All solution graphs in multidimensional screening

Kokovin, Sergey; Nahata, Babu and Zhelobodko, Evgeny
Sobolev Institute of Mathematics SB RAS, University of Louisville, Novosibirsk State University, Economics Department

November 2010

All Solution Graphs in Multidimensional Screening

Sergey Kokovin
Department of Economics,
Novosibirsk State University
Novosibirsk 630090, Russia.
e-mail: kokovin@math.nsc.ru,
Babu Nahata
Department of Economics
University of Louisville
Louisville, Kentucky 40292, USA.
e-mail: nahata@louisville.edu
and
Evgeny Zhelobodko
Department of Economics,
Novosibirsk State University
Novosibirsk 630090, Russia.
email: ezhel@ieie.nsc.ru

All Solution Graphs in Multidimensional Screening

By

Sergey Kokovin[a], **Babu Nahata**[b], **and Evgeny Zhelobodko**[a]

[a]Department of Economics, Novosibirsk State University, Novosibirsk 630090, Russia,

e-mail: skokovin@math.nsc.ru, ezhel@ieie.nsc.ru[1]

[b]Department of Economics University of Louisville, Louisville, Kentucky 40292, USA.

e-mail: nahata@louisville.edu.

Abstract.

We study general discrete-types multidimensional screening without any noticeable restrictions on valuations, using instead ε-relaxation of the incentive-compatibility constraints. Any active (becoming equality) constraint can be perceived as "envy" arc from one type to another, so the set of active constraints is a digraph. We find that: (1) any solution has an in-rooted acyclic graph ("river"); (2) for any logically feasible river there exists a screening problem resulting in such river. Using these results, any solution is characterized both through its spanning-tree and through its Lagrange multipliers, that can help in finding solutions and their efficiency/distortion properties.

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

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

1 Introduction

Targets of modern screening theory are quite numerous: optimal taxation, optimal hiring policy, and notably non-linear pricing or second-degree price discrimination. All economic situations of this kind have essentially the same mathematical representation based on discrete or continuous *heterogeneous* population of agents; screening means treating different types differently on self-selection basis. Recent screening literature (see reviews by

[1]We are grateful for the financial support from the University of Louisville and from the Economics Education and Research Consortium, Inc. (EERC) grant No 06-056, with funds provided by the Eurasia Foundation (with funding from the US Agency for International Development), the World Bank Institution, the Global Development Network and the Government of Sweden. We are thankful for the assistance and comments from Richard Ericson, Victor Polterovich, Alexei Savvateev and Jean-Charles Rochet.

Rochet and Stole (2003), Armstrong (2005), Stole (2007)) leaves aside textbook cases satisfying Spence-Mirrlees single-crossing condition (SCC) on agents' types.[2] New approaches include into analysis many realistic situations when goods or services are multidimensional, or just consumers' valuations for the commodity are not ordered in any vertical or horizontal sense. For instance, a teenager can consume a lot of Internet traffic, but cannot pay too much for the first minute of connection, while a businessman would buy several minutes a day very expensively for E-mail, but cannot consume a lot. That means that these two demand curves do cross and SCC is violated. Moreover, when the commodity characteristics are multidimensional, like power and color of a car, the SCC becomes just senseless. Still, the producers do somehow design their product lines, i.e., menus of quantity/tariff bundles. Therefore, the economists should understand what happens, notably, upward or downward distortion of quantity/quality and ideas for diminishing the deadweight loss. However, theory becomes complicated without SCC.

To explain the key ideas, we define the *envy-graph of solution* as the list of those incentive-compatibility and participation constraints which are active, i.e., become equalities; agent types are the nodes of the graph and the arcs (constraints) connect them. Each equality means that one agent is indifferent between the two bundles and almost eager to switch, or (almost) "envies" another bundle. Standardly, under SCC the solution graph is linear, i.e., a path: The highest-demand consumer type envies the second-highest one, who envies the third one, and so forth. This simple chain structure enables a textbook method to obtain solutions and their important economic properties, notably informational rent for all types except the lowest one, efficiency at-the-top and distortion below, that means socially-optimal quantity for the highest-demand type and non-optimal quantities for others.

We want to know, what are the feasible solution structures, and related distortion properties without SCC? In answering these questions, we extensively apply notions from graph theory (see Section 3 for all definitions and Fig.1 for illustration). In addition to

[2] When i's consumer type has monetary valuation $V_i(x)$ for quantity/quality $x \geq 0$, SCC means "vertical" order of types such that $V'_{i+1}(x) > V'_i(x)\ \forall i, x,\ V_i(0) = 0$, i.e., higher type likes the commodity stronger. Another popular simplifying assumption originating from Hotelling's linear-city model is "horizontal" order: all agents are identical except for locations of their bliss points in some unidimensional space of quantity/quality.

further elaboration of such methodology, the novelty applied in this paper is the formulation of the screening model in a slightly generalized way: each constraint is relaxed by a small parameter $\rho \geq 0$. As explained in Section 3, such relaxation luckily overcomes several technical obstacles for clear results, including dicycles, bunching and problematic existence of the Lagrange multipliers (dicycle is a closed directed path in the graph, bunching means same bundles for different types, see special sudy of this phenomenon in Rochet and Chone, 1998).

The background results on solution structures without SCC are formulated here in terminology of this paper, since there are no standard terms for envy-graphs so far. Guesnerie and Seade (1982) show that under one-dimensional quality and strictly concave utility functions dicycles in the solution structure cannot contain more than two different nodes (bundles). Besides dicycles among non-bunched bundles either inexist or can be eliminated without any loss in profit (see also Brito et al. (1990)). Rochet (1987) addresses the opposite question: What kinds of structures do exist? He studies the implementation problems using graph terms; these results are summarized in Rochet and Stole (2003) as Lemmas 1 and 2 which, among other statements, state indirectly that any node is connected by a path to the root of the solution graph (such graph is called *in-rooted*).

Our results include Lemma 1 in Section 4 which under weaker assumptions than in Rochet and Stole states in-rooted property directly: Any solution has an in-rooted graph, thereby it contains a spanning-tree (a tree containing all nodes). In addition, under strict relaxation ($\rho > 0$) Proposition 1 states that envy-graph does not have dicycles, so it is a *river*, which is an acyclic in-rooted graph. Therefore, bunching among predecessors and successors in the graph is also formally excluded. It may remain only as an accident among disconnected agent types, and there can be almost-bunching, that means ρ-close bundles.

Envy-graph being a river is, surprisingly, the *only* property of the solution structure that holds without special assumptions on utilities, because our Proposition 2 states under modest conditions that every river is a solution structure for some screening problem.[3] Lemma 3 enumerates all possible rivers, i.e., different structures of screening. They are quite numerous: 5 rivers for two consumers, 79 for three consumers, and so on. These findings bury the hope for easy screening theory without SCC or similar restrictions, but

[3]We are thankful to Alexei Savvateev for posing this question and to Charles Rochet and for inspiring us to publish the answer.

something can be done.

Based on pervious statements, Proposition 3 characterizes any solution through its spanning-tree without first-order conditions (FOC). In contrast, under differentiability Proposition 4 characterizes any relaxed solution through its FOC, relying on the existence of Lagrange multipliers under rivers, and this class of graphs serves as a road-map to find all solutions.

Extensions and economic fruits of our approach are postponed to our next paper, but we mention them here to motivate extensive mathematics of this paper. First, even without SCC we generally have efficiency at-the-top, distortion below and informational rent for all types except the lowest one. Only "the top" means now any top (or "leaf") in the spanning-tree. Second, though algorithmic questions in non-convex optimization like general screening are difficult, our propositions shows the way to practically find solutions through solving FOC for all possible rivers and then comparing the resulting local optima (this discovers the first practical solution method without SCC). Third, there is an economically meaningful tractable class of consumer populations – "spatial" populations – which includes both textbook special classes: horizontal and vertical populations. In this class the solution properties including distortion are more easy to describe than in general screening. This enables taxonomy of all possible regimes of screening, and predicting changes of market outcomes with the population parameters. Finally, it is worth mentioning that our approach can be valid not only in screening but also in broader mechanism-design or "implementation" settings with incentive-compatibility, so when possible we make statements in the form suitable for broad applications.

Section 2 formulates the model, Section 3 presents and motivates our approach to graph theory in screening. Section 4 presents results on solution structures, Section 5 characterizes solutions, and Section 6 concludes. Appendices contain proofs which are non-intuitive and technical.

2 Model

Our model is almost standard in screening of discrete consumers, only restrictions on functions are relaxed and a constraints-relaxation parameter added. We interpret the model mostly in terms of a monopolistic seller and buyers, 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 being similar in more general areas of mechanism design than 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 appear in the market, or the total number or mass of such agents (consumers). Multiple agents of the same type can also mean multiple purchases of one individual. The quantity- or quality-tariff bundles are (x_i, t_i), where $x_i \in X$ denotes the l-dimensional vector of attributes of the bundle purchased by 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 monetary valuation of a purchase. In most usual particular case of common outside option $0 \in X$, valuations can be normalized as $V_i(0) = 0$, but in a more general case non-participation level $V_i(a_i) = 0$ relates to any outside option a_i. For Proposition 4, we additionally assume differentiability, but otherwise do not restrict V_i, X.[4]

A monopolist selects a subset $I^n \subseteq I^N$ of $n \leq N$ types to be served and offers a product or service using a menu of several packages of different quantities or qualities at some fixed tariffs on 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 types and their probabilies but cannot discriminate personally. We generalize the usual linear cost to a more general-form cost function $C(m, x) : \mathbb{R}^{n+nl} \rightarrow \mathbb{R}$, but in Propositions 2 and 4 it takes a special *fixed-plus-separable* form

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

[4]Weak restrictions on X and V allow to model many interesting and realistic situations, say, satiable demands and/or discrete characteristics. Positivity of consumption and tariffs can be modelled through positivity restrictions on X, V, whereas decreasing V_i or negative x_i are appropriate for modelling effort spent in a principal-agent setting.

where $f_0 \geq 0$ stands for some fixed cost and $c(\cdot) : \mathbb{R}^l \to \mathbb{R}$ is cost function per-package.[5] We use the standard assumption that the producer designs only one package for each type, thereby planning an *assignment*, $(x, t) = \{(x_i, t_i)\}_{i \in I^n}$ and from equivalent choices an agent selects whatever the principal prefers (friendly behavior). The profit π is the difference between total tariffs and total costs. After introducing a constraint-relaxation parameter $\rho \geq 0$ for technical reasons, we can formulate the seller's *relaxed assignment-optimization* program as follows.

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

$$\forall i \in I^n \Rightarrow V_i(x_i) - t_i + \rho \geq V_i(x_k) - t_k \quad \forall k \in I^n \setminus \{i\}, \tag{2}$$

$$\forall i \in I^n \Rightarrow V_i(x_i) - t_i \geq 0 \;. \tag{3}$$

Here (2) represent the incentive-compatibility (IC) constraints, and participation constraints are (3). A plan (x, t) satisfying (2)–(3) is called *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 $\rho \geq 0$ a solution $(\bar{x}, \bar{t})$ to (1)–(3) is a called here a *relaxed ρ-specific solution*, or just relaxed solution or *ρ-solution*.

Though Propositions 1 and 3 cover general case $\rho \geq 0$, the main focus of our study is on ρ-solutions with small $\rho > 0$, because such relaxation implies acyclic solutions without sacrificing appropriate modelling of reality.[6] Moreover, we have found in numerous examples that generally a relaxed solution converges under $\rho \to 0$ to related non-relaxed solution, though proof of such convergence remains an open question.

To complete the setting, it should be added that under fixed-and-separable costs $C(m, x) = f_0 + \sum_{i=1}^{n} m_i c(x_i)$, it is possible and standard to normalize the model. It means considering

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

[6] Economically speaking, a relaxation parameter ρ can be interpreted as the "cost of switching" for 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 $\rho \geq 0$, for modelling a premium to the agent for non-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 undermine existence of solutions.

normalized, i.e., net-of-cost valuations $v_i(x_i) = V_i(x_i) - c(x_i)$ or social surpluses instead of initial valuations, and seek for net-of-cost tariffs $\tau_i = t_i - c(x_i)$, named also as per-package profits τ_i. Then initial screening problem (1)-(3), obviously, amounts to the *relaxed normalized screening program* allowing for strongest results in the sequel:[7]

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

$$\forall i \in I^n \Rightarrow v_i(x_i) - \tau_i + \rho \geq v_i(x_k) - \tau_k \quad \forall k \in I^n \setminus \{i\}, \tag{5}$$

$$\forall i \in I^n \Rightarrow v_i(x_i) - \tau_i \geq 0 \ . \tag{6}$$

3 Graph notions for incentive compatible structures

Now we introduce some notions from graph theory and suggest our way of applying such notions to screening or, more generally, to incentive-compatibility. The terminology and methodology in such applications has not been standard so far. For example, Brito et al. (1990) speak of eliminating "cycles of *binding* incentive constraints among separated types," which is different from our terminology below (see also other terminologies in Guesnerie and Seade (1982), and in Vohra (2008)). 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 but for limited use of term "binding", reversed direction of arcs and some new notions.

First we reiterate standard terms for graphs and add some new terms, followed by relating graphs to screening problems and motivate our approach.

Standard terms for digraphs. A *directed graph* or *digraph* G (hereafter just "graph") is a collection of nodes (vertices) denoted here $i \in G$ 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

[7]It is worth recalling here that welfare-maximizing screening under such restriction on total costs is an equivalent problem, reciprocal to profit-maximization (see e.g. Brito et al., 1990, Rochet and Stole, 2003). So, our results can be transferred to welfare-maximization .

