Supermodularity and Affine Policies in Dynamic Robust Optimization

Dan A. Iancu*, Mayank Sharma† and Maxim Sviridenko‡

March 20, 2012

Abstract

This paper considers robust dynamic optimization problems, where the unknown parameters are modeled as uncertainty sets. We seek to bridge two classical paradigms for solving such problems, namely (1) Dynamic Programming (DP), and (2) policies parameterized in model uncertainties (also known as decision rules), obtained by solving tractable convex optimization problems.

We provide a set of unifying conditions (based on the interplay between the convexity and supermodularity of the DP value functions, and the lattice structure of the uncertainty sets) that, taken together, guarantee the optimality of the class of *affine* decision rules. We also derive conditions under which such affine rules can be obtained by optimizing simple (e.g., linear) objective functions over the uncertainty sets. Our results suggest new modeling paradigms for dynamic robust optimization, and our proofs, which bring together ideas from three areas of optimization typically studied separately (robust optimization, combinatorial optimization - the theory of lattice programming and supermodularity, and global optimization - the theory of concave envelopes), may be of independent interest.

We exemplify our findings in an application concerning the design of flexible contracts in a two-echelon supply chain, where the optimal contractual pre-commitments and the optimal ordering quantities can be found by solving a single linear program of small size.

1 Introduction

Dynamic optimization problems under uncertainty have been present in numerous fields of science and engineering, and have elicited interest from diverse research communities, on both a theoretical and a practical level. As a result, many solution approaches have been proposed, with various degrees of generality, tractability, and performance guarantees. One such methodology, which has received renewed interest in recent years due to its ability to provide workable solutions for many real-world problems, is robust optimization and robust control.

*Graduate School of Business, Stanford University, Stanford, CA 94305, daniancu@stanford.edu. Part of this research was conducted while the author was with the Department of Mathematical Sciences of the IBM T.J. Watson Research Center, whose financial support is gracefully acknowledged.

†IBM T.J.Watson Research Center, Yorktown Heights, NY 10598, mxsharma@us.ibm.com.

‡Computer Science Department, University of Warwick, Coventry, CV4 7AL, United Kingdom, M.I.Sviridenko@warwick.ac.uk.

The topics of robust optimization and robust control have been studied, under different names, by a variety of academic groups, in operations research (Ben-Tal and Nemirovski [1999, 2002], Ben-Tal et al. [2002], Bertsimas and Sim [2003, 2004]), engineering (Bertsekas and Rhodes [1971], Fan et al. [1991], El-Ghaoui et al. [1998], Zhou and Doyle [1998], Dullerud and Paganini [2005]), economics (Hansen and Sargent [2001], Hansen and Sargent [2008]), with considerable effort put into justifying the assumptions and general modeling philosophy. As such, the goal of the current paper is not to *motivate* the use of robust (and, more generally, distribution-free) techniques. Rather, we take the modeling approach as a given, and investigate questions of tractability and performance guarantees in the context of a specific class of dynamic optimization problems.

More precisely, we are interested in problems over a finite planning horizon, in which a system with state $\boldsymbol{x}_t$ is evolving according to some prescribed dynamics, influenced by the actions (or controls) $\boldsymbol{u}_t$ taken by the decision maker at every time t, and also subject to unknown disturbances $\boldsymbol{w}_t$. In a traditional (min-max or max-min) robust paradigm, the modeling assumption is that the uncertain quantities $\boldsymbol{w}_t$ are only known to lie in a specific uncertainty set $\mathcal{W}_t$, and the goal is to compute non-anticipative policies $\boldsymbol{u}_t$ so that the system obeys a set of pre-specified constraints robustly (i.e., for any possible realization of the uncertain parameters), while minimizing a worst-case performance measure (see, e.g., Löfberg [2003], Bemporad et al. [2003], Kerrigan and Maciejowski [2003], Ben-Tal et al. [2004, 2005a] and references therein).

Numerous problems in operations research result in models that exactly fit this description. For instance, multi-echelon inventory management, airline revenue management or dynamic pricing problems can all be formulated in such a framework, for suitable choices of the states, actions and disturbances (refer to Zipkin [2000], Porteus [2002] or Talluri and van Ryzin [2005] for many other instances). One such example, which we use throughout our paper to both motivate and exemplify our results, is the following supply chain contracting model, originally considered (in a slightly different form) in Ben-Tal et al. [2005b] and Ben-Tal et al. [2009].

Problem 1. *Consider a two-echelon supply chain, consisting of a retailer and a supplier. The retailer is selling a single product over a finite planning horizon, and facing unknown demands from customers. She is allowed to carry inventory and to backlog unsatisfied demand, and she can replenish her inventory in every period by placing orders with the supplier.*

The supplier is facing capacity investment decisions, which must be made in advance, before the selling season begins. In an effort to smoothen the production, the supplier enters a contract with the retailer, whereby the latter pre-commits (before the start of the selling season) to a particular sequence of orders. The contract also prescribes payments that the retailer must make to the supplier, whenever the realized orders deviate from the pre-commitments.

The goal is to determine, in a centralized fashion, the capacity investments by the supplier and the pre-commitments and ordering policies by the retailer that would minimize the overall, worst-case system cost (consisting of capacity investments, inventory, ordering, and contractual penalties).

The typical approach for solving such problems is via a Dynamic Programming (DP) formulation (Bertsekas [2001]), in which the optimal state-dependent policies $\boldsymbol{u}_t^\star(\boldsymbol{x}_t)$ and value functions $J_t^\star(\boldsymbol{x}_t)$ are characterized going backwards in time. DP is a powerful and flexible technique, enabling the modeling of complex problems, with nonlinear dynamics, incomplete information structures, etc. For certain "simpler" (low-dimensional) problems, the DP approach also allows an exact characterization of the optimal actions; this has lead to numerous celebrated results in operations research, a classic example being the optimality of basestock or (s, S) policies in inventory systems (Scarf [1960], Clark and Scarf [1960], Veinott [1966]). Furthermore, the DP approach often entails very useful comparative statics analyses, such as monotonicity results of the optimal policy with respect to particular problem parameters or state variables, monotonicity or convexity of the value functions, etc. (see, e.g., the classical texts Zipkin [2000], Topkis [1998], Heyman and Sobel [1984], Simchi-Levi et al. [2004], Talluri and van Ryzin [2005] for numerous such examples). We critically remark that such comparative statics results are often possible even for complex problems, where the optimal policies cannot be completely characterized (e.g., Zipkin [2008], Huh and Janakiraman [2010]).

The main downside of the DP approach is the well-known "curse of dimensionality", in that the complexity of the underlying Bellman recursions explodes with the number of state variables [Bertsekas, 2001], leading to a limited applicability of the methodology in practical settings. In fact, an example of this phenomenon already appears in the model for Problem 1: after the first period (in which the capacity investment and the pre-commitment decisions are made), the state of the problem consists of the vector of capacities, pre-commitments, and inventory available at the retailer. As Ben-Tal et al. [2005b, 2009] remark, even though the DP optimal ordering policy might have a simple form (e.g., if the ordering costs were linear, it would be a base-stock policy), the methodology would run into trouble, as (i) one may have to discretize the state variable and the actions, and hence produce only an approximate value function, (ii) the DP would have to be solved for any possible choice of capacities and pre-commitments, (iii) the value function depending on capacities and pre-commitments would, in general, be non-smooth, and (iv) the DP solution would provide no subdifferential information, leading to the use of zero-order (i.e., gradient-free) methods to solve the resulting first-stage problem, which exhibit notoriously slow convergence. This drawback of the DP methodology has motivated the development of numerous schemes for computing approximate solutions, such as Approximate Dynamic Programming, Model Predictive Control, and others (see, e.g., Bertsekas and Tsitsiklis [1996], Bertsekas [2001], de Farias and Van Roy [2003] or Powell [2007]).

An alternative approach is to give up solving the Bellman recursions altogether (even approximately), and instead focus on particular classes of policies that can be found by solving tractable optimization problems. One of the most popular such approaches is to consider *decision rules* directly parameterized in the observed disturbances, i.e.,

$$\boldsymbol{u}_t : \mathcal{W}_1 \times \mathcal{W}_2 \times \cdots \times \mathcal{W}_{t-1} \to \mathbb{R}^m, \tag{1}$$

where m is the number of control actions at time t. One such example, of particular interest due to its tractability and empirical success, has been the class of *affine* decision rules. Originally suggested in the stochastic programming literature [Charnes et al., 1958, Garstka and Wets, 1974], these rules have gained tremendous popularity in the robust optimization literature due to their tractability and empirical success (see, e.g., Löfberg [2003], Ben-Tal et al. [2004], Ben-Tal et al. [2005a, 2006], Bemporad et al. [2003], Kerrigan and Maciejowski [2003, 2004], Skaf and Boyd [2010], and the book Ben-Tal et al. [2009] and survey paper Bertsimas et al. [2011] for more references). Recently, they have been reexamined in stochastic settings, with several papers (Nemirovski and Shapiro [2005], Chen et al. [2008], Kuhn et al. [2009]) providing tractable methods for determining optimal policy parameters, in the context of both single-stage and multi-stage linear stochastic programming problems.

One central question when restricting attention to a particular subclass of policies (such as affine) is whether this induces large optimality gaps as compared to the DP solution. One such attempt was Bertsimas and Goyal [2010], which considers a two-stage linear optimization problem and shows that affine policies are optimal for a simplex uncertainty set, but can be within a factor of $\mathcal{O}\big(\sqrt{\dim(\mathcal{W})}\big)$ of the DP optimal objective in general, where $\dim(\mathcal{W})$ is the dimension of the (first-stage) uncertainty set. The work that is perhaps most related to ours is Bertsimas et al. [2010], where the authors show that affine policies are provably optimal for a considerably simpler setting than Problem 1 (without capacity or pre-commitment decisions, with linear ordering costs, and with the uncertainty set described by a hypercube). The proofs in the latter paper rely heavily on the problem structure, and are not easily extendable to any other settings (including that of Problem 1). Other research efforts have focused on providing tractable dual formulations, which allow a computation of lower or upper bounds, and hence an assessment of the sub-optimality level (see Kuhn et al. [2009] for details).

However, these (seemingly weak) theoretical results stand in contrast with the considerably stronger empirical observations. In a thorough simulation conducted for an application very similar to Problem 1, Ben-Tal et al. [2009] (Chapter 14, page 392) report that affine policies are *optimal* in all 768 instances tested, and Kuhn et al. [2009] find similar results for a related example.

In view of this observation, the goal of the present paper is to enhance the understanding of the modeling assumptions and problem structures that underlie the optimality of affine policies. We seek to do this, in fact, by bridging the strengths of the two approaches suggested above. Our contributions are as follows.

- We provide a set of unifying conditions on (a) the uncertainty sets, and (b) the optimal policies and value functions of a DP formulation, which, taken together, guarantee that affine decision rules are optimal in a dynamic optimization problem. The reason why such conditions are useful is that one can often conduct meaningful comparative statics analyses, even in situations when a DP formulation is computationally intractable. If the optimal value functions and policies happen to match our conditions, then one can forget about numerically solving the DP, and can instead simply focus attention on affine decision rules, which can often be computed by solving particular tractable (convex) mathematical programs.

  Our conditions critically rely on the convexity and supermodularity of the objective functions in question, as well as the lattice structure of the uncertainty set $\mathcal{W}$. To the best of our knowledge, these are the first results suggesting that lattice uncertainty sets might play a central role in constructing dynamic robust models, and that they bear a close connection with the optimality of affine forms in the resulting problems. Our proof techniques combine ideas from three areas of optimization typically studied separately - robust optimization, combinatorial optimization (the theory of lattice programming and supermodularity), and global optimization (the theory of concave envelopes) - and may be of independent interest.

- Using these conditions, we reexamine Problem 1, and show that - once the capacities and pre-commitment decisions are fixed - affine ordering policies are *provably* optimal, under *any convex ordering, inventory and contractual penalty costs*. Furthermore, the worst-case optimal ordering policy has a natural interpretation in terms of fractional satisfaction of backlogged demands. This is a considerable generalization and simplification of the results in Bertsimas et al. [2010], and it enforces the notion that optimal decision rules in robust models can retain a simple form, even as the cost structure of the problem becomes more complex: for instance, when ordering costs are convex, ordering policies that are affine in historical demands remain optimal, whereas policies parameterized in inventory become considerably more complex (see the discussion in Section 3.3.1).

