Working paper

PARTICIPATION CONSTRAINTS IN DISCONTINUOUS ADVERSE SELECTION MODELS[1]

October 13, 2017

David Martimort and Lars Stole

Abstract. We present a set of necessary and sufficient conditions for a class of optimal control problems with pure state constraints for which the objective function is linear in the state variable but the objective function is otherwise only restricted to be upper semi-continuous in the control variable. We apply those conditions to a number of economic environments in contract theory where discontinuities in objectives prevail. Examples include nonlinear pricing of digital goods and common agency models of lobbying where interest groups compete with contracts.

Keywords. Optimal control, non-smooth optimization, convex analysis, type-dependent participation constraints, principal-agent models.

1. INTRODUCTION

The textbook treatment of optimal screening contracts typically takes the agent's outside option as a fixed constant, independent of type.[1] More complex settings which allow for competition by rival principals, non-trivial ownership rights on productive assets and type-dependent fixed costs require a departure from this restrictive assumption. Lewis and Sappington (1989) initiated the seminal study of screening contracts in this more general setting by constructing the solution to a class of optimal control problems with type-dependent participation constraints. This class of problems was further enriched by Maggi and Rodriguez (1995), among others, with the most general statement of the problem and its solution culminating in the analysis offered by Jullien (2000).

These techniques have allowed modelers to apply the optimal contracting paradigm to more general economic contexts, unveiling new features of optimal contracts. Applications have spread through many fields including the design of nonlinear prices under the threat of bypass (Curien, Jullien and Rey (1998)), competition in nonlinear prices under various environments (Martimort and Stole (2009), Stole (1995, 2003), Calzolari and DeNicolo (2013)), trade policy in open economies (Brainard and Martimort (1996)), regulation of privately-owned monopolies (Auriol and Picard (2011)), and optimal contracting under liability constraints (Ollier and Thomas (2013)) or in some dynamic environments (Deb and Said (2015)), to name some interesting examples.

Unfortunately, the existing techniques also have their own limits. In particular, the need for tractability has led authors to restrict their analysis to economic environments which

[1]We are especially thankful to John Birge for many helpful discussions. A less general version of the main theorem in this paper appears in an earlier, unpublished note, *"Necessary and sufficient conditions for non-smooth linear-state optimal control problems"* (2009). The present paper provides the more general result along with a geometric intuition for its proof and several relevant applications. The usual disclaimer applies.

[a]Paris School of Economics-EHESS, david.martimort@parisschoolofeconomics.eu.

[b]University of Chicago Booth School of Business, lars.stole@chicagobooth.edu.

[1]See Laffont and Martimort (2002, Chapter 3) for instance.

are sufficiently *smooth.* In many circumstances, such as when firms face non-trivial fixed costs or sunk investments, the environment is inherently discontinuous. In other settings, such as when principals compete against one another using positive payment schedules, equilibria may emerge which exhibit discontinuities in each player's payoff function. In both cases, we are left uncertain about the consequences of such discontinuities for optimal contracts and the generality of results when environments (and equilibria) are not assumed to be smooth *a priori.* Important economic insights may have gone unnoticed because of our restricted attention. Developing the required techniques for *non-smooth* environments and showing how they apply in practice are the purposes of this paper.

More precisely, our goal is twofold. First, we present a set of necessary and sufficient conditions for a class of optimal control problems with pure state constraints for which the objective function is linear in the state variable but the objective function may exhibit kinks and discontinuities in the control variable. Second, we apply these techniques to (what we believe are) quite natural contracting environments where existing techniques are inadequate.

Our first application deals with nonlinear pricing of a digital good under the threat of competition by a low-quality fringe. Because the provider of a high-quality good needs to build capacity for extra services, the cost function is discontinuous. We show that, under the threat of competition, a dominant firm may choose to leave additional rents, even for some buyers with intermediate valuations for the service, in order to avoid incurring the extra capacity cost. The second and third examples deal with two variations of a lobbying game in which interest groups offer nonlinear (and possibly discontinuous) contribution schedules to influence a central decision-maker. In both variations, interest groups may offer discontinuous contribution schedules which render objective functions discontinuous in the decision-maker's action. In such cases, our general control theorem provides a full characterization of best responses. In the first scenario, we assume that each interest group can enforce an exclusivity clause and only pay the decision-maker when he chooses decisions that favorable to the group. We show that the exclusive equilibrium in this game splits the type space between opposite groups and equilibrium payments to the decision-maker exhibit an upward jump. In the second scenario we assume that exclusivity clauses are not allowed and show that the previous split outcome is no longer part of an equilibrium outcome.

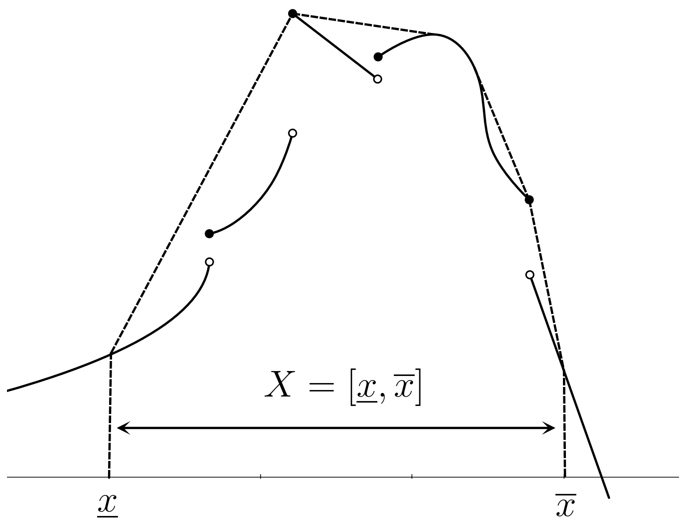
Section 2 gives a brief overview of optimization techniques in non-smooth environments, providing the key intuition for our main theorem. Section 3 presents our main result: A theorem that characterizes solutions to optimal control problems in which the objective function is only restricted to be upper semi-continuous. This theorem builds on earlier work by Vinter and Zheng (2000), but it refines its application to the case of quasi-linear objectives which is prevalent in contract theory. While Vinter and Zheng (2000) focuses on necessary conditions for optimality, we prove that these conditions are also sufficient in our context. We also discuss there to which extent this theorem extends the existing literature and especially the work by Jullien (2000) under much weaker conditions. Section 4 develops our applications. Proofs are relegated to the Appendix.

2. NON-SMOOTH OPTIMIZATION

Before specializing our inquiry, it is useful to explore necessary and sufficient conditions for the general problem of maximizing an upper semi-continuous function, $h : \mathbb{R} \to \mathbb{R}$, over a compact set $X \subset \mathbb{R}$. These conditions provide the intuition for our analysis when we develop our applications in Section 4.

A generalization of the first-order condition for smooth, concave optimization programs can be obtained for general upper semi-continuous programs by introducing a few concepts from non-smooth convex analysis. The basic idea is that any solution to the original upper semi-continuous program must lie on the minimal concave envelope or *concavification* of the objective. Consider, for example, the upper semi-continuous function graphed in Figure 1 in bold.

Figure 1:



This function is defined over the real line, but the restricted domain of interest is $X = [\underline{x}, \overline{x}]$. The minimal-concave envelope over this domain is depicted by the dashed lines in the graph. Notice its value is negative infinity outside of $[\underline{x}, \overline{x}]$. Obviously, the maximum of *this* concave envelope is a solution to the original program. More generally, in the case in which there is a continuum of solutions (i.e., the maximum is achieved on a horizontal component of the majorization), there exist two solutions to the original program – the endpoints of the majorization. It is, in this sense, without loss to convert an upper semi-continuous program over a compact set into a concave (but possibly non differentiable) program over the same set. Formally, we will denote $\overline{co}_X(h)$ to refer to the concavification of an objective function, h, over a domain, X, and $\overline{co}_X(h)(x)$ to refer to the value of this envelope evaluated at x.[2]

Having reached the conclusion that we may focus on the concave envelope of the program, we can now import the generalized notion of derivative from convex analysis. For-

[2]When $X = \mathbb{R}$, we simplify notation and use $\overline{co}(h)(x)$ in place of $\overline{co}_{\mathbb{R}}(h)(x)$. In the non-smooth optimization literature, often one considers the minimal concave envelope of h over the real line instead of some domain X, but in this case with a penalty function, $\Psi_X(x)$ which equals 0 for $x \in X$ and $-\infty$ for $x \notin X$. Thus, in our notation, $\overline{co}_X(h) = \overline{co}(h + \Psi_X)$.

mally, we will define a set of gradients at any point to be all those vectors which "support" the graph at the given point, and we refer to this set-valued notion of derivative as the generalized gradient or the *super-differential*, denoted $\partial h(x)$ when applied to a concave function h at point x.[3] Where h is differentiable, the super-differential is single-valued and corresponds to the gradient. If h exhibits a kink and $X \subseteq \mathbb{R}$, the super-differential is an interval of gradients with endpoints corresponding to the left- and right-side derivatives at the point. More generally, if $X \subseteq \mathbb{R}^n$, then

$$\partial h(x) = \{\tau \in \mathbb{R}^n \mid h(y) \leq h(x) + \langle \tau, y - x \rangle \ \forall\, y \in \mathbb{R}^n\}.$$

Using this generalization of gradient, we can now state the necessary and sufficient conditions for x^* to be a maximum of an upper semi-continuous function, h, over some given domain $x \in X$: if x^* is a solution to the maximization program, then the following first-order condition must be satisfied:

$$0 \in \partial \overline{co}_X(h)(x^*). \tag{2.1}$$

Furthermore, if x^* satisfies (2.1) and the envelope coincides with h at x^*, i.e.,

$$\overline{co}_X(h)(x^*) = h(x^*) \tag{2.2}$$

then x^* solves the maximization program.

These conditions can be further tightened when a component of the objective function is affine. To this end, suppose that $h = g + f$ where g is affine (and slightly abusing notations, let us write $g(x) = gx$). Well-known identities from convex analysis give us:

$$\overline{co}_X(h)(x) = gx + \overline{co}_X(f)(x) \text{ and } \partial \overline{co}_X(h)(x) = g + \partial \overline{co}_X(f)(x).$$

Thus, the linear linear part of the objective can be factored out and the "first-order" necessary and sufficient condition for the optimality of x^* reduces to

$$-g \in \partial \overline{co}_X(f)(x^*).$$

This property will be repeatedly used throughout our analysis, first, to derive generalized first-order conditions for our infinite-dimensional optimal control problem and, second, to tackle applications in contract theory where such a decomposition is frequently available.

The simple example above is meant to illustrate how tools from non-smooth convex analysis can be used to convert a pointwise optimization problem with an upper-semi-continuous objective into a corresponding concave (albeit, non-smooth) optimization

[3] In our context of maximizing a concave function, it is perhaps more accurate to say that the graph of a concave function "supports" its gradients. Nonetheless, we use the term "support" from convex analysis given it is evocative and familiar. The term *sub-differential* is the parallel notion of super-differential when applied to convex functions. When we refer to the generalized gradient of a function that is understood to be convex, we will abuse notation slightly by again using the notation $\partial h(x)$, where it is understood that when h is convex, then

$$\partial h(x) = \{\tau \in \mathbb{R}^n \mid h(y) \geq h(x) + \langle \tau, y - x \rangle \ \forall\, y \in \mathbb{R}^n\}.$$

See Ferrera (2014) for an introduction to non-smooth analysis and an in-depth discussion of super- and sub-differentials.

problem that allows for the use of generalized first-order conditions. Furthermore, these first-order conditions can be further reduced to gradient conditions on the majorization of the nonlinear component of the objective function by subtracting away any linear components. A similar intuition lies beneath the mechanics of the proof to the main theorem in this paper and behind the analysis of the applications developed in Section 4.

Unfortunately, the general problems we are considering are more complex. First, we deal with principal-agent models which are essentially infinite-dimensional, optimal control problems; we thus want to maximize a functional that depends linearly on the information rent left to the privately-informed agent while this rent itself evolves according to incentive compatibility conditions. Finding the solutions for such problem *a priori* relies on optimal control techniques. Therefore, the non-smooth techniques above are only useful if one can find the solutions of such optimization problems as pointwise optima. This is where the assumption of linearity of the objective function takes its full strength. Linearity allows such transformation; a point which is already well-known from principal-agent screening models in quasi-linear and smooth environments where this familiar step is achieved through a simple integration by parts.[4] Things are more complicated with a state-dependent reservation utility but the basic intuition remains valid, provided that the type-dependent participation constraints are conveniently "weighted". Second, a major difficulty of such optimization problems is to deal with pure-state constraints – the agent's information's rent having to be above any reservation payoff that he may have got elsewhere. This possibility introduces a shadow cost of that participation constraint whose impact on the optimal solution enters linearly. The force of this linearity is to introduce a dichotomy between incentive and participation concerns, both of which enter linearly in the maximand, and the possible non-smoothness of the nonlinear part of the objective.

