situations, when designing a menu for all blocks, each firm considers the existing price/location bundles of other firms as given *multiple outside options*, which becomes a feature of our paper. We study mainly product lines (screening) in one- or two-dimensional quality-spaces, both for a general case and also under a specific "spatial" class of preferences, somewhat different from two most standard classes of preferences (Spence-Mirrlees or Hotelling).

More specifically, this paper considers a discrete product line for discrete consumer types.[5] The setting is almost standard but for the two features. The first is different outside options for different consumers. A rare example of multiple outside options in screening theory is Rochet and Chone (1998), who study bunching—same bundles for different types of consumers. We use this feature to build a bridge from monopolistic to oligopolistic free-entry screening that remains inadequately explored. Our second and the main novelty is the ε-relaxation of the incentive-compatibility constraints. It allows us to get rid off "essential" bunching (see definitions in Section 3). Then, Proposition 1 guarantees the existence of the Lagrange multipliers (we enforce similar proposition from Kokovin et al., 2011). Because of bunching the existence of the Lagrange multipliers has remained problematic so far in the screening theory. This question is important, because the multipliers are the key to finding solutions and in characterizing distortion in non-trivial situations. Furthermore, the relaxation enables us to completely characterize the class of possible graph structures for screening solutions: they are in-rooted acyclic graphs (Kokovin et al., 2011). The "envy-graph" of a solution is the list of its active incentive-compatibility constraints, perceived as arcs directed from the "envying" (almost eager to switch) agent to the envied quality-tariff bundle.

After some preliminaries, Theorem 1, without any essential restrictions on valuations, states that similar to SCC the direction of distortion is always governed by "envy" directed from a high-demand consumer to a low-demand consumer, the lower types get the distorted bundles and the higher types enjoy the informational rent. Extending this result from SCC to the general case may seem trivial. However, our necessary and sufficient conditions and the mistakes in the previous literature (discussed together with the theorem) show that there are complications. In particular, in defining "envy" the literature broadly confuses between "active" and "binding" con-

[5] Discrete types should be understood as approximation, what we may have in mind here is a continuous population interval served by a *continuous* interval of shops, continuity being approximated by many discrete points.

straints. The other complication arises in defining a "higher" demand consumer when valuations are non-ordered. For example, suppose an Internet provider serving several types of consumers designs a product line characterized by the traffic volume per month. Assume that the adults' maximal valuation for the first minute (or the chock-price) is higher than for teenagers, but the latter are eager to consume more traffic. In this case: Who has the "higher" demand? Is such family of preferences vertical or horizontal? We suggest it to be judged by the market outcome: "horizontal" market should mean no envy *at the solution*, whereas "vertical" one means a linear structure of envy. From our propositions one can see that no envy is the outcome, if and only if, the peaks of net-of cost valuations are not strictly below each other, and it is (generically) the only case when overall efficiency and zero informational rent appear. However, we believe that real life rarely provides such clear-cut horizontal or vertical outcomes and this motivates our study.

Having above in mind, in Section 5 we supplement the known relaxed Spence-Mirrlees conditions (see Araujo and Moreira, 2010) with one more. Our definition of "Hotelling-Spence-Mirrlees preferences" includes both polar cases, vertical and horizontal, and bridges them together with all intermediate realistic situations. This family of preferences can be called "spatial," because each consumer type is characterized mainly by her bliss point and her personal reservation utility—personal outside option. Under such (not too specific) restriction on preferences, the graph structures become much more specific than "all in-rooted acyclic graphs", revealed without the restriction. Namely, in a one-dimensional quality space the graphs are shown to be linear or weakly linear (Theorem 2), and this not only is true for profit-maximizing solutions, but also for any incentive-compatible plan. This enables us to understand envy-structure of socially-efficient or oligopolistic solutions. Similarly, in a two-dimensional quality space Theorem 3 establishes that any incentive-compatible envy-graph is a planar one, i.e., the arcs of envy on the plane do not cross. Thereby, the distortion caused by envy is transferred only to the neighbor of any bundle, and the direction of distortion becomes understandable. Corollaries to Theorems 2 and 3 state that the profit-maximizing solutions under monopoly should have the centrifugal pattern of distortion—"from the hills to the valleys", i.e., from the locations (bliss points) of high-willingness-to-pay consumers towards the areas with lower willingness-to-pay consumers (see Figures 2, 3). This location pattern looks counter-intuitive because it means that generally the service-points should be biased towards low-income areas away from the high-income areas.

Generally, our examples and ideas express our doubts in the applicability of strict SCC or purely horizontal preferences to any real-life product lines. Moreover, our approach opens a question for empirical economists. Which product line observed in real markets relates to what type of solution structure, and where can efficiency/distortion be a plausible diagnosis? We add that our findings for monopolistic screening can be extended to mechanism design problems and other situations with incentive compatibility, because the envy-graphs methodology developed here applies there as well (see Vohra, 2008).

Section 2 formulates the screening model with relaxation, Section 3 presents our approach to graph theory in screening and the background results: no-bunching and no cycles under relaxed constraints. Section 4 presents the general results related to efficiency, distortion and informational rent, for any types of preferences. Section 5 studies the specific solution properties under the "spatial" preferences: specific graphs, patterns of distortion and examples. Section 6 concludes and the Appendix contains some proofs.

2 Model

Our discrete screening model is somewhat more general than the standard one, because the restrictions on functions are relaxed, a constraints-relaxation parameter is added and multiple outside options are allowed. We formulate the model for a monopolistic seller, but have in mind all other usual interpretations and applications of screening, including principal-agent relations, Pareto-efficient allocations, etc. (see Rochet and Chone (1998) and Rochet and Stole (2003)). Moreover, we expect the structures of incentive-compatible solutions to be similar in the other areas of mechanism design and not just screening.

Consumer types are indexed by $i \in I^N = \{1, ..., N\}$; and $m_i > 0$ is the *frequency* of type i, which can be either the probability to participate in the market, or the total number or mass of such agents (consumers). Multiple agents of the same type can also mean multiple purchases by one individual. The quantity- or the quality-tariff bundles are denoted by (x_i, t_i), where $x_i \in X$ denotes the l-dimensional vector of attributes of the bundle purchased by the agent i. Here $X \subset \mathbb{R}^l$ denotes a consumption set, which can be discrete or continuous, and the product of such sets is $X^N = X \times X \times ... \times X \subset \mathbb{R}^{Nl}$. When $0 \in X$, this zero bundle may denote the common outside option which is non-participation, otherwise outside options may be multiple. Tariff t_i is the monetary transfer from consumer i to the firm. We assume quasi-linear utility functions-??

$$U_i(x_i, t_i) = V_i(x_i) - t_i,$$

where V_i is the monetary valuation of a purchase. In the particular case of a common outside option of non-participation $0 \in X$, valuations can be normalized as $V_i(0) = 0$. For a more general case we assume $k \geq 1$ outside options which are some fixed quantity-tariff bundles produced by other firms and non-participation amounts to outside options set $K \equiv \{(a_1, b_1)..., (a_k, b_k)\} \ni (0, 0)$ available to each consumer (see Figure 2). For some propositions we additionally assume differentiability, but otherwise do not restrict V_i, X.[6]

A monopolist selects a subset $I^n \subseteq I^N$ of $n \leq N$ types of consumers to be served and offers a product or a service using a menu of several packages of different quantities or qualities at some fixed tariffs on a take-it-or-leave-it basis (under $0 \in X$ the monopolist can set $n \equiv N$ and just assign $x_i = 0$ to agents not served). Afterwards the agents self-select. The seller knows the possible characteristics of the types and their probabilities but cannot discriminate personally. The cost function is *quasi-separable*:

$$C(m, x) = f_0 + \sum_{i \in I^n} m_i c(x_i),$$

where $f_0 \geq 0$ stands for some fixed cost and $c(\cdot) : \mathbb{R}^l \to \mathbb{R}$ is the cost function per-package.[7] We use the standard assumption that the producer designs only one package for each type, thereby plans an *assignment*, $(x, t) = \{(x_i, t_i)\}_{i \in I^n}$, and from the equivalent choices an agent selects whatever the principal prefers (friendly behavior). The profit π is the difference between the total tariffs and the total costs. After introducing a constraint-relaxation parameter $\rho \geq 0$

[6]Weak restrictions on X and V allow us to model many interesting and realistic situations, for example, satiable demands and discrete characteristics. Positive consumption and tariffs can be modelled through positivity restrictions on X, V. Instead, decreasing valuations V_i or negative quantities x_i are appropriate for modelling efforts spent in a principal-agent setting. By treating multiple outside options as options offered by the competitor(s) allows us to analyze oligopolistic markets.

[7]As shown in Kokovin et al. (2010, 2013), too general cost functions, including convex ones (decreasing returns), sometimes can undermine the applicability of the screening setting.

for technical reasons, we can formulate the seller's *relaxed assignment-optimization* program as follows.

$$\pi(x,t,\rho) \;=\; \sum_{i\in I^n} m_i t_i - C(m,x) \to \max_{I^n\subset I^N,(x,t)\in(X^n,\mathbb{R}^n)}, \textbf{ s.t.} \quad (1)$$

$$V_i(x_i) \;-\; t_i+\rho \ge V_i(x_j)-t_j \;\; (\forall i\in I^n, \forall j\in I^n\setminus\{i\}), \quad (2)$$

$$V_i(x_i)-t_i \;\ge\; V_i(a_l)-b_l \;\; (\forall i\in I^n, \forall l\in K). \quad (3)$$

Here (2) and (3) represent the incentive-compatibility (IC) constraints, and the participation constraints respectively. A plan (x,t) satisfying (2)–(3) is called *ρ-feasible*. The admissible set for (x,t) defined by these constraints is denoted as $Z(\rho)\subset(X^n,\mathbb{R}^n)$.

A solution $(\bar{x},\bar{t})$ to the problem (1)–(3) under $\rho=0$ is the standard *screening solution*. More generally, under $\rho\ge 0$ a solution $(\bar{x},\bar{t})$ to (1)–(3) is called here a *relaxed ρ-specific solution*, or just a *ρ-solution*.

The main focus of our study further is on ρ-solutions with $\rho>0$, because relaxation implies acyclic solution graphs, without sacrificing modelling of reality (under small ρ).[8] Moreover, we have found (see Kokovin et al., 2011) that when $\rho\to 0$, the relaxed solutions converge to the non-relaxed solutions.

