URAL MATHEMATICAL JOURNAL, Vol. 2, No. 2, 2016

OPTIMAL MULTIATTRIBUTE SCREENING[1]

Thomas A. Weber

Chair of Operations, Economics and Strategy
Swiss Federal Institute of Technology, Lausanne, Switzerland
thomas.weber@epfl.ch

Abstract: We provide a technique for constructing optimal multiattribute screening contracts in a general setting with one-dimensional types based on necessary optimality conditions. Our approach allows for type-dependent participation constraints and arbitrary risk profiles. As an example we discuss optimal insurance contracts.

Key words: Asymmetric Information, Incentive Contracting, Maximum Principle, Nonlinear Pricing.

1. Introduction

Starting with the seminal contribution by Mirrlees [1], optimal screening contracts have found many applications in economics, including taxation, nonlinear pricing, and the regulation of monopolies [1–4]. The underlying game of asymmetric information contains two periods. In the first period, the principal announces a menu of attribute bundles to an agent who possesses private information about his utility function, and who can select an attribute bundle by sending a message to the principal. In the second period, allocations are made according to a publicly known enforceable mapping from the message space to both attribute bundles and monetary transfers from agent to principal. Subject to the agent's participation constraint (individual rationality) and the agent's self-interested choice behavior (incentive compatibility), an optimal screening menu maximizes the principal's expected payoff. In this paper, which is related to [5,6], we provide an explicit method to construct optimal screening contracts with continuously distributed, one-dimensional "types" representing the agents' private information and multiple instruments (or attributes), based on necessary optimality conditions. The problem of finding optimal screening contracts can be formulated as an optimal control problem, for which we derive necessary optimality conditions using the maximum principle by Pontryagin et al. [7] and the technique of successive approximation [8]. In this vein, the method approximates an optimal schedule by directly computable solutions to a sequence of relaxed screening problems. We allow for type-dependent participation constraints and payoff functions that are nonlinear in each contract instrument. This is useful when considering the effects of variable outside options and/or risk-aversion on the optimal contracts. The set of participating types is also subject to optimization.

The paper is organized as follows. In Section 2, we introduce a general model for multiattribute screening and formulate the principal's optimization problem in an optimal-control setting. In Section 3, we provide a full set of necessary optimality conditions that can be used to construct a solution to the screening problem, including an iterative approximation method. In Section 4, we discuss the design of a menu of optimal insurance contracts for an agent with unknown risk aversion as a practical example. Section 5 concludes.

[1]This paper was presented as a plenary lecture in October 2016 in Ekaterinburg, at the International Conference on "Systems Analysis: Modeling and Control" in memory of Academician Arkady Kryazhimskiy.

2. Model

2.1. Screening Problem

We consider a standard screening setup with a principal ("she") and an agent ("he"). The agent's possible type θ, known only to him, lies in the type space $\Theta = [0,1]$.[2] The type θ summarizes all the private information the agent has. From the principal's viewpoint it is distributed with the continuous probability density $f = \dot{F}$ on the support Θ.[3] The principal designs a schedule of instruments (or attributes) $x = (x_0, x_1, \ldots, x_n) : \Theta \to \mathbb{R}_+^{n+1}$ so as to maximize her expected payoff,

$$\bar{V}(x;\theta_0) = \int_{\theta_0}^{1} V(x(\theta),\theta) f(\theta)\, d\theta, \tag{2.1}$$

where $n \geq 1$ is a given integer which denotes the number of attributes characterizing the bundle, and $\Theta_0 = [\theta_0, 1] \subseteq \Theta$ is the set of participating types, with the marginal type $\theta_0 \in \Theta$ subject to optimization.[4] In order to ensure implementability of a favored attribute schedule, the principal's optimization problem is subject to the agent's self-interested behavior, which manifests itself in the form of two constraints. First, given that an agent of type θ has net utility $U(\xi,\theta)$ for a bundle $\xi \in \mathbb{R}_+^{n+1}$, the optimal type announcement ϑ satisfies the *incentive-compatibility constraint*

$$U(x(\vartheta),\theta) \geq U(x(\hat{\theta}),\theta), \quad \forall\, \vartheta, \hat{\theta} \in \Theta_0. \tag{2.2}$$

Second, since the agent is free to walk away from the principal's proposed menu of contracts (attribute schedule), the participation set Θ_0 is defined by the agent's *participation constraint*

$$\vartheta, \theta \in \Theta_0 \quad \Leftrightarrow \quad U(x(\vartheta),\theta) \geq U(0,\theta). \tag{2.3}$$

That is, an agent of type θ participates if by pretending to be any other participating agent type ϑ (including himself) he achieves a utility that is at least equal to the utility derived from a zero bundle (corresponding to nonparticipation). The principal's screening problem is to find an (absolutely continuous) attribute schedule $x(\cdot)$ which maximizes the objective in (2.1) subject to the implementability constraints (2.2) and (2.3).

2.2. Key Assumptions

The attribute x_0 corresponds to a *numéraire* good such as a monetary transfer from the agent to the principal.[5] The principal's payoff function $V : \mathbb{R}^{n+1} \times \Theta \to \mathbb{R}$ is continuously differentiable in the attribute vector x and continuous in the type θ. In addition, we make the following two assumptions on V.

P1. $V_{x_0} > 0 > V_{x_i}$ for all $i \in \{1, \ldots, n\}$ (P-Monotonicity).[6]

P2. $V(0,\theta) \geq 0$ for all $\theta \in \Theta$ (Possibility of Inaction).

[2] Using a simple affine transformation this does in fact allow for *any* compact interval on the real line.

[3] Throughout we use the dot-notation for total derivatives with respect to θ, e.g., $\dot{F} = \frac{dF}{d\theta}$.

[4] The problem of maximizing the principal's expected utility can be viewed as a "dynamic optimization problem" if $\bar{V}(x;\theta_0)$ represents an average payoff generated by a trajectory $x(\theta)$ on the time interval $[\theta_0, 1]$, where the starting time θ_0 is subject to optimization. The methods presented in this paper extend to intervals of the form $[\theta_0, \theta_1]$ where both boundaries are subject to optimization (cf. footnote 23).

[5] We sometimes use x to denote bundles (i.e., points) in the attribute space $\mathbb{R}_+^{n+1} = \{\hat{x} \in \mathbb{R}^{n+1} : \hat{x} \geq 0\}$ rather than full schedules (i.e., functions); the same 'notational flexibility' is used for components x_i of x.

[6] Throughout the text, we denote partial derivatives using subscripts, e.g., $V_{x_0} = \partial V / \partial x_0$.

Agent type θ's net utility from the bundle x is $U(x,\theta)$, where $U : \mathbb{R}^{n+2} \to \mathbb{R}$ is a twice continuously differentiable function. His outside option (which is realized after not agreeing to any of the principals' proposed contracts) yields the *reservation utility* $r(\theta) = U(0,\theta)$, where $r : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function. We make the following three assumptions on U.

A1. For all $(x,\theta) \in \mathbb{R}^{n+1}_+ \times \Theta$: (i) $U_\theta(0,x_1,\ldots,x_n,\theta) \geq \dot{r}(\theta)$, and (ii) $U_{x_i}(x,\theta) > 0 > U_{x_0}(x,\theta)$ for all $i \in \{1,\ldots,n\}$ (A-Monotonicity).

A2. For all $(x_1,\ldots,x_n,\theta) \in \mathbb{R}^n_+ \times \Theta$: (i) $U(0,x_1,\ldots,x_n,\theta) \geq r(\theta)$ (Attribute Desirability), and (ii) $r(\theta) > \lim_{x_0\to\infty} U(x_0,x_1,\ldots,x_n,\theta)$ (Transfer Sensitivity).

A3. For all $(x,\theta) \in \mathbb{R}^{n+1}_{++} \times \Theta$: $|U_{x_{i'}\theta}(x,\theta)| > 0$ for some $i' \in \{1,\ldots,n\}$ (Incentive Regularity).[7]

Finally, to exclude unbounded screening contracts, the gains from trade between the principal and the agent need to be bounded. To formalize this notion, we first introduce the principal's *equivalent variation* $E(x_1,\ldots,x_n,\theta)$, which corresponds to the minimum payment in terms of the numéraire good x_0 she would be willing to accept to provide the attribute bundle $(x_1,\ldots,x_n)$ to the agent, i.e.,

$$E(x_1,\ldots,x_n,\theta) = \inf\{x_0 \in \mathbb{R}_+ : V(x_0,x_1,\ldots,x_n,\theta) \geq 0\}, \tag{2.4}$$

where we adopt the convention that $\inf \emptyset = \infty$. On the other hand, agent type θ's *compensating variation* $C(x_1,\ldots,x_n,\theta)$ is his maximum willingness to pay (in terms of x_0) for the attribute bundle $(x_1,\ldots,x_n)$ provided by the principal, i.e.,

$$C(x_1,\ldots,x_n,\theta) = \sup\{x_0 \in \mathbb{R}_+ : U(x_0,x_1,\ldots,x_n,\theta) \geq r(\theta)\}, \tag{2.5}$$

where we adopt the convention that $\sup \emptyset = -\infty$. The following assumption guarantees that gains from trade between principal and agent exist and the contracts between the two involve only bounded attribute bundles. Let

$$\mathcal{X} = \{(x_0,x_1,\ldots,x_n) \in \mathbb{R}^{n+1}_+ : C(x_1,\ldots,x_n,\theta) \geq x_0 \geq E(x_1,\ldots,x_n,\theta) \quad \text{and} \quad \theta \in \Theta\}$$

be the (feasible) contract space.[8]

T1. $\mathcal{X}$ is bounded (Bounded Contract Space).

We now comment on the six assumptions P1, P2, A1–A3, and T1. Assumption P2 bounds the principal's value of the screening problem from below by zero, since the zero schedule $x = 0$ is always feasible. Assumption P1 means that the principal's preferences are nonsatiated in the numéraire good. The monotonicity of the principal's payoff with respect to the non-numéraire attributes $(x_1,\ldots,x_n)$ is unimportant and can be relaxed in situations which involve cooperation between the principal and the agent. Correspondingly, we assume in A1(ii) that the agent dislikes paying the numéraire to the principal, but finds all other attributes desirable. The first inequality in P1 and A1(ii) can be always satisfied by relabelling and simple sign-transformations, as long as there is one attribute that both the principal and the agent like, so it is possible for the agent to compensate the principal for her actions regarding the other attributes. Assumption A1(i) guarantees that the set of participating types is convex (of the form $\Theta_0 = [\theta_0, 1]$ for some $\theta_0 \in$

[7] Without any loss of generality, we assume in what follows that $i' = 1$; furthermore, $\mathbb{R}^{n+1}_{++} = \text{int}(\mathbb{R}^{n+1}_+)$.

[8] The equivalent and compensating variations are classical welfare measures [9]. A normative relationship between them, in particular a lack of exchange due to the endowment effect (when $E > C$) for identical individuals, is discussed by [10]. Here, assumptions A1 and P1 imply that the contracting parties' preferences differ, and assumptions A2 and P2 that gains from trade exist.

$[0,1]$) and includes the highest type $\theta = 1$,[9] as long as it is nonempty. The latter is guaranteed by P2, since the principal can always provide the zero-attribute bundle at no disbenefit. The first inequality in A2 means that any agent always accepts the bundle $(0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1}_+$ when he obtains the non-numéraire attributes for free. The second inequality means that for any attribute vector $(x_1, \ldots, x_n) \in \mathbb{R}^n_+$ there is a price (in terms of x_0) that is too high. The incentive-regularity condition A3 requires that the agent perceives increasing (or decreasing) differences with respect to one non-numéraire attribute and his type. This condition is needed only in a neighborhood of points where the agents' incentive compatibility becomes a binding constraint. Finally, the assumption T1 stipulating that the gains from trade be bounded is naturally satisfied in any practical situation, e.g., when both U and V are bounded and there exists an ε such that $U_{x_0} < -\varepsilon < 0 < \varepsilon < V_{x_0}$. Intuitively it is enough when the principal's marginal costs $-V_{x_i}$ of providing a bundle (to a given agent type) increase fast enough in the attributes, and at the same time the agent's marginal utilities U_{x_i} for these attributes decrease. Assumption T1 guarantees that the solution to the principal's screening problem behaves as if it were constrained to attribute schedules with values in the set $\mathcal{X}$ without the need for an explicit consideration of this constraint. Note that assumptions P2 and A2 imply that the contract space is nonempty, as it must contain the zero bundle, i.e., $0 \in \mathcal{X}$.