3. THE THEOREM

In what follows, we will denote control problems in which the state variable, u, is restricted to be an absolutely continuous function on Θ. We denote the feasible set of such functions by $AC(\Theta, \mathbb{R})$, where $\Theta \equiv [\underline{\theta}, \overline{\theta}]$. As a motivation, in the context of principal-agent models, the state variable is typically the agent's rent as a function of type. Incentive compatibility, together with mild technical conditions on utility, implies that such a function is absolutely continuous.[5] We will also restrict attention to problems in which that state variable must satisfy a non-negativity constraint:

$$u(\theta) \geq 0 \quad \forall \theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]. \tag{3.1}$$

Using again our motivation of principal-agent problems, the nonnegativity constraint (3.1) corresponds to a participation constraint inducing the agent to accept an offer rather than take the outside option, normalized to zero. When the state variable u is both absolutely continuous and nonnegative, we will say it is *admissible*.

We are interested in the following pure-state control program:

$$(\mathcal{P}) : \text{Maximize}_{u \in AC(\Theta,\mathbb{R})} \ \Lambda(u) \equiv \int_{\underline{\theta}}^{\overline{\theta}} (s(\theta, \dot{u}(\theta)) - u(\theta) f(\theta))\, d\theta \text{ subject to (3.1).}$$

[4]Myerson (1981, Lemma 3), Baron and Myerson (1982, Lemma 2) and Laffont and Martimort (2002, Chapter 3) for a textbook treatment. Cases where this familiar trick does not work involve the optimal taxation model of Mirrlees (1971) among other possible examples where quasi-linearity is not assumed.

[5]See Milgrom and Segal (2002), for example.

Regularity. Importantly, we only assume that $s(\theta, \cdot)$ is an upper semi-continuous function, bounded from above, and that $f(\theta)$ is a positive, bounded function giving rise to an absolutely-continuous definite integral $F(\theta) \equiv \int_{\underline{\theta}}^{\theta} f(\theta)d\theta$. Without loss of generality, we normalize f such that $F(1) = 1$. In the principal-agent context, this allows us to interpret F as a continuous probability distribution and f as its associated density. We also make a minimal technical assumption that $s(\cdot, \cdot)$ is $\mathcal{L} \times \mathcal{B}$-measurable, where $\mathcal{L}$ denotes the set of Lebesgue measurable subsets of Θ and $\mathcal{B}$ is the set of Borel measurable subsets of $\mathbb{R}$.

Linearity. We define the Lagrangian for program $(\mathcal{P})$ as $L(\theta, u, q) \equiv s(\theta, q) - uf(\theta)$. We should make clear that the key restriction we have placed on the control problem $(\mathcal{P})$ is that, for any θ and q, the Lagrangian is a linear function of the state variable u. As we will see below in our applications, this linearity echoes the nature of a number of economic problems, especially in contract theory where agents are risk-neutral and payoffs are linear in money. The function $s(\theta, q)$ can there be viewed as a surplus function (up to a normalization) while u is then the share of this surplus that is captured by the agent – his information rent. From a technical viewpoint, the linearity restriction is the primary source of many sharp results in the analysis that follows, including the ability for us to relax the continuity of s, the characterization of the solution through a simple generalized gradient condition, and the fact that necessary conditions for optimality are also sufficient.

We now present our main result for this class of problems.

Theorem 1 *$\overline{u}$ is a solution to program $(\mathcal{P})$ if and only if $\overline{u}$ is admissible and there exists a probability measure μ defined over the Borel subsets of Θ with an associated adjoint function, $\overline{\gamma} : \Theta \to [0, 1]$, defined by $\overline{\gamma}(\underline{\theta}) = 0$ and for $\theta > \underline{\theta}$,*

$$\overline{\gamma}(\theta) \equiv \int_{[\underline{\theta}, \theta)} \mu(d\tilde{\theta}),$$

such that the following conditions are satisfied:

$$\int_{\underline{\theta}}^{\overline{\theta}} \overline{u}(\tilde{\theta})\mu(d\tilde{\theta}) = 0, \tag{3.2}$$

$$0 \in F(\theta) - \overline{\gamma}(\theta) + \partial\overline{co}(s)(\theta, \dot{\overline{u}}(\theta)) \text{ for a.e. } \theta \in \Theta, \tag{3.3}$$

$$\overline{co}(s)(\theta, \dot{\overline{u}}(\theta)) = s(\theta, \dot{\overline{u}}(\theta)) \text{ for a.e. } \theta \in \Theta. \tag{3.4}$$

Furthermore, if

$$v(\theta, \sigma) \equiv \arg\max_{q \in \mathbb{R}} \; s(\theta, v) + (F(\theta) - \sigma)v \tag{3.5}$$

is single-valued and continuous over the domain $(\theta, \sigma) \in \Theta \times [0, 1]$, then the solution $\overline{u}$ to $(\mathcal{P})$ is continuously differentiable.

Interpretations. The conditions in Theorem 1 are very similar to those in Theorem 1 in Jullien (2000). In both theorems, necessary and sufficient conditions are stated in terms of a probability measure which serves to express a "complementary slackness condition" (3.2) and an optimality condition (3.3). Moreover, both theorems use a similar condition to establish the continuity of $\dot{\bar{u}}(\theta)$ in the solution to $(\mathcal{P})$. Yet, the present Theorem requires substantially weaker assumptions on the primitive function s. To import results in Seierstadt and Sydsaeter (1987, Theorems 2 and 3, Chapt. 5), who provide necessary and sufficient conditions satisfied by an optimal path for an optimal control problem with pure state constraints, Jullien (2000) makes the stronger hypothesis that $s(\theta,\cdot)$ is twice continuously differentiable in v, an assumption used in his Lemma 7 (p.32) in tandem with strict concavity of the objective to obtain a smooth version of (3.3). Our contribution is to demonstrate the broader validity of these conditions for problems with upper semi-continuous integrands.

• It is worth noting the importance of the linearity of the objective function in u, and its usefulness in isolating all the consequences of non-smoothness as coming from possible discontinuities in the surplus function $s(\theta, q)$. This allows us to focus on the concave envelope of this function. Importantly, this concave envelope is not necessarily differentiable; indeed, it will fail to be differentiable wherever $s(\theta, q)$ has an upward discontinuity. In multi-principal games, it will also fail to be differentiable whenever the equilibrium tariffs of some principals (which we will see in our applications become embedded in the equivalent of s) are discontinuous.

To illustrate, consider the case where the Lagrangian is reduced to $L(\theta, u, v) \equiv s(\theta, v)$. In the contract theory applications below, this assumption corresponds to an assumption of complete information on the agent's preferences. The optimization problem so constructed can be solved pointwise and any solution is given by the optimality condition:

$$0 \in \partial\overline{co}(s)(\theta, v^*(\theta)) \text{ a.e.} \tag{3.6}$$

The solution to our original program $(\mathcal{P})$ differs from v^* in (3.6) as a result of the addition of the linear term in its maximand. Condition (3.3) indicates how the solution needs to be modified.

• As in Jullien (2000), the measure $\mu(d\theta)$ reflects the shadow cost of the participation constraint (3.1) around θ. The adjoint function $\overline{\gamma}(\theta)$ can thus be interpreted as the sum of these shadow costs for all types less than θ. Replacing the right-hand side of (3.1) uniformly by $\epsilon < 0$ for all $\tilde{\theta} \leq \theta$ would relax the optimization problem and increase its value by $\overline{\gamma}(\theta)$.

Observe that this adjoint function so constructed is right-continuous rather than left-continuous. In Section 4 below, we show that the right-continuous adjoint constructions allow for the possibility that the probability measure μ has mass points where the participation constraint begins binding. $\overline{\gamma}(\theta)$ has upward jumps at such points. This possibility must be considered for programs in which s exhibits a discontinuity.

• The condition that $v(\theta, \sigma)$ is single-valued and continuous is implied by the strict concavity of $s(\theta, \cdot)$. It is also implied by the weaker condition in Jullien (2000, Assumption 2) that $s(\theta, v) - (\sigma - F(\theta))v$ is strictly quasi-concave in v for any $\sigma \in [0, 1]$.

Remarks about Applying this Result: We comment below on various features of the results and offer a guideline for reading our applications in Section 4 in view of this theorem.

• Principal-agent problems with type-dependent participation constraints, as studied in the pathbreaking works of Lewis and Sappington (1989), Maggi and Rodriguez (1995) and Jullien (2000), are often expressed under the form $(\mathcal{P}')$ below so as to make the nature of the agent's outside option, $\hat{U}(\theta)$, and its associated participation constraint more explicit.

$$(\mathcal{P}') : \text{Maximize}_{\{U \in AC(\Theta, \mathbb{R}), q\}} \ \tilde{\Lambda}(U, q) \equiv \int_{\underline{\theta}}^{\overline{\theta}} (\tilde{s}(\theta, q(\theta)) - U(\theta)) f(\theta) d\theta$$
$$\text{subject to } \dot{U}(\theta) = g(q(\theta), \theta) \text{ a.e., and } U(\theta) \geq \hat{U}(\theta) \text{ for all } \theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}],$$

where the control variable q is measurable. This variable is generally interpreted as a quantity vector that belongs to a feasible set $\mathcal{Q} \subseteq \mathbb{R}^k$ (for $k \in \mathbb{N}$). Note that the domain of definition of q can be easily included into the objective to fit with the formalism of Theorem 1 if we set $\tilde{s}(\theta, q(\theta)) = -\infty$ for $q \notin \mathcal{Q}$. The differential equation that defines $\dot{U}(\theta)$ follows from incentive compatibility. The participation constraint $U(\theta) \geq \hat{U}(\theta)$ reflects the possibility that different types face different outside options.

The program $(\mathcal{P}')$ can easily be transformed into the canonical program $(\mathcal{P})$ we explore by defining $u(\theta) = U(\theta) - \hat{U}(\theta)$ and assuming that $\hat{U}(\theta)$ is differentiable:

$$s(\theta, \dot{u}(\theta)) = \max_{q \in \mathcal{Q}} \left\{ \tilde{s}(\theta, q(\theta)) f(\theta) \text{ subject to } \dot{u}(\theta) = \dot{\hat{U}}(\theta) + g(q(\theta), \theta) \text{ and } q(\theta) \in \mathcal{Q} \right\}.$$

This reduction is particularly easy when g is itself a bijection between q and $\dot{u}$, for any θ. Such a restriction on g implies that the control q can be expressed as a function of $\dot{u}$ and θ: $q(\theta) = g_q^{-1}(\dot{u}(\theta) - \dot{\hat{U}}(\theta), \theta)$. In this case, we may substitute and obtain

$$s(\theta, \dot{u}(\theta)) = \tilde{s}(\theta, g_q^{-1}(\dot{u}(\theta) - \dot{\hat{U}}(\theta), \theta))) f(\theta).$$

This path for proceeding when the control variable v is multi-dimensional is illustrated in Sections 4 when we describe equilibria with exclusivity clauses. Observe that, because the function $s(\theta, \dot{u}(\theta))$ so obtained is now a maximum over all possible controls that lead to the same derivative $\dot{u}(\theta)$ of the state variable, it may be *"smoother"* than $\tilde{s}(\theta, q(\theta))$ itself. Nevertheless, this extra degree of smoothness might not suffice to guarantee the assumption of twice-continuous differentiability relied upon in Jullien (2000).

4. APPLICATIONS

This Section presents simple economic settings for which the optimization tools developed in this paper are definitively needed.

4.1. *Nonlinear Pricing by Digital Firm under Competitive Pressure*

Motivation. Pricing for digital products and other technology-based products and services is notoriously complex. One source of complexity is that variable increases in demand for those products are fulfilled by the addition of blocks of computing or network infrastructure. Variable costs for firms in such industry are zero while any extra block of services requires a costly investment to provide additional capacity. The nature of these investments introduces discontinuities in costs and surplus functions that calls for the use of the tools we develop in this paper. Because of the associated technical difficulties, the literature in the field has been rather sparse. Spulber (1993) and Thomas (2001) have analyzed nonlinear pricing by a monopolist with a cost function that entails a fixed capacity constraint. On top of the usual information distortions, the optimal output profile depends on the shadow cost of this capacity. Huang and Sundararajan (2011) take stock of such analysis to determine the set of types who are bunched at the capacity level in a model of nonlinear pricing for digital products. Our analysis below is complementary to theirs since they focus on the monopolistic case and with analyze the case of a competition with a fringe.