To complete the setting, it should be added that under a quasi-separable cost $f_0+\sum_{i=1}^n m_i c(x_i)$, it is possible and standard to normalize. It means considering the normalized net-of-cost valuations $v_i(x_i)=V_i(x_i)-c(x_i)$ or social surpluses instead of the initial valuations, and seek for net-of-cost tariffs $\tau_i=t_i-c(x_i)$, or per-package profits τ_i.[9] Similarly, $u_{0i}\equiv\max_{l\in K}\{V_i(a_l)-b_l\}$ becomes the reservation utility of each consumer, the seller cannot serve her by giving less utility. Then the initial screening problem (1)-(3), obviously, amounts to the *normalized screening program* to be studied further:

$$\tilde{\pi}(x,\tau,\rho) \;=\; -f_0+\sum_{i=1}^n m_i\tau_i \to \max_{I^n\subset I^N,(x,\tau)\in(X^n,\mathbb{R}^n)}, \textbf{ s.t.} \quad (4)$$

$$v_i(x_i) \;-\; \tau_i+\rho\ge v_i(x_k)-\tau_k \quad (\forall i\in I^n, \forall j\in I^n\setminus\{i\}), \quad (5)$$

$$v_i(x_i)-\tau_i \;\ge\; u_{0i} \;\; (\forall i\in I^n, \forall l\in K). \quad (6)$$

3 Graph notions, graph structures and Lagrange multipliers

Now we introduce some graph theory notions and our approach to applying them to screening and incentive-compatibility problems. The terminology and the methodology are not standard so far. For example, Brito et al. (1990) speak of eliminating "cycles of *binding* incentive constraints among separated types," some different terminologies appear in Guesnerie and Seade

[8]Economically speaking, a relaxation parameter ρ can be interpreted as the "cost of switching" for the agent i from her usual package (x_i,t_i) to some new package k. One could try to make ρ negative instead of our $\rho\ge 0$, for modelling a premium to the agent for not switching and designing a *strictly* incentive-compatible menu that ensures strictly-dominant-strategy implementation of solutions. Unfortunately, $\rho<0$ does not exclude dicycles, and often undermines the existence of solutions.

[9]It is worth recalling that welfare-maximizing screening under restriction on total costs is an equivalent problem, reciprocal to profit-maximization (see e.g. Brito et al., 1990, Rochet and Stole, 2003).

(1982), in Vohra (2008) and others. More importantly, mixing binding with active constraints is rather common in the screening literature (see Brito et al. (1990), Rochet and Stole (2003) and Andersson (2005)), even though the distinction matters as we show in this section. We mainly follow Rochet and Stole's terminology, except for the term "binding", a reversed direction of arcs and our new notions. First we define the terms; relate graphs to screening and then motivate our approach.

Standard terms for digraphs. A *directed graph* or *digraph* G (hereafter just "graph") is a collection of nodes (vertices) denoted as $i \in G$ and of arcs (oriented edges) $(i,j) \in G$. Each arc, denoted as $i \to j$ or equivalently (i,j), describes an active constraint of our screening problem so that multiple arcs in direction i, j and loops (i,i) are excluded. In each $i \to j$, the arc's tail i is the adjacent *predecessor* of j, and the arc's head j is the adjacent *successor* of i. A *source* is a node without predecessors (with 0 in-degree). A (local) *sink* is a node without successors (0 out-degree). If the sink is unique and is reached from all nodes, it is called an *in*-root or, hereafter, just *root* of this (rooted) graph. A node without adjacent arcs is *disconnected*. A *walk* is a sequence of adjacent nodes and edges $\{i_1, e_{12}, i_2, e_{23}, i_3, ..., i_n\} = \{i_1 \to i_2 \to i_3 \to ..., i_n\}$; a *path* is a directed nonempty walk with distinct nodes, i.e., not a loop (not $i \to i$). When there is a unique directed path from any node to the root, then this graph is called an *in-tree*, hereafter just a *tree*, and the simplest tree is a star $\{i_1 \to i_0, i_2 \to i_0, ..., i_n \to i_0\}$. A s*panning-tree* of graph G is a subgraph—a tree containing all nodes of G. An (*in-*)*rooted* graph is a digraph with a unique sink (in-root) when this root is reachable from every node through a path. Obviously, any in-rooted graph contains one or more spanning-trees. A closed directed path $\{i_1 \to i_2 \to i_3 ... \to i_1\}$ is a *dicycle,* and a digraph is *acyclic* if there are no dicycles. A partial order among nodes $i_1, ..., i_n$ can be viewed as an acyclic digraph when order relation $i \succ j$ is equivalent to arc $i \to j$ ... In addition, the following notions and the notion of *preorder* defined in Appendix.

New terms: rivers and flows. Any in-rooted acyclic digraph is called a *river*. Obviously, all trees are rivers but the latter may also contain *bypasses* defined as two directed paths $(i_1 \to i_2 \to ... \to i_k)$, $(i_1 \to i_3 \to ... \to i_k)$ with the same source and the same sink (see Fig.1 below for illustration). *A flow-graph* in our context is a 2-colored digraph such that *all* sinks and maybe some other nodes are colored as *drains,* the remaining nodes becoming *non-drains.* Obviously, after connecting all drains of any flow-graph "to" some additional node (root), this flow-graph becomes a river. Thereby each acyclic flow-graph can be perceived as a river without its root. There is one-to-one correspondence between rivers and acyclic flow-graphs.

We call a digraph a *directed chain* when it consists of unique directed path $\{i_1 \to i_2, \to, ..., i_n\}$ having all nodes distinct (no repetition or branching). We call a graph (piece-wise) *linear* or when each of its connected component is a chain or has an underlying undirected chain.

Graphs application in screening. In applying graphs to screening, all agents' identities #1,...,#n are treated as nodes whereas constraints are interpreted as *envy* arcs within a related *envy graph.* In this graph, the non-participation option is considered as an additional node with the label #0. It must succeed all sinks and can succeed other nodes. More precisely, our optimization program (1)–(3) has $n \times (n-1) + n = n^2$ inequalities and all can become active, i.e., equalities. For any feasible plan (x,t) we define its *envy A-graph* $\bar{\bar{G}}(x,t)$ as the list $\bar{\bar{G}}(x,t) = \{(i_1,j_1),(i_2,j_2),...\}$ of all constraints that are active at (x,t) (double-bar over G highlights equalities as the basis of definition and for a non-feasible plan (x,t) we similarly define the strict-envy graph $G^{<}(x,t)$ as the list of all violated constraints). The direction of any active constraint (i,k): $V_i(x_i) - t_i \geq V_i(x_k) - t_k$ is represented as an arc $(i \to k)$ going *from i to k*, i.e., in the direction of a possible choice of consumer switching. It means that an agent i (weakly) envies package $\#k$, being indifferent between her package and $\#k$, almost eager to switch to $\#k$. The opposite direction of arcs, chosen in Rochet and Stole, seems inconvenient for this interpretation and for the use of "flow networks" in screening. Finally, the notation $\bar{\bar{G}}_{-0}(x,t) = \bar{\bar{G}}(x,t) \setminus \{\#0\}$ means further the unrooted graph, where the root node $\#0$ is deleted, but the related arcs remain as the indicator of "drain." Thereby, this graph $\bar{\bar{G}}_{-0}(x,t)$ is the flow-scheme uniquely related to the plan (x,t).

Fig. 1 illustrates these notions through an example violating SCC but having the common outside option $(0,0)$. Three valuations $\{v_1(x), v_2(x), v_3(x)\} = \{(2x - 2x^2),\ x - 0.75x^2,\ (0.72x - 0.36x^2)\}$ are shown in green. Their peaks (marked by the green squares) are the bundles $(x,t) = \{(0.5, 0.5),\ (0.666667, 0.33333),\ (1.0, 0.36)\}$ which are the first-best for the monopolist when the IC constraints are ignored. In contrast, the red circles show the actual profit-maximizing solution $(\bar{x}, \bar{t}) = \{(0.5, 0.432255),\ (0.725275, 0.330757),\ (1.0, 0.36)\}$. One can see a rightward distortion at $\bar{x}_2 > 0.666667$ and consumer surplus for the agent $\#1$: $\bar{t}_1 = 0.432255 < 0.5$. The A-envy graph (in red) results from the *active* indifference curves that connect the envying bundle and the envied bundle. Here it is a river with root (0,0), namely, $\bar{\bar{G}}(\bar{x}, \bar{t}) = \{1 \to 2, 3 \to 2, 2 \to 0, 3 \to 0\}$. But, is it also the graph of *binding* constraints (those that influence the optimal value when relaxed or eliminated)? No, the participation constraint $3 \to 0$ is excessive, it became active just occasionally.

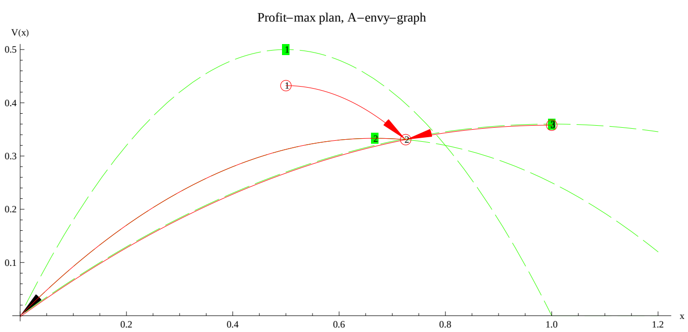

Figure 1: How A-envy-graph results from a solution.

We now introduce B-graphs and LA-graphs related to solutions, and explain their relationships to A-graphs under our ρ-relaxation.

First note that even under concave valuations V a screening problem (1)-(2) is typically non-convex. It is so because concave functions enter into both sides of the inequalities. Therefore, for any non-convex optimization, a distinction becomes important between an active constraint and a *binding* constraint—the one which influences the optimal value when relaxed or eliminated. Generally, a binding constraint need not be active and an active one need not be binding, see example (7) below. So, screening may also need B-graphs representing all binding constraints, not only A-graphs.

In addition, there could also be a need for a LA-graph, which is defined as the list of all LA-constraints—those having strictly positive Lagrange multipliers (see our Proposition 1). This LA-graph generally may differ both from A-graph and B-graph, and even from their intersection. The typical reason for the discrepancy among these graphs is due to the so-called *bunching* situation. Bunching means identical packages $(x_i, t_i) = (x_j, t_j) = ...$ are assigned to different agents i, j,... at the optimum. Such an outcome is known to be quite a regular case in standard screening with $\rho = 0$, see Rochet and Chone (1998) for a thorough treatment of bunching. In a bunch, naturally, all the bunched agents do envy each other, thereby creating a dicycle in the A-graph $\bar{\bar{G}}(x,t)$ and an "over-constrained" situation. Bunching and more general dicycles create major hardships in characterizing and finding solutions, mainly because the usual constraint-qualification conditions fail and then the existence and finding the Lagrange multipliers become problematic.

In contrast, under positive relaxation ($\rho > 0$), dicycles and bunching among predecessors and successors are excluded in A-graphs as shown in Lemma 2 below. The Lagrange multipliers do exist and most often become unique. Additionally, based on our experience with solutions, (only) under positive relaxation, A-graph "almost always" coincides with the LA-graph. The latter is most useful one for solution characterization, whereas the former is more easily observable at any admissible plan.