3. Optimal Screening Contracts

We treat the screening problem in an optimal-control framework. Accordingly, the *admissible schedules* $x : \Theta \to \mathbb{R}^{n+1}_+$ are in the Sobolev space $\mathbf{W}_{1,\infty}$ of absolutely continuous functions with essentially bounded derivatives, and the corresponding class of *admissible controls* $u : \Theta \to \mathbb{R}^n$ is the Lebesgue space $\mathbf{L}_\infty$ of all essentially bounded functions. An admissible marginal type is any $\theta_0 \in \Theta$. Correspondingly, let $\mathcal{D} = \mathbf{L}_\infty(\Theta, \mathbb{R}^{n+1}_+) \times \mathbf{W}_{1,\infty}(\Theta, \mathbb{R}^n) \times \Theta$ be the *domain of admissibility* for solutions (x^*, u^*, θ_0^*) to the screening problem.[10]

Theorem 1. *Under assumptions P1 and A1–A2, the principal's screening problem can be written in the form:*

$$\sup_{(x,u,\theta_0)\in\mathcal{D}} \bar{V}(x;\theta_0), \tag{P}$$

subject to

$$\dot{x} = \Phi(x,\theta)u,\ U(x(\theta_0),\theta_0) = r(\theta_0), \tag{3.1}$$

and

$$\min_{\hat{\theta}\in\Theta_0} \left\{ (\hat{\theta} - \theta) \Big(U_\theta(x(\hat{\theta}),\theta) - U_\theta(x(\theta),\theta) \Big) \right\} \geq 0, \tag{3.2}$$

for all $\theta \in \Theta_0 = [\theta_0, 1]$, *where*[11]

$$\Phi(x,\theta) = \begin{bmatrix} \varphi(x,\theta) \\ \mathbf{I}_n \end{bmatrix} \in \mathbb{R}^{(n+1)\times n}, \quad \varphi(x,\theta) = -\frac{(U_{x_1}(x,\theta),\ldots,U_{x_n}(x,\theta))}{U_{x_0}(x,\theta)} \geq 0, \tag{3.3}$$

and $\mathbf{I}_n$ *denotes an* $(n \times n)$*-identity matrix.*

P r o o f. Given an admissible schedule $x : \Theta \to \mathbb{R}^{n+1}$, we conclude from A1 that $U(x(\hat{\theta}),\theta_1) - r(\theta_1) \geq U(x(\hat{\theta}),\theta_0) - r(\theta_0)$ for any $\hat{\theta}, \theta_1, \theta_0 \in \Theta$ with $\theta_1 \geq \theta_0$. Thus, type θ_0's participation

[9] Convexity of Θ_0 can be achieved also when $U - r$ is only quasiconcave in θ [11], in which case the upper marginal type becomes subject to optimization also.

[10] At an optimal solution (x^*, u^*, θ_0^*), the schedule x^* and control u^* need only be defined on the optimal participation set $\Theta_0^* = [\theta_0^*, 1] \subseteq \Theta$.

[11] A vector $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ satisfies $y \geq 0$ if and only if $y_i \geq 0$ for all $i \in \{1, \ldots, n\}$.

implies participation for all types $\theta \in \Theta_0 = [\theta_0, 1]$. Since by P1 and A1 it is $V_{x_0} > 0 > U_{x_0}$, the lowest participating type θ_0 cannot get any surplus, so necessarily $U(x(\theta_0), \theta_0) = r(\theta_0)$. The lowest type θ_0 is itself subject to optimization. Equation (3.1) states that $d(x_1, \dots, x_n)/d\theta = u$, where u is subject to optimization. When searching for schedules $x : \Theta \to \mathbb{R}^{n+1}$ that maximize expected profits $\bar{V}(x; \theta_0)$, by the revelation principle [12, 13] the principal can restrict attention to schedules under which all types report truthfully,[12] so

$$U(x(\theta), \theta) \geq U(x(\hat{\theta}), \theta) \tag{3.4}$$

for all $\theta, \hat{\theta} \in \Theta$. The incentive-compatibility condition (3.4) is equivalent to (3.1)–(3.3). To prove this, we first show that (3.4) implies (3.1)–(3.3). Indeed, by subtracting $U(x(\hat{\theta}), \hat{\theta})$ from (3.4) and switching the labels for θ and $\hat{\theta}$ we obtain that

$$U(x(\theta), \theta) - U(x(\hat{\theta}), \hat{\theta}) \geq U(x(\hat{\theta}), \theta) - U(x(\hat{\theta}), \hat{\theta})$$

and

$$U(x(\hat{\theta}), \hat{\theta}) - U(x(\theta), \theta) \geq U(x(\theta), \hat{\theta}) - U(x(\theta), \theta).$$

Combining the last two inequalities yields

$$U(x(\theta), \theta) - U(x(\theta), \hat{\theta}) \geq U(x(\theta), \theta) - U(x(\hat{\theta}), \hat{\theta}) \geq U(x(\hat{\theta}), \theta) - U(x(\hat{\theta}), \hat{\theta}) \tag{3.5}$$

for all $\theta, \hat{\theta} \in \Theta$. Selecting any $\hat{\theta}, \theta \in \operatorname{int} \Theta$ and taking the limit for $\hat{\theta} \to \theta^+$ and $\theta \to \theta^-$ we get

$$\frac{dU(x(\theta), \theta)}{d\theta} = U_\theta(x(\theta), \theta)$$

almost everywhere (a.e.) on Θ, which is equivalent to

$$U_x(x(\theta), \theta)\, \dot{x}(\theta) = 0. \tag{3.6}$$

The last equation defines the system dynamics $\dot{x}(\theta) = \Phi(x(\theta), \theta)\, u(\theta)$ on Θ_0, as specified in (3.1) and (3.3). Note that $\dot{x}_0(\theta) = \varphi(x(\theta), \theta)$ by solving (3.6) for $\dot{x}_0$. The fact that

$$\varphi(x, \theta) = (\varphi_1(x, \theta), \dots, \varphi_n(x, \theta)) \geq 0$$

(componentwise) follows directly from A1. Since by admissibility of the menu x and smoothness of U, the function $U(x(\theta), \theta)$ is absolutely continuous in θ, by the fundamental theorem of calculus [14, p. 134] it is

$$U(x(\hat{\theta}), \hat{\theta}) - U(x(\theta), \theta) = \int_\theta^{\hat{\theta}} U_\theta(x(\vartheta), \vartheta)\, d\vartheta. \tag{3.7}$$

The fundamental theorem of calculus also implies that

$$U(x(\hat{\theta}), \hat{\theta}) - U(x(\hat{\theta}), \theta) = \int_\theta^{\hat{\theta}} U_\theta(x(\hat{\theta}), \vartheta)\, d\vartheta, \tag{3.8}$$

so by subtracting (3.7) from (3.8) one can rewrite the first inequality in (3.5) in the form

$$\int_\theta^{\hat{\theta}} \Big[U_\theta(x(\hat{\theta}), \vartheta) - U_\theta(x(\vartheta), \vartheta) \Big]\, d\vartheta \geq 0$$

[12] The revelation principle describes the (almost trivial) fact that when the principal is able to commit to a mechanism (i.e., an allocation function $\hat{x} : \Theta \to \mathcal{X}$) by solving the agent's problem (2.2) of announcing a type $\vartheta = \vartheta^*(\theta) \in \Theta_0$ that maximizes his expected utility, she can simply choose x, with $x(\theta) \mapsto \hat{x}(\vartheta^*(\theta))$, instead of $\hat{x}$ to obtain a mechanism that reveals the agent's type truthfully.

for all $\theta, \hat{\theta} \in \Theta$. This last inequality holds for all $\theta, \hat{\theta} \in \Theta$ if and only if

$$\begin{cases} U_\theta(x(\hat{\theta}), \theta) \geq U_\theta(x(\theta), \theta), & \text{if } \hat{\theta} > \theta, \\ U_\theta(x(\hat{\theta}), \theta) \leq U_\theta(x(\theta), \theta), & \text{if } \hat{\theta} < \theta, \end{cases}$$

i.e., if and only if the single-crossing condition (3.2) is satisfied for all $\theta \in \Theta_0$. Hence, condition (3.4) is necessary for conditions (3.1)–(3.3) in Theorem 1. In order to show that conditions (3.1)–(3.3) are also sufficient for the incentive compatibility constraint (3.4), consider any $\hat{\theta}, \theta \in \Theta$ and rewrite (3.4) using the integral representation (3.7) and the fundamental theorem of calculus as

$$U(x(\theta), \theta) - U(x(\hat{\theta}), \hat{\theta}) = \int_{\hat{\theta}}^{\theta} U_\theta(x(\vartheta), \vartheta)\, d\vartheta \geq \int_{\hat{\theta}}^{\theta} U_\theta(x(\hat{\theta}), \vartheta)\, d\vartheta = U(x(\hat{\theta}), \theta) - U(x(\hat{\theta}), \hat{\theta}),$$

or in the more compact form

$$\int_{\theta}^{\hat{\theta}} \Big[U_\theta(x(\hat{\theta}), \vartheta) - U_\theta(x(\vartheta), \vartheta) \Big]\, d\vartheta \geq 0.$$

This last inequality holds for all $\hat{\theta}, \theta \in \Theta$ if and only if $U_\theta(x, \theta)$ exhibits the single-crossing property (3.2) for all $\theta \in \Theta$, which yields the desired sufficiency of (3.1)–(3.3) for (3.4). As alluded to earlier, the initial condition in (3.1) is due to the agent's participation constraint, $U(x(\theta), \theta) \geq r(\theta)$ for all $\theta \in \Theta_0$. By A2, the set of participating types Θ_0 is nonempty (given that $\mathcal{X} \neq \emptyset$). □

The intuition for the system dynamics in (3.1) is that, except for the component x_0 (which contains the agent's payment in the numéraire good), the rate of change of the principal's schedule x as a function of the type θ is governed by the control variable u. The dynamics of the system are constrained by the fact that a participating agent $\theta \in [\theta_0, 1]$ needs to find it optimal to report his type truthfully, such that $U(x(\theta), \theta) \geq U(x(\hat{\theta}), \theta)$. In other words, by consuming a bundle $x(\hat{\theta})$ possibly different from the bundle $x(\theta)$ designed for him, agent θ cannot be better off. This incentive compatibility (or "implementability") is responsible for the dynamics of x_0 and the constraint (3.2). The latter constraint renders problem (P) a nonstandard optimal control problem,[13] for it involves the schedule x generically at different points of the type space. Hence, the corresponding necessary optimality conditions, summarized by the following result, differ from standard versions of the maximum principle. The existence of a solution to (P) is implied by [11, Thm. 2]. We now provide a set of necessary optimality conditions.