Technology. We focus on the case in which an extra block of capacity is needed to supply any output in excess of $q = 1$. The cost function of a dominant firm (sometimes referred to as *the principal*) producing such a digital product exhibits an upward jump discontinuity in cost at $q = 1$:

$$C(q) = \begin{cases} 0 & q \in [0, 1] \\ k & q \in (1, \overline{Q}]. \end{cases}$$

We assume that the upper bound on possible outputs, $\overline{Q}$, is large enough to cover all possible realizations of demand.

On the demand side, a consumer (hereafter, *the agent*) has a valuation for q units of services which is given by:

$$V(\theta, q) = (S - \theta)q - \frac{q^2}{2}$$

where S is the intercept of his demand function and θ is an heterogeneity parameter representing a demand shock that has a uniform distribution, $F(\theta) = \theta - \underline{\theta}$, on $\Theta = [\underline{\theta}, \overline{\theta}]$ with corresponding density $f(\theta) = \overline{\theta} - \underline{\theta} = 1$.

Efficiency. The cost discontinuity implies that the total surplus function, $s(\theta, q) = v(\theta, q) - C(q)$, is itself discontinuous with a downward jump at $q = 1$. Following the optimality conditions for a non-smooth objective which were given in Section 2, the efficient consumption level $q^*(\theta)$ solves the pair of conditions:

$$\theta \in \partial\, \overline{co}_{\mathbb{R}_+} \left(Sq - \frac{q^2}{2} - k\delta_{q>1} \right) (q^*(\theta)),$$

$$Sq^*(\theta) - \frac{(q^*(\theta))^2}{2} - k\delta_{q^*(\theta)>1} = \overline{co} \left(Sq - \frac{q^2}{2} - k\delta_{q>1} \right) (q^*(\theta)),$$

where $\delta_{q>1}(q)$ is the indicator function,

$$\delta_{q>1} = \begin{cases} 0 & \text{if } q < 1 \ , \\ 1 & \text{otherwise.} \end{cases}$$

From these conditions, it is immediate to derive the efficient output as

$$q^*(\theta) = \begin{cases} 1 & \text{if } \theta \in \left[S - 1 - \sqrt{2k}, S - 1\right], \\ S - \theta & \text{otherwise} \end{cases}$$

where we assume $\underline{\theta} + \sqrt{2k} < S - 1 < \overline{\theta}$ to maintain positive outputs everywhere and to ensure that there are always high-demand types for whom incurring the fixed cost k is socially valuable.

Competitive fringe. The customer may also buy from a competitive fringe which, although it provides an alternative lower-quality version of the digital good, does not face any capacity constraint. The customer's valuation for this alternative good is

$$\hat{V}(\theta, q) = (\hat{S} - \theta)q - \frac{q^2}{2}.$$

We assume that the difference in marginal valuations between the dominant firm and the competitive fringe is positive: $\Delta S = S - \hat{S} > 0$. Taking this outside option, the customer buys $\hat{q}(\theta) = \hat{S} - \theta$ units of service which yields a payoff worth

$$\hat{U}(\theta) = \max_{q \in \mathcal{Q}} (\hat{S} - \theta)q - \frac{q^2}{2} \equiv \frac{(\hat{S} - \theta)^2}{2}, \tag{4.1}$$

where we assume $\underline{\theta} < \hat{S} - 1 < \overline{\theta}$ to ensure non-negative consumption for any type that chooses to buy from the fringe suppliers.

Incentive Compatibility and Participation constraints. We are interested in finding the optimal nonlinear price $T(q)$ that the dominant firm should design to extract the consumer' surplus under the threat of switching to the competitive fringe. From the Taxation Principle,[6] focusing on this class of strategy is without loss of generality. Define $U(\theta)$ to be the agent's value, or *informational rent*, conditional on acceptance:

$$U(\theta) \equiv \max_{q \in [0, \overline{Q}]} V(\theta, q) - T(q).$$

Because $\overline{Q}$ is finite, it is straightforward to check that $U(\theta)$ is absolutely continuous. $U(\theta)$ is also a minimum of linear functions of θ and therefore is convex and admits a sub-differential $\partial U(\theta)$ which is single-valued, almost everywhere. In particular, at any point of differentiability

$$\dot{U}(\theta) = -q(\theta). \tag{4.2}$$

At points of non-differentiability, we may arbitrarily choose q to equal either the left or right derivative of U.

[6] Rochet (1987).

The customer chooses either to buy exclusively from the dominant firm or from the fringe. He thus accepts the dominant firm's offer whenever the following type-dependent participation constraint holds:

$$U(\theta) \geq \hat{U}(\theta) \quad \forall \theta \in \Theta. \tag{4.3}$$

From here, it is straightforward to reformulate the problem to fit Theorem 1. First, we set $\dot{u}(\theta) = \dot{U}(\theta) - \dot{\hat{U}}(\theta) = \hat{q}(\theta) - q(\theta)$ so as (4.3) is characterized as in (3.1). Expressed in terms of $v = \dot{u}$, the surplus function is

$$s(\theta, v) = (S - \theta)(\hat{q}(\theta) - v) - \frac{1}{2}(\hat{q}(\theta) - v)^2 - k\delta_{q>1}(\hat{q}(\theta) - v).$$

The concave envelope of this surplus function admits the following subdifferential

$$\partial_v \overline{\mathrm{co}}(s)(\theta, v) = \begin{cases} -\Delta S - v & \text{for } v \leq \hat{q}(\theta) - 1 - \sqrt{2k}\ , \\ -\Delta S - \hat{q}(\theta) + 1 + \sqrt{2k} & \text{for } \hat{q}(\theta) - 1 - \sqrt{2k} \leq v < \hat{q}(\theta) - 1, \\ [-\Delta S - \hat{q}(\theta) + 1, -\Delta S - \hat{q}(\theta) + 1 + \sqrt{2k}] & \text{for } v = \hat{q}(\theta) - 1, \\ -\Delta S - v & \text{for } v > \hat{q}(\theta) - 1\ . \end{cases}$$

We are now equipped to derive the optimal solution but to do so, and to compare our techniques and those used in the smooth scenario by Jullien (2000), we first start with the case $k = 0$.

Smooth surplus function: $k = 0$. We start with the simplest scenario where $s(\theta, v)$ is smooth and strictly concave in its second argument and, thus, admits a subgradient everywhere with a single element, the derivative. Under these conditions, our Theorem 1 takes the same form as Theorem 1 in Jullien (2000). The adjoint function $\overline{\gamma}_0(\theta) = \int_{\underline{\theta}}^{\theta} \mu_0(ds)$ and the optimal output $\overline{q}_0(\theta) = \hat{q}(\theta) - \dot{\overline{u}}_0(\theta)$ must satisfy:

$$\overline{\gamma}_0(\theta) - \theta + \underline{\theta} = \frac{\partial s}{\partial v}(\theta, \dot{\overline{u}}_0(\theta)) = -S + \theta + \hat{q}(\theta) - \dot{\overline{u}}_0(\theta) \quad a.e.,$$

which can be rewritten as

$$\overline{q}_0(\theta) = S - 2\theta + \underline{\theta} + \overline{\gamma}_0(\theta). \tag{4.4}$$

Had the incumbent firm been a monopoly, its optimal output would still be given by the above formula, with the well-known feature that the customer's participation constraint would bind only at $\overline{\theta}$. In that case, $\overline{\gamma}_0(\theta) = 0$ for all $\theta \in [\tilde{\theta}, \overline{\theta})$. The monopoly schedule would exhibit such a strong downwards output distortion that customers with high types would prefer to switch to the higher consumption available from the competitive fringe. We thus conjecture that the participation constraint (4.3) is binding on an interval $[\tilde{\theta}, \overline{\theta}]$ and so

$$\overline{\gamma}_0(\theta) = 0 \quad \forall \theta \in [\underline{\theta}, \tilde{\theta}). \tag{4.5}$$

From this conjecture, the support of the measure μ would be $[\tilde{\theta}, \overline{\theta}]$. In other words, the optimal output satisfies

$$\overline{q}_0(\theta) = \max\{S - 2\theta + \underline{\theta}, \hat{q}(\theta)\}. \tag{4.6}$$

Accordingly, we look for a measure μ_0 which is absolutely continuous with respect to the Lebesgue measure on $[\tilde{\theta}, \overline{\theta})$ but has a charge at $\overline{\theta}$. Since (4.3) is binding on $[\tilde{\theta}, \overline{\theta}]$, differentiating the equality $\overline{u}_0(\theta) = 0$ yields $\hat{q}(\theta) = \overline{q}_0(\theta)$ a.e. on the interior on this interval. Inserting into (4.4) immediately gives us the following expression of $\overline{\gamma}_0(\theta)$:

$$\overline{\gamma}_0(\theta) = -\Delta S + \theta - \underline{\theta}, \quad \forall \theta \in [\tilde{\theta}, \overline{\theta}). \tag{4.7}$$

By differentiating the latter expression, we obtain that μ_0 has a positive density, $d\mu_0(\theta) = d\theta$, on $[\tilde{\theta}, \overline{\theta}]$. Because we must have $\overline{\gamma}(\overline{\theta}) = 1$, it follows that μ must also have a mass point at $\overline{\theta}$. The required charge is $\mu(\{\overline{\theta}\}) = \Delta S < 1$. Finally, $\tilde{\theta}$ is defined by a "smooth-pasting condition" $S - 2\tilde{\theta} + \underline{\theta} = \hat{q}(\tilde{\theta})$; output is continuous at $\tilde{\theta}$ and $\overline{u}_0$ is differentiable at that point. We find

$$\tilde{\theta} = \underline{\theta} + \Delta S, \tag{4.8}$$

which is interior because $\underline{\theta} + \Delta S < \overline{\theta} = \underline{\theta} + 1$.

From the specifications above, it becomes easy to check that conditions (3.2), (3.3) and (3.4) all hold. From Theorem 1, these conditions are indeed sufficient for optimality. This validates our *"guess-and-verify"* approach.

NON-SMOOTH SURPLUS FUNCTION: $k > 0$. Under this scenario, the result in Jullien (2000) cannot be applied, and so we turn to Theorem 1 of the present paper. To streamline exposition and limit the number of possible cases to study, suppose that

$$\hat{S} - \tilde{\theta} > 1.$$

This condition ensures that the dominant firm incurs some extra cost k if it wants to offer the same output as the fringe everywhere on a right-neighborhood of $\tilde{\theta}$ since such output is above 1 on that neighborhood. To avoid paying such cost, the dominant firm would like to keep output just below 1 and induce a bunch of types to take that option, as in the efficient scenario. Yet, this condition is in conflict with the requirement of a binding participation constraint everywhere on the right of $\tilde{\theta}$. The economic intuition here is that there are efficiency gains not to pay for the fixed cost k and part of these gains are redistributed to the customer under the form of information rent when his type lies in such a neighborhood. In other words, the existing capacity constraint of the dominant firm forces it to give to some types more than what they would get with the fringe. The market is segmented with four different connected subsets of types as we see below.

To prove this result, we proceed as above with a *"guess-and-verify"* approach. Of course, a good candidate for the adjoint function $\overline{\gamma}$ is to start from $\overline{\gamma}_0$ and modify this adjoint according to the economic intuition we just outlined for the shape of the solution. Consider the following adjoint:

$$\overline{\gamma}(\theta) = \begin{cases} 0 & \text{for } \theta \in [\underline{\theta}, \tilde{\theta}), \\ -\Delta S + \theta - \underline{\theta} & \text{for } \theta \in [\tilde{\theta}, \theta_1) \text{ and } \theta \in [\theta_2, \overline{\theta}), \\ \gamma_1 & \text{for } \theta \in [\theta_1, \theta_2), \end{cases} \tag{4.9}$$

for some θ_1, θ_2 and γ_1 yet to be determined. Compared with $\overline{\gamma}_0$, the new adjoint function $\overline{\gamma}$ has thus a similar expression on $[\underline{\theta}, \theta_1)$ but it also has an upward jump at the point θ_1, followed by a plateau. This shape means that the participation constraint binds over $[\tilde{\theta}, \theta_1)$ but that the measure μ has now a mass at θ_1. In a right-neighborhood of θ_1, namely (θ_1, θ_2), the participation constraint is instead slack, before it binds again on the top interval $[\theta_3, \overline{\theta}]$ for some θ_3 soon to be defined.