To appreciate the difference between A, B, LA constraints and related hardships with characterizing optima, consider a simplest over-constrained non-convex example, where the constraints display all three kinds of importance:

$$\max x \in \mathbb{R} \qquad \text{s.t. } (i): \;\; x^2 \geq 1, \;\; (ii): \;\; x^4 \geq 1, \; (iii): \; x \leq 0. \tag{7}$$

Clearly, here the optimum is $\bar{x} = -1$, and the constraint (iii) is binding, because it cannot be dropped and keep the optimum intact, but (iii) is not active or LA. In contrast, the two constraints (i) and (ii) are active but not binding, because any one of these two constraints can be removed without changing the solution. Each can either be LA or not, because any Lagrange multipliers $\lambda_A, \lambda_B \geq 0$ such that $\lambda_A + \lambda_B = 1$ are admissible. Unfortunately, none of these multipliers λ_i reflect the sensitivity of the objective function to the related constraint, as it should. However, for a small price for accuracy, we can exclude this indeterminacy and weakness of λ_i. We can remove the over-constrained situation by slightly relaxing one of the constraints, (i) or (ii). Such harmless trick is common in linear programming to exclude cycles.

In screening, like in linear programming, our ρ-relaxation helps to overcome all over-constrained situations and cycles. This discussion motivates our focus mainly on the *relaxed* screening problems and on envy A-graphs $\bar{\bar{G}}(x,t)$. Hereafter, what we have in mind is these kind of graphs when we drop "A" and mention just envy graphs.

3.1 Background facts on solution structures: all envy-graphs are rivers

Now we repeat from Kokovin et al. (2011) the necessary lemmas on solution structures.

The lemmas below state the most general properties of the solution structures, guaranteed solely by quasi-linearity of utilities.

LEMMA 1: (IN-ROOTED ENVY-GRAPH). *For any ρ-solution $(\bar{x},\bar{t})$ its envy-graph $\bar{\bar{G}}(\bar{x},\bar{t})$ is in-rooted, i.e., each node i is connected to the root (#0) by a directed path $i \to ... \to 0$. Thus, $\bar{\bar{G}}(\bar{x},\bar{t})$ contains a spanning-tree.*

LEMMA 2: (PROFITS ORDER). *Take any ρ-solution $(\bar{x},\bar{t})$ under quasi-separable costs ($C(m,x) = f_0 + \sum_{i=1}^{n} m_i c(x_i)$ ($f_0 \geq 0$)), then:* (i) *the profit contribution $\tau_i = t_i - c(x_i)$ from any agent is not lower than the contribution from any of her successor in the envy-graph, i.e., $i \to ... \to j \Rightarrow \;\bar{\tau}_i \geq \bar{\tau}_j$,;* (ii) *under ($\rho > 0$) this inequality is strict: $i \to ... \to j \Rightarrow \bar{\tau}_i > \bar{\tau}_j$, and for the adjacent couples $i \to j$ it has the particular form $\bar{\tau}_i \geq \bar{\tau}_j + \rho$, whereas bunching among predecessors and successors ($x_i = x_j$) and other dicycles are excluded.*

The above two lemmas imply the following lemma on acyclic solution structures.

LEMMA 3: (ENVY-GRAPHS ARE RIVERS).[10] *For any ρ-solution $(\bar{x}, \bar{t})$ to a screening problem with quasi-separable costs and positive relaxation $\rho > 0$, its envy-graph $\bar{\bar{G}}(\bar{x}, \bar{t})$ is a river.*

Note that bunching ($x_i = x_j$) among the predecessors and the successors is excluded, it remains possible only for the disconnected packages that coincide accidentally.[11] Unlike the usual bunching, the accidental bunching can be ignored because it has no impact on characterizing solutions.

Proposition 2 in Kokovin et al. (2011) shows also that all rivers can be envy-graphs under some valuations and this class of graphs is enumerated.

3.2 Solution characterization through FOC and Lagrange multipliers

Now we show how one can use the envy-graphs in characterizing and finding solutions. Since the existence of multipliers is guaranteed under positive relaxation $\rho > 0$ (a big reward of the relaxation), relying on Lemmas 1-3, in Kokovin et al. (2011) under relaxation $\rho > 0$ any optimal solution with its first-order conditions, i.e., Lagrange multipliers can be characterized. We expand this proposition now onto the case $\rho \geq 0$, by denoting λ_{i0} the Lagrange multiplier for the i-th participation constraint and λ_{ij} for the $i \to j$ incentive-compatibility constraint.

PROPOSITION 1 (FOC CHARACTERIZATION). *Assume quality space $X = \mathbb{R}^l$, quasi-separable costs, continuously differentiable net valuations v_i bringing positive net surplus somewhere: $\exists x : v_i(x) > u_0$.*

(a) *When qualities bringing positive net surplus $\{x : v_i(x) > u_0\}$ are bounded, then a solution to the normalized problem (4)-(6) does exist.*

(b) *Additionally, when the solution $(\bar{x}, \bar{\tau})$ is unique, then:* (i) *There exist some Lagrange multipliers $\lambda = (\lambda_{1,0}, \lambda_{1,2}, ..., \lambda_{n,n-2}, \lambda_{n,n-1}) \in \mathbb{R}_+^{n*n}$, satisfying the following first-order conditions of Lagrangian $\mathcal{L}(.)$ and supplementary inequalities for finding $(\bar{x}, \bar{\tau}, \lambda)$ from a hypothetical LA-graph G_+^λ:*

$$\frac{\partial \mathcal{L}(\bar{x}, \bar{\tau}, \lambda)}{\partial \tau_i} = m_i - \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} + \sum_{k \in P_i^{ad}(G_+^\lambda)} \lambda_{ki} = 0 \ \forall i > 0, \tag{8}$$

$$\nabla_{x_i} \mathcal{L}(\bar{x}, \bar{\tau}, \lambda) = \nabla_{x_i} v_i(\bar{x}_i) \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} - \sum_{k \in P_i^{ad}(G_+^\lambda)} \lambda_{ki} \nabla_{x_i} v_k(\bar{x}_i) = 0; \forall i > 0, \tag{9}$$

$$0 = v_i(\bar{x}_i) - \bar{\tau}_i - v_i(\bar{x}_j) + \bar{\tau}_j + \rho_{ij} \ \ \forall (i,j) \in G_+^\lambda, \tag{10}$$

$$0 \leq v_i(\bar{x}_i) - \bar{\tau}_i - v_i(\bar{x}_j) + \bar{\tau}_j + \rho_{ij} \ \ \forall (i,j) \notin G_+^\lambda, \text{ where} \tag{11}$$

$$G_+^\lambda = \{(ij) | \lambda_{ij} > 0\}, \ \ \bar{x}_0 := 0, \ \ \bar{\tau}_0 := 0. \tag{12}$$

(ii) *The Lagrange multipliers of the constraints successive to any i are bounded as*

[10] Reducibility of cycles in A-graph of the main problem (with more restrictions on v_i, C than here) was proven in Guesnerie and Seade (1982) through the same simple Lemma 2, and is repeated in subsequent papers.

[11] Such solution can be called *regular*; bunching is excuded among prdecessors and successor (it may occur only ocassionally among nodes not connected by a path).

$$\sum_{j\in S_i^{ad}(G_+^\lambda)} \lambda_{ij} \leq M_i^{PG_+^\lambda} := \sum_{j\in P(i,G_+^\lambda)\cup\{i\}} m_j \qquad \forall i; \tag{13}$$

moreover, when the river G_+^λ is a tree, the positive multiplier for the unique successor of i is found as

$$\lambda_{is_i^1(G_+^\lambda)} = M_i^{PG_+^\lambda}.$$

Proof: see Appendix.

In essence, the proposition above provides FOC and a method for practically finding solutions under relaxation, though it does not formulate a sufficient condition for the optima, it only gives the *necessary* one. Typically the reason is a *non-convex optimization* in screening, even under strictly concave net valuations $v_i(\cdot)$ (see Section 3). Therefore, for finding a solution through this characterization, one should explore *all* possible rivers G_+^λ, and then compare profits from these locally-optimal solutions. So far, this method is the only practical way for arbitrary valuations, and Proposition 1 provides justification for it.

Interestingly for finding the solutions, any screening problem can be interpreted as a "flow-network". In our context it is an acyclic flow-graph F supplemented with incoming flows $m_i \geq 0$ assigned to all nodes, ultimate out flowing magnitudes $\lambda_{jj} < 0$ assigned to certain nodes (drains), and current-flow magnitudes $\lambda_{ij} \geq 0$ assigned to all arcs. Then the equation (8) is interpreted as the balance of inflows to and outflows from each node. Respectively, the following "conservation law" $\sum_{i\in F} m_i = \sum_{j\in F} \lambda_{jj}$ holds (see Vohra (2008), Berg and Ehtamo, 2010)), i.e,, the whole network have balanced inflows and outflows. This interpretation provides an interesting analogy from physics and helps one to understand why the Lagrange multipliers are bounded from the above in claim (ii).

4 Distortion as a result of envy

"For where you have envy..., there you find disorder and ... evil practice." /James 3:16/

Under the usual SCC, it is a common knowledge that whenever a bundle is envied it is distorted and conversely, a bundle free of envy is free of distortion. Economic intuition suggests that such equivalency should hold also without SCC. This section generally supports this conjecture but with some cautions, and provides rather comprehensive results on distortion. They turn out to be dependent on the above analysis.

We use the following definition of distortion, rather standard for a separable screening problem like (4)–(6).[12]

Definition: An allocation $\bar{x}_i$ designed for the i-th agent is called (*partially*) *efficient* or *non-distorted* when $\bar{x}_i$ maximizes the joint welfare of this agent and the principal, regardless of all other packages in the menu $(\bar{x}, \bar{\tau})$. That is,

$$\bar{x}_i \in \arg\max_{x_i\in R^l} v_i(x_i) = \arg\max_{x_i\in R^l}(V_i(x_i) - c(x_i)).$$

In the opposite case, the package and the allocation $\bar{x}_i$ are called *distorted* for the i-th agent (the same quantity $\bar{x}_i$ can be efficient for i, but distorted for some bunched $j : \bar{x}_j = \bar{x}_i$). The

[12]In contrast, without separability or/and quasi-linearity of utilities (as in Guesnerie and Seade (1982)), the distortion notion becomes tedious, dependent on other packages.

related bundle $(\bar{x}_i, \bar{\tau}_i)$ is also called distorted. An allocation $\bar{x}$ is called *overall-efficient* or socially the first-best, when $\bar{x}$ maximizes the total social welfare $\sum_i m_i v_i(x_i)$ without regard to incentive-compatibility constraints.

Now we formulate a sufficient condition for partial efficiency of a bundle in three different versions, only the third one is non-obvious.