Theorem 2. *Given that assumption T1 is satisfied, let (x^*, u^*, θ_0^*) be an optimal solution to the screening problem* (P) *and let $\Theta_0^* = [\theta_0^*, 1]$ be the corresponding set of participating types. Then there exists a function $\psi = (\psi_0, \ldots, \psi_n) : \Theta_0^* \to \mathbb{R}^{n+1}$ of bounded variation, a constant $\lambda_0 \geq 0$, and a nonnegative Borel measure ν, such that the following optimality conditions are satisfied.*

C1. Adjoint Equation: $\psi(\theta) \in \mathbb{R}^{n+1}$ satisfies

$$\psi(\theta) = \int_{\theta}^{1} (\lambda_0 V_x(x^*, \vartheta) f(\vartheta) + \psi_0 \varphi_x(x^*, \vartheta) u^*)\, d\vartheta - \int_{\theta}^{1} (\rho(x^*, \vartheta) - \vartheta) U_{x\theta}(x^*, \vartheta)\, d\nu,$$

for all $\theta \in \Theta_0^*$, where the measurable selection ρ satisfies[14]

$$\rho(\bar{x}, \theta) \in \arg\min_{\hat{\theta} \in \Theta_0^*} \Big\{ (\hat{\theta} - \theta) \Big(U_\theta(x^*(\hat{\theta}), \theta) - U_\theta(\bar{x}, \theta) \Big) \Big\},$$

[13] In the remainder of the text, references to problem (P) (or to its relaxation (P') introduced below) include the complete problem setting with assumptions and relevant constraints.

[14] A set-valued minimizer of a continuous function over a compact set is nonempty, compact-valued, and upper semi-continuous in θ. It therefore has a measurable selection [15, p. 44].

for any $(\bar{x}, \theta) \in \mathbb{R}^{n+1} \times \Theta_0^*$.

C2. Maximality:

$$u_i^*(\theta) \neq 0 \quad \Rightarrow \quad \psi_0(\theta)\varphi_i(x^*(\theta), \theta) + \psi_i(\theta) = 0,$$

a.e. on Θ_0^*, *for all* $i \in \{1, \ldots, n\}$.

C3. Transversality:

$$\textit{(i)}\ \ \theta_0^* = 0 \quad \Rightarrow \quad \exists\, \lambda_1 \in \mathbb{R} :\ \psi^*(0) + \lambda_1 U_x(x^*(0), 0) = 0;$$

$$\textit{(ii)}\ \ \theta_0^* > 0 \quad \Rightarrow \quad \psi(\theta_0^*)\left(U_\theta(x^*(\theta_0^*), \theta_0^*) - \dot{r}(\theta_0^*)\right) + \lambda_0 V(x^*(\theta_0^*), \theta_0^*) U_x(x^*(\theta_0^*), \theta_0^*) = 0.$$

C4. Complementary Slackness:

$$\textit{supp}(\nu) \subseteq \left\{ \theta \in \Theta : \min_{\hat{\theta} \in \Theta_0^*} \left\{ (\hat{\theta} - \theta) \left(U_\theta(x^*(\hat{\theta}), \theta) - U_\theta(x^*(\theta), \theta) \right) \right\} = 0 \right\}.$$

C5. Nontriviality: $\lambda_0 + \sup_{\theta \in \Theta_0^*} \|\psi(\theta)\| > 0$.

P r o o f. See appendix. □

The adjoint variable $\psi(\theta)$ corresponds to the shadow value of the optimal attribute schedule at type θ, given that the evolution of the optimal schedule satisfies the ordinary differential equation (ODE) in (3.1) as well as the incentive-compatibility constraint (3.2). The adjoint equation C1 describes the evolution of ψ on the set of participating types. In particular, the shadow value of the attribute schedule vanishes for the highest type, $\theta = 1$. The maximality condition C2 requires that the optimal control u^* can be essentially bounded only if, while maximizing the principal's expected payoff, the gradient of Φu with respect to u (corresponding to the right-hand side of (3.1)) vanishes. The transversality condition C3 is implied by the optimality of the marginal agent of type θ_0^*, who is indifferent between participating or not. If θ_0^* is not a boundary solution (i.e., when $\theta_0^* > 0$), condition C3 (ii) means that the total change in value, as measured by the agent's marginal surplus through the change in his type movement (evaluated at the shadow value $\psi(\theta_0^*)$) plus his marginal utility for the attribute bundle (evaluated at the principal's net payoff $V(x^*(\theta_0^*), \theta_0^*)$) must vanish for the indifferent type θ_0^*. If it is optimal to serve all agents (i.e., when $\theta_0^* = 0$), one obtains a distortion of the shadow values described by the transversality condition C3 (i). The complementary slackness condition C4 shows that the support of the measure ν must be inside the set of types for which the incentive-compatibility constraint (3.2) is binding. Thus, if (3.2) is never binding, then ν vanishes. Condition C5 ensures that the necessary optimality conditions are nontrivial in the sense that λ_0 and ν cannot vanish together. This can imply an important simplification: if $\nu = 0$, then necessarily $\lambda_0 > 0$, so that, without any loss in generality, we can set $\lambda_0 = 1$, since all other optimality conditions are positively homogeneous in λ_0.

It is generally difficult to construct a measure ν without the precise knowledge of x^*, since ρ is defined using x^* and θ_0^*. When taking the limit $\hat{\theta} \to \theta$ in the maximand of the left-hand side in (3.2), the following constraint is implied:

$$\gamma \cdot u \geq 0, \tag{3.2'}$$

where $\gamma = (1, \gamma_2, \ldots, \gamma_n)$ and $\gamma_j = U_{x_j\theta}/U_{x_1\theta}$ corresponds to the agent's marginal rate of substitution (with respect to U_θ) between x_1 and x_j for $j \in \{2, \ldots, n\}$. We refer to (P) with constraint (3.2) replaced by (3.2') therefore as the *relaxed screening problem* (P'). Contrary to the screening problem (P), the relaxed screening problem (P') can be solved explicitly.

Theorem 3. *Under assumptions P1, P2, and A1–A3, if* (x^*, u^*, θ_0^*) *is an optimal solution to the relaxed screening problem* (P')*, then there exists an absolutely continuous function* $\psi : \Theta \to \mathbb{R}^{n+1}$ *such that the transversality condition C3 is satisfied, and*[15]

$$-\dot{\psi} = V_x(x^*, \theta) f(\theta) + (\psi_0 \varphi_x(x^*, \theta) + \mu \gamma_x(x^*, \theta))\, u^*, \quad \psi(1) = 0, \tag{3.9}$$

and

$$[\psi_0 \varphi_x + \mu \gamma_x, \Phi] u^* = \psi_0 \varphi_\theta + \mu \gamma_\theta - f V_x \Phi, \tag{3.10}$$

where $\mu = \psi_0 \varphi_1(x^*, \theta) + \psi_1$*, and* $\mu \gamma \cdot u^* = 0$*.*

P r o o f. Apply the optimality conditions in Theorem 2 to the optimal control problem (P'). For this, consider the Hamiltonian $H = \lambda_0 V f + \psi \cdot \Phi u$, and first examine the case where the constraint (3.2') is not binding, i.e., where $\gamma \cdot u^* > 0$. If the optimal control u^* is "proper," i.e., independent of $\bar{u}$, then the maximality condition with respect to the relaxed screening problem (P') implies that $H_u = \psi \Phi = 0$ on an optimal state-control trajectory. Differentiating both sides with respect to θ, taking into account the adjoint equation, yields

$$-\dot{\psi}_i = -\left(V_{x_0} f + \psi_0 \varphi_{x_0} \cdot u^*\right) \varphi_i + \psi_0 \varphi_{i,x} \Phi u^* + \psi_0 \varphi_{i,\theta} = V_{x_i} f + \psi_0 \varphi_{x_i} \cdot u^* \tag{3.11}$$

for any $i \in \{1, \ldots, n\}$. These n equations can be rewritten more compactly in the form

$$[\psi_0 \varphi_x, \Phi]\, u^* = \psi_0 \varphi_\theta - f V_x \Phi. \tag{3.12}$$

In the case where $\gamma \cdot u^* = 0$, maximization of the Hamiltonian subject to (3.2') yields the optimality condition $\psi \Phi = \mu \gamma$, where $\mu \geq 0$ is the corresponding Lagrange multiplier. The first component of this condition yields that $\mu = \psi_0 \varphi_1(x^*, \theta) + \psi_1$ as claimed. □

The Lagrange multiplier μ is associated with the relaxed incentive-compatibility constraint (3.2'). By eliminating u^* from the above relations and from (3.1), a solution to the relaxed screening problem can therefore be obtained by solving a system of $n + 3$ ODEs.

Corollary 1. *Let assumptions P1, P2, and A1–A3 be satisfied, and let* (x^*, u^*, θ_0^*) *be an optimal solution to the relaxed screening problem* (P')*. (i) If the matrix* $R = [\varphi_x, \Phi]$ *is nonsingular at the optimal solution, then there exist absolutely continuous functions* $\psi_0, \psi_1 : \Theta_0^* \to \mathbb{R}$ *such that*

$$\left\{ \begin{array}{rcl} \dot{x}^* & = & \Phi R^{-1} \left(\varphi_\theta - \frac{f V_x \Phi}{\psi_0} \right), \\ -\dot{\psi}_0 & = & V_{x_0} f + \varphi_{x_0} R^{-1} \left(\psi_0 \varphi_\theta - f V_x \Phi \right), \end{array} \right. \tag{3.13}$$

provided that $\gamma \cdot R^{-1} \left(\varphi_\theta - \frac{f V_x \Phi}{\psi_0} \right) > 0$*. (ii) Otherwise, setting* $\hat{\gamma} = (\gamma_2, \ldots, \gamma_n)$ *and* $\hat{u} = (u_2, \ldots, u_n)$*), there exist absolutely continuous functions* $\psi_0, \psi_1 : \Theta_0^* \to \mathbb{R}$ *such that*

$$\left\{ \begin{array}{rcl} \dot{x}^* & = & \hat{\Phi} \left[\psi_0 \hat{\varphi}_x + \psi_1 \hat{\gamma}_x, \hat{\Phi} \right]^{-1} \left(\psi_0 \hat{\varphi}_\theta + \psi_1 \hat{\gamma}_\theta - f V_x \hat{\Phi} \right), \\ -\dot{\psi}_i & = & V_{x_i} f + \left(\psi_0 \hat{\varphi}_{x_i} + \psi_1 \hat{\gamma}_{x_i} \right) \left[\psi_0 \hat{\varphi}_x + \psi_1 \hat{\gamma}_x, \hat{\Phi} \right]^{-1} \left(\psi_0 \hat{\varphi}_\theta + \psi_1 \hat{\gamma}_\theta - f V_x \hat{\Phi} \right), \end{array} \right. \tag{3.14}$$

for $i \in \{0, 1\}$*. (iii) The boundary conditions for both (3.13) and (3.14) are* $U(x^*(\theta_0^*), \theta_0^*) = r(\theta_0^*)$ *and* $\psi(1) = 0$*.*

[15] Given two matrices $A \in \mathbb{R}^{m \times l}$ and $B \in \mathbb{R}^{l \times m}$, where m, l are positive integers, the *Lie bracket* of A and B is given by $[A, B] = AB - (AB)^T = AB - B^T A^T \in \mathbb{R}^{m \times m}$, where $(\cdot)^T$ denotes the transpose of $(\cdot)$.

To obtain the $n+2$ initial values $x(\theta_0^*)$ and θ_0^* for the system (3.13), one can use the $n+1$ transversality conditions C3 in conjunction with the initial condition in (3.1).

Remark 1. In the case where all payoff functions are quasilinear in the numéraire, it is $\varphi_{x_0} = 0$ and $V_{x_0} = 1$, which implies that $\psi_0(\theta) = 1 - F(\theta)$ by virtue of (3.13), provided that

$$\left(1, \frac{U_{x_2\theta}}{U_{x_1\theta}}, \ldots, \frac{U_{x_n\theta}}{U_{x_1\theta}}\right) \cdot [\varphi_x, \Phi]^{-1} \left(\varphi_\theta - \frac{f}{1-F} V_x \Phi\right) \geq 0. \tag{3.15}$$

The last inequality can be checked *ex ante*. It generalizes the standard Spence—Mirrlees sorting condition [16, Eq. (7), p. 155]. Note that the last inequality (3.15) features all primitives of the problem, and it is satisfied if the relaxed incentive-compatibility constraint (3.2') is not binding.

