

# Maximizing the Effect of Altruism: Resource-Constrained Mechanism Design\*

Nicholas Wu <sup>†</sup>      Nicole Immorlica<sup>‡</sup>      Brendan Lucier <sup>§</sup>

October 25, 2023

## Abstract

We study the problem faced by a policymaker who wants to influence the actions of a population of agents, subject to a per-capita transfer constraint. Each agent has a private type that determines their costs, and their action is verifiable. This paper introduces a new perspective by endogenizing the value of the resource that drives incentives; the transferable resource holds no inherent worth to the policymaker but is restricted by finite availability. In the base model, we characterize the optimal mechanism, showcasing the emergence of a pooling region where the transfer constraint binds for sufficiently low-cost types. We describe the threshold governing this region, and highlight the shadow costs that arise. We then reintroduce a linear value for the transferable; as the policymaker’s resource value increases, the policymaker demands more from agents whose transfer constraint is binding but demands less from others. Our findings contribute to understanding mechanism design in resource-constrained settings, with other applications in effective altruism and non-monetary incentives.

---

\*We would like to thank Dirk Bergemann, Laura Doval, Johannes Hörner, Bart Lipman, Kai Hao Yang, Jidong Zhou, Tan Gan, Yingkai Li, and participants at the Yale microeconomic theory lunch seminar and breakfasts for helpful comments and discussions.

<sup>†</sup>Department of Economics, Yale University, New Haven, CT 06511, [nick.wu@yale.edu](mailto:nick.wu@yale.edu)

<sup>‡</sup>Microsoft Research

<sup>§</sup>Microsoft Research

# 1 Introduction

With the rise of government, corporate and individual social responsibility, nontrivial sums of money are donated to address environmental and social issues. Individuals and corporations seeking to address environmental issues by offsetting carbon emissions will offer money to private agents such as biochar projects, conservation efforts, and nascent carbon removal technologies. Others donate to their alma maters, seeking to advance research and education, or to poverty alleviation organizations. Governments subsidize production of socially beneficial technologies such as the COVID-19 vaccine, provide school or housing vouchers to the poor, or give developers incentives to provide low-income housing; in fact, subsidies issued to profit-maximizing firms and are often deemed necessary to spur growth in nascent industries.<sup>1</sup> The amount of money given away is nontrivial; the top 25 individual givers, according to Forbes, have given over \$169 billion to charitable causes over their lifetimes.<sup>2</sup> Given a budgeted sum of money, how does one maximize impact?

We model this problem as a strategic interaction between a principal, endowed with a limited amount of a resource, and an agent with private information about its own costs. The scenario differs from the standard mechanism design literature in two important ways. Firstly, the principal is budget-constrained; that is, the principal cannot spend more than an upper bound regardless of the private information of the firm. Secondly, the value of money is endogenous to its ability to influence the agent; the principal has no value for the budget themselves, and only cares about their impact through the agent. This model is sufficiently general to also characterize many related design problems; we provide additional examples in Section 2.

We study an ex-post budget constraint because it arises in many natural scenarios. In many institutional settings, the allocation of funds is separate from their usage; one government body might determine the amount of funding to go towards a certain goal, while a different government agency oversees the usage of those funds. In other situations, the resource being used to incentivize an agent can be non-monetary; with prizes or prestige, the agent naturally has a bounded value for the prize, and the principal's value for the prize is endogenous and determined by the mechanism.

In this paper, we extend the mechanism design and delegation literature by proposing a model for providing incentives under a budget constraint. We show that the optimal solution

---

<sup>1</sup>See <https://www.politico.com/news/agenda/2021/12/22/carbon-removal-advance-market-commitments-525988>

<sup>2</sup>See <https://www.forbes.com/sites/forbeswealthteam/2022/01/19/americas-top-givers-2022-the-25-most-philanthropic-billionaires/>

is implementable as a subsidy schedule, where the altruist compensates the firm based on the observed change in production. We show that the schedule partitions the agent types into more and less efficient types, offering the entire budget for types that meet a threshold and discriminating on less efficient types.

Intuitively, the primary concern for output induces a fully *endogenous* value of money for the principal. The principal’s schedule can be understood as selecting some threshold type and splitting the agent types. All the more efficient agents are pooled; they are required to satisfy a single output threshold and given the entire budget. The more efficient agent types determine the shadow value of money; the less efficient agents face a schedule that optimizes with respect to this shadow value. The threshold trades off a welfare loss from withholding the budget from the less efficient agents with information rents it must pay the more efficient types. A positive measure of the most efficient types are always pooled, even as the budget increases. To illustrate this point starkly, we consider agent cost functions that satisfy a separability condition. We show that under this separability condition, the shape of the schedule is invariant to the size of the budget. In other words, no matter how large the budget is, the threshold for the pooling region never changes.

We extend the model to allow a competitive “outside option” for the altruist; that is, the unused residual budget has a linear value. We show that provided the value of the budget is not high, the same pooling region of the optimal schedule persists. The linear value case highlights the importance of the threshold type; as the principal’s value for the resource increases, the principal asks more from types more efficient than the threshold, and asks less from types less efficient than the threshold. We can extend our methodology to a broader class allowing nonlinear, convex budget costs as well. However, the ex post budget constraint feature is critical to the results; the pooling region does not arise if the budget only needs to be satisfied ex ante.

The paper is structured as follows. In the next subsection, we preview the related literature in mechanism design and delegation and distinguish this particular situation from similar work. Section 2 details the model and assumption. Section 3 formalizes the general mechanism design problem and highlights the important features of the model. Section 4 discusses the model with a linear value for the resource. Section 5 characterizes other variants of the model, and Section 6 concludes.

## 1.1 Literature

Our model is related to the literature on contract design in settings with moral hazard and private actions going back to Grossman and Hart (1983) and Holmström (1979). In this line of work, a principal contracts with an agent to take a costly action that is not directly observable, but that results in an observable outcome that is payoff-relevant to the principal and upon which payments can be contingent. In contrast to this line of work, in our model the principal can observe a deterministic outcome (i.e., production) and the uncertainty is over the cost function of the agent. Our work also introduces a constraint on the maximum transfer.

Additionally, our work is related to literature on monopoly regulation. Closely related work to ours is that of Baron and Myerson (1982), which studies a social planner problem for regulating a monopolist, and Laffont and Tirole (1986), which introduces noisy cost observations. However, our work is distinct in multiple ways. Firstly, our environment considers a budget-constrained principal, whereas there are no constraints on transfers in the other models in the literature. Additionally, the only objective for the principal in our scenario is to maximize production, where as the principal in the monopoly regulation environment typically maximizes a weighted sum of consumer surplus and producer profits. There are other minor distinctions in our model; we allow for a broad class of cost functions for the agent, whereas the literature typically assumes particular functional forms.

Given that our work involves nonstandard value for transfers, we complement the literature on delegation. The original delegation framework was formulated to Holmstrom (1980), which introduced the general class of delegation problems and provides conditions for existence of a solution. Delegation models typically assume that contingent transfers are unavailable to the principal, and instead the principal retains the right to take a state-dependent action that impacts utilities of both the principal and agent. Holmstrom (1980) characterizes optimal delegation sets for a series of examples assuming that the delegation sets take an interval form; later delegation literature generalized the interval assumption Alonso and Matouschek (2008), added money burning Amador and Bagwell (2013); Amador et al. (2018), and added costly verification Halac and Yared (2020). In contrast to much of the delegation literature, our model does allow for contingent transfers between the principal and agent, but in a special way; the transfers do not factor into the principal objective and face a budget constraint.

Given that the altruist has no intrinsic value for the budget herself, our problem relates to the literature on budget-feasible mechanism design (see Singer (2014) for a survey). That

literature studies a procurement problem where the designer wishes to purchase a utility-maximizing set items from sellers subject to a budget constraint Singer (2010). The main result shows that the designer can achieve a bounded approximation to the first-best so long as the utility function is submodular. Recent work in this area has investigated approximation ratios under alternate assumptions on the utility function Amanatidis et al. (2019); Gravin et al. (2020); Leonardi et al. (2017). Our work is differentiated from this literature in two significant ways. First, the altruist is maximizing production from a single firm rather than maximizing a utility function over a combinatorial set of items. Second, other papers typically impose budget constraints on the agent side, whereas the constraints in our environment are on the principal’s side.

Also related is the literature on contest design. This literature studies a setting in which a designer with a fixed budget wishes to solicit submissions to a contest. The designer can split the budget among potentially multiple prizes. The objective can be to optimize the aggregate quality of submissions Archak and Sundararajan (2009); Moldovanu and Sela (2001), or the quality of the best submission Chawla et al. (2019); Che and Gale (2003); Ghosh and Kleinberg (2016); Moldovanu and Sela (2006). Maximizing aggregate quality is similar to our altruist’s objective of maximizing production. Furthermore, contests with multiple agents is reminiscent of the extension of our problem to multiple firms, discussed in Section 5. In the contest design literature, agents compete over a scarce resource, the prize(s) through their choice of quality level. In our model, in contrast, the optimal solution provides firms guaranteed payouts independent of others choices (as the altruist partitions the budget between firms based on her beliefs about their types). The quality choice, and corresponding menu choice, of the firm then impacts the prices of the competition in the downstream market.

Our paper is topically related to the large literature on mechanism design for social good Abebe and Goldner (2018). This literature seeks to use mechanism design to solve social problems. In particular, our primary motivation of an altruist funding carbon removal technologies relates to the environmental economics literature (see Greaker and Popp (2022) for a survey). Papers on carbon reduction specifically address a range of related issues, including optimal cap and trade policies for firms with uncertain production costs Goldner et al. (2019), the design of carbon credits in the presence of an uncontrolled downstream market Babaioff et al. (2023), and optimal contract design for a budget-constrained altruist attempting to maximize afforestation Dai Li et al. (2022).

One interpretation of our model focusing on the impact of markets on an altruist objective

is echoed in Kang (2022). In that paper, the social welfare optimizer – a policymaker – determines the quality and allocation of a good that she produces with an eye towards redistribution. The policymaker must contend with the fact that its production impacts the market for inputs to the production process and hence the price (and demand) for the privately-provided goods in the downstream market. This can create distortions that run against the policymaker’s objective. In contrast, our work focuses on a funder, rather than an additional producer, of goods, and assumes these goods are produced subject to a type-dependent production cost function.

## 2 Model

We first lay out the formalism for the model, and then discuss applications at the end of this section. Consider a principal-agent design problem where the principal would like to provide incentives for an agent to take a costly action. The action space is  $X = \mathbb{R}_+$ ; an action  $x \in X$  can be interpreted as an amount of a good produced by the agent. Without loss, we normalize the space  $X$  so that the principal derives utility  $x$  from the agent taking action  $x$ . The agent has a private cost function  $\Psi : X \times \Theta \rightarrow \mathbb{R}_+$ , which determines the cost the agent incurs from producing  $x \in X$  given the agent’s type  $\theta \in \Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+$ . Without loss, we normalize the cost function  $\Psi(0, \theta) = 0$  for all  $\theta$ .

The agent’s type is drawn prior to the start of the game by Nature from a full-support distribution  $\mu \in \Delta(\Theta)$ . The type is privately observed only by the agent, but the principal has belief  $\mu$  over the agent type. We assume the belief  $\mu$  admits a probability density function  $f$ , with corresponding cumulative distribution function  $F$ .

The principal has a budget  $T$  of a transferable resource that the agent values. The agent’s utility is quasilinear in the cost and the transfer: that is, if the agent of type  $\theta$  produces  $x$  and receives a transfer  $t$ , his utility is  $u_A(x, \theta) = t - \Psi(x, \theta)$ . To be clear, we suppose the principal can observe the action  $x$ , as opposed to the moral hazard problem in Holmström (1979) and Grossman and Hart (1983); instead, the principal must decide how to award funds given the observed action, without precisely knowing the agent’s cost *function*.

To highlight the key economic forces, we will present the baseline model where the principal has no value for the transferable resource. In Section 4 we consider allow the principal to have a constant value the resource, and consider a generalized value function in the extensions in Section 5.

**Assumption 1** *We assume the cost function  $\Psi$  is nonnegative, twice continuously differ-*

entiable, supermodular, and strictly increasing in both arguments. Additionally, we assume that  $\Psi$  and its partial derivative in  $\theta$ ,  $\Psi_\theta$ , are both convex in  $x$ .