The optimal output is given as follows:

$$\overline{q}(\theta) = \begin{cases} S - 2\theta + \underline{\theta} & \text{for } \theta \in [\underline{\theta}, \tilde{\theta}), \\ \hat{S} - \theta & \text{for } \theta \in [\tilde{\theta}, \theta_1) \text{ and } \theta \in [\theta_3, \overline{\theta}), \\ 1 & \text{for } \theta \in [\theta_1, \theta_2), \\ S - 2\theta + \underline{\theta} + \gamma_1 & \text{for } \theta \in [\theta_2, \theta_3). \end{cases} \tag{4.10}$$

This optimal output remains equal to $\overline{q}_0(\theta)$ for the bottom interval $[\underline{\theta}, \theta_1)$, i.e., for those types with the highest consumption levels for which the incumbent finds it worth to incur the extra capacity. This optimal output has also a downward jump at θ_1. Types in the right-neighborhood $[\theta_1, \theta_2)$ are all bunched and consume one unit although they would like to get more. However, the dominant firm prefers not to pay the fixed cost for providing such extra consumption. Those types nevertheless enjoy a price discount to compensate for the constrained consumption and, as a result, they get a payoff above what they would get with the fringe. Yet, types with a lower valuation for the good consume less than 1 unit of service up to the point where the lowest valuations in the interval $[\theta_3, \overline{\theta}]$ again consume the same amount as with the fringe.

At type θ_1, the dominant firm is indifferent between selling the same quantity than the fringe if this quantity is above 1 and saving on the cost of the capacity but selling only 1 unit. This condition is

$$\hat{S} - \theta_1 = 1 + \sqrt{2k}.$$

This type θ_1 is above $\tilde{\theta}$ when $\hat{S} - 1 - \sqrt{2k} > \tilde{\theta}$.

Type θ_2 is the highest type to receive one unit of services and it thus verifies:

$$S - 2\theta_2 + \underline{\theta} + \gamma_1 = 1.$$

Type θ_3 is the highest type to receive more and get more rent than what he would get from the fringe:

$$\theta_3 = \Delta S + \underline{\theta} + \gamma_1.$$

Finally, the parameter γ_1, is chosen so that $\overline{u}(\theta_1) = \overline{u}(\theta_3) = 0$. This condition implies that $\int_{\theta_1}^{\theta_3} (\hat{q}(\theta) - \overline{q}(\theta))d\theta = 0$. Using the expression of $\overline{q}$ given in (4.10) finally gives us the following condition:

$$\int_{\theta_1}^{\theta_2} (\hat{q}(\theta) - 1)d\theta + \int_{\theta_2}^{\theta_3} (\hat{q}(\theta) - (S - 2\theta + \underline{\theta} + \gamma_1))d\theta = 0.$$

Geometrically, this condition just says that the algebraic area between the curves $\hat{q}$ and $\overline{q}$ is zero over $[\theta_1, \theta_3]$. Using this fact, we find after simplifications

$$\gamma_1 = 2\hat{S} - S + 2\sqrt{k} - \underline{\theta} - 1 < 1,$$

where the last inequality follows when k is small enough from all our previous conditions on parameter values. Observe that $\overline{\gamma}$ has a upward jump at θ_1 which means that the measure μ puts a Dirac mass worth $(2 + \sqrt{2})\sqrt{k}$ at θ_1.

It is straightforward to check that $(\overline{q} = \hat{q} - \dot{\overline{u}}, \overline{\gamma})$ altogether satisfy the optimality conditions (3.2), (3.3) and (3.4). Hence, from Theorem 1, we have the solution of our problem.

4.2. *Lobbying Competition*

Motivation. Consider two competing interest groups (the principals, sometimes referred to as P_i for $i = 1, 2$ in the sequel) with conflicting preferences $S_1(q) = -S_2(q) = q$ over a policy decision q. That policy belongs to an interval $\mathcal{Q} = [-\overline{Q}, \overline{Q}]$ where $\overline{Q}$ is supposed to be large enough to avoid corner solutions under all circumstances below. Principals compete by means of transfer schedules $t_i(q)$ (bribes or campaign contributions) for the service of a decision-maker (the agent) whose decisions are on *"sale"*. The decision-maker's preferences over the policies, q, and payments, t, are expressed as $t - \theta q - \frac{q^2}{2}$. The agent's bliss point, $q = -\theta$, is private information. For simplicity, we assume that the agent's bliss point is uniformly distributed over $[-\delta, \delta]$ with $\delta < 1$. By choosing his bliss point, the agent would get a reservation payoff $U_0(\theta) = \frac{\theta^2}{2}$. Since principals are symmetrically biased in opposite directions, if they were able to cooperate, they would collectively induce the agent to make this choice.

Exclusive Equilibria. The nature of the principals' conflicting preferences suggests that we may be able to construct an equilibrium outcome with fierce *head-to-head* competition for the exclusivity of the agent's services. Indeed, principal P_1 enjoys higher policies and is ready to cajole types closer to $-\delta$ since those types find it more attractive to increase the policy. This is the reverse for principal P_2 who prefers types closer to δ. Our goal here is to construct such an equilibrium outcome in which principals split the type space; each keeping influence on a subset of nearby types.

To this end, we conjecture that each principal will offer a schedule that induce exclusive dealing for half the type space (the marginal agent being located at $\theta = 0$), and that each principal will offer marginal tariffs that correspond with a monopolist's marginal tariffs over the exclusive region. This leads one to hypothesize that principal 2, for example, will offer a tariff of the form

$$t_2(q) = \alpha_2 - \frac{1-\delta}{2}q + \frac{q^2}{4},$$

for some α_2. We further conjecture that competition between the principals will be entirely focused on the marginal agent and will result in positive rents for this type, leading to an optimality condition for marginal influence. In particular, we conjecture that Principal 2 will offer the following upper semi-continuous contribution schedule:

$$\overline{t}_2(q) = \begin{cases} \frac{5}{4}(1-\delta)^2 - \frac{1-\delta}{2}q + \frac{q^2}{4} & \text{if } q \le -1 + \delta \\ 0 & \text{otherwise.} \end{cases} \qquad (4.11)$$

We then verify (using Theorem 1) that this does indeed comprise an equilibrium strategy.

Observe that $\bar{t}_2(-1+\delta) = 2(1-\delta)^2 > 0$ so that $\bar{t}_2(q)$ has a downward jump at $-1+\delta$ and is strictly decreasing, convex and positive for $q \leq -1+\delta$. Because payments are non-negative, the agent always finds it optimal to take P_2's contribution even when he is not subject to P_1's influence. Doing so gives him a reservation payoff $\hat{U}(\theta)$ while choosing a policy $\hat{q}(\theta)$ such that

$$\hat{U}(\theta) = \max_{q \in \mathcal{Q}} \bar{t}_2(q) - \theta q - \frac{q^2}{2} \text{ and } \hat{q}(\theta) \in \partial \hat{U}(\theta)$$

Easy but tedious computations lead to:

$$\hat{q}(\theta) = \begin{cases} -2\theta - 1 + \delta & \text{if } \theta \geq 0 \\ -1+\delta & \text{otherwise.} \end{cases} \quad (4.12)$$ [7]

Suppose now that principal P_1 can react to P_2's offer by himself imposing an exclusivity clause in his relationship with the agent. To model such setting, we allow principal P_1 to suggest to the agent when he should contract with him and when he should instead only deal with principal P_2. Formally, a direct and truthful revelation mechanism offered by P_1 is now a triplet $\left\{t_1(\hat{\theta}), q(\hat{\theta}), p(\hat{\theta})\right\}$ that specifies respectively a payment t_1, a policy q and a probability p of dealing exclusively with P_1 as a function of the agent's message $\hat{\theta}$ on his bliss point.[8] With such scheme, principal P_1 also induces a rent profile defined as:

$$U(\theta) = \max_{\hat{\theta} \in \Theta} \; p(\hat{\theta}) \left(t_1(\hat{\theta}) - \theta q(\hat{\theta}) - \frac{q^2}{2}(\hat{\theta}) \right) + (1 - p(\hat{\theta}))\hat{U}(\theta).$$

Since the possibility of exclusion is already encapsulated into the definition of $U(\theta)$, the agent's participation constraint can now be written as:

$$U(\theta) \geq \hat{U}(\theta) \quad \forall \theta \in [-\delta, \delta]\,. \quad (4.13)$$

From Theorem 2 in Milgrom and Segal (2002, p. 586), $U(\theta)$ so defined is absolutely continuous and thus almost everywhere differentiable with

$$\dot{U}(\theta) = -\left(p(\theta)q(\theta) + (1-p(\theta))\hat{q}(\theta)\right) \quad a.e.. \quad (4.14)$$

In the sequel, we will reduce the requirement of incentive compatibility to the necessary condition (4.14) (thus leaving it to check ex post that this condition is also sufficient).

With such a mechanism, principal P_1 gets a payoff when dealing with an agent of type θ which is worth:

$$p(\theta)\left((1-\theta)q(\theta) - \frac{q^2(\theta)}{2}\right) + (1-p(\theta))\hat{q}(\theta) + (1-p(\theta))\hat{U}(\theta) - U(\theta).$$

[7] The rent $\hat{U}(\theta)$ is convex and actually everywhere differentiable since $\hat{q}(\theta)$ is itself continuous so that $\partial\hat{U}(\theta)$ is always single-valued.

[8] We are using here the Revelation Principle to compute best responses to the nonlinear tariff $\bar{t}_2(q)$ offered by P_2. In a second step, we will recover from the optimal direct revelation mechanism so obtained a nonlinear $\bar{t}_1(q)$. This approach is legitimate to characterize all equilibrium allocations as discussed in Peters (2001) and Martimort and Stole (2002).

To write the optimization problem in a form consistent with our canonical formulation, we introduce the new state variable $u(\theta) = U(\theta) - \hat{U}(\theta)$ and the new control variable $z(\theta) = q(\theta) - \hat{q}(\theta)$. Because principal P_1 is looking for policies which are *a priori* greater than those implemented by P_2, we shall also impose the restriction that z is non-negative. With this notation, we may then rewrite (4.14) and (4.13), respectively, as

$$\dot{u}(\theta) = -p(\theta)z(\theta) \quad a.e., \tag{4.15}$$

$$u(\theta) \geq 0 \quad \forall \theta \in [-\delta, \delta]\,. \tag{4.16}$$

Principal P_1's optimization problem becomes now

$$(\mathcal{P}): \quad \max_{(p\in[0,1],v,z)} \int_{-\delta}^{\delta} \left[\hat{q}(\theta) + p(\theta)\left((1-\theta-\hat{q}(\theta))z(\theta) - \frac{1}{2}z^2(\theta) - \bar{t}_2(\hat{q}(\theta))\right) - u(\theta)\right]\frac{d\theta}{2\delta},$$

subject to (4.15) and (4.16).