PROPOSITION 2. *Consider a solution $(\bar{x}, \bar{\tau})$ to the problem (4)–(6), then: (i) Whenever all IC constraints $(j \to i)$ leading* ***to*** *consumer type i are non-binding (can be dropped without changing the optimal profit), then $\bar{x}_i$ is non-distorted;*

(ii) An allocation $\bar{x}_i$ of consumer type i would be distorted if and only if this distortion helps the principal to increase profit by relaxing the related IC constraint;[13]

(iii) Under convex X and concave net valuations, if a quantity-tariff bundle $(\bar{x}_i, \bar{\tau}_i)$ appears (strictly) inferior for other agents (i.e., there is no active IC constraint $(j \to i)$ leading ***to*** *an agent i in the A-graph $\bar{\bar{G}}(\bar{x}, \bar{\tau})$), then the bundle $(\bar{x}_i, \bar{\tau}_i)$ is non-distorted.*

Proof: see Appendix.

Surprisingly in (iii) there is the need for convexity/concavity assumption. Similar claim is proved within Proposition 2 of Guesnerie and Seade (1982) under strict concavity of utilities, locally-different consumers and one-dimensional X. However, their brief proof uses concavity only implicitly, which tempted Brito, Hamilton, Slutsky and Stiglitz (1990) to mistakenly drop the concavity assumption in their Proposition 3, repeated also in Andersson (2005) as Lemma 3 (using many dimensions). Such relaxation is incorrect as shown by our counter-example, Example 1 below. The reason for the mistakes was the confusion between active and binding constraints. Indeed, for claim (i) or its version (ii) no concavity assumptions are needed, unlike the sufficient condition (iii).

EXAMPLE 1. Let two agent types have equal frequencies $m_1 = m_2$ and the net valuations $v_1 = \max\{4 - 4(1-z)^2, \min\{4z, 4, 8-z\}, 5 - (5/16)(4-z)^2\}$, $v_2 = 7 - (7/16)(4-z)^2$. Here v_2 is concave but v_1 is only quasi-concave. One can check that the socially efficient quantities are: $x_1 = 4$, $x_2 = 4$. However, the profit-maximizing menu is $(\bar{x}_1, \bar{\tau}_1) = (1, 4)$, $(x_2, \tau_2) = (4, 7)$ with profit equal to 11, and it has no active IC constraints, though constraint $\#2 \to \#1$ is binding, preventing a better incentive-incompatible plan $(x_1, \tau_1) = (4, 5)$, $(x_2, \tau_2) = (4, 7)$. A similar socially efficient incentive-compatible plan $(x_1, \tau_1) = (4, 5)$, $(x_2, \tau_2) = (4, 5)$ brings less profit, only 10, compared to $(\bar{x}, \bar{\tau})$. Thus, "only participation constraints active" is not a sufficient condition for overall efficiency without concavity or strict quasi-concavity.

Now we turn to more complicated *necessary and sufficient* conditions for distortion in terms of active or LA constraints. These can be formulated as *aggregate envy* to a given packageas follows.

ASSUMPTION DC: The net-valuations v_i are continuously differentiable and concave on an admissible space $X = \mathrm{R}^l$. The solution $(\bar{x}, \bar{\tau})$ studied is characterized by the first-order conditions (8-12), the set of admissible multipliers supporting this solution is denoted as $\Lambda = \Lambda(\bar{x}, \bar{\tau})$.

THEOREM 1 (DISTORTION AND LAGRANGE-ACTIVE CONSTRAINTS): *Let the assumption* (DC) *hold at some solution $(\bar{x}, \bar{\tau})$. If the gradients of valuations satisfy the inequality*

$$\sum_{k \neq j_0} \lambda_{kj_0} \nabla v_k(\bar{x}_{j_0}) \neq 0 \tag{14}$$

[13]We are grateful to Larry Samuelson for suggesting this very intuitive formulation/ interpretation.

for all supporting Lagrange multipliers $\lambda \in \Lambda(\bar{x}, \bar{\tau})$, then the package $(\bar{x}_{j_0}, \bar{\tau}_{j_0})$ is distorted. Conversely, when this relation becomes an equality for some supporting $\lambda \in \Lambda(\bar{x}, \bar{\tau})$, then this package is non-distorted.

COROLLARY (DISTORTION DIRECTION):[14] *Suppose that only one agent k shows LA-envy towards a package $(\bar{x}_{j_0}, \bar{\tau}_{j_0})$ in the sense $\min_{\lambda\in\Lambda} \lambda_{kj_0} > 0$, $\lambda_{ij_0} = 0 \, \forall i \neq k$, and argmaxima for these two net-valuations do not coincide: $\arg\max_z v_{j_0}(z) \neq \arg\max_z v_k(z)$. Then the allocation $\bar{x}_{j_0}$ is distorted. Moreover, for a unidimensional commodity ($l = 1$) a bigger envying package ($\bar{x}_k > \bar{x}_{j_0}, k \to j_0$) implies that the envied package $\bar{x}_{j_0}$ is undersized ($\bar{x}_{j_0} < \arg\max_{z\in R} v_{j_0}(z)$), and the opposite relation ($\bar{x}_k < \bar{x}_{j_0}, k \to j_0$) implies an oversized package $\bar{x}_{j_0}$.*[15]

Proof: see Appendix.

To complete the efficiency analysis under non-specific valuations, we should mention two simple, generally known in the literature, facts that follow from Lemma 2 and Theorem 1:

(1) *Overall distortion cannot occur, at least one bundle is efficient*; (2) *Assume convex X, concave v, and only the participation constraints being active (i.e., A-graph being a "star"), then overall efficiency results.*

Again, the need for concavity/convexity here is surprising but supported by the same counter-example (Example 1). All the claims above are illustrated by our Figures 1, 2 and 3.

5 Special Case: Spatially Heterogeneous Population

In this section we assume a one- or a two-dimensional continuous real space X of quality/quantity characteristics, and population having a "spatial" structure. It means that all agents have the same or approximately the same shape of net-of-cost valuations v_i, but each agent i has her individual (socially efficient) bliss point $b_i = \arg\max_{x\in X} v_i(x) \in X$ and her individual maximum $h_i = \max_{x\in X} v_i(x) = v_i(b_i)$ the height of the valuation that she can pay if participating. Based on such parametrization in a one- or a two-dimensional space X of characteristics, we impose a restriction on the family of preferences that can be called a Hotelling-Spence-Mirrlees condition. Then, at the solution, the population of the agents gets partitioned into groups, each group being ordered in the spirit of Spence-Mirrlees single-crossing condition. Such regularity allows to reduce the domain of possible solution structures dramatically to very simple (linear or planar) classes of graphs.

ASSUMPTION HSM. We assume that the net valuation functions $v_i(\cdot)$ of all agents are continuous, strictly concave and have non-coinciding bliss points $b_i = \arg\max_{z\in \mathrm{R}} v_i(z)$ such that for any couple of types i, j and the direction $\Delta = b_j - b_i \in X$, the valuations satisfy the non-normalized single-crossing condition:

$$\frac{v_i(z+\Delta) - v_i(z)}{|\Delta|} < \frac{v_j(z+\Delta) - v_j(z)}{|\Delta|} \; \forall z, \tag{15}$$

so that along this direction the difference $v_j(z) - v_i(z)$ is a strictly increasing function, whereas these two curves cannot intersect more than once.

In a one-dimensional space $X = R^1$, this condition differs from the usual SCC only in the sense that normalization is absent, i.e., we do not require $v_i(0) = 0$, because valuations now

[14] For applicability of the Kuhn-Tucker and Envelope Theorems see Appendix.

[15] Compare our claim (ii) with earlier special cases, namely with examples in Andersson (2008), showing ambiguous direction of distortion.

need not intersect at 0, or even need not intersect at all. We shall need also a more special assumption of this kind, amounting to "spatial" preferences, formulated as follows.

Assumption HSM+. Let all agents' net valuation functions $v_i(\cdot)$ be generated by some common strictly convex non-negative *distance function* w, such that

$$v_i(z) \equiv h_i - w(|b_i - z|) \ \ \forall i, \ \max_{z \in X} w(z) = w(0) \equiv 0$$

where b_i are the bliss points and h_i denote the highest possible net tariff. Thus, agents differ only in their bliss points and the demand heights, but not in the shape of their valuations.

These two versions of "spatial" preferences have a clear interpretation in one-dimensional space as follows.

5.1 One-dimensional quality

The example of quadratic net valuations like $v_i(z) = h_i - (b_i - z)^2$ helps to explain how the preferences under Spence-Mirrlees's are bridged with Hotelling's under HSM+ preferences. When specific heights are $h_i = b_i^2 \ \forall i$, this quadratic preference profile $v_i(z) = 2b_i z - z^2$ becomes *vertical*, i.e., it satisfies the usual normalized Spence-Mirrlees condition[16]. In contrast, the Hotelling's *horizontal* preference family is defined by equal heights $h_i = h_j \ \forall i, j$. These two distinct classes have contrasting properties: the well-known outcome of vertical profile is its *linear* solution graph $\bar{\bar{G}}$, which is a single path from n to 1. It happens because each agent i *can envy only her left neighbor* $i-1$ and nobody else. In contrast, as shown in Andersson (2008) and Nahata et al. (2003), the horizontal profile yields a simple star-graph of the solution— no one envies anybody.

We are ready now to bridge these two classic cases together and prove that, for the same reasons as for these two extremes (vertical and horizontal), all other preferences satisfying HSM+ or HSM also generate rather simple class of flow-graphs, which are linear but for bunching.

We defined above a *piecewise-linear* graph so that each of its connected component becomes a chain when we neglect the directions. Thereby it does not have any (directed or non-directed) cycles or branching. Now we modify this definition as follows.

Definition. A flow-graph $\bar{\bar{G}}_{-0}$ is called *weakly-linear* when its non-ordered underlying graph is *weakly-linear*, i.e., it becomes linear when any bunch of nodes is perceived as one node.

At the expense of one additional new notion, we call such graph a multi-centipede" because it consists of ordered connected chains connected either tail-to-tail or head-to-head. They always stay on their heads, since each sink is a drain (see Fig.2). The drains look like legs touching the ground. This analogy helps us to discuss solutions and their properties. All tails are non-distorted, the heads are almost-always distorted, and the intermediate segments are distorted always.

Now, for a given number of agents served, we prove weak linearity of our graphs through arguments similar to the usual SCC case. The difference lies only in a special treatment of bunching and varying directions of envy.

Theorem 2. *Assume* HSM *preferences in a one-dimensional space* $X = R$ *and agents ordered according to their bliss points* $b_1 < b_2 < ... < b_n$, *so that the difference* $v_{i+1}(z) - v_i(z)$ *is a strictly increasing function. Consider any incentive-compatible plan* $(\bar{x}, \bar{\tau}) \in R^{2n}$ *with* n *agents served under any relaxation* $\rho \geq 0$. *Then:*

[16]Similarly, for any function w, the heights are adjusted as $h_i : v_i(0) = 0$ to give SCC.