A couple of remarks about the assumptions and game structure are in order. The supermodularity assumption is indispensable, and intuitively requires that higher types suffer steeper costs. The  $C^2$  differentiability assumption is technical and allows us to use cross-derivatives. The convexity assumptions ensure validity and sufficiency of first-order optimization conditions. The interpretation of the convexity condition in  $\Psi_\theta$  is that the *convexity* of the agent's cost function is increasing in the type. For instance, in an example where an altruist offers money to a firm, the assumption requires that the higher  $\theta$  firms are less scalable (i.e. their production costs grow faster). In this setting, the type  $\theta$  can be interpreted as an inefficiency parameter.

By standard arguments, we invoke the revelation principle, so we consider direct mechanisms where the agent reports his type to the principal subject to incentive compatibility and individual rationality constraints. (We will see later that all feasible direct mechanisms will be implementable by a subsidy schedule.)

More precisely, the mechanism design problem requires the principal to select functions  $x : \Theta \rightarrow X$  and  $t : \Theta \rightarrow \mathbb{R}_+$ , in the following optimization problem:

$$\begin{aligned}
 \max_{x,t} \quad & \int_{\theta \in \Theta} x(\theta) f(\theta) d\theta & (1) \\
 \text{subject to} \quad & t(\theta) - \Psi(x(\theta), \theta) \geq t(\theta') - \Psi(x(\theta'), \theta) & \forall \theta, \theta' \in \Theta & \text{(IC)} \\
 & t(\theta) - \Psi(x(\theta), \theta) \geq 0 & \forall \theta \in \Theta & \text{(IR)} \\
 & t(\theta) \leq T & \forall \theta \in \Theta & \text{(B)}
 \end{aligned}$$

To understand the problem, we provide three example applications of the model.

**Example (Emissions Reduction)** Consider a policymaker trying to reduce emissions from a large population. Individuals in the population have private emissions abatement costs  $\theta \sim F$  which determine how severely emissions reductions impact their utility; that is,  $\Psi(x, \theta)$  is how costly an individual of type  $\theta$  finds it to reduce their emissions by  $x$ . The policymaker cannot pay any single individual more than an upper bound  $T$ , but would like to induce the largest possible collective emissions reduction. The optimal mechanism can be implemented by a subsidy schedule, which determines how much compensation an individual receives as a function of the individuals's emissions reduction.

**Example (Altruism)** Suppose an altruist or a government has pledged a budgeted sum  $T$  to spur production of a product (vaccine, baby formula) during a shortage. The altruist is the principal, and the monopoly producing the product is the agent, whose private type characterizes their production costs. The action  $x$  corresponds to the amount of the product produced, and the resource is the budgeted money. The altruist seeks to maximize production conditional on respecting the budget constraint.

**Example (Non-monetary Incentives)** Consider an academic institution interested in increasing its research output by hiring a new researcher. The institution can grant the researcher a non-monetary award (title, prize, tenure) which has no intrinsic value to the institution itself. The researcher values the award at  $T$ , and has a private cost of producing output  $\Psi(\cdot, \theta)$ . The institution can assess the output of the researcher and decide with what probability to give the award. The mechanism in this example corresponds to the institution committing to a schedule that maps research output into a probability of receiving the award.

### 3 Optimal Mechanisms

In this section, we first present the mathematical characterization of the optimal mechanism. We then discuss some of the qualitative features of the model, notably the emergence of a pooling region. We provide some economic intuition for the optimal mechanism as splitting the type space into “efficient” and “inefficient” types and optimizing distinct shadow cost functions for the two. Finally, we show that the optimal mechanism admits a closed-form under a separability assumption on  $\Psi$ , and in this case, the size of the pooling region is invariant to  $T$ .

Solving our mechanism relies on a lemma which characterizes the feasible schedules, or the set of allocations  $x : \Theta \rightarrow \mathbb{R}_+$  for which there exists a transfer schedule  $t : \Theta \rightarrow [0, T]$  that satisfies (IC), (IR), and (B). The lemma follows the standard mechanism design techniques to derive the feasible set.

**Lemma 1**  $x : \Theta \rightarrow X$  is a feasible schedule if and only if the following two conditions hold:

1.  $x$  is nonincreasing.
2.  $\Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds \leq T$



Further, given a feasible schedule  $x$ , a transfer function that supports the schedule is given by

$$t(\theta) = \Psi(x(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds \quad (2)$$

The first condition is a monotonicity constraint, and the second condition constrains the magnitude of  $x$ . Note that the normalization constraint (2) comes from the transfer bound, the IC constraints, and the envelope theorem, and so the feasible set characterization does not depend on the prior. Note the first part of the Lemma implies that these two conditions are necessary and sufficient for a feasible schedule, and the second part explicitly determines the transfer schedule that implements any feasible schedule.

The proof of Lemma 1 is standard; we use supermodularity to show that  $x$  must be nonincreasing, and then apply the envelope theorem to the IC constraints to derive the transfer schedule. The transfer schedule in Lemma 1 is analogous to the Myersonian transfer schedule, since the transfer schedule is also derived from using the envelope theorem to integrate out interim utility of the firm from the incentive compatibility conditions.

The integral expression in the transfer expression is the information rent that type  $\theta$  receives; since  $\Psi(x(\theta), \theta)$  is exactly the loss the firm experiences from producing  $x(\theta)$ , the transfer compensates the firm for the loss and provides additional rent, which is larger for more efficient types (lower  $\theta$ ).

As a corollary, since  $\Psi$  and  $\Psi_{\theta}$  are convex, conditions (1) and (2) jointly determine a strictly convex subset of the space of functions from  $\Theta \rightarrow \mathbb{R}_+$ . Thus, as a corollary of our lemma, since the mechanism design problem maximizes a linear functional over a strictly convex set, the optimal mechanism is unique.

**Corollary 1** *If a solution exists, the solution to the generalized mechanism design problem in (1) is unique.*

Further, Lemma 1 implies that the mechanism is implementable by a subsidy schedule or action-contingent transfers; that is, the optimal mechanism  $x(\theta), t(\theta)$  parametrizes a well-defined function from  $X \rightarrow \mathbb{R}_+$ .

**Corollary 2** *Suppose  $\theta < \theta'$ , and let  $x, t$  be a feasible mechanism. If  $x(\theta) = x(\theta')$ , then  $t(\theta) = t(\theta')$ .*

**Proof.** By Lemma 1, since  $x$  must be nonincreasing and  $x(\theta') = x(\theta)$ ,  $x$  must be constant

on  $(\theta, \theta')$ . As a minor abuse of notation, let  $x$  be that value on  $(\theta, \theta')$ . Then

$$\begin{aligned}
t(\theta) &= \Psi(x, \theta) + \int_{\theta}^{\bar{\theta}} \Psi_{\theta}(x(t), t) dt \\
&= \Psi(x, \theta) + \int_{\theta}^{\theta'} \Psi_{\theta}(x, t) dt + \int_{\theta'}^{\bar{\theta}} \Psi_{\theta}(x(t), t) dt \\
&= \Psi_{\theta}(x, \theta') + \int_{\theta'}^{\bar{\theta}} \Psi_{\theta}(x(t), t) dt = t(\theta')
\end{aligned}$$

■

Since  $x, t$  are both nonincreasing, the corollary implies that any feasible mechanism  $x(\cdot), t(\cdot)$  can be implemented as a subsidy schedule  $\hat{t} : X \rightarrow \mathbb{R}_+$ .

With some more work, we can show the existence of a solution and characterize it with a differential equation system.

**Theorem 1** *An optimal mechanism  $(x, t)$  for (1) exists and is unique. The optimal mechanism, together with a nonnegative Lagrange multiplier  $\lambda$  and a nonnegative, absolutely continuous costate function  $\rho$ , jointly solve the following system of differential equations with boundary constraints:*

$$\rho(\theta) > 0 \implies \dot{x}(\theta) = 0 \tag{3}$$

$$\dot{\rho}(\theta) = \lambda \Psi_{x\theta}(x(\theta), \theta) - f(\theta) \tag{4}$$

$$\rho(\bar{\theta}) = 0 \tag{5}$$

$$\rho(\underline{\theta}) = \lambda \Psi_x(x(\underline{\theta}), \underline{\theta}) \tag{6}$$

$$t(\underline{\theta}) = \Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds = T \tag{7}$$

where (4) holds wherever  $\rho$  is differentiable.

To understand the result, note that the feasible set characterization in Lemma 1 allows us to focus solely on characterizing  $x$ , subject to a nonincreasing condition and a rewritten budget constraint. We handle the budget constraint by writing the Lagrangian, and deal with the nonincreasing condition by taking an optimal control approach<sup>3</sup> where ‘time’ corresponds to the type interval, the control state variable is the required action  $x$ , and control variable is the derivative of  $x$ . We can interpret the conditions in Theorem 1 through this control framework. The costate  $\rho$  denotes the shadow cost of the monotonicity constraint. The first

---

<sup>3</sup>See Laffont and Tirole (1986)

condition (3) is the complementary slackness condition; wherever the constraint's shadow cost  $\rho^*$  is positive, the constraint is binding (i.e.  $x$  is constant). The second condition (4) dictates the costate evolution. The constraints (5) and (6) establish boundary conditions. The final constraint (7) establishes that the budget binds at the most efficient / least costly type.

The proof intuition follows three steps. First, we write out the Lagrangian relaxation of the problem using the feasible set characterization from Lemma 1. We formulate the optimization as an optimal control problem, where the control variable governs the derivative of  $x$  and is constrained to be nonpositive wherever  $x$  is differentiable.

Second, we show that the conditions in Theorem 1 are necessary and sufficient for an optimal mechanism. Fixing  $\lambda$ , the first four conditions (3)-(6) are necessary and sufficient for a solution to the Lagrangian relaxation control problem; necessity follows by invoking the Pontryagin maximum principle, and sufficiency follows from Arrow's sufficiency condition as proved in Kamien and Schwartz (1971). Sufficiency of  $\lambda > 0$  and (7) follows from Lagrangian duality; we show necessity by proving no solution to (3)-(6) exists for  $\lambda \leq 0$ .

Finally, we prove existence. We use the theory of ordinary differential equations to show that (3)-(6) has a solution for  $\lambda > 0$ . We then use the intermediate value theorem to show that there must exist a  $\lambda$  such that (7) holds. Since we proved in the second step that these conditions were necessary and sufficient for a solution, we thus showed that an optimal mechanism exists. Corollary 1 implies that this solution must also be unique.

To understand how the characterization in Theorem 1 can be used to determine the optimal mechanism, suppose  $\lambda$  were exogenously fixed. Consider the illustration in Figure 1. Complementary slackness implies that when  $\rho$  is positive,  $x$  must be constant, and whenever  $x$  is decreasing,  $\rho$  must be constant at zero, and hence  $x$  must follow the trajectory implied by the implicit equation derived from  $\dot{\rho} = 0$ . The boundary conditions dictate the start and end values of the costate  $\rho$ . Note that altering  $\lambda$  either uniformly increases or decreases  $x$ ; intuitively, at the solution value of the Lagrange multiplier, the budget constraint exactly binds.

A remark on the convexity assumption on  $\Psi_\theta$  is in order. The result in Theorem 1 relies on the Pontryagin maximum principle providing a unique solution; this occurs given that the maximized Hamiltonian is concave. In the model, the Arrow sufficiency condition is that for all  $\theta$ ,

$$xf(\theta) - \lambda\Psi_\theta(x, \theta)$$

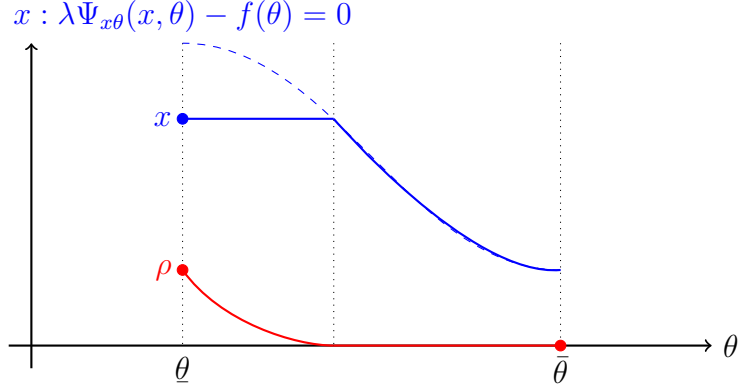


Figure 1: Illustration of the conditions in Theorem 1. The dotted blue line satisfies the implicit equation. When  $\rho$  is zero,  $x$  must coincide with the dotted blue line.

is concave in  $x$ . Assuming  $\Psi_\theta$  is convex in  $x$  is equivalent to this sufficiency condition.<sup>4</sup>

In the next sections, we discuss qualitative properties of the optimal mechanism and economic insights derived from the solution characterization in Theorem 1.