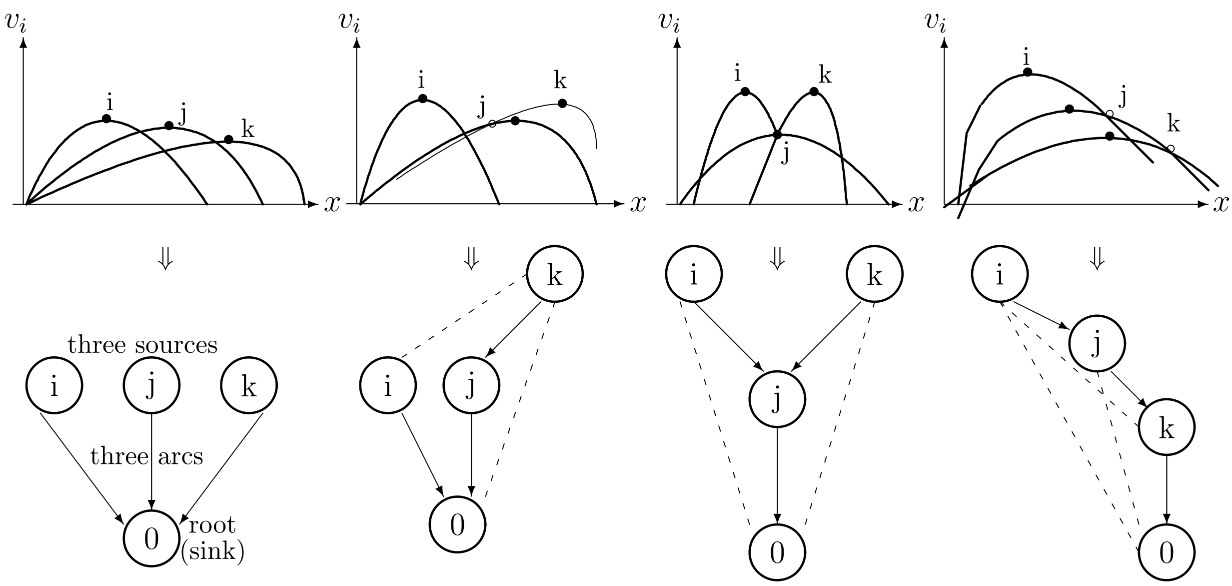
each $i \to j$ the arc's tail i is the adjacent *predecessor* of j, and the arc's head j is the adjacent *successor* of i. A *source* is a connected node (connected to some other nodes) without predecessors, while a connected node without successors is a *sink*; if the sink is unique, it is called an *in-root* or, hereafter just *root*. A node without adjacent arcs is *disconnected.* A *walk* is a sequence of adjacent nodes and edges $i_1, e_{12}, i_2, e_{23}, i_3...$; 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*. A *spanning-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 Appendix 1 see more explanations, together with the *preorder* $\succeq$ definition).

New terms. An in-rooted acyclic digraph is called hereafter a *river*. Obviously, all trees are rivers but the latter may have *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).

Envy-graphs in screening. In applying graphs to screening, all types' identities #1,...,#n are treated as nodes, whereas constraints are interpreted as *envy* arcs within related *envy graph.* In this graph, common or individual non-participation option is considered as an additional common node with label #0. More precisely, our optimization program (1)–(3) has $n \times (n-1) + n = n^2$ inequalities that can become active, i.e, becoming equalities. For any feasible or incentive-compatible 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 active at (x, t) (double-bar over G highlights equalities as the basis of definition and $j = 0$ means non-participation option). For a non-feasible plan (x, t) we can similarly define the strict-envy graph $G^{<}(x, t)$ representing the list of all violated constraints and the list $G^{\geq}(x, t) = G^{>}(x, t) \cup \bar{\bar{G}}(x, t)$ of satisfied 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*, in the direction of possible switching. It means that type 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 (2003), seems inconvenient in this respect.

Fig. 1 illustrates these notions and ideas in the case of *three* agent types. Sample profiles of three valuations shown in the upper panel generate all four possible classes of spanning-trees: star, spider, fork and chain. The star cannot have bypasses. However, adding or not various bypasses (shown by dash lines) to other trees can make 4 different unlabelled rivers from a spider, 3 rivers from a fork and 9 rivers from a chain. Thereby, 17 qualitatively different screening regimes are possible when three types are served; distortion or inefficiency means that the equilibrium bundles are not the summits of the net valuations (here such distorted bundles are empty dots). Among these 17 regimes only star structure guarantees social efficiency, as we show in another paper, but incidental efficiency may occur. Further, when labelling these 17 graphs, various permutations of labels increase the number of possible solution structures under three given types to 79, as Lemma 3 ensures.



(a) Star-tree (b) Spider-tree (c) Fork-tree (d) Chain-tree

Figure 1: Possible configurations of net valuations and 4 resulting classes of envy-graphs.

LA-graphs, B-graphs and motivation for ρ-relaxation. First note that a screening problem (1)-(2) is quite often non-convex, even under concave valuations V. It happens because concave functions enter both sides of the inequalities. Then, as for any non-convex optimization, a distinction becomes important between an active constraint and a *binding* constraint (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 equation (7) below. So, a B-graph representing all binding constraints may be needed.

Besides, there could also be a need for a LA-graph, which is defined as the list of all LA constraints, i.e, those having strictly positive Lagrange multipliers (see our Proposition 4). This LA-graph generally may differ from A-graph and B-graph, and even from their intersection. The typical reason for the discrepancy among A-,B- and LA-graphs is the so-called *bunching* situation. Bunching means identical packages $(x_i, t_i) = (x_j, t_j) = ...$ assigned to different agent types $i, j, ...$ 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 such cases, naturally, all bunched agents do envy each other, thereby creating a dicycle in A-graph $\bar{\bar{G}}(x,t)$ and "over-constrained" situation. Bunching and other dicycles create substantial hardships in characterizing and finding solutions, mainly because constraint-qualification conditions fail. Then 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, under positive relaxation A-graph generally coincides with 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 the most simple over-constrained non-convex example, where the constraints display all three kinds of importance:

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

Clearly, the optimum here is $\bar{x} = -1$, and the constraint (iii) is binding, because it cannot be dropped keeping 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 reflects the sensitivity of the objective function to related constraint, as it should. However, for a small price in 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 with data 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 $\bar{\bar{G}}\ (x,t)$. From now we have in mind this kind of graphs when we drop "A" and mention just envy-graphs. Only in Proposition 4 and some proofs we use LA-graphs $G_+^\lambda\ (x,t) \subseteq \bar{\bar{G}}\ (x,t)$.

4 Solution structures: all essential envy-graphs are rivers and vice versa

In this Section we show which types of solution structures are possible and which are not. This helps in characterizing the solutions in Section 5.

4.1 All envy-graphs are rivers

To prepare Propositions 1, 2 and 3, a lemma below states the most general property of solution structures which is guaranteed solely by quasi-linearity of utilities for both the non-relaxed ($\rho = 0$) and the relaxed screening.[8]

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.*

Proof: In $\bar{\bar{G}}(\bar{x}, \bar{t})$, suppose the root (#0) is free of envy, i.e., no constraint $(i \to 0)$ is active. The specific form of the constraints (2) shows that in this case, with quantities $\bar{x}$ remaining unchanged, all tariffs $(\bar{t}_1, ..., \bar{t}_n)$ could be increased simultaneously by some (same) amount without violating any constraint, because variables t_i enter all IC

[8] Previous somewhat similar statements that we know, are Lemmas 1 and 2 in Rochet and Stole (2003), and Proposition 1 in Andersson (2005), but we distinguish between binding and active constraints, consider quite general functions and ρ-relaxation. To generalize the in-rooted property further, a similar lemma for non-quasi-linear utility functions $u(x,t)$ should assume them continuous, strictly decreasing in t, $u(x, \infty) = -\infty$.

inequalities at both sides. This contradicts profit maximization at $(\bar{x},\bar{t})$. So, the set $L_1^\wedge = \{i|\ (i,0) \in \bar{\bar{G}}(\bar{x},\bar{t})\}$ of agent types adjacently connected to the root is non-empty: $L_1^\wedge \neq \emptyset$. If the compliment-set $I^n \setminus L_1^\wedge$ is empty, then the lemma is proved. Otherwise, suppose that the compliment-set is not connected *to* #0 and to $L_1^\wedge$ by envy (not mentioning the connection *from*). Then, for the same reason, it again contradicts the optimality of $(\bar{x},\bar{t})$, so the second-high layer $L_2^\wedge = \{i|\ (i,j) \in \bar{\bar{G}}(\bar{x},\bar{t}) : j \in L_1^\wedge\}$ is also nonempty, thereby connected indirectly to #0. We repeat this logic for all layers, and by induction, the in-rooted property is proved. □

Now, under additional assumptions we can ensure acyclic property of in-rooted envy-graphs under relaxation $\rho > 0$. It is done using Lemma 2. Its idea and the simple version of statement (i) originates in Guesnerie and Seade (1982) being also used in several papers. We extend it for the case $\rho > 0$ and enforce it.[9] In our terms, the lemma states that the preorder of profit-contributions cannot contradict the partial order $\succ$ of the envy-graph.

Lemma 2: (profits order). *Take any ρ-solution $(\bar{x},\bar{t})$ under fixed-and 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., $i \to \ldots \to j \Rightarrow\ \bar{\tau}_i \geq \bar{\tau}_j$,;* (ii) *under ($\rho > 0$) this inequality is strict: $i \to \ldots \to j \Rightarrow \bar{\tau}_i > \bar{\tau}_j$, 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.*

Proof. Using separable-cost assumption, we can argue in terms of per-package profit contributions $\tau_i = t_i - c(x_i)$. Let us prove claim (ii) for the case $\rho > 0$. Assume that there could be a couple of adjacent agents ($i \succeq_{\to} j, i \to j$) with the reverse order of profit contributions: ($\tau_i \leq \tau_j$). Then we could increase the objective function for amount $m_i\rho > 0$ by replacing the envier's package $(\bar{x}_i, \bar{\tau}_i)$ by the envied package $(\bar{x}_j, \bar{\tau}_j)$, i.e., by assigning a new package $(\tilde{x}_i, \tilde{\tau}_i) := (\bar{x}_j, \bar{\tau}_j + \rho)$. This new menu $((\bar{x}_1, \bar{\tau}_1), \ldots, (\tilde{x}_i, \tilde{\tau}_i), \ldots, (\bar{x}_n, \bar{\tau}_n))$ remains incentive-compatible because no new quantity-tariff packages arise, and the only affected agent i is indifferent between her old and new packages, by the assumption $i \to j$ which means $\tilde{v}_i(\bar{x}_i) - \bar{\tau}_i = \tilde{v}_i(\bar{x}_j) - \bar{\tau}_j - \rho$. Besides, with a bigger net tariff ($\tilde{\tau}_i = \bar{\tau}_j + \rho > \bar{\tau}_i$)

[9] Other versions of claim (i), in Brito et al. (1990) and Andersson (2005), are proved under stronger restrictions on costs and valuations than here, and consider only adjacent agents, without using graph notions. Extension of such claim onto non-separable concave cost functions see in Kokovin et al. (2010).

now the new menu brings a larger profit compared to the initial one. This contradicts the optimality of initial $(\bar{x}, \bar{\tau})$ menu. Thus, $i \to j \Rightarrow \bar{\tau}_i \geq \bar{\tau}_j + \rho$. By transitivity and using chains like $\bar{\tau}_i > \bar{\tau}_k ... > \bar{\tau}_j$, such ordering conclusion follows also for non-adjacent successive agents $i \to ...j$. The same logic proves the case (i) with $\rho = 0$, only the inequality proved is not strict: $\bar{\tau}_i \geq \bar{\tau}_k ... \geq \bar{\tau}_j$. $\square$

The above two lemmas imply the following proposition on acyclic solution structures.[10]

PROPOSITION 1: (ALL ENVY-GRAPHS ARE RIVERS).[11] *For any ρ-solution $(\bar{x}, \bar{t})$ to a screening problem with fixed-and separable costs and relaxation $\rho > 0$, its envy-graph $\bar{\bar{G}}(\bar{x}, \bar{t})$ is a river.*

Proof. In-rooted property of $\bar{\bar{G}}$ follows from Lemma 1. As to dicycles, by Lemma 2 under $\rho > 0$, profit contributions $\bar{\tau}$ are well coordinated with partial order of A-graph $\bar{\bar{G}}$, i.e., predecessors always bring strictly higher profit contributions $\bar{\tau}_i$ than their successors. Therefore any (bunched or not) dicycle $i \to j \to ... \to i$ in $\bar{\bar{G}}$ would mean a cycled order $\bar{\tau}_i > \bar{\tau}_j > ... > \bar{\tau}_i$ among real numbers $\bar{\tau}_i \in \mathbb{R}$, which is impossible. $\square$

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

Additionally, observe that we have obtained here the restrictions on endogenously emerging structure of the solutions, in contrast with the exogenous structure (unique path) imposed by SCC in many papers. Of course, the river-structure revealed is too broad a characterization, but it is the *only* possible knowledge that can be obtained without some specific assumptions on valuations v_i like SCC. Now we show why it is the case.

[10]Moreover, observe that the above proof simplifies a given plan through "unifying" some agents, i.e., deleting a package from the menu and endowing its owner with another package already existing in this menu. Then it becomes obvious that regarding any ρ-feasible plan we can also state that it is either acyclic or can simplified in this way into a weakly more profitable ρ-feasible plan satisfying the natural profit ordering and thereby acyclic.

[11] 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.

4.2 All rivers can be envy-graphs

Let us show that *any* river with n nodes plus the root can be a solution's envy-graph for some screening problem where n agent types are served. Having explained the main idea by Fig.1, we now enumerate all n_{+0}-rivers formally (sign := means assigning a value).

