

## MIT Open Access Articles

*Ironing without control*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Toikka, Juuso. "Ironing Without Control." Journal of Economic Theory 146, no. 6 (November 2011): 2510–2526.

**As Published:** <http://dx.doi.org/10.1016/j.jet.2011.06.003>

**Publisher:** Elsevier

**Persistent URL:** <http://hdl.handle.net/1721.1/103942>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-NonCommercial-NoDerivs License



# Ironing without control\*

Juuso Toikka

Department of Economics, Massachusetts Institute of Technology  
50 Memorial Dr, Cambridge, MA 02142, USA

E-mail address: `toikka@mit.edu`

Phone: (617) 324-3666; Fax: (617) 253-1330

June 17, 2011

## Abstract

I extend Myerson's [R. Myerson, Optimal auction design, Math. Oper. Res. 6 (1981) 58–73] ironing technique to more general objective functions. The approach is based on a generalized notion of virtual surplus which can be maximized pointwise even when the monotonicity constraint implied by incentive compatibility binds. It is applicable to quasilinear principal-agent models where the standard virtual surplus is weakly concave in the allocation or appropriately separable in the allocation and type. No assumptions on allocation rules are required beyond monotonicity.

**Keywords:** Optimization, Monotonicity constraint, Mechanism Design, Ironing

**JEL classification codes:** C60, C70, D80

## 1 Introduction

In quasilinear principal-agent models where the agent's type is one-dimensional and her payoff function satisfies a single-crossing condition, optimal contracts are found by maximizing expected virtual surplus subject to the constraint that allocation be monotone in type (see, e.g., Fudenberg and Tirole [4], or Salanié [13]). Formally, the problem is one of maximizing a real-valued functional

$$\int_0^1 J(\phi(\theta), \theta) dF(\theta)$$

over the set of nondecreasing functions  $\phi : \Theta \rightarrow X$  mapping types to allocations. The standard approach to solving the problem when the monotonicity

---

\*I am indebted to Ilya Segal for time, advice, and encouragement. I thank Manuel Amador, Aaron Bodoh-Creed, Albie Bollard, Ken Judd, Carlos Lever, Romans Pancs, an associate editor, and two referees for comments and discussions. Financial support from the Yrjö Jahnsson Foundation is gratefully acknowledged. Errors are mine.

constraint binds—known as “ironing” because of the nature of the solution—is to apply optimal control theory [4, 13]. In this paper I build on Myerson’s [10] characterization of optimal auctions to develop an alternative method.

More specifically, I first observe that Myerson’s approach can be used in problems where the function  $J$  is appropriately separable in its arguments. For example, this class of problems includes the price discrimination model of Mussa and Rosen [9] and the regulation model of Laffont and Tirole [7]. I then extend the technique to problems where  $J$  is weakly concave in its first argument, i.e., where the virtual surplus of each type exhibits weakly decreasing marginal returns in the allocation. As discussed in Section 4, such concave problems include the aforementioned examples of separable problems as well as many models with interdependent values that generally fail separability.

Heuristically, the approach is based on a generalized notion of virtual surplus that takes into account total surplus, the agent’s information rent, and any distortions the current type causes to other types through the monotonicity constraint. As standard virtual surplus accounts only for the first two, it fails to capture the impact on the principal’s payoff of global incentive compatibility constraints that correspond to the monotonicity constraint. The generalized virtual surplus incorporates them through a convexification procedure due to Myerson [10]. It amounts to averaging the marginal contributions of types affected by the change in the allocation of the current type. In separable problems, such as in Myerson’s auction model, this can be done independently of the allocation. I show that the idea can be extended to concave problems by applying the procedure to each allocation. This yields the generalized virtual surplus which can be simply maximized pointwise even when global incentive compatibility constraints bind.

The results in this paper improve on the standard approach of using optimal control in that no restrictions are placed on admissible allocation rules beyond monotonicity. To see how such restrictions come about in the standard approach, I briefly review how the maximum principle is applied to the problem.<sup>1</sup> The challenge there is how to incorporate the monotonicity constraint, which can not be imposed on the control variable. The trick is to let the allocation  $\phi(\theta)$  be the state and take its derivative  $\phi'(\theta)$  as the control. Monotonicity is then simply the requirement that the control be non-negative. However, this formulation is unsatisfactory as it assumes that the allocation rule  $\phi$ —an endogenous object—is absolutely continuous.<sup>2</sup> For example, this assumption is violated in auctions and other trading problems with linear utility, where optimal allocations are discontinuous. Furthermore, even in models where the optimal allocation is differentiable, showing that non-differentiable rules offer no improvement requires formulating the problem in a larger space where such rules are feasible and which is no longer covered by standard results. In contrast, the approach in this paper allows for all monotone allocation rules and

<sup>1</sup>Jullien [6] provides a general exposition. For more on optimal control, see, e.g., Léonard and van Long [8], or the advanced treatment by Vinter [15].

<sup>2</sup>The state of the art in necessary conditions in the literature on optimal control requires the state  $\phi$ , also known as the *arc*, to be absolutely continuous—see Clarke [3].

thus provides a unified treatment of problems with or without jumps in the optimal allocation. It offers a possible way of assessing whether the properties assumed in the standard approach hold in applications.

Contemporaneous work by Hellwig [5] also allows for jumps in the allocation rule by extending the maximum principle to a class of problems with monotonicity constraints while relaxing absolute continuity. He builds on the generalized maximum principle of Clarke [2], which relies on non-smooth analysis. In contrast, the approach in the present paper is elementary and self-contained.

Nöldeke and Samuelson [11] develop an approach to ironing without optimal control, which is quite different from the present paper. Roughly put, they work with the inverse of the allocation rule. The monotonicity constraint on the inverse turns out to be non-binding more generally than the one on the allocation rule. However, their approach requires virtual surplus to be strictly concave in the allocation, which rules out jumps in the allocation rule. The approach in this paper is more general for problems where the participation constraint binds only at an extreme type, whereas Nöldeke and Samuelson also consider models where it may bind at intermediate types.<sup>3</sup>

I set up the problem in the next section. I then review Myerson's approach and observe that it applies to separable problems in Section 3. The readers familiar with the technique can proceed directly to Section 4 where I present the extension to concave problems, which is the main contribution of the paper. An appendix collects the proofs omitted from the main text.

## 2 The problem

Let  $X := [0, \bar{x}]$  and  $\Theta := [0, 1]$ . Let  $J : X \times \Theta \rightarrow \mathbb{R}$ , and let  $F : \Theta \rightarrow [0, 1]$  be a cdf with density  $f := F' > 0$ .<sup>4</sup> Let  $\mathcal{M} := \{\phi : \Theta \rightarrow X \mid \phi \text{ is nondecreasing}\}$ . This paper is concerned with problems of the form<sup>5</sup>

$$\sup_{\phi \in \mathcal{M}} \left\{ \int_0^1 J(\phi(\theta), \theta) dF(\theta) \right\}. \quad (\text{P})$$

The leading example is the expected virtual surplus maximization problem from single-agent mechanism design. With that in mind I use the following terminology:  $x \in X$  is an *allocation*,  $\theta \in \Theta$  is a *type*,  $F$  is the principal's *belief* about  $\theta$ ,  $J$  is the *virtual surplus function*, and  $\phi$  is an *allocation rule*. An allocation rule is *optimal* if it attains the supremum in (P).

<sup>3</sup>The approach presented here can be used in the latter type of problems once the types with binding participation constraints have been identified.