### 3.1 Pooling

We first discuss when the principal awards the maximum  $T$  to the agent (i.e., for what types  $t(\theta) = T$ ). From the budget binding constraint, we know that the most efficient type  $\underline{\theta}$  always receives the full transfer. In this section, we provide three insights. First, we show that a positive measure of types always receive the entire transfer; we say the optimal mechanism “pools” the most efficient types. Second, we characterize the threshold type; agents more efficient than the threshold are pooled and receive the full transfer  $T$ , and the principal optimally withholds the resource from agents less efficient than the threshold. Third, we characterize when the optimal mechanism pools all types; in this case, the optimal mechanism simply pays the entire  $T$  to the agent whenever the agent meets a base action level.

**Proposition 1 (Pooling At The Top)** *There exists a  $\hat{\theta} > \underline{\theta}$  such that the optimal mechanism  $x^*$  is constant on  $[\underline{\theta}, \hat{\theta}]$ . The transfer  $t^*$  is equal to  $T$  for these types.*

**Proof.** Consider the conditions in Theorem 1. Since  $\rho(\underline{\theta}) > 0$ , complementary slackness (3) implies that  $x^*$  must be constant until  $\rho$  reaches zero. Corollary 2 implies that the optimal

<sup>4</sup>There are other sufficiency conditions for Pontryagin’s maximum principle with weaker assumptions on the Hamiltonian. Relaxing assumptions on  $\Psi$  would require another sufficiency argument for the uniqueness of a solution characterized by the maximum principle.

$t^*$  is also constant on this set of types, and the budget binding condition (7) implies that the transfer is  $T$ . ■

To intuitively understand why this pooling region always occurs, consider the loss in principal utility that occurs in the optimal mechanism relative to perfect information. With perfect information, the principal always gives away the entirety of the budget  $T$  and obtains the maximum action  $x$  from the agent such that  $\Psi(x, \theta) = T$ . The optimal mechanism, relative to perfect information, differs in two ways. Either the principal cannot obtain the highest action because the principal deliberately withholds the resource, or the principal must concede information rent to the agent. Hence, the trade-off between the relative loss from these two effects dictates the size of the pooling region of the most efficient types.

**Proposition 2 (Pooling Threshold)** *Suppose the optimal  $x^*$  is not constant on the entire interval of types. Let  $\underline{x} \equiv x^*(\underline{\theta})$ . Then the largest type  $\hat{\theta}$  that receives the full transfer satisfies the following:*

$$\frac{\Psi_{x\theta}(\underline{x}, \hat{\theta})}{\Psi_x(\underline{x}, \hat{\theta})} = \frac{f(\hat{\theta})}{F(\hat{\theta})} \quad (8)$$

**Proof.** By optimality, there exists some Lagrange multiplier and costate function  $\lambda, \rho$  that satisfy the conditions of Theorem 1. At the cutoff, since  $x^*$  is not constant on the entire interval, it must be that  $x^*$  stops being constant at  $\hat{\theta}$ . Complementary slackness (3) implies that  $\rho(\hat{\theta}) = 0$ . Integrating out the costate evolution (4) gives

$$\lambda \Psi_x(\underline{x}, \hat{\theta}) = F(\hat{\theta})$$

Since we assumed  $f$  was differentiable,  $x^*$  is continuous at  $\hat{\theta}$  but the right derivative must be negative; hence,  $\dot{\rho}(\hat{\theta}) = 0$  since complementary slackness implies  $\rho$  must be constant at zero for some interval above  $\hat{\theta}$ . Using the costate evolution (4) and taking the upper limit  $\theta \rightarrow \hat{\theta}$ , we get

$$\lambda \Psi_{x\theta}(\underline{x}, \hat{\theta}) = f(\hat{\theta})$$

Substituting out  $\lambda$  gives (8). ■

To interpret this condition, rewrite (8) as

$$\Psi_x(\underline{x}, \hat{\theta}) = \Psi_{x\theta}(\underline{x}, \hat{\theta}) \frac{F(\hat{\theta})}{f(\hat{\theta})}$$

Under this rewriting, the condition is equivalent to the marginal benefit from paying the

threshold type to be equal to the marginal information rent cost demanded.

It is possible that the principal always awards the entire  $T$  to the agent, regardless of the type. The following result characterizes when this happens. Let  $\bar{x}$  be such that  $T = \Psi(\bar{x}, \bar{\theta})$  (that is, the highest action that least efficient type  $\bar{\theta}$  would be willing to take if given the entire budget  $T$ ). Then

**Proposition 3 (Complete Pooling)** *The optimal mechanism pools all types (that is, the optimal  $x$  is constant) if and only if for all  $\theta$ ,*

$$\Psi_x(\bar{x}, \theta) \geq F(\theta)\Psi_x(\bar{x}, \bar{\theta}).$$

*Further, if the optimal mechanism pools all types, then  $x(\theta) = \bar{x}$  and  $t(\theta) = T$  for all  $\theta$ .*

**Proof.** First, suppose that  $x^*(\theta) = x^*$  is constant. Let  $\lambda, \rho$  be the Lagrange multiplier and costate variable satisfying Theorem 1. By the budget binding condition (7),

$$\begin{aligned} T &= \Psi(x^*, \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_{\theta}(x^*, s) ds \\ &= \Psi(x^*, \underline{\theta}) + \Psi(x^*, \bar{\theta}) - \Psi(x^*, \underline{\theta}) \\ &= \Psi(x^*, \bar{\theta}) \end{aligned}$$

Hence,  $\bar{x} = x^*$ . By the upper bound condition (5), integrating the costate evolution equation (4), we must have

$$\rho(\bar{\theta}) = \lambda\Psi_x(\bar{x}, \bar{\theta}) - 1 = 0$$

So  $\lambda = 1/\Psi_x(\bar{x}, \bar{\theta})$ . Then since  $\rho$  satisfies the costate evolution condition (4),

$$\rho(\theta) = \lambda\Psi_x(\bar{x}, \theta) - F(\theta) \geq 0$$

and so the result follows.

For the converse, suppose that  $\bar{\lambda}\Psi_x(\bar{x}, \theta) \geq F(\theta)$  for all  $\theta$ . Then construct  $\rho(\theta) = \bar{\lambda}\Psi_x(\bar{x}, \theta) - F(\theta)$ , and define  $x^*(\theta) = \bar{x}$  and  $\lambda = 1/\Psi_x(\bar{x}, \bar{\theta})$ . It is easy to check that these satisfy the condition of Theorem 1. Hence the optimal mechanism is  $x(\theta) = \bar{x}$  and  $t(\theta) = T$  for all  $\theta$ . ■

**Example Revisited (Emissions Reduction)** Revisit the emissions reduction example from before. As Corollary 2 implies, the optimal mechanism for the regulator is to offer a

subsidy schedule, that pays some amount to each firm depending on their emissions reduction. Proposition 1 implies that the regulator pays the maximum amount  $T$  for a sufficiently large reduction, which is taken by a *positive measure* of firms. Proposition 3 then provides conditions for which the optimal regulator’s policy has a “bang-bang” feature; that is, the regulator never gives any partial transfer, but only pays  $T$  or zero. This latter case is equivalent to providing an unconditional transfer and mandating a fixed level of emissions reduction.

### 3.2 Shadow Costs

Having established that the mechanism always admits a pooling region up to some  $\hat{\theta}$ , we provide some economic intuition for the optimal  $x$  given  $\hat{\theta}$ . That is, throughout this section, suppose the optimal mechanism has a cutoff type  $\hat{\theta} < \bar{\theta}$  determined by Proposition 2. We characterize the schedule above and below the cutoff, using the shadow value interpretation of the Lagrange multiplier  $\lambda$ .

First, define  $\bar{x} \equiv x(\theta)$  for  $\theta \in [\underline{\theta}, \hat{\theta}]$ . The following result intuitively shows that the optimal  $\bar{x}$  solves a maximization problem for the principal, where the marginal benefit is given by the probability that the agent meets the cutoff, and the marginal cost is  $\lambda\Psi(\cdot, \hat{\theta})$ , the shadow cost of the transfer that must be paid to the threshold type.

**Corollary 3** *Given the optimal shadow value of money  $\lambda$ , the highest action  $\bar{x}$  solves the following maximization:*

$$\bar{x} = \arg \max_x \left\{ xF(\hat{\theta}) - \lambda\Psi(x, \hat{\theta}) \right\}.$$

**Proof.** Integrating out the costate evolution from Theorem 1, we find that

$$\lambda\Psi_x(\bar{x}, \hat{\theta}) - F(\hat{\theta}) = 0$$

which is exactly the FOC for the maximization. The FOC exactly characterizes the solution since  $\Psi$  is convex in  $x$ . ■

Corollary 3 characterizes the tradeoff for the highest action  $\bar{x}$ ; intuitively, by increasing the highest action, the principal gets more production from  $F(\hat{\theta})$  types of agents, but pays the shadow cost  $\lambda\Psi(\cdot, \hat{\theta})$ . As mentioned before, the value of the Lagrange multiplier  $\lambda$  can be interpreted as the shadow value of the resource to the principal.

Now, we show that the types from  $(\hat{\theta}, \bar{\theta}]$  intuitively face a *different* shadow cost to the principal. This insight helps provide us an intuition for what the optimal mechanism looks like for types  $\theta > \hat{\theta}$ .

**Proposition 4** Suppose  $\Psi_{x\theta}(x, \theta)/f(\theta)$  is weakly increasing in  $\theta$  and  $\hat{\theta} < \bar{\theta}$ . Then on  $[\hat{\theta}, \bar{\theta}]$ , the optimal  $x$  satisfies the point-wise maximization:

$$x(\theta) = \max_x \{x f(\theta) - \lambda \Psi_\theta(x, \theta)\} \quad (9)$$

**Proof.** Recall that by definition, since pooling ends at  $\hat{\theta}$ ,  $\rho(\hat{\theta}) = 0$ . By the Bellman principle of optimality, the restriction of  $x$  to the domain  $[\hat{\theta}, \bar{\theta}]$ , must be optimal for the following control subproblem:

$$\begin{aligned} \max_x \quad & \int_{\hat{\theta}}^{\bar{\theta}} (x(\theta)f(\theta) - \lambda \Psi_\theta(x(\theta), \theta)) \, d\theta \\ \text{subject to} \quad & \dot{x} = u \leq 0 \end{aligned} \quad (\text{Monotonicity})$$

Note that the solution to (9) point-wise maximizes the objective of the subproblem. To confirm that this is optimal, it suffices to show that  $x(\theta)$  satisfies the monotonicity constraint. Note that since  $\Psi_\theta$  is convex, the first-order condition for  $x$  gives

$$1 = \lambda \frac{\Psi_{x\theta}(x, \theta)}{f(\theta)}$$

Note that the RHS is an increasing function of  $x$ . Since the RHS is assumed to be increasing in  $\theta$  and increasing in  $x$ , it follows that  $x(\theta)$  must be decreasing in  $\theta$ . ■

To interpret the maximization problem in (9), note that by increasing  $x(\theta)$ , the principal gains a marginal benefit of  $f(\theta)$  but must pay  $\Psi_\theta(x(\theta), \theta)$  more *information rent* to the best types (or equivalently, must procure  $\Psi_\theta(x(\theta), \theta)$  of the resource). Note that the shadow resource cost is  $\lambda$ , so the cost is exactly the shadow information rent cost that must be paid.

As a remark, note that the main obstacle to point-wise maximization is the monotonicity constraint; hence, one can obtain a more general variant of Proposition 4 by ironing the shadow cost function  $\Psi_\theta(x, \theta)/f(\theta)$  (and in fact, this is what the costate variable does in Theorem 1).

In sum, the optimal mechanism can be interpreted as the principal splitting the agent types into two populations, and maximizing  $x$  against different shadow cost functions. For the efficient types, the principal maximizes against the shadow production cost,  $\lambda \Psi$ ; for the less efficient types, the principal maximizes against the shadow information rent cost,  $\lambda \Psi_\theta$ .