Lemma 3 (number of rivers).[12] *Consider all labelled rivers having n non-root nodes $\{1, ..., n\}$ and the root (sink) with label #0. The number $r_0(n)$ of such rivers can be found recursively as*

$$\begin{aligned} r_0(n) &:= \sum_{k=1}^{n} 2^{n-k} a_{n,k}, \text{ where} \\ a_{n,k} &:= \sum_{m=1}^{n-k} (2^k - 1)^m 2^{k(n-m-k)} \binom{n}{k} a_{n-k,m}, \qquad a_{j,j} = 1, \end{aligned}$$

and $\binom{n}{k}$ denotes the number of all k-element subsets of $\{1, 2, ..., n\}$. In particular, $r_0(1) = 1$, $r_0(2) = 5$, $r_0(3) = 79$, $r_0(4) = 2865$, $r_0(5) = 254111$.

Proof: see Appendix 2.

Note that so far we have got the number $r_0(n)$ of possible graphs for those agent types who are served. We can tell also the number of possible envy graphs for any N-agents screening problem where some agents may remain unserved and excluded from the graph. Here we should just summarize numbers $r_0(n)$ for all possible n-subsets $(n = 1, 2, ..., N)$ from the labelled set I^N and get

$$r_1(N) = \sum_{n=1}^{N} \binom{N}{n} r_0(n).$$

In the typical special case when $0 \in X$ is a common outside option for everybody, this zero bundle is just assigned to all non-served types, so formally everybody is always served: $n = N$. Then zero and non-zero quantities need not be distinguished in the graph representation of such solution (see also Remark after Proposition 4). In this case the number of possible resulting rivers is only $r_0(N) < r_1(N)$.

Proposition 2 (all rivers can be envy-graphs). *Consider a given dimensionality $l \geq 1$ of real commodity space $X = \mathbb{R}^l_+$, dimensionality n of population served, and a relaxation parameter $\rho \in [0, \frac{1}{2}(\sqrt{n+1} - \sqrt{n}))$. For each river $\bar{G}$ among all $r_0(n)$ logically*

[12] In Appendix 2 one can find also enumeration of all labelled trees with root #0, in particular $T_1 = 1$, $T_2 = 3$, $T_3 = 16$, $T_4 = 125$, $T_5 = 1296$.

possible rivers, there exists a profile of continuous concave net-of-cost valuations $(v_1, ..., v_n)$: $v_i(0) = 0$ *and frequencies* $(m_1, ..., m_n)$ *generating such a river as the envy-graph* $\bar{\bar{G}}$ *of the solution for the* ρ*-relaxed normalized screening program* (4)–(5).

PROOF: see Appendix 3.

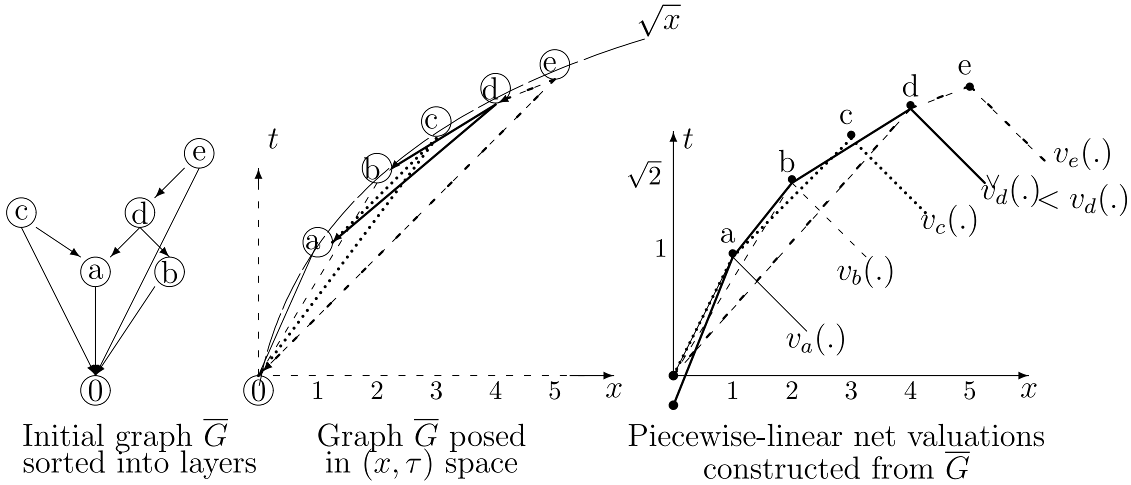


Figure 2: Construction of valuations generating any river.

This long and tedious proof is based on direct construction of the needed net-valuation curves, as illustrated by example in Fig 2. Namely, first we sort the given river $\bar{G}$ into layers according to the longest path to the root. Second, graph $\bar{G}$ is positioned in quantity-tariff space on some strictly concave curve like $\sqrt{\cdot}$ so that none of the nodes is below any arc, and all arcs go downwards. Third, the arcs going from any node become the parts of the active piece-wise linear indifference curve belonging to this agent type. Forth, the curves non-adjacent to the root (like active curve $\check{v}_d$ in Fig.2 which is lower than related valuation v_d) are connected to some point below the root and lifted up to become valuations. Besides, all curves are extended to $+\infty$ as downward sloping lines. Fifth, we prove that the nodes at the peaks of the valuations built are really optima under these valuations and some frequencies m_i adjusted to this graph. This stage uses a version of our Proposition 4 with Lagrange characterization of solutions.

Discussing this result, one can object against too specific non-smooth valuations constructed. However, it is not a problem, we understand how to extend the proof of Proposition 2 to strictly convex smooth valuations. Moreover, for $n = 2$ types (see Nahata et al. 2002) and for $n = 3$ types (in unpublished paper) we have found that the needed valuations generating any river can be taken quadratic, and importantly, each of possible $r_0(n)$ rivers results from a non-degenerate region of parameters $a_i, b_i : v_i(x) = a_i x - b_i x^2$. These special cases support the conjecture that all rivers are not only *possible*, but also can be expected to result from screening solutions with *substantial probabilities* under quite reasonable utilities.

These conclusions appear surprising and somewhat disappointing. Rather big number $r_0(n)$ of qualitatively different outcomes dims the hope for a quick and simple analysis of general screening since natural restrictions on valuations like concavity do not reduce the variety. Proposition 2 motivates the necessity for more restrictive assumptions to predict more definite structures and properties. But at least we know how many switches between different screening regimes are possible when valuations or frequencies do change.

5 Solution characterization: spanning-trees and Lagrange multipliers

This section shows how one can use the envy-graphs in characterizing solutions, and gives a hint for finding them. Relying on Lemma 1 and Proposition 1, we characterize now any optimal solution (1) through its spanning-trees and tree-specific programs; (2) through its river-specific first-order conditions with Lagrange multipliers (in the case of normalized problem, smooth valuations and $\rho > 0$). Each method enables to replace the initial package-optimization program with an *equivalent* finite family of graph-specific programs, each specific problem being simple and allowing for direct exact solution.

5.1 Characterizing solutions by spanning-trees

The spanning-tree solution method optimizes only quantity-variables after expressing all tariffs through quantities. Such reduction of variables is well known under SCC since Spence (1980), its general idea is called "virtual-surplus approach." However, theoretically justified path-optimizing implementation of this approach was referred by professionals in the topic as impractical without SCC, see Rochet and Stole (2003) subsection 3.1. For practically solving their examples with small n Rochet and Stole (2003) and also Armstrong and Rochet (1999) optimize only spanning-trees instead of all paths. Indeed, studying all spanning-trees among a complete $n \times n$ graph of constraints looks doable, unlike studying all possible paths. Now Proposition 3 theoretically justifies their idea in general case, whereas other general methods are unknown.

In other words, Propositions 3 and 4 under very weak restrictions state the *necessary* conditions for optima in the absence of usual SCC. To formulate these conditions, we introduce new notations related to any node k in a graph G:

- $P_k(G) \subset \{0, 1, ..., n\}$ is the set of all predecessors of node k;
- $P_k^{ad}(G) \subseteq P_k(G) \subset \{0, 1, ..., n\}$ is the set adjacent predecessors of k;
- $S_k(G)$ is the set of all successors of k;
- $S_k^{ad}(G)$ is the set of adjacent successors;
- $s_k^1(G)$ is the unique adjacent successor, i.e., $S_k^{ad}(G) = \{s_i^1(G)\}$.

Sign := everywhere means assigning some value and (common or not for all agents) outside option or non-participation is denoted as node #0, the root of $\bar{\bar{G}}$.

Proposition 3 (Spanning-tree characterization). *Let* $(\bar{x}, \bar{t}) \in X^n \times \mathbb{R}^n$ *be a* ρ*-solution to the general screening program (1)-(3) where* n *consumers are served (*$\rho \geq 0$*). Then:* (i) *There exists a spanning-tree* $G_T \subseteq \bar{\bar{G}}(\bar{x}, \bar{t})$ *within this envy-graph such that the assignment* $(\bar{x}, \bar{t})$ *is also a solution to the following tree-specific optimization program (8)-(11) solved w.r.t. variables* x:

$$\tilde{\pi}(x, G_T) \quad : \quad = \sum_{i=1}^{n} m_i\, \boldsymbol{\theta}_i(x, G_T) - C(m, x) \to \max_{x \in X^n} \quad \text{s.t.} \tag{8}$$

$$V_i(x_i) - \boldsymbol{\theta}_i(x, G_T) \quad \geq \quad V_i(x_j) - \boldsymbol{\theta}_j(x, G_T) - \rho \qquad \forall (i, j) \notin G_T, \tag{9}$$

$$V_i(x_i) - \boldsymbol{\theta}_i(x, G_T) \quad \geq \quad 0, \tag{10}$$

tariff functions $\boldsymbol{\theta}_i$ *and tariffs* t_i *being determined by* G_T *as*

$$t_k = \boldsymbol{\theta}_k(x, G_T) := \sum_{j \in S_k(G_T) \cup \{k\}} [V_j(x_j) - V_j(x_{s_j^1(G_T)}) + \rho] \quad \forall k \geq 1,\ \boldsymbol{\theta}_0(.) \equiv 0. \tag{11}$$

(ii) *For any spanning-tree* $G_T \subseteq \bar{\bar{G}}(\bar{x}, \bar{t})$ *resulting from a solution, all solutions to the* G_T*-specific program (8)-(11) are also the solutions to initial program (1)-(3).*

(iii) *Under a fixed-plus-separable cost function, the tree-specific normalized objective function becomes separable w.r.t.* x_i*, namely:*

$$\tilde{\pi}(x, G_T) = \sum_{i=1}^{n}[m_i v_i(x_i) + \sum_{j \in P_i^{ad}(G_T)} (v_i(x_i) - v_j(x_i)) M_j^{PG_T}], \quad (12)$$

$$M_j^{PG_T} := \sum_{k \in P_j(G_T) \cup \{j\}} m_k, \quad (13)$$

$M_j^{PG_T}$ being the sum of predecessors' frequencies.

PROOF: see Appendix 3.

5.2 Characterizing solutions by FOC

We characterize now any optimal solution through its its first-order conditions and Lagrange multipliers in the case of normalized problem, smooth functions and $\rho > 0$. These necessary conditions of optima enable to replace the initial package-optimization program with a finite family of river-specific programs, each having a direct exact solution (by Lemma 3, as many as $r_1(N)$ rivers should be explored). Existence of the Lagrange multipliers for non-relaxed problem ($\rho = 0$) in the absence of usual SCC remains an uneasy open question, while the relaxed problem and related guaranteed existence of the Lagrange multipliers is our novelty.

PROPOSITION 4 (FOC-FOR-RIVERS CHARACTERIZATION). *Assume $X = \mathbb{R}^l$, fixed-and-separable costs, continuously differentiable net valuations $v_i : v_i(0) = 0$, and relaxation $\rho > 0$. Take any solution $(\bar{x}, \bar{\tau})$ to the normalized problem (4)-(6) and its LA-graph $G_+^\lambda = G_+^\lambda(\bar{x}, \bar{\tau}) \subseteq \bar{\bar{\bar{G}}}(\bar{x}, \bar{\tau})$, then:* (i) *There exist Lagrange multipliers $\lambda = (\lambda_{1,0}, \lambda_{1,2}, ...,$ $\lambda_{n,n-2}, \lambda_{n,n-1}) \in \mathbb{R}_+^{n*n}$, satisfying the first-order conditions on Lagrangian $\mathcal{L}(.)$ and supplementary inequalities as follows:*[13]

[13]In these equations all four sums of Lagrange multipliers can be simplified as $\sum_{\forall j} \lambda_{ij}$, but our river-specific formulation is more convenient for practical applications. This formulation enables to find *all solutions* $(\bar{x}, \bar{\tau}, \lambda)$ from hypothesic rivers G_+^λ.

$$\frac{\partial\mathcal{L}(\bar{x},\bar{\tau},\lambda)}{\partial t_i} = m_i - \sum_{j\in S_i^{ad}(G_+^\lambda)}\lambda_{ij} + \sum_{j\in P_i^{ad}(G_+^\lambda)}\lambda_{ji} = 0 \;\; \forall i>0, \tag{14}$$

$$\nabla_{x_i}\mathcal{L}(\bar{x},\bar{\tau},\lambda) = \nabla_{x_i}v_i(\bar{x}_i)\sum_{j\in S_i^{ad}(G_+^\lambda)}\lambda_{ij} - \sum_{j\in P_i^{ad}(G_+^\lambda)}\lambda_{ji}\nabla_{x_i}v_j(\bar{x}_i) = 0; \;\; \forall i>0, \tag{15}$$

$$0 = v_i(\bar{x}_i) - \bar{\tau}_i - \; v_i(\bar{x}_j) + \bar{\tau}_j + \rho \quad \forall (i,j)\in G_+^\lambda, \tag{16}$$

$$0 = v_i(\bar{x}_i) - \bar{\tau}_i \quad \forall (i,0)\in G_+^\lambda, \tag{17}$$

$$0 \le v_i(\bar{x}_i) - \bar{\tau}_i - \; v_i(\bar{x}_j) + \bar{\tau}_j + \rho \quad \forall (i,j)\notin G_+^\lambda, \tag{18}$$

$$0 \le v_i(\bar{x}_i) - \bar{\tau}_i \quad \forall (i,0)\notin G_+^\lambda, \; \bar{x}_i\in X, \text{ where} \tag{19}$$

$$G_+^\lambda = \{(ij)|\lambda_{ij}>0\}. \tag{20}$$

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

$$m_i + \sum_{j\in P_i^{ad}(G_+^\lambda)}\lambda_{ji} = \sum_{j\in S_i^{ad}(G_+^\lambda)}\lambda_{ij} \le M_i^{PG_+^\lambda} := \sum_{j\in P(i,G_+^\lambda)\cup\{i\}} m_j \quad ; \tag{21}$$

moreover, when the river G_+^λ is a tree, the positive multiplier for the unique successor $s_i^1(G_+^\lambda)$ of i is found as

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

Proof: see Appendix 3.