Remark 2. In the case where $R = 0$, as in the optimal insurance example discussed in Section 4, relation (3.10) immediately implies that

$$\psi_0 \varphi_\theta + \mu \gamma_\theta = f V_x \Phi. \tag{3.16}$$

Solving this n-dimensional optimality condition, one can determine real-valued functions q_i, such that

$$x_i = q_i(x_0, \theta, \psi_0), \quad i \in \{1, \ldots, n\}, \tag{3.17}$$

which then allows for the solution of the Hamiltonian system consisting of the state equation in (3.1) and the adjoint equation (3.9).

If the solution to the relaxed screening problem (P') is feasible in the screening problem (P), then it is also a solution to the principal's screening problem. Failing that, we can approximate an optimal solution to (P) by solutions to appropriate relaxed problems. For this we introduce a sequence of relaxed screening problems $\{(P'_{kl})\}_{k,l\geq 1}$. For any $k, l \geq 1$, problem (P'_{kl}) is identical to problem (P') with V replaced by

$$V^{kl}(x, \theta) = V(x, \theta) - k \left(g_-(x, \theta; x^{k-1,l-1})\right)^2 - l \left\|x - x^{k-1,l-1}\right\|^2,$$

and

$$g_-(\bar{x}, \theta; x^{k-1,l-1}) = \min_{\hat{\theta} \in [\theta_0^k, 1]} \left\{0, (\hat{\theta} - \theta) \left(U_\theta(x^{k-1,l-1}(\hat{\theta}), \theta) - U_\theta(\bar{x}, \theta)\right)\right\}$$

for any $(\bar{x}, \theta) \in \mathbb{R}^{n+1} \times \Theta$. We denote a solution to problem (P'_{kl}) by $(x^{kl}, u^{kl}, \theta_0^{kl})$. To initialize the sequence of relaxed screening problem, we set $V^{00} = V$, thus adding a problem (P'_{00}) which is identical to problem (P'); its solution $(x^{00}, u^{00}, \theta_0^{00})$ is therefore described by our earlier results.

Theorem 4. *(i) For any given $k \geq 1$, the sequence $\left\{(x^{kl}, u^{kl}, \theta_0^{kl})\right\}_{l\geq 1}$ of solutions to (P'_{kl}) converges to $(x^k, u^k, \theta_0^k) \in \mathbf{W}_{1,\infty} \times \mathbf{L}_\infty \times \Theta$. (ii) The sequence $\left\{(x^k, u^k, \theta_0^k)\right\}_{k\geq 1}$ converges to a solution (x^*, u^*, θ_0^*) of* (P).

P r o o f. (i) Fix any $k \geq 1$. The sequence $\{x^{kl}\}_{l\geq 1}$ is by construction a Cauchy sequence in the Banach space $\mathbf{L}_2(\Theta)$, and therefore converges strongly. This implies weak convergence of the sequence $\{u^{kl}\}_{l\geq 1}$. Lastly, by the Bolzano—Weierstrass theorem the sequence $\{\theta_0^{kl}\}_{l\geq 1} \subset \Theta$ contains a convergent subsequence with limit in Θ. Consider the limits $\theta_0^k \leq \hat{\theta}_0^k$ of any two such convergent subsequences. Then $U(x^k(\theta_0^k), \theta_0^k) = r(\theta_0^k)$ and $U(x^k(\hat{\theta}_0^k), \hat{\theta}_0^k) = r(\hat{\theta}_0^k)$. On the other hand, as already noted in the proof of Theorem 1, by A1 for any $\theta \in [\theta_0^k, 1]$ it is $U(x(\theta), \theta_0^k) - r(\theta_0^k) \leq$

$U(x(\theta),\hat{\theta}_0^k)-r(\hat{\theta}_0^k)$, so necessarily $U(x(\theta),\theta)-r(\theta)=0$ for all $\theta\in[\theta_0^k,\hat{\theta}_0^k]$. Without loss of generality, we can therefore take $\theta_0^k=\liminf_{l\to\infty}\theta_0^{kl}$. The limit (x^k,u^k,θ_0^k) solves the relaxed screening problem (P') with V replaced by $V^k(x,\theta)=V(x,\theta)-k\left(g_-(x,\theta;x^k)\right)^2$. (ii) The convergence of the sequence $\{\theta_0^k\}_{k\geq 1}\subset\Theta$ to $\theta_0^*\in\Theta$ obtains as in part (i). Furthermore, any limit (x^*,u^*,θ_0^*) of the sequence $\left\{(x^k,u^k,\theta_0^k)\right\}_{k\geq 1}$ satisfies $g_-(x^*(\theta),\theta;x^*)=0$ a.e. on Θ_0^*, and is thus a feasible solution to the screening problem (P), provided that $(x^*,u^*)\in\mathbf{W}_{1,\infty}\times\mathbf{L}_\infty$. The latter follows from the existence of a solution to (P). □

4. Application: Optimal Insurance

Consider the problem of designing a nonlinear insurance contract with multiple contingencies, which dates back at least to Stiglitz [17]. An agent has constant absolute risk aversion $\theta\in\Theta=[0,1]$.[16] The type parameter θ belongs to the agent's private information, and—from the principal's point of view—it is distributed with the differentiable probability density $f>0$ on the type space Θ.[17] The agent's utility for any real-valued monetary payoff ξ is

$$v(\xi,\theta)=-\exp\left(-\theta\xi\right),$$

and his current wealth is zero. The agent faces n possible, mutually exclusive loss events $L_1,\ldots,L_n$, which are ordered by magnitude such that $0<L_1<\cdots<L_n<\infty$. The probability of loss event L_i is $p_i>0$, so $p_0=1-(p_1+\cdots+p_n)\in(0,1)$ is the probability that no loss event occurs.[18] An insurance contract subsidizes the agent by an amount x_i when loss event L_i occurs, and asks the agent for a net payment of x_0 in the absence of a loss. Given an insurance contract $x=(x_0,\ldots,x_n)$, the agent's expected utility is

$$U(x,\theta)=p_0v(-x_0,\theta)+\sum_{i=1}^n p_iv(x_i-L_i,\theta).$$

The agent's reservation utility without contract is therefore

$$r(\theta)=U(0,\theta)=-p_0+\sum_{i=1}^n p_iv(-L_i,\theta).$$

On the other hand, the principal's expected payoff is

$$V(x,\theta)=p_0x_0-\sum_{i=1}^n p_ix_i,$$

independent of θ. It is straightforward to verify that both the agent and the principal have utility functions which satisfy assumptions A1–A3 and P1–P2, respectively. The principal's equivalent variation in (2.4) is

$$E(x_1,\ldots,x_n,\theta)=\sum_{i=1}^n\left(\frac{p_i}{p_0}\right)x_i,$$

while the agent's compensating variation in (2.5) takes the form

$$C(x_1,\ldots,x_n,\theta)=\left(\frac{1}{\theta}\right)\ln\left[1+\sum_{i=1}^n\left(\frac{p_i}{p_0}\right)e^{\theta L_i}\left(1-e^{-\theta x_i}\right)\right].$$

[16] By a change of units (i.e., renormalization) this is without loss of generality; cf. also footnote 2.

[17] To satisfy assumption T1, the agent's risk aversion needs to be strictly positive. By considering $\theta/\bar{\theta}$ instead of θ, the analysis generalizes to positive risk aversions in $[0,\bar{\theta}]$, for any $\bar{\theta}>0$.

[18] The vector $(p_0,\ldots,p_n)$ lies in the interior of the n-simplex $\Delta_n=\{(\hat{p}_0,\ldots,\hat{p}_n)\in\mathbb{R}_{++}^{n+1}:\hat{p}_0+\cdots+\hat{p}_n=1\}$.

Thus, assumption T1 is satisfied as long as $f(\theta) = 0$ on $[0, \varepsilon]$ for some minimum risk-aversion $\varepsilon \in (0, 1)$. When the agent is risk-neutral, then the principal would be willing to offer an actuarially fair contract to the agent. By continuous completion for $\varepsilon \to 0^+$ (or, alternately, by imposing a very small capital cost on the principal), we can include risk-neutral agents who are then offered a zero contract. Using the optimality conditions in Theorem 2 and Corollary 1, we find that the Lie product $R = [\varphi_x, \Phi]$ vanishes identically, so by Remark 3:

$$(x_0 - y_i)\,\psi_0 = \frac{f(p_0\varphi_i - p_i) - \mu\gamma_{i\theta}}{-\varphi_i} = \frac{\big(1 - e^{-\theta(x_0 - y_i)}\big)p_0 f + (p_0/p_i)\mu\gamma_{i\theta}}{e^{-\theta(x_0-y_i)}}, \quad i \in \{1, \dots, n\}, \quad (4.18)$$

where $\varphi_i = -(p_i/p_0)\,v(x_0 - y_i, \theta)$ and $\gamma_{i\theta} = -(p_i/p_1)(y_i/y_1)(y_1 - y_i)v(y_1 - y_i, \theta)$ using the abbreviation $y_i \equiv L_i - x_i \geq 0$.

Full Coverage. One immediate solution of (4.18) is $y_i = L_i - x_i = x_0$, for all $i \in \{1, \dots, n\}$, which yields *full coverage* for the participating agent types.[19] The lowest participating type θ_0 in the full-coverage scenario is determined by setting that agent's insurance premium equal to his "certainty equivalent." The latter corresponds to the agent's compensating variation for the insurance contract, so necessarily $x_0 = C(L_1 - x_0, \dots, L_n - x_0, \theta_0)$, which in turn implies that

$$x_0 = g(\theta_0) \equiv \begin{cases} \ln(-r(\theta_0))/\theta_0, & \text{if } \theta_0 \in (0, 1], \\ \bar{L}, & \text{if } \theta_0 = 0, \end{cases}$$

where $\bar{L} = \sum_{i=1}^n p_i L_i$ denotes the agent's expected loss.[20] The principal's expected payoff under full coverage is $\bar{V}(x; \theta_0) = (x_0 - \bar{L})(1 - F(\theta_0))$, so that the optimal participation threshold becomes the global solution of a scalar maximization problem on an interval (for details, see [18]):

$$\theta_0^* \in \arg\max_{\theta_0 \in [0,1]} \left\{(g(\theta_0) - \bar{L})(1 - F(\theta_0))\right\}.$$

As a result, the optimal (constant) schedule is $x^* = (x_0^*, L_1 - x_1^*, \dots, L_n - x_n^*)$, where $x_0^* = g(\theta_0^*)$ and $x_i^* = L_i - x_0^*$ for $i \in \{1, \dots, n\}$. The full-coverage solution leads to no information revelation at all, as all the agent types are offered the same contract. This is also referred to as "bunching" [4].

Partial Coverage. Based on the available optimality conditions it may be possible to construct another solution to the optimal insurance problem, which involves at least partial information revelation. Indeed, for a given $\theta \in (0, 1]$, provided that $\mu = 0$ and $\phi(\theta) \equiv p_0 f(\theta)/\psi_0(\theta) > 1/\theta$, there is a negative solution to (4.18), i.e., there exists a $\zeta = \zeta(\theta) < 0$ such that

$$\zeta = \left(e^{\theta\zeta} - 1\right)\phi. \quad (4.19)$$

In this case, the solution $\zeta = x_0 - y_i = x_0 + x_i - L_i < 0$ is independent of $i \in \{1, \dots, n\}$. For $\phi(\theta) \in [0, 1/\theta]$, the only solution to (4.19) is $\zeta = 0$, reverting back to the full-insurance regime (for that agent type θ). Because by the transversality condition (C1) it is $\psi_0(1) = 0$, this implies that for large enough agent types θ the principal may find it optimal to use partial coverage.