### 3.3 Separable Cost Function

In this section, we show that for multiplicatively separable  $\Psi$  the mechanism admits a closed form solution, where the shape of the solution, including the pooling region from Proposition 1, is determined by the belief. In particular, the multiplicatively separable case is remarkable in that the size of the pooling region is *independent* of  $T$ , which we show in Proposition 6.

Firstly, we say  $\Psi$  is separable in  $x$  and  $\theta$  if  $\Psi(x, \theta) = B(\theta)\Gamma(x)$ , where  $\Gamma$  is convex by assumption. Without loss, since  $\Psi$  is assumed increasing in  $\theta$ , we can renormalize the type space and redefine  $\Theta$  so that  $\Psi(x, \theta) = \theta\Gamma(x)$  where  $\Gamma$  is a strictly convex function. In this case, we can actually provide a closed-form characterization of the shape of the schedule. Let  $\tilde{f}$  denote the left derivative of  $\text{cav } F$ , the concavification of the cumulative distribution function  $F$  on  $[0, \infty)$ , and denote by  $(\Gamma')^{-1}$  the inverse function associated with  $\Gamma'$  (which is increasing by assumption).

**Proposition 5** *Suppose that  $\Psi(x, \theta) = \theta\Gamma(x)$ , where  $\Gamma$  is convex. Then the optimal mechanism  $(x^*, t^*)$  induces a production schedule*

$$x^*(\theta) = (\Gamma')^{-1} \left( \frac{\tilde{f}(\theta)}{\lambda} \right)$$

where  $\lambda$  is the nonnegative Lagrange multiplier chosen so that

$$\theta\Gamma(x^*(\theta)) + \int_{\theta}^{\bar{\theta}} \Gamma(x^*(\theta)) d\theta = T$$

The transfer schedule is still given by Lemma 1.

Intuitively, this holds because the linearity of  $\Psi$  in  $\theta$  allows the separation of the monotonicity constraint into a standard ironing problem. Thus, for this specific case, the optimal schedule depends on the density of the concavified cumulative distribution function of the belief,  $\tilde{f}$ . Since Proposition 5 shows that  $x^*$  is an increasing transformation of  $\tilde{f}$ , the content of Proposition 5 is that for separable cases, the shape of the optimal mechanism is entirely determined by the *belief* of the principal and not by the agent's cost function. Indeed, the pooling behavior of the optimal mechanism is completely identical for  $\Psi^A(x, \theta) = \theta\Gamma^A(x)$  and  $\Psi^B(x, \theta) = \theta\Gamma^B(x)$  for any two convex  $\Gamma^A, \Gamma^B$ .

We illustrate how this insight applies to a discrete, two-type example. Suppose  $\theta$  is either 1 or 2, and the cost function takes the form  $\Psi(x, \theta) = \theta x^2$ . The principal believes that  $\theta = 1$  with probability  $\mu$ . Figure 2 illustrates  $\tilde{f}$ ; for the two-type case, whether  $\mu$  is

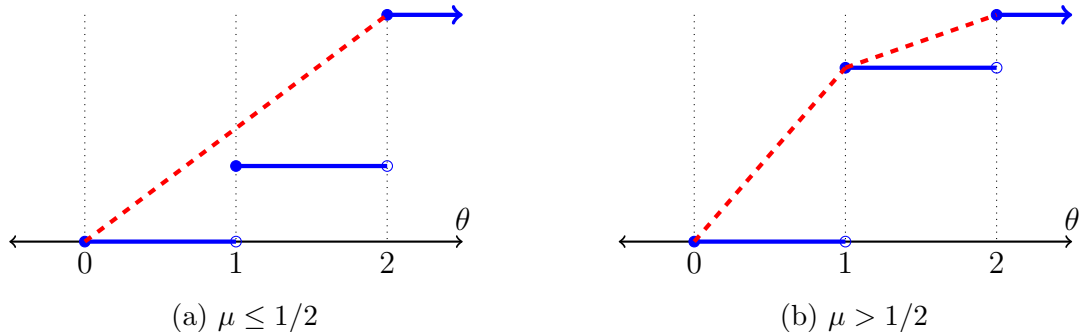


Figure 2: The distribution function corresponding to the belief of the principal is plotted in solid lines, and the dashed line denotes the concavification. In the first case, the concavified distribution function is constant, and so the optimal mechanism pools the types. In the second case, the left-derivative of the concavified distribution function takes different values at 1 and 2, so the optimal mechanism separates types.

larger than or less than  $1/2$  determines whether  $\tilde{f}(1) = \tilde{f}(2)$  or not. By Proposition 5,  $x^*$  is an increasing transformation of  $\tilde{f}$ , so the shape of  $\tilde{f}$  also determines the shape of  $x^*$ . The optimal mechanism separates types 1 and 2 if and only if  $\mu > 1/2$ .<sup>5</sup>

Note that for the separable case, Propositions 2 and 3 simplify considerably:

**Corollary 4** *Suppose  $\Psi(x, \theta) = \theta\Gamma(x)$ , and suppose the optimal  $x^*$  is not constant on the entire interval of types. Let  $\underline{x} \equiv x^*(\underline{\theta})$ . Then the largest type  $\hat{\theta}$  that receives the full transfer satisfies the following:*

$$\frac{1}{\hat{\theta}} = \frac{f(\hat{\theta})}{F(\hat{\theta})} \quad (10)$$

**Corollary 5** *Suppose  $\Psi(x, \theta) = \theta\Gamma(x)$ . Then the optimal mechanism pools all types if and only if for all  $\theta$ ,*

$$\theta/\bar{\theta} \geq F(\theta)$$

In some sense, the upper bound on the schedule implied by Proposition 1 follows from the existence of an upper bound on transfers. An initial intuition might be that the pooling region occurs because the constraint is binding, and that the binding region shrinks if  $T$  is large. However, this is *not* the case. The pooling arises because of the *relative* informational cost of separating low and high  $\theta$  to *relative* production ability. In fact, for the separable case, we have the following result:

---

<sup>5</sup>In our modeling assumptions, we took  $F$  as continuous and differentiable. We can extend this to scenarios where  $F$  is discontinuous (equivalently,  $f$  admits point masses) by taking the limits of a sequence of continuous  $F_n$  converging to  $F$  and observing that for the separable case, the optimizer function is continuous in  $\tilde{f}$ , which is continuous in  $f$  in the sup norm.

**Proposition 6** *Suppose  $\Psi(x, \theta) = \theta\Gamma(x)$ . Let  $\hat{\theta}$  be the largest type that receives the full transfer. Then  $\hat{\theta}$  is invariant to  $T$  and  $\Gamma$ .*

**Proof.** By Proposition 5, consider  $x_T$  and  $x_{T'}$ , the optimal schedules for budgets  $T$  and  $T'$ . Since  $\tilde{f}$  is unchanged, and  $x^*$  is an increasing transformation of  $\tilde{f}$ , the pooling region for  $x_T$  and  $x_{T'}$  are the same, and hence  $\hat{\theta}$  is invariant to altering the size of the budget. ■

That is, the size of the pooling region does not change at all as  $T$  changes. This implies that the *shape* of the contract is determined primarily by the belief of the principal, and the *scale* of the contract is determined by the size of the budget. Hence, the existence of the pooling region at the top stems from the principal's relative tradeoff between giving up information rents and withholding a non-valuable resource.

### 3.4 Linear Subsidy Schedule

Suppose the principal only offered some linear subsidy, paying for  $x$  at a constant rate  $r$ . By setting such a subsidy schedule, very efficient types exhaust the budget and take  $x = T/r$  and less efficient types choose some  $x$  that equalizes their marginal cost with the rate:  $\Psi_x(x, \theta) = r$ . We will show that the mechanism design approach can always do better; that is, some feasible schedule Pareto dominates the schedule resulting from a linear subsidy.

**Proposition 7** *Let  $x_r$  be the outcome schedule from a linear subsidy at rate  $r$ . There exists a feasible schedule  $x^*$  in the mechanism design problem such that for all  $\theta$ ,  $x^*(\theta) \geq x_r(\theta)$  with inequality holding strictly on a positive measure of  $\theta$ .*

**Proof.** We construct a transfer schedule  $t$  that implements  $x_r$  and then show that that transfer schedule does not exhaust the budget, implying that we could increase production on some measure of types. Let  $\hat{\theta}_r$  be the largest  $\theta$  such that  $\Psi_x(T/r, \theta) \leq r$ ; that is, the largest  $\theta$  whose marginal cost of producing at  $T/r$  is less than  $r$ .

First, suppose that  $\hat{\theta}_r = \bar{\theta}$ . Then under  $r$ , any type of firm produces  $T/r$ . So

$$\Psi(x_r(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_{\theta}(x_r(\theta), \theta) d\theta = \Psi(T/r, \bar{\theta}) < \frac{T}{r} \Psi_x(T/r, \hat{\theta}_r) \leq \frac{T}{r} r = T$$

where the first inequality follows from convexity of  $\Psi$  in  $x$ , and the third line follows from the definition of  $\hat{\theta}_r$ . Thus, the transfer schedule given by (2) implements  $x_r$  without exhausting the budget.

Now, suppose  $\hat{\theta}_r < \bar{\theta}$ . Then

$$\begin{aligned}
\Psi(x_r(\underline{\theta})) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_{\theta}(x_r(\theta), \theta) d\theta &= \int_{\underline{\theta}}^{\hat{\theta}_r} \Psi_{\theta}(T/r, \hat{\theta}_r) d\theta + \int_{\hat{\theta}_r}^{\bar{\theta}} \Psi_{\theta}(x_r(\theta), \theta) d\theta \\
&= \Psi(T/r, \hat{\theta}_r) + \int_{\hat{\theta}_r}^{\bar{\theta}} [\Psi_{\theta}(x_r(\theta), \theta) + \Psi_x(x_r(\theta), \theta)x'_r(\theta)] d\theta \\
&\quad - \int_{\hat{\theta}_r}^{\bar{\theta}} \Psi_x(x_r(\theta), \theta)x'_r(\theta) d\theta \\
&= \Psi(T/r, \hat{\theta}_r) + \left[ \Psi(x_r(\bar{\theta}), \bar{\theta}) - \Psi(x_r(\hat{\theta}_r), \hat{\theta}_r) \right] - \int_{\hat{\theta}_r}^{\bar{\theta}} \Psi_x(x_r(\theta), \theta)x'_r(\theta) d\theta \\
&= \Psi(x_r(\bar{\theta}), \bar{\theta}) - r \int_{\hat{\theta}_r}^{\bar{\theta}} x'_r(\theta) d\theta \tag{11} \\
&= \Psi(x_r(\bar{\theta}), \bar{\theta}) - r \left( x_r(\bar{\theta}) - x_r(\hat{\theta}_r) \right) \\
&= \Psi(x_r(\bar{\theta}), \bar{\theta}) - r \left( x_r(\bar{\theta}) - T/r \right) \\
&= T + \Psi(x_r(\bar{\theta}), \bar{\theta}) - \Psi_x(x_r(\bar{\theta}), \bar{\theta})x_r(\bar{\theta}) < T
\end{aligned}$$

where (11) uses the fact that the agent equalizes marginal cost with the subsidy rate,  $\Psi_x(x, \theta) = r$ . ■

From Proposition 7, we thus see that for any outcome implementable by a linear subsidy schedule, there is a feasible schedule implementable in the mechanism design problem that yields more output. So the linear subsidy schedule is dominated by the optimal outcomes that are implementable in the mechanism design problem.

## 4 Linear Value

In the baseline model, we supposed that the principal's objective was independent of  $t$ , which we interpret as the principal having no value for the motivational resource. In certain applications, it may be reasonable to suppose the principal does have some value for the resource (or alternatively, obtaining the resource incurs a cost to the principal per unit of resource).

To motivate the model, we revisit our primary application, the emissions reduction example from before.

**Example (Emissions Reduction)** Let us return to the policymaker trying to induce a large number of firms to take reduce emissions. Now, in addition to providing a subsidies for emissions reductions, the policymaker could also use some of the funds to purchase

emissions offsetting credits in a competitive market, where the market price for offsetting a unit of emissions is some constant  $k$ . The policymaker would like to maximize the net emissions reduction across both channels, via offsetting or reducing emissions.