Remark: When condition $\bar{x}_i\in X$ takes most usual form $\bar{x}_i\ge 0$ (and $0\in X$ denotes the common outside option which is non-participation) it can be included into Lagrangian in usual way, then Lagrangian is maximized unconditionally. However, under natural valuations: $v_i(x_i)<0\;\forall x_i<0$, negative x_i brings negative tariff and cannot be locally-optimal. Then positivity constraints become redundant, as well as special handling of exclusion of some types from trade (otherwise possibility of exclusion could increase the number of iterations, i.e., exceed number $r_0(n)$ given by Lemma 3).

The use of these two propositions for finding the solutions is straightforward (we do not discuss computational efficiency). In essence, Proposition 3 states that *each* solution to the initial problem can be found through some spanning-tree. So, whenever we know a method to globally optimize tree-specific function $\tilde{\pi}(x,G_T)$ subject to constraints (8)-(11), it is sufficient to try all possible trees G_T, i.e., only 3 trees under two agent types, 16 under three types and so forth (see comment to Lemma 3). Generally a tree-specific program is non-convex but still much easier than the initial one because it often amounts to unconstraint optimization of $\tilde{\pi}(x, G_T)$. When some not-in-tree constraint becomes binding, it should be included into Lagrangian as in Proposition 4 and optimization repeated. But optimization cannot stop with the first local optimum found. We emphasize that we do not provide a sufficient condition for optima, only the *necessary* one, because typically we have a *non-convex optimization* in screening, even under strictly concave net valuations $v_i(\cdot)$ (see Section 3). Therefore, theoretically for finding a true solution, one should explore *all* possible trees G_T using Proposition 3 or all possible rivers G_+^λ using Proposition 4. Naturally, usual branches-and-bounds technique can help to considerably reduce this search, because when a tree $G_T = G_+^\lambda$ gives a feasible solution worse than another tree, all related rivers including G_T can be neglected, being non-optimal.

In the spirit of Vohra (2008), an interesting extension of our graph approach to screening can be the flow-network approach. Indeed, using our Proposition 4 one can use the Lagrange multipliers in the role of current-flows to and from any node i, frequencies m_i being the source magnitudes and equation (21) serving as a "conservation law" for these flows. Probably, this enables to use the special methods of flow-network theory in screening.

Another, more close physical interpretation of the screening problem arising from Proposition 4 is that frequencies m_i are the gravity forces pulling each bundle (x_i, τ_i) upwords together with the active indifference curve $v_i(\cdot)$ that must contain this point. The incentive-compatibility and participation constraints mean that neither point, including (0,0), can lie strictly below any curve. In this sense, Proposition 4 describes the local maximum of a screening program as the equilibrium between the forces pulling the whole net of curves and points up and its bearing (0,0). Then, the conservation law (21) expresses the balance between the forces pulling each point i upwards and the reaction of its bearings – successors. This direct physical analogy helps in finding efficient heuristic algorithms for solutions and, more importantly, in clear understanding distortion of particular bundles pulled aside from their peaks. For example, in Fig.1 the accidental efficiency of the middle point j in the case of "fork" means that the force pushing from the left occurs equal to one pushing from the right. In other cases, envy-arcs of the solution graphs generally bring distortion to the envied bundles and we prove this in another paper through Proposition 4.

6 Conclusions

We study a general setting for discrete multidimensional or unidimensional screening without SCC and obtain four mathematical results. (1) Without any restrictive assumptions on costs and valuations, a solution structure can be represented by a graph which is in-rooted and acyclic, i.e., a “river.” (2) More surprisingly, for any logically feasible river there exists a screening problem that results in such graph. Therefore in general one should explore all rivers to find or characterize solutions. These qualitatively different “regimes” in screening are enumerated, this number measuring computational complexity of screening without SCC. (3) We obtain characterization of solutions in terms of spanning-trees. (4) Under constraints relaxation, any solution is characterized in terms of Lagrange multipliers for each possible river. Based on these four results, any solution can be found by comparing finitely-many river-specific exact solutions (so far general practical solution method without SCC was absent).

Methodological novelties include our elaborated way of using graph theory in screening and the relaxation of incentive-compatibility constraints. The latter helps to overcome hardships with bunching, cycles, and existence of Lagrange multipliers that prevented finding solutions and their properties without SCC.

Our non-published subsequent paper derives efficiency or distortion properties from the graph structures, and imposes a reasonable restriction weaker than SCC to narrow down the class of outcomes. Another interesting extension inspired by Rochet and Chone (1998) could be a study of continuous setting without SCC, describing all possible manifolds of envy instead of envy-graphs studied here.

Appendix 1: two preorders

Partial order and preorders. In constructing and classifying graphs we need the following notions of partial order and preorder in a graph.

Any digraph G generates a unique *arc-induced partial order* $\succ_{\rightarrow}$ among the nodes, where the notation “$i \succ_{\rightarrow} j$” (“i is higher than j”) means that i is the predecessor of j, but j is not the predecessor of i. All couples (i, j) without $i \succ_{\rightarrow} j$ or $j \succ_{\rightarrow} i$ relation are perceived as non-comparable, which is denoted as $i?j$. This $i?j$ relates to couples (i, j) not connected by a path or connected by a dicycle. Clearly, an acyclic digraph is uniquely restored from its partial order.

A "preorder" $\succeq$ is a complete transitive binary relation allowing for equivalence denoted $i \sim j$ between any couple (i, j) but not allowing for non-comparability relation $i?j$. In other words, preorder is a partitioning of the nodes' set I into ordered classes, called here *layers*. Several preorders can be induced by any graph, the most useful being the following two.

For any (acyclic or not) digraph, the *shortest*-path *preorder* $\succeq_\wedge$ is constructed by sorting the nodes into layers, according to the shortest distance $d^\wedge(i, \#0)$ from i to the root. This $d^\wedge(i, \#0)$ is the number of arcs in the shortest available path to the root ($\wedge$ denotes minimum). The nodes that are equidistant from #0 are equivalent. The root is not a layer, all nodes adjacent to the root constitute the first layer $L_1^\wedge$, the nodes adjacent to $L_1^\wedge$ constitute the second layer $L_2^\wedge$, and so on; the more distant layer is higher (see use of this preorder $\succeq_\wedge$in Lemma 1).

In contrast, the *longest*-path *preorder* $\succeq_\vee$ is defined only for acyclic digraphs, and often differs from $\succeq_\wedge$. It reflects the length $d^\vee(i, \#0)$ of the longest available path from each i to the root ($\vee$ denotes maximum). A node having a longer maximal path is considered higher in preorder $\succeq_\vee$, but nodes with equal lengths are equivalent. Namely, the first layer $L_1^\vee$ contains all nodes connected to #0 by no more than one arc; layer $L_2^\vee$ contains all nodes connected to #0 by no more than two arcs, and so on. We use preorder $\succeq_\vee$ in proving Proposition 2 and illustrate it in Fig.1. One can see that only four $\succeq_\vee$-preorder types are possible among rivers with 3 non-zero nodes: $\{L_1^\vee \ni (i, j, k)\}$, $\{L_1^\vee \ni (i, j), L_2^\vee \ni k\}$, $\{L_1^\vee \ni i, L_2^\vee \ni (k, j)\}$, $\{L_1^\vee \ni i, L_2^\vee \ni j, L_3^\vee \ni k\}$.

Both preorders $\succeq_\vee$,$\succeq_\wedge$ do not contradict the partial order $\succ$ of a graph, but the first preorder $\succeq_\wedge$ often includes a predecessor and its successor into the same layer, unlike $\succeq_\vee$ which is "finer."

Appendix 2: enumeration of trees and rivers[14]

6.0.1 Number of trees

For solving a screening problem we should know the number of all in-rooted trees with n non-root nodes labelled $1, 2, ..., n$, and a root (sink) #0. We derive it from the following lemma about non-directed labelled trees.

Lemma 4 [Cayley (1889)].[15] *The number C_n of labelled trees of order n (with n nodes) is equal to $C_n = n^{n-2}$, first numbers of this sequence being $C_1 = 1$, $C_2 = 1$, $C_3 = 3$, $C_4 = 16$, $C_5 = 125$...*

Corollary. *The number T_n of labelled in-rooted trees with the sink labelled #0 and n non-root nodes $1, 2, ..., n$ is equal to*

$$T_n = C_{n+1} = (n+1)^{n-1},$$

first numbers of this sequence being $T_1 = 1$, $T_2 = 3$, $T_3 = 16$, $T_4 = 125$, $T_5 = 1296$, $T_6 = 16807$, $T_7 = 262144$...

Proof of Corollary. Directing a non-directed tree G (i.e., making an in-rooted tree from G) amounts to choosing any node in the role of the root which determines the direction of all arcs. Similarly, when labelling an unlabelled tree we can first assign any node to become the lowest number out of all labels used (#0 in our case) and then assign somehow other labels. Any labelled n-tree has a *unique* counterpart among all labelled n-trees directed towards the lowest number, which is #0 in this paper where the direction is towards #0. These classes are equivalent. Hence, our problem of counting in-trees of n nodes with labels $\{1, ..., n\}$ plus the root #0 is equivalent to Cayley's problem for non-directed labelled trees of the order $n + 1$. $\square$

Now we can use similar ideas for counting rivers.

6.0.2 Number of rivers

We need to count all *labelled rivers* with n non-root nodes $\#1, \#2, ..., \#n$ plus the root #0. This class of n_{+0}-rivers include trees and rivers with bypasses. Adding bypasses to trees is

[14] A more poetical title would be: "Enumeration of Stars, Trees and Rivers" but, unfortunately, there is only *one* star among n-rivers.

[15] See Harary and Palmer (1973), Section 1.7.

a good method to build rivers maintaining the longest-path preorder unchanged (see Fig 1). We use this idea to derive what we need from Robinson's lemma below. The lemma considers any acyclic digraphs, which is a broader class than rivers because the sink need not be unique.

Lemma 5 [Robinson (1970)].[16] *Denote by $a_{n,k}$ the number of labelled acyclic digraphs with n nodes, exactly $k \leq n$ of these nodes lacking arcs "to" them (because being sources or disconnected). Denote by $\binom{n}{k}$ the number of k-element subsets of the set $\{1, 2, ..., n\}$. Any number $a_{n,k}$ can be found recursively from the lower-order elements of matrix a as $a_{n,k} = \sum_{m=1}^{n-k}(2^k - 1)^m 2^{k(n-m-k)}\binom{n}{k}a_{n-k,m}$. Using matrix a, the total number $\rho_{ac}(n)$ of acyclic labelled digraphs of order n is found as $\rho_{ac}(n) = \sum_{k=1}^{n} a_{n,k}$, the first 7 numbers of this sequence being $\rho_{ac}(\cdot) =$ 1, 3, 25, 479, 22511, 2349987, 569684123.*

We use Robinson's lemma and details of its proof to derive our Lemma 3, which considers a different class of graphs but gives a similar formula, except for a multiplier 2^{n-k}.

Proof of Lemma 3 (number of rivers). We need to apply Robinson's lemma and prove that any magnitude $r_0(n)$ can be found recursively from the Robinson's matrix a as

$$r_0(n) := \sum_{k=1}^{n} 2^{n-k} a_{n,k}. \tag{22}$$

To this end, note that our task of enumerating rivers is rather similar to Robinson's enumeration of all acyclic digraphs because any acyclic (connected or disconnected) digraph G of n nodes $i = 1, ..., n$, we can extend by adding there a root #0. Namely, we take each node not connected *to* anything (i.e., take all sinks and disconnected nodes) and connect it *to* #0 by an arc *from* this node, thus "extending" the graph G downstream (see Harary and Palmer, 1973, Section 1.6, about extensions). Obviously this "extension" of graph G always makes a river from the initial acyclic graph G. Further, all rivers constructed in this way are different because their basic graphs were different and they were supplemented by #0 in a unique way. The class of acyclic digraphs supplemented by a root in this "minimal" way constitutes a family of rivers called further $R_1 = R_1(n)$. This class is equivalent to the class A_n of all acyclic digraphs and its number of elements is $|A_n| = \rho_{ac}(n)$ by Robinson's lemma.