We now continue to follow the solution algorithm outlined in Remark 3, via (3.1) and (3.9). Indeed, the law of motion in (3.1), together with $\dot{x}_i = \dot{\zeta} - \dot{x}_0$ for all $i \in \{1, \dots, n\}$, implies that

$$\dot{x}_0 = \varphi(x, \theta) \cdot (\dot{x}_1, \dots, \dot{x}_n) = \left(\dot{\zeta} - \dot{x}_0\right)\sum_{i=1}^n \varphi_i(x, \theta) = -\frac{\frac{1-p_0}{p_0}\,v(\zeta, \theta)\,\dot{\zeta}}{1 - \frac{1-p_0}{p_0}\,v(\zeta, \theta)} = \frac{\frac{1-p_0}{p_0}\,\dot{\zeta}}{e^{\theta\zeta} + \frac{1-p_0}{p_0}}. \quad (4.20)$$

[19] By the adjoint equation (C1), it is $\psi_0 = (1 - F)p_0$ and $\psi_i = -(1 - F)p_i$ on $[\theta_0^*, 1]$ for all $i \in \{1, \dots, n\}$. (Thus, nontriviality (C5) holds.) Theorem 3 yields $\mu(\theta) = \psi_0(\theta)\varphi_1(x^*(\theta), \theta) + \psi_1(\theta) = 0$ for all $\theta \in [\theta_0^*, 1]$.

[20] By l'Hôpital's rule and the definition of r, it is $\lim_{\theta_0 \to 0^+} \ln(-r(\theta_0))/\theta_0 = r'(0)/r(0) = \bar{L}$.

Using again the law of motion, the first component of the adjoint equation (3.9) becomes

$$-\frac{\dot{\psi}_0}{\psi_0} = \frac{p_0 f}{\psi_0} + (\varphi_x(x,\theta)u)_0 = \phi + \frac{\theta v(\zeta,\theta)}{p_0}\sum_{i=1}^{n} p_i u_i = \phi - \theta\dot{x}_0. \tag{4.21}$$

Since on the one hand $\dot{\psi}_0/\psi_0 = (\dot{f}/f) - (\dot{\phi}/\phi)$, by the definition of ϕ, and since on the other hand

$$\frac{\dot{\phi}}{\phi} = \left(1 - \theta\phi e^{\theta\zeta}\right)\frac{\dot{\zeta}}{\zeta} + \frac{\zeta e^{\theta\zeta}}{1-e^{\theta\zeta}},$$

by virtue of (4.19), one obtains—taking account of (4.20)—that relation (4.21) is equivalent to

$$-\frac{\dot{f}}{f} = -\frac{\dot{\phi}}{\phi} + \phi - \theta\dot{x}_0 = -\left(1 + \frac{\theta\zeta\, e^{\theta\zeta}}{1-e^{\theta\zeta}}\right)\frac{\dot{\zeta}}{\zeta} - \frac{\zeta}{1-e^{\theta\zeta}} - \frac{\theta\dot{\zeta}}{1+\frac{p_0}{1-p_0}e^{\theta\zeta}}.$$

But the last equation implies an initial-value problem,

$$\dot{\zeta} = \left[\frac{1}{\zeta} + \frac{\theta e^{\theta\zeta}}{1-e^{\theta\zeta}} + \frac{(1-p_0)\theta}{(1-p_0)+p_0 e^{\theta\zeta}}\right]^{-1}\left(\frac{\dot{f}}{f} - \frac{\zeta}{1-e^{\theta\zeta}}\right), \quad \zeta(\theta_0) = \zeta_0, \tag{4.22}$$

where the initial value is equal to the certainty equivalent of the marginal agent's exposure conditional on a loss,[21]

$$\zeta_0 = -\frac{\ln\left(\sum_{i=1}^{n}\frac{p_i e^{\theta L_i}}{1-p_0}\right)}{\theta} < 0,$$

thus rendering the contract worthless for the type θ_0, and consequently: $x_0(\theta_0) = 0$.

5. Conclusion

The solution to multiattribute screening problems with one-dimensional types and type-dependent participation constraints can be obtained using optimality conditions derived from a nonstandard version of Pontryagin's maximum principle.[22] Contrary to the extant literature on screening, we do not assume representations of preferences that are quasilinear in the numéraire attribute, thus allowing for arbitrary risk profiles. We also do not require payoff functions to be supermodular in all nonmonetary attributes but impose incentive compatibility as a nonlocal constraint. We have shown that a solution to the multiattribute screening problem, in the case where the incentive-compatibility constraint is binding, can be obtained by solving a sequence of relaxed screening problems, in which constraint violations are increasingly penalized. The results depend in essence only on the convexity of the participation set[23] and on (local) incentive regularity (where needed). Thus, even if A1–A2 and P1–P2 are not satisfied everywhere, one may use our results to construct solutions and then verify the assumptions in a neighborhood of the obtained solutions *ex post*. We further show that the dimensionality of the $2(n+1)$-dimensional Hamiltonian system can often be reduced significantly in concrete problems (e.g., to a 1-dimensional differential equation in our optimal-insurance application in Section 4).

[21] The agent's expected utility is $U(x,\theta) = p_0 v(-x_0,\theta) + (1-p_0)v(\zeta - x_0,\theta) = \left(p_0 + (1-p_0)e^{-\rho\zeta}\right)v(-x_0,\theta)$.

[22] Type-dependent participation constraints arise when agents have heterogeneous outside options, as illustrated in Section 4.

[23] If $U - r$ is quasiconcave in θ (instead of nondecreasing) the participation set is of the form $[\theta_0,\theta_1]$ and θ_1 becomes subject to optimization, leading to an additional transversality condition analogous to C3.

Acknowledgements

The author would like to thank Sergey Aseev and Arkardy Kryazhimskiy, as well as participants of the 2015 INFORMS Annual Meeting in Philadelphia and the 5th International Conference on Continuous Optimization (ICCOPT 2016) in Tokyo, for helpful discussions on related research. A number of very useful comments by an anonymous referee, Ksenia Melnikova, and participants of the 2016 International Conference on "System Analysis: Modeling and Control" in Yekaterinburg are gratefully acknowledged. This paper is dedicated to the memory of Academician Arkady Kryazhimskiy, an outstanding researcher and very kind individual, who will be missed.

REFERENCES

1. **Mirrlees, J.A.** An exploration in the theory of optimal income taxation // Rev. Econ. Stud. 1971. Vol. 38, no. 2, P. 175–208.
2. **Baron, D.P., Myerson, R.B.** Regulating a monopolist with unknown costs // Econometrica. 1982. Vol. 50, no. 4, P. 911–930.
3. **Mirman, L.J., Sibley, D.** Optimal nonlinear prices for multiproduct monopolies // Bell J. Econ. 1980. Vol. 11, no. 2, P. 659–670.
4. **Mussa, M., Rosen, S.** Monopoly and product quality // J. Econ. Theory. 1978. Vol. 18, no. 2, P. 301–317.
5. **Guesnerie, R., Laffont, J.-J.** A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm // J. Public Econ. 1984. Vol. 25, no. 3, P. 329–369.
6. **Matthews, S., Moore, J.** Monopoly provision of quality and warranties: an exploration in the theory of multidimensional screening // Econometrica. 1987. Vol. 55, no. 2, P. 441–467.
7. **Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.** The mathematical theory of optimal processes, New York: Wiley Interscience, 1962. 360 p.
8. **Aseev, S.** Methods of regularization in nonsmooth problems of dynamic optimization // J. Math. Sci. 1999. Vol. 94, no. 3, P. 1366–1393.
9. **Hicks, J.R.** Value and capital, Oxford: Clarendon Press, 1939. 331 p.
10. **Weber, T.A.** Hicksian welfare measures and the normative endowment effect // Am. Econ. J.: Microecon. 2010. Vol. 2, no. 4, P. 171–194.
11. **Weber, T.A.** Screening with externalities. Working Paper, Stanford University, 2005.
12. **Gibbard, A.** Manipulation of voting schemes: a general result // Econometrica. 1973. Vol. 41, no. 4, P. 587–601.
13. **Myerson, R.B.** Incentive compatibility and the bargaining problem // Econometrica. 1979. Vol. 47, no. 1, P. 61–74.
14. **Rudin, W.** Principles of mathematical analysis (3rd edition). New York: McGraw-Hill, 1976. 342 p.
15. **Jayne, J.E., Rogers, C.A.** Selectors. Princeton: Princeton University Press, 2002. 167 p.
16. **Laffont, J.-J.** The economics of uncertainty and information. Cambridge: MIT Press, 1989. 289 p.
17. **Stiglitz, J.E.** Monopoly, non-linear pricing and imperfect information: the insurance market // Rev. Econ. Stud. 1977. Vol. 44, no. 3, P. 407–430.
18. **Weber, T.A.** Global optimization on an interval // J. Optimiz. Theory App. 2016. Forthcoming. [doi:10.1007/s10957-016-1006-y]
19. **Arutyunov, A.V.** Optimality conditions: abnormal and degenerate problems. Dordrecht: Kluwer, 2000. 299 p.
20. **Weber, T.A.** Optimal control theory with applications in economics. Cambridge: MIT Press, 2011. 360 p. (Preface by A.V. Kryazhimskiy)
21. **Gelfand, I.M., Fomin, S.V.** Calculus of variations. Englewood-Cliffs: Prentice-Hall, 1963. 240 p.
22. **Kolmogorov, A.N., Fomin, S.V.** Elements of the theory of functions and functional analysis, parts I & II. Rochester: Graylock Press, 1957. 288 p.
23. **Megginson, R.E.** An introduction to Banach space theory. New York: Springer, 1998. 596 p.
24. **Dunford, N., Schwartz, J.T.** Linear operators, part I: general theory. New York: Wiley Interscience, 1958. 858 p.

25. **Milyutin, A.A., Osmolovskii, N.P.** Calculus of variations and optimal control. Providence: American Mathematical Society, 1998. 372 p.
26. **Zorich, V.A.** Mathematical analysis, vol. I. New York: Springer, 2004. 574 p.
27. **Kirillov, A.A., Gvishiani, A.D.** Theorems and problems in functional analysis. New York: Springer, 1982. 347 p.
28. **Taylor, A.E.** General theory of functions and integration. New York: Blaisdell Publishing, 1965. 437 p.
29. **Giaquinta, M., Modica, G., Souček, J.** Cartesian currents in the calculus of variations, vol. I. New York: Springer, 1998. 711 p.
30. **Riesz, F., Sz.-Nagy, B.** Functional analysis, New York: Ungar Publishing, 1955. 491 p.
31. **Schechter, M.** Principles of functional analysis (2nd edition). Providence: American Mathematical Society, 2002. 425 p.

Appendix: Proof of Theorem 2

P r o o f. The argument proceeds in six steps (cf. [7, 8, 19, 20]). For any $(\bar{x}, \theta) \in \mathbb{R}^{n+1} \times \Theta$, let

$$g(\bar{x}, \theta) = \min_{\hat{\theta} \in \Theta_0^*} \left\{ (\hat{\theta} - \theta) \left(U_\theta(x^*(\hat{\theta}), \theta) - U_\theta(\bar{x}, \theta) \right) \right\}.$$

Step 1: Approximate the problem (P) *by a sequence of relaxed problems* $\left\{(\bar{\mathrm{P}}_k)\right\}_{k \geq 1}$.