Formally, consider a variation of the design problem (1):

$$\begin{aligned}
\max_{x,t} \quad & \int_{\theta \in \Theta} [x(\theta) - kt(\theta)] f(\theta) d\theta & (12) \\
\text{subject to} \quad & t(\theta) - \Psi(x(\theta), \theta) \geq t(\theta') - \Psi(x(\theta'), \theta) & \forall \theta, \theta' \in \Theta & \text{(IC)} \\
& t(\theta) - \Psi(x(\theta), \theta) \geq 0 & \forall \theta \in \Theta & \text{(IR)} \\
& t(\theta) \leq T & \forall \theta \in \Theta & \text{(B)}
\end{aligned}$$

where  $k \geq 0$  is some constant. That is, the important distinction is in the objective, where there is an additional  $-kt(\theta)$  term that appears. This term could correspond to the principal obtaining some value from the unused  $T - t$  measure of resource, or the principal incurring some explicit cost to obtain the resource, with this problem degenerating to the original problem when  $k = 0$ . We can characterize the optimal solution as follows:

**Theorem 2** *An optimal mechanism  $(x, t)$  for design problem (12) exists, and it satisfies Lemma 1. There exists a nonnegative Lagrange multiplier  $\lambda$ , a nonnegative, absolutely continuous costate function  $\rho$ , and a nonnegative, absolutely continuous costate  $w$ , such that  $(x, t, \lambda, \rho, w)$  uniquely solve complementary slackness (3), border conditions (5) and (6), a modified costate evolution given by*

$$\dot{\rho}(\theta) = \lambda \Psi_{x\theta}(x^*(\theta), \theta) - f(\theta) \left( 1 - k \left( \Psi_x(x^*(\theta), \theta) + \Psi_{x\theta}(x^*(\theta), \theta) \frac{F(\theta)}{f(\theta)} \right) \right) - w(\theta) \quad (13)$$

that holds wherever  $\rho$  is differentiable, and an additional complementary slackness constraint

$$w(\theta) > 0 \implies x(\theta) = 0. \quad (14)$$

Further, the budget binding condition (7) must hold if  $\lambda > 0$ .

A couple of comments on the modifications are in order. Note that the altered costate evolution features two new terms. First,

$$\psi(x, \theta) = \Psi_x(x, \theta) + \Psi_{x\theta}(x, \theta) \frac{F(\theta)}{f(\theta)} \quad (15)$$

is the *virtual* marginal cost of the agent, similarly to the Myersonian virtual value. The

second,  $w$ , is a complementary slackness variable that arises because the lower bound on  $x$  can bind and types of agents can be excluded in this case.

## 4.1 Pooling

Given the additional term in the objective of the problem, we show that some features of the baseline solution still hold. Firstly, we have the following analogues of Proposition 1 and 2:

**Proposition 8** *If  $\lambda > 0$ , there exists a  $\hat{\theta} > \underline{\theta}$  such that the mechanism is constant on  $[\underline{\theta}, \hat{\theta}]$ . The transfer  $t^*$  is equal to  $T$  on this interval. Further, if  $\hat{\theta} < \bar{\theta}$ , then  $\hat{\theta}$  satisfies the following implicit equation:*

$$\frac{\Psi_{x\theta}(\underline{x}, \hat{\theta})}{\Psi_x(\underline{x}, \hat{\theta}) \left(1 - k\Psi_x(\underline{x}, \hat{\theta})\right)} = \frac{f(\hat{\theta})}{F(\hat{\theta})} \quad (16)$$

where  $\underline{x} = x(\underline{\theta})$ .

**Proof.** If  $\lambda > 0$ , since the budget must bind,  $x(\underline{\theta}) > 0$ , and so it follows that  $\rho(\underline{\theta}) > 0$ , so complementary slackness implies that  $x$  must be constant and positive. Integrating out the new costate evolution, we get that since  $x(\underline{\theta}) > 0$ ,  $w(\underline{\theta}) = 0$ , and so

$$\rho(\underline{\theta}) = \lambda\Psi_x(\underline{x}, \underline{\theta}) - F(\underline{\theta})(1 - k\Psi_x(\underline{x}, \underline{\theta}))$$

At the threshold  $\hat{\theta}$ ,  $\rho(\hat{\theta}) = 0$ , and by a similar argument as Proposition 2,  $\dot{\rho}(\hat{\theta}) = 0$ , so

$$\lambda\Psi_{x\theta}(\underline{x}, \hat{\theta}) = f(\hat{\theta}) \left(1 - k \left(\Psi_x(\underline{x}, \hat{\theta}) + \Psi_{x\theta}(\underline{x}, \hat{\theta}) \frac{F(\hat{\theta})}{f(\hat{\theta})}\right)\right)$$

$$\lambda + F(\hat{\theta})k = \frac{f(\hat{\theta})}{\Psi_{x\theta}(\underline{x}, \hat{\theta})} \left(1 - k\Psi_x(\underline{x}, \hat{\theta})\right)$$

Combining, we get (16). ■ With a linear value, the pooling region depends on whether the budget constraint binds. For sufficiently high values of  $k$  or high values of  $T$ , the optimal schedule may be such that the principal's optimal schedule in the absence of the budget constraint is actually feasible. Note that this could not happen in the baseline model, where the budget always binds; since the principal did not obtain any value from the resource in the baseline case, the baseline optimization problem without a budget constraint is unbounded (giving infinite resource to the agent).

Also, with the principal having a value for the resource, it is possible that a positive measure of types are excluded from the mechanism. We provide a sufficient condition for

such exclusion to happen.

**Proposition 9** *Suppose  $\Psi_{x\theta}(x, \theta)/f(\theta)$  is increasing in  $\theta$  and  $k > 1/\psi(0, \bar{\theta})$ . Then the optimal mechanism excludes a positive measure of types.*

**Proof.** Firstly, if  $k > 1/\Psi_x(0, \underline{\theta}) = 1/\psi(0, \underline{\theta})$ , it is not hard to check that the mechanism is degenerate:  $x(\cdot) = t(\cdot) = 0$  (since intuitively, even the lowest cost type  $\underline{\theta}$  has a marginal cost of more than  $1/k$ ). So suppose  $k < 1/\psi(0, \underline{\theta})$ . By the costate evolution (13), on a sequence of  $\theta \rightarrow \tilde{\theta}$  from the left, continuity of  $x$  implies that  $x$  is positive and decreasing. So  $w$  must be 0 and  $\rho = 0$ , so

$$0 = \lambda \Psi_{x\theta}(0, \tilde{\theta}) - f(\tilde{\theta}) \left( 1 - k \left( \Psi_x(0, \tilde{\theta}) + \Psi_{x\theta}(0, \tilde{\theta}) \frac{F(\tilde{\theta})}{f(\tilde{\theta})} \right) \right)$$

$$1 - k\psi(0, \tilde{\theta}) = \lambda \frac{\Psi_{x\theta}(0, \tilde{\theta})}{f(\tilde{\theta})}$$

Note by our assumptions, the RHS is increasing, and the LHS is decreasing in  $\tilde{\theta}$ . Further, the LHS goes from positive to negative from  $\underline{\theta}$  to  $\bar{\theta}$  and the RHS is positive, so the intermediate value theorem implies there exists a  $\tilde{\theta} < \bar{\theta}$  such that the equality holds. Then the optimal mechanism excludes all types  $\theta > \tilde{\theta}$ . ■

Intuitively, this arises because the principal has a positive value for the resource; thus, intuitively, if some type of agent has an information rent-adjusted marginal cost of  $x$  as more than  $1/k$ , the principal finds it best to exclude this agent.

## 4.2 Comparison to Naive Benchmark

Recall in the baseline model, the problem was unbounded without the budget constraint, since the principal did not value the resource. With a linear value for the resource  $k$ , the problem is no longer unbounded; thus, we can compare the optimal mechanism to a naive solution, where the principal designs the mechanism (i.e. solves (12)) without factoring the budget constraint, and instead “runs out” of money if the mechanism is supposed to offer  $t > T$ .

The solution to the naive problem corresponds to the problem where the Lagrange multiplier  $\lambda$  is set to zero. As a result, from Theorem 2, the condition pinning down  $x$  when not in a flat region is missing a  $\lambda \Psi_{x\theta}$  term, which implies that  $x$  is larger when the budget constraint does not bind. In consequence, the budget binds at a higher type than in the

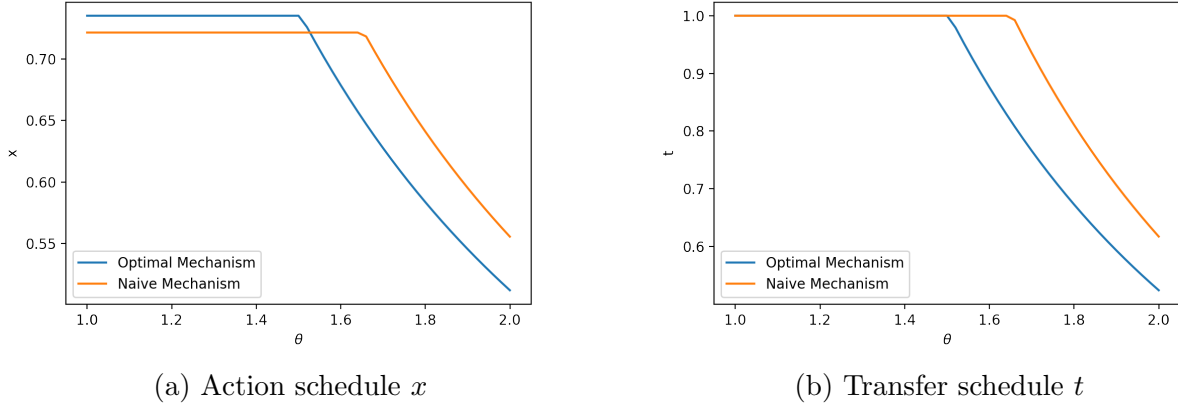


Figure 3: Comparison of naive and optimal mechanisms. Values are distributed uniformly on  $[1, 2]$ ,  $k = 0.3$ , the cost function is  $\Psi(x, \theta) = \theta x^2$ , and the budget is  $T = 1$ . The naive mechanism uses more of the resource in expectation, and obtains less  $x$  from more efficient types than the optimal mechanism.

optimal mechanism (i.e. the naive solution spends more of the resource in expectation), and the naive solution obtains a lower  $x$  from the most efficient types of agents. Figure 3 illustrates this with an example, where values are distributed uniformly on  $[1, 2]$ ,  $k = 0.3$ , the cost function is  $\Psi(x, \theta) = \theta x^2$ , and the budget is  $T = 1$ .

### 4.3 Comparative Statics

We characterize how the optimal mechanism changes in terms of  $k$  in the following result. Define  $\bar{x} = x(\bar{\theta})$ ; equivalently,  $\bar{x}$  is the  $x$  needed for the agent to get  $T$  in the mechanism.

**Proposition 10** *As  $k$  increases,  $\bar{x}$  increases and  $\lambda$  decreases:  $d\bar{x}/dk > 0$  and  $d\lambda/dk < 0$ .*

We leave the proof to the appendix; the key idea is that we identify three equations from Theorem 2 that pin down  $(\bar{x}, \hat{\theta}, \lambda)$ , and apply the implicit function theorem to derive the comparative statics. To illustrate the implications of the proposition, Figure 4 plots the optimal mechanism for  $\Psi(x, \theta) = \theta x^2$ ,  $T = 2$ , and  $\theta$  uniformly distributed on  $[0, 1]$ . As  $k$  increases, the  $x$  required to receive the full transfer  $T$  in the optimal mechanism decreases, and the mechanism excludes more types. However, since  $\lambda$  increases in  $k$ , it also is true that  $x(\bar{\theta})$  decreases; that is, the principal becomes more lenient on the less efficient types, reducing  $x$  for higher types.