However, to construct the complete family of rivers called $R_{all}(n)$ from the smaller family R_1, we should supplement R_1 by some other rivers. To comprehend this fact, consider a

[16] See Harary and Palmer (1973), Section 1.6.

river named $r_{kj} \in R_1$ with $k < n$ nodes adjacent to the root. Robinson's proof of his lemma tells us that there are $a_{n,k}$ such rivers: $j = 1, ..., a_{n,k}$. (To make quite correct reference here, we apply notation $a_{n,k}$ to k *sinks* of an acyclic digraph with n nodes, instead of Robinson's k *sources* used in Section 1.6. in Harary and Palmer (1973), but changing + for - in all directions of all graphs in a family makes no difference for their number). When this river r_{kj} has high (different from $L_1^\vee$) layers, they can be supplemented by some bypasses from these $L_2^\vee, L_3^\vee$,... directly to the root, such bypasses being absent in R_1 by construction (see Fig.1 for examples of extending a graph with bypasses). Therefore, enumerating all such modifications of any river $r_{kj} \in R_1$ amounts to 2^{n-k} combinations of "yes" or "no" possible for all $N(r_{kj}) = n - k$ high nodes of this river, previously non-adjacent to the root; each high node may become adjacent or non-adjacent to the root. Thus, each initial river r_{kj} becomes included into a richer family constructed from it through adding bypasses $i \to 0$. A river like "star" in Fig.1 may lack high layers $L_2^\vee, L_3^\vee, ...$, then $r_{kj} : k = n$, and it is the only member in its family: $2^{n-n} = 1$. More generally, the possibility of bypasses/no bypasses to #0 increases the number of rivers similar to initial r_{kj} by 2^{n-k} times. This multiplier with number $a_{n,k}$ of initial rivers yields the desired formula (22).

However, to use this formula without doubt, we must be sure that our extension of a river $r_{kj} \in R_1$ (one or several bypasses $i \to 0$ added) generates always a *new* river, different from any river, generated from some other element $r_{\hat{k}i}$ of class R_1. When $\hat{k} \neq k$ the difference is obvious because longest-path preorder $\succeq_\vee$ of a graph is not changed by any additional bypasses to #0, whereas such $r_{\hat{k}i}$ and r_{kj} belong to different classes, their first layers being different: $|L_1^\vee(r_{\hat{k}i})| = \hat{k} \neq k = |L_1^\vee(r_{kj})|$. More generally, even when $\hat{k} = k$ suppose that the same river $r_{ks} \in R_{all}$ resulted form adding bypasses $l \to 0$ to two different initial rivers $r_{ki} \in R_1$ and $r_{kj} \in R_1$. However, reverse operation (removing from the constructed river r_{ks} all bypasses $l \to 0$ of high layers $L_2^\vee, L_3^\vee$,...) is an operation with the unique outcome. Therefore, two constructed rivers could coincide only when their initial rivers did coincide: $r_{ki} = r_{kj} \in R_1$. Thus, we never arrive at the same river by extending (with bypasses to #0) different initial rivers from R_1. In other words, there arise no duplicates in R_{all} during our extension of R_1 and the lemma is proved. $\square$

Appendix 3: proofs of Propositions 2–4

We prove Propositions 3 and 4 before proving the complicated Proposition 2 which uses some version of Proposition 4 (Propositions 3 and 4 do not rely on Proposition 2).

Proof of Proposition 3.[17] (i) Based on Lemma 1, a (non-unique) spanning-tree can always be chosen from the envy-graph $\bar{\bar{G}}(\bar{x},\bar{t})$ of the solution $(\bar{x},\bar{t})$ to the initial problem. Select any such tree $G_T = G_T(\bar{x},\bar{t})$. Compare the formulations and observe that new problem (8)-(11) differs from the initial problem (1)-(2) *only* in one additional requirement: equation (11) requires that those constraints $(i,j) \in G_T$ which were active (equalities) at point $(\bar{x},\bar{t})$ would remain equalities at all feasible points; thereby it calculates the tariffs specifically as $\bar{t}_k = T_k(x_i) := \sum_{i\in S_k(G_T)\cup\{k\}}[V_i(x_i) - V_i(x_{s_i^1(G_T)})]$ (using the named equalities recursively). This new requirement makes the set of constraints *more* restrictive, though $(\bar{x},\bar{t})$ still satisfies it and at $(\bar{x},\bar{t})$ the new objective function takes the same optimal value $\tilde{\pi}(\bar{x},G_T)$ as the old function: $\pi(\bar{x},\bar{t}) = \tilde{\pi}(\bar{x},G_T)$. Hence in this new G_T-specific problem no admissible plan $(\tilde{x},\tilde{t})$ can be better than the initial plan $(\bar{x},\bar{t})$, which therefore remains optimal for the new tree-specific problem (8)-(11).

(ii) Any other optimal G_T-tree-specific solution $(\tilde{x},\tilde{t})$ must give the same value to new and old objective functions $\tilde{\pi}$ and π, as well as initial optimal plan $(\bar{x},\bar{t})$. Besides, $(\tilde{x},\tilde{t})$ must satisfy all constraints (8)-(11) that are stronger than (1)-(2). Therefore, such $(\tilde{x},\tilde{t})$ is also a solution to the initial program (1)-(2).

(iii) In separable case, derivation of our expressions $\tilde{\pi}, M_k^{PG}$ in (12)-(13) from initial formulae (8)-(11) is performed directly by recursively substituting tariffs along the tree. □

Proof of Proposition 4. When the Lagrange multipliers exist, these FOC follow straightforwardly from differentiating the Lagrangian $\mathcal{L}(x,\tau,\lambda)$ of our relaxed problem (4)-(5). We have only reformulated the usual FOC in graph terms with summation over $\forall j \in P_i^{ad}(G_+^\lambda)$, $\forall j \in S_i^{ad}(G_+^\lambda)$ for predecessors and successors instead of summation over simpler equivalent index-set $(\forall j \neq i)$. (The reason is that graph expressions help for practically finding local maxima from hypotheses on G_+^λ).

More serious question is the applicability of the Kuhn-Tucker theorem itself and related

[17]Under differentiability and other restrictions, one can derive the needed proof from Proposition 4, but this general proof is simpler.

existence of the Lagrange multipliers λ for a solution $(\bar{x}, \bar{\tau})$. Here the assumption $(\rho > 0)$ helps, because Proposition 2 ensures that there are no cycles, and Lemma 2 gives the strict order of profits. To characterize our non-convex optimization program, we exploit a non-convex version of the Kuhn-Tucker theorem with the Mangasarian-Fromovitz constraint-qualification, namely, Theorem 4.2 in Rockafellar (1993), reformulated as the following Lemma 6.

Lemma 6. Take an optimization program $\{\max_{z \in \mathbb{R}^{\bar{n}}} u(z);\ g_k(z) \geq 0,\ k = 1, ..., m\}$ with continuously differentiable functions, and a locally optimal point $\bar{z}$. Assume the cone of admissible directions at $\bar{z}$ is solid, i.e., there exists a direction-vector $w \in \mathbb{R}^{\bar{n}}$ with a positive scalar product to the gradient of each active inequality-constraint so that $w \nabla g_k(\bar{z}) > 0 \quad \forall k : g_k(\bar{z}) = 0$. Then there exist dual vector $\lambda \in \mathbb{R}^m_+$ satisfying the first-order conditions with this optimum $\bar{z}$.

To apply this lemma, we reformulate its terms in our notation: $\bar{z} = (\bar{x}, \bar{\tau})$, $\bar{n} = 2n$, $m = n \times n$, $(k) = (i, j)$, $g_{ij}(x, \tau) := v_i(x_i) - v_i(x_j) - \tau_i + \tau_j \geq 0$. Then the solid-cone restriction on the active constraints means

$$\begin{aligned} \exists w \ &\in \ \mathbb{R}^{2n} : \ w \nabla g_{ij}(\bar{x}, \bar{\tau}) > 0 \ \ \forall (i, j) \in \bar{G}(\bar{x}, \bar{\tau}) : \ \text{where} \\ \bar{G}(\bar{x}, \bar{\tau}) \ &: \ = \{(i, j) : g_{ij}(\bar{x}, \bar{\tau}) := v_i(\bar{x}_i) - v_i(\bar{x}_j) - \bar{\tau}_i + \bar{\tau}_j = 0\}. \end{aligned} \tag{23}$$

So, to apply Lemma 6, we should build a vector w satisfying condition (23). In other words, for all small $\epsilon > 0$, the direction $w\epsilon$ would lie in the admissible cone's interiority, so that $(\bar{x}, \bar{\tau}) + w\epsilon$ would be a strictly admissible plan (one showing only strict inequalities).

First, we renumber the agents in the order of their profit-contributions (net tariffs), so that $\bar{\tau}_{(1)} < \bar{\tau}_{(2)} < ... < \bar{\tau}_{(j)} < ... < \bar{\tau}_{(n)}$ (these numbers cannot coincide by Lemma 2). Second, we construct the needed specific vector $\bar{w}$ only from the net tariffs:

$$\bar{w} = (0^n, \bar{w}_{n+1}, \bar{w}_{n+2}, ..., \bar{w}_{2n}) = (0, ..., 0, -\bar{\tau}_{(1)}, -\bar{\tau}_{(2)}, -\bar{\tau}_{(3)}, ..., -\bar{\tau}_{(n)}) \in \mathbb{R}^{2n}. \tag{24}$$

One can easily check that such proportional decrease in all tariffs makes a feasible point $(\bar{x}, \bar{\tau}) + \bar{w}\epsilon$ *strictly* incentive compatible. So, our $\bar{w}$ satisfies the regularity condition (23):

$$\bar{w} \nabla g_{ij}(\bar{x}, \bar{\tau}) = 0 \cdot \dot{v}_i(\bar{x}_i) - 0 \cdot \dot{v}_i(\bar{x}_j) - 1 \cdot (-\bar{\tau}_i) + 1 \cdot (-\bar{\tau}_j) > 0. \tag{25}$$

Hence, Lemma 6 is applicable and the Lagrange multipliers exist.

(ii) To estimate the multipliers in the case of a tree, we can recursively find all λ_{ij} substituting them along our tree G_T into formula (14) of the river-specific program. In a tree this is possible because each node has a unique successor $S_i^{ad}(G) = \{s_i^1(G)\}$. Starting from the sources ("top" nodes without predecessors), we find their $\lambda_{is_i^1}$ as $\lambda_{is_i^1} = m_i$. Then we exclude them from the graph, similarly find λ_{js_j} for the remaining sources, and continue recursively finding for each i the Lagrange multiplier $\lambda_{is_i^1} = M_i^{PG_T}$ of the unique IC constraint leading from i downwards as required in (21).

For the case of a river which is not a tree, we see from (14) that the sum of succeeding multipliers λ_{ij} exceeds the sum of preceding ones, exactly by magnitude m_i:

$$m_i + \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} = \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij}.$$

If we search for $\max_{\lambda \geq 0}\{\sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij}\}$ among non-negative λ_{ij} satisfying all these equations, we find the needed upper bound (21). □

For Proposition 2 we need the following version of Proposition 4 with non-differentiable functions, left and right derivatives $v_i'^-(\cdot), v_i'^+(\cdot)$. For simplicity, we formulate it only for the unidimensional commodity and common outside option 0 for all agents, exactly as needed in proving our Proposition 2.

Proposition 4* (FOC FOR RIVERS UNDER PIECEWISE LINEAR v). *Assume unidimensional $X \subset R^1$, fixed-and-separable costs and piecewise-linear net valuations $v(\cdot)$. Take any solution $(\bar{x}, \bar{\tau})$ to the normalized problem (4)-(5), then:* (i) *There exist a 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 generalized first-order conditions of Lagrangian $\mathcal{L}(.)$ and supplementary inequalities for finding $(\bar{x}, \bar{\tau}, \lambda)$ from any hypothesis on G_+^λ:*

$$\frac{\partial \mathcal{L}(\bar{x}, \bar{\tau}, \lambda)}{\partial t_i} = m_i - \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} + \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} = 0 \ \ \forall i > 0, \tag{26}$$

$$\frac{\partial \mathcal{L}(\bar{x}, \bar{\tau}, \lambda)}{\partial^- x_i} = v_i'^-(\bar{x}_i) \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} - \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} v_j'(\bar{x}_i) \geq 0; \ \ \forall i > 0, \tag{27}$$

$$\frac{\partial \mathcal{L}(\bar{x}, \bar{\tau}, \lambda)}{\partial^+ x_i} = v_i'^+(\bar{x}_i) \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} - \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} v_j'(\bar{x}_i) \leq 0; \ \ \forall i > 0, \tag{28}$$

$$0 = v_i(\bar{x}_i) - \bar{\tau}_i - v_i(\bar{x}_j) + \bar{\tau}_j + \rho_{ij} \quad \forall (i,j) \in G_+^\lambda, \tag{29}$$

$$0 \leq v_i(\bar{x}_i) - \bar{\tau}_i - v_i(\bar{x}_j) + \bar{\tau}_j + \rho_{ij} \quad \forall (i,j) \notin G_+^\lambda, \tag{30}$$

$$G_+^\lambda = \{(ij) | \lambda_{ij} > 0\}, \quad \bar{x}_0 := 0, \quad \bar{\tau}_0 := 0, \tag{31}$$

where the left derivative is $v_i'^{-}(\bar{x}_i) := \frac{\partial v_i(\bar{x}_i)}{\partial^- x_i} := \lim_{t\to+0} \frac{v_i(\bar{x}_i)-v_i(\bar{x}_i-t)}{t}$ *whereas the right derivative is* $v_i'^{+}(\bar{x}_i) := \frac{\partial v_i(\bar{x}_i)}{\partial^+ x_i} := \lim_{t\to+0} \frac{-v_i(\bar{x}_i)+v_i(\bar{x}_i+t)}{t}$.

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

$$\sum_{j\in S_i^{ad}(G_+^\lambda)} \lambda_{ij} \le M_i^{PG_+^\lambda} := \sum_{j\in P(i,G_+^\lambda)\cup\{i\}} m_j \qquad \forall i; \tag{32}$$

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. Generally this modification of Proposition 4 is proved in a way similar to initial version. The only hardship is to ensure the existence of the Lagrange multipliers without guaranteed solid admissible set. However, under piecewise-linear valuations, our admissible set (convex or non-convex) is anyway a polyhedron, obviously combined of several convex polyhedrons. The objective function maximized is linear. Therefore, any local or global optimum $(\bar{x}, \bar{\tau})$ can be looked upon as a maximum of the function π on a convex polyhedron locally coinciding with the global admissible set. Such local program amounts to linear programming. Thereby, by linear (polyhedron) version of the Kuhn-Tucker theorem, any optimum can be characterized by some Lagrange multipliers, i.e., they exist.