(i) *The order of the incentive-compatible qualities* $\bar{x}_i$ *weakly preserves the order of the bliss points in the sense* $\forall i, j: \; b_i < b_j \Rightarrow \bar{x}_i \leq \bar{x}_j$. *Moreover, under* $\rho > 0$, *strict inequality* $\bar{x}_i < \bar{x}_j$ *holds for all* i, j *connected in the A-graph (no bunching, except maybe for the disconnected neighbor sinks).*[17]

(ii) *The graph ordering is also predetermined by the bliss points; the solution flow-graph* $\bar{\bar{G}}_{-0}(\bar{x}, \bar{\tau})$ *is weakly linear, being positioned on the quality axis through connecting all distinct qualities* $\bar{x}_i \neq \bar{x}_j$ *by their envy-arcs* $i \to j$ *(if any) and joining equal quantities into a bunched node. Under* $\rho > 0$ *this graph is linear, moreover, each agent* i *cannot envy anybody except the two neighbors of types* $i-1$ *and* $i+1$.[18]

Proof. In our arguments, we exploit types numbering $b_i < b_j \Leftrightarrow i < j$.

(i) under $\rho = 0$, to check the types' order preservation for quantities for any i, j, we use their incentive compatibility constraints $v_i(\bar{x}_i) - \bar{\tau}_i \geq v_i(\bar{x}_j) - \bar{\tau}_j$ and $v_j(\bar{x}_j) - \bar{\tau}_j \geq v_j(\bar{x}_i) - \bar{\tau}_i$ that are satisfied at $(\bar{x}, \bar{\tau})$, and summarize them as: $v_j(\bar{x}_i) - v_i(\bar{x}_i) \leq v_j(\bar{x}_j) - v_i(\bar{x}_j)$. Comparing this inequality with (15) we reach the conclusions $\bar{x}_i < \bar{x}_j \Rightarrow b_i < b_j$ and $b_i < b_j \Rightarrow \bar{x}_i \leq \bar{x}_j$ because the difference $v_j(z) - v_i(z)$ is a strictly increasing function. Under $\rho > 0$ the logic is the same, the relaxation does not change it. Further, under $\rho > 0$ there is no bunching by Lemma 2 among the adjacent nodes: $\bar{x}_i \neq \bar{x}_j$. We postpone the remaining claim in the parenthesis because it needs the graph structure.

(ii) To check the order of the types is preserved in the graph, one can use the same logic of increasing differences $v_{i+1}(z) - v_i(z)$. We conclude that when the incentive-compatibility constraint $i \to i+1$ is satisfied at $(\bar{x}, \bar{\tau})$ then together with the satisfied constraints $i+1 \to i+2, \ldots$, it amounts to satisfying the envy constraint from i to any higher than $i+1$ type j (with $j \neq i$) as a strict equality: $v_i(\bar{x}_i) - \bar{\tau}_i > v_i(\bar{x}_j) - \bar{\tau}_j$. The same logic works for the lower types $i-2, \ldots$. Thus, all non-bunched with i non-neighbor types are not envied by i (moreover, these constraints can be eliminated from the initial problem and replaced by $x_1 \leq x_2 \leq \ldots \leq x_n$).

From the same logic it follows the absence of any free (disconnected) node x_k lying strictly between connected nodes $i \to j$: $\nexists k : \; x_i < x_k < x_j$.

We have thus found that among distinct nodes only the neighbors in peaks can be adjacent in the A-graph, and the nodes $\#1 < \#2 < \ldots < \#n$ are linearly ordered on the quality axis corresponding to $x_1 \leq x_2 \leq \ldots \leq x_n$. Thereby, connecting these points x_i with relevant arcs of envy from $\bar{\bar{G}}_{-0}(\bar{x}, \bar{\tau})$ we must get a (piecewise) linear graph on this axis, and (ii) is proved. What remains to be shown is that neighboring tails of the centipedes cannot be bunched by accident. Since they are the tops of the graph, they remain non-envied. This confirms that they are non-distorted, $x_i = b_i$ (see Corollary below for more details) and from the assumption of different peaks $b_i < b_{i+1}$ it follows their non-bunched quantities. □

To appreciate the reduction in graphs' variety that the assumption HSM brings, note that here active can be only ties among the neighbors: left arrow or/and right arrow in the graph ($i \to i+1$ or/and $i \leftarrow i+1$) and nothing else. Therefore, *the number of possible (linear)*

[17] Actually, as one can see from the proof, this claim "i" is true for any incentive compatible plan, optimal or not. Another enforcement is the claim that, for any valuations family v parametrized with the bliss points b_i and heights h_i, under $\rho > 0$ any (even nonessential) bunching is a zero measure case. That is, it appears with probability 0. To show this, it is sufficient to disturb the bunched bliss points or heights in any direction and the bunch disappears.

[18] It follows that when any two distinct nodes $\bar{x}_i \neq \bar{x}_j$ are adjacent in this graph ($i \to j$), there does not exist any b_k between the bliss points b_i, b_j ($\nexists \, b_k : \; b_i < b_k < b_j$ or $b_i > b_k > b_j$), or such an intermediary k is bunched with j: $\bar{x}_k = \bar{x}_j$.

solution graphs is only 3^{n-1}, much smaller than Proposition 2 predicts for the general case.

Now, using the proposition obtained and Theorem 1 about distortion, we get the natural conclusion about how distortion/efficiency depends upon the bundle's position in the graph. We consider some agent i's bundle (x_i, t_i) "envied from both sides," when arcs $x_{i-1} \to x_i$ and $x_{i+1} \to x_i$ are present in the A-graph.

Corollary. *If the plan* $(\bar{x}, \bar{t})$ *is profit-maximizing under* HSM *and valuations* v_i *are differentiable, then: (a) all non-envied nodes (sources) in* $\bar{\bar{G}}_{-0}(\bar{x}, \bar{t})$ *are non-distorted, (b) a node envied from both sides may or may not be distorted, (c) each node envied from one side with a positive Lagrange multiplier is distorted.*

Proof. Here (a) and (c) follow immediately from our theorem, but for the nodes envied from both sides we must show examples of distorted and non-distorted outcomes. It is sufficient to take valuations $v_1(z) = 2 - (-1 - z)^2$, $v_2(z) = 0.5 - (0 - z)^2$, $v_3(z) = 2 - (1 - z)^2$ with $m_1 = m_2 = m_3 = 1$ and a socially optimal plan $(\bar{x}_1, \bar{x}_2, \bar{x}_3) = (-1, 0, 1) = (b_1, b_2, b_3)$. Simple calculation show that it is also profit-maximizing, and the symmetry of left and right neighbors of agent #2 entails equality (14) in the form $\lambda_{12}\dot{v}_1(\bar{x}_2) + \lambda_{32}\dot{v}_3(\bar{x}_2) = 0$ and hence non-distorted $\bar{x}_2$ (in spite of envy and $\lambda_{12} > 0, \lambda_{32} > 0$). When we use the same logic in the reverse direction by introducing any asymmetry in this example, say, $v_3(z) = 2 + \varepsilon - (1 - z)^2$, we get distortion. □

It appears rather plausible from arbitrary $\varepsilon > 0$ here, that the envied but non-distorted bundles are the rare degenerate cases.

Now we should compare the variety of distortion outcomes with two classic polar classes: *vertical* and *horizontal* profiles of preferences. The former was already shown to be a special case of our assumption HSM+, the case generating the linear graph $n \to (n-1) \to ... \to 1 \to 0$ with distortion everywhere except n. The latter profile is another special case with uniform heights of valuations: $h_1 = h_2 = ... = h_n$. It generates the disjoint graph $\bar{\bar{G}}_{-0}$ resulting in overall efficiency of bundles. More generally, even without the uniform heights, a profile can be called *quasi-horizontal* when it generates the disjoint graph $\bar{\bar{G}}_{-0}$, and one can easily realize that for any function w there is a non-degenerate region of parameters b, h that generates overall efficiency. However, our class HSM+ includes many other interesting outcomes in addition to these polar two, and appears almost as tractable as these two. We illustrate such analysis and possible distortion outcomes by Example 2 below.

Example 2. Consider a product line (x, τ) designed for the population of 6 consumer groups. Let the frequencies (sizes of the subpopulations) $m \equiv (m_1, ..., m_6) = (1, 2, 1, 1, 1, 1)$ and quadratic valuations $v_i(x_i) \equiv h_i - 0.2(b_i - x_i)^2$ for a one-dimensional quality x, depicted in Figure 2. In the upper panel, three *non-common* outside options are shown by yellow. Option (#1, #2) is available for these two agents and thus determines their reservation utilities. Option (#3, #5) is available for agents #3, #5. Somehow, options #4 and #6 are available only to these two groups. These outside options may be the bundles designed by a competitor and assumed to be given for the monopolist. Agent groups' masses are $m = 3, 1, 1.5, 4.4, 1.5, 1$. The valuations have peaks at the bliss points $b = (1, 2, 3, 4, 5, 6)$ with heights $h = (1.0, 1.45, 1.5, 1.1, 1.5, 1.1)$. These points (b_i, h_i) are shown in green. Black lines with arrows describe the (strict and non-strict) envy-graph resulting from such incentive-incompatible plan. Vertical arrows describe the participation constraints (drains). We observe a linear flow-graph. If the drains were connected to the same outside option, this graph would become a river.

In the lower panel the profit-maximizing points $(\bar{x}_i, \bar{t}_i)$ are shown in red and it occasionally happens that only $x_1 = 0.66666$ gets distorted. The reason is that the point $(\bar{x}_4, \bar{t}_4)$ is symmetrically envied from both the sides, whereas another envied (but non-distorted) point $(\bar{x}_6, \bar{t}_6)$ just

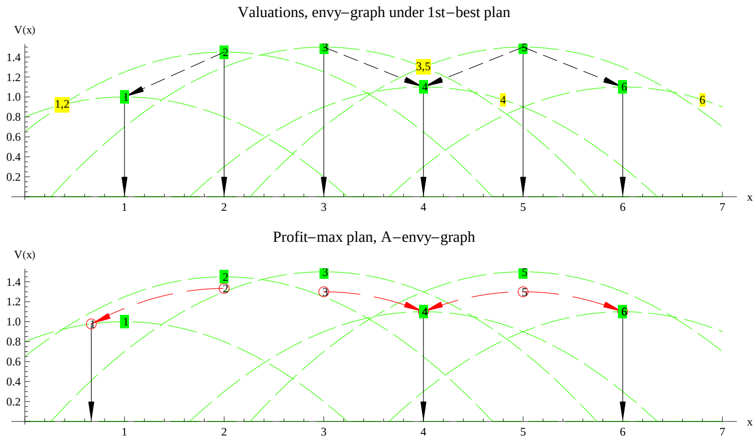

Figure 2: One-dimensional distortion under HSM preferences (Example 2).

happens to lie on the active curve v_5. The red lines with arrows describe the resulting A-envy-graph that mainly preserves the structure of the graph of the first-best incentive-incompatible plan, being a reduction from it.[19] The Lagrange multipliers, numbered naturally, are $\lambda_{11} = 4.0$, $\lambda_{44} = 7.4$, $\lambda_{66} = 1.0$, $\lambda_{21} = 1.0$, $\lambda_{34} = 1.5$, $\lambda_{54} = 1.5$, $\lambda_{56} = 0$, so, here LA-graph differs from A-graph. The main conclusion is that the flow graph of solution must be linear, but this line can consist of directed chains connected head-to-head or tail-to-tail. The line may break into pieces and distortion of any envied bundle is directed from the agent envying from the outside.