Revisiting the emissions reduction example, the comparative statics in  $k$  can be interpreted as changes in the optimal subsidy schedule as the emissions offset market becomes



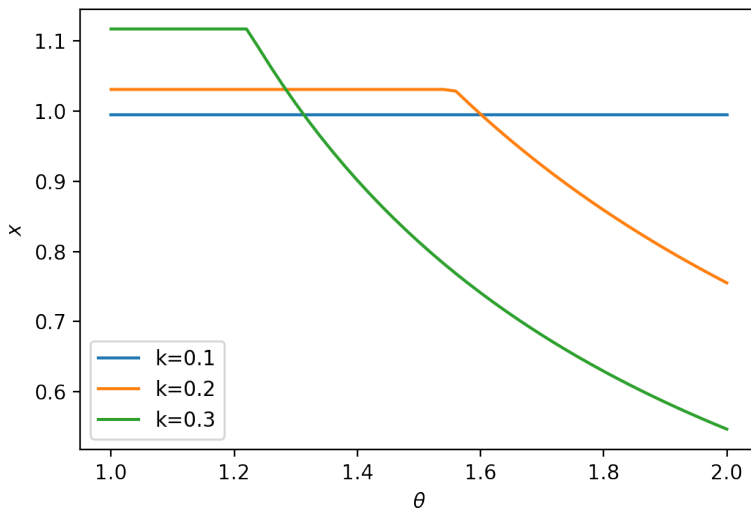


Figure 4: Optimal mechanisms for  $\Psi(x, \theta) = \theta x^2$ ,  $T = 2$ ,  $\theta$  uniformly distributed on  $[1, 2]$ , as  $k$  varies.

more or less effective. The comparative statics highlight the role of the threshold  $\hat{\theta}$ : as the outside option increases, the policymaker becomes more demanding (asks for more aggressive emissions reductions  $x$ ) from firms that are more efficient than the threshold, but becomes more lenient (asks for less emissions reductions  $x$ ) from types that are less efficient than the threshold.

## 5 Extensions

In this section, we discuss some other extensions of our model. We first consider a general arbitrary convex resource cost  $c(t)$ , as opposed to the linear cost examined in the previous section. We show that the optimal control technique extends to this case as well. In the second subsection, we characterize the problem under an ex-ante budget constraint, and contrast the optimal mechanism to the solution characterized by Theorem 1.

## 5.1 Generalized Transfer Value

Suppose the principal has an arbitrary continuous, weakly convex resource value  $c(t)$ . That is, suppose the designer's problem is now

$$\begin{aligned} \max_{x,t} \quad & \int_{\theta \in \Theta} [x(\theta) - c(t(\theta))] f(\theta) d\theta & (17) \\ \text{subject to} \quad & t(\theta) - \Psi(x(\theta), \theta) \geq t(\theta') - \Psi(x(\theta'), \theta) & \forall \theta, \theta' \in \Theta & \text{(IC)} \\ & t(\theta) - \Psi(x(\theta), \theta) \geq 0 & \forall \theta \in \Theta & \text{(IR)} \\ & t(\theta) \leq T & \forall \theta \in \Theta & \text{(B)} \end{aligned}$$

The optimal mechanism characterization for this extension follows the same form as before:

**Proposition 11** *An optimal mechanism  $(x, t)$  for design problem (17) exists and is unique. Further, there exists nonnegative, absolutely continuous costate functions  $\rho_x, \rho_t, \lambda$  such that  $x, t, \rho_x, \rho_t, \lambda$  satisfy Lemma 1 and solve*

$$\rho_x(\theta) + \rho_t(\theta)\Psi_x(x(\theta), \theta) > 0 \implies \dot{x}(\theta) = 0 \quad (18)$$

$$\dot{\rho}_x(\theta) = -\rho_t(\theta)\dot{x}(\theta)\Psi_{xx}(x(\theta), \theta) - f(\theta) \quad (19)$$

$$\lambda(\theta) > 0 \implies t(\theta) = T \quad (20)$$

$$\dot{\rho}_t(\theta) = c'(t(\theta))f(\theta) - \lambda(\theta) \quad (21)$$

$$\rho_x(\bar{\theta}) + \rho_t(\bar{\theta})\Psi_x(x(\bar{\theta}), \bar{\theta}) = 0 \quad (22)$$

$$\rho_x(\underline{\theta}), \rho_t(\underline{\theta}) = 0 \quad (23)$$

Note that even though the state space  $(x, t)$  is two-dimensional, the variable  $u = \dot{x}$  jointly controls both state variables, so the control is still one-dimensional; the boundary condition (22) then comes from the orthogonality constraint for the boundary surface at  $\bar{\theta}$  implied by the binding IR constraint  $t(\bar{\theta}) = \Psi(x(\bar{\theta}), \bar{\theta})$ . Hence, the optimal control technique we use for the general case extends when we have a more general resource value for the principal.

## 5.2 Ex Ante Budget Constraints

An alternative formulation might require that the budget only needs to bind ex-ante (i.e., the expected transfer cannot exceed  $T$ ). That is, consider the following maximization

$$\begin{aligned}
\max_{x,t} \quad & \int_{\theta \in \Theta} x(\theta) f(\theta) d\theta & (24) \\
\text{subject to} \quad & t(\theta) - \Psi(x(\theta), \theta) \geq t(\theta') - \Psi(x(\theta'), \theta) & \forall \theta, \theta' \in \Theta & \text{(IC)} \\
& t(\theta) - \Psi(x(\theta), \theta) \geq 0 & \forall \theta \in \Theta & \text{(IR)} \\
& \int_{\underline{\theta}}^{\bar{\theta}} t(\theta) f(\theta) d\theta \leq T & \forall \theta \in \Theta & \text{(B')}
\end{aligned}$$

The optimal solution is characterized by the following proposition.

**Proposition 12** *An optimal mechanism  $(x, t)$  for (24) exists, and is unique. Further, there exists a nonnegative Lagrange multiplier  $\lambda$  and a nonnegative, absolutely continuous costate functions  $\rho$  such that  $(x, \lambda, \rho)$  is unique solution that satisfies Lemma 1 and complementary slackness (3) holds, with the modified costate evolution equation*

$$\dot{\rho}(\theta) = f(\theta) \left( 1 - \lambda \left( \Psi_x(x(\theta), \theta) + \Psi_{x\theta}(x(\theta), \theta) \frac{F(\theta)}{f(\theta)} \right) \right) \quad (25)$$

*upper boundary condition (5), modified lower bound condition  $\rho(\underline{\theta}) = 0$ , and modified budget binding condition*

$$\int f(\theta) \left( \Psi(x(\theta), \theta) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds \right) d\theta = T \quad (26)$$

Notably, the lower boundary condition is modified to  $\rho(\underline{\theta}) = 0$ , and hence the pooling region that appears for the ex-post budget constraint does *not* appear here. Thus, the key feature of our model, the pooling region where the principal trades off between utilizing the budget less or providing information rents, does not emerge in the case where the budget is only required ex-ante.

## 6 Conclusion

We analyze a mechanism design model where a principal interacts with an agent, where the principal has a budget of a resource it can provide as an incentive. The design problem fea-

tures an ex-post budget constraint and a first-order objective maximizing the agent's action. We derive the optimal mechanism, and show that the mechanism pools sufficiently efficient types and withholds the budget from less efficient types. The pooling region threshold is crucial to understanding how the mechanism changes when the principal has a resource value; as the resource value increases, the principal demands more or less from agents depending on their type relative to the threshold.

There are several promising avenues for future research. First, extending our framework to multiple agents could more accurately model certain scenarios, such as procurement from an oligopolistic industry. Another direction involves exploring dynamic incentives within this paper's framework. Adding temporal dimensions could help illustrate how optimal mechanisms evolve over time, with changing resource availability and agent behavior. Additionally, our model's applicability to various domains suggests insights for real-world implementations and empirical validations. Lastly, extending the model to allow for more general uncertainty, such as an uncertain resource value or a richer space of cost functions, could enhance the practical relevance of this paper's insights.

## References

- Abebe, R. and Goldner, K. (2018), ‘Mechanism design for social good’, *AI Matters* **4**(3), 27–34.
- Alonso, R. and Matouschek, N. (2008), ‘Optimal delegation’, *The Review of Economic Studies* **75**(1), 259–293.
- Amador, M. and Bagwell, K. (2013), ‘The theory of optimal delegation with an application to tariff caps’, *Econometrica* **81**(4), 1541–1599.  
**URL:** <http://www.jstor.org/stable/23524185>
- Amador, M., Bagwell, K. and Frankel, A. (2018), ‘A note on interval delegation’, *Economic Theory Bulletin* **6**(2), 239–249.
- Amanatidis, G., Kleer, P. and Schäfer, G. (2019), Budget-feasible mechanism design for non-monotone submodular objectives: Offline and online, *in* ‘Proceedings of the 2019 ACM Conference on Economics and Computation’, pp. 901–919.
- Archak, N. and Sundararajan, A. (2009), ‘Optimal design of crowdsourcing contests’, *ICIS 2009 proceedings* p. 200.
- Babaioff, M., Immorlica, N., Li, Y. and Lucier, B. (2023), ‘Making auctions robust to after-markets’, *ITCS* .
- Baron, D. P. and Myerson, R. B. (1982), ‘Regulating a monopolist with unknown costs’, *Econometrica* **50**(4), 911–930.  
**URL:** <http://www.jstor.org/stable/1912769>
- Chawla, S., Hartline, J. D. and Sivan, B. (2019), ‘Optimal crowdsourcing contests’, *Games and Economic Behavior* **113**, 80–96.
- Che, Y.-K. and Gale, I. (2003), ‘Optimal design of research contests’, *American Economic Review* **93**(3), 646–671.
- Dai Li, W., Immorlica, N. and Lucier, B. (2022), Contract design for afforestation programs, *in* ‘Web and Internet Economics: 17th International Conference, WINE 2021, Potsdam, Germany, December 14–17, 2021, Proceedings’, Springer, pp. 113–130.
- Ghosh, A. and Kleinberg, R. (2016), ‘Optimal contest design for simple agents’, *ACM Transactions on Economics and Computation (TEAC)* **4**(4), 1–41.

- Goldner, K., Immorlica, N. and Lucier, B. (2019), ‘Reducing inefficiency in carbon auctions with imperfect competition’, *ITCS* .
- Gravin, N., Jin, Y., Lu, P. and Zhang, C. (2020), ‘Optimal budget-feasible mechanisms for additive valuations’, *ACM Transactions on Economics and Computation (TEAC)* **8**(4), 1–15.
- Greaker, M. and Popp, D. (2022), Environmental economics, regulation, and innovation, Working Paper 30415, National Bureau of Economic Research.  
**URL:** <http://www.nber.org/papers/w30415>
- Grossman, S. J. and Hart, O. D. (1983), ‘An analysis of the principal-agent problem’, *Econometrica* **51**(1), 7–45.  
**URL:** <http://www.jstor.org/stable/1912246>
- Halac, M. and Yared, P. (2020), ‘Commitment versus flexibility with costly verification’, *Journal of Political Economy* **128**(12), 4523–4573.
- Holmström, B. (1979), ‘Moral hazard and observability’, *The Bell journal of economics* pp. 74–91.
- Holmstrom, B. (1980), On the theory of delegation, Technical report, Discussion Paper.
- Kamien, M. I. and Schwartz, N. L. (1971), ‘Sufficient conditions in optimal control theory’, *Journal of Economic Theory* **3**(2), 207–214.  
**URL:** <https://www.sciencedirect.com/science/article/pii/0022053171900184>
- Kang, Z. Y. (2022), The public option and optimal redistribution, Technical report, Working paper.
- Laffont, J.-J. and Tirole, J. (1986), ‘Using cost observation to regulate firms’, *Journal of Political Economy* **94**(3), 614–641.  
**URL:** <http://www.jstor.org/stable/1833051>
- Leonardi, S., Monaco, G., Sankowski, P. and Zhang, Q. (2017), Budget feasible mechanisms on matroids, in ‘International conference on integer programming and combinatorial optimization’, Springer, pp. 368–379.
- Moldovanu, B. and Sela, A. (2001), ‘The optimal allocation of prizes in contests’, *American Economic Review* **91**(3), 542–558.

- Moldovanu, B. and Sela, A. (2006), ‘Contest architecture’, *Journal of Economic Theory* **126**(1), 70–96.
- Singer, Y. (2010), Budget feasible mechanisms, in ‘2010 IEEE 51st Annual Symposium on foundations of computer science’, IEEE, pp. 765–774.
- Singer, Y. (2014), ‘Budget feasible mechanism design’, *ACM SIGecom Exchanges* **12**(2), 24–31.
- Toikka, J. (2011), ‘Ironing without control’, *Journal of Economic Theory* **146**(6), 2510–2526.  
**URL:** <https://www.sciencedirect.com/science/article/pii/S0022053111000871>
- Walczak, S. (2001), ‘Well-posed and ill-posed optimal control problems’, *Journal of Optimization Theory and Applications* **109**(1), 169–185.