Another distinction from the basic version of Proposition 4 here is the use of the directional derivatives instead of the usual v'. However, this generalization is rather standard: at the maximum the left derivative should be non-negative and the right one should be non-positive. When both coincide it means the usual condition $v'(x) = 0$. This generalization is explained and illustrated by the equations (34) in the proof of Proposition 2. □

Now we are ready for our longest proof.

Proof of Proposition 2.

It is sufficient to prove the proposition for the dimensionality $l = 1$ because a function $v_i(x_{i1})$ of one argument can be viewed as a function $v_i(x_{i1}, x_{i2}, ...)$ formally dependent upon a higher-dimensional argument. Therefore an example, i.e., a profile of functions $v_1(x_{11}), ..., v_n(x_{n1})$ yielding any arbitrarily given river answers our question for any dimensionality $l \ge 1$. Moreover, constructing the needed example in a higher dimensionality is much easier because of more freedom.

We take only nontrivial case $n \geq 2$ and exploit geometric reasoning. For any given parameter $\rho \geq 0$ and any given river $\bar{G} = \{i_1 \to i_2, i_3 \to i_4, ...\}$, we show how to construct the valuations and the frequencies resulting in such A-graph $\bar{G}$. Our plan during this construction is: (1) to establish some preorder among the nodes of our graph $\bar{G}$ and renumber our agents respectively; (2) under $\rho = 0$, to choose some quantities/tariffs menu $(\bar{x}, \bar{\tau})$ coordinated with this numeration and with the graph $\bar{G}$ in the sense of Lemma 2; (3) to find some net valuations v and frequencies m that generate this menu $(\bar{x}, \bar{\tau})$ as a ρ-solution such that $\bar{\bar{G}}(\bar{x}, \bar{\tau}) = \bar{G}$ (see Fig.4.2); (4) to prove optimality of $(\bar{x}, \bar{\tau})$ under these v, m; (5) to extend our reasoning onto the case $\rho > 0$.

(1). We have excluded bunching and other dicycles by considering only rivers. So, we can establish a longest-path preorder $\succeq_\vee$ (see definition in Section 3) among agents $1, 2, ..., n$ perceived as quantity/tariff packages or nodes of the graph $\bar{G}$. We sort all these nodes into several layers $L_1^\vee, L_2^\vee$,..., according to this preorder $\succeq_\vee$. In particular, the layer $L_1^\vee$ contains all agents connected to #0 by no more than one arc; all nodes connected to #0 by no more than two arcs belong to $L_2^\vee$, and so on. Such preordered partitioning of population I^n is suitable for the needed construction of packages because it is well coordinated with the arc-induced partial order $\succ_\to$ within the initial graph $\bar{G}$.

To renumber the agents, we start from the layer $L_1^\vee$. Assume there are some $k_1 \geq 1$ nodes in $L_1^\vee$. We assign the smallest successive integer numbers 1, 2, 3,..., k_1 to all agents from the layer $L_1^\vee$ chosen in any arbitrary order from $L_1^\vee$. Similarly, assuming k_2 elements in the next layer $L_2^\vee$, we assign the subsequent (higher) integer numbers: $k_1{+}1$, $k_1{+}2$,...,$k_1{+}k_2$ to all agents from the layer $L_2^\vee$ taken in any order from this layer. Afterwards, we proceed similarly by labelling the layer $L_3^\vee$ with subsequent numbers $k_1{+}k_2{+}1$, $k_1{+}k_2{+}2$,...,$k_1{+}k_2{+}k_3$ and so on. In this way we can renumber elements of all layers and exhaust population I^n. From now on, without loss of generality, we assume for simplicity of notation that agents initially had the same labels (indices) as these freshly assigned numbers 1,...,n.

(2). We stick to special case $\rho = 0$ until step 5 and then generalize the construction.

We should choose now a solution menu (x, τ) satisfying Lemma 2 for the graph $\bar{G}$, i.e., for the predecessors we should construct higher net tariffs τ_i than for the successors. Geometrically, this task of choosing an appropriate (x, τ) means drawing our graph $\bar{G}$ in the quantity-tariff space $\mathbb{R}^2$ in a specific manner (see Fig.4.2). To do this, first we assign the agents' labels (indices) according to their consumption quantities, i.e., take

$\bar{x}_1 := 1, \bar{x}_2 := 2, \bar{x}_3 := 3, ..., \bar{x}_n := n$.

Now, to assign the needed net tariffs τ_i, it is sufficient to take a square root function (or any other strictly concave strictly increasing function) of the quantities, i.e., $\bar{\tau}_i := \sqrt{\bar{x}_i}$. The needed assignment $(\bar{x}, \bar{\tau}) \in \mathbb{R}^2_+$ is constructed from the given graph $\bar{G}$ in such a way that all predecessors always have bigger net tariffs $\bar{\tau}_i$ and bigger quantities $\bar{x}_i$ than their successors (i.e., the predecessors lie to the right and above). Now we connect the obtained points $(\bar{x}_i, \bar{\tau}_i), (\bar{x}_j, \bar{\tau}_j), ...$ by the line segments which are all arcs ij of the initial graph $\bar{G}$. Thus, the needed allocation of graph $\bar{G}$ in quantity/tariff space is completed, and each arc of the re-allocated graph $\bar{G}$ goes to a *lower* net tariff, as Lemma 2 requires.

To explain this construction through an example, we illustrate a sample graph $\bar{G} = \{0 \leftarrow c \rightarrow a \rightarrow 0 \leftarrow b \leftarrow d \leftarrow e \rightarrow 0,\ d \rightarrow a\}$, $\rho = 0$, in Fig.2. The left panel shows step 1: this graph is sorted into 3 layers according to the longest-path preorder $\succeq_{\vee}$. The middle panel shows the next step of enumerating the nodes according to their layers ($a := \#1, b := \#2, ...$) and assigning them the menu $(\bar{x}, \bar{\tau}) := ((1, 1), (2, \sqrt{2}), ...)$. In essence, by steps 1 and 2 our initial graph $\bar{G}$ is only turned and stretched, i.e., specifically positioned in the quantity-tariff space. The arrows (arcs) are drawn with different types of dashes to distinguish connections between each node and to highlight that all arcs go leftward and downwards. By taking $\sqrt{\cdot}$, we have ensured such a position of graph $\bar{G}$ where an arc can never go above any point. One can understand now that such a reallocation is possible for *any* river $\bar{G}$. The right panel illustrates making some valuations from the arcs obtained.

(3). Now from the arcs obtained we construct the active indifference curves $\check{v}_i$ and such net valuations v_i that could generate our assignment $(\bar{x}, \bar{\tau})$ as an optimal solution. We make these active curves $\check{v}_i$ piecewise-linear by taking the existing arcs and adding some segments as illustrated in the right panel of Fig.2. The general requirement for $(\bar{x}, \bar{\tau})$ to be a solution is that the i-th active indifference curve $\check{v}_i$ under construction should go through *all* points $(\bar{x}_j, \bar{\tau}_j)$ to which i is *adjacently* connected as a predecessor ($i \rightarrow j$). We first explain in detail such construction and then summarize it in equation (26) below.

3.1. Each package-point $(\bar{x}_i, \bar{\tau}_i)$ is already connected by some directed linear segments (arcs) denoted now as $A_{ik} := [(\bar{x}_i, \bar{\tau}_i), (\bar{x}_k, \bar{\tau}_k)]$ to one or more lower packages according to arcs $i \rightarrow k, i \rightarrow j, ...$ of initial graph $\bar{G}$. We construct now the graph U_i under the i-th active indifference curve $\check{v}_i$ by including all these segments into $U_i \ni A_{ik} \forall k$. Further, U_i should be extended leftward and rightward by some additional segments to define curve $\check{v}_i$

on the whole half-space $\mathbb{R}_+$.

3.2. Specifically, to extend the curve $\check{v}_i$ to the *root*, no supplement is needed if $(\bar{x}_i, \bar{\tau}_i)$ is already adjacently connected to (0,0), which is the case for points $i = a, b, c, e$ in Fig.4.2. But when $(\bar{x}_i, \bar{\tau}_i)$ is not connected to #0 so far, like point d in Fig.4.2, it is sufficient to connect $(\bar{x}_i, \bar{\tau}_i)$ by a line segment to some point strictly below the root $(0, 0)$, lower for some small magnitude $\delta > 0$. We use from now on specific value $\delta = \frac{1}{2n^{2n}}$ and show that it is sufficiently small for our purpose. We connect one destination point $(0, -\delta)$ of this kind to all agents non-adjacent to (0,0) and respectively construct a starting-segment $A_{i0} := [(\bar{x}_i, \bar{\tau}_i), (0, -\delta)]$ for each node $i : i \not\to 0$.

3.3. As to the *upper* interval $(\bar{x}_i, +\infty)$ of the domain for the valuation v_i, it is sufficient to take any non-increasing ray $A_{i\infty}$ to make a maximum or a summit out of $(\bar{x}_i, \bar{\tau}_i)$. In particular, for each node i we can take now any strictly decreasing ray, for instance, $A_{i\infty} := [(\bar{x}_i, \bar{\tau}_i), (\bar{x}_i + z, \bar{\tau}_i - 1 * z)) : \ \forall z > 0$ (see Fig.4.2). Such a non-increasing ray $A_{i\infty}$ similar to initial segments do not go above any other nodes $j \neq i$ because high-labelled nodes lie higher. The same property is true for the starting-segments A_{i0} due to strict concavity of $\sqrt{\cdot}$. Any initial segment $A_{i,i+k} := [(\bar{x}_i, \bar{\tau}_i), (\bar{x}_{i+k}, \bar{\tau}_{i+k})]$ also goes below the intermediate points because its ends and all points belong to strictly concave curve $(x, \sqrt{x})$. This means that the new segments of $\check{v}_i$ as well as the old ones do not violate any IC constraints.

3.4. Now, for any given agent i, we use all initial and additional segments $\{A_{i0}, A_{i1}, A_{i2}, ..., A_{i\infty}\}$ going radially out of her package $(\bar{x}_i, \bar{\tau}_i)$ for constructing the under-graph set U_i of her active valuation curve $\check{v}_i$. Namely we define U_i as the convex hull $U_i = Hull\{A_{i0}, A_{i1}, ..., A_{i\infty}\}$ of all points belonging to these segments. The needed active indifference curve $\check{v}_i$ is defined as the upper envelope of this set U_i. Finally, if the node is adjacent to (0,0), then the i-th net valuation is assigned as the active indifference curve itself: $v_i := \check{v}_i$. In the opposite case the valuation is taken higher for δ than the active curve $\check{v}_i$. The final formula summarizes this construction of net valuations as follows:

$$v_i(z) := \left[\begin{array}{ll} \check{v}_i(z) & \text{if } (i \to 0) \in \bar{G}; \\ \check{v}_i(z) + \delta & \text{if } (i \to 0) \notin \bar{G}. \end{array} \right. ; \quad \check{v}_i(z) := \max_{(z,\tau) \in Hull\{A_{i0}, A_{i1}, ..., A_{i\infty}\}} \tau \ .$$

To comprehend this construction, compare the middle and the right panels in Fig.2. Here point $\#1 = a$ was connected only to the root by the segment $[(0, 0), (1, 1)]$, so its valuation according to the above formula becomes $v_a(z) := \check{v}_a(z) := \min\{1 * z, \ \frac{4}{3} - \frac{1}{3}z\} \ \forall z \geq 0$, and shown by thin solid lines. Similarly, for $\#2 = b$ its valuation becomes $v_b(z) := \check{v}_b(z) :=$

$\min\{\frac{\sqrt{2}}{2}z,\ \sqrt{2}+\frac{2}{3}-\frac{1}{3}z\}$, depicted by thin dashed lines. Similar is the dotted curve $v_c(z) := \check{v}_c(z) := \min\{z,\ \frac{3-\sqrt{3}}{2}+\frac{\sqrt{3}-1}{2}z,\ \sqrt{3}+1-\frac{1}{3}z\}$, and the thick-dashed curve $v_e(z) := \check{v}_e(z) := \min\{\frac{1}{2}z,\ 10-4\sqrt{5}+\frac{\sqrt{5}-2}{1}z,\ \sqrt{5}+\frac{5}{3}-\frac{1}{3}z\}$. Node $\#4 = d$ not adjacent to (0,0) is different, its active indifference curve $\check{v}_d(z) = \min\{-\frac{1}{2}+\frac{3}{2}z, 2-\sqrt{2}+\frac{\sqrt{2}-1}{1}z, 2\sqrt{2}-2+\frac{2-\sqrt{2}}{2}z, \frac{10}{3}-\frac{1}{3}z\}$ (drawn as thick solid lines) contains artificial point $(0,-\delta)$, and the related valuation $v_d(z) := \frac{1}{2}+\check{v}_d(z)$ is somewhat higher than $\check{v}_d$.

By using the convex hull operation, we always arrive at *concave* net valuations v_i. Besides, each active indifference curve $\check{v}_i$ connects its peak $(\bar{x}_i, \bar{\tau}_i)$ only with those points j whose arcs (i,j) belong to the initial graph $\bar{G}$. Other constraints not included into $\bar{G}$ are not active, i.e., they are satisfied as strict inequalities because their peaks $(\bar{x}_k, \bar{\tau}_k)$ lie higher than $\check{v}_i$. Thus, the obtained system of active indifference curves exactly generates initial graph $\bar{G}$ of the active constraints. We have seen already that none of the IC constraints are violated and so our plan $(\bar{x}, \bar{\tau})$ is *feasible* for the obtained net valuations $v_i(.)$ and indeed it generates the needed almost–envy graph in the sense $\bar{G} = \bar{\bar{G}}(\bar{x}, \bar{\tau})$.