We approximate the principal's screening problem (P) by a sequence of problems $(\bar{\mathrm{P}}_k)$, $k = 1, 2, \ldots$, in each of which the constraints are relaxed. The sequence of relaxed problems approximates the original problem, since deviations from the constraints and the optimal control u^* are penalized using successively increasing weights. We first fix the positive numbers ε, δ, and

$$\bar{u} \geq 2 + \operatorname*{ess\,sup}_{\theta \in \Theta_0^*} \|u^*(\theta)\|,$$

relative to which the sequence $\{(\bar{\mathrm{P}}_k)\}_{k=1}^{\infty}$ of relaxed problems will be defined. For any $k \geq 1$, let

$$V^k(x, u, \theta) = V(x, \theta) - \delta \|u(\theta) - u^*(\theta)\|^2 - k g_-(x, \theta)^2, \tag{5.1}$$

where $g_- = \min\{g, 0\}$ takes on nonzero (negative) values whenever the relevant constraint (3.2) of the original problem (P) is violated. Similarly, for any $\theta_0 \in \Theta$ and $\underline{x} \in \mathbb{R}_+^{n+1}$ we set

$$K(\underline{x}, \theta_0) = (\min\{\theta_0, 0\})^2 + \left(U(\underline{x}, \theta_0) - r(\theta_0)\right)^2 \tag{5.2}$$

to penalize deviations from the endpoint constraints (including $\theta_0 \geq 0$) imposed in the original problem. We are now ready to formulate the relaxed problem $(\bar{\mathrm{P}}_k)$ for any $k \geq 1$, given an optimal solution $(x^*, u^*, \theta_0^*) \in \mathcal{D}$ to the original problem (P):

$$\left\{ \begin{array}{l} \sup_{(x,u,\theta_0,\underline{x}) \in \hat{\mathcal{D}}} \left\{ \int_{\theta_0}^1 V^k(x(\theta), u(\theta), \theta)\, dF(\theta) - (\underline{x} - \underline{x}^*)^2 - (\theta_0 - \theta_0^*)^2 - k K(\underline{x}, \theta_0) \right\} \\ \text{s.t.} \\ \quad \dot{x} = \Phi u, \quad x(\theta_0) = \underline{x}, \\ \quad \varepsilon \geq \|x - x^*\|_\infty + (\underline{x} - \underline{x}^*)^2 + (\theta_0 - \theta_0^*)^2, \\ \quad u \in \mathcal{U}, \end{array} \right\} \tag{$\bar{\mathrm{P}}_k$}$$

where $\hat{\mathcal{D}} = \mathcal{D} \times \mathbb{R}_+^{n+1}$ is an augmented domain of admissibility, $\mathcal{U} = \{\hat{u} \in \mathbb{R}^n : \|\hat{u}\| \leq \bar{u}\}$ is a control-constraint set, and $\underline{x}^* = x^*(\theta_0^*)$ is the bundle offered to the optimal marginal agent type. We denote a solution to the relaxed problem $(\bar{\mathrm{P}}_k)$ by $(x^k, u^k, \theta_0^k, \underline{x}^k)$. By setting u^k to zero for all $\theta \notin [\theta_0^k, 1]$ we extend any such solution to all types in Θ (containing the optimal interval $[\theta_0^k, 1]$), so the state trajectory x^k is constant on $[0, \theta_0^k]$.

Step 2: Show that each problem $(\bar{\mathrm{P}}_k)$ *has a solution.*
For any $k \geq 1$, there exists a solution $(x^k, u^k, \theta_0^k, \underline{x}^k)$ to the relaxed problem $(\bar{\mathrm{P}}_k)$. Let

$$\{(x^{k,j}, u^{k,j}, \theta_0^{k,j}, \underline{x}^{k,j})\}_{j=1}^{\infty}$$

be an admissible *maximizing sequence* [21, p. 193] for the problem $(\bar{\mathrm{P}}_k)$. Since $u^{k,j}$ takes values in the closed ball of $\mathbb{R}^n$ at 0 of radius $\bar{u}$, and *a forteriori* the contract space (which contains all bundles that can actually be transacted between principal and agent) is bounded by T1,[24] this maximizing sequence is uniformly bounded, which allows the following three conclusions for an appropriate subsequence (and for simplicity we identify our original maximizing sequence with this subsequence by relabelling indices if necessary). First, from the definition of an admissible sequence $\{x^{k,j}\}_{j=1}^{\infty} \subset \mathbf{W}_{1,\infty}$ and the uniform boundedness of $\{\dot{x}^{k,j}\}_{j=1}^{\infty}$ this sequence of state trajectories is equicontinuous, so by the Arzelà—Ascoli theorem [22, Part I, p. 54] it converges uniformly to $\hat{x}^k$. Second, we obtain pointwise convergence of $(\underline{x}^{k,j}, \theta_0^{k,j})$ to $(\hat{\underline{x}}^k, \hat{\theta}_0^k)$ as $j \to \infty$. Third, since the space of admissible controls $\mathbf{L}_\infty$ is a subset of the Hilbert space $\mathbf{L}_2$ (which can be identified with its dual), $u^{k,j}$ converges weakly to $\hat{u}^k$ as $j \to \infty$.[25]

We now show that in fact the above limits coincide with the solution to $(\bar{\mathrm{P}}_k)$, i.e.,

$$(x^k, u^k, \theta_0^k, \underline{x}^k) = (\hat{x}^k, \hat{u}^k, \hat{\theta}_0^k, \hat{\underline{x}}^k). \tag{5.3}$$

For any $\theta \in [\theta_0^k, 1]$ it is

$$x^{k,j}(\theta) = x^{k,j}(\theta_0^{k,j}) + \int_{\theta_0^{k,j}}^{\theta} \Phi(x^{k,j}(\vartheta), \vartheta) u^{k,j}(\vartheta)\, d\vartheta,$$

so by taking the limit for $j \to \infty$:

$$\hat{x}^k(\theta) = \hat{x}^k(\hat{\theta}_0^k) + \int_{\hat{\theta}_0^k}^{\theta} \Phi(\hat{x}^k(\vartheta), \vartheta) \hat{u}^k(\vartheta)\, d\vartheta.$$

Hence, the limiting tuple $(\hat{x}^k, \hat{u}^k, \hat{\theta}_0^k, \hat{\underline{x}}^k)$ is consistent with the Cauchy problem for the state evolution, i.e., it satisfies

$$\dot{\hat{x}}^k = \Phi \hat{u}^k, \quad \hat{x}^k(\hat{\theta}_0^k) = \hat{\underline{x}}^k.$$

The state constraint $\varepsilon \geq \|\hat{x}^k - x^*\|_\infty + \left\|\hat{\underline{x}}^k - \underline{x}^*\right\|^2 + \left(\hat{\theta}_0^k - \theta_0^*\right)^2$ is satisfied by uniform convergence of the maximizing sequence. Lastly, the control constraint $\hat{u}^k \in \mathcal{U}$ is satisfied, since each $u^{k,j}$, $j = 1, 2, \ldots$, is feasible ($\bar{u}$ has been chosen appropriately large). The weak convergence $u^{k,j} \overset{w}{\to} \hat{u}^k$ as $j \to \infty$ implies, by Mazur's compactness theorem [23, p. 254], that there exists a sequence $\{v^{k,j}\}_{j=1}^{\infty}$ with elements in the convex hull $\operatorname{co}\{u^{k,j}\}_{j=1}^{\infty}$ which converges strongly to $\hat{u}^k$ in $\mathbf{L}_2^n(\Theta)$. We therefore obtain that equation (5.3) holds, i.e., the limit point $(\hat{x}^k, \hat{u}^k, \hat{\theta}_0^k, \hat{\underline{x}}^k)$ of the maximizing sequence describes an admissible solution to the relaxed problem $(\bar{\mathrm{P}}_k)$.

Step 3: Show that the solutions of $(\bar{\mathrm{P}}_k)_{k\geq 1}$ *converge to the solution of* (P).
As before, there exists an admissible tuple $(\hat{x}, \hat{u}, \hat{\theta}_0) \in \mathcal{D}$ such that $x^k \rightrightarrows \hat{x}^k$, $u^k \overset{w}{\to} \hat{u}$, and $\theta_0^k \to \hat{\theta}_0^k$. We now show that

$$(\hat{x}, \hat{u}, \hat{\theta}_0) = (x^*, u^*, \theta_0^*), \tag{5.4}$$

i.e., in particular that $x^k \rightrightarrows x^*$, $u^k \overset{w}{\to} u^*$, and $\theta_0^k \to \theta_0^*$.

[24] The assumption T1 is used here to guarantee the boundedness of the first (numéraire) component of the attribute schedule, as its dynamics in (3.1) are not directly governed by the control variable.

[25] By the Banach—Alaoglu theorem the unit ball in L_2 is weakly* (and therefore weakly) compact, so by the Eberlein—Šmulian theorem it is also weakly sequentially compact [23, p. 229/248]. This property of reflexive Banach spaces can also be deduced from the uniform boundedness principle [24, Ch. 2].

Let $\pi^k(x,u,\theta_0,\underline{x}) = \int_{\theta_0}^1 V^k(x(\theta),u(\theta),\theta)\,dF(\theta) - \|\underline{x}-\underline{x}^*\|^2 - (\theta_0-\theta_0^*)^2 - kK(\underline{x},\theta_0)$. As a consequence of the uniform boundedness of the state-control trajectories, there exists a constant $M>0$ such that $M \geq \pi^k(x^k,u^k,\theta_0^k,\underline{x}^k)$ for all k. Since $\pi^k(x^k,u^k,\theta_0^k,\underline{x}^k) - \pi^k(x^*,u^*,\theta_0^*,\underline{x}^*) \geq 0$, it is

$$\frac{M}{k} \geq \int_{\theta_0^k}^1 \left(\left(\frac{\delta}{k}\right)\|u^k-u^*\|^2 + g_-^2\right)\,dF(\theta) + \left(\frac{1}{k}\right)\left[\|\underline{x}^k-\underline{x}^*\|^2 + (\theta_0^k-\theta_0^*)^2\right] + K(\underline{x}^k,\theta_0^k) \geq 0.$$

Taking the limit for $k\to\infty$, by continuity of K it is $K(\hat{\underline{x}},\hat{\theta}_0)=0$, i.e., $(\hat{\underline{x}},\hat{\theta}_0)$ satisfies the endpoint constraints $U(\hat{\underline{x}},\hat{\theta}_0) = r(\hat{\theta}_0)$ and $\hat{\theta}_0 \geq 0$. Moreover,

$$\lim_{k\to\infty}\int_{\theta_0^k}^1 \left(g_-(x^k,\theta)\right)^2\,dF(\theta) = 0.$$

Since the type distribution F has full support Θ (no type can be excluded for sure), it is

$$g_-(x^k,\theta) = 0.$$

Hence, $g(x^k,\theta)=0$, which is equivalent to

$$\begin{cases} \hat{\theta}>\theta & \Rightarrow \quad U_\theta(x^*(\hat{\theta}),\theta) \geq U_\theta(x^k(\theta),\theta), \\ \hat{\theta}\leq\theta & \Rightarrow \quad U_\theta(x^*(\hat{\theta}),\theta) \leq U_\theta(x^k(\theta),\theta), \end{cases}$$

for all $\theta,\hat{\theta}\in\Theta$. Because U is by assumption twice differentiable, and $U_\theta(x,\theta)$ is Lipschitz-continuous in x, there exists a constant $L>0$ such that $|U_\theta(\bar{x},\theta) - U_\theta(\bar{y},\theta)| \leq L\|\bar{x}-\bar{y}\|$ for all $\bar{x},\bar{y}$ in the relevant (by assumption T1 bounded) contract space $\mathcal{X}$ and all $\theta\in\Theta$. For $\hat{\theta}>\theta$ this means that

$$U_\theta(x^k(\hat{\theta}),\theta) + L\|x^*-x^k\| \geq U_\theta(x^k(\hat{\theta}),\theta) + \Big(U_\theta(x^*(\hat{\theta}),\theta) - U_\theta(x^k(\hat{\theta}),\theta)\Big) \geq U_\theta(x^k(\theta),\theta),$$

and, for $\hat{\theta}<\theta$, that

$$U_\theta(x^k(\hat{\theta}),\theta) - L\|x^*-x^k\| \geq U_\theta(x^k(\hat{\theta}),\theta) + \Big(U_\theta(x^*(\hat{\theta}),\theta) - U_\theta(x^k(\hat{\theta}),\theta)\Big) \leq U_\theta(x^k(\theta),\theta).$$

Hence, it is

$$\min_{\hat{\theta}\in\Theta_0^*}\left\{(\hat{\theta}-\theta)\left(U_\theta(x^*(\hat{\theta}),\theta) - U_\theta(x^*(\theta),\theta)\right)\right\} \geq -L\varepsilon,$$

and for $\varepsilon\to 0^+$ the limit $(\hat{x},\hat{u},\hat{\theta}_0,\hat{\underline{x}})$ becomes admissible in problem (P). This implies

$$\pi(x^*,u^*,\theta_0^*,\underline{x}^*) \geq \pi(\hat{x},\hat{u},\hat{\theta}_0,\hat{\underline{x}}). \tag{5.5}$$

On the other hand, $\pi^k(x^k,u^k,\theta_0^k,\underline{x}^k) \geq \pi^k(x^*,u^*,\theta_0^*,\underline{x}^*) =: \pi(x^*,u^*,\theta_0^*,\underline{x}^*) = \bar{V}(x^*;\theta_0^*)$, whence

$$\pi(x^k,u^k,\theta_0^k,\underline{x}^k) - \|\underline{x}^k-\underline{x}^*\|^2 - (\theta_0^k-\theta_0^*)^2 - \delta\int_{\theta_0^k}^1 \|u^k-u^*\|^2\,dF(\theta) \geq \pi(x^*,u^*,\theta_0^*,\underline{x}^*)$$

for all $k\geq 1$. Hence, taking the limit for $k\to\infty$ yields

$$\pi(\hat{x},\hat{u},\hat{\underline{x}},\hat{\theta}_0) - \left(\hat{\theta}_0-\theta_0^*\right)^2 - \delta\lim_{k\to\infty}\int_{\Theta_0}\|u^k-u^*\|^2\,dF(\theta) \geq \pi(x^*,u^*,\theta_0^*,\underline{x}^*),$$

which together with (5.5) implies that $\hat{\theta}_0=\theta_0^*$, $\hat{u}=u^*$, and

$$\lim_{k\to\infty}\int_\Theta \|u^k-u^*\|^2\,dF(\theta) = 0,$$

so the sequence $\{u^k\}_{k=1}^{\infty}$ converges to u^* a.e. on Θ.