5.2 Two-dimensional spatial preferences

Theorem 2 about a flat and an order-preserving solution graph is generalized now onto a two-dimensional quality space. However, we use specific assumption HSM+ with $X = \mathrm{R}^2$ and quadratic distance function, i.e, quadratic valuations

$$v_i(x) = h_i - (b_{i1} - x_1)^2 - (b_{i2} - x_2)^2 \tag{16}$$

sometimes also called "gravity preferences".

A graph is called a *planar* one when it can be displayed on a plane without any intersecting arcs. We call it *weakly-planar* when it becomes planar after treating *each bunch as a single node*. This property is established as follows.

THEOREM 3. *Assume two-dimensional quadratic preferences (16), and any ρ-incentive-compatible plan $(\bar{x}, \bar{t})$, then: (i) It generates a weakly-planar flow-graph $\bar{\bar{G}}_{-0}(\bar{x}, \bar{t})$, which can be positioned on the quality space $X = \mathrm{R}^2$ through connecting qualities $\bar{x}_j$ perceived as nodes by (linear) envy-arcs, whereas under $\rho > 0$ the graph is planar.*[20]

(ii) When $i \to j$ in this graph, then the envied point $\bar{x}_j$ in this graph, the bliss-point b_j of this agent and the envying agent's bliss-point b_i belongs to some right-angled triangle where b_i

[19] One can derive a conjecture that the profit-maximizing envy-graph should always preserve the structure of the "first-best" envy-graph that emerges under the first-best plan (x^*, t^*) with the relaxation $\rho = \infty$ where only participation constraints are considered (or at least be a *reduction* of the first-best graph). However, only sometimes such preservation holds.

[20] This claim very probably can be extended to *all* spatial prefernces, not only quadratic, but the envy arcs will loose the linear shape.

and $\bar{x}_j$ are the ends of the hypotenuse and b_j belongs to the cathetus starting at b_i.[21]

Proof. Take any ρ-incentive-compatible plan (x, t) and display its envy-graph $\bar{G}$ on a quality plane, each couple (x_i, x_j) becoming an arc when $(i \to j)$. Our goal is to show that these linear arcs do not intersect. Whenever $(x_i \to x_j)$, both points must lie on the active surface W_i (and the envied point j belongs also to surface W_j). It is easy to show that the projection of the intersection of any couple of surfaces, W_i and W_k, is a straight line $\{(x_{i1}, x_{i2})|\, v_i(x_{i1}, x_{i2}) - t_i = v_k(x_{i1}, x_{i2}) - t_k\} \Rightarrow h_i - (b_{i1} - x_{i1})^2 - (b_{i2} - x_{i2})^2 - t_i = h_k - (b_{k1} - x_{i1})^2 - (b_{k2} - x_{i2})^2$. Indeed, powers 2 here cancel each other and the equation becomes linear. Naturally, all points envied by i lie on one side from this line (closer to i because belonging to its surface), but all nodes envied by k lie on the opposite side (closer to k), see Fig. 3. Thereby, non-intersection of any couple of envy-arcs $i \to j$ and $k \to l$ becomes clear. The arcs $x_i \to x_j$ and $x_k \to x_l$ just belong to different half-spaces of the plane, each containing its envying agent: x_i or x_k respectively. For any ρ-incentive-compatible plan (x, t), the logic is exactly the same only *all* active surfaces become a little (for ρ) lower than under $\rho = 0$. $\square$.

We now illustrate the use of envy-structures for discussing the direction of distortion among profit-maximizing packages or locations.

Example 3: (monopolistic-location rule). In Figure 3 the green dots describe nine locations $\{b_1, ..., b_9\} = \{(0.35, 0.35), (0.55, 0.9), (0.5, 1.5), (0.9, 0.55), (1.0, 1.0), (1.15, 1.48), (1.57, 0.5), (1.45, 1.0), (1$ on the square $[0, 2] \times [0, 2] \subset R^2$ which may represent nine small towns. They are populated with 9 related consumer groups having valuation heights $\{h_1, ..., h_9\} = \{0.55, 1, 1, 1, 1.3, 0.95, 1, 1, 0.86\}$ which are the maximal tariffs that can be paid if served at home, having in mind also the personalized outside options (not presented here explicitly unlike in Figure 2). The towns have populations $\{m_1, ..., m_9\} = \{1, 1, 1, 1, 1.4, 1, 1, 1, 1\}$. The agents have gravity valuations $v_i(x_i) = h_i - (b_i - x_i)^2$ as in (16). A monopolistic seller (e.g., a chain store) chooses the locations and price levels for 9 facilities (supermarkets) within or near these 9 towns. The first-best facility position $x_i = b_i$ inside each town is "non-distorted" one, such positions are numbered accordingly to $\{b_1, ..., b_9\}$. However, such lucky outcome for all 9 towns could result only under quasi-horizontal preferences that provide disconnected flow-graph, but that is not the case here. Instead, the profit-maximizing facility locations calculated numerically are $\{x_1, ..., x_9\} = \{(0.278333, 0.278333), (0.457353, 0.879412), (0.5, 1.5), (0.879412, 0.457353),$
$(1., 1.), (1.1788, 1.57215), (1.1788, 1.57215), \quad (1.55833, 1.), (1.29, 1.6)\}$.
They are shown by nine red circles connected by arrows of the envy-graph in the direction of envy. One can see that distortion (deviation of the profit-maximizing circle from related socially-optimal dot) always obeys Theorem 1: the direction *from* the envying fist-best location *to* the envied fist-best location similar to the direction of the envy-arcs: the envier is *pushing* the envied bundle outside. In particular, location 2 is envied by two agent groups, so, by eqn. (14) the sum of the pushing gradients $(\nabla v_2, \nabla v_4)$ determines the direction of distortion.

One can observe that the graph structure obeys Theorem 3: it is planar and is order-preserving in the sense that the higher peaks remain higher in the graph here. When there is a unique envier, *the direction of distortion* of any envied location is exactly opposite to

[21] This claim expresses a sort of shape-preservation between the net of the bliss-points $(b_1, ..., b_n)$ and the net of resulting qualities $(x_1, ..., x_n)$: each envied node lies approximately in the direction from b_i to b_j but farther away. Since the non-envied nodes (the graph summits) remain non-distorted $(x_i = b_i)$, the whole net-of-profit-maximizing qualities $(x_1, ..., x_n)$ looks like a continuous deformation of initial net $(b_1, ..., b_n)$, these points to become x_i being pushed away from the graph summits (see our figures).

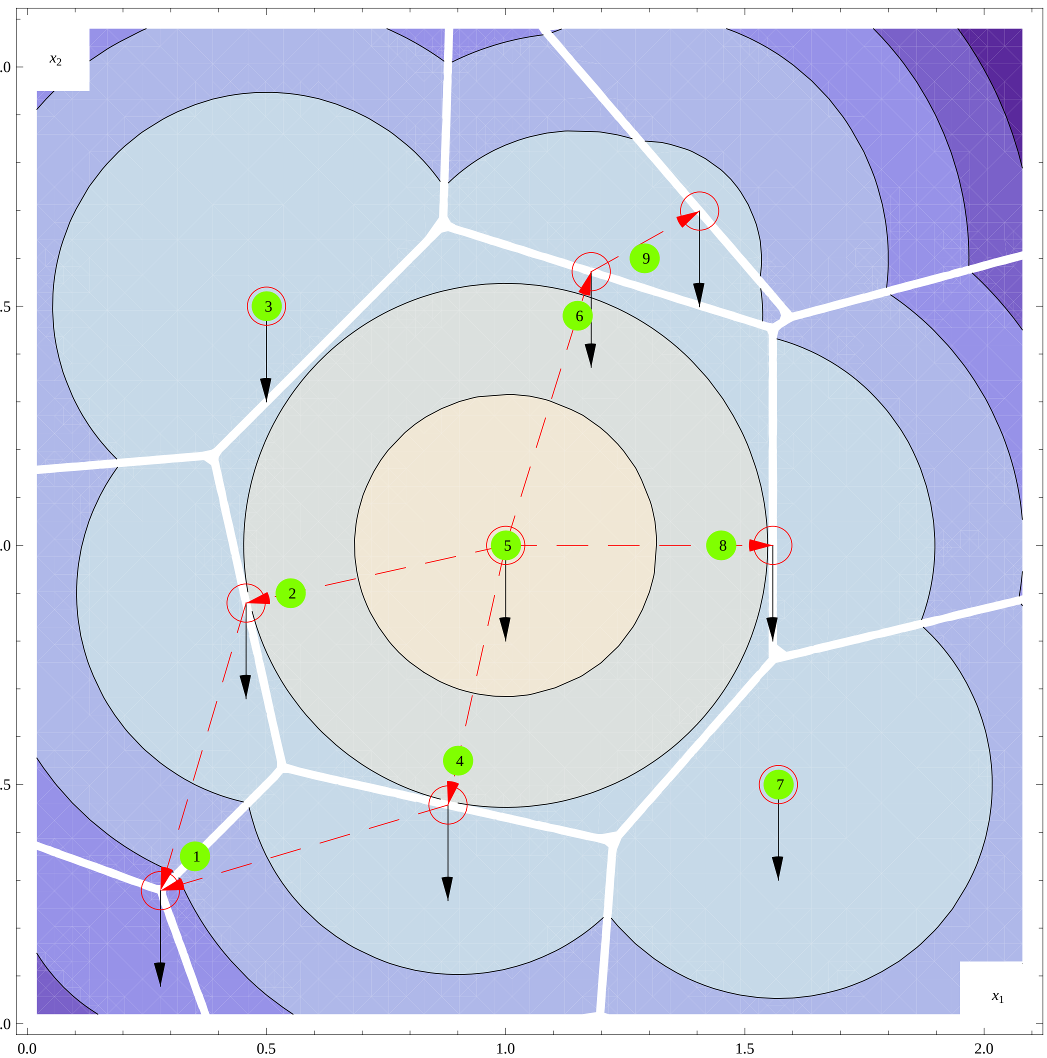

Figure 3: Two-dimensional locational distortion under HSM preferences (Example 3).

the envier's summit, and thereby can be called "envy pressure" from the envier. Under two enviers, the direction of distortion reflects their weighted "envy pressure" on the envied node. Moreover, *the amount of distortion* is larger when the weight of the envier is bigger (that can be derived from the optimality conditions). Thus, assuming similar costs at all locations, the rule of monopolistic location distortion in such situations can be formulated as centrifugal—"from the hills to the valleys", i.e., the network of supply points is distorted relatively to the network of the demand points *towards least willingness to pay and away from the highest willingness to pay.*

This seems counter-intuitive: we expect more shops located on the street that is populated by the rich rather than on the one populated by the poor. However, our expectations may result from (frequently observed) other market structures: oligopoly or free-entry oligopoly similar to monopolistic competition. These may have the opposite location rule. If it is true, then by observing pro-centric or anti-centric location pattern we can conjecture about the underlying market structure, is it essentially monopoly (tacit collusion) or oligopoly. It could help to rationalize the location choices of producers.