- Recognizing that, even knowing that affine policies *are* optimal, one could still face the conundrum of solving complex mathematical programs, we provide a set of conditions under which the maximization of a sum of several convex and supermodular functions on a lattice can be replaced with the maximization of a single linear function. With these conditions, we show that, if all the costs in Problem 1 are jointly convex and piece-wise affine (with at most m pieces), then *all* the decisions in the problem (capacity investments, pre-commitments and optimal ordering policies) can be obtained by solving *a single* linear program, with $\mathcal{O}(mT^2)$ variables. This explains the empirical results in Ben-Tal et al. [2005b, 2009], and identifies the sole modeling component that renders affine decision rules suboptimal in the latter models.

The rest of the paper is organized as follows. Section 2 contains a precise mathematical description of the two main problems we seek to address. Section 3 and Section 4 contain our main results, with the answers to each separate question, and a detailed discussion of their application to Problem 1 stated in the introduction. Section 5 concludes the paper. The Appendix contains relevant background material on lattice programming, supermodularity and concave envelopes (Section 7.1 and Section 7.2), as well as some of the technical proofs (Section 7.3).

1.1 Notation

Throughout the text, vector quantities are denoted in bold font. To avoid extra symbols, we use concatenation of vectors in a liberal fashion, i.e., for $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^k$, we use $(\boldsymbol{a}, \boldsymbol{b})$ to denote either the row vector $(a_1, \ldots, a_n, b_1, \ldots, b_k)$ or the column vector $(a_1, \ldots, a_n, b_1, \ldots, b_k)^T$. The meaning should be clear from context. The operators min, max, $\geq$ and $\leq$ applied to vectors should be interpreted in component-wise fashion.

For a vector $\boldsymbol{x} \in \mathbb{R}^n$ and a set $S \subseteq \{1, \ldots, n\}$, we use $\boldsymbol{x}(S) \stackrel{\text{def}}{=} \sum_{j \in S} x_j$, and denote by $\boldsymbol{x}_S \in \mathbb{R}^n$ the vector with components x_i for $i \in S$ and 0 otherwise. In particular, $\mathbf{1}_S$ is the characteristic vector of the set S, $\mathbf{1}_i$ is the i-th unit vector of $\mathbb{R}^n$, and $\mathbf{1} \in \mathbb{R}^n$ is the vector with all components equal to 1. We use $\Pi(S)$ to denote the set of all permutations on the elements of S, and $\pi(S)$ or $\sigma(S)$ denote particular such permutations. We let $S^C = \{1, \ldots, n\} \setminus S$ denote the complement of S, and, for any permutation $\pi \in \Pi(S)$, we write $\pi(i)$ for the element of S appearing in the i-th position under permutation π.

For a set $P \subseteq \mathbb{R}^n$, we use $\text{ext}(P)$ to denote the set of its extreme points, and $\text{conv}(P)$ to denote its convex hull.

2 Problem Statement

As discussed in the introduction, both the DP formulation and the decision rule approach have well-documented merits. The former is general-purpose, and allows very insightful comparative statics analyses, even when the DP approach itself is computationally intractable. For instance, one can check the monotonicity of the optimal policy or value function with respect to particular problem parameters or state variables, or prove the convexity or sub/supermodularity of the value function. A recent such example in the inventory literature are the monotonicity results concerning the optimal ordering policies in single or multi-echelon supply chains with positive lead-time and lost sales (Zipkin [2008] and Huh and Janakiraman [2010]). For more examples, we refer the interested reader to several classical texts on inventory and revenue management: Zipkin [2000], Topkis [1998], Heyman and Sobel [1984], Simchi-Levi et al. [2004], Talluri and van Ryzin [2005].

In contrast, the decision-rule approach does not typically allow such structural results, but instead takes the pragmatic view of focusing on practical decisions, which can be efficiently computed by convex optimization techniques (see, e.g., Chapter 14 of Ben-Tal et al. [2009]).

The goal of the present paper is to provide a link between the two analyses, and to enhance the understanding of the modeling assumptions and problem structures that underlie the optimality of affine decision rules. More precisely, we pose and address two main problems, the first of which is the following.

Problem 2. *Consider a one-period game between a decision maker and nature*

$$\max_{\boldsymbol{w}\in\mathcal{W}} \min_{\boldsymbol{u}(\boldsymbol{w})} f(\boldsymbol{w},\boldsymbol{u}), \tag{2}$$

where $\boldsymbol{w}$ denotes an action chosen by nature from an uncertainty set $\mathcal{W}\subseteq\mathbb{R}^n$, $\boldsymbol{u}$ is a response by the decision maker, allowed to depend on nature's action $\boldsymbol{w}$, and f is a total cost function. With $\boldsymbol{u}^\star(\boldsymbol{w})$ denoting the **Bellman-optimal** *policy, we seek conditions on the set $\mathcal{W}$, the policy $\boldsymbol{u}^\star(\boldsymbol{w})$, and the function $f\big(\boldsymbol{w},\boldsymbol{u}^\star(\boldsymbol{w})\big)$ such that there exists an* **affine** *policy that is* **worst-case optimal** *for the decision maker, i.e.,*

$$\exists\, Q\in\mathbb{R}^{m\times n},\ \boldsymbol{q}\in\mathbb{R}^m \text{ such that } \max_{\boldsymbol{w}\in\mathcal{W}} \min_{\boldsymbol{u}(\boldsymbol{w})} f(\boldsymbol{w},\boldsymbol{u}) = \max_{\boldsymbol{w}\in\mathcal{W}} f(\boldsymbol{w}, Q\boldsymbol{w}+\boldsymbol{q}).$$

To understand the question, imagine separating the objective into two components, $f(\boldsymbol{w},\boldsymbol{u}) = h(\boldsymbol{w}) + J(\boldsymbol{w},\boldsymbol{u})$. Here, h summarizes a sequence of historical costs (all depending on the unknowns $\boldsymbol{w}$), while J denotes a cost-to-go (or value function). As such, the outer maximization in (2) can be interpreted as the problem solved by nature at a particular stage in the decision process, whereby the total costs (historical + cost-to-go) are being maximized. The inner minimization exactly captures the decision maker's problem, of minimizing the cost-to-go.

As remarked earlier, an answer to this question would be most useful in conjunction with comparative statics results obtained from a DP formulation: if the optimal value (and policies) matched the conditions in the answer to Problem 2, then one could forget about numerically solving the DP, and could instead simply focus attention on disturbance-affine policies, which could be computable by efficient convex optimization techniques (see, e.g., Löfberg [2003], Ben-Tal et al. [2004], Ben-Tal et al. [2005a], Ben-Tal et al. [2009] or Skaf and Boyd [2010]).

We remark that the notion of *worst-case optimal* policies in the previous question is different than that of *Bellman-optimal* policies [Bertsekas, 2001]. In the spirit of DP, the latter requirement would translate in the policy $\boldsymbol{u}^\star(\boldsymbol{w})$ being the optimal response by the decision maker for *any* revealed $\boldsymbol{w}\in\mathcal{W}$, while the former notion only requires that $\boldsymbol{u}(\boldsymbol{w}) = Q\boldsymbol{w}+\boldsymbol{q}$ is an optimal response at points $\boldsymbol{w}$ that result in the overall worst-case cost (while keeping the cost for all other $\boldsymbol{w}$ below the worst-case cost). This distinction has been drawn before [Bertsimas et al., 2010], and is one of the key features distinguishing robust (mini-max) models from their stochastic counterparts, and allowing the former models to potentially admit optimal policies with simpler structure than those for the latter class. While one could build a case against worst-case optimal policies by arguing that a rational decision maker should never accept policies that are not Bellman optimal (see, e.g., Epstein and Schneider [2003] and Cheridito et al. [2006] for pointers to the literature in economics and risk theory on this topic), we adopt the pragmatic view here that, provided there is no degeneracy in the optimal policies (i.e., there is a unique set of optimal policies in the problem), one can always replicate the true Bellman-optimal policies through a receding horizon approach [Bertsekas, 2001], by applying the first-stage decisions and resolving the subproblems in the decision process.

While answering the above question is certainly very relevant, the results might still remain existential in nature. In other words, even armed with the knowledge that affine policies *are* optimal, one could be faced with the conundrum of solving complex mathematical programs to find such policies. To partially alleviate this issue, we raise the following related problem.

Problem 3. *Consider a maximization problem of the form*

$$\max_{\boldsymbol{w}\in\mathcal{W}} \sum_{t\in\mathcal{T}} h_t(\boldsymbol{w}),$$

where $\mathcal{W}\subseteq\mathbb{R}^n$ denotes an uncertainty set, and $\mathcal{T}$ is a finite index set. Let $J^\star$ denote the maximum value in the problem above (assumed finite). We seek conditions on $\mathcal{W}$ and/or h_t such that there exist **affine** *functions $\boldsymbol{z}_t(\boldsymbol{w}), \forall\, t\in\mathcal{T}$, such that*

$$z_t(\boldsymbol{w}) \geq h_t(\boldsymbol{w}),\ \forall\, \boldsymbol{w}\in\mathcal{W},$$
$$J^\star = \max_{\boldsymbol{w}\in\mathcal{W}} \sum_{t\in\mathcal{T}} z_t(\boldsymbol{w}).$$

In words, the latter problem requires that substituting a set of true historical costs h_t with potentially larger (but *affine*) costs z_t results in no change of the worst-case cost. Since one can typically optimize linear functionals efficiently over most uncertainty sets of practical interest (see, e.g., Ben-Tal et al. [2009]), an answer to this problem, combined with an answer to Problem 2, might enable very efficient mathematical programs for computing worst-case optimal affine policies that depend on disturbances.

3 Discussion of Problem 2

We begin by considering Problem 2 in the introduction. With $\boldsymbol{u}^\star(\boldsymbol{w}) \in \arg\min_{\boldsymbol{u}} f(\boldsymbol{w},\boldsymbol{u})$ denoting a Bellman-optimal response by the decision maker, the latter problem can be summarized compactly as finding conditions on $\mathcal{W}$, $\boldsymbol{u}^\star$ and f such that

$$\exists\, Q\in\mathbb{R}^{m\times n},\ \boldsymbol{q}\in\mathbb{R}^m \,:\, \max_{\boldsymbol{w}\in\mathcal{W}} f\big(\boldsymbol{w}, u^\star(\boldsymbol{w})\big) = \max_{\boldsymbol{w}\in\mathcal{W}} f(\boldsymbol{w}, Q\boldsymbol{w}+\boldsymbol{q}).$$

To the best of our knowledge, two partial answers to this question are known in the literature. If $\mathcal{W}$ is a simplex or a direct product of simplices, and $f(\boldsymbol{w}, Q\boldsymbol{w}+\boldsymbol{q})$ is convex in $\boldsymbol{w}$ for any $Q,\boldsymbol{q}$, then a worst-case optimal policy can be readily obtained by computing $Q,\boldsymbol{q}$ so as to match the value of $u^\star(\boldsymbol{w})$ on all the points in $\operatorname{ext}(\mathcal{W})$ (see Bertsimas and Goyal [2010] and Lemma 14.3.6 in Ben-Tal et al. [2009] for the more general case). This is not a surprising result, since the number of extreme points of the uncertainty set exactly matches the number of policy parameters (i.e., the degrees of freedom in the problem). A separate instance where the construction is possible is provided in Bertsimas et al. [2010], where $\mathcal{W} = \mathcal{H}_n \overset{\text{def}}{=} [0,1]^n$ is the unit hypercube in $\mathbb{R}^n$, $u:\mathcal{W}\to\mathbb{R}$, and f has the specific form

$$f(\boldsymbol{w},u) = a_0 + \boldsymbol{a}^T\boldsymbol{w} + c\cdot u + g(b_0 + \boldsymbol{b}^T\boldsymbol{w} + u),$$

where $a_0, b_0, c\in\mathbb{R}$, $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^n$ are arbitrary, and $g:\mathbb{R}\to\mathbb{R}$ is any convex function. The proof for the latter result heavily exploits the particular functional form above, and does not lend itself to any extensions or interpretations. In particular, it fails even if one replaces $c\cdot u$ with $c(u)$, for some convex function $c:\mathbb{R}\to\mathbb{R}$.