3.5. Now we should choose such frequencies m_i that make our feasible plan $(\bar{x}, \bar{\tau})$ also *optimal* for these net valuations v_i. To make it easy, each quantity $\bar{x}_i$ have been constructed at the peak of its net valuation v_i, thereby quantities $\bar{x}$ are the first-best ones (i.e., $v_i(\bar{x}_i) = \arg\max_{x_i} v_i(x_i)\ \forall i$). Moreover, within the same A-graph $\bar{G}$, profit cannot be improved by changing the whole menu (x, τ) because variables $\bar{x}_i$ are first-best, while $\bar{\tau}_i$ cannot be increased any more without violating some IC constraints under this $\bar{x}$ (more detailed explanation follows). However, a more delicate task is to choose now appropriate frequencies m_i that ensure that any *other* graph is not better for profit. In particular, it should not be profitable to delete (in the sense $(x_i, \tau_i) := 0$) any package from the menu for achieving an increase in other tariffs.

To construct weights m_i guaranteeing such impossibility of improvement, first we denote the left derivative of v_i at x_i as $v_i'^-(x_i) := \lim_{t\to+0} \frac{v_i(x_i)-v_i(x_i-t)}{t}$ and the right derivative as $v_i'^+(x_i) := \lim_{t\to+0} \frac{-v_i(x_i)+v_i(x_i+t)}{t}$ (both directional derivatives are finite at all x since v_i is piece-wise linear and each one coincides with one of the slopes among the segments A_{ik}). At $x_i = 0$ we artificially define the left derivative as $v_i'^-(0) = v_i'^+(0)$. Now, studying all the derivatives, we can find the maximal slope $\check{s}$ and the a lower estimate $\hat{s}$ of the minimal slope among all linear segments of all valuation curves:

$$\check{s} := 1 = \max_i \{ \sup_{x_i \in (0,\infty)} |v_i'^{-}(x_i)|, \sup_{x_i \in (0,\infty)} |v_i'^{+}(x_i)| \}, \quad \hat{s} := (\sqrt{n} - \sqrt{n-1}).$$

To comprehend the formula, note that both ends of these segments typically belong to the curve $\sqrt{.}$, except the left ends of the type $(0, -\delta)$. The steepest segment is the first one going from (0,0) to (1,1). The most flat segment can go from $(n-1, \sqrt{n-1})$ to $(n, \sqrt{n})$ at least when there is an arc $(n \to (n-1))$ in the graph, otherwise the flattest segment is steeper. Thus, instead of the minimal slope $\min_i \{\inf_{x_i \in (0,\infty)} |v_i'^{-}(x_i)|, \inf_{x_i \in (0,\infty)} |v_i'^{+}(x_i)|\}$, we can take its lower estimate $\hat{s} := \sqrt{n} - \sqrt{n-1}$ which is weakly smaller than the slope of any segment.

For subsequent construction we use magnitude

$$\hat{\varepsilon} := \frac{\hat{s}}{2n\check{s}} = \frac{\sqrt{n} - \sqrt{n-1}}{2n}$$

that decreases in n and satisfies bound $\hat{\varepsilon} < 1/8$ for all $n \geq 2$.

Now we can assign the first frequency $\bar{m}_1 = 1$ and other frequencies as

$$\bar{m}_i := \hat{\varepsilon}^{(i-1)} = \left(\frac{\sqrt{n} - \sqrt{n-1}}{2n} \right)^{i-1} \quad \forall i.$$

Each $\bar{m}_i < \bar{m}_{i-1}$ because of higher power $(i-1)$ for higher i and small $\hat{\varepsilon}$. This construction intends to ensure the needed doubled domination of any $(i < n)$ frequency over the sum of all higher frequencies:

$$\frac{\bar{m}_i}{2} = \frac{\hat{\varepsilon}^{(i-1)}}{2} > \sum_{j=i+1}^{n} \bar{m}_j = \sum_{j=i+1}^{n} \hat{\varepsilon}^{(j-1)} = \frac{\hat{\varepsilon}^i (1 - \hat{\varepsilon}^{n-i})}{1 - \hat{\varepsilon}},$$

where we have used the formula for the sum of finite geometric progression, and the inequality holds true because $\hat{\varepsilon} < 1/8$.

(4). Why our menu $(\bar{x}, \bar{\tau}) = ((1,1), (2, \sqrt{2}), ...)$ is *optimal* under these $v, \bar{m}$? The general idea is that under these frequencies $\bar{m}$ we can optimize the components of the menu separately, one-by-one or layer by layer from below to upwards. Very small higher weights $\bar{m}_i$ and piecewise-linear net-valuation curves enable such sequence. Since agents with higher indices are comparatively rare, their influence on the lower packages is negligible, so they are not essential for choosing (x_i, τ_i) at the lower layers. In this case the optimal choice of each (x_i, τ_i) becomes obvious. Just each x_i should be taken at the peak of the related curve $v_i(.)$ exactly as our $\bar{x}_i$, and after choosing x_i all tariffs τ_i should be maximized subject to all constraints and already fixed quantities $\bar{x}_i$. This simple algorithm results exactly in our $\bar{\tau}$.

To implement this general idea, we should only check are the weights $\bar{m}_i$ of higher-numbered packages small enough to make x_i, τ_i irrelevant for the optimal positions of x_k, τ_k with lower numbers? We start with discussing any optimal menu $(\tilde{x}, \tilde{\tau})$ and show that it cannot differ from our menu $(\bar{x}, \bar{\tau})$.

4.1. Bounds on vertical deviation from $\bar{\tau}$ and graph-order of $(\tilde{x}, \tilde{\tau})$.

First we construct the *upper bound* on the optimal value of the profit function maximized through designing the so-called "first-best-optimal" menu. This means optimization w.r.t. the participation constraints $v_j(x_j) \geq 0$ only. The structure of the "first-best" menu $(\breve{x}, \breve{\tau}) = (\bar{x}, \breve{\tau})$ is obvious:

$$(\breve{x}, \breve{\tau}) = ((1, \sqrt{1}), (2, \sqrt{2} + \xi_2), ..., (n, \sqrt{n} + \xi_n)), \quad \xi_j = \left[\begin{array}{ll} 0 & \forall (j \to 0) \in \bar{G} \\ \delta & \forall (j \to 0) \notin \bar{G} \end{array} \right. .$$

Here a small addition ξ_j to the basic tariff $\sqrt{j}$ is chosen as $\xi_j = 0$ if this node j is connected to the root (0,0) in the sense $(j, 0) \in \bar{G}$, or $\xi_j = \delta$ for the opposite case. The latter case can be understood from Fig.4.2; here any active indifference curve starting from point $(0, -\delta)$ below the origin can be lifted up for δ until the participation constraint becomes binding. Thus, the upper bound on the optimal payoff is $\breve{\pi} = 1 + \sum_{i=2}^{n} (\bar{m}_i \sqrt{i} + \bar{m}_i \xi_i) \leq 1 + \sum_{i=2}^{n} (\bar{m}_i \sqrt{i} + \bar{m}_i \delta)$. Additionally, due to small δ, the tariffs order $\breve{\tau}_1 < \breve{\tau}_2 < ... < \breve{\tau}_n$ in the first-best menu is the same as in our $(\bar{x}, \bar{\tau})$.

Further, to find a *lower bound* on the optimal value of the objective function, recall that our constructed menu $(\bar{x}, \bar{\tau}) = ((1, \sqrt{1}), (2, \sqrt{2}), ..., (n, \sqrt{n}))$ is feasible and δ-close in all dimensions to the first-best menu. So, its payoff $\bar{\pi} = \pi(\bar{x}, \bar{\tau}) = \sum_{i=1}^{n} \bar{m}_i \sqrt{i}$ differs from the first-best payoff $\breve{\pi} = \pi(\breve{x}, \breve{\tau}) = \sum_{i=1}^{n} \bar{m}_i (\xi_i + \sqrt{i})$ only by a small amount

$$\bar{l} = \breve{\pi} - \bar{\pi} \leq \delta \sum_{i=2}^{n} \bar{m}_i < \frac{1}{2} \delta = \frac{1}{4n^{2n}}.$$

The latter equality follows from step 3.2, while $\sum_{i=2}^{n} \bar{m}_i < 1/2$ is ensured by step 3.5.

We are going to show that any feasible menu $(\tilde{x}, \tilde{\tau})$ brings a bigger loss than $\bar{l}$, if $(\tilde{x}, \tilde{\tau})$ "essentially" differs from $(\bar{x}, \bar{\tau})$ in the following sense: some package $(\tilde{x}_i, \tilde{\tau}_i)$ steps in vertical direction outside some ζ-vicinity of the current package $(\bar{x}_i, \bar{\tau}_i) = (i, \sqrt{i})$, namely, $|\tilde{\tau}_i - \bar{\tau}_i| > \zeta_i = \frac{\sqrt{i} - \sqrt{i-1}}{2}$. Here ζ_i denotes half of the distance from $\bar{\tau}_i = \sqrt{i}$ towards the lower neighbor. We show now impossibility of such a big "distortion" of $\tilde{\tau}_i$ compared to the first-best plan.

Any big amount $\zeta_i > \delta$ of upward distortion is impossible because of the participation constraint. To study the downward distortion, let us estimate the loss $\tilde{l}$ in profit from a downward shift of $\tilde{\tau}_i$ for the amount ζ_i, compared to the first-best solution $(\breve{x}, \breve{\tau})$. When $\tilde{\tau}_i - \breve{\tau}_i \geq \zeta_i$ it follows that

$$\tilde{l} = \breve{\pi} - \tilde{\pi} \geq \zeta_i \bar{m}_i = n\left(\frac{\sqrt{n} - \sqrt{n-1}}{n}\right)^i > n\left(\frac{1 - \sqrt{1 - 1/n}}{\sqrt{n}}\right)^n$$

even when all other variables $\tilde{\tau}_j$ are first-best optimal in the sense $\tilde{\tau}_j = \breve{\tau}_j\ \forall j \neq i$. We simplify $n\left(\frac{1-\sqrt{1-1/n}}{\sqrt{n}}\right)^n = n\left(\frac{(1-(1-1/n))}{(1+\sqrt{1-1/n})\sqrt{n}}\right)^n > \frac{1}{(2\sqrt{n})^n}$ and check comparison

$$\bar{l} < \frac{1}{2}\delta = \frac{1}{4n^{2n}} = \frac{1}{((2n)^2)^n} < \frac{1}{(2\sqrt{n})^n} < \tilde{l},$$

that holds for any package i. So, any plan ζ_i-different in tariffs from the first-best $(\breve{x}, \breve{\tau})$ cannot be optimal. Therefore, the order of any optimal $\tilde{\tau}_1 < \tilde{\tau}_2 < ... < \tilde{\tau}_n$ is the same as the order of our current plan; high-numbered packages lie higher in space. Then, by Lemma 2, the high-numbered packages must lie higher in the A-graph of any optimal $(\tilde{x}, \tilde{\tau})$, as they do in our $\bar{G}$.

4.2. **Generalized FOC**. Under our $v, \bar{m}$, we are going to prove the absence of any horizontal distortion in any optimal menu $(\tilde{x}, \tilde{\tau})$. This means proving $\tilde{x}_i = i$. For this goal, we use Proposition 4* that generalizes Proposition 4 (which is proved without Proposition 2 considered now). The generalization extends Proposition 4 from differentiable to piece-wise linear functions, like in our construction $v, \bar{m}$. By Proposition 4*, the Lagrange multipliers should exist for the Lagrangian $\mathcal{L}$ which is now a piece-wise differentiable function:

$$\mathcal{L}(x, \tau, \lambda) := \sum_{i=1}^{n}(m_i \tau_i - \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} \cdot (v_i(x_i) - \tau_i + \rho - v_i(x_j) + \tau_j)).$$

Usual first-order conditions (14)-(20) from Proposition 4 are naturally modified by Proposition 4* into the following generalized first-order conditions (GFOC) that use the directional derivatives $v_i'^{-}(.), v_i'^{+}(.)$ of the Lagrangian instead of usual derivatives to characterize any optimal $(\tilde{x}, \tilde{\tau})$:

$$\frac{\partial \mathcal{L}(\tilde{x}, \tau, \lambda)}{\partial \tau_i}(\tilde{\tau}) = m_i - \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} + \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} = 0 \;\; \forall i > 0, \tag{33}$$

$$\frac{\partial \mathcal{L}(x, \tilde{\tau}, \lambda)}{\partial^- x_i}(\tilde{x}) := v_i'^-(\tilde{x}_i) \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} - \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} v_j'^-(\tilde{x}_i) \geq 0 \;\; \forall i > 0, \tag{34}$$

$$\frac{\partial \mathcal{L}(x, \tilde{\tau}, \lambda)}{\partial^+ x_i}(\tilde{x}) := v_i'^+(\tilde{x}_i) \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij} - \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} v_j'^+(\tilde{x}_i) \leq 0 \;\; \forall i > 0, \tag{35}$$

$$0 = v_i(\tilde{x}_i) - \tilde{\tau}_i - v_i(\tilde{x}_j) + \tilde{\tau}_j + \rho_{ij} \;\; \forall (i,j) \in G_+^\lambda := \{(ij) | \lambda_{ij} > 0\}, \tag{36}$$

$$0 \leq v_i(\tilde{x}_i) - \tilde{\tau}_i - v_i(\tilde{x}_j) + \tilde{\tau}_j + \rho_{ij} \;\; \forall (i,j) \notin G_+^\lambda, \; \tilde{x}_0 := 0, \; \tilde{\tau}_0 := 0. \tag{37}$$

The first equation (33) here has the obvious meaning, it summarizes all multipliers of τ_i in the Lagrangian. The second condition (34) says that it is unprofitable to decrease variable x_i below the optimal value $\tilde{x}_i$, while (35) states that it is unprofitable to increase x_i. Here we have used left and right derivatives $v_i'^-(\tilde{x}_i), v_i'^+(\tilde{x}_i)$ coinciding with the left and right slopes of the piece-wise linear function v_i. These conditions (34)-(35) naturally replace one condition (15) from Proposition 4. The latter two equations (16)-(18) just repeat the supplementary slackness conditions and inactive inequalities.