Example 4: Can we empirically reveal the kind of envy-graph for some observed product line? For instance, in a liquor shop a typical menu of packaging many brands of whisky contains 0.75, 1.0, 1.75 liters in quantity dimension and "young", "middle" and "very old" in quality dimension (measured by the age of the whisky). This overall amounts to 9 points of service like in Figure 3.[22] Our intuitive conjecture is that 1.0 liter bottle of the middle quality has the highest price-cost margin or net tariff. Thereby, this middle package should serve as the "top" of the graph, whereas "border" packages should have quality and quantity distortion in the opposite directions: too small sizes for small bottles and too big for the bigger one, similar to the distortion in Figure 3. Empirical study can show if it is really the case or not.

6 Conclusions

We have used a quite general setting to study discrete screening without the single-crossing condition (SCC). Our technical novelties include constraints relaxation that makes the analysis more tractable, and extensive application of graph theory to screening problems. A modelling novelty is multiple outside options in screening, which enables one to extend the screening methodology and "envy-graphs" onto product lines in oligopoly and reach several conclusions.

(1) Regarding distortion in general case, we confirm the commonly held belief among economists that the usual "efficiency at the top and distortion below" remains true even without SCC, but we provide some important clarifications. The "top" now means any "source" of the solution graph, and "below" refers to its successors. Specifically, *a bundle is distorted if and only if non-zero is the aggregate envy* (the sum of envying utility gradients weighted by their Lagrange multipliers) towards this bundle. This aggregate gradient also determines the distortion direction: it is *opposite to the envy pressure.*

(2) To get more definite predictions about distortion (for the price of some additional restrictions), we introduce a new and promising "spatial" class of preferences bridging Hotelling and Spence-Mirrlees assumptions. These preferences enable us to characterize the solution graphs as linear or planar, i.e., every bundle envies only its neighbor(s). The direction of distortion in

[22]However, unlike our 9-towns example, consumer types in the whisky example can be of continuous nature and preferences need not follow the gravity pattern, so our model does not apply strictly.

monopolistic product lines becomes clear: *distortion is centrifugal in quality space*—from the bliss points of the high-demand consumers towards the low demands.

Further we plan to study an extension to "oligopoly with competences:" does it show the same pattern of distortion as monopoly? Also, it would be interesting to study in the same fashion a free-entry oligopoly, where consumers are continuously spatially heterogenous (between Hotelling and Spence-Mirrlees cases), with discrete points of service.

Generally, by our examples and theorems we would like to convince the reader that considering *non-vertical and non-horizontal envy structures in product lines* can be practical for firms to design their optimal menus, and for researchers to rationalize some specific patterns observed.

7 Appendix: proofs

PROOF OF PROPOSITION 1 (existence of the Lagrange multipliers in relaxed and non-relaxed cases).

The claim (a) about the existence of solution is simple. A continuous objective function should have a maximum on the compact set. To construct the compact set for admissible (x, τ), we use our assumption on bounded qualities for positive surplus and artificially compactify the admissible set as $Z = \{x \in R^{ln}, \tau \in R^n_+ : v_i(x_i) \geq u_{i0}\, \forall i\}$. Any negative tariffs or qualities bringing negative surplus cannot be optimal, so essentially this compactification does not restrict the admissible set.

In proving (b) we rely on Kokovin et al. (2011), Proposition 4 about the existence of the Lagrange multipliers.[23] It is very similar to what we are proving now but for positive relaxation $\rho > 0$ and our idea is to expand this statement onto the case $\rho = 0$ by a limit transition that uses additional assumption of unique maximum. We construct an infinite sequence of all local argmaxima with their Lagrange multipliers and study any such couple $(z^{(n)}, \lambda^{(n)})_{\rho \to 0} \to (z^*, \lambda^*)$. This limit (z^*, λ^*) (of the sequence or some of its subsequence) must exist because variables $z = (x, \tau)$ are bounded by our assumption on bounded $x : v_i(x) > 0$. Also bounded are variables $\lambda \in [0, \check{M}] : \check{M} = \max_G\{M_G\}$ due to relation (13) holding for all $\rho > 0$ (Proposition 4 from Kokovin et al., (2011)). One of such limits (related to one of local maxima) must coincide with our global maximum $\bar{z} = z^*$ of the unconstrained problem. It is so, because the relaxation $\rho > 0$ keeps the global maximum feasible and the objective function cannot increase in ρ discontinuously at $\bar{z} : \rho = 0$, since $\bar{z}$ is supposed to be unique (or at least isolated). It is common in optimization that when we continuously expand the admissible set, the argmaxima changes continuously at the points where it is unique. Thus, there is a sequence $(z^{(n)}, \lambda^{(n)})_{\rho \to 0} \to (\bar{z}, \lambda^*)$ converging to our global maximum. Further, since the number of possible graphs is finite, the sequence must contain a subsequence with the same list $G^{\lambda^{(n)}}$ of LA-constraints for whole tail of the sequence. So, all the needed equalities and non-strict inequalities (8)–(13) hold true for the whole tail and thereby for the limiting point $\bar{z}$. This means that by this sequence we have constructed the list of LA-constraints and related vector λ^* of the Lagrange multipliers, satisfying the FOC, i.e., the Lagrange necessary conditions for this maximum $\bar{z}$. The essence of this proof is that all excessive constraints (that could be bunched at $\bar{z}$ and thereby prevent the

[23]Its proof exploits Lemma 2 and the specific feature of our screening problem: a linear objective function and separable constraints w.r.t. $v_i(x_i)$ and τ_i.

cone of the admissible direction being solid at $\bar{z}$) are ignored, not included into the needed list G^{λ^*} of LA-constraints by construction through the sequence. $\square$.

PROOF OF PROPOSITION 2.

Claim (i) means: [B-envy-free $i_0 \Rightarrow$ efficient x_{i_0}]. Take the optimization problem in terms of net-valuations $v_i(x_i) = V_i(x_i) - C(x_i)$. No-B-envy assumption means that all constraints like $(j \to i_0)\forall j$ could be eliminated from the optimization program and the objective function $\pi = \sum_{i=1}^{n} m_i t_i$ is maximized w.r.t. x, t under the remaining constraints. They include variables x_{i_0}, t_{i_0} only in the *left* side of the inequalities in the form $v_{i_0}(x_{i_0}) - t_{i_0} \geq v_{i_0}(x_j) - t_j$, $v_{i_0}(x_{i_0}) - t_{i_0} \geq u_{i0}$. The bigger the magnitude $v_{i_0}(x_{i_0})$, the bigger t_{i_0} can become, but t_{i_0} is maximized and no other constraint restricts profit contribution t_{i_0} from the *above*. Therefore some constraint of this type is binding, and at the argmaximum $(\bar{x}, \bar{t})$ of profit, the function $v_{i_0}(.)$ also reaches its *unconstrained* maximum at the quantity $\bar{x}_{i_0}$. This means that $\bar{x}_{i_0}$ is non-distorted . Claim (ii) is obvious.

Further, we shall need the following auxiliary claim: [concavity and LA-envy-free $i_0 \Rightarrow$ efficient x_{i_0}]. To prove by contradiction, suppose $\bar{x}_{i_0}$ is *not* the unconstrained argmaximum of $v_{i_0}(.)$. Then, under concavity of v_{i_0}, in any close vicinity of $\bar{x}_{i_0}$ there is a point $\check{x}_{i_0}$ (actually, many points) bringing higher value $v_{i_0}(\check{x}_{i_0}) > v_{i_0}(\bar{x}_{i_0})$ (an alternative assumption of strict quasi-concavity works similarly). By no-LA assumption, there exists some $\varepsilon > 0$ such that relaxation of all constraints of the type $(j, i_0) : v_j(x_j) - t_j \geq v_j(x_{i_0}) - t_{i_0}$ for this amount ε, the solution remains unchanged.

Then the additional welfare $v_{i_0}(\check{x}_{i_0}) - v_{i_0}(\bar{x}_{i_0}) > 0$ from the new better point $\check{x}_{i_0}$ situated in ε-vicinity of $\bar{x}_{i_0}$ could be distributed between the agent i_0 and the seller. In fact, by constructing a new package $(\check{x}_{i_0}, \check{t}_{i_0})$ one can increase the profit π without violating any constraints. This can be done by slightly increasing the net tariff $\check{t}_{i_0} = \bar{t}_{i_0} + \delta$ enough to not violate constraints with direction (i_0, j): $v_{i_0}(\check{x}_{i_0}) - t_{i_0} \geq v_{i_0}(x_j) - t_j$. These constraints have some slack $v_{i_0}(\check{x}_{i_0}) - v_{i_0}(\bar{x}_{i_0})$ now, and constraints (j, i_0) have some slack by LA-free assumption. But, the increased profit contradicts the optimality of $\bar{x}$. This proves that $\bar{x}_{i_0} \in \arg\max_{x_i} v_{i_0}(x_i)$.

Claim (iii): [concavity, weakly-A-envy free $i_0 \Rightarrow$ efficient x_{i_0}]. By weakly-A-envy free we mean absence of envy from any agents not bunched with the one studied. Under the no-bunching case, obviously, if a package $(\bar{x}_{i_0}, \bar{t}_{i_0})$ is A-envy-free, it also is LA-envy-free, so the claim just proved applies (one can also repeat similar concavity arguments for an independent proof).

Now we prove the same no-distortion claim for the case of a *group* of consumers $K = \{i_0, ..., k\} : \bar{x}_k = \bar{x}_{k-1} = ... = \bar{x}_{i_0}$, bunched together with this package $i_0 : v_j(x_j) - t_j = v_j(x_{i_0}) - t_{i_0}\,\forall j \in K$ and not envied from outside. Can their incentive-compatibility constraints comprise a cycle causing a distortion? Suppose there are one or more agents from this group whose welfare function v_j does not attain maximum at the equilibrium point, i.e., $\bar{x}_j \notin \arg\max_z v_j(z)$.