In the current paper, we also focus our attention on uncertainty sets $\mathcal{W}$ that are polytopes in $\mathbb{R}^n$. More precisely, with $V = \{1, \ldots, n\}$, we consider any directed graph $G = (V, E)$, where $E \subseteq V^2$ is any set of directed edges, and are interested in uncertainty sets of the form:

$$\mathcal{W} = \{\boldsymbol{w} \in \mathcal{H}_n \,:\, w_i \geq w_j, \,\forall\, (i, j) \in E\} \tag{3}$$

It can be shown (see Tawarmalani et al. [2010] and references therein for details) that the polytope $\mathcal{W}$ in (3) has boolean vertices, since the matrix of constraints is totally unimodular. As such, any $\boldsymbol{x} \in \operatorname{ext}(\mathcal{W})$ can be represented as $\boldsymbol{x} = \mathbf{1}_{S_x}$, for some set $S_x \subseteq V$. Furthermore, it can also be checked that the set $\operatorname{ext}(\mathcal{W})$ is a sublattice of[1] $\{0,\, 1\}^n$ [Topkis, 1998],

$$\forall\, \boldsymbol{x}, \boldsymbol{y} \in \operatorname{ext}(\mathcal{W}) \,:\, \min(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{1}_{S_x \cap S_y} \in \operatorname{ext}(\mathcal{W}), \;\; \max(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{1}_{S_x \cup S_y} \in \operatorname{ext}(\mathcal{W}).$$

Among the uncertainty sets typically considered in the modeling literature, the hypercube is one example that fits the description above. Certain hyper-rectangles, as well as any simplices or cross-products of simplices could also be reduced to this form via a suitable change of variables[2] (see, e.g., Tawarmalani et al. [2010]). For an example of such an uncertainty set and its corresponding graph G, we direct the reader to Figure 1.

For any polytope of the form (3), we define the corresponding set $\Pi^{\mathcal{W}}$ of permutations of V that are consistent with the pre-order induced by the lattice $\operatorname{ext}(\mathcal{W})$, i.e.,

$$\Pi^{\mathcal{W}} = \big\{\pi \in \Pi(V) \,:\, \pi^{-1}(i) \leq \pi^{-1}(j), \,\forall\, (i, j) \in E\big\}. \tag{4}$$

In other words, if $(i, j) \in E$, then i must appear before j in any permutation $\pi \in \Pi^{\mathcal{W}}$. Furthermore, with any permutation $\pi \in \Pi(V)$, we also define the simplex Δ_π obtained from vertices of $\mathcal{H}$ in the order determined by the permutation π, i.e.,

$$\Delta_\pi \stackrel{\text{def}}{=} \operatorname{conv}\Big(\big\{\mathbf{0} + \sum_{j=1}^{k} \mathbf{1}_{\pi(j)} \,:\, k = 0, \ldots, n\big\}\Big). \tag{5}$$

It can then be checked (see, e.g., Tawarmalani et al. [2010]) that any vertex $\boldsymbol{w} \in \operatorname{ext}(\mathcal{W})$ belongs to several such simplices. More precisely, with $\boldsymbol{w} = \mathbf{1}_{S_w}$ for a particular $S_w \subseteq V$, we have

$$\boldsymbol{w} \in \Delta_\pi, \,\forall\, \pi \in \mathscr{S}_w \stackrel{\text{def}}{=} \big\{\, \pi \in \Pi^{\mathcal{W}} \,:\, \{\pi(1), \ldots, \pi(|S_w|)\} = S_w \,\big\}. \tag{6}$$

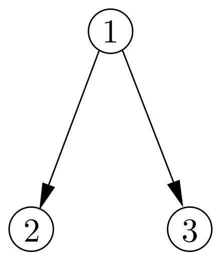
In other words, $\boldsymbol{w}$ is contained in any simplex corresponding to a permutation π that (a) is consistent with the pre-order on $\mathcal{W}$, and (b) has the indices in S_w in the first $|S_w|$ positions. An example in included in Figure 1.

Since $\operatorname{ext}(\mathcal{W})$ is a lattice, we can also consider functions $f : \mathcal{W} \to \mathbb{R}$ that are *supermodular on* $\operatorname{ext}(\mathcal{W})$, i.e.,

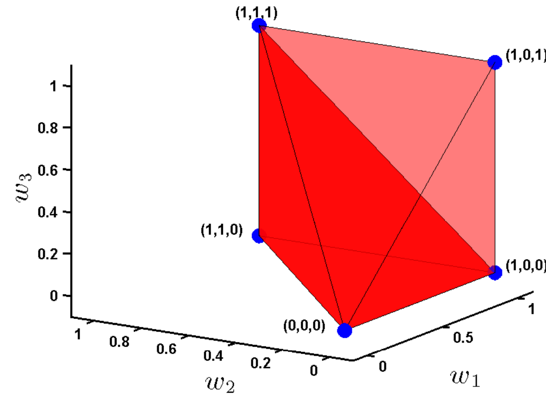
$$f\big(\min(\boldsymbol{x}, \boldsymbol{y})\big) + f\big(\max(\boldsymbol{x}, \boldsymbol{y})\big) \geq f(\boldsymbol{x}) + f(\boldsymbol{y}), \,\forall\, \boldsymbol{x}, \boldsymbol{y} \in \operatorname{ext}(\mathcal{W}).$$

[1] We could also state these results in terms of $\mathcal{W}$ itself being a sublattice of $\mathcal{H}_n$. However, the distinction will turn out to be somewhat irrelevant, since the convexity of all the objectives will dictate that only the structure of the extreme points of $\mathcal{W}$ matters.

[2] For a simplex, if $\mathcal{W}^\Gamma = \big\{\boldsymbol{w} \geq \mathbf{0} \,:\, \sum_{i=1}^n w_i \leq \Gamma\big\}$, then, with the change of variables $y_k \stackrel{\text{def}}{=} (\sum_{i=1}^k w_i)/\Gamma$, $\forall\, k \in \{1, \ldots, n\}$, the corresponding uncertainty set in the $\boldsymbol{y}$ variables is $\mathcal{W}_y = \big\{\boldsymbol{y} \in [0, 1]^n \,:\, 0 \leq y_1 \leq y_2 \leq \cdots \leq y_n \leq 1\big\}$.



(a) $G = (V, E)$, where $V = \{1, 2, 3\}$ and $E = \{(1,2), (1,3)\}$.



(b) $\mathcal{W} = \{\boldsymbol{w} \in \mathcal{H}_3 \,:\, w_1 \geq w_2,\, w_1 \geq w_3\}$.

Figure 1: Example of a sublattice uncertainty set. (a) displays the graph of precedence relations, and (b) plots the corresponding uncertainty set. Here, $\Pi^{\mathcal{W}} = \{(1,2,3), (1,3,2)\}$, and the two corresponding simplicies are $\Delta_{(1,2,3)} = \operatorname{conv}\big(\{(0,0,0), (1,0,0), (1,1,0), (1,1,1)\}\big)$ and $\Delta_{(1,3,2)} = \operatorname{conv}\big(\{(0,0,0), (1,0,0), (1,0,1), (1,1,1)\}\big)$, shown in different shades in (b). Also, $\mathscr{S}_{(0,0,0)} = \mathscr{S}_{(1,0,0)} = \mathscr{S}_{(1,1,1)} = \Pi^{\mathcal{W}}$, while $\mathscr{S}_{(1,0,1)} = \{(1,3,2)\}$, and $\mathscr{S}_{(1,1,0)} = \{(1,2,3)\}$.

The properties of such functions have been studied extensively in combinatorial optimization and economics (see, e.g., Fujishige [2005], Schrijver [2003] and Topkis [1998] for detailed treatments and references). The main results that are relevant for our purposes are summarized in Section 7.1 and Section 7.2 of the Appendix.

With these definitions, we can now state our first main result, providing a set of sufficient conditions guaranteeing the desired outcome in Problem 2.

Theorem 1. *Consider any optimization problem of the form*

$$\max_{\boldsymbol{w} \in \mathcal{W}} \min_{\boldsymbol{u}(\boldsymbol{w})} f(\boldsymbol{w}, \boldsymbol{u}),$$

where $\mathcal{W}$ is of the form (3). Let $\boldsymbol{u}^\star : \mathcal{W} \to \mathbb{R}^m$ denote a Bellman-optimal response of the decision maker, and $f^\star(\boldsymbol{w}) \stackrel{\text{def}}{=} f\big(\boldsymbol{w}, \boldsymbol{u}^\star(\boldsymbol{w})\big)$ be the corresponding optimal cost function. Assume the following conditions are met:

[A1] *$f^\star(\boldsymbol{w})$ is convex on $\mathcal{W}$ and supermodular in $\boldsymbol{w}$ on $\operatorname{ext}(\mathcal{W})$.*

[A2] *For $Q \in \mathbb{R}^{m \times n}$ and $\boldsymbol{q} \in \mathbb{R}^n$, the function $f\big(\boldsymbol{w},\, Q\boldsymbol{w} + \boldsymbol{q}\big)$ is convex in $(Q, \boldsymbol{q})$ for any fixed $\boldsymbol{w}$.*

[A3] *There exists $\hat{\boldsymbol{w}} \in \arg\max_{\boldsymbol{w} \in \mathcal{W}} f^\star(\boldsymbol{w})$ such that, with $\mathscr{S}_{\hat{\boldsymbol{w}}}$ given by (6), the matrices $\{Q^\pi\}_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}}$ and vectors $\{\boldsymbol{q}^\pi\}_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}}$ obtained as the solutions to the systems of linear equations*

$$\forall\, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}} \;:\; Q^\pi \boldsymbol{w} + \boldsymbol{q}^\pi = \boldsymbol{u}^\star(\boldsymbol{w}),\; \forall\, \boldsymbol{w} \in \operatorname{ext}(\Delta_\pi), \tag{7}$$

are such that the function $f\big(\boldsymbol{w},\, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}}\big)$ is convex in $\boldsymbol{w}$ and supermodular on $\operatorname{ext}(\mathcal{W})$, for any *$\bar{Q}$ and $\bar{\boldsymbol{q}}$ obtained as*

$$\bar{Q} = \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi Q^\pi, \quad \bar{\boldsymbol{q}} = \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \boldsymbol{q}^\pi, \qquad \text{where } \lambda_\pi \geq 0,\; \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi = 1. \tag{8}$$

Then, there exist $\{\lambda_\pi\}_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}}$ *such that, with* $\bar{Q}$ *and* $\bar{\boldsymbol{q}}$ *given by* (8),

$$\max_{\boldsymbol{w}\in\mathcal{W}} f^\star(\boldsymbol{w}) = \max_{\boldsymbol{w}\in\mathcal{W}} f(\boldsymbol{w}, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}}).$$

Before presenting the proof of the theorem, we provide a brief explanation and intuition for the conditions above. A more detailed discussion, together with relevant examples, is included immediately after the proof.

The interpretation and the test for conditions **[A1]** and **[A2]** are fairly straightforward. The idea behind **[A3]** is to consider every simplex Δ_π that contains the maximizer $\hat{\boldsymbol{w}}$; there are exactly $|\mathscr{S}_{\hat{\boldsymbol{w}}}|$ such simplices, characterized by (6). For every such simplex, one can compute a corresponding affine decision rule $Q^\pi \boldsymbol{w} + \boldsymbol{q}^\pi$ by linearly interpolating the values of the Bellman-optimal response $u^\star(\boldsymbol{w})$ at the extreme points of Δ_π. This is exactly what is expressed in condition (7), and the resulting system is always compatible, since every such matrix-vector pair has exactly m rows, and the $n+1$ variables on each row participate in exactly $n+1$ linearly independent constraints (one for each point in the simplex). Now, the key condition in **[A3]** considers affine decisions rules obtained as *arbitrary* convex combinations of the rules $Q^\pi\boldsymbol{w}+\boldsymbol{q}^\pi$, and requires that the resulting cost function, obtained by using such rules, remains convex and supermodular in $\boldsymbol{w}$.