# A Appendix: Supplemental Material

## A.1 Omitted Proofs

**Proof of Lemma 1.** Consider any  $x$  that satisfies the two conditions. Take  $t$  given by (2). Since  $\Psi_\theta$  is nonnegative,  $t(\theta) \geq \Psi(x(\theta), \theta)$ , so IR is satisfied. Now, we show that  $t$  must be nonincreasing. Suppose  $\theta \leq \theta'$ . Then

$$\begin{aligned}
 t(\theta) &= \Psi(x(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \\
 &= \Psi(x(\theta), \theta) + \int_{\theta}^{\theta'} \Psi_\theta(x(s), s) ds + \int_{\theta'}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \\
 &\geq \Psi(x(\theta), \theta) + \int_{\theta}^{\theta'} \Psi_\theta(x(\theta'), s) ds + \int_{\theta'}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \\
 &= \Psi(x(\theta), \theta) + \Psi(x(\theta'), \theta') - \Psi(x(\theta'), \theta) + \int_{\theta'}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \\
 &= [\Psi(x(\theta), \theta) - \Psi(x(\theta'), \theta)] + t(\theta')
 \end{aligned}$$

where we used the fact that  $x$  is nonincreasing and supermodularity of  $\Psi$  in the 3rd step. Also, since  $x$  is nonincreasing, and  $\Psi$  is increasing in its first argument, the bracketed part is nonnegative, and so we have that  $t$  is nonincreasing. Hence, to check the budget constraint, we just have to check that the budget constraint holds at  $t(\underline{\theta})$ :

$$t(\underline{\theta}) = \Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}} \Psi_\theta(x(s), s) ds = \int_{\underline{\theta}} \Psi_\theta(x(s), s) ds \leq T$$

Finally, to check that IC is satisfied, consider IC for type  $\theta$ . We consider the deviation of  $\theta$  to  $\theta'$ . We first show that  $\theta$  does not want to deviate up, or  $\theta \leq \theta'$ :

$$\begin{aligned}
 t(\theta) - \Psi(x(\theta), \theta) &= \int_{\theta}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \\
 &= \int_{\theta}^{\theta'} \Psi_\theta(x(s), s) ds + \int_{\theta'}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \\
 &= \int_{\theta}^{\theta'} \Psi_\theta(x(s), s) ds + t(\theta') - \Psi(x(\theta'), \theta') \\
 &\geq \int_{\theta}^{\theta'} \Psi_\theta(x(\theta'), s) ds + t(\theta') - \Psi(x(\theta'), \theta') \\
 &= \Psi(x(\theta'), \theta') - \Psi(x(\theta'), \theta) + t(\theta') - \Psi(x(\theta'), \theta') \\
 &= t(\theta') - \Psi(x(\theta'), \theta)
 \end{aligned}$$



Now we check that  $\theta$  does not want to deviate down to type  $\theta' \leq \theta$ :

$$\begin{aligned}
t(\theta) - \Psi(x(\theta), \theta) &= \int_{\theta}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds \\
&= - \int_{\theta'}^{\theta} \Psi_{\theta}(x(s), s) ds + \int_{\theta'}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds \\
&= - \int_{\theta'}^{\theta} \Psi_{\theta}(x(s), s) ds + t(\theta') - \Psi(x(\theta'), \theta') \\
&\geq - \int_{\theta'}^{\theta} \Psi_{\theta}(x(\theta'), s) ds + t(\theta') - \Psi(x(\theta'), \theta') \\
&= -\Psi(x(\theta'), \theta) + \Psi(x(\theta'), \theta') + t(\theta') - \Psi(x(\theta'), \theta') \\
&= t(\theta') - \Psi(x(\theta'), \theta)
\end{aligned}$$

So IC is satisfied. Thus, if  $x$  satisfies the conditions, it is feasible.

Now, we show the two conditions are necessary. Consider any feasible  $x : \Theta \rightarrow \mathbb{R}_+$ . Suppose  $\theta \leq \theta'$ , and  $\theta, \theta' \in \Theta$ . From rewriting the IC constraints, we get that

$$\Pi(x(\theta); \theta) - \Pi(x(\theta'); \theta) \geq t(\theta') - t(\theta)$$

$$\Pi(x(\theta'); \theta') - \Pi(x(\theta); \theta') \geq t(\theta) - t(\theta')$$

Note that the  $\Pi_{LF}(\theta)$  and  $\Pi_{LF}(\theta')$  terms will cancel on the left hand sides. Adding these together, we get

$$-\Psi(x(\theta), \theta) + \Psi(x(\theta'), \theta) - \Psi(x(\theta'), \theta') + \Psi(x(\theta), \theta') \geq 0$$

Rearranging

$$\Psi(x(\theta'), \theta) + \Psi(x(\theta), \theta') \geq \Psi(x(\theta), \theta) + \Psi(x(\theta'), \theta')$$

And so supermodularity of  $\Psi$  implies that  $x(\theta') \leq x(\theta)$ , and so the first condition must hold. To show that the transfers must take the form given, define the interim utility

$$U(\theta) = t(\theta) - \Psi(x(\theta), \theta)$$

IC implies that

$$U(\theta) = \max_{\theta'} t(\theta') - \Psi(x(\theta'), \theta)$$

By the envelope theorem,

$$U'(\theta) = -\Psi_\theta(x(\theta), \theta)$$

$$U(\theta) = U(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} \Psi_\theta(x(s), s) ds$$

So transfers must be of the form

$$t(\theta) = U(\theta)$$

$$= U(\bar{\theta}) + \Psi(x(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_\theta(x(s), s) ds$$

In particular, the transfer for type  $\underline{\theta}$  satisfies

$$T \geq t(\underline{\theta})$$

$$= U(\bar{\theta}) + \Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_\theta(x(s), s) ds$$

$$\geq \Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_\theta(x(s), s) ds$$

since  $U(\bar{\theta}) \geq 0$  by IR. Hence, the two conditions are necessary and sufficient for a feasible schedule. ■

**Proof of Theorem 1.** By Lemma 1, we can rewrite the design problem as

$$\begin{aligned} \max_x \quad & \int_{\Theta} x(\theta) f(\theta) d\theta & (27) \\ \text{subject to} \quad & \Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \leq T & \text{(Normalization)} \\ & x \text{ nonincreasing} & \text{(Monotonicity)} \end{aligned}$$

The Lagrangian relaxation is

$$\begin{aligned} \max_x \quad & \int_{\Theta} x(\theta) f(\theta) d\theta + \lambda \left( T - \Psi(x(\underline{\theta}), \underline{\theta}) - \int_{\underline{\theta}}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \right) \\ \text{subject to} \quad & x \text{ nonincreasing} & \text{(Monotonicity)} \end{aligned}$$

By monotonicity,  $x$  must be differentiable almost everywhere. The monotonicity constraint then requires that wherever  $x$  is differentiable,  $\dot{x} \leq 0$ . Letting  $u = \dot{x}$  and rewriting the

objective function, the problem becomes

$$\begin{aligned} \max_x \quad & \int_{\Theta} (x(\theta)f(\theta) - \lambda\Psi_{\theta}(x(\theta), \theta)) \, d\theta + \lambda T - \lambda\Psi(x(\underline{\theta}), \underline{\theta}) \quad (28) \\ \text{subject to} \quad & \dot{x} = u \leq 0 \quad (\text{Monotonicity}) \end{aligned}$$

The problem now is an optimal control problem with a scrap value constraint;  $u$  is the control variable, and  $x$  is the state variable. To argue that this rewriting is without loss, we adopt the Laffont and Tirole (1986) argument. The original problem optimizes over the space of nonincreasing  $x$ . Given an optimal solution  $x^*$  to (28) that is absolutely continuous and nonincreasing,  $x^*$  must also be optimal over the original space, since the space of absolutely continuous nonincreasing functions is dense in the space of nonincreasing functions under the weak norm topology, and the maximization objective is continuous.

We first show that an optimal mechanism  $(x, t)$  satisfies the conditions (3)-(7). For a fixed  $\lambda > 0$ , Pontryagin's maximum principle implies that any optimal solution admits a nonnegative, absolutely continuous control variable such that (3) - (6) are true. Further, note that if  $\lambda = 0$ , (6) implies that  $\rho(\underline{\theta}) = 0$ , but (4) implies that  $\dot{\rho}(\underline{\theta}) = -f(\underline{\theta}) < 0$ , which is impossible since  $\rho$  must be nonnegative. Hence,  $\lambda > 0$ , so by Lagrangian duality we must have that the budget binds, so (7) holds.

Second, we show that any solution that satisfies (3)-(7) must be optimal. Note that fixing  $\lambda$ , the maximized Hamiltonian corresponding to the control problem is concave in the state variable  $x$ , so the Arrow sufficiency condition holds.<sup>6</sup> Therefore, the solution maximizes the Lagrangian and the budget constraint (7) binds, the solution must be optimal in the original design problem.

Finally, we argue that an optimal mechanism exists (uniqueness holds from Corollary 1). To show an optimal mechanism exists, we show that a solution to conditions (3)-(7) exists. Fixing  $\lambda$ , the conditions (3)-(6) admits a solution<sup>7</sup>  $(x_{\lambda}, \rho_{\lambda})$  by the Caratheodory existence theorem. Define  $t_{\lambda}$  as the transfer schedule for  $x_{\lambda}$  by equation (2). We now argue that a  $\lambda$  exists such that (7) holds. To show such a  $\lambda$  exists, we argue that  $t_{\lambda}(\underline{\theta})$  is continuous in  $\lambda$ ,  $t_{\lambda}(\underline{\theta}) \rightarrow 0$  for  $\lambda \rightarrow \infty$ , and  $t_{\lambda}(\underline{\theta}) \rightarrow \infty$  for  $\lambda \rightarrow 0$ ; the intermediate value theorem then implies the result.

For continuity of  $t_{\lambda}(\underline{\theta})$ , note that taking any sequence  $\{\lambda_k\} \rightarrow \lambda$ , the family of control problems (28) admits an optimal  $x_{\lambda_k}$ , which must converge  $x_{\lambda_k} \rightarrow x_{\lambda}$  weakly in  $H^1$  by

<sup>6</sup>See Kamien and Schwartz (1971) for details.

<sup>7</sup>More precisely, a solution in the extended sense, i.e. absolutely continuous and satisfying the differential equations almost everywhere.

Theorem 3.1 from Walczak (2001). Hence, by construction of  $t_\lambda$ ,  $t_{\lambda_k}(\underline{\theta}) \rightarrow t_\lambda(\underline{\theta})$ .

To characterize the boundary, note that as  $\lambda \rightarrow \infty$ , the limit control problem is simply

$$\begin{aligned} \max_x \quad & \lambda \left( \int_{\Theta} -\Psi_\theta(x(\theta), \theta) d\theta + T - \Psi(x(\underline{\theta}), \underline{\theta}) \right) + o(\lambda) \\ \text{subject to} \quad & \dot{x} = u \leq 0 \end{aligned} \quad (\text{Monotonicity})$$

which has an optimal solution of  $x = 0$  everywhere in the limit  $\lambda \rightarrow \infty$ , and hence  $t_\lambda(\underline{\theta}) \rightarrow 0$  as  $\lambda \rightarrow \infty$ . Finally, note that as  $\lambda \rightarrow 0$ , the problem becomes

$$\begin{aligned} \max_x \quad & \int_{\Theta} x(\theta) f(\theta) d\theta \\ \text{subject to} \quad & \dot{x} = u \leq 0 \end{aligned} \quad (\text{Monotonicity})$$

which is unbounded, and hence  $t_\lambda(\underline{\theta}) \rightarrow \infty$  as  $\lambda \rightarrow 0$ . Thus, by the intermediate value theorem, there exists a  $\lambda$  such that  $t_\lambda(\underline{\theta}) = T$ . ■

**Proof of Proposition 5.** As in the proof of Theorem 1, we write out the Lagrangian:

$$\begin{aligned} \max_x \quad & \int_{\Theta} x(\theta) f(\theta) d\theta + \lambda \left( T - \underline{\theta} \Gamma(x(\underline{\theta})) - \int_{\underline{\theta}}^{\bar{\theta}} \Gamma(x(s)) ds \right) \\ \text{subject to} \quad & x \text{ nonincreasing} \end{aligned} \quad (\text{Monotonicity})$$