<sup>4</sup>It is straightforward to adapt the results to problems where  $X$  or  $\Theta$  (or both) is discrete. Details are available from the author upon request. Furthermore, it is without loss to take  $F$  to be the uniform distribution. Namely, let  $F$  be a cdf on  $[\underline{\theta}, \bar{\theta}]$ . Let  $\tilde{\phi} := \phi \circ F^{-1}$  and let  $\tilde{J}(x, \cdot) := J(x, F^{-1}(\cdot))$  for all  $x \in X$ . A change of variables then gives  $\int_{\underline{\theta}}^{\bar{\theta}} J(\phi(\theta), \theta) dF(\theta) = \int_0^1 J(\phi(F^{-1}(q)), F^{-1}(q)) dq = \int_0^1 \tilde{J}(\tilde{\phi}(q), q) dq$ .

<sup>5</sup>For the problem to be well defined, the mapping  $J(\phi(\cdot), \cdot) : \Theta \rightarrow \mathbb{R}$  has to be integrable for any  $\phi \in \mathcal{M}$ . A sufficient conditions for this is that  $J : X \times \Theta \rightarrow \mathbb{R}$  is measurable and bounded.

In the canonical principal-agent model virtual surplus takes the form

$$J(x, \theta) = v(x, \theta) + u(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_2(x, \theta), \quad (2.1)$$

where  $v$  and  $u$  are the principal's and the agent's payoff functions, and  $u_2(x, \theta)$  denotes the partial derivative of  $u$  with respect to its second argument. However, as the methods apply more broadly, I take  $J$  as the primitive and provide sufficient conditions on  $u$  and  $v$  that deliver the required properties of  $J$ .

### 3 The separable case

It is instructive to start by reviewing the technique of Myerson [10]. I show that it is applicable to problems that are separable in the following sense.

**Definition 3.1.**  $J : X \times \Theta \rightarrow \mathbb{R}$  is *separable* if there exist functions  $a, b : X \rightarrow \mathbb{R}$  and  $k, l : \Theta \rightarrow \mathbb{R}$ , where  $a$  is strictly increasing, such that for all  $(x, \theta) \in X \times \Theta$ ,

$$J(x, \theta) = a(x)k(\theta) + b(x) + l(\theta). \quad (3.1)$$

The problem (P) is *separable* if  $J$  is separable.

**Example 3.2** (Mussa and Rosen [9], Myerson [10]). *In the classic model of monopolistic price discrimination by Mussa and Rosen [9] the virtual surplus from consumer of type  $\theta$  takes the form*

$$J(x, \theta) = x \left( \theta - \frac{1 - F(\theta)}{f(\theta)} \right) - c(x),$$

where  $x$  is quality and  $c(x)$  is the cost of producing it. Taking  $c(x) = \theta_0 x$  and interpreting  $x$  as the probability of sale gives the virtual surplus in a single-agent version of the optimal auction problem of Myerson [10].

**Example 3.3** (Laffont and Tirole [7]). *In a simplified version of the workhorse model of regulation by Laffont and Tirole [7], a principal hires an agent to work on a project that generates a benefit  $b$  at an observable cost  $c(e, \theta) = \theta - e$ . Both the parameter  $\theta$  and effort  $e$  are the agent's private information. The agent's payoff is  $t - \gamma(e)$ , where  $t$  is a transfer and  $\gamma(x)$  is the cost of effort; the principal's payoff is  $b - c(e, \theta) - t$ . The virtual surplus is<sup>6</sup>*

$$J(e, \theta) = b - \theta + e - \gamma(e) - \frac{F(\theta)}{f(\theta)} \gamma'(e).$$

Letting  $x = -e$  this satisfies Definition 3.1 as long as  $\gamma'' > 0$ .

---

<sup>6</sup>Take  $c$  to be the choice variable. Then payoffs are  $t - \gamma(\theta - c)$  and  $b - c - t$ , so the virtual surplus is  $b - c - \gamma(\theta - c) - \frac{F(\theta)}{f(\theta)} \gamma'(\theta - c)$ . Using  $c = \theta - e$  we can write it in terms of effort.

It can readily be verified that the canonical virtual surplus (2.1) is separable if  $u$  is separable and values are private (i.e.,  $v$  is independent of  $\theta$ ).

Let  $J$  be separable with  $k : \Theta \rightarrow \mathbb{R}$  as in Definition 3.1. For all  $q \in [0, 1]$ , let

$$h(q) := k(F^{-1}(q)), \quad (3.2)$$

and

$$H(q) := \int_0^q h(r) dr. \quad (3.3)$$

Let  $\text{conv}H$  be the convex hull of  $H$  on  $[0, 1]$  (see, e.g., Rockafellar [12]). Define

$$G := \text{conv}H. \quad (3.4)$$

Then  $G$  is the highest convex function on  $[0, 1]$  such that  $G \leq H$ .<sup>7</sup> Since  $G$  is convex, it is continuously differentiable except possibly at countably many points. Define  $g : [0, 1] \rightarrow \mathbb{R}$  as follows. For all  $q \in (0, 1)$  such that  $G'(q)$  exists, let

$$g(q) := G'(q), \quad (3.5)$$

and extend  $g$  to all of  $[0, 1]$  by right-continuity. Finally, define  $\bar{k} : \Theta \rightarrow \mathbb{R}$  by

$$\bar{k}(\theta) := g(F(\theta)). \quad (3.6)$$

Note that  $\bar{k}$  is nondecreasing by construction.

Armed with the function  $\bar{k}$  defined above, define the *generalized virtual surplus*

$$\bar{J}(x, \theta) := a(x)\bar{k}(\theta) + b(x) + l(\theta), \quad (3.7)$$

where  $a$ ,  $b$  and  $l$  are as in Definition 3.1. The generalized virtual surplus  $\bar{J}$  differs from the virtual surplus  $J$  only in that  $k$  is replaced with  $\bar{k}$ . Define the maximizer correspondence  $\Phi : \Theta \rightarrow \mathcal{P}(X)$ , where  $\mathcal{P}(X)$  is the power set of  $X$ , by

$$\Phi(\theta) := \left\{ x \in X \mid \bar{J}(x, \theta) = \sup_{y \in X} \bar{J}(y, \theta) \right\}. \quad (3.8)$$

A function  $\phi$  is a *selection* from  $\Phi$  if  $\phi(\theta) \in \Phi(\theta)$  for all  $\theta \in \Theta$ . By Topkis [14], a monotone (i.e., nondecreasing) selection exists if  $\Phi$  is nonempty-valued, since the generalized virtual surplus  $\bar{J}$  has increasing differences in  $(x, \theta)$  by construction.

*Remark 3.4.*  $J = \bar{J}$  if and only if  $k$  is nondecreasing. (Then  $H$  is convex so that  $H = G$ ,  $h = g$  and  $k = \bar{k}$ .) In that case the optimization problem in (3.8) reduces to the standard practise of maximizing virtual surplus pointwise.

The effect of the monotonicity constraint is captured by the following condition.

---

<sup>7</sup>I.e.,  $\text{conv}H(q) := \min\{\lambda H(q_1) + (1-\lambda)H(q_2) \mid (\lambda, q_1, q_2) \in [0, 1]^3 \text{ and } \lambda q_1 + (1-\lambda)q_2 = q\}$ .

**Definition 3.5.**  $\phi : \Theta \rightarrow X$  has the *pooling property* if for all open intervals  $I \subset \Theta$ ,

$$G(F(\theta)) < H(F(\theta)) \text{ for all } \theta \in I \implies \phi \text{ is constant on } I.$$

*Remark 3.6.* Since  $F$ ,  $G$  and  $H$  are continuous, the pooling property requires  $\phi$  to be constant on any set where  $G \circ F$  and  $H \circ F$  differ.