3.1 Proof of Theorem 1

In view of these remarks, the strategy behind the proof of Theorem 1 is quite straightforward: we seek to show that, if conditions **[A1-3]** are obeyed, then one can find suitable convex coefficients $\{\lambda_\pi\}_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}}$ so that the resulting affine decision rule $\bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}}$ is worst-case optimal. To ensure the latter fact, it suffices to check that the global maximum of the function $f(\boldsymbol{w}, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}})$ is still reached at the point $\hat{\boldsymbol{w}}$, which is one of the maximizers of $f^\star(\boldsymbol{w})$. Unfortunately, this is not trivial to do, since both functions $f(\boldsymbol{w}, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}})$ and $f^\star(\boldsymbol{w})$ are convex in $\boldsymbol{w}$ (by **[A1,3]**), and it is therefore hard to characterize their global maximizers, apart from stating that they occur at extreme points of the feasible set [Rockafellar, 1970].

The first key idea in the proof is to examine the *concave envelopes* of $f(\boldsymbol{w}, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}})$ and $f^\star(\boldsymbol{w})$, instead of the functions themselves. Recall that the concave envelope of a function $f : P \to \mathbb{R}$ on the domain P, which we denote by $\mathrm{conc}_P(f) : P \to \mathbb{R}$, is the point-wise smallest concave function that over-estimates f on P [Rockafellar, 1970], and always satisfies $\arg\max_{\boldsymbol{x}\in P} f \subseteq \arg\max_{\boldsymbol{x}\in P} \mathrm{conc}_P(f)$ (the interested reader is referred to Section 7.2 of the Appendix for a short overview of background material on concave envelopes, and to the papers Tardella [2008] or Tawarmalani et al. [2010] for other useful references).

In this context, a central result used repetitively throughout our analysis is the following characterization for the concave envelope of a function that is convex *and supermodular* on a polytope of the form (3).

Lemma 1. *If* $f^\star : \mathcal{W} \to \mathbb{R}$ *is convex on* $\mathcal{W}$ *and supermodular on* $\mathrm{ext}(\mathcal{W})$*, then*

1. *The concave envelope of* $f^\star$ *on* $\mathcal{W}$ *is given by the Lovász extension of* $f^\star$ *restricted to* $\mathrm{ext}(\mathcal{W})$*:*

$$\mathrm{conc}_{\mathcal{W}}(f^\star)(\boldsymbol{w}) = f^\star(\boldsymbol{0}) + \min_{\pi\in\Pi^{\mathcal{W}}} \sum_{i=1}^{n} \left[f^\star\Big(\sum_{j=1}^{i} \boldsymbol{1}_{\pi(j)}\Big) - f^\star\Big(\sum_{j=1}^{i-1} \boldsymbol{1}_{\pi(j)}\Big) \right] w_{\pi(i)}. \tag{9}$$

2. *The inequalities* $(\boldsymbol{g}^{\pi})^T \boldsymbol{w} + g_0 \geq f^{\star}(\boldsymbol{w})$ *defining non-vertical facets of* $\operatorname{conc}_{\mathcal{W}}(f^{\star})$ *are given by the set* $\operatorname{ext}(\mathcal{D}_{f^{\star},\mathcal{W}}) = \{(\boldsymbol{g}^{\pi}, g_0) \in \mathbb{R}^{n+1} : \pi \in \Pi^{\mathcal{W}}\}$*, where*

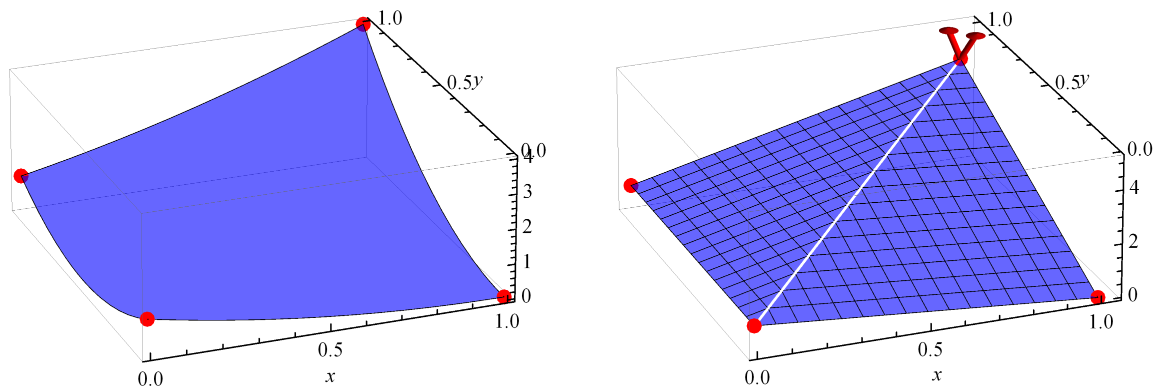
$$g_0 \stackrel{\text{def}}{=} f^{\star}(\mathbf{0}), \quad \boldsymbol{g}^{\pi} \stackrel{\text{def}}{=} \sum_{i=1}^{n} \left[f^{\star}\Big(\sum_{j=1}^{i} \mathbf{1}_{\pi(j)}\Big) - f^{\star}\Big(\sum_{j=1}^{i-1} \mathbf{1}_{\pi(j)}\Big)\right] \mathbf{1}_{\pi(i)}, \ \forall \pi \in \Pi^{\mathcal{W}}. \tag{10}$$

3. *The polyhedral subdivision of* $\mathcal{W}$ *yielding the concave envelope is given by the restricted Kuhn triangulation,*

$$\mathcal{K}^{\mathcal{W}} \stackrel{\text{def}}{=} \{ \Delta_{\pi} : \pi \in \Pi^{\mathcal{W}} \}.$$

Proof. This result is a particular instance of Corollary 3 in the Appendix, which in itself is a restatement of Theorem 3.3 and Corollary 3.4 in Tawarmalani et al. [2010], to which we direct the interested reader for more details. □

The previous lemma essentially establishes that the concave envelope of a function $f^{\star}$ that is convex and supermodular on an integer sublattice of $\{0,1\}^n$ is determined by the Lovász extension [Lovász, 1982]. The latter function is *polyhedral* (i.e., piece-wise affine), and is obtained by affinely interpolating the function $f^{\star}$ on all the simplicies in the Kuhn triangulation $\mathcal{K}^{\mathcal{W}}$ of the hypercube (see Section 7.2.1 of the Appendix). A plot of such a function f and its concave envelope is included in Figure 2.



(a) $f : \mathcal{H}_2 \to \mathbb{R}$, $f(x,y) = (x + 2y - 1)^2$ (b) $\operatorname{conc}_{\mathcal{H}_2}(f)$

Figure 2: A convex and supermodular function (a) and its concave envelope (b). Here, $\mathcal{W} = \mathcal{H}_2$, $\Pi^{\mathcal{W}} = \{(1,2),(2,1)\}$, and $\mathcal{K}^{\mathcal{W}} = \{\Delta_{(1,2)}, \Delta_{(2,1)}\}$, where $\Delta_{(1,2)} = \operatorname{conv}(\{(0,0),(1,0),(1,1)\}$ and $\Delta_{(2,1)} = \operatorname{conv}(\{(0,0),(0,1),(1,1)\})\}$. The plot in (b) also shows the two normals of non-vertical facets of $\operatorname{conc}_{\mathcal{W}}(f)$, corresponding to $\boldsymbol{g}^{(1,2)}$ and $\boldsymbol{g}^{(2,1)}$.

With this powerful lemma, we can now provide a result that brings us very close to a complete proof of Theorem 1.

Lemma 2. *Suppose* $f^{\star} : \mathcal{W} \to \mathbb{R}$ *is convex on* $\mathcal{W}$ *and supermodular on* $\operatorname{ext}(\mathcal{W})$*. Consider an arbitrary* $\hat{\boldsymbol{w}} \in \operatorname{ext}(\mathcal{W}) \cap \arg\max_{\boldsymbol{w}\in\mathcal{W}} f^{\star}(\boldsymbol{w})$*, and let* $\boldsymbol{g}^{\pi}$ *be given by* (10)*. Then,*

1. *For any* $\boldsymbol{w} \in \mathcal{W}$*, we have*

$$f^{\star}(\boldsymbol{w}) \leq f^{\star}(\hat{\boldsymbol{w}}) + (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^{\pi}, \ \forall \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}. \tag{11}$$

2. There exists a set of convex weights $\{\lambda_\pi\}_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}}$ *such that* $\boldsymbol{g} = \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \, \boldsymbol{g}^\pi$ *satisfies*

$$(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} \leq 0, \, \forall \, \boldsymbol{w} \in \mathcal{W}. \tag{12}$$

Proof. The proof is rather technical, and we defer it to Section 7.3 of the Appendix. □

For a geometric intuition of these results, we refer to Figure 2. In particular, the first claim simply states that the vectors $\boldsymbol{g}^\pi$ corresponding to simplicies that contain $\hat{\boldsymbol{w}}$ are valid supergradients of the function $f^\star$ at $\hat{\boldsymbol{w}}$; this is a direct consequence of Lemma 1, since any such vectors $\boldsymbol{g}^\pi$ are also supergradients for the concave envelope $\text{conc}_{\mathcal{W}}(f^\star)$ at $\hat{\boldsymbol{w}}$. The second claim states that one can always find a convex combination of the supergradients $\boldsymbol{g}^\pi$ that yields a supergradient $\boldsymbol{g}$ that is not a direction of increase for the function $f^\star$ when moving in *any* feasible direction away from $\hat{\boldsymbol{w}}$ (i.e., while remaining in $\mathcal{W}$).

With this lemma, we can now complete the proof of our main result.

Proof of Theorem 1. Consider any $\hat{\boldsymbol{w}}$ satisfying the requirement **[A3]**. Note that the system of equations in (7) is uniquely defined, since each row of the matrix Q^π and the vector $\boldsymbol{q}^\pi$ participate in exactly $n+1$ constraints, and the corresponding constraint matrix is non-singular. Furthermore, from the definition of Δ_π in (5), we have that $\boldsymbol{0} \in \text{ext}(\Delta_\pi)$, $\forall \, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}$, so that the system in (7) yields $\boldsymbol{q}^\pi = \boldsymbol{u}^\star(\boldsymbol{0})$, $\forall \, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}$.

By Lemma 2, consider the set of weights $\{\lambda_\pi\}_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}}$, such that $\boldsymbol{g} = \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \, \boldsymbol{g}^\pi$ satisfies $(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} \leq 0, \, \forall \, \boldsymbol{w} \in \mathcal{W}$. We claim that the corresponding $\bar{Q} = \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \, Q^\pi$, and $\bar{\boldsymbol{q}} = \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \, \boldsymbol{q}^\pi$ provide the desired affine policy $\bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}}$ such that

$$\max_{\boldsymbol{w} \in P} f^\star(\boldsymbol{w}) = \max_{\boldsymbol{w} \in P} f(\boldsymbol{w}, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}}).$$

To this end, note that, by **[A3]**, the functions $f(\boldsymbol{w}, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}})$ and $f^\pi(\boldsymbol{w}) \stackrel{\text{def}}{=} f(\boldsymbol{w}, Q^\pi \boldsymbol{w} + \boldsymbol{q}^\pi)$, $\forall \, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}$ are convex in $\boldsymbol{w}$ and supermodular on $\text{ext}(\mathcal{W})$. Also, by construction,

$$\begin{aligned} &\forall \, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}, \;\; f^\pi(\boldsymbol{w}) = f^\star(\boldsymbol{w}), \, \forall \, \boldsymbol{w} \in \text{ext}(\Delta_\pi), \\ &\quad\quad f(\hat{\boldsymbol{w}}, \bar{Q}\hat{\boldsymbol{w}} + \bar{\boldsymbol{q}}) = f^\star(\hat{\boldsymbol{w}}). \end{aligned} \tag{13}$$

Thus, for any $\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}$, the supergradient $\boldsymbol{g}^\pi$ defined for the function $f^\star$ in (10) remains a valid supergradient for f^π at $\boldsymbol{w} = \hat{\boldsymbol{w}}$. As such, relation (11) also holds for each function f^π, i.e.,

$$f^\pi(\boldsymbol{w}) \leq f^\pi(\hat{\boldsymbol{w}}) + (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^\pi, \;\; \forall \, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}. \tag{14}$$

The following reasoning then concludes our proof

$$\begin{aligned} \forall \, \boldsymbol{w} \in P, \;\; f(\boldsymbol{w}, \bar{Q}\boldsymbol{w} + \bar{\boldsymbol{q}}) &\stackrel{\textbf{[A2]}}{\leq} \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi f(\boldsymbol{w}, Q^\pi \boldsymbol{w} + \boldsymbol{q}^\pi) \\ &\stackrel{(14)}{\leq} \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \Big[f^\pi(\hat{\boldsymbol{w}}) + (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^\pi \Big] \\ &\stackrel{(13)}{=} f(\hat{\boldsymbol{w}}, \bar{Q}\hat{\boldsymbol{w}} + \bar{\boldsymbol{q}}) + \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \, (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^\pi \\ &\stackrel{(12)}{\leq} f(\hat{\boldsymbol{w}}, \bar{Q}\hat{\boldsymbol{w}} + \bar{\boldsymbol{q}}). \end{aligned}$$

□

3.2 Examples and Discussion of Existential Conditions

We now proceed to discuss the conditions in Theorem 1, and relevant examples of functions satisfying them. Condition [**A1**] can be generally checked by performing suitable comparative statics analyses. For instance, $f^\star(\boldsymbol{w})$ will be convex in $\boldsymbol{w}$ if $f(\boldsymbol{w}, \boldsymbol{u})$ is jointly convex in $(\boldsymbol{w}, \boldsymbol{u})$, since partial minimization preserves convexity [Rockafellar, 1970]. For supermodularity of $f^\star$, more structure is typically needed on $f(\boldsymbol{w}, \boldsymbol{u})$. The following proposition provides one such example, which proves particularly relevant in our analysis of Problem 1.

Proposition 1. *Let $f(\boldsymbol{w}, u) = c(u) + g(b_0 + \boldsymbol{b}^T\boldsymbol{w} + u)$, where $c, g : \mathbb{R} \to \mathbb{R}$ are arbitrary convex functions, and $\boldsymbol{b} \geq 0$ or $\boldsymbol{b} \leq 0$. Then, condition* [**A1**] *is satisfied.*

Proof. Since f is jointly convex in $\boldsymbol{w}$ and u, $f^\star$ is convex. Furthermore, note that $f^\star$ only depends on $\boldsymbol{w}$ through $\boldsymbol{b}^T\boldsymbol{w}$, i.e., $f^\star(\boldsymbol{w}) = \tilde{f}(\boldsymbol{b}^T\boldsymbol{w})$, for some convex $\tilde{f}$. Therefore, since $\boldsymbol{b} \geq 0$ or $\boldsymbol{b} \leq 0$, $f^\star$ is supermodular (see Lemma 2.6.2 in Topkis [1998]). □

Condition [**A2**] can also be tested by directly examining the function f. For instance, if f is jointly convex in $\boldsymbol{w}$ and $\boldsymbol{u}$, then [**A2**] is trivially satisfied, as is the case in the example of Proposition 1.

In practice, the most cumbersome condition to test is undoubtedly [**A3**]. Typically, a combination of comparative statics analyses and structural properties on the function f will be needed. We exhibit how such techniques can be used by making reference, again, to the example in Proposition 1.

Proposition 2. *Let $f(\boldsymbol{w}, u) = c(u) + g(b_0 + \boldsymbol{b}^T\boldsymbol{w} + u)$, where $c, g : \mathbb{R} \to \mathbb{R}$ are arbitrary convex functions, and $\boldsymbol{b} \geq 0$ or $\boldsymbol{b} \leq 0$. Then, condition* [**A3**] *is satisfied.*

Proof. Let $h(x, y) \stackrel{\text{def}}{=} c(y) + g(x + y)$. It can be shown (see Lemma 6 of the Appendix, or Heyman and Sobel [1984] and Theorem 3.10.2 in Topkis [1998]) that $\arg\min_y h(x, y)$ is decreasing in x, and $x + \arg\min_y h(x, y)$ is increasing in x.

Consider any $\hat{\boldsymbol{w}} \in \arg\max_{\boldsymbol{w} \in P} f^\star(\boldsymbol{w}) \cap \operatorname{ext}(P)$. In this case, the construction in (7) yields

$$\forall\, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}} \;:\; (\boldsymbol{q}^\pi)^T\boldsymbol{w} + q_0^\pi = u^\star(\boldsymbol{w}) \equiv y^\star(b_0 + \boldsymbol{b}^T\boldsymbol{w}),\, \forall\, \boldsymbol{w} \in \operatorname{ext}(\Delta_\pi),$$

for some $y^\star(x) \in \arg\min_y h(x, y)$. We claim that:

$$\boldsymbol{b} \geq 0 \;\Rightarrow\; \boldsymbol{q}^\pi \leq 0 \text{ and } \boldsymbol{b} + \boldsymbol{q}^\pi \geq 0,\, \forall\, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}} \tag{15a}$$

$$\boldsymbol{b} \leq 0 \;\Rightarrow\; \boldsymbol{q}^\pi \geq 0 \text{ and } \boldsymbol{b} + \boldsymbol{q}^\pi \leq 0,\, \forall\, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}. \tag{15b}$$

We prove the first claim (the second follows analogously). Since $\mathbf{0} \in \operatorname{ext}(\Delta_\pi)$, we have $q_0^\pi = y^\star(b_0)$. If $\boldsymbol{b} \geq 0$, then the monotonicity of $y^\star(x)$ implies that

$$q_0^\pi + q_i^\pi = y^\star(b_0 + b_i) \leq y^\star(b_0),\, \forall\, i \in \{1, \ldots, n\},$$

which implies that $\boldsymbol{q}^\pi \leq 0$. Similarly, the monotonicity of $x + y^\star(x)$ implies that $\boldsymbol{b} + \boldsymbol{q}^\pi \geq 0$.

With the previous two claims, it can be readily seen that the functions

$$f^\pi(\boldsymbol{w}) = c\big((\boldsymbol{q}^\pi)^T\boldsymbol{w} + q_0^\pi\big) + g\big(b_0 + q_0^\pi + (\boldsymbol{b} + \boldsymbol{q}^\pi)^T\boldsymbol{w}\big)$$

are convex in $\boldsymbol{w}$ and supermodular on $\operatorname{ext}(P)$, and that the same conclusion holds for affine policies given by arbitrary convex combinations of $(\boldsymbol{q}^\pi, q_0^\pi)$, hence [**A3**] must hold. □

In view of Proposition 1 and Proposition 2, we have the following example where Theorem 1 readily applies, and which will prove essential in the discussion of the two-echelon example of Problem 1.

Lemma 3. *Let $f(\boldsymbol{w}, u) = h(\boldsymbol{w}) + c(u) + g(b_0 + \boldsymbol{b}^T\boldsymbol{w} + u)$, where $h : \mathcal{H}_n \to \mathbb{R}$ is convex and supermodular on $\{0, 1\}^n$, and $c, g : \mathbb{R} \to \mathbb{R}$ are arbitrary convex functions. Then, if either $\boldsymbol{b} \geq 0$, $\boldsymbol{b} \leq 0$ or h is affine, there exist $\boldsymbol{q} \in \mathbb{R}^n$, $q_0 \in \mathbb{R}$ such that*

$$\max_{\boldsymbol{w}\in P} f^\star(\boldsymbol{w}) = \max_{\boldsymbol{w}\in P} f(\boldsymbol{w}, \boldsymbol{q}^T\boldsymbol{w} + q_0) \tag{16a}$$

$$\operatorname{sign}(\boldsymbol{q}) = -\operatorname{sign}(\boldsymbol{b}) \tag{16b}$$

$$\operatorname{sign}(\boldsymbol{b} + \boldsymbol{q}) = \operatorname{sign}(\boldsymbol{b}) \tag{16c}$$

Proof. The optimality for the case of $\boldsymbol{b} \geq 0$ and $\boldsymbol{b} \leq 0$ follows directly from Proposition 1 and Proposition 2 (note that adding the convex and supermodular function h does not change any of the arguments there). The proofs for the sign relations concerning $\boldsymbol{q}$ follow from (15a) and (15b), by recognizing that the same inequalities hold for any convex combination of the vectors $\boldsymbol{q}^\pi$.

When h is affine, then the case with an arbitrary $\boldsymbol{b}$ can be transformed, by a suitable linear change of variables for $\boldsymbol{w}$, to a case with $\boldsymbol{b} \geq 0$, and modified b_0 and affine h. ☐

3.3 Application to Problem 1

In this section, we revisit the production planning model discussed in Problem 1 of the introduction, where the full power of the results introduced in Section 3 can be used to derive the optimality of ordering policies that are affine in historical demands.

As remarked in the introduction, a very similar model has been originally considered in Ben-Tal et al. [2005b] and Ben-Tal et al. [2009]; we first describe our model in detail, and then discuss how it relates to that in the other two references.

Let $1, \ldots, T$ denote the finite planning horizon, and introduce the following variables:

- K_t : the installed capacity at the supplier in period t. With $\boldsymbol{K} = (K_1, \ldots, K_T)$, let $r(\boldsymbol{K})$ denote the investment cost for installing capacity vector $\boldsymbol{K}$.
- p_t : the contractual pre-commitment for period t. Let $\boldsymbol{p} = (p_1, \ldots, p_T)$ be the vector of pre-committed orders.
- q_t : the realized order quantity from the retailer in period t. The corresponding cost incurred by the retailer is $c_t(q_t, \boldsymbol{K}, \boldsymbol{p})$, and includes the actual purchasing cost, as well as penalties for over or under ordering.
- I_t : the inventory on the premises of the retailer at the *beginning* of period $t \in \{1, \ldots, T\}$. Let $h_t(I_{t+1})$ denote the holding/backlogging cost incurred at the *end* of period t.
- d_t : unknown customer demand in period t. We assume that the retailer has very limited information about the demands, so that only bounds are available, $d_t \in \mathcal{D}_t = [\underline{d}_t, \overline{d}_t]$.

The problem of designing investment, pre-commitment and ordering decisions that would minimize the system-level cost in the worst case can then be re-written as

$$\min_{\boldsymbol{K},\boldsymbol{p}} \Bigg[r(\boldsymbol{K}) + \min_{q_1\geq 0} \Big[c_1(q_1, \boldsymbol{K}, \boldsymbol{p}) + \max_{d_1\in\mathcal{D}_1} \Big[h_1(I_1) + \cdots + \min_{q_T\geq 0} \big[c_T(q_T, \boldsymbol{K}, \boldsymbol{p}) + \max_{d_T\in\mathcal{D}_T} \ h_T(I_T) \big] \ldots \Big] \Big] \Bigg]$$
$$\text{s.t. } I_{t+1} = I_t + q_t - d_t, \quad \forall\, t \in \{1, 2, \ldots, T\}.$$

By introducing the class of ordering policies that depend on the history of observed demands,

$$q_t : \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_{t-1} \to \mathbb{R}, \tag{17}$$

we claim that the theorems of Section 3 can be used to derive the following structural results.

Theorem 2. *Assume the inventory costs h_t are convex, and the ordering costs $c_t(q_t, \boldsymbol{K}, \boldsymbol{p})$ are convex in q_t for any fixed $\boldsymbol{K}, \boldsymbol{p}$. Then, for any fixed $\boldsymbol{K}, \boldsymbol{p}$,*

1. *Ordering policies that depend <u>affinely</u> on the history of demands are worst-case optimal.*
2. *Each affine order occurring after period t is partially satisfying the demands that are still backlogged in period t.*

Before presenting the proof, we discuss the meaning of the result, and comment on the related literature. The first claim confirms that ordering policies depending affinely on historical demands are (worst-case) optimal, as soon as the first-stage capacity and order pre-commitment decisions are fixed. The second claim provides a structural decomposition of the affine ordering policies: every such order placed in or after period t can be seen as partially satisfying the demands that are still backlogged in period t, with the free coefficients corresponding to safety stock that is built in anticipation for future increased demands.

The model is related to that in Ben-Tal et al. [2005b, 2009] in several ways. With the exception of the investment cost $r(\boldsymbol{K})$, the latter model has all the ingredients of ours, with the costs having the specific form

$$\begin{aligned} c_t(q_t, \boldsymbol{K}, \boldsymbol{p}) &= \tilde{c}_t \cdot q_t + \alpha_t^- \max(0,\, p_t - q_t) + \alpha_t^+ \max(0,\, q_t - p_t) + \\ &\qquad \beta_t^- \max(0,\, p_{t-1} - p_t) + \beta_t^+ \max(0,\, p_t - p_{t-1}), \\ h_t(I_{t+1}) &= \max(\tilde{h}_t\, I_{t+1},\, -b_t\, I_{t+1}). \end{aligned} \tag{18}$$

Here, $\tilde{c}_t$ is the per-unit ordering cost, $\alpha_t^{+/-}$ are the penalties for over/under-ordering (respectively) relative to the pre-commitments, $\beta_t^{+/-}$ are penalties for differences in pre-commitments for consecutive periods, $\tilde{h}_t$ is the per-unit holding cost, and b_t is the per-unit backlogging cost. Such costs are clearly piece-wise affine and convex, and hence fit the conditions of Theorem 2. Note that our model allows more general convex production costs, for instance, reflecting the purchase of units beyond the installed capacity at the supplier (e.g., from a different supplier or an open market), resulting in an extra cost $c_t^{om} \max(0,\, q_t - K_t)$.

The one feature present in Ben-Tal et al. [2005b], but absent from our model, are cumulative order bounds, of the form

$$\hat{L}_t \le \sum_{k=1}^{t} q_t \le \hat{H}_t,\ \forall\, t \in \{1, \ldots, T\}.$$

Such constraints have been shown to preclude the optimality of ordering policies that are affine in historical demands, even in a far simpler model (without investments, pre-commitments, and with linear ordering cost) - see Bertsimas et al. [2010]. Therefore, the result in Theorem 2 shows that these constraints are, in fact, the *only* modeling component in Ben-Tal et al. [2005b, 2009] that hinders the optimality of affine ordering policies.

We also mention some related literature in operations management to which our result might bear some relevance. A particular demand model, which has garnered attention in various operational problems, is the Martingale Model of Forecast Evolution (see Hausman [1969], Heath and

Jackson [1994], Graves et al. [1998], Chen and Lee [2009], Bray and Mendelson [2012] and references therein), whereby demands in future periods depend on a set of external demand shocks, which are observed in each period. In such models, it is customary to consider so-called Generalized Order-Up-To Inventory Policies, whereby orders in period t depend in an *affine* fashion on demand signals observed up to period t (see Graves et al. [1998], Chen and Lee [2009], Bray and Mendelson [2012]). Typically, the affine forms are considered for simplicity, and, to the best of our knowledge, there are no proofs concerning their optimality in the underlying models. In this sense, if we interpret the disturbances in our model as corresponding to particular demand shocks, our results may provide evidence that affine ordering policies (in historical demand shocks) are provably optimal for particular finite horizon, robust counterparts of the models.

3.3.1 Dynamic Programming Solution

In terms of solution methods, note that Problem 1 can be formulated as a Dynamic Program (DP) [Ben-Tal et al., 2005b, 2009]. In particular, for a fixed $\boldsymbol{K}$ and $\boldsymbol{p}$, the state-space of the problem is one-dimensional, i.e., the inventory I_t, and Bellman recursions can be written to determine the underlying optimal ordering policies $q_t^\star(I_t, \boldsymbol{K}, \boldsymbol{p})$ and value functions $J_t^\star(I_t, \boldsymbol{K}, \boldsymbol{p})$,

$$\begin{aligned} J_t(I, \boldsymbol{K}, \boldsymbol{p}) &= \min_{q \geq 0} \big[\, c_t(q, \boldsymbol{K}, \boldsymbol{p}) + g_t(I + q, \boldsymbol{K}, \boldsymbol{p}) \,\big], \\ g_t(y, \boldsymbol{K}, \boldsymbol{p}) &\stackrel{\text{def}}{=} \max_{d \in \mathcal{D}_t} \big[\, h_t(y - d) + J_{t+1}^\star(y - d) \,\big], \end{aligned} \tag{19}$$

where $J_{T+1}(I, \boldsymbol{K}, \boldsymbol{p})$ can be taken to be 0 or some other *convex* function of I, if salvaging inventory is an option (see Ben-Tal et al. [2005b] for details). With this approach, one can derive the following structural properties concerning the optimal policies and value functions.

Lemma 4. *Consider a fixed $\boldsymbol{K}$ and $\boldsymbol{p}$. Then,*

1. *Any optimal order quantity is non-increasing in starting inventory, i.e., $q_t^\star(I_t, \boldsymbol{K}, \boldsymbol{p})$ is non-increasing in I_t.*

2. *The optimal inventory position after ordering is non-decreasing in starting inventory, i.e., $I_t + q_t^\star(I_t, \boldsymbol{K}, \boldsymbol{p})$ is non-decreasing in I_t.*

3. *The value functions $J_t^\star(I_t, \boldsymbol{K}, \boldsymbol{p})$ and $g_t(y, \boldsymbol{K}, \boldsymbol{p})$ are convex in I_t and y, respectively.*

Proof. These properties are well-known in the literature on inventory management (see Example 8-15 in Heyman and Sobel [1984], Proposition 3.1 in Bensoussan et al. [1983] or Theorem 3.10.2 in Topkis [1998]), and follow by backwards induction, and a repeated application of Lemma 6 in the Appendix. We omit the complete details due to space considerations. □

When the convex costs c_t are also piece-wise affine, the optimal orders follow a *generalized basestock policy*, whereby a different base-stock is prescribed for every linear piece in c_t (see Porteus [2002]).

In terms of completing the solution of the original problem, once the value function $J_1(I_1, \boldsymbol{K}, \boldsymbol{p})$ is available, one can solve the problem $\min_{\boldsymbol{K},\boldsymbol{p}} J_1(I_1, \boldsymbol{K}, \boldsymbol{p})$. However, as outlined in Ben-Tal et al. [2005b, 2009], such an approach would encounter several difficulties in practice: (i) one may have to discretize I_t and q_t, and hence only produce an approximate value for J_1, (ii) the DP would have to be solved for any possible choice of $\boldsymbol{K}$ and $\boldsymbol{p}$, (iii) $J_1(I_1, \boldsymbol{K}, \boldsymbol{p})$ would, in general, be non-smooth, and (iv) the DP solution would provide no subdifferential information for J_1, leading to the use of zero-order (i.e., gradient-free) methods for solving the resulting first-stage problem, which exhibit notoriously slow convergence.

These results are in stark contrast with Theorem 2, which argues that affine ordering policies remain optimal for *arbitrary* convex ordering cost, i.e., the complexity of the policy does not increase with the complexity of the cost function. Furthermore, as we argue in Section 4, the *exact* solution for the case of piece-wise affine costs (such as those considered in Ben-Tal et al. [2005b, 2009]) can actually be obtained by solving a single LP, with manageable size.

3.3.2 Proof of Theorem 2

To simplify the notation, let $\boldsymbol{d}_{[t]} \stackrel{\text{def}}{=} (d_1, \dots, d_{t-1})$ denote the vector of demands known at the beginning of period t, residing in $\mathcal{D}_{[t]} \stackrel{\text{def}}{=} \mathcal{D}_1 \times \cdots \times \mathcal{D}_{t-1}$. Whenever $\boldsymbol{K}$ and $\boldsymbol{p}$ are fixed, we suppress the dependency on $\boldsymbol{K}$ and $\boldsymbol{p}$ for all quantities of interest, such as $q_t^\star$, $J_t^\star$, c_t, g_t, etc. The following lemma essentially proves the desired result in Theorem 2.

Lemma 5. *Consider a fixed $\boldsymbol{K}$ and $\boldsymbol{p}$. For every period $t \in \{1, \dots, T\}$, one can find an affine ordering policy $q_t^{\text{aff}}(\boldsymbol{d}_{[t]}) = \boldsymbol{q}_t^T \boldsymbol{d}_{[t]} + q_{t,0}$ such that*

$$J_1^\star(I_1) = \max_{\boldsymbol{d}_{[t+1]} \in \mathcal{D}_{[t+1]}} \left[\sum_{k=1}^{t} \Big(c_k(q_k^{\text{aff}}) + h_k(I_{k+1}^{\text{aff}}) \Big) + J_{t+1}^\star(I_{t+1}^{\text{aff}}) \right], \tag{20}$$

where $I_k^{\text{aff}}(\boldsymbol{d}_{[k]}) = \boldsymbol{b}_k^T \boldsymbol{d}_{[k]} + b_{k,0}$ denotes the affine dependency of the inventory I_k on historical demands, for any $k \in \{1, \dots, t\}$. Furthermore, we also have

$$\boldsymbol{b}_t \le \boldsymbol{0}, \quad \boldsymbol{q}_t \ge \boldsymbol{0}, \quad \boldsymbol{q}_t + \boldsymbol{b}_t \le \boldsymbol{0}. \tag{21}$$

Let us first interpret the main statements. Equation (20) guarantees that using the affine ordering policies in periods $k \in \{1, \dots, t\}$ (and then proceeding with the Bellman-optimal decisions in periods $t+1, \dots, T$) does not increase the overall optimal worst-case cost. As such, it essentially proves the first part of Theorem 2.

Relation (21) confirms the structural decomposition of the ordering policies: if a particular demand d_k no longer appears in the backlog at the beginning of period t (i.e., $\boldsymbol{b}_t^T \boldsymbol{1}_k = 0$), then the current ordering policy does not depend on d_k (i.e., $\boldsymbol{q}_t^T \boldsymbol{1}_k = 0$). Furthermore, if a fraction $-b_{t,k} \in (0, 1]$ of demand d_k is still backlogged in period t, the order q_t^{aff} will satisfy a fraction $q_{t,k} \in [0, -b_{t,k}]$ of this demand. Put differently, the affine orders decompose the fulfillment of any demand d_k into (a) existing stock in period k and (b) partial orders in periods $k, \dots, T$, which is exactly the content of the second part of Theorem 2.

Proof of Lemma 5. The proof is by forward induction on t. At $t = 1$, an optimal *constant* order is available from the DP solution, $q_1^{\text{aff}} = q_1^\star(I_1)$. Also, since $I_2 = I_1 + q_1^{\text{aff}} - d_1$, we have $\boldsymbol{b}_2 \le 0$.

Assuming the induction is true at stages $k \in \{1, \dots, t-1\}$, let us consider the problem solved by nature at time $t-1$, given by (20). The cumulative historical costs in stages $1, \dots, t-1$ are given by

$$\tilde{h}_t(\boldsymbol{d}_{[t]}) \stackrel{\text{def}}{=} \sum_{k=1}^{t-1} \Big(c_k(q_k^{\text{aff}}) + h_k(I_{k+1}^{\text{aff}}) \Big) = \sum_{k=1}^{t-1} \Big[c_k(\boldsymbol{q}_t^T \boldsymbol{d}_{[k]} + q_{k,0}) + h_k(\boldsymbol{b}_{k+1}^T \boldsymbol{d}_{[k+1]} + b_{k+1,0}) \Big].$$

By the induction hypothesis, $\boldsymbol{q}_k \geq \boldsymbol{0}$, $\boldsymbol{b}_k \leq \boldsymbol{0}$, $\forall\, k \in \{1,\dots,t-1\}$, and $\boldsymbol{b}_t \leq \boldsymbol{0}$. Therefore, since c_k and h_k are convex, the function $\tilde{h}_t$ is convex and *supermodular* in $\boldsymbol{d}_{[t]}$. Recalling that $J_t^\star$ is derived from the Bellman recursions (19), i.e.,

$$J_t^\star(I_t) = \min_{q\geq 0}\big[\, c_t(q) + g_t(I_t + q)\,\big].$$

we obtain that equation (20) can be rewritten equivalently as

$$J_1^\star(I_1) = \max_{\boldsymbol{d}\in\mathcal{D}_{[t]}}\Big[\,\tilde{h}_t(\boldsymbol{d}) + \min_{q_t\geq 0}\big[\, c_t(q_t) + g_t(\boldsymbol{b}_t^T\boldsymbol{d} + b_{t,0} + q_t)\,\big]\,\Big]. \tag{22}$$

In this setup, we can directly invoke the result of Lemma 3. Since the non-negativity constraints can be emulated via a suitable convex barrier, we readily obtain that there exists an *affine* ordering policy $q_t^{\text{aff}}(\boldsymbol{d}_{[t]}) \stackrel{\text{def}}{=} \boldsymbol{q}_t^T\boldsymbol{d}_{[t]} + q_{t,0}$, that is worst-case optimal for problem (22) above. Furthermore, Lemma 3 also states that $\operatorname{sign}(\boldsymbol{q}_t) = -\operatorname{sign}(\boldsymbol{b}_t)$ and $\operatorname{sign}(\boldsymbol{q}_t + \boldsymbol{b}_t) = \operatorname{sign}(\boldsymbol{b}_t)$, which completes the proof. $\square$

4 Discussion of Problem 3

As suggested in the introduction, the sole knowledge that affine decision rules are optimal might not necessarily provide a "simple" computational procedure for generating them. An immediate example of this is Problem 1 itself: to find optimal affine ordering policies $q_t^{\text{aff}}(\boldsymbol{d}_{[t]}) = \boldsymbol{q}_t^T\boldsymbol{d}_{[t]} + q_{t,0}$ for any fixed $\boldsymbol{K}$ and $\boldsymbol{p}$, we would have to solve the following optimization problem:

$$\min_{\{\boldsymbol{q}_t,\, q_{t,0}\}_{t=1}^T} \; \max_{\boldsymbol{d}_{[T+1]}\in\mathcal{D}_{[T+1]}} \; \sum_{t=1}^{T}\Big[\, c_t(q_t^{\text{aff}}) + h_t\Big(I_1 + \sum_{k=1}^{t}(q_t^{\text{aff}} - d_k)\Big)\Big] \tag{23a}$$

$$\text{s.t.} \qquad q_t^{\text{aff}}(\boldsymbol{d}_{[t]}) \geq 0,\; \forall\, \boldsymbol{d}_{[t]} \in \mathcal{D}_{[t]},\; \forall\, t \in \{1,\dots,T\}. \tag{23b}$$

While the constraints (23b) can be handled via standard techniques in robust optimization [Ben-Tal et al., 2009], the objective function is seemingly intractable, even when the convex costs c_t and h_t take the piece-wise affine form (18), also considered in Ben-Tal et al. [2005b, 2009].

With this motivation in mind, we now recall Problem 3 stated in the introduction, and note that it is exactly geared towards simplifying objectives of the form (23a). In particular, if the inner expression in (23a) depended *bi-affinely*[3] on the decision variables and the uncertain quantities, then standard techniques in robust optimization could be employed to derive tractable robust counterparts for the problem (see Ben-Tal et al. [2009] for a detailed overview). The following theorem summarizes our main result of this section, providing sufficient conditions that yield the desired outcome.

Theorem 3. *Consider an optimization problem of the form*

$$\max_{\boldsymbol{w}\in P}\Big[\,\boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} h_i(\boldsymbol{w})\,\Big],$$

where $P \subset \mathbb{R}^k$ is any polytope, $\boldsymbol{a} \in \mathbb{R}^n$ is an arbitrary vector, $\mathcal{I}$ is a finite index set, and $h_i : \mathbb{R}^n \to \mathbb{R}$ are functions satisfying the following properties

[3] That is, it would be affine in one set of variables when the other set is fixed.

[P1] *h_i are concave extendable from* $\mathrm{ext}(P)$, $\forall\, i \in \mathcal{I}$,

[P2] $\mathrm{conc}_P\,(h_i + h_j) = \mathrm{conc}_P\,(h_i) + \mathrm{conc}_P\,(h_j)$, *for any* $i \neq j \in \mathcal{I}$.

Then there exists a set of affine functions $z_i(\boldsymbol{w})$, $i \in \mathcal{I}$, *satisfying* $z_i(\boldsymbol{w}) \geq h_i(\boldsymbol{w})$, $\forall\, \boldsymbol{w} \in P$, $\forall\, i \in \mathcal{I}$, *such that*

$$\max_{\boldsymbol{w}\in P}\Big[\,\boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} z_i(\boldsymbol{w})\,\Big] \;= \max_{\boldsymbol{w}\in P}\Big[\,\boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} h_i(\boldsymbol{w})\,\Big].$$

Proof. The proof is slightly technical, so we choose to relegate it to Section 7.3 of the Appendix. □

Let us discuss the statement of Theorem 3 and relevant examples of functions satisfying the conditions therein. **[P1]** requires the functions h_i to be concave-extendable; by the discussion in Section 7.2 of the Appendix, examples of such functions are any convex functions or, when $P = \mathcal{H}_n$, any component-wise convex functions. More generally, concave-extendability can be tested using the sufficient condition provided in Lemma 7 of the Appendix.

Apriori, condition **[P2]** seems more difficult to test. Note that, by Theorem 8 in the Appendix, it can be replaced with any of the following equivalent requirements:

[P3] $\mathrm{conc}_P\,(h_i) + \mathrm{conc}_P\,(h_j)$ is concave-extendable from vertices, for any $i \neq j \in \mathcal{I}$

[P4] For any $i \neq j \in \mathcal{I}$, the linearity domains $\mathcal{R}_{h_i,P} = \{F_k \,:\, k \in \mathcal{K}\}$ and $\mathcal{R}_{h_j,P} = \{G_\ell \,:\, \ell \in \mathcal{L}\}$ of $\mathrm{conc}_P\,(h_i)$ and $\mathrm{conc}_P\,(h_j)$, respectively, are such that $F_k \cap G_\ell$ has all vertices in $\mathrm{ext}(P)$, $\forall\, k \in \mathcal{K}$, $\forall\, \ell \in \mathcal{L}$.

The choice of which condition to include should be motivated by what is easier to test in the particular application of interest. A particularly relevant class of functions satisfying both requirements **[P1]** and **[P2]** is the following.

Example 1. *Let P be a polytope of the form (3). Then, any functions h_i that are convex and supermodular on* $\mathrm{ext}(P)$ *satisfy the requirements* **[P1]** *and* **[P2]**.

The proof for this fact is the subject of Corollary 4 of the Appendix. An instance of this, which turns out to be particularly pertinent in the context of Problem 1, is $h_i(\boldsymbol{w}) = f_i(b_{i,0} + \boldsymbol{b}_i^T\boldsymbol{w})$, where $f_i : \mathbb{R} \to \mathbb{R}$ are convex functions, and $\boldsymbol{b}_i \geq \boldsymbol{0}$ or $\boldsymbol{b}_i \leq \boldsymbol{0}$. A further subclass of the latter is $P = \mathcal{H}_n$ and $\boldsymbol{b}_i = \boldsymbol{b} \geq 0$, $\forall\, i \in \mathcal{I}$, which was the object of a central result in Bertsimas et al. [2010] (Section 4.3 in that paper, and in particular Lemmas 4.8 and 4.9).

We remark that, while maximizing convex functions on polytopes is generally NP-hard (the max-cut problem is one such example [Pardalos and Rosen, 1986]), maximizing *supermodular* functions on lattices can be done in polynomial time [Fujishige, 2005]. Therefore, our result does not seem to have direct computational complexity implications. However, as we show in later examples, it does have the merit of drastically simplifying particular computational procedures, particularly when combined with outer minimization problems such as those present in many robust optimization problems.

As another subclass of Example 1, we include the following.

Example 2. *Let $P = \mathcal{H}_n$, and $h_i(\boldsymbol{w}) = \prod_{k\in\mathcal{K}_i} f_k(\boldsymbol{w})$, where $\mathcal{K}_i$ is a finite index set, and f_k are nonnegative, supermodular, and increasing (decreasing), for all $k \in \mathcal{K}_i$. Then h_i are convex and supermodular.*

This result follows directly from Lemma 2.6.4 in Topkis [1998]. One particular example in this class are all polynomials in $\boldsymbol{w}$ with non-negative coefficients. In this sense, Theorem 3 is useful in deriving a simple (linear-programming based) algorithm for the following problem.

Corollary 1. *Consider a polynomial p of degree d in variables $\boldsymbol{w} \in \mathbb{R}^n$, such that any monomial of degree at least two has positive coefficients. Then, there is linear programming formulation of size $\mathcal{O}(n^d)$ for solving the problem $\max_{\boldsymbol{w}\in[0,1]^n} p(\boldsymbol{w})$.*

Proof. Note first that the problem is non-trivial due to the presence of potentially negative affine terms. Representing p in the form $p(\boldsymbol{w}) = \boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} h_i(\boldsymbol{w})$, where each h_i has degree at least two, we can use the result in Theorem 3 to rewrite the problem equivalently as follows:

$$\begin{aligned}\max_{\boldsymbol{w}\in[0,1]^n} p(\boldsymbol{w}) = \min_{t,\{\boldsymbol{z}_i, z_{i,0}\}_{i\in\mathcal{I}}} \; & t \\ \text{s.t.} \;\; & t \geq \boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} (z_{i,0} + \boldsymbol{z}_i^T\boldsymbol{w}),\ \forall\, \boldsymbol{w} \in [0,1]^n \qquad (*) \\ & h_i(\boldsymbol{w}) \leq z_{i,0} + \boldsymbol{z}_i^T\boldsymbol{w},\ \forall\, \boldsymbol{w} \in [0,1]^n. \qquad (**)\end{aligned}$$

By Theorem 3, the semi-infinite LP on the right-hand side has the same optimal value as the problem on the left. Furthermore, standard techniques in robust optimization can be invoked to reformulate constraints (*) in a tractable fashion (see Ben-Tal et al. [2009] for details), and constraints (**) can be replaced by a finite enumeration over at most 2^d extreme points of the cube (since each monomial term h_i has degree at most d). Therefore, the semi-infinite LP can be rewritten as an LP of size $\mathcal{O}(n^d)$. □

4.1 Application to Problem 1

To exhibit how Theorem 3 can be used in practice, we again revisit Problem 1. More precisely, recall that one had to solve the seemingly intractable optimization problem in (23a) and (23b) in order to find the optimal affine orders q_t^{aff} for any fixed first-stage decisions $\boldsymbol{K}$, $\boldsymbol{p}$, and this was the case even when all the problem costs were piecewise-affine.

In this context, the following result in a direct application of Theorem 3.

Theorem 4. *Assume the costs c_t, h_t and r are jointly convex and piece-wise affine, with at most m pieces. Then, the optimal $\boldsymbol{K}$, $\boldsymbol{p}$, and a set of worst-case optimal ordering policies $\{q_t^{\text{aff}}\}_{t\in\{1,\ldots,T\}}$ can be computed by solving a single linear program with $\mathcal{O}(m \cdot T^2)$ variables and constraints.*

Proof. Consider first a fixed $\boldsymbol{K}$ and $\boldsymbol{p}$. The expression for the inner objective in (23a) is

$$\sum_{t=1}^{T} \Big[c_t(q_t^{\text{aff}}, \boldsymbol{K}, \boldsymbol{p}) + h_t(I_{t+1}^{\text{aff}}) \Big],$$

where $I_t^{\text{aff}}(\boldsymbol{d}_{[t]}) = I_1 + \sum_{k=1}^{t-1}(q_k^{\text{aff}} - d_k) \overset{\text{def}}{=} \boldsymbol{b}_t^T\boldsymbol{d}_{[t]} + b_{t,0}$ is the expression for the inventory under affine orders. The functions c_t and h_t are convex. Furthermore, by Lemma 4, there exist worst-case optimal affine rules $q_t^{\text{aff}}(\boldsymbol{d}_{[t]}) = \boldsymbol{q}_t\boldsymbol{d}_{[t]} + q_{t,0}$ such that

$$\boldsymbol{q}_t \geq 0, \quad \boldsymbol{b}_{t+1} \leq 0, \quad \forall\, t \in \{1,\ldots,T\}.$$

Therefore, $c_t\big(q_t^{\text{aff}}(\boldsymbol{d}_{[t]}), \boldsymbol{K}, \boldsymbol{p}\big)$ and $h_t\big(I_{t+1}(\boldsymbol{d}_{[t+1]})\big)$, as functions of $\boldsymbol{d}_{[T+1]}$, are convex and *supermodular* on $\text{ext}(\mathcal{D}_{[T+1]})$, and fall directly in the realm of Theorem 3.

In particular, an application of the latter result implies the existence of a set of *affine* ordering costs $c_t^{\mathrm{aff}}(\boldsymbol{d}_{[t]}) = \boldsymbol{c}_t^T \boldsymbol{d}_{[t]} + c_{t,0}$ and *affine* inventory costs $z_t^{\mathrm{aff}}(\boldsymbol{d}_{[t+1]}) = \boldsymbol{z}_t^T \boldsymbol{d}_{[t]} + z_{t,0}$ such that:

$$\max_{\boldsymbol{d}_{[T+1]} \in \mathcal{D}_{[T+1]}} \sum_{t=1}^{T} \left[\, c_t(q_t^{\mathrm{aff}}, \boldsymbol{K}, \boldsymbol{p}) + h_t(I_{t+1}^{\mathrm{aff}}) \,\right] = \max_{\boldsymbol{d}_{[T+1]} \in \mathcal{D}_{[T+1]}} \sum_{t=1}^{T} \left(c_t^{\mathrm{aff}} + z_t^{\mathrm{aff}} \right)$$

$$c_t^{\mathrm{aff}}(\boldsymbol{d}_{[t+1]}) \geq c_t(q_k^{\mathrm{aff}}, \boldsymbol{K}, \boldsymbol{p}), \; \forall\, \boldsymbol{d}_{[t]} \in \mathcal{D}_{[t]} \qquad (*)$$

$$z_t^{\mathrm{aff}}(\boldsymbol{d}_{[t+1]}) \geq h_t\Big(I_{t+1}^{\mathrm{aff}}(\boldsymbol{d}_{[t+1]})\Big), \; \forall\, \boldsymbol{d}_{[t+1]} \in \mathcal{D}_{[t+1]}. \qquad (**)$$

With this transformation, the objective is a bi-affine function of the uncertainties $\boldsymbol{d}_{[T+1]}$ and the decision variables $\{\boldsymbol{c}_t, \boldsymbol{z}_t\}$. Furthermore, if the costs c_t and h_t are piece-wise affine, the constraints $(*)$ and $(**)$ can also be written as bi-affine functions of the uncertainties and decisions. For instance, suppose

$$c_t(q, \boldsymbol{K}, \boldsymbol{p}) = \max_{j \in \mathcal{J}_t} \Big\{ \boldsymbol{\alpha}_j^T (\, q, \boldsymbol{K}, \boldsymbol{p}\,) + \beta_j \Big\}, \; \forall\, t \in \{1, \ldots, T\},$$

for suitably sized vectors $\boldsymbol{\alpha}_j$, $j \in \cup_t \mathcal{J}_t$. Then, $(*)$ are equivalent to

$$\boldsymbol{c}_t^T \boldsymbol{d}_{[t]} + c_{t,0} \geq \boldsymbol{\alpha}_j^T \left(\, \boldsymbol{q}_t^T \boldsymbol{d}_{[t]} + q_{t,0}, \boldsymbol{K}, \boldsymbol{p} \,\right) + \beta_j,$$

which are bi-affine in $\boldsymbol{d}_{[T+1]}$ and the vector of decision variables $\boldsymbol{x} \overset{\mathrm{def}}{=} \big(\boldsymbol{K}, \boldsymbol{p}, \boldsymbol{q}_t, q_{t,0}, \boldsymbol{c}_t, c_{t,0}, \boldsymbol{z}_t, z_{t,0}\big)_{t \in \{1,\ldots,T\}}$. As such, the problem of finding the optimal capacity and order pre-commitments and the worst-case optimal policies can be written as a robust LP (see, e.g., Ben-Tal et al. [2005b] and Ben-Tal et al. [2009]), in which a typical constraint has the form

$$\lambda_0(\boldsymbol{x}) + \sum_{t=1}^{T} \lambda_t(\boldsymbol{x}) \cdot d_t \leq 0, \qquad \forall\, \boldsymbol{d} \in \mathcal{D}_{[T+1]}\,,$$

where $\lambda_i(\boldsymbol{x})$ are affine functions of the decision variables $\boldsymbol{x}$. It can be shown (see Ben-Tal et al. [2009] for details) that the previous semi-infinite constraint is equivalent to

$$\begin{cases} \lambda_0(\boldsymbol{x}) + \sum_{t=1}^{T} \Big(\lambda_t(\boldsymbol{x}) \cdot \frac{\underline{d}_t + \overline{d}_t}{2} + \frac{\overline{d}_t - \underline{d}_t}{2} \cdot \xi_t \Big) \leq 0 \\ -\xi_t \leq \lambda_t(\boldsymbol{x}) \leq \xi_t, \quad t = 1, \ldots, T\,, \end{cases} \tag{24}$$

which are linear constraints in the decision variables $\boldsymbol{x}, \boldsymbol{\xi}$. Therefore, the problem of finding the optimal parameters can be reformulated as an LP with $O(m\,T^2)$ variables and $O(m\,T^2)$ constraints, which can be solved very efficiently using commercially available software. □

5 Conclusions

In this paper, we strived to bridge two well-established paradigms for solving robust dynamic problems. The first is Dynamic Programming - a methodology with very general scope, which allows insightful comparative statics analyses, but suffers from the curse of dimensionality, which limits its use in practice. The second involves the use of decision rules, i.e., policies parameterized in model uncertainties which are typically obtained by restricting attention to particular functional forms and solving tractable convex optimization problems. The main downside to the latter approach is the lack of control over the degree of suboptimality of the resulting decisions.

In this paper, we focus on the popular class of *affine* decision rules, and discuss sufficient conditions on the value functions of the dynamic program and the uncertainty sets, which ensure their optimality. We exemplify our findings in an application concerning the design of flexible contracts in a two-echelon supply chain, where the optimal contractual pre-commitments and the optimal ordering quantities can be found by solving a single linear program of small size. From a theoretical standpoint, our results emphasize the interplay between the convexity and supermodularity of the value functions, and the lattice structure of the uncertainty sets, suggesting new modeling paradigms for dynamic robust optimization.

6 Acknowledgements

The authors would also like to thank Mohit Tawarmalani and Jean-Philippe Richard for fruitful conversations and help with the proof of Lemma 8.

7 Appendix

7.1 Lattice Theory and Supermodularity

The proofs in the current paper use several concepts from the theory of lattice programming and supermodular functions, which we formally define here. The presentation follows closely Milgrom and Shannon [1994] and Topkis [1998], to which we direct the interested reader for proofs and a detailed treatment of the subject.

Let X be any set equipped with a transitive, reflexive, antisymmetric order relation $\geq$. For elements $\boldsymbol{x}, \boldsymbol{y} \in X$, let $\boldsymbol{x} \vee \boldsymbol{y}$ denote the least upper bound (or the join) of $\boldsymbol{x}$ and $\boldsymbol{y}$ (if it exists), and let $\boldsymbol{x} \wedge \boldsymbol{y}$ denote the greatest lower bound (or the meet) of $\boldsymbol{x}$ and $\boldsymbol{y}$ (if it exists).

Definition 1. *The set X is a* lattice *if for every pair of elements $\boldsymbol{x}, \boldsymbol{y} \in X$, the join and the meet exist and are elements of X.*

Similarly, $S \subset X$ is a *sublattice* if it is closed under the join and meet operations. In our treatment, the typical lattices under consideration are subsets of the hypercube $\mathcal{H}_n = [0, 1]^n$. Therefore, the operations $\geq$ and $\leq$ are understood in component-wise fashion, and $\wedge$ ($\vee$) are given by component-wise minimum (maximum).

Our analysis requires stating when the sets of maximizers (or minimizers) of a function is increasing or decreasing in particular state variables. To compare two such sets, we use the *strong set order* introduced by Veinott [1989]. If X is a lattice with the relation $\geq$, and Y, Z are elements of the power set of X, we say that $Y \geq Z$ if, for every $\boldsymbol{y} \in Y$ and $\boldsymbol{z} \in Z$, $\boldsymbol{y} \vee \boldsymbol{z} \in Y$ and $\boldsymbol{y} \wedge \boldsymbol{z} \in Z$. For instance, $[2, 4] \geq [1, 3]$, but $[1, 5] \not\geq [2, 4]$ and $[2, 4] \not\geq [1, 5]$. Analogous definitions hold for the $\leq$ relation.

Definition 2. *For a lattice $S \subseteq \mathbb{R}^n$, a function $f : S \to \mathbb{R}$ is said to be supermodular if $f(\boldsymbol{x}' \wedge \boldsymbol{x}'') + f(\boldsymbol{x}' \vee \boldsymbol{x}'') \geq f(\boldsymbol{x}') + f(\boldsymbol{x}'')$, for all $\boldsymbol{x}'$ and $\boldsymbol{x}'' \in S$.*

Similarly, a function f is called *submodular* if $-f$ is supermodular. Supermodular and submodular functions have been studied extensively in various fields, such as physics [Choquet, 1954], economics (Schmeidler [1986], Topkis [1998], Milgrom and Shannon [1994]) combinatorial optimization (Lovász [1982], Schrijver [2003], Fujishige [2005]), or mathematical finance [Föllmer and

Schied, 2004], to name only a few. They also play a central role in our treatment, since they admit a compact characterization for their concave envelopes.

Apart from the definition, several methods are known for testing whether a function is supermodular. One such test, applicable to functions $f : \mathbb{R}^n \to \mathbb{R}$ that are twice continuously differentiable, is $\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0, \forall\, i \neq j \in \{1, \ldots, n\}$. Two particular examples that occur often throughout our analysis are the following.

Example 3 (Lemma 2.6.2 in Topkis [1998]). *Suppose $Y \subseteq \mathbb{R}$ is a convex set, X is a sublattice of $\mathbb{R}^n$, $\boldsymbol{a} \in \mathbb{R}^n$ is a vector that satisfies $\boldsymbol{a}^T\boldsymbol{x} \in Y, \forall\, \boldsymbol{x} \in X$, $g : Y \to \mathbb{R}$, and $f(\boldsymbol{x}) = g(\boldsymbol{a}^T\boldsymbol{x})$. Then, f is supermodular in $\boldsymbol{x}$ on X if one of the following conditions holds:*

- *$\boldsymbol{a} \geq 0$ and g is convex on Y.*
- *$n = 2$, $\operatorname{sign}(a_1) = -\operatorname{sign}(a_2)$, and g is concave on Y.*

We note that the results above hold even when g is not twice continuously differentiable. For an overview of many other relevant classes of supermodular functions, we direct the interested reader to Topkis [1998] and Fujishige [2005].

As suggested earlier, we are interested in characterizing conditions when the set of maximizers (or minimizers) of a function is increasing (decreasing) with particular problem parameters. The following result provides a fairly general set of such conditions.

Theorem 5 (Theorem 2.8.2 in Topkis [1998]). *If X and T are lattices, S is a sublattice of $X \times T$, S_t is the section of S at $\boldsymbol{t} \in T$, and $f(\boldsymbol{x}, \boldsymbol{t})$ is supermodular in $(\boldsymbol{x}, \boldsymbol{t})$ on S, then $\arg\max_{x \in S_t} f(\boldsymbol{x}, \boldsymbol{t})$ is increasing in $\boldsymbol{t}$ on $\{\boldsymbol{t} : \arg\max_{\boldsymbol{x} \in S_t} f(\boldsymbol{x}, \boldsymbol{t}) \neq \emptyset\}$.*

We note that more general conditions are known in the literature, based on concepts such as quasisupermodular functions (see, e.g., Milgrom and Shannon [1994]). However, the result above suffices for our purposes in the present paper. To see how it can be used in a concrete setting, we include the following example, which is a well-known result in operations research (see, e.g., Example 8-15 in Heyman and Sobel [1984], Proposition 3.1 in Bensoussan et al. [1983] or Theorem 3.10.2 in Topkis [1998]), which is very useful in our analysis of Problem 1. We include its derivation here for completeness.

Lemma 6. *Let $f(x, u) = c(u) + g(x + u)$, where $c, g : \mathbb{R} \to \mathbb{R}$ are arbitrary convex functions. Then, $\arg\min_u f(x, u)$ is decreasing in x, and $x + \arg\min_u f(x, u)$ is increasing in x.*

Proof. Note first that

$$\min_u \big[c(u) + g(x + u)\big] = -\max_u \big[-c(u) - g(x + u)\big] \overset{(r \overset{\text{def}}{=} -u)}{=} -\max_r \big[-c(-r) - g(x - r)\big]$$

Since g is convex, the function $-c(-r) - g(x - r)$ is supermodular in (x, r) on the lattice $\mathbb{R} \times \mathbb{R}$. Therefore, by Theorem 5, $\arg\max