The control variable (p, z) is bi-dimensional. To transform this problem into our canonical form, we proceed as suggested in Section 3 and define

$$2\delta s(\theta, v) = \max_{p,z}\ \hat{q}(\theta) + p\left((1-\theta-\hat{q}(\theta))z - \frac{1}{2}z^2 - \bar{t}_2(\hat{q}(\theta))\right) \quad \text{s.t. } v = -pz \text{ and } p \in [0,1]\,, z \geq 0. \tag{4.17}$$

In the Appendix, we show that this surplus function can be expressed as

$$2\delta s(\theta, v) = \hat{q}(\theta) - (1-\theta-\hat{q}(\theta))v + \begin{cases} -\frac{v^2}{2} - \bar{t}_2(\hat{q}(\theta)) & \text{if } v \leq -\sqrt{2\bar{t}_2(\hat{q}(\theta))}, \\ \sqrt{2\bar{t}_2(\hat{q}(\theta))}v & \text{if } v \in [-\sqrt{2\bar{t}_2(\hat{q}(\theta))}, 0]. \end{cases} \tag{4.18}$$

Although $s(\theta, v)$ is concave and is continuously differentiable, it is not twice so as required in Jullien (2000), so we apply our Theorem 1. In particular, the optimality condition (3.3) takes a simple form:

$$\overline{\gamma}(\theta)) - \frac{\theta+\delta}{2\delta} = \frac{\partial s}{\partial v}(\theta, \dot{\overline{u}}(\theta)) = -(1-\theta-\hat{q}(\theta)) + \begin{cases} -\dot{\overline{u}}(\theta) & \text{if } \dot{\overline{u}}(\theta) \leq -\sqrt{2\bar{t}_2(\hat{q}(\theta))}, \\ \sqrt{2\bar{t}_2(\hat{q}(\theta))} & \text{if } \dot{\overline{u}}(\theta) \in [-\sqrt{2\bar{t}_2(\hat{q}(\theta))}, 0]. \end{cases} \tag{4.19}$$

Again, we guess the structure of the solution and, thanks to the sufficiency part of our Theorem, we only check that the optimality conditions (3.2), (3.3) and (3.4) are satisfied by a pair $(\overline{u}(\theta), \overline{\gamma}(\theta))$. To this end, consider the following optimal policy choice and exclusion probability induced by P_1:

$$\overline{q}(\theta) = \begin{cases} 1-\delta-2\theta & \text{for } \theta \in [-\delta, 0) \\ \hat{q}(\theta) & \text{for } \theta \in (0, \delta]; \end{cases} \quad \text{and } \overline{p}(\theta) = \begin{cases} 1 & \text{for } \theta \in [-\delta, 0) \\ 0 & \text{for } \theta \in (0, \delta]. \end{cases} \tag{4.20}$$

Together $\overline{q}$ and $\overline{p}$ implies that:

$$\dot{\overline{u}}(\theta) = \begin{cases} 2(-1+\delta+\theta) & \text{for } \theta \in [-\delta, 0) \\ 0 & \text{for } \theta \in (0, \delta]. \end{cases} \tag{4.21}$$

The following adjoint function

$$\overline{\gamma}(\theta) = \begin{cases} 0 & \text{for } \theta \in [-\delta, 0) \\ 1 & \text{for } \theta \in [0, \delta]. \end{cases} \tag{4.22}$$

together with $\overline{v} = \dot{\overline{u}}$ just defined satisfy the optimality condition (4.19). With such adjoint function, the participation constraint (4.16) is binding on $[0, \delta]$ which determines $\overline{u}(0) = 0$. The corresponding measure μ has a Dirac mass at 0. From an economic viewpoint, it means that all competition to attract the agent's services takes place for that type in the middle of the type space.

The *bang-bang* nature of the optimal probability of exclusion allows us to transform the optimal direct mechanism found above into a simple nonlinear instrument; a tariff $\bar{t}_1(q)$ that is offered to the agent only if he does not serve the rival principal P_2. It is actually easy to check that this tariff satisfies $\bar{t}_1(q) = \bar{t}_2(-q)$ and thus:

$$\bar{t}_1(q) = \begin{cases} \frac{5}{4}(1-\delta)^2 + \frac{1-\delta}{2}q + \frac{q^2}{4} & \text{if } q \geq -1+\delta \\ 0 & \text{otherwise.} \end{cases} \tag{4.23}$$

In other words, our optimization techniques have thus allowed us to characterize a (symmetric) equilibrium of the lobbying game where principals insist on exclusivity. With those contributions, the principals' respective dominance areas are on each side of the type space with almost all types dealing with only one principal. On his own dominance area, each principal behaves as a monopoly and it can be readily verified that the policy $\overline{q}(\theta)$ is the same as that that would be offered in the monopoly scenario where the competing lobbying group would be absent. Schedules nevertheless differ because principals might bid up for the services of the marginal agent located at 0. The discontinuity in $\bar{t}_2(q)$ at $-1+\delta$ is indeed chosen so that principal P_1 is indifferent between inducing his "monopoly" policy $\overline{q}(0) = 1-\delta > 0$ from that type and compensating this agent for the extra payment he would get from P_2 when choosing $\hat{q}(0) = -1+\delta < 0$. The equilibrium policy is thus discontinuous at $\theta = 0$, with that type being indifferent between choosing $1-\delta > 0$ and $-1+\delta < 0$ and the participation constraint (4.13) (and a similar one obtained by solving P_2's own best-response optimization problem) being binding at that point.

Relationship with the I.O. literature on competitive screening. The features of the exclusive equilibria we have just unveiled are reminiscent of some models of competition in contracts that were developed in the I.O. literature. Champsaur and Rochet (1989) have analyzed how the seller of a high-quality good may restrict the quality spectrum when he competes with a fringe offering a lower quality in a model of vertical differentiation à la Mussa and Rosen (1978). Stole (1995) has instead analyzed how sellers who are horizontally differentiated may compete one with the other by attracting the marginal customer in their clienteles. In these models, the space of contracts is *a priori* limited in the sense that deviations available to a principal never allow for the possibility that this principal may induce a stochastic acceptance of the competing seller's own contract. Our analysis allows for such a possibility and, remarkably, shows that equilibrium contracts are non-stochastic. A second difference between those models and the lobbying environment we are analyzing here is the public good nature of the agent's decision. This means that, even if a principal is not active on a subset of types, he earns some payoff in the lobbying environment under scrutiny. This possibility entails a possible discontinuity in the surplus function that was not stressed in the I.O. environment since a seller makes no profits if the customer is served by his rival.

Overlapping Influence. Suppose now that principal P_1 cannot enforce any exclusivity clause. The decision-maker may now contract with both lobbying groups. The game becomes now a delegated common agency game in the parlance of Bernheim and Whinston (1986). Our goal here is not to characterize all equilibria of such games. For the purpose of the present paper and to illustrate the power of the techniques, we are simply going to test whether the equilibrium contract of the exclusive scenario remains an equilibrium contract when exclusivity can no longer be enforced. This exercise requires us to solve for the optimal tariff that principal P_1 offers at a best response to $\bar{t}_2(q)$.

To this end, first observe that, because $\bar{t}_2(q)$ is non-negative, the agent is always at least weakly better off accepting this offer, possibly choosing a policy greater than $-1+\delta$ and not being paid for that by principal P_2. By the same token, principal P_1 certainly never induces a policy choice corresponding to a negative payment since the agent would be better off refusing such contract. Hence, P_1 offers a non-negative nonlinear tariff $t_1(q)$. With such tariff, P_1 now induces a rent profile for the common agent which is worth:

$$U(\theta) = \max_{q\in\mathcal{Q}} \; t_1(q) + \bar{t}_2(q) - \theta q - \frac{q^2}{2}.$$

By a by-now standard argument, $U(\theta)$ is a convex function, absolutely continuous and thus almost everywhere differentiable with a derivative at those points still given by (4.2). Finally, the agent's participation constraint is still given by (4.13).

We thus rewrite P_1's problem as

$$(\mathcal{P}) : \max_{(q,U)} \int_{-\delta}^{\delta} \left[(1-\theta)q(\theta) - \frac{q^2(\theta)}{2} + \bar{t}_2(q(\theta)) - U(\theta) \right] \frac{d\theta}{2\delta},$$

subject to (4.2) a.e., and (4.13).

Adopting our previous notation, i.e., $u(\theta) = U(\theta) - \hat{U}(\theta)$ and $v(\theta) = q(\theta) - \hat{q}(\theta)$, we may transform this problem into its canonical form by defining the surplus function as

$$2\delta s(\theta, v) = \hat{q}(\theta) - (1-\theta-\hat{q}(\theta))v - \frac{v^2}{2} - \bar{t}_2(\hat{q}(\theta)) + \bar{t}_2(\hat{q}(\theta) - v). \tag{4.24}$$

This surplus function is upper semi-continuous with a downward jump at $v = \hat{q}(\theta) + 1 - \delta$. Again, we rely on Theorem 1 to characterize principal P_1's best response. The first step is to compute $\overline{\text{co}}(s)$. Tedious computations show that the concave envelope of this surplus function can be expressed as

$$2\delta\overline{\text{co}}(s)(\theta, v) = \begin{cases} \hat{q}(\theta) + \frac{2\theta-1-\delta+\hat{q}(\theta)}{2}v - \frac{v^2}{4} & \text{if } \hat{q}(\theta)+1-\delta \le v \\ \hat{q}(\theta) + \frac{(\hat{q}(\theta)-1+\delta)^2}{4} + (\theta-\delta)v & \text{if } \hat{q}(\theta)-1+\delta \le v \le \hat{q}(\theta)+1-\delta \\ \hat{q}(\theta) - (1-\theta-\hat{q}(\theta))v - \frac{v^2}{2} - \bar{t}_2(\hat{q}(\theta)) & \text{if } v \le \hat{q}(\theta)-1+\delta. \end{cases}$$

This concave envelope is made of three pieces, two being quadratic and the last one being linear. It admits the following subdifferential

$$2\delta\partial_v\overline{\text{co}}(s)(\theta,v)=\begin{cases}\hat{q}(\theta)+\frac{2\theta-1-\delta+\hat{q}(\theta)}{2}-\frac{v}{2} & \text{if } \hat{q}(\theta)+1-\delta\leq v\\ \theta-\delta & \text{if } \hat{q}(\theta)-1+\delta\leq v\leq\hat{q}(\theta)+1-\delta\\ -(1-\theta-\hat{q}(\theta))-v & \text{if } v\leq\hat{q}(\theta)-1+\delta.\end{cases}$$

To get a simple characterization of the influence area where P_1 actively impacts on the decision, and especially to stress the most interesting scenarios where the participation constraint (4.13) is binding on an interval with non-empty interior, we shall also assume from now on that $\delta\geq\frac{1}{2}$. Under this condition, the following ouput profile and adjoint function summarize P_1's best response:

$$\overline{q}(\theta)=\begin{cases}1-\delta-2\theta & \text{for } \theta\in[-\delta,1-\delta],\\ -1+\delta=\hat{q}(\theta) & \text{for } \theta\in[1-\delta,\delta];\end{cases}\tag{4.25}$$

$$2\delta\overline{\gamma}(\theta)=\max\{\theta-1+\delta,0\}.\tag{4.26}$$

It is indeed easy to check that all conditions (3.2), (3.3) and (3.4) are satisfied by the pair $(\overline{q},\overline{\gamma})$ and the associated rent profile $\overline{u}$. Moreover, the participation constraint is now binding on on the upper tail interval $[1-\delta,\delta]$. Interestingly, we obtain from the characterization above:

$$\dot{\overline{u}}(\theta)=\begin{cases}2(1-\delta-\theta) & \text{for } \theta\in[-\delta,1-\delta],\\ 0 & \text{for } \theta\in[1-\delta,\delta].\end{cases}\tag{4.27}$$

This example showcases two important phenomena that do not arise in the smooth setting in Jullien (2000). First, even though the optimal output $\overline{q}$ is continuous, the rent profile $\overline{U}$ is no longer smooth at $1-\delta$, the lowest bound of the support of the measure μ. This lack of smoothness is not due to an upward jump in $\overline{\gamma}$ (i.e., a Dirac mass of the measure μ at that point). Instead, it comes from the fact that the surplus function has a downward jump and the optimum is found at an extreme point of the affine piece in $\overline{co}(s)(\theta,v)$ for $\theta\in[1-\delta,\delta]$, namely at $q=-1+\delta$. Second, at such point, $\partial_v\overline{co}(s)(\theta,v)$ is an interval when $q=-1+\delta$. There is some leaway in specifying the adjoint function. Any other increasing function $\overline{\gamma}'$ such that $\overline{\gamma}'(\theta)\in[\overline{\gamma}(\theta),1]$ for $\theta\in[1-\delta,\delta]$ and $\overline{\gamma}'(\theta)=0$ otherwise supports the optimal allocation.

From an economic viewpoint, the best responses with and without an exclusivity clauses are radically different. Without exclusivity, the influence area of P_1 is enlarged. All types in $[0,1-\delta]$ now receive a positive payment from P_1 and modify their choice accordingly by taking decisions $\overline{q}(\theta)$ that are above those that they would be taking by only dealing with principal P_2, namely $\hat{q}(\theta)$. With an exclusivity clause, P_1 can enforce a policy that is even closer to its own preferences for those types $\theta\in[-1+\delta,0]$. It is achieved by raising payments so as to make this exclusive option more attractive. Without an exclusivity clause, such raise would be wasted. P_1 prefers to keep payments low but expands influence beyond 0. In other words, the equilibrium allocation found with exclusivity is no longer an equilibrium outcome when exclusivity clauses cannot be enforced. The strategic nature of the game with and without exclusivity are thus quite different.