Step 4: Show that the problem $(\bar{\mathrm{P}}_k)$ *becomes a standard OCP* $(\bar{\mathrm{P}}'_k)$ *for large* k.
Because of the uniform convergence of the optimal state-trajectories x^k and the pointwise convergence of the boundary points θ_0^k (as $k \to \infty$) to the corresponding optimal state trajectory x^* and optimal boundary point θ_0^* of the original problem (P), the state constraint in the relaxed problem $(\bar{\mathrm{P}}_k)$ is strictly not binding for large enough k, i.e.,

$$\varepsilon > \|x^k - x^*\|_\infty + \left(\theta_0^k - \theta_0^*\right)^2,$$

as long as k is sufficiently large. Hence, for fixed constants ε, δ, and $\bar{u}$ there exists a $k_0 = k_0(\varepsilon, \delta, \bar{u}) \geq 1$ such that for all $k \geq k_0$ the problem $(\bar{\mathrm{P}}_k)$ can be rewritten equivalently in the form

$$\left\{ \begin{array}{l} \sup_{(x,u,\theta_0,\underline{x})\in\hat{\mathcal{D}}} \left\{ \int_{\theta_0}^1 V^k(x(\theta),u(\theta),\theta)\, dF(\theta) - \|\underline{x} - \underline{x}^*\|^2 - (\theta_0 - \theta_0^*)^2 - kK(\underline{x},\theta_0) \right\} \\ \text{s.t.} \\ \quad \dot{x} \;=\; \Phi u, \quad x(\theta_0) \;=\; \underline{x}. \end{array} \right\} \qquad (\bar{\mathrm{P}}'_k)$$

Necessary optimality conditions based on the maximum principle [7] for problem $(\bar{\mathrm{P}}'_k)$ are readily available.

Step 5: Obtain necessary optimality conditions for $(\bar{\mathrm{P}}'_k)$.
We provide here a version of the maximum principle by Milyutin and Osmolovskii [25, P. 24–25].[26] Let

$$H^k(x,u,\theta,\lambda_0^k,\psi^k) = \lambda_0^k V^k f + \psi^k \cdot \Phi u$$

be the Hamiltonian function associated with problem $(\bar{\mathrm{P}}'_k)$, where $\lambda_0^k \in \mathbb{R}$ is a constant multiplier, $\psi^k \in \mathbb{R}^{n+1}$ is an adjoint variable, and x is the state of the system. The Hamiltonian corresponds to the instantaneous payoff to the principal including the current benefit of state velocities. The shadow prices of these are measured by the adjoint variables λ_0^k and ψ^k respectively.

Maximum Principle for Problem $(\bar{\mathrm{P}}'_k)$. *If* $(x^k, u^k, \theta_0^k, \underline{x}^k)$ *is an optimal solution for the problem* $(\bar{\mathrm{P}}'_k)$*, then there exists an absolutely continuous function* $\psi^k : [\theta_0^k, 1] \to \mathbb{R}^{n+1}$*, and a constant* $\lambda_0^k > 0$*, such that the following relations hold.*

1. *Adjoint Equation:*

$$-\dot{\psi}^k(\theta) = H_x^k(x^k(\theta), u^k(\theta), \theta, \lambda_0^k, \psi^k(\theta)) \qquad (5.6)$$

2. *State Transversality:*

$$\psi^k(\theta_0^k) \;=\; 2k\left(U(\underline{x}^k,\theta_0^k) - r(\theta_0^k)\right) U_x(\underline{x}^k,\theta_0^k) + 2\left(\underline{x}^k - \underline{x}^*\right) \qquad (5.7)$$

$$\psi^k(1) \;=\; 0 \qquad (5.8)$$

3. *Maximality:*

$$u^k(\theta) \in \arg\max_{u\in\mathcal{U}} H^k(x^k(\theta), u, \theta, \lambda_0^k, \psi^k(\theta)) \quad \textit{a.e. on } [\theta_0^k, 1] \qquad (5.9)$$

One can extend the standard maximum principle comprised by the above conditions, and obtain the following type-transversality condition.

4. *Type Transversality:*

$$\left\{ \begin{array}{c} 2\left(k\min\{\theta_0^k, 0\} + k\left(U(\underline{x}^k,\theta_0^k) - r(\theta_0^k)\right)\left(U_\theta(\underline{x}^k,\theta_0^k) - \dot{r}(\theta_0^k)\right) + (\theta_0^k - \theta_0^*)\right) \\ + \sup_{u\in\mathcal{U}} H^k(\underline{x}^k, u, \theta_0^k, \lambda_0^k, \psi^k(\theta_0^k)) = 0. \end{array} \right\} \qquad (5.10)$$

[26] They consider a slightly more general Mayer problem. Problem $(\bar{\mathrm{P}}'_k)$ is a Bolza problem on a variable interval, which can be reduced to a Mayer problem (in which the objective function depends only on the endpoints of the state trajectory) by introducing an additional real-valued state variable.

The type-transversality condition is crucial for determining the missing multiplier associated with the optimization of the participation set in the principal's screening problem. To prove type transversality, consider some real τ (with $|\tau|$ sufficiently small), and let

$$\hat{\omega}_0^k(\tau) = (x^k(\theta_0^k + \tau), \theta_0^k + \tau),$$

and

$$\hat{\Pi}^k(\tau) = \int_{\theta_0^k+\tau}^1 \lambda_0^k V^k(x^k(\theta), u^k(\theta), \theta)\, dF(\theta) - \|\hat{\omega}_0^k(\tau) - (\underline{x}^*, \theta_0^*)\|^2 - kK(\hat{\omega}_0^k(\tau)),$$

where if necessary we extend the state-control trajectory beyond the lowest participating type. Then by optimality of θ_0^k,

$$\begin{aligned}
0 &\leq \hat{\Pi}_2^k(0) - \hat{\Pi}_2^k(\tau) \\
&= \int_{\theta_0^k}^{\theta_0^k+\tau} \lambda_0^k V^k(x^k, u^k, \theta)\, dF(\theta) + \left(\|\hat{\omega}_0^k(\tau) - (\underline{x}^*, \theta_0^*)\|^2 - \|(\underline{x}^k - \underline{x}^*, \theta_0^k - \theta_0^*)\|^2\right) \\
&\qquad\qquad + k\left(K(\hat{\omega}_0^k(\tau)) - K(\underline{x}^k, \theta_0^k)\right) \\
&= \int_{\theta_0^k}^{\theta_0^k+\tau} \left(\lambda_0^k V^k f + \left(2\left(x^k(\theta) - \underline{x}^*\right) + kK_{\underline{x}}(x^k(\theta), \theta)\right) \cdot \dot{x}^k(\theta) + 2(\theta - \theta_0^*) + kK_{\theta_0}(x^k(\theta), \theta)\right) d\theta \\
&= \int_{\theta_0^k}^{\theta_0^k+\tau} \left(H^k(x^k, u^k, \theta, \lambda_0^k, \psi^k)\big|_{\psi^k = 2(x^k(\theta) - \underline{x}^*) + kK_{\underline{x}}(x^k(\theta), \theta)} + kK_{\theta_0}(x^k(\theta), \theta)\right) d\theta + 2(\theta_0^k - \theta_0^*)\tau + \tau^2 \\
&\leq \int_{\theta_0^k}^{\theta_0^k+\tau} \left(\sup_{u\in\mathcal{U}} H^k(x^k, u, \theta, \lambda_0^k, 2\left(x^k - \underline{x}^*\right) + kK_{\underline{x}}(x^k, \theta)) + kK_{\theta_0}(x^k(\theta), \theta)\right) d\theta + 2(\theta_0^k - \theta_0^*)\tau + \tau^2.
\end{aligned}$$

By the first mean-value theorem for the integral [26, p. 352], the last integral can be written in the form

$$\left(\sup_{u\in\mathcal{U}} H^k\left(x(\check{\theta}_0^k), u, \check{\theta}_0^k, \lambda_0^k, 2\left(x^k(\check{\theta}_0^k) - \underline{x}^*\right) + kK_{\underline{x}}(x^k(\check{\theta}_0^k), \check{\theta}_0^k)\right) + kK_{\theta_0}(x^k(\check{\theta}_0^k), \check{\theta}_0^k)\right)\tau$$

for some appropriate $\check{\theta}_0^k \in [\theta_0^k, \theta_0^k + \tau]$. Hence, dividing the previous inequality by $\tau > 0$ and subsequently taking the limit for $\tau \to 0^+$ yields

$$0 \leq \sup_{u\in\mathcal{U}} H^k(x^k, u, \theta, \lambda_0^k, 2\left(\underline{x}^k - \underline{x}^*\right) + kK_{\underline{x}}(\underline{x}^k, \theta_0^k)) + kK_{\theta_0}(\underline{x}^k, \theta_0^k) + 2(\theta_0^k - \theta_0^*). \tag{5.11}$$

Consider now the case where we extend the type interval by an increment $\hat{\tau} > 0$ to the left beyond the optimal lowest type θ_0^k, which corresponds to $\tau < 0$ in the above relations. If we set $\hat{\tau} = -\tau > 0$, then it is (using our earlier computations)

$$0 \leq -\int_{\theta_0^k-\hat{\tau}}^{\theta_0^k} \left(H^k(x^k, u^k, \theta, \lambda_0^k, \psi^k)\Big|_{\psi^k = 2\left(x^k(\theta) - \underline{x}^*\right) + kK_{\underline{x}}(x^k(\theta), \theta)} + kK_{\theta_0}(x^k(\theta), \theta)\right) d\theta - 2(\theta_0^k - \theta_0^*)\hat{\tau} + \hat{\tau}^2,$$

where we can extend $u^k(\theta)$, for all $\theta \in [\theta_0^k - \hat{\tau}, \theta_0^k]$ to the left. In particular, the latter extension can be performed such that

$$\rho \geq \left[\sup_{u\in\mathcal{U}} H^k\left(x^k, u, \theta, \lambda_0^k, kK_{\underline{x}}(x^k(\theta), \theta)\right) - H^k\left(x^k, u^k, \theta, \lambda_0^k, kK_{\underline{x}}(x^k(\theta), \theta)\right)\right]$$

on the interval $[\theta_0^k - \hat{\tau}, \theta_0^k]$ for some arbitrary $\rho > 0$. Hence,

$$0 \leq -\left(\sup_{u\in\mathcal{U}} H^k(x^k(\check{\theta}_0^k), u, \check{\theta}_0^k, \lambda_0^k, 2\left(x^k(\check{\theta}_0^k) - \underline{x}^*\right) + kK_{\underline{x}}(x^k(\check{\theta}_0^k), \check{\theta}_0^k)) - \rho + kK_{\theta_0}(x^k(\check{\theta}_0^k), \check{\theta}_0^k) + 2(\theta_0^k - \theta_0^*)\right)\hat{\tau} + \hat{\tau}^2,$$

for some $\check{\theta}_0^k \in [\theta_0^k - \hat{\tau}, \theta_0^k]$, so by dividing through $\hat{\tau} > 0$ and taking the limits for $\hat{\tau} \to 0^+$ and $\rho \to 0^+$, we get

$$0 \geq \sup_{u \in \mathcal{U}} H^k(\underline{x}^k, u, \theta_0^k, \lambda_0^k, 2\left(\underline{x}^k - \underline{x}^*\right) + kK_{\underline{x}}(\underline{x}^k, \theta_0^k)) + kK_{\theta_0}(\underline{x}^k, \theta_0^k) + 2(\theta_0^k - \theta_0^*). \tag{5.12}$$

Relations (5.11) and (5.12), together with (5.7), are equivalent to (5.10).