4.3. **No horizontal deviation from** $\bar{x}$. Implementing our general idea of sequential optimization, consider now optimization of the first package (x_1, τ_1). If we had zero higher weights $m_i = 0 \; \forall i > 1$, then we could ignore all high-numbered variables $x_i, \tau_i : i > 1$ and choose exactly the constructed package $(\bar{x}_1, \bar{\tau}_1) = (1, \sqrt{1})$ which is "non-distorted" in the sense $v_i(\bar{x}_i) = \arg\max_{x_i} v_i(x_i)$ ("first-best"). Besides, the entire welfare is appropriated as profit $\bar{\tau}_1 = v_1(\bar{x}_1)$, so the profit contribution $\pi_1 = \bar{m}_1 \tau_1$ from this package cannot be improved. Now let us see that the same outcome $(\bar{x}_1, \bar{\tau}_1)$ of optimization occurs even under non-zero frequencies $\bar{m}_{i>1} > 0$, because these $\bar{m}_i$ are designed sufficiently small to not influence lower package (x_1, τ_1).

To ensure $\tilde{x}_1 = \bar{x}_1 = 1$, note that some optimal menu should exist. Indeed, in our example all functions v_i are continuous, first increasing from 0 to some peak and then decreasing to 0 and below, whereas we maximize the weighted sum $\sum m_i \tau_i$ under constraints $\tau_i \leq v_i(x_i)$ from above and additional constraints. Therefore, the admissible set for (x, τ) can be artificially bounded from below by $0 \in R_{++}^{2n}$ and from the above by some big $M \in R_{++}^{2n}$ without changing the solutions. The new admissible set is equivalent, non-empty, bounded and compact. This suffices for existence of optima.

Based on the existence, to check the optimality of non-distorted value $\bar{x}_1 = 1$ in our example we can ignore second-order conditions. It is sufficient to apply the necessary conditions (33)-(37) to any optimal point $\tilde{x}$ and observe these conditions satisfied only under $\tilde{x}_1 = \bar{x}_1$.

First check relation (35) holding at $\tilde{x}_1 = \bar{x}_1$. It means that increasing x_1 beyond $\bar{x}_1$ is unprofitable. Indeed, using (33) reformulated as $m_i + \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ji} = \sum_{j \in S_i^{ad}(G_+^\lambda)} \lambda_{ij}$, the "unprofitable-increasing" relation (35) for any active graph $G_+^\lambda = G(\tilde{x}, \tilde{\tau})$ can be reformulated as

$$v_1'^{+}(\tilde{x}_1)m_1 \leq \sum_{j \in P_1^{ad}(G_+^\lambda)} (v_j'^{+}(\tilde{x}_i) - v_1'^{+}(\tilde{x}_1))\lambda_{j1} \ \leq \sum_{j>1} (v_j'^{+}(\tilde{x}_i) - v_1'^{+}(\tilde{x}_1))m_j. \tag{38}$$

To ensure these inequalities, we construct a chain:

$$\frac{m_1}{2} > \sum_{j>1} m_j \geq \sum_{j \in P_i^{ad}(G_+^\lambda)} m_j \geq \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ij}. \tag{39}$$

The first strict inequality here reflects the choice of $\bar{m}_i$ in step 3.5 (irrespective of the unknown graph $G_+^\lambda = G(\tilde{x}, \tilde{\tau})$ of the solution). The latter inequality is true because of the upper bound on $\sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ij}$ established in Proposition 4* and expressed now as $\sum_{k \in S_1^{ad}(G_+^\lambda)} \lambda_{1k} = m_1 + \sum_{j \in P_1^{ad}(G_+^\lambda)} \lambda_{j1} \leq \sum_{j \in P(1, G_T)} m_j + m_1$.

Using chain (39), we start checking inequality (38). At the peak, by construction the right derivative is negative: $v_1'^{+}(\tilde{x}_1) = -1$, and $(\sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ij} \geq \sum_{j \in P_i^{ad}(G_+^\lambda)} \lambda_{ij} (\frac{v_j'^{+}(\tilde{x}_i)}{v_1'^{+}(\tilde{x}_1)} - 1))$ because $|v_j'^{+}(.)| \leq 1, |v_j'^{-}(.)| \leq 1$ always holds in our example and $|v_j'^{+}(\tilde{x}_i) - v_1'^{+}(\tilde{x}_1)| \leq 2$ holds based on step 3.5. So, multiplying the chain (39) by $v_1'^{+} < 0$, we get (38) which holds true at $\bar{x}_1 = 1$ and at a higher $x_1 > \bar{x}_1$. That is, any increase of x_1 at points $\tilde{x}_1 \geq \bar{x}_1$ rightward from $\bar{x}_1$ is unprofitable. But increase is profitable at any point $\tilde{x}_1 < \bar{x}_1$ to the left of $\bar{x}_1$, because there the right derivative is positive: $v_1'^{+}(\tilde{x}_1) > 0$. Using our chain as before, we ensure this profitability because inequality $m_1 \leq \sum_{j \in P_1^{ad}(G_+^\lambda)} \lambda_{j1}(v_j'^{+}(\tilde{x}_i) - v_1'^{+}(\tilde{x}_1))/v_1'^{+}(\tilde{x}_1)$ is violated. Thus, any optimal point $\tilde{x}_1$ cannot be smaller than $\bar{x}_1 = 1$ (any leftward distortion is excluded).

To exclude rightward distortion, we reformulate the condition (34) as

$$v_1^{-}(\tilde{x}_1)m_1 \geq \sum_{j \in P_1^{ad}(G_+^\lambda)} \lambda_{j1}(v_j'^{-}(\tilde{x}_i) - v_1'^{-}(\tilde{x}_1)) \ .$$

and check it similarly to leftward distortion. The condition holds at $\tilde{x}_1 = \bar{x}_1 = 1$ and to the left of $\bar{x}_1$ but appears violated at all points $\tilde{x}_1 > \bar{x}_1$, i.e., rightward from $\bar{x}_1$ where $v_1'^{-}(\tilde{x}_1) < 0$. Summarizing, the only optimal horizontal position for the first package can be $\tilde{x}_1 = \bar{x}_1 = 1$.

Now we can check in a similar fashion the absence of distortion for any other package $i > 1$. Based on step 4.1, we are sure that its predecessors $j \in P_i^{ad}(G_+^\lambda)$ in the A-graph has higher numbers than i and substantially smaller weights $\bar{m}_j$ than $\bar{m}_i$. Then the logic of excluding horizontal distortion works similarly for all i. Therefore, $\tilde{x} = \bar{x}$.

4.4. **Finding tariffs** $\tilde{\tau}$. After fixing the optimal quantities at $\tilde{x} = \bar{x} = (1, 2, ..., n)$, the optimization of remaining variables $\tau = (\tau_1, ..., \tau_n)$ becomes a very simple linear-optimization program: $\max_\tau \sum_i \bar{m}_i \tau_i$ s.t. $v_{ii} - \tau_i \leq v_{ij} - \tau_j \; \forall i, j$, where magnitudes $v_{ik} = v_i(x_k)$ are already known and the list of possible active constraints is already tightly restricted by step 4.1. This fact helps to find the optimal $\tilde{\tau}_i$ rather easily and recursively; first for $i = 1$, then for $i = 2$ and so on. Namely for $i = 1$ the only possible active arc is known as $1 \to 0$. So, we find $\tilde{\tau}_1 = \bar{\tau}_1 = \sqrt{1}$, this segment $S_{10} = [(0,0),(1,1)]$ of active indifference curve v_1 being built from arc $1 \to 0$ of the initial graph G. Then, we find $\tilde{\tau}_2 = \bar{\tau}_2 = \sqrt{2}$, because we have built this active curve either from segment S_{20}, or from S_{21}. The segment goes from point $(2, \sqrt{2})$ to one of the two lower already fixed points (0,0) or $(1, \sqrt{1})$ of our plan $(\bar{x}, \bar{\tau})$. An active indifference curve v_2 cannot let tariff $\tilde{\tau}_2$ go higher than $\sqrt{2}$, whereas there is no need for making it lower. Knowing from step 4.1 that all other constraints $(2 \to j)$ $j \neq 0, 1$ do not matter in choosing $\tilde{\tau}_2$. Similarly, in choosing $\tilde{\tau}_3$ only the arcs $3 \to 0$, $3 \to 1$, $3 \to 2$ may matter and one of these constraints is already constructed as active, and the other two are satisfied at point $\tilde{\tau}_3 = \bar{\tau}_3 = \sqrt{3}$. Proceeding in similar way for all i, we come to the needed equivalence $\tilde{\tau}_i = \bar{\tau}_i = \sqrt{i}$.

To summarize our steps 4.1-4.4, all quantities are non-distorted ($\tilde{x}_i = \bar{x}_i$) and all tariffs in any optimal menu also coincide with our constructed menu: $\tilde{\tau}_i = \bar{\tau}_i$. Thus, our plan $(\bar{x}, \bar{\tau})$ appears optimal under our valuations v_i and frequencies $\bar{m}_i$. Thus we have built valuations and frequencies resulting in the optimal plan with a given A-graph. This proves our proposition for the case $\rho = 0$.

(5) Relaxation $\rho > 0$. So far in steps 2, 3 and 4 we have studied the unrelaxed case of $\rho = 0$. Now we extend our construction and proof to the case $\rho > 0$.

In this case, for any solution the geometry of active indifference curves is as follows.

Feasibility of (x, τ) under relaxed constraint (i, j) of the type $v_i(x_i) - t_i + \rho \geq v_i(x_j) - t_j$ means that the active indifference curve $\hat{v}_i$ containing the "envying" package (x_i, τ_i) can go slightly *above* the envied package (x_j, τ_j), namely higher for magnitude ρ. In other words, each package (x_j, τ_j) has its higher duplicate $(x_j, \tau_j + \rho)$, and a segment $(i \to j)$ of active indifference curve $\hat{v}_i$ instead of connecting (x_i, τ_i) and (x_j, τ_j) connects i to j-th upper duplicate $(x_j, \tau_j + \rho)$.

Having this in mind, we construct the same menu $(\bar{x}, \bar{\tau}) = ((1, 1), (2, \sqrt{2}), ..., (n, \sqrt{n}))$ as previously, but slightly different active indifference curves $\hat{v}_i$. They go somewhat higher than the previous curves everywhere, except for their peaks $(\bar{x}_i, \bar{\tau}_i)$ and the declining segments $S_{i\infty}$. More precisely, each segment S_{ij} of active indifference curve instead of connecting point $(\bar{x}_i, \bar{\tau}_i)$ with lower point $(\bar{x}_j, \bar{\tau}_j)$ connects now $(\bar{x}_i, \bar{\tau}_i)$ with a slightly higher point $(\bar{x}_j, \bar{\tau}_j + \rho)$. The assumption $\rho \in [0, \frac{1}{2}\sqrt{n+1} - \frac{1}{2}\sqrt{n})$ ensures that each segment S_{ij} still goes downward. Then, such a slight modification of the active curves and valuations makes no difference for the rest of the proof, which remains exactly the same as for $\rho = 0$.

The proposition is proved. $\square$

We can suggest the following plausible enforcement of Proposition 2. We have a sketch of its proof through ε-approximations of valuations v_i used in Proposition 2.

Conjecture 1. *Proposition 2 can be enforced in two respects: (i) net valuations yielding the given A-graph $\bar{G}$ can be constructed strictly concave and smooth, (ii) resulting LA-graph also can coincide with $\bar{G}$.*

This conjecture is worth studying because the set of LA-constraints (those with positive Lagrange multipliers) often do differ from the active constraints under bunching or non-smooth valuations. In particular, one can see that LA-graph in our example depicted in Fig.2 does not coincide with the A-graph $\bar{\bar{G}}(\bar{x}, \bar{t})$. Indeed, relaxing or abolishing the two constraints $c \to a$, $e \to d$ cannot enhance the optimal profit because the related packages $(x_c, \tau_c), (x_e, \tau_e)$ are attached to (0,0) by the bypass constraints $c \to 0$, $e \to 0$. So, the active constraints $c \to a$, $e \to d$ should have zero Lagrange multipliers. It is well know that under smooth optimization such "over-constrained" situations are "rare," and this is why the above conjecture could help in characterizing the solutions.

References

[1] Andersson, T., 2005. "Profit maximizing nonlinear pricing." *Economic letters* 88, 135-139.

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

[3] Armstrong, M., Rochet, J-C., 1999. "Multi-dimensional Screening:" A User's Guide, *European Economic Review* 43, 959-79.

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

[7] Harary, F. and E.M. Palmer., 1973. "*Graphical Enumeration*," Academic press, New York and London.

[8] Kokovin, S., Nahata, B., and Zhelobodko, E., 2010. "Multidimensional screening under nonlinear costs: Limits of standard approach," *Economic letters*, forthcoming.

[9] Nahata, B., Kokovin, S., and Zhelobodko, E., 2002. "Package Sizes, Tariffs, Quantity Discount and Premium," http://econpapers.repec.org/paper/wpawuwpgt/0307002.htm.

[10] Rochet, J.-C., 1987. "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context," *Journal of Mathematical Economics*, 16(2), 191-200.

[11] 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.

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

[13] Rockafellar, R.T., 1993. “Lagrange Multipliers and Optimality,” *SIAM Review* v.35, No.2

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

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

[16] Vohra, R.V., 2008 “Paths, Cycles and Mechanism Design,” mimeo, Kellogg School of Business.