The following theorem is a minor generalization of the single-agent version of Myerson's [10] characterization of optimal auctions; the proof is presented in the Supplementary material for completeness.<sup>8</sup>

**Theorem 3.7** (Myerson [10]). *Let  $J$  be separable and let  $\phi \in \mathcal{M}$ . Then  $\phi$  achieves the supremum in (P) if and only if  $\phi$  has the pooling property and  $\phi(\theta) \in \Phi(\theta)$  a.e.*

This result identifies solutions to (P) with particular monotone selections from  $\Phi$ . It implies the standard ironing result: Outside pooling intervals  $J = \bar{J}$  so the solution there coincides with pointwise maximization of virtual surplus.

For the purposes of finding optimal allocation rules, the remarkable part of Theorem 3.7 is given by two corollaries, which show that the transformed problem can be solved by pointwise maximization even when the monotonicity constraint binds.

**Corollary 3.8.** *Let  $J$  be separable. Assume  $\Phi(\theta)$  is non-empty and compact for all  $\theta \in \Theta$ , and let  $\phi^*(\theta) := \max \Phi(\theta)$  and  $\phi_*(\theta) := \min \Phi(\theta)$ . Then the selections  $\phi^*$  and  $\phi_*$  are monotone and achieve the supremum in (P).*

**Corollary 3.9.** *Let  $J$  be separable. Assume  $\Phi$  is single-valued except at countably many points. Then any selection from  $\Phi$  attains the supremum in (P).*

That is, the infinite-dimensional constrained maximization problem (P) is equivalent to the family of independent one-dimensional optimization problems in (3.8) in the sense that the highest and the lowest optimal allocation rule can be found by maximizing the generalized virtual surplus pointwise (without having to consider the pooling property or the monotonicity of the selection). This is true of all optimal allocation rules unless there is a non-degenerate region of types for which multiple allocations maximize the generalized virtual surplus.

## 4 The concave case

I assume in this section that  $J$  is defined and once continuously differentiable on an open set containing  $X \times \Theta$ . This implies that the integral in problem (P) is well-defined for any  $\phi \in \mathcal{M}$ . In order to simplify notation, I normalize the distribution  $F$  to be uniform.<sup>9</sup>

<sup>8</sup>Myerson only shows sufficiency, but necessity is straightforward—see Supplementary material.

<sup>9</sup>As noted in footnote 4, this entails no loss of generality. Alternatively, the normalization can be incorporated in the function  $h$  as in the separable case. Finally, if  $f$  is continuously differentiable, then it can simply be included in  $J$  by redefining  $\tilde{J}(x, \theta) := J(x, \theta)f(\theta)$ .

The results of this section require the virtual surplus  $J$  to be weakly concave in the allocation  $x$ . Examples 3.2 and 3.3 satisfy this assumption provided that the cost function  $c$  is weakly convex in Example 3.2 and we have  $\gamma''' \geq 0$  in Example 3.3. The following is an example of a non-separable concave problem.

**Example 4.1** (Mussa and Rosen [9] with interdependent values). *Example 3.2 shows that the price discrimination model of Mussa and Rosen [9] leads to a separable problem. If the seller's cost depends directly on the buyer's type (say, because some types are more demanding and hence more costly to serve), then the problem is in general no longer separable. The virtual surplus takes the form*

$$J(x, \theta) = x \left( \theta - \frac{1 - F(\theta)}{f(\theta)} \right) - c(x, \theta),$$

*which is weakly concave in  $x$  as long as the cost function  $c$  is weakly convex in  $x$  for each  $\theta$ . A version of the model is solved in section 4.4.*

More generally, the canonical virtual surplus (2.1) is not separable even under private values if the agent's utility function is not separable. It will nevertheless be concave in  $x$  if the social surplus  $v(x, \theta) + u(x, \theta)$  is weakly concave in  $x$  and the agent's marginal utility  $u_1(x, \theta)$  is supermodular (i.e.,  $u_{112} \geq 0$  so that changes in marginal utility when  $x$  varies are weakly greater for higher types).

## 4.1 Definitions

I use essentially the same notation as in the separable case to highlight the analogy between the constructions. The main difference is that now convexification has to be done allocation-by-allocation.

For all  $(x, \theta) \in X \times \Theta$ , let

$$h(x, \theta) := J_1(x, \theta), \quad (4.1)$$

and let

$$H(x, \theta) := \int_0^\theta h(x, r) dr. \quad (4.2)$$

By assumption  $h$  is continuous (and hence integrable on  $\{x\} \times [0, \theta]$ ) and thus  $H(x, \cdot)$  is continuously differentiable on  $(0, 1)$  for any fixed  $x$ . For all  $x \in X$ , let

$$G(x, \cdot) := \text{conv}H(x, \cdot). \quad (4.3)$$

$G(x, \cdot)$  is the highest convex function on  $[0, 1]$  such that  $G(x, \cdot) \leq H(x, \cdot)$ .<sup>10</sup> As the convex hull of a differentiable function,  $G(x, \cdot)$  is continuously differentiable on  $(0, 1)$ . Its derivative, denoted  $G_2(x, \theta)$ , is nondecreasing in  $\theta$ . For all  $\theta \in (0, 1)$ , let

$$g(x, \theta) := G_2(x, \theta), \quad (4.4)$$

<sup>10</sup>I.e.,  $\text{conv}H(x, \theta) = \min\{\lambda H(x, \theta_1) + (1 - \lambda)H(x, \theta_2) \mid \lambda \theta_1 + (1 - \lambda)\theta_2 = \theta, (\lambda, \theta_1, \theta_2) \in [0, 1]^3\}$ .



and extend  $g(x, \cdot)$  to all of  $[0, 1]$  by continuity.

Define the *generalized virtual surplus*  $\bar{J} : X \times \Theta \rightarrow \mathbb{R}$  by<sup>11</sup>

$$\bar{J}(x, \theta) := J(0, \theta) + \int_0^x g(s, \theta) ds. \quad (4.5)$$

Define the correspondence  $\Psi : \Theta \rightarrow \mathcal{P}(X)$  by

$$\Psi(\theta) := \left\{ x \in X \mid \bar{J}(x, \theta) = \sup_{y \in X} \bar{J}(y, \theta) \right\}. \quad (4.6)$$

Finally, let  $\Gamma := \{\phi : \Theta \rightarrow X \mid \phi \text{ is measurable}\}$ .

*Remark 4.2.* Suppose  $J$  has increasing differences (i.e.,  $J_1(x, \theta)$  is increasing in  $\theta$ ) so that maximizing virtual surplus pointwise gives a solution that is nondecreasing in  $\theta$ . Then  $H(x, \cdot)$  is convex (because  $H_2(x, \theta) = J_1(x, \theta)$ ). But then  $H = G$  and  $h = g$  so that  $J = \bar{J}$ . Thus, under the standard sufficient conditions for pointwise maximization the generalized virtual surplus reduces to the virtual surplus.

*Remark 4.3.* If  $J$  is separable (but not necessarily concave), then the above construction yields the same generalized virtual surplus (up to a constant) as the one defined in (3.7) for the separable case (see the Supplementary material for details). In this sense the above approach is a proper generalization of Myerson's.

## 4.2 The results

With the definitions in place, I am now ready to state the main results of the paper, the proofs of which are in the next subsection.

**Theorem 4.4** (Values). *If  $J$  is weakly concave in  $x$ , then*

$$\sup_{\phi \in \mathcal{M}} \left\{ \int_0^1 J(\phi(\theta), \theta) d\theta \right\} = \sup_{\phi \in \Gamma} \left\{ \int_0^1 \bar{J}(\phi(\theta), \theta) d\theta \right\}.$$

That is, in order to find the maximized expected value of the virtual surplus function over all nondecreasing allocation rules in (P), it suffices to maximize the expected generalized virtual surplus over all (measurable) allocation rules. The latter can be done pointwise. Indeed, by definition any selection from  $\Psi$  attains the supremum on the right-hand side.

In terms of economics, Theorem 4.4 shows that if the agent's utility function in the canonical principal-agent model has the single-crossing property, then the generalized virtual surplus captures all implications of incentive compatibility on the principal's problem. In this sense it generalizes the standard notion of virtual surplus which only accounts for local incentive compatibility constraints.

---

<sup>11</sup>I show below that the function  $g(\cdot, \theta)$  is continuous for all fixed  $\theta$ . It is thus bounded on the compact set  $[0, x]$  and hence integrable.

It turns out that while any solution to (P) is (almost everywhere equal to) a monotone selection from  $\Psi$ , there may be monotone selections from  $\Psi$  that are not solutions to (P).<sup>12</sup> However, the smallest and largest optimal allocation rule can be obtained by maximizing the generalized virtual surplus independently for each type. In order to state this result, define for all  $\theta \in \Theta$

$$\phi^*(\theta) := \max \Psi(\theta) \quad \text{and} \quad \phi_*(\theta) := \min \Psi(\theta). \quad (4.7)$$

**Theorem 4.5** (Maximizers). *Let  $J$  be weakly concave in  $x$ . Then  $\phi^*$  and  $\phi_*$  are monotone and attain the supremum in (P). Furthermore, if  $\phi \in \mathcal{M}$  attains the supremum in (P), then  $\phi_*(\theta) \leq \phi(\theta) \leq \phi^*(\theta)$  and  $\phi(\theta) \in \Psi(\theta)$  a.e.*

Note that this result also establishes the existence of a maximizer in problem (P). If  $\Psi$  is single-valued except at countably many points, then all selections are monotone and agree with  $\phi^*$  a.e. This gives the following analog of Corollary 3.9.

**Corollary 4.6.** *Let  $J$  be weakly concave in  $x$ . Assume  $\Psi$  is single-valued except at countably many points. Then any selection from  $\Psi$  attains the supremum in (P).*

*Remark 4.7.* The supplementary material contains an example showing that in general weak concavity of  $J$  in  $x$  cannot be dropped from the assumptions in Theorems 4.4 and 4.5. This is because the construction here considers only small changes in the allocation as it builds upon the derivative  $J_1(x, \theta)$ . Under concavity this is sufficient as then local optimality implies global optimality.<sup>13</sup>

*Remark 4.8.* Since  $\bar{J}$  is weakly concave in  $x$ , the extremal selections satisfy

$$\phi^*(\theta) = \max \{x \in X \mid g(x, \theta) \geq 0\} \quad \text{and} \quad \phi_*(\theta) = \min \{x \in X \mid g(x, \theta) \leq 0\},$$

where  $\max \emptyset = 0$  and  $\min \emptyset = \bar{x}$ .

Subsection 4.4 contains a fully solved example. While obtaining closed-form solutions can be tedious, the results can be used to derive comparative statics and other properties of the maximizers to problem (P). I illustrate this with two applications: I first argue that strict concavity of  $J$  in  $x$  is a sufficient condition for the maximizer to be unique and hence continuous. I then show that the solutions have the familiar ironing property.

<sup>12</sup>Maximizers can be characterized by a pooling property as in the separable case. A function  $\phi \in \mathcal{M}$  has the *generalized pooling property (gpp)* if for all  $x \in X$  and all open intervals  $I \subset \Theta$ ,

$$\phi^{-1}(x) \in I \quad \text{and} \quad G(x, \theta) < H(x, \theta) \quad \text{for all } \theta \in I \quad \implies \quad \phi \text{ is constant on } I,$$

where  $\phi^{-1}$  is the generalized inverse defined in (4.8). It can be shown that if  $J$  is weakly concave in  $x$ , then  $\phi \in \mathcal{M}$  attains the supremum in (P) if and only if  $\phi$  has gpp and  $\phi(\theta) \in \Psi(\theta)$  a.e. Checking gpp is in general quite complicated. Furthermore, unlike in the separable case, the result does not immediately characterize optimal pooling intervals since the intervals depend on the allocation rule  $\phi$ . For these reasons I do not pursue this direction further.

<sup>13</sup>Quasi- or pseudo-concavity does not suffice as the construction here involves taking integrals and the sum of two quasiconcave functions need not be quasiconcave.

Suppose  $J$  is strictly concave in  $x$ . The argument in the proof of Lemma 4.11 can be adapted to show that then  $\bar{J}$  is also strictly concave in  $x$ . Hence (4.6) is single-valued and thus continuous by the Maximum theorem. Thus there is a unique, continuous optimal allocation rule. The general point illustrated here is that it can be relatively easy to show that the convexification procedure preserves certain properties of  $J$  which then translate to properties of solutions.

In order to show the ironing result, note first that if  $J$  is weakly concave and continuously differentiable in  $x$ , so is  $\bar{J}$  (see Lemmas 4.10 and 4.11 below). Thus first-order conditions are necessary and sufficient when maximizing  $\bar{J}(x, \theta)$ . For simplicity, suppose that for each  $\theta$  there exists a unique interior allocation  $\phi(\theta)$  that maximizes the generalized virtual surplus  $\bar{J}(x, \theta)$ . If  $\phi(\theta)$  does not also maximize the virtual surplus  $J(x, \theta)$ , then by weak concavity of  $J$  in  $x$  we have

$$J_1(\phi(\theta), \theta) = h(\phi(\theta), \theta) \neq 0 = g(\phi(\theta), \theta) = \bar{J}_1(\phi(\theta), \theta).$$

But then  $H(\phi(\theta), \cdot)$  and  $G(\phi(\theta), \cdot)$  differ in some open neighborhood  $I$  of  $\theta$ . Thus  $g(\phi(\theta), \cdot)$  is flat on  $I$  so that  $g(\phi(\theta), \tau) = 0$  for all  $\tau \in I$ . Since we have assumed a unique maximum for each type, this implies that  $\phi(\theta)$  is assigned to all types in  $I$ . That is, if the allocation  $\phi(\theta)$  does not maximize virtual surplus from type  $\theta$ , then  $\theta$  belongs to a pooling interval. Outside these intervals the solution coincides with pointwise maximization of virtual surplus.

### 4.3 Proofs

I first establish three technical lemmas, the first two of which do not require concavity. Their proofs are somewhat tedious but not particularly illuminating, and hence relegated to the Appendix.

**Lemma 4.9.** *The function  $g$  defined in (4.4) is continuous in  $x$ .*

**Lemma 4.10.** *The generalized virtual surplus  $\bar{J}$  defined in (4.5) is continuous, continuously differentiable in  $x$ , and has increasing differences in  $(x, \theta)$ .*

**Lemma 4.11.** *If the virtual surplus  $J$  is weakly concave in  $x$ , then the generalized virtual surplus  $\bar{J}$  is weakly concave in  $x$ .*

Given the properties of  $\bar{J}$  established in Lemma 4.10, Berge's Maximum theorem [1] implies that the correspondence  $\Psi$  defined by (4.6) is upper hemicontinuous, nonempty, and compact-valued. As  $\bar{J}$  has increasing differences, the selections  $\phi^*$  and  $\phi_*$  defined in (4.7) exist and are nondecreasing.

It is convenient to introduce two more pieces of notation. For all  $\phi \in \mathcal{M}$ , define the generalized inverse  $\phi^{-1} : X \rightarrow \Theta$  by

$$\phi^{-1}(x) := \inf \{ \theta \in \Theta : \phi(\theta) \geq x \}, \quad (4.8)$$

where  $\inf \emptyset = 1$  by convention. Denote the difference between the expected virtual surplus and the expected generalized virtual surplus from allocation rule  $\phi \in \mathcal{M}$  by

$$\Delta(\phi) := \int_0^1 (J(\phi(\theta), \theta) - \bar{J}(\phi(\theta), \theta)) d\theta.$$

Since  $J(x, \theta) = \bar{J}(x, \theta) + (J(x, \theta) - \bar{J}(x, \theta))$ , the objective in (P) can be written as

$$\int_0^1 J(\phi(\theta), \theta) d\theta = \int_0^1 \bar{J}(\phi(\theta), \theta) d\theta + \Delta(\phi).$$

The following two lemmas about  $\Delta(\phi)$  are the key to the proof.

**Lemma 4.12.**  $\sup_{\phi \in \mathcal{M}} \Delta(\phi) = 0$ .

*Proof.* For all  $\phi \in \mathcal{M}$  we have

$$\begin{aligned} \Delta(\phi) &= \int_0^1 (J(\phi(\theta), \theta) - \bar{J}(\phi(\theta), \theta)) d\theta \\ &= \int_0^1 \left[ J(0, \theta) + \int_0^{\phi(\theta)} h(s, \theta) ds - \left( J(0, \theta) + \int_0^{\phi(\theta)} g(s, \theta) ds \right) \right] d\theta \\ &= \int_0^1 \int_0^{\phi(\theta)} (h(s, \theta) - g(s, \theta)) ds d\theta \\ &= \int_0^{\bar{x}} \int_{\phi^{-1}(x)}^1 (h(x, \theta) - g(x, \theta)) d\theta dx \\ &= \int_0^{\bar{x}} (G(x, \phi^{-1}(x)) - H(x, \phi^{-1}(x))) dx \leq 0, \end{aligned} \tag{4.9}$$

where the last equality follows since  $G(x, 1) = H(x, 1)$  for all  $x \in X$ . The inequality is by definition of  $G$ . The constant function  $\phi \equiv 0$  achieves the bound.  $\square$

**Lemma 4.13.** *If  $J$  is weakly concave in  $x$ , then  $\Delta(\phi^*) = \Delta(\phi_*) = 0$ .*

Given any  $\phi \in \mathcal{M}$ , let

$$A_\phi := \{x \in (0, \bar{x}) \mid G(x, \phi^{-1}(x)) - H(x, \phi^{-1}(x)) < 0\}.$$

Inspecting the last line in (4.9) it is seen that in order to prove Lemma 4.13 it suffices to show that  $A_{\phi^*}$  and  $A_{\phi_*}$  have at most countably many elements. In showing this I use the following observation.

**Claim.** *Let  $\phi$  be a monotone selection from  $\Psi$  and let  $x \in A_\phi$ . Then there exists an open neighborhood  $U$  of  $\theta_x := \phi^{-1}(x)$  such that  $x \in \Psi(\theta)$  for all  $\theta \in U$ .*

*Proof of the Claim.* Fix a monotone selection  $\phi$  and  $x \in A_\phi$ . Since  $\phi$  is monotone, we have  $\lim_{\theta \uparrow \theta_x} \phi(\theta) \leq x \leq \lim_{\theta \downarrow \theta_x} \phi(\theta)$ . Upper hemi-continuity of  $\Psi$  then implies  $\lim_{\theta \uparrow \theta_x} \phi(\theta) \in \Psi(\theta_x)$  and  $\lim_{\theta \downarrow \theta_x} \phi(\theta) \in \Psi(\theta_x)$ . But  $\bar{J}(\cdot, \theta_x)$  is concave by Lemma 4.11 and hence  $\Psi(\theta_x)$  is convex. Thus  $x \in \Psi(\theta_x)$ . Since  $x$  is interior and  $\bar{J}(\cdot, \theta_x)$  is continuously differentiable in  $x$  by Lemma 4.10, we have the first order condition

$$\bar{J}_1(x, \theta_x) = g(x, \theta_x) = 0.$$

As  $G(x, \cdot)$  and  $H(x, \cdot)$  are continuous,  $G(x, \theta_x) - H(x, \theta_x) < 0$  implies that there is an open neighborhood  $U$  of  $\theta_x$  such that  $G(x, \cdot) - H(x, \cdot) < 0$  on  $U$ , and hence  $g(x, \cdot)$  is constant on  $U$ . Combined with the above first order condition this gives

$$\bar{J}_1(x, \theta) = g(x, \theta) = g(x, \theta_x) = 0 \quad \text{for all } \theta \in U.$$

Since  $\bar{J}$  is concave in  $x$ , this implies that  $x \in \Psi(\theta)$  for all  $\theta \in U$ .  $\square$

*Proof of Lemma 4.13.* Let  $x \in A_{\phi_*}$ . By the above Claim, there exists an open neighborhood  $U$  of  $\theta_x := \phi_*^{-1}(x)$  such that  $x \in \Psi(\theta)$  for all  $\theta \in U$ . Thus  $\phi_*(\theta) = \min \Psi(\theta) \leq x$  for all  $\theta \in U$ . But by definition of the inverse,  $\phi_*(\theta) \geq x$  for all  $\theta > \theta_x$ . Thus  $\phi_*(\theta) = x$  for all  $\theta \in U$  such that  $\theta > \theta_x$ . But then  $x$  is a point of discontinuity of  $\phi_*^{-1}$ . Since  $\phi_*^{-1}$  is monotone, there can be at most countably many such points. Hence  $A_{\phi_*}$  has at most countably many elements.

Let  $y \in A_{\phi^*}$ . Again there is an open neighborhood  $V$  of  $\theta_y$  such that  $y \in \Psi(\theta)$  for all  $\theta \in V$ . Thus  $\phi^*(\theta) = \max \Psi(\theta) \geq y$  for all  $\theta \in V$ . But the definition of the inverse implies  $\phi^*(\theta) < y$  for all  $\theta < \theta_y$ , a contradiction. Thus  $A_{\phi^*}$  is empty.  $\square$

We are now ready to complete the proofs of the theorems.

*Proof of Theorem 4.4.* Since  $\mathcal{M} \subset \Gamma$ , we have

$$\begin{aligned} \sup_{\phi \in \Gamma} \left\{ \int_0^1 \bar{J}(\phi(\theta), \theta) d\theta \right\} &\geq \sup_{\phi \in \mathcal{M}} \left\{ \int_0^1 \bar{J}(\phi(\theta), \theta) d\theta \right\} \\ &\geq \sup_{\phi \in \mathcal{M}} \left\{ \int_0^1 \bar{J}(\phi(\theta), \theta) d\theta + \Delta(\phi) \right\} \\ &= \sup_{\phi \in \mathcal{M}} \left\{ \int_0^1 J(\phi(\theta), \theta) d\theta \right\}, \end{aligned}$$

where the second line is by Lemma 4.12 and the last by definition of  $\Delta(\phi)$ .

In the other direction, if  $J$  is weakly concave in  $x$ , we have

$$\begin{aligned} \sup_{\phi \in \mathcal{M}} \left\{ \int_0^1 J(\phi(\theta), \theta) d\theta \right\} &\geq \int_0^1 J(\phi^*(\theta), \theta) d\theta \\ &= \int_0^1 \bar{J}(\phi^*(\theta), \theta) d\theta = \sup_{\phi \in \Gamma} \left\{ \int_0^1 \bar{J}(\phi(\theta), \theta) d\theta \right\}, \end{aligned}$$

where the second line is by Lemma 4.13 and the definition of  $\phi^*$ .  $\square$

*Proof of Theorem 4.5.* By inspection of the proof of Theorem 4.4,  $\phi^*$  attains the supremum in (P). The same argument shows that  $\phi_*$  attains the supremum in (P).

Assume then that  $\phi \in \mathcal{M}$  attains the supremum in (P). Since  $\Delta(\phi) \leq 0$  by Lemma 4.12, we have

$$\int_0^1 \bar{J}(\phi(\theta), \theta) d\theta \geq \int_0^1 J(\phi(\theta), \theta) d\theta = \sup_{\gamma \in \mathcal{M}} \left\{ \int_0^1 J(\gamma(\theta), \theta) d\theta \right\}.$$

Thus  $\phi(\theta) \in \Psi(\theta)$  a.e. by Theorem 4.4. The inequalities follow by (4.7).  $\square$

#### 4.4 An example

I conclude by presenting an example of a concave (but non-separable) problem where the optimal allocation rule is discontinuous and hence it can not be solved using the standard optimal control approach. Nevertheless, the virtual surplus is weakly concave in the allocation and the technique developed in this paper applies.

Consider a model of monopolistic price discrimination with interdependent values as in Example 4.1. The buyer's valuation is given by  $u(x, \theta) = x(\theta^2 + \frac{1}{6})$ , and the seller's cost is given by

$$c(x, \theta) = \begin{cases} 0 & \text{if } x < 1, \\ \frac{1}{2}(\theta + 1)(x - 1)^2 & \text{if } x \geq 1. \end{cases}$$

That is, the seller's marginal cost is constant (and normalized to zero) for the first unit, and it is increasing thereafter. The type  $\theta$  is distributed uniformly on the unit interval. The set of feasible allocations is  $X = \mathbb{R}_+$ . Virtual surplus takes the form

$$\begin{aligned} J(x, \theta) &= \begin{cases} x(\theta^2 + \frac{1}{6}) - (1 - \theta)2x\theta & \text{if } x < 1, \\ x(\theta^2 + \frac{1}{6}) - (1 - \theta)2x\theta - \frac{1}{2}(\theta + 1)(x - 1)^2 & \text{if } x \geq 1 \end{cases} \\ &= (3\theta^2 - 2\theta + \frac{1}{6})x - \frac{1}{2}(\theta + 1)(x - 1)^2 \mathbf{1}_{\{x \geq 1\}}. \end{aligned}$$

This specification captures a situation where higher types of the buyer have a higher marginal willingness to pay for the good or service, but at the same time are more expensive to serve than the lower types. One can think of  $x$  as the quality of a restaurant meal or of some other service, where the customers who are willing to pay the most for given quality are also the most demanding and require the most attention from the staff. Because of these countervailing effects, ironing is needed: Maximizing virtual surplus pointwise suggests the allocation rule

$$\hat{\phi}(\theta) = \begin{cases} \frac{3\theta^2 - \theta + \frac{7}{6}}{\theta + 1} & \text{if } \theta \in [0, \frac{2 - \sqrt{2}}{6}] \cup [\frac{2 + \sqrt{2}}{6}, 1], \\ 0 & \text{if } \theta \in (\frac{2 - \sqrt{2}}{6}, \frac{2 + \sqrt{2}}{6}). \end{cases}$$

In a sense then, the middle types are the worst. The highest types, while being the most expensive to serve, have a high enough marginal willingness to pay to be attractive to the seller. Similarly, the lowest types are attractive despite their low willingness to pay since serving them is inexpensive. Note that because of the initially constant marginal cost,  $\hat{\phi}$  is discontinuous.

In order to apply Theorems 4.4 and 4.5, we use the definitions in (4.1)–(4.5) to construct the generalized virtual surplus. We have

$$h(x, \theta) = \begin{cases} 3\theta^2 - 2\theta + \frac{1}{6} & \text{if } x < 1, \\ 3\theta^2 - (x + 1)\theta - x + \frac{7}{6} & \text{if } x \geq 1, \end{cases}$$

and

$$H(x, \theta) = \begin{cases} \theta^3 - \theta^2 + \frac{1}{6}\theta & \text{if } x < 1, \\ \theta^3 - \frac{x+1}{2}\theta^2 - (x - \frac{7}{6})\theta & \text{if } x \geq 1. \end{cases}$$

Note that, for  $x$  small enough,  $H(x, \cdot)$  is initially concave and then convex. This makes convexifying simple as it suffices to “iron out” the initial concave part. Straightforward calculations give the convex hulls

$$G(x, \theta) = \begin{cases} -\frac{1}{12}\theta & \text{if } \theta < \frac{1}{2}, \\ \theta^3 - \theta^2 + \frac{1}{6}\theta & \text{if } \theta \geq \frac{1}{2}, \end{cases} \quad \text{for all } x < 1,$$

and

$$G(x, \theta) = \begin{cases} (-\frac{x^2}{16} - \frac{9}{8}x + \frac{53}{48})\theta & \text{if } \theta < \frac{x+1}{4}, \\ \theta^3 - \frac{x+1}{2}\theta^2 - (x - \frac{7}{6})\theta & \text{if } \theta \geq \frac{x+1}{4}, \end{cases} \quad \text{for all } x \geq 1.$$

Thus we have

$$g(x, \theta) = \begin{cases} -\frac{1}{12} & \text{if } \theta < \frac{1}{2}, \\ 3\theta^2 - 2\theta + \frac{1}{6} & \text{if } \theta \geq \frac{1}{2}, \end{cases} \quad \text{for all } x < 1, \quad (4.10)$$

and

$$g(x, \theta) = \begin{cases} -\frac{x^2}{16} - \frac{9}{8}x + \frac{53}{48} & \text{if } \theta < \frac{x+1}{4}, \\ 3\theta^2 - (x+1)\theta - x + \frac{7}{6} & \text{if } \theta \geq \frac{x+1}{4}. \end{cases} \quad \text{for all } x \geq 1. \quad (4.11)$$

Recall from Remark 4.8 that the maximal optimal allocation rule is given by

$$\phi^*(\theta) = \max \{x \in X \mid g(x, \theta) \geq 0\},$$

where  $\max \emptyset = 0$ . Since  $g$  is weakly decreasing  $x$ , we immediately see from (4.10) that  $\phi^*(\theta) = 0$  for all  $\theta < \frac{2-\sqrt{2}}{6}$ . We may then note that the first case in (4.11) is negative. Hence the optimal allocation for  $\theta \geq \frac{2-\sqrt{2}}{6}$  can be solved for from the second case in (4.11). Thus we have

$$\phi^*(\theta) = \begin{cases} 0 & \text{if } \theta \in [0, \frac{2+\sqrt{2}}{6}), \\ \frac{3\theta^2 - \theta + \frac{7}{6}}{\theta + 1} & \text{if } \theta \in [\frac{2+\sqrt{2}}{6}, 1]. \end{cases}$$

That is, the lowest types are excluded, then there is a jump from 0 to 1 at  $\frac{2+\sqrt{2}}{6}$  after which the solution coincides with pointwise maximization of standard virtual surplus.

## Appendix: Proofs of the technical lemmas

*Proof of Lemma 4.9.* The claim is established by showing successively that all the functions defined in (4.1)–(4.4) are continuous (in  $x$ ).

By definition  $h = J_1$  is continuous, since  $J$  is assumed once continuously differentiable on an open set containing  $X \times \Theta$ .

For continuity of  $H$ , fix  $(x, \theta) \in X \times \Theta$ . For any sequence converging to  $(x, \theta)$ ,

$$\lim_{n \rightarrow \infty} H(x_n, \theta_n) = \lim_{n \rightarrow \infty} \int_0^1 h(x_n, r) \mathbb{1}_{\{r \leq \theta_n\}} dr = \int_0^1 h(x, r) \mathbb{1}_{\{r \leq \theta\}} dr = H(x, \theta),$$

where the second equality follows by Lebesgue's dominated convergence theorem since  $|h(x_n, r) \mathbb{1}_{\{r \leq \theta_n\}}| \leq \sup_{(x, \theta) \in X \times \Theta} |h(x, \theta)| < \infty$  for all  $(x_n, \theta_n, r) \in X \times \Theta^2$  by continuity of  $h$  and compactness of its domain, and since for a.e.  $r \in \Theta$ , we have  $\lim_{n \rightarrow \infty} h(x_n, r) \mathbb{1}_{\{r \leq \theta_n\}} = h(x, r) \mathbb{1}_{\{r \leq \theta\}}$  by continuity of  $h$ . Thus  $H$  is continuous.

I then show that  $G$  is continuous in  $x$ . By definition,

$$\begin{aligned} -G(x, \theta) &= \max\{-\lambda H(x, \theta_1) - (1 - \lambda)H(x, \theta_2) : \\ &\quad (\lambda, \theta_1, \theta_2) \in [0, 1]^3 \text{ and } \lambda\theta_1 + (1 - \lambda)\theta_2 = \theta\}. \end{aligned}$$

Holding  $\theta \in \Theta$  fixed, the objective function in the above maximization problem is continuous in  $(\lambda, \theta_1, \theta_2, x)$  and the feasible set is compact and independent of  $x$ . Thus  $G$  is continuous in  $x$  by the Maximum theorem.

Consider then  $g$ . It is convenient to extend  $G$  to  $X \times D$ , where  $D \subset \mathbb{R}$  is an open set containing  $[0, 1] = \Theta$ . Since  $G(\cdot, \theta)$  is continuous for all  $\theta \in \Theta$ , and  $G(x, \cdot)$  is continuously differentiable on  $(0, 1)$  and convex on  $[0, 1]$  for all  $x \in X$ , the extension can be chosen such that  $G(\cdot, \theta)$  is continuous for all  $\theta \in D$ , and  $G(x, \cdot)$  is continuously differentiable and convex on  $D$  for all  $x$ .<sup>14</sup> We then have  $g(x, \theta) = G_2(x, \theta)$  for all  $(x, \theta) \in X \times \Theta$ .

Fix  $\theta \in [0, 1]$ . Let  $x \in X$  and let  $(x_n)$  be a sequence in  $X$  such that  $x_n \rightarrow x$ . We have

$$\begin{aligned} \limsup_{x_n \rightarrow x} g(x_n, \theta) &= \limsup_{x_n \rightarrow x} \lim_{t \downarrow 0} \frac{G(x_n, \theta + t) - G(x_n, \theta)}{t} \\ &= \limsup_{x_n \rightarrow x} \inf_{t > 0} \frac{G(x_n, \theta + t) - G(x_n, \theta)}{t} \\ &\leq \limsup_{x_n \rightarrow x} \frac{G(x_n, \theta + s) - G(x_n, \theta)}{s} \quad \forall s > 0 : \theta + s \in D \\ &= \frac{G(x, \theta + s) - G(x, \theta)}{s} \quad \forall s > 0 : \theta + s \in D, \end{aligned}$$

where the first equality is by definition of  $g$ , the second by convexity of  $G$  in  $\theta$ , and the last by continuity of  $G$  in  $x$ . Therefore,

$$\limsup_{x_n \rightarrow x} g(x_n, \theta) \leq \lim_{s \downarrow 0} \frac{G(x, \theta + s) - G(x, \theta)}{s} = g(x, \theta).$$

<sup>14</sup>For example, for all  $x \in X$ , let  $G(x, \cdot)$  be affine on  $D \setminus [0, 1]$  with  $G_2(x, \theta) = \lim_{\tau \downarrow 0} G_2(x, \tau)$  for all  $\theta < 0$  and  $G_2(x, \theta) = \lim_{\tau \uparrow 0} G_2(x, \tau)$  for all  $\theta > 1$ , where the affine parts are chosen such that  $G(x, \cdot)$  is continuous on  $D$ . By construction,  $G(x, \cdot)$  so extended is convex and continuously differentiable. Continuity of  $G(\cdot, \theta)$  for all  $\theta \in D \setminus [0, 1]$  follows from the continuity of  $G(\cdot, 0)$  and  $G(\cdot, 1)$ , since, e.g.,  $|g(x, \theta) - g(y, \theta)| = |g(x, 0) - g(y, 0)|$  for all  $\theta < 0$  and all  $(x, y) \in X^2$ .



An analogous argument gives

$$\liminf_{x_n \rightarrow x} g(x_n, \theta) \geq \frac{G(x, \theta) - G(x, \theta - s)}{s} \quad \forall s > 0 : \theta - s \in D,$$

which implies

$$\liminf_{x_n \rightarrow x} g(x_n, \theta) \geq \lim_{s \downarrow 0} \frac{G(x, \theta) - G(x, \theta - s)}{s} = g(x, \theta).$$

Thus  $\lim_{x_n \rightarrow x} g(x_n, \theta) = g(x, \theta)$  so that  $g(\cdot, \theta)$  is continuous at  $x$ . Since  $x$  and  $\theta$  were arbitrary, the claim follows.  $\square$

*Proof of Lemma 4.10.* For any fixed  $\theta$ , the function  $\bar{J}$  is an integral function in  $x$ , and the integrand  $g(\cdot, \theta)$  is continuous by Lemma 4.9. Hence  $\bar{J}$  is continuously differentiable in  $x$  for all  $\theta \in \Theta$ .

For continuity, fix  $(x, \theta) \in X \times \Theta$  and consider a sequence in  $X \times \Theta$  such that  $(x_n, \theta_n) \rightarrow (x, \theta)$ . Note that  $g$  is nondecreasing and continuous in  $\theta$  by construction, and it is continuous in  $x$  by Lemma 4.9. Hence for all  $(s, \theta_n) \in X \times \Theta$ ,

$$-\infty < \inf_{y \in X} g(y, 0) \leq g(s, \theta_n) \leq \sup_{z \in X} g(z, 1) < \infty.$$

We thus have

$$\begin{aligned} \lim_{n \rightarrow \infty} \bar{J}(x_n, \theta_n) &= \lim_{n \rightarrow \infty} J(0, \theta_n) + \lim_{n \rightarrow \infty} \int_0^{\bar{x}} g(s, \theta_n) \mathbb{1}_{\{s \leq x_n\}} ds \\ &= J(0, \theta) + \int_0^{\bar{x}} g(s, \theta) \mathbb{1}_{\{s \leq x\}} ds = \bar{J}(x, \theta), \end{aligned}$$

where the second equality is by continuity of  $J$  and the Lebesgue's dominated convergence theorem as  $|g(s, \theta_n) \mathbb{1}_{\{s \leq x_n\}}| \leq \max \{|\inf_{y \in X} g(y, 0)|, |\sup_{z \in X} g(z, 1)|\}$  for all  $(s, x_n, \theta_n) \in X^2 \times \Theta$ , and since for a.e.  $s \in X$ ,  $\lim_{n \rightarrow \infty} g(s, \theta_n) \mathbb{1}_{\{s \leq x_n\}} = g(s, \theta) \mathbb{1}_{\{s \leq x\}}$  by continuity of  $g(s, \cdot)$ . Thus  $\bar{J}$  is continuous.

To show increasing differences in  $(x, \theta)$ , take any  $(x, x') \in X^2$  and  $(\theta, \theta') \in \Theta^2$  such that  $x' \geq x$  and  $\theta' \geq \theta$ . Then

$$\bar{J}(x', \theta') - \bar{J}(x, \theta') = \int_x^{x'} g(s, \theta') ds \geq \int_x^{x'} g(s, \theta) ds = \bar{J}(x', \theta) - \bar{J}(x, \theta),$$

where the inequality follows from  $g$  being nondecreasing in  $\theta$ .  $\square$

*Proof of Lemma 4.11.* The generalized virtual surplus  $\bar{J}$  is differentiable in  $x$  by Lemma 4.10. Recalling the definition from (4.5) we have  $\bar{J}_1(x, \theta) = g(x, \theta)$ . So it suffices to show that  $g$  is nonincreasing in  $x$  for any fixed  $\theta \in \Theta$ .

To this end, note first that  $h(x, \theta) = J_1(x, \theta)$  is nonincreasing in  $x$  for any  $\theta \in \Theta$  by the weak concavity of  $J$  in  $x$ . By (4.2) we have for all  $x' > x$  and  $\theta' > \theta$ ,

$$H(x', \theta') - H(x', \theta) = \int_\theta^{\theta'} h(x', r) dr \leq \int_\theta^{\theta'} h(x, r) dr = H(x, \theta') - H(x, \theta).$$

That is,  $H$  has increasing differences in  $(-x, \theta)$ .

Suppose then to the contrary of  $g$  being nonincreasing in  $x$  that there exists  $(x, x', \theta_o) \in X^2 \times \Theta$  such that  $x' > x$  and  $g(x', \theta_o) > g(x, \theta_o)$ . Since  $g(x', \cdot)$  and  $g(x, \cdot)$  are continuous, it is without loss to assume that  $\theta_o$  is interior. Let  $L : [0, 1] \rightarrow \mathbb{R}$  be the unique affine function tangent to  $G(x, \cdot)$  at  $\theta_o$ . Similarly, let  $L' : [0, 1] \rightarrow \mathbb{R}$  be the affine function tangent to  $G(x', \cdot)$  at  $\theta_o$ . Denote  $\rho := G(x, \theta_o) - G(x', \theta_o)$ .

**Claim 1.**  $\exists \theta \leq \theta_o : H(x, \theta) - H(x', \theta) \geq \rho$ .

*Proof of Claim 1.* By definition  $G(x, \cdot)$  is convex and lies everywhere below  $H(x, \cdot)$  so that

$$H(x, \theta) \geq G(x, \theta) \geq L(\theta) \quad \text{for all } \theta \in \Theta.$$

If  $H(x', \theta_o) = G(x', \theta_o)$ , we are done since then the first of the above inequalities (evaluated at  $\theta_o$ ) implies  $H(x, \theta_o) - H(x', \theta_o) \geq G(x, \theta_o) - G(x', \theta_o) = \rho$ . Otherwise  $H(x', \theta_o) > G(x', \theta_o)$ . But then  $G(x', \cdot)$  is affine and coincides with  $L'$  in a neighborhood of  $\theta_o$ . Moreover, the point  $(\theta_o, G(x', \theta_o))$  is a convex combination of two points from the graph of  $H(x', \cdot)$  so that there exists  $\theta_1 < \theta_o$  such that  $H(x', \theta_1) = L'(\theta_1)$ . By construction  $\frac{d}{d\theta}(L(\theta) - L'(\theta))|_{\theta=\theta_o} = g(x, \theta_o) - g(x', \theta_o) < 0$ . This together with the second of the above inequalities implies  $H(x, \theta_1) - H(x', \theta_1) \geq L(\theta_1) - L'(\theta_1) > L(\theta_o) - L'(\theta_o) = G(x, \theta_o) - G(x', \theta_o) = \rho$ .  $\square$

**Claim 2.**  $\exists \theta > \theta_o : H(x, \theta) - H(x', \theta) < \rho$ .

*Proof of Claim 2.* Assume first that  $H(x, \theta_o) = G(x, \theta_o)$ . The difference  $H(x, \theta) - L'(\theta)$  is a continuously differentiable function of  $\theta$  with  $H(x, \theta_o) - L'(\theta_o) = G(x, \theta_o) - G(x', \theta_o) = \rho$ . Moreover,  $\frac{d}{d\theta}(H(x, \theta) - L'(\theta))|_{\theta=\theta_o} = g(x, \theta_o) - g(x', \theta_o) < 0$  so that there exists  $\theta_2 > \theta_o$  such that  $\rho > H(x, \theta_2) - L'(\theta_2) \geq H(x, \theta_2) - H(x', \theta_2)$  as we wanted to show.

Assume then that  $H(x, \theta_o) > G(x, \theta_o)$ . Then  $G(x, \cdot)$  coincides with  $L$  in a neighborhood of  $\theta_o$  and there exists  $\theta_3 > \theta_o$  such that  $H(x, \theta_3) = L(\theta_3)$ . But then  $H(x, \theta_3) - H(x', \theta_3) \leq L(\theta_3) - L'(\theta_3) < L(\theta_o) - L'(\theta_o) = \rho$ .  $\square$

We have arrived at the desired contradiction since Claims 1 and 2 are incompatible with  $H$  having increasing differences in  $(-x, \theta)$ . Hence  $g$  is nonincreasing in  $x$  implying that  $\bar{J}$  is weakly concave in  $x$ .  $\square$

## References

- [1] C. Berge, Topological Spaces, Oliver & Boyd, London, UK, 1963.
- [2] F. Clarke, The maximum principle under minimal hypothesis, SIAM Journal of Control and Optimization 14 (1976), 1078–1091.
- [3] F. Clarke, A general theorem on necessary conditions in optimal control, Discrete and Continuous Dynamical Systems 29 (2011), 485–503.

- [4] D. Fudenberg, J. Tirole, *Game Theory*, The MIT Press, Cambridge, MA, 1991.
- [5] M. F. Hellwig, A maximum principle for control problems with monotonicity constraints, Working Paper, Max Planck Institute, 2009.
- [6] B. Jullien, Participation constraints in adverse selection models, *J. Econ. Theory* 93 (2000), 1–47.
- [7] J. Laffont, J. Tirole, Using Cost Observation to Regulate Firms, *J. Polit. Economy* 94 (1986), 614–641.
- [8] D. Léonard, N. van Long, *Optimal Control Theory and Static Optimization in Economics*, Cambridge Univ. Press, Cambridge, UK, 1992.
- [9] M. Mussa, S. Rosen, Monopoly and product quality, *J. Econ. Theory* 18 (1978), 301–317.
- [10] R. Myerson, Optimal auction design, *Math. Oper. Res.* 6 (1981), 58–73.
- [11] G. Nöldeke, L. Samuelson, Optimal bunching without optimal control, *J. Econ. Theory* 134 (2007), 405–420.
- [12] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.
- [13] B. Salanié, *The Economics of Contracts: A Primer*, The MIT Press, Cambridge, MA, 1997.
- [14] D. M. Topkis, Minimizing a submodular function on a lattice, *Operations Research* 26 (1978), 305–321.
- [15] R. Vinter, *Optimal Control*, Birkhäuser, Boston, MA, 2010.