Define  $f(\theta) = 0$  for  $\theta < \underline{\theta}$ . Then note that the optimization can be rewritten equivalently as

$$\begin{aligned} \max_x \quad & \int_0^{\bar{\theta}} x(\theta) f(\theta) d\theta + \lambda \left( T - \int_0^{\bar{\theta}} \Gamma(x(s)) ds \right) \\ \text{subject to} \quad & x \text{ nonincreasing} \end{aligned} \quad (\text{Monotonicity})$$

To handle the monotonicity constraint, we use a generalized ironing technique. By Theorem 4.4 in Toikka (2011), the optimal  $x(\theta)$  then maximizes  $\bar{J}(\cdot, \theta; \lambda)$  at the optimal  $\lambda$ , where

$$\bar{J}(x, \theta; \lambda) = \int_0^s \frac{\partial}{\partial \theta^-} \text{cav}_\theta [F(\theta) - \lambda \theta \Gamma'(x)] d\theta$$

and  $\text{cav}_\theta$  denote the concavification operation in  $\theta$ , where the differentiation is taken from the left (as the concavification need not be smooth in  $\theta$ ). Note that since  $(\theta - \underline{\theta})$  is linear in

$\theta$ , we can rewrite  $\bar{J}$  as

$$\begin{aligned}\bar{J}(x, \theta; \lambda) &= \int_0^x \frac{\partial}{\partial \theta^-} \text{cav}_\theta [F(\theta)] d\theta - \lambda \Gamma'(x) \\ &= \tilde{f}(\theta) - \lambda \Gamma'(x)\end{aligned}$$

Hence, the problem is reduced to a point-wise optimization at each  $\theta$ : inverting  $\Gamma'$  implies that

$$x^*(\theta) = (\Gamma')^{-1} \left( \frac{\tilde{f}(\theta)}{\lambda} \right)$$

It remains to argue that a  $\lambda$  exists such that the budget constraint binds. Note that as  $\lambda \rightarrow 0$ ,  $x^* \rightarrow \infty$ , and as  $\lambda \rightarrow \infty$ ,  $x^* \rightarrow 0$ . Hence, the intermediate value theorem implies some finite value of  $\lambda$  makes the budget constraint bind. ■

**Proof of Theorem 2.** The analysis proceeds exactly as in Theorem 1; in this case, we plug in the transfers from Lemma 1 into the objective and integrate by parts. The Hamiltonian becomes

$$\mathcal{H} = x f(\theta) - \lambda \Psi_\theta(x, \theta) - k \Psi(x, \theta) f(\theta) - k \Psi_\theta(x, \theta) F(\theta) + \rho u$$

Applying Pontryagin, the costate evolution (13) follows. ■

**Proof of Proposition 10.** Since  $\Psi_{x\theta}(x, \theta)/f(\theta)$  is weakly increasing in  $\theta$ , it is easy to check that the  $x$  that solves  $\dot{\rho} = 0$  is nonincreasing. First, define the subproblem on  $[\hat{\theta}, \bar{\theta}]$  with no initial conditions:

$$\max_{x, t} \int_{\hat{\theta}}^{\bar{\theta}} \left[ x(\theta) - k \left( \Psi(x(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_\theta(x(s), s) ds \right) - \lambda \Psi_\theta(x(\theta), \theta) \right] f(\theta) d\theta \quad (29)$$

subject to  $x$  nonincreasing (Monotonicity)

Note that any solution to the conditions in Theorem 2 restricted to  $[\hat{\theta}, \bar{\theta}]$  also solve the Pontryagin necessary conditions for this subproblem, and hence by Arrow's sufficiency theorem, the subsolution must also optimize this subproblem.

Then by Proposition 8, the objects  $\bar{x}$ , the threshold type  $\hat{\theta}$ , and  $\lambda$  jointly solve the

following three equations:

$$\begin{aligned}
G_1(\bar{x}, \hat{\theta}, \lambda; k) &= \Psi(\bar{x}, \hat{\theta}) + \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds - T &= 0 \\
G_2(\bar{x}, \hat{\theta}, \lambda; k) &= \lambda \Psi_x(\bar{x}, \hat{\theta}) - F(\hat{\theta})(1 - k \Psi_x(\bar{x}, \hat{\theta})) &= 0 \\
G_3(\bar{x}, \hat{\theta}, \lambda; k) &= \lambda \Psi_x(\bar{x}, \hat{\theta}) + k(\Psi_x(\bar{x}, \hat{\theta})f(\hat{\theta}) + \Psi_{x\theta}(\bar{x}, \hat{\theta})F(\hat{\theta})) - f(\hat{\theta}) &= 0
\end{aligned}$$

where  $x(s)$  solves the point-wise first-order condition from the maximized Hamiltonian of the subproblem (29) at the optimal  $u$ :

$$H'(x(s), s, \lambda; k) = \lambda \Psi_x(x(s), s) + k(\Psi_x(x(s), s)f(s) + \Psi_{x\theta}(x(s), s)F(s)) - f(s) = 0$$

We apply the implicit function theorem to find the derivatives. The Jacobean of the system evaluated at  $(\bar{x}, \hat{\theta}, \lambda)$  is

$$J = \begin{bmatrix} \Psi_x & 0 & \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \frac{dx(s)}{d\lambda} ds \\ (\lambda + kF)\Psi_{xx} & 0 & \Psi_x \\ k(\Psi_{xx}f + \Psi_{xx\theta}F) + \lambda \Psi_{x\theta} & k(2\Psi_{x\theta}f + \Psi_x f' + \Psi_{x\theta\theta}F) + \lambda \Psi_{x\theta\theta} - f' & \Psi_{x\theta} \end{bmatrix}$$

where the functions are understood to be evaluated at  $(\bar{x}, \hat{\theta}, \lambda)$  where appropriate. The Jacobean is invertible since the partial  $\partial G_1/\partial \lambda$  is negative because  $dx(s)/d\lambda$  is negative, and the other nonzero entries are positive. Note that

$$\frac{\partial G_1}{\partial k} = \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \frac{dx(s)}{dk} ds < 0$$

$$\frac{\partial G_2}{\partial k} = \Psi_x F$$

Then from the implicit function theorem, we have

$$\begin{aligned}
\frac{d\bar{x}}{dk} &= \frac{\Psi_x \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \left( \frac{dx(s)}{d\lambda} F(\hat{\theta}) - \frac{dx(s)}{dk} \right) ds}{(\Psi_x)^2 - (\lambda + kF)\Psi_{xx} \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \frac{dx(s)}{d\lambda} ds} \\
\frac{d\lambda}{dk} &= \frac{(\lambda + kF)\Psi_{xx}F \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \frac{dx(s)}{dk} ds - (\Psi_x)^2 F}{(\Psi_x)^2 - (\lambda + kF)\Psi_{xx} \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \frac{dx(s)}{d\lambda} ds}
\end{aligned}$$

From the observation that  $dx(s)/d\lambda < 0$ , it follows that the denominators are strictly positive, so the sign of each of these derivatives is the same as the sign of the numerator. It

follows then that  $d\lambda/dk < 0$ , since  $dx(s)/dk$  is negative. To determine the sign of  $d\bar{x}/dk$ , we apply the implicit function theorem for  $dx(s)/d\lambda$  and  $dx(s)/dk$ , and obtain that:

$$\begin{aligned}
& \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \left( \frac{dx(s)}{d\lambda} F(\hat{\theta}) - \frac{dx(s)}{dk} \right) ds \\
&= \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \left( \frac{-\frac{\partial H}{\partial \lambda}}{\frac{\partial H}{\partial x}} F(\hat{\theta}) - \frac{-\frac{\partial H}{\partial k}}{\frac{\partial H}{\partial x}} \right) ds \\
&= \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \left( \frac{-\Psi_x(x(s), s)}{\frac{\partial H}{\partial x}} F(\hat{\theta}) + \frac{\Psi_x(x(s), s)f(s) + \Psi_{x\theta}(x(s), s)F(s)}{\frac{\partial H}{\partial x}} \right) ds \\
&= \int_{\hat{\theta}}^{\bar{\theta}} \Psi_{x\theta}(x(s), s) \left( \frac{\Psi_x(x(s), s)f(s) + \Psi_{x\theta}(x(s), s)(F(s) - F(\hat{\theta}))}{\frac{\partial H}{\partial x}} \right) ds
\end{aligned}$$

Since  $\partial H/\partial x$  is positive, and  $s \geq \hat{\theta}$  it follows that this quantity is positive; hence  $d\bar{x}/dk$  is positive. ■

**Proof of Proposition 12.** By Lemma 1, we can reduce (IC) and (IR) into a monotonicity constraint and a transfer schedule. Adding a Lagrange multiplier  $\lambda$  to the budget constraint, the control problem becomes

$$\begin{aligned}
\max_x \quad & \int_{\Theta} x(\theta)f(\theta) d\theta + \lambda \left( T - \int_{\underline{\theta}}^{\bar{\theta}} \left( \Psi(x(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds \right) f(\theta) d\theta \right) \\
& \dot{x} = u \leq 0
\end{aligned}$$

Integrating by parts, and writing the integral in one term, we get

$$\begin{aligned}
\max_x \quad & \int_{\Theta} \left( x(\theta) - \lambda \left( \Psi(x(\theta), \theta) + \Psi_{\theta}(x(\theta), \theta) \frac{F(\theta)}{f(\theta)} \right) \right) f(\theta) d\theta + \lambda T \\
& \dot{x} = u \leq 0
\end{aligned}$$

Applying Pontryagin and the Arrow sufficiency result, we get the conditions in the proposition. ■

**Proof of Proposition 11.** Applying Lemma 1, and noting that (IR) binds at the

highest type  $\bar{\theta}$ , the problem becomes

$$\begin{aligned} \max_x \quad & \int_{\Theta} (x(\theta) - c(t(\theta)))f(\theta) d\theta \\ \text{subject to} \quad & t(\theta) = \Psi(x(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_{\theta}(x(s), s) ds \\ & x \text{ nonincreasing} \end{aligned}$$

As before, we interpret this as a control problem; however, we now have two state variables  $x, t$ . Note that unlike the linear case, we cannot substitute  $t$  into the objective and integrate out by parts; rather, the control problem maintains  $t$  in the state. To formulate the law of motion for  $t$ , note that

$$\dot{t}(\theta) = \Psi_x(x(\theta), \theta)\dot{x}(\theta)$$

Hence, as before, let  $u = \dot{x}$  be the control variable. Then  $u$  simultaneously controls both  $x$  and  $t$  according to the following laws of motion:

$$\begin{bmatrix} \dot{x}(\theta) \\ \dot{t}(\theta) \end{bmatrix} = \begin{bmatrix} u(\theta) \\ \Psi_x(x(\theta), \theta)u(\theta) \end{bmatrix}$$

Then since IR binds at the top, we obtain the following boundary surface constraint:

$$t(\bar{\theta}) - \Psi(x(\bar{\theta}), \bar{\theta}) = 0$$

So all together, the control problem is

$$\begin{aligned} \max_x \quad & \int_{\Theta} (x(\theta) - c(t(\theta)))f(\theta) d\theta \\ & \begin{bmatrix} \dot{x}(\theta) \\ \dot{t}(\theta) \end{bmatrix} = \begin{bmatrix} u(\theta) \\ \Psi_x(x(\theta), \theta)u(\theta) \end{bmatrix} \\ & t(\bar{\theta}) - \Psi(x(\bar{\theta}), \bar{\theta}) = 0 \\ & t(\theta) \leq T \end{aligned}$$

We apply Pontryagin again, and note that because  $c$  is convex, the Hamiltonian is concave in both state variables, and invoke the Arrow sufficiency conditions. The Hamiltonian is

$$\mathcal{H} = (x - c(t))f(\theta) + \lambda t + \rho_x u + \rho_t \Psi_x(x, \theta)u$$



Since  $u$  maximizes the unconstrained Hamiltonian, we get the complementary slackness conditions (18) and (20). The costate equations (19) and (21) follow. The boundary condition lies on a surface, and so the orthogonality condition implies that the costate vector at the upper terminal must lie in the direction of the gradient vector:

$$\begin{bmatrix} \rho_x \\ \rho_t \end{bmatrix}(\bar{\theta}) \propto \begin{bmatrix} -\Psi_x(x(\bar{\theta}), \bar{\theta}) \\ 1 \end{bmatrix}$$

Using  $\rho_t$  to substitute out the constant of proportionality, we get the boundary condition (22). Since the lower endpoint is free, we get the other boundary condition (23). ■