Relationship with the lobbying literature. The analysis of the game that was developed above with the exclusivity requirement is novel. In Martimort and Stole (2015), we use the tools of non-smooth optimal control to provide a set of necessary conditions that are satisfied by all equilibria in delegated common agency games, including discontinuous ones. Those equilibria may sometimes Pareto dominate the equilibrium outcome that one would obtain assuming smooth tariffs; an important reason for developing the tools presented in this note.[9]

REFERENCES

Auriol, E. and P.M. Picard (2009). Government Outsourcing: Public Contracting with Private Monopoly. *The Economic Journal,* **119** 1464–1493.

Baron, D. and R. Myerson (1982). Regulating a Monopolist with Unknown Costs. *Econometrica*, **50** 911–930.

Bernheim, B. and M. Whinston (1986). Menu Auctions, Resource Allocation, and Economic Influence. *The Quarterly Journal of Economics* **10** 1–31.

Brainard, S. and D. Martimort (1997). Strategic Trade Policy with Incompletely Informed Policymakers. *Journal of International Economics,* **42**, 33–65.

Calzolari, G. and V. Denicolo (2013). Competition with Exclusive Contracts and Market-Share Discounts. *The American Economic Review,* **103** 2384–2411.

Champsaur, P. and J.C. Rochet (1989). Multiproduct Duopolists. *Econometrica* **57** 533–557.

Curien, N., B. Jullien and P. Rey (1998). Pricing Regulation under Bypass Competition. *The RAND Journal of Economics,* **29** 259–279.

Deb, R. and M. Said (2015). Dynamic Screening with Limited Commitment. *Journal of Economic Theory,* **159** 891–928.

Ferrera, J. (2014). *An Introduction to Nonsmooth Analysis.* Academic Press.

Huang, K. and A. Sundararajan (2011). Pricing Digital Goods: Discontinuous Costs and Shared Infrastructure. *Information Systems Research,* **22** 721–738.

Jullien, B. (2000). Participation Constraints in Adverse Selection Models. *Journal of Economic Theory* **93** 1–47.

Laffont, J.J. and D. Martimort (2002). *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press.

Lewis, T. and D. Sappington (1989). Countervailing Incentives in Agency Problems. *Journal of Economic Theory* , **49** 294–313.

Maggi, G. and A. Rodriguez-Clare (1995). On Countervailing Incentives. *Journal of Economic Theory.* **66** 238–263.

Martimort, D. and A. Semenov (2008). Ideological Uncertainty and Lobbying Competition. *Journal of Public Economics*, **92** 456–481.

Martimort, D. and L. Stole (2002). The Revelation and Delegation Principles in Common Agency Games. *Econometrica.* **70** 1659–1673.

Martimort, D. and L. Stole (2009). Market Participation in Delegated and Intrinsic CommonAgency Games. *The RAND Journal of Economics*, **40**, 78–102.

Martimort, D. and L. Stole (2015). Menu Auctions and Influence Games. Mimeo Chicago Booth

[9]Martimort and Semenov (2008) have analyzed such smooth equilibria under alternative specifications of the decision-maker's objective.

School of Business.

Milgrom, P. and I. Segal (2002). Envelope Theorems for Arbitrary Choice Sets. *Econometrica,* **70** 583–601.

Mirrlees, J. (1971). An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies,* **38** 175–208.

Mussa, M. and S. Rosen (1978). Monopoly and Product Quality. *Journal of Economic Theory,* **18** 301–317.

Myerson, R. (1981). Optimal Auction Design. *Mathematics of Operations Research.* **6** 58–73.

Ollier, S. and L. Thomas (2013). Ex Post Participation Constraint in a PrincipalAgent Model with Adverse Selection and Moral Hazard. *Journal of Economic Theory,* **148** 2383–2403.

Peters, M. (2001). Common Agency and the Revelation Principle. *Econometrica.* **69** 1349–1372.

Rochet, J. C. (1985). The Taxation Principle and Multi-Time Hamilton-Jacobi Equations. *Journal of Mathematical Economics.* **14** 113–128.

Rochet, J. C. (1987). A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context. *Journal of Mathematical Economics.* **16** 191–200.

Rockafellar, T. and R. Wets (2004). *Variational Analysis.* Berlin: Springer.

Royden, H.L. (1988). *Real Analysis.* New York: Prentice Hall.

Seierstad, A. and K. Sydsaeter (1987). *Optimal Control Theory with Economic Applications.* Amsterdam: North Holland.

Spulber, D. (1993). Monopoly Pricing of Capacity Usage under Asymmetric Information. *Journal of Industrial Economics* **41** 241–257.

Stole, L. (1995). Nonlinear Pricing and Oligopoly. *Journal of Economics and Management Strategy,* **4** 529–562.

Stole, L. (2003). Price Discrimination and Imperfect Competition. In *Handbook of Industrial Organization,* M. Armstrong and R. Porter eds., **3** 34–47.

Thomas, L. (2001). Cost Functions in Non-Linear Pricing. *Economics Letters* **72** 53–59.

Vinter, R. (2000). *Optimal Control.* Boston: Birkhauser.

Vinter, R. and H. Zheng (1995). Necessary Conditions for Optimal Control Problem with State Constraints. *Transactions of the American Mathematical Society* **350** 1181–1204.

5. PROOFS

5.1. *Theorem 1*

5.1.1. *Preliminaries for Non-Smooth Analysis*

We draw heavily from Vinter and Zheng (1998) in the following presentation. A complete treatment can be found in the monograph of Vinter (2000). Theorem 3 from Vinter and Zheng (1998) appears as Theorem 10.2.1 in Vinter (2000).

Take a closed set $A \subseteq \mathbb{R}^n$ and a point $x \in A$. A vector $r \in \mathbb{R}^n$ is a *limiting normal* to A at x if there exists a sequence $(x_i, r_i) \to (x, r)$ with $x_i \in A$ and a constant $M \geq 0$ such that for each i in the sequence $r_i \cdot (x_i - x) \leq M||x_i - x||^2$, where $|| \cdot ||$ denotes Euclidean distance. The cone of limiting normal vectors to A at x is denoted $N_A(x)$. Given a lower semi-continuous function $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ and a point $x \in \mathbb{R}$ such that $g(x) < +\infty$, the *limiting sub-differential* of g at x is defined as

$$\partial g(x) \equiv \{\xi \,|\, (\xi, -1) \in N_{\text{epi}\{g\}}(x, g(x))\},$$

where epi$\{g\}$ is the *epigraph* of the function g defined as

$$\text{epi}\{g\} \equiv \{(x,\alpha) \in \mathbb{R} \times \mathbb{R} \,|\, \alpha \geq g(x)\}.$$

The *asymptotic limiting sub-differential* of g at x, written $\partial^\infty g(x)$, is defined as

$$\partial^\infty g(x) \equiv \{\xi \,|\, (\xi, 0) \in N_{\text{epi}\{g\}}(x, g(x))\}.$$

Finally, we define:

$$\partial_x^> h(t,x) \equiv \overline{co}\{\lim_i \xi_i \,|\, \exists t_i \to t, x_i \to x \;\text{ s.t. } h(t_i, x_i) > 0 \text{ and } \xi_i \in \partial_x h(t_i, x_i)\ \forall i\}.$$

Two results from non-smooth analysis (Vinter (2000), Propositions 4.3.3 and 4.3.4) that we use are (1) $\partial^\infty g(x) = \{0\}$ if g is Lipschitz continuous and (2) for any x such that $g(x)$ is finite,

$$N_{\text{epi}\{g\}}(x, g(x)) = \{(\xi d, -\xi) \,|\, \xi > 0\,, d \in \partial g(x)\} \cup \{\partial^\infty g(x) \times \{0\}\}.$$

A *local maximizer* of $\Lambda(x)$ is a feasible arc, $\overline{x}$, which maximizes $\Lambda(x)$ over all feasible arcs $x \in AC(\Theta, \mathbb{R}_+)$ within an ε neighborhood of $\overline{x}$, $||\overline{x} - x||_{ac} \leq \varepsilon$ where we denote the norm on the space of absolutely-continuous functions by $||x||_{ac} \equiv ||x(\underline{\theta})|| + \int_\Theta ||\dot{x}(\theta)|| d\theta$. A *local minimizer* is defined analogously.

5.1.2. *Necessity*

First, and for completeness, we reproduce here Theorem 3 of Vinter and Zheng (1998) which provides necessary conditions for solutions to the following minimization program:

$$(\mathcal{P}') : \qquad \text{Minimize } J(x) \equiv \int_{\underline{\theta}}^{\overline{\theta}} L(\theta, x(\theta), \dot{x}(\theta)) d\theta$$

$$\text{subject to } x \in AC(\Theta, \mathbb{R}) \text{ and } h(\theta, x(\theta)) \leq 0 \text{ for all } \theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}].$$ [10]

We will prove necessity for Theorem 1 by specializing Theorem 3 from Vinter and Zheng (1998), exploiting fact that our integrand in Λ is a linear function of x and $h(\theta, x) = -x$.

Theorem A.1 *(Vinter and Zheng (1998), Theorem 3) Let $\overline{x}$ be local minimizer for $(\mathcal{P}')$ in $AC(\Theta, \mathbb{R})$ such that $J(\overline{x}) < +\infty$. Assume that the following hypotheses are satisfied:*

- *H_1. $L(\cdot, x, \cdot)$ is $\mathcal{L} \times \mathcal{B}$ measurable for each x and $L(\theta, \cdot, \cdot)$ is lower semi-continuous for a.e. $\theta \in \Theta$.*
- *H_2. For every $N > 0$ there exists $\delta > 0$ and $k \in L^1$ such that*

$$||L(\theta, x', v) - L(\theta, x, v)|| \leq k(\theta)||x' - x||, \quad L(\theta, \overline{x}(\theta), v) \geq -k(\theta)$$

  for a.e. $\theta \in \Theta$, for all $x, x' \in \overline{x}(\theta) + \delta B$ and $v \in \dot{\overline{x}}(\theta) + NB$, where B is a unit Euclidean ball.
- *H_3. h is upper semi-continuous near $(\theta, \overline{x}(\theta))$ for all $\theta \in \Theta$, and there exists a constant k_h such that*

$$||h(\theta, x') - h(\theta, x)|| \leq k_h ||x' - x||$$

  for all $\theta \in \Theta$ and all $x', x \in \overline{x}(\theta) + \delta B$.

Then there exist an arc $p \in AC(\Theta, \mathbb{R})$, a constant $\lambda \geq 0$, a non-negative measure μ on the Borel subsets of Θ and a μ-integrable function $\zeta : \Theta \to \mathbb{R}$, such that

[10] We specialize their theorem to our present problem in which the range of $x(\theta)$ is one-dimensional and there is no endpoint cost function.

(i). $\lambda + \max_{\theta\in\Theta} |p(\theta)| + \int_\Theta \mu(d\tilde{\theta}) = K > 0$ *(where K is an arbitrary normalization constant),*[11]

(ii).

$$\dot{p}(\theta) \in \overline{co}\left\{\eta \,|\, (\eta, p(\theta) + \int_{[\underline{\theta},\theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta}), -\lambda) \in N_{\text{epi}\{L(\theta,\cdot,\cdot)\}}(\overline{x}(\theta), \dot{\overline{x}}(\theta), L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)))\right\} \text{ a.e.,}$$

(iii).

$$p(\underline{\theta}) = p(\overline{\theta}) - \int_\Theta \zeta(\tilde{\theta})\mu(d\tilde{\theta}) = 0,$$

(iv).

$$\left(p(\theta) + \int_{[\underline{\theta},\theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta})\right)\dot{\overline{x}}(\theta) - \lambda L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) \in \arg\max_{v\in\mathbb{R}} \left(p(\theta) + \int_{[\underline{\theta},\theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta})\right) v - \lambda L(\theta, \overline{x}(\theta), v),$$

(v). $\zeta(\theta) \in \partial_x^{>} h(\theta, \overline{x}(\theta))$ *μ-a.e. and* $\text{supp}\{\mu\} \subseteq \{\theta \,|\, h(\theta, \overline{x}(\theta)) = 0\}$.

We apply this result to our setting by substituting $xf(\theta) - s(\theta, v)$ in program $(\mathcal{P})$ in place of $L(\theta, x, v)$ and thereby converting the maximization functional Λ in program $(\mathcal{P})$ to the minimization functional J in program $(\mathcal{P}')$. We complete the transformation by requiring that $h(\theta, x) = -x$, and that $L(\theta, x, v)$ is a linear function of x for any (θ, v).