Take a small $\varepsilon > 0$ and denote a small ε-vicinity of $\bar{x}_{i_0}$ as: $B(\bar{x}_{i_0}, \varepsilon) := \{z \in R^l |\ ||\bar{x}_{i_0} - z|| \leq \varepsilon\}$, small enough so that all IC constraints to i_0, which are strict inequalities at the point $\bar{x}_{i_0}$ (those (j, i_0): $j \notin K$) remain satisfied under all $z \in B(\bar{x}_{i_0}, \varepsilon)$ also, with $\bar{t}_i$ remaining fixed. Continuity of v_i (which follows from concavity on R^l) allows us to build such B. Now maximize among agents and points and denote an agent by k whose welfare function v_k attains the maximum value within $B(\bar{x}_{i_0}, \varepsilon)$ among all $\{i, ..., k\}$, so that

$$\check{x}_k := \arg\max_{j \in K} \max_{z \in B(\bar{x}_{i_0}, \varepsilon)} (v_j(z) - v_j(\bar{x}_{i_0})).$$

As in the proof (no-LA+C) above, from the assumptions of distorted $\bar{x}_{i_0}$ and concavity (or strict quasi-concavity) there exists $\tilde{x}_{i_0} \in B(\bar{x}_{i_0}, \varepsilon)$ such that $v_{i_0}(\tilde{x}_{i_0}) > v_{i_0}(\bar{x}_{i_0})$. Combining this with maximal position of $\check{x}_k$, we get $v_k(\check{x}_k) - v_k(\bar{x}_{i_0}) \geq v_{i_0}(\tilde{x}_{i_0}) - v_{i_0}(\bar{x}_{i_0}) > v_{i_0}(\bar{x}_{i_0}) - v_{i_0}(\bar{x}_{i_0}) = 0 = v_k(\bar{x}_{i_0}) - v_k(\bar{x}_{i_0})$, so $v_k(\check{x}_k) > v_k(\bar{x}_{i_0})$. Then, replacing the equilibrium assignment $\bar{x}_k$ by $\check{x}_k$ for this agent, we can again choose a new tariff $\check{t}_k = \bar{t}_k + v_k(\check{x}_k) - v_k(\bar{x}_{i_0})$ (as in version no-LA+C) to increase both welfare and profit without violating any constraint with this new $(\check{x}_k, \check{t}_k)$. Indeed, the position of $\check{x}_k$ in B guarantees that the outside agents $\forall j \notin K$ will not switch to $\check{x}_k$ even under the old tariff $\bar{t}_k$, and so they will not switch for a bigger new $\check{t}_k$ also. Our k-th agent herself is indifferent between the old and the new package: $v_k(\check{x}_k) - \check{t}_k = v_k(\bar{x}_{i_0}) - \bar{t}_k$, so she does not switch. Other agents ($\forall j \in K$, i.e., bunched) will not wish to switch to package $(\check{x}_k, \check{t}_k)$, because k is chosen to maximize the benefit from switching among $\forall j \in K$. In other words, by recalling $\bar{t}_j = \bar{t}_k = \bar{t}_{i_0}$, we can ensure incentive-compatibility constraint j, k as satisfied:

$$v_j(\bar{x}_{i_0}) - \bar{t}_{i_0} = v_j(\bar{x}_{i_0}) - \check{t}_k + v_k(\check{x}_k) - v_k(\bar{x}_{i_0}) \geq v_j(\bar{x}_{i_0}) - \check{t}_k + v_j(\check{x}_k) - v_j(\bar{x}_{i_0}) = v_j(\check{x}_k) - \check{t}_k.$$

So, we have increased the profit with a new feasible package $(\check{x}_k, \check{t}_k)$ and this contradicts the profit-maximizing $(\bar{x}_k, \bar{t}_k)$, so the distortion assumption was wrong. This proves that there is no-distortion now for *all* bunched agents not envied from outside: $\bar{x}_j \in \arg\max_z v_j(z)\ \forall j \in K$.

This completes the proof of the proposition. □

Proof of Theorem 1: [aggregate LA-envy towards $j_0 \Leftrightarrow$ distorted j_0].

We have assumed that the Kuhn-Tucker theorem is applicable to our profit maximization program formulated in terms of net-valuations (see the conditions for this in Proposition 1). So, there must exist Lagrangian multipliers $\bar{\lambda}_{is} \geq 0$ related to all constraint (i, s) such that the profit maximum can be characterized at the point $(\bar{x}, \bar{t}, \bar{\lambda})$ by the first-order conditions of the following Lagrangian:

$$L(x, t, \lambda) := \sum_{i=1}^{n} m_i t_i + \sum_{i=1}^{n} \sum_{s=0}^{n} \lambda_{is}[v_i(x_i) - v_i(x_s) - t_i + t_s],$$

where we have denoted the package #0 representing non-participation as $(x_0, t_0) := (0, 0)$. If there are multiple dual variables λ satisfying the FOC, we fix one of them and discuss only it further. Taking the FOC w.r.t. t_{j_0}, we can collect all terms with envy directed *from* j_0 as $\sum_{s=0}^{n} \lambda_{j_0 s}$ and another sum $\sum_{k \neq j_0} \bar{\lambda}_{k j_0}$ represents all terms with envy *to* j_0:

$$\partial L(\bar{x}, \bar{t}, \bar{\lambda}) / \partial t_{j_0} = m_{j_0} - \sum_{s=0}^{n} \bar{\lambda}_{j_0 s} + \sum_{k \neq j_0} \bar{\lambda}_{k j_0} = 0$$

(of course, $\bar{\lambda}_{ij} = 0$ for non-active constraints).

From the condition $\bar{\lambda}_{ij} \geq 0$ we have $m_{j_0} + \sum_{k \neq j_0} \bar{\lambda}_{k j_0} > 0$, therefore $\sum_{s=0}^{n} \lambda_{j_0 s} > 0$. Now taking the FOC w.r.t. any component $x_{j_0 r}$ of x_{j_0}, we get two similar sums of multipliers λ_{ij} directed to and from j_0 (using the non-restricted domain of $x \in R^{nl}$, and denote derivative as $\dot{v}_{ir}(z) := \frac{d}{dz_r} v_i(z)$):

$$\partial L(x, t, \lambda) / \partial x_{j_0 r} = \dot{v}_{j_0 r}(\bar{x}_{j_0}) \sum_{s=0}^{n} \lambda_{j_0 s} - \sum_{k \neq j_0} \lambda_{k j_0} \dot{v}_{kr}(\bar{x}_{j_0}) = 0 \quad \forall r = 1, ..., l.$$

On the other hand, by concavity of v_{j_0}, the point $\bar{x}_{j_0}$ is non-distorted (it is an unrestricted maximum of v_{j_0}) if and only if $\nabla v_{j_0}(\bar{x}_{j_0}) = 0 \in R^l$. So, in view of $\sum_{s=0}^{n} \lambda_{j_0 s} > 0$, we have proved that the relation $\sum_{k\neq j_0}\lambda_{kj_0}\dot{v}_{kr}(\bar{x}_{j_0})=0\ \forall r$ is the necessary and sufficient condition for no-distortion. If this equality holds for any λ satisfying the FOC, then it holds for all such λ. In other words, inequality $\sum_{k\neq j_0}\lambda_{kj_0}\nabla v_{kr}(\bar{x}_{j_0})\neq 0\in R^l$ for all λ (aggregate LBA-envy from some agents) implies distortion. □

Corollary (distortion direction) follows.

A program finding profit-maximizing menus

In Kokovin's personal page at HSE (http://www.hse.ru/en/user/?_r=5563158.703607496726/Other) we present our program in language Wolfram-Mathematika, that finds profit-maximizing menus under any quadratic valuations of three agent types (it can be modified to cope with non-quadratic valuations also and to more types).

References

[1] Andersson, T., 2005. "Profit maximizing nonlinear pricing." Economics Letters, 88, 135-139.

[2] Andersson, T., 2008. "Efficiency properties of nonlinear pricing schedules without the single-crossing condition," Economics Letters, 99(2), 364-66.

[3] Araujo, A., Moreira, H., 2010. "Adverse selection problems without the Spence-Mirrlees condition," Journal of Economic Theory, 145(3), 1113-1141.

[4] Armstrong, M., 2006. "Recent Developments in the Economics of Price Discrimination," in: Blundell, Newey, and Persson, (Eds), *Advances in Economics and Econometrics: Theory and Applications,* Ninth World Congress of the Econometric Society, Volume II, 97-141, Cambridge: Cambridge University Press.

[5] Brito, D.L., Hamilton, J.H., Slutsky, J.E., and Stiglitz, J., 1990. "Pareto Efficient Tax Structures," Oxford Economic Papers, 42, 61–77.

[6] Berg, K., Ehtamo, H. (2010) "Interpretation of the Lagrange multipliers in non-linear pricing problem."- Optimization Letters 4(2), 275-285.

[7] Guesnerie, R., Seade, J., 1982. "Nonlinear Pricing in a Finite Economy," Journal of Public Economics, 17, 157-179.

[8] Iskakov, M., Iskakov, A., 2012. "Equilibrium in Secure Strategies — Intuitive Formulation," HSE Working Paper WP7/2012/06 Series WP7.

[9] Kokovin, S., Nahata, B., Zhelobodko, E., 2010. "Multidimensional screening under nonlinear costs: Limits of standard approach." Economics Letters, 107(2), 263–265.

[10] Kokovin, S., Nahata, B., Zhelobodko, E., 2011. "All Solution Graphs in Multidimensional Screening [in Russian]." Journal of New Economic Association, Moscow, issue 11, pp.10-38. (In English: Preprint at RePEC (2011), Electronic Resource.)

[10] Kokovin, S., Nahata, B., Zhelobodko, E., 2013. “Dupuit conjecture for constrained screening: overall distortion and type-partitioning.” Economics Letters, 118 (2013), pp.439-441.

[11] Nahata, B., Kokovin, S., Zhelobodko, E., 2003. “Self-selection under non-ordered valuations: type-splitting, envy-cycles, rationing and efficiency.” RePEC discussion paper, http://EconPapers.repec.org/RePEc:wpa:wuwpgt:0307001.

[12] Rochet, J.-C., Stole, L., 2003. “The Economics of Multidimensional Screening,” *Advances in Economics and Econometrics: Theory and Applications: Eighth World Congress Vol I.* (Eds.) Mathias Dewatripont, Lars Peter Hansen and Stephen J. Turnovsky. Cambridge: Cambridge University Press.

[13] Rochet, J.-C., Chone, P. (1998). “Ironing, Sweeping and Multi-dimensional Screening,” Econometrica, 66(4), 174-93.

[14] Stole, L.A. 2007. “Price Discrimination and Competition” in: *Handbook of Industrial Organization*, 2007, v. 3, eds. M.Armstrong, R.H.Porter, pp. 2221-2299 Elsevier: Amsterdam.

[15] Vohra, R.V., 2008 “Paths, Cycles and Mechanism Design,” http://www.kellogg.northwestern.edu/faculty/vohra/ftp/duke.pdf.

[16] Wilson, R., 1993. “*Nonlinear Pricing*,” Oxford University Press, Oxford, UK

Sergey Kokovin
Sobolev Institute of Mathematics, Novosibirsk State University and NRU Higher School of Economics
NSU, Novosibirsk 630090, Russia.
e-mail: skokov7@gmail.com