Using the above extended maximum principle (5.6)–(5.10) for the relaxed problem $(\bar{\mathrm{P}}_k')$, one obtains the adjoint equation

$$-\dot{\psi}^k = \lambda_0^k V_x f + \nu^k g_x + \psi^k \cdot \left(\Phi_x u^k\right) = \lambda_0^k V_x f + \nu^k g_x + \psi_0^k \varphi_x u^k \tag{5.13}$$

on Θ, where

$$\nu^k(\theta) = \begin{cases} -2k\lambda_0^k g_-(x^k, \theta), & \theta \in [\theta_0^k, 1], \\ 0, & \text{otherwise.} \end{cases} \tag{5.14}$$

The maximality condition (5.9) constitutes a constrained optimization problem, for which there exists a Lagrange multiplier $\varsigma^k = \varsigma^k(\theta)$ such that

$$\psi_0 \varphi_i + \psi_i - 2\lambda_0^k \delta(u_i^k - u_i^*) + 2\varsigma^k u_i^k = 0$$

for all $i \in \{1, \ldots, n\}$, with complementary slackness $\varsigma^k \left(\|u^k\| - \bar{u}\right) = 0$. By Step 3 we know that $u^k \to u^*$ a.e. on Θ. Hence, by Egorov's theorem [27, p. 24] for any $\bar{\delta} > 0$ there is a subset $\Theta_{\bar{\delta}}$ of Θ, such that $\int_{\Theta \setminus \Theta_{\bar{\delta}}} d\theta < \bar{\delta}$ and $u^k \rightrightarrows u^*$ uniformly on $\Theta_{\bar{\delta}}$. Since u^* is feasible, this uniform convergence implies that $u^k \notin \partial\mathcal{U}$ on $\Theta_{\bar{\delta}}$ for k large enough. By virtue of complementary slackness, the corresponding Lagrange multipliers ς^k and ρ^k therefore vanish on $\Theta_{\bar{\delta}}$, as long as k is large enough. In other words,

$$\psi_0 \varphi_i + \psi_i - 2\lambda_0^k \delta(u^k - u^*) = 0 \tag{5.15}$$

a.e. on Θ for all $i \in \{1, \ldots, n\}$, as long as k is large enough.

Step 6: Derive necessary optimality conditions for (P).
The sequence $\{\lambda_0^k\}_{k=1}^{\infty}$ is uniformly bounded and $\{\psi_k\}_{k=1}^{\infty}$ is also equicontinuous. Hence we conclude, as before (in Step 2), that there exist $\psi_{\varepsilon,\bar{u}}$ and $\lambda_{\delta,\bar{u}}$, such that

$$\psi^k \rightrightarrows \psi_{\delta,\bar{u}}, \quad \lambda_0^k \to \lambda_{\delta,\bar{u}}.$$

As already indicated through the notation, the limits $\psi_{\delta,\bar{u}}$ and $\lambda_{\delta,\bar{u}}$ generically depend on the constants δ and $\bar{u}$. More specifically, these limits correspond to the optimal solution to the screening problem (P) if we replace V by $V - \delta\|u - u^*\|^2$ and introduce the additional constraint $\|u\| \leq \bar{u}$.

Adjoint Equation. Since by the maximum principle for problem $(\bar{\mathrm{P}}_k')$ (cf. Step 5) it is $\lambda_0^k > 0$, relations (5.6)–(5.8) are positively homogeneous of degree one in ψ^k/λ_0^k, and relation (5.9) is positively homogeneous of degree zero, it is possible to multiply equations (5.6)–(5.8) with positive numbers (and relabel the variables λ_0^k and ψ^k back), such that

$$0 < \lambda_0^k + \max_{\theta \in [\theta_0^k, 1]} \|\psi^k(\theta)\|^2 + \int_{\theta_1^k}^1 \left(\nu^k(\theta)\right)^2 d\theta \leq 1. \tag{5.16}$$

Integrating the components of the adjoint equation (5.13) yields (using the state-transversality condition (5.8))

$$\psi^k(\theta) \;=\; \int_\theta^1 \left(\lambda_0^k V_x f + \psi_0^k \varphi_x u^k\right) d\theta + \int_\theta^1 \nu^k g_x d\theta \tag{5.17}$$

for all $\theta \in \Theta$ (using the standard extension from $[\theta_0^k, 1]$ to Θ explained in Step 1).

Consider now a Borel measure with density given ν^k by (5.14).[27] By abuse of notation we denote this measure also by ν^k. The Borel measure ν^k is nonnegative by construction and bounded as a consequence of (5.16). Thus, there exists a Borel measure ν, such that (passing to a subsequence if necessary) $\nu^k \xrightarrow{w} \nu$ as $k \to \infty$.

Since the total variation of ψ^k on Θ is uniformly bounded for all k by (5.16) (for every absolutely continuous function is of bounded variation on a compact interval [28, p. 412]), and the sequence $\{\psi^k\}$ is also uniformly bounded as a consequence of (5.17), by Helly's selection principle [28, p. 398] there exists a function $\hat{\psi}$, so (a subsequence of) the sequence $\{\psi^k\}$ converges to $\hat{\psi}$. By taking the limit in (5.17) for $k \to \infty$ we thus obtain

$$\hat{\psi}_x(\theta) = \int_\theta^1 \left(\lambda_0 V_x f + \hat{\psi} \cdot \Phi_x\right) d\theta + \int_\theta^1 g_x d\nu, \tag{5.18}$$

for almost all $\theta \in \Theta_0^*$, where $\hat{\psi} = \psi_{\delta,\bar{u}}$. The adjoint equation C1 then follows by noting that (using the envelope theorem) $g_x(x^*(\theta),\theta) = -\left(\rho(x^*(\theta),\theta) - \theta\right) U_{x\theta}(x^*(\theta),\theta)$.

Maximality. Consider the maximality condition (5.15) for Problem $(\bar{\mathrm{P}}_k)$, which holds for k large enough. Since $x^k \rightrightarrows x^*$ and $u^k \to u^*$ (a.e. on Θ) for $k \to \infty$, we obtain the maximality condition C2.

Transversality. Since $\theta_0^k \to \theta_0^*$ as k tends to infinity (cf. Step 3), by setting

$$\lambda_1 = -\lim_{k\to\infty} 2k\left(U(\underline{x}^k,\theta_0^k) - r(\theta_0^k)\right) \text{ and } \lambda_2 = -\lim_{k\to\infty} 2k \min\{\theta_0^k, 0\}$$

we obtain from the state-transversality condition (5.7) for $k \to \infty$ (taking into account the maximality condition C2) that

$$\psi(\theta_0^*) = -\lambda_1 U_x(\underline{x}^*,\theta_0^*), \tag{5.19}$$

and from the type transversality condition (5.10) that

$$-\lambda_1\left(U_\theta(\underline{x}^*,\theta_0^*) - \dot{r}(\theta_0^*)\right) - \lambda_2 + \lambda_0 V(\underline{x}^*,\theta_0^*) = 0. \tag{5.20}$$

If $\theta_0^* > 0$, then by the definition of λ_2 it is $\lambda_2 = 0$. Combining relations (5.19) and (5.20) for $\lambda_2 = 0$ yields the transversality condition C3 (ii). If on the other hand $\theta_0^* = 0$, then we obtain the transversality condition C3 (i) from relation (5.19) directly.

[27] Recall [29, Ch. 1] that all Borel measures ν on Θ form a linear normed space with norm

$$\|\nu\| = \sup_{h\in C[0,1]:\|h\|_\infty=1} \int_0^1 h(\vartheta)\, d\nu(\vartheta).$$

Interpreting the above Stieltjes integral as a scalar product, we can think of any Borel measure ν as an element of the dual space $(C[0,1])^*$. A sequence of Borel measures $\nu^1, \nu^2, \ldots$ weak*-converges to ν if for all $h \in C[0,1]$,

$$\int_0^1 h(\vartheta)\, d\nu^k(\vartheta) \to \int_0^1 h(\vartheta)\, d\nu(\vartheta),$$

as $k \to \infty$. Since $C[0,1]$ is separable, by the theorem of choice every (bounded) sequence of measures has a weak*-convergent subsequence [30, p. 64; 31, p. 189]. Since for any $h \in C[0,1]$ the function $\int_0^\theta h(\vartheta)\, d\nu(\vartheta)$ is continuous a.e. on Θ, for any sequence of points $\{\theta_k\}_{k=1}^\infty$ with limit θ, for almost every such $\theta \in \Theta$ it is:

$$\lim_{k\to\infty} \int_0^{\theta_k} h(\vartheta)\, d\nu(\vartheta) = \int_0^\theta h(\vartheta)\, d\nu(\vartheta).$$

Moreover, if in addition the sequence $\{\nu^k\}_{k=0}^\infty$ of Borel measures has the Borel measure ν as its limit, then

$$\lim_{k\to\infty} \int_0^{\theta_k} h(\vartheta)\, d\nu^k(\vartheta) = \int_0^\theta h(\vartheta)\, d\nu(\vartheta)$$

for almost every $\theta \in \Theta$.

Complementary Slackness. The definition (5.14) of the measure ν^k, and the fact that $\nu^k \xrightarrow{w} \nu$ as $k \to \infty$ (cf. also footnote 27) yields the complementary slackness condition C4.

Nontriviality. If λ_0 and ψ are trivial, then (λ_0, ψ) must vanish identically on Θ_0^*. From the adjoint equation C1 it follows that

$$\int_\theta^1 g_x(x^*(\vartheta), \vartheta)\, d\nu(\vartheta) = 0$$

for all $\theta \in \Theta_0^*$. By the complementary slackness condition C4, whenever $\nu(\theta) \neq 0$ it is also $g(x^*(\theta), \theta) = 0$. This implies that

$$\int_{\theta_0^*}^1 d\nu = 0.$$

On the other hand, it is possible to renormalize the middle term in (5.16) in the solution of the approximate problem $(\bar{\mathrm{P}}_k)$ such that

$$\lambda_0^k + \max_{\theta \in [\theta_0^k, 1]} \|\psi^k(\theta)\|^2 + \int_{\theta_0^k}^1 + \left(\nu^k(\theta)\right)^2 d\theta = 1$$

for any $k \geq 1$. By taking the limit for $k \to \infty$,

$$\int_{\theta_0^*}^1 \left(\nu(\theta)\right)^2 d\theta = 1.$$

By the nonnegativity of the measure ν, this yields a contradiction. Hence, the incentive regularity condition A3 ensures that (λ_0, ψ) does not vanish identically on the optimal participation set Θ_0^*, which implies the nontriviality condition C5. □