First, we verify that hypotheses H_1-H_3 are satisfied for our program $(\mathcal{P})$. Because $s(\theta, \cdot)$ is upper semi-continuous and $\mathcal{B}$-measurable, and because $L(\theta, x, v)$ is linear in x, H_1 is satisfied. H_2 requires that $L(\theta, \cdot, v)$ is Lipschitz continuous, which is trivial given that L is linear in x with coefficient $f(\theta)$. Because the transformed program has $h(\theta, x) = -x$, h is a continuous linear function of x and thus H_3 is also satisfied.

Next, we specialize the conclusions of Vinter and Zheng (1998) by making use of the additional restrictions on $L(\cdot)$ and $h(\cdot)$. We present this in the following Lemma.

LEMMA 1 *Suppose that $L(\theta, x, v)$ is a linear function of x and that $h(\theta, x) = -x$. Then the conclusions (i)-(v) of Theorem A.1 imply*

(a). $\lambda + \max_{\theta\in\Theta} |p(\theta)| + \int_\Theta \mu(d\tilde{\theta}) = K$,

(b). $\dot{p}(\theta) = \lambda f(\theta)$ *a.e.,*

(c). $p(\underline{\theta}) = p(\overline{\theta}) + \int_\Theta \zeta(\tilde{\theta})\mu(d\tilde{\theta}) = 0$

(d). $\dot{\overline{x}}(\theta) \in \arg\max_{v\in\mathbb{R}} \left(p(\theta) + \int_{[\underline{\theta},\theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta})\right) v + \lambda s(\theta, v)$, *a.e.,*

(e). $\zeta(\theta) = -1$ *μ-a.e. and* $\text{supp}\{\mu\} \subseteq \{\theta \,|\, \overline{u}(\theta) = 0\}$.

[11] We choose to state the Theorem using $K > 0$ as an arbitrary normalization rather than $K = 1$, which is the normalization chosen in Vinter and Zheng (1998). Later, by setting $K = 3$, we will succeed in normalizing μ to a probability measure which is a more familiar object.

Proof of Lemma 1: Implications (i) and (a) are identical. Implication (ii) requires

$$\dot{p}(\theta) \in \overline{co}\left\{\eta \mid \left(\eta, p(\theta) + \int_{[\underline{\theta},\theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta}), -\lambda\right) \in N_{\text{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x}(\theta), \dot{\overline{x}}(\theta), L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta))\right)\right\}, a.e.$$

Because $L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) = f(\theta)\overline{x}(\theta) - s(\theta, \dot{\overline{x}}(\theta))$ is finite, the limiting normal cone in the above expression can be written as

$$\begin{aligned} &N_{\text{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x}(\theta), \dot{\overline{x}}(\theta), L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta))\right) \\ &\quad = \left\{(\xi d_1, \xi d_2, -\xi) \mid \xi > 0,\ (d_1, d_2) \in \partial\left(f(\theta)\overline{x}(\theta) - s(\theta, \dot{\overline{x}}(\theta))\right)\right\} \\ &\qquad \bigcup \left\{\partial^\infty\left(f(\theta)\overline{x}(\theta) - s(\theta, \dot{\overline{x}}(\theta))\right) \times \{0\}\right\}. \end{aligned}$$

Using the fact that $L(\cdot)$ is additively separable in x and $\dot{x}$ yields (Rockafellar and Wets (2004, Proposition 10.5))

$$\begin{aligned} \partial\left(f(\theta)\overline{x}(\theta) - s(\theta, \dot{\overline{x}}(\theta))\right) &= \partial\left(f(\theta)\overline{x}(\theta)\right) \times \partial\left(-s(\theta, \dot{\overline{x}}(\theta))\right) \\ &= \{f(\theta)\} \times \partial\left(-s(\theta, \dot{\overline{x}}(\theta))\right) \end{aligned}$$

and

$$\begin{aligned} \partial^\infty\left(f(\theta)\overline{x}(\theta) - s(\theta, \dot{\overline{x}}(\theta))\right) &\subseteq \partial^\infty\left(f(\theta)\overline{x}(\theta)\right) \times \partial^\infty\left(-s(\theta, \dot{\overline{x}}(\theta))\right) \\ &= \{0\} \times \partial^\infty\left(-s(\theta, \dot{\overline{x}}(\theta))\right), \end{aligned}$$

where the last equality uses the fact that a linear function is Lipschitz continuous and hence $\partial^\infty(f(\theta)\overline{u}(\theta)) = \{0\}$. Substituting these sub-differentials into the expression for the limiting normal cone, we have a simple inclusion:

$$\begin{aligned} N_{\text{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x}(\theta), \dot{\overline{x}}(\theta), L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta))\right) &\subseteq \left\{(\xi f(\theta), \xi d_2, -\xi) \mid \xi > 0,\ d_2 \in \partial\left(-s(\theta, \dot{\overline{x}}(\theta))\right)\right\} \\ &\qquad \bigcup \left\{\{0\} \times \partial^\infty\left(-s(\theta, \dot{\overline{x}}(\theta))\right) \times \{0\}\right\}. \end{aligned}$$

This simplifies again to the inclusion

$$\begin{aligned} &N_{\text{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x}(\theta), \dot{\overline{x}}(\theta), \overline{L}(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta))\right) \\ &\quad \subseteq \left\{(\xi f(\theta), \xi d_2, -\xi) \mid \xi \geq 0,\ d_2 \in \partial\left(-s(\theta, \dot{\overline{x}}(\theta))\right) \bigcup \partial^\infty\left(-s(\theta, \dot{\overline{x}}(\theta))\right)\right\}. \end{aligned}$$

The key point to note is that any vector in the limiting normal cone must point in the same direction in the $(\overline{x}, \overline{L})$ plane, regardless of d_2. Returning to implication (ii), we see that any point η in the given convex hull must satisfy $(\eta, \cdot, -\lambda) = (\xi f(\theta), \cdot, -\xi)$ for some $\xi \geq 0$, and hence the convex hull reduces to $\{\lambda f(\theta)\}$. We conclude that implication (ii) simplifies to implication (b) given that $L(\cdot)$ is both additively separable and linear in x.

Implication (iii) is identical to implication (c).

Using the transformation $L(\theta, x, v) = xf(\theta) - s(\theta, v)$, implication (iv) simplifies to implication (d). Lastly, the fact that $h(\theta, x) = -x$ yields $\partial_x h(\theta, \overline{u}(\theta)) = \partial_x^> h(\theta, \overline{x}(\theta)) = \{-1\}$. Thus, implication (v) simplifies to $\zeta(\theta) = -1$ μ-a.e. and

(A1) $\quad supp\{\mu\} \subseteq \{\theta \mid \overline{u}(\theta) = 0\}.$

This is implication (e). *Q.E.D.*

An immediate inspection of conditions (a)-(e) suggest further simplifications by combining these conditions. Conditions (b) and (c) jointly yield

$$p(\theta) = \lambda F(\theta).$$

Because $p(\overline{\theta}) = \lambda$ and $\zeta(\theta) = -1$ a.e. with respect to μ, condition (c) also implies

$$\int_{\Theta} \mu(d\tilde{\theta}) = \lambda.$$

Because we also have $\max_{\theta\in\Theta} |p(\theta)| = \lambda$, condition (a) implies $\lambda > 0$ and in particular $\lambda = \frac{K}{3}$. Because the choice of K is arbitrary, we choose $K = 3$ as a normalization, yielding $\lambda = 1$ and $\int_{\Theta} \mu(d\tilde{\theta}) = 1$. Thus, up to this normalization, μ is a probability measure on Θ. Defining now $\overline{\gamma}(\theta) = \int_{[\underline{\theta},\theta)} \mu(d\tilde{\theta})$, the implication in (d) is therefore

$$\text{(A2)} \qquad \dot{\overline{x}}(\theta) \in \arg\max_{v\in\mathbb{R}} \; s(\theta, v) + (F(\theta) - \overline{\gamma}(\theta))\, v, \; a.e..$$

This condition can finally be expressed as (3.3) and (3.4) of Theorem 1. Lastly, implication of (e) delivers the complementary slackness condition (3.2). We have therefore proven the necessity of the conditions in Theorem 1.

5.1.3. *Sufficiency*

Sufficiency is proven by generalizing Arrow's Sufficiency Theorem to non-smooth optimal control problems and specializing the theorem to the case in which the objective integrand is a linear function of x. We adapt the argument of Arrow's Sufficiency Theorem using the basic approach of Seierstad and Sydsaeter (1987) but relaxing their continuity and smoothness assumptions. The regularity of the optimal solution follows from arguments involving the necessary conditions.

Let x be any admissible arc satisfying thus $x \in AC(\Theta, \mathbb{R})$ and $x(\theta) \geq 0$ for all $\theta \in \Theta$. Define

$$\Delta = \int_{\Theta} \left\{ \left(s(\theta, \dot{\overline{x}}(\theta)) - \overline{x}(\theta) f(\theta)\right) - \left(s(\theta, \dot{x}(\theta)) - x(\theta) f(\theta)\right) \right\} d\theta.$$

We will demonstrate that, under conditions (A1) and (A2) of Theorem 1, $\Delta \geq 0$.

To this end, it is useful to define the Hamiltonian for program (P) with $\overline{\gamma}(\theta)$ being the adjoint equation which satisfies conditions(A1) and (A2):

$$H(\theta, x, v) \equiv s(\theta, v) - x f(\theta) - (\overline{\gamma}(\theta) - F(\theta)) v.$$

Note that $\overline{\gamma}(\theta)$ is defined for $\theta \in (\underline{\theta}, \overline{\theta}]$ and thus $H(\cdot)$ inherits the same domain. Nonetheless, because μ is not part of expression of Δ and F is absolutely continuous, we can ignore the point $\underline{\theta}$ in the integral and conclude that

$$\Delta = \int_{(\underline{\theta},\overline{\theta}]} \left(H(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) - H(\theta, x(\theta), \dot{x}(\theta))\right) d\theta + \int_{\Theta} (F(\theta) - \overline{\gamma}(\theta)) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta.$$

Define the optimized Hamiltonian as

$$\hat{H}(\theta, x) \equiv \sup_{v\in\mathbb{R}} \; H(\theta, x, v).$$

Because $\overline{\gamma}(\theta) - F(\theta)$ is bounded on $(\underline{\theta}, \overline{\theta}]$ and $s(\theta, \cdot)$ is bounded from above by assumption, $\hat{H}(\cdot)$ must be finite. Condition (A2) implies that

$$\hat{H}(\theta, \overline{x}(\theta)) = H(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta))$$

and for any admissible $x \in AC(\Theta; \mathbb{R}_+)$,

$$\hat{H}(\theta, x(\theta)) \geq H(\theta, x(\theta), \dot{x}(\theta)).$$

Combining these facts, we obtain

$$\begin{aligned} H(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) - H(\theta, x(\theta), \dot{x}(\theta)) &\geq \hat{H}(\theta, \overline{x}(\theta)) - \hat{H}(\theta, x(\theta)) \\ &= f(\theta)(x(\theta) - \overline{x}(\theta)). \end{aligned}$$

The last statement relies fundamentally on the linearity of $H(\cdot)$ in x. Substituting into the previous statement for Δ, we have

$$\begin{aligned} \Delta &\geq \int_{(\underline{\theta},\overline{\theta}]} f(\theta)(x(\theta) - \overline{x}(\theta))d\theta + \int_{\Theta} (F(\theta) - \overline{\gamma}(\theta)) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= \int_{\Theta} \left(f(\theta)(x(\theta) - \overline{x}(\theta)) + F(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right)\right) d\theta - \int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= \int_{\Theta} \frac{d}{d\theta}[F(\theta)(x(\theta) - \overline{x}(\theta))]d\theta - \int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= (x(1) - \overline{x}(1)) - \int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta. \end{aligned}$$

It follows that $\Delta \geq 0$ if

$$(x(1) - \overline{x}(1)) - \int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \geq 0.$$

If $\overline{\gamma}$ were absolutely continuous, we would be able to integrate the second term by parts and reach such a conclusion. Because $\overline{\gamma}$ has possibly countable upward discontinuities, we must proceed more carefully. Note that $\overline{\gamma}$ is non-decreasing on $(\underline{\theta}, \overline{\theta}]$ with at most a countable number of upward jump discontinuities. $\overline{\gamma}$ is thus the sum of a countable number of singular jump functions plus a measure $d\mu(\theta)$ which is absolute continuous with respect to the Lebesgue measure and thus write as $d\mu(\theta) = \nu(\theta)d\theta$.[12] Denote the set of jump discontinuities by $\{\tau_1, \tau_2, \ldots\}$, a possibly infinite but countable set. Let $\mathcal{I}$ be the index set of τ_i and let define the size of the jump discontinuity at any τ_i by $\Delta\mu(\tau_i) = \overline{\gamma}(\tau_i^+) - \overline{\gamma}(\tau_i) > 0$. We thus write:

$$\overline{\gamma}(\theta) = \sum_{\tau_i \leq \theta, i \in \mathcal{I}} \Delta\mu(\tau_i) + \int_{\underline{\theta}}^{\theta} \nu(\theta)d\theta.$$

Since $\overline{\gamma}$ is absolutely continuous outside the discontinuities, we can integrate by parts between any pair of discontinuities. Also note that at any such upward jump point, τ, $\overline{\gamma}$ is left- and right-continuous with $\overline{\gamma}(\tau) < \overline{\gamma}(\tau^+)$ and (by condition (A1)) we have $\overline{x}(\tau^+) = 0$.

Between any two points τ_i and τ_{i+1}, we know

$$\begin{aligned} &\int_{(\tau_i,\tau_{i+1}]} \overline{\gamma}(\theta) \left(\dot{u}(\theta) - \dot{\overline{v}}(\theta)\right) d\theta \\ &\qquad = \overline{\gamma}(\theta)(x(\theta) - \overline{x}(\theta))\Big|_{t=\tau_i^+}^{\tau_{i+1}} - \int_{(\tau_i,\tau_{i+1})} (x(\theta) - \overline{x}(\theta))\nu(\theta)d\theta \\ &\qquad = \overline{\gamma}(\tau_{i+1})(x(\tau_{i+1}) - \overline{x}(\tau_{i+1})) - \overline{\gamma}(\tau_i^+)(x(\tau_i) - \overline{x}(\tau_i)) \\ &\qquad\quad - \int_{(\tau_i,\tau_{i+1})} (x(\theta) - \overline{x}(\theta))\nu(\theta)d\theta. \end{aligned}$$

[12]Royden (1988).

The second equality above uses the fact that x and $\overline{x}$ are continuous on Θ.

Then we may write

$$\begin{aligned}
&\int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta)\left(\dot{x}(\theta)-\dot{\overline{x}}(\theta)\right)d\theta \\
&= \sum_{i\in\mathcal{I}} \overline{\gamma}(\tau_{i+1})(x(\tau_{i+1})-\overline{x}(\tau_{i+1})) - (\Delta\mu(\tau_i)+\overline{\gamma}(\tau_i))(x(\tau_i)-\overline{x}(\tau_i)) \\
&- \int_{(\tau_i,\tau_{i+1})}(x(\theta)-\overline{x}(\theta))\nu(\theta)d\theta \\
&\qquad = (x(1)-\overline{x}(1)) - \sum_{i\in\mathcal{I}}\Delta\mu(\tau_i)(x(\tau_i)-\overline{x}(\tau_i)) - \int_{(\tau_i,\tau_{i+1})}(x(\theta)-\overline{x}(\theta))\nu(\theta)d\theta.
\end{aligned}$$

By complementary slackness in condition (A1), we know $\overline{x}(\theta)\nu(\theta)=0$ and at any jump point τ_i we must have $\overline{x}(\tau_i)=0$. Thus,

$$\int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta)\left(\dot{x}(\theta)-\dot{\overline{x}}(\theta)\right)d\theta = (x(1)-\overline{x}(1)) - \sum_{i\in\mathcal{I}}\Delta\mu(\tau_i)x(\tau_i) - \int_{(\tau_i,\tau_{i+1})} x(\theta)\nu(\theta)d\theta.$$

We deduce

$$\begin{aligned}
\Delta &\geq (x(1)-\overline{x}(1)) - \int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta)\left(\dot{x}(\theta)-\dot{\overline{x}}(\theta)\right)d\theta \\
&= \sum_{i\in\mathcal{I}}\Delta\mu(\tau_i)x(\tau_i) + \int_{(\tau_i,\tau_{i+1})} x(\theta)\nu(\theta)d\theta.
\end{aligned}$$

Because $x(\theta)\geq 0$, μ is a non-negative measure, and jump discontinuities $\Delta\mu(\tau_i)$ are positive, we conclude that $\Delta\geq 0$ as claimed. We have proven that conditions (A1) and (A2) are sufficient for a solution.

5.1.4. *Smoothness of the Solution $\overline{x}$.*

We add the hypothesis that

$$v(\theta,\sigma)\equiv \arg\max_{v\in\mathbb{R}}\ s(\theta,v)+(F(\theta)-\sigma)v$$

is single-valued and continuous for $(\theta,\sigma)\in\Theta\times[0,1]$. It follows that $v(\theta,\sigma)$ is non-increasing in σ and from condition (A2), that $\dot{\overline{x}}(\theta)=v(\theta,\overline{\gamma}(\theta))$ a.e.

Suppose to the contrary that $\dot{\overline{x}}$ is discontinuous at some point $\tau\in\Theta$. Initially, suppose that Condition (A2) is extended to hold for all $\theta\in\Theta$ rather than for a.e. $\theta\in(\underline{\theta},\overline{\theta}]$; call this Condition (A2'). Condition (A2') and the additional hypothesis that $v(\theta,\sigma)$ is continuous in (θ,σ) jointly imply that $\dot{\overline{x}}(\theta)$ is discontinuous at τ only if $\overline{\gamma}$ is also discontinuous at τ. Any discontinuity in $\overline{\gamma}$, however, must be an upward jump, $\overline{\gamma}(\tau^+)-\overline{\gamma}(\tau)>0$, implying that $\dot{\overline{x}}(\theta)$ must jump downwards. Complementary slackness (Condition (A1), however, imposes that $\overline{x}(\tau)=0$, with the implication that a downward discontinuity at τ would violate the state constraint $u(\theta)\geq 0$ in the neighborhood to the immediate right of τ. Hence, continuity must hold for all points $\theta\in[\underline{\theta},\theta)$ under Condition (A2'). Furthermore, because $\overline{\gamma}$ is left continuous at $t=1$, no jump in $\dot{\overline{x}}(\theta)$ is possible at this endpoint. We conclude that Condition (A2') implies that $\dot{\overline{x}}(\theta)$ is continuous for all $\theta\in\Theta$. The weaker Condition (A2) allows $\dot{\overline{x}}(\theta)$ to violate the maximization condition on sets of measure zero, including at $\theta=\underline{\theta}$. But such violations have no effect on the solution $\overline{x}$ which is absolutely continuous. Thus, $\overline{x}$ is smooth as posited.

5.2. *Lobbying Game*

Exclusive Equilibria. Notice that $v = 0$ implies either $p = 0$ or $z = 0$. Inserting those values into the maximand of (4.17), we get:

$$\text{(B1)} \qquad 2\delta s(\theta, 0) = \hat{q}(\theta).$$

When $v \neq 0$, we necessarily have $p \neq 0$ and we can thus write $z = -\frac{v}{p}$. Inserting again into the maximand of (4.17), we obtain the following definition of $s(\theta, v)$ for $v < 0$:

$$\text{(B2)} \qquad 2\delta s(\theta, v) = \max_p \left\{ \hat{q}(\theta) - (1 - \theta - \hat{q}(\theta))v - \frac{v^2}{2p} - p\bar{t}_2(\hat{q}(\theta)) \text{ s.t. } p \in (0, 1] \right\}.$$

The maximand on the right-hand side of of (B2) is concave in p and its derivative is worth:

$$\frac{v^2}{2p^2} - \bar{t}_2(\hat{q}(\theta)).$$

Therefore, a corner solution $p = 1$ obtains when:

$$v^2 > 2\bar{t}_2(\hat{q}(\theta)).$$

In that case, we finally get:

$$\text{(B3)} \qquad 2\delta s(\theta, v) = \hat{q}(\theta) - (1 - \theta - \hat{q}(\theta))v - \frac{v^2}{2} - \bar{t}_2(\hat{q}(\theta)).$$

Suppose instead that

$$v^2 \leq 2\bar{t}_2(\hat{q}(\theta)).$$

The maximum on the right-hand side of (B2) is then interior and, since $v < 0$, achieved at

$$p = -\frac{v}{\sqrt{2\bar{t}_2(\hat{q}(\theta))}}.$$

We thus rewrite the value of this maximand as:

$$\text{(B4)} \qquad 2\delta s(\theta, v) = \hat{q}(\theta) - \left(1 - \theta - \hat{q}(\theta) - \sqrt{2\bar{t}_2(\hat{q}(\theta))}\right) v.$$

Using our *"guess-and-try"* approach, we now look for a solution $(\overline{v}(\theta), \overline{\gamma}(\theta))$ with the following features.

- On $[-\delta, 0)$, we have:

$$\text{(B5)} \qquad \overline{\gamma}(\theta) = 0, \overline{u}(\theta) > 0 \text{ and } \overline{v}(\theta) < 0 \text{ with } \overline{v}^2(\theta) > 2\bar{t}_2(-1 + \delta) = 4(1 - \delta)^2.$$

The latter condition implies:

$$\text{(B6)} \qquad p(\theta) = 1 \quad \forall \theta \in [-\delta, 0).$$

Under those conditions, and taking into account that the distribution of types is uniform, the optimality requirement (3.3) can now be rewritten as:

$$\overline{v}(\theta) = \dot{\overline{u}}(\theta) \in \arg\max_{v \leq 0} \left\{ \hat{q}(\theta) - (2 - \delta - 2\theta - \hat{q}(\theta))v - \frac{v^2}{2} - \bar{t}_2(\hat{q}(\theta)) \right\}$$

or

(B7) $\quad \overline{q}(\theta) = 1 - \delta - 2\theta \quad \forall \theta \in [-\delta, 0).$

We check that the last condition in (B5) then holds, as postulated. Inserting the value of $\overline{q}(\theta)$ and $\hat{q}(\theta)$ which are respectively obtained from (B7) and (4.12), we obtain:

$$\overline{v}^2(\theta) = 4(1-\delta-\theta)^2 > 4(1-\delta)^2$$

which trivially holds for $\theta < 0$.

- On $(0, \delta]$, we have:

(B8) $\quad \overline{\gamma}(\theta) = 1, \overline{u}(\theta) = 0$ and $\overline{v}(\theta) = 0$ with $\overline{v}^2(\theta) < 2\overline{t}_2(\hat{q}(\theta))$.

The first condition means that the adjoint $\overline{\gamma}(\theta)$ has an upward discontinuity at 0, corresponding to a Dirac mass of the measure μ at that point. The second condition implies that (B4) applies.

Under those conditions, and taking into account that the distribution of types is uniform, the optimality requirement (3.3) can be rewritten as:

$$\overline{v}(\theta) = \dot{\overline{u}}(\theta) \in \arg\max_{v \leq 0} \left\{ \hat{q}(\theta) - \left(1 + \delta - 2\theta - \hat{q}(\theta) - \sqrt{2\overline{t}_2(\hat{q}(\theta))}\right) v \right\}.$$

The maximand above is achieved for $\overline{v}(\theta) = 0$ if

(B9) $\quad 1 + \delta - 2\theta - \hat{q}(\theta) - \sqrt{2\overline{t}_2(\hat{q}(\theta))} \leq 0 \Leftrightarrow 2 \geq \overline{t}_2(\hat{q}(\theta)) \quad \forall \theta \in (0, \delta]$

where the last inequality is obtained after using (4.12). Now observe that $\overline{t}_2(q)$ being decreasing implies $\overline{t}_2(\hat{q}(\theta)) \leq \overline{t}_2(\hat{q}(\delta)) = 2 - 2\delta + \delta^2$ for all $\theta \in (0, \delta]$ which proves that the right-hand side inequality of (B9) holds since $2 - 2\delta + \delta^2 \leq 2$ for $\delta \leq 1$.

The condition $\overline{v}(\theta) = 0$ then implies:

(B10) $\quad \overline{q}(\theta) = \hat{q}(\theta) \quad \forall \theta \in (0, \delta].$

Note that the specifications of $\overline{\gamma}(\theta)$ and $\dot{\overline{u}}(\theta)$ given in (B5) and (B8) altogether imply that the participation constraint (4.13) is binding over the interval $[0, \delta]$ with a Dirac mass at 0 for the measure μ.

It is straightforward to check that the optimal policy given by (B7) for $\theta < 0$ and (B10) for $\theta > 0$ is implemented by P_1 with an exclusion rule given by (4.20). Using the above specification of the probability d'exclusion, it is also routine to check that the condition (4.14) is also sufficient for incentive compatibility since $\overline{q}(\theta)$ is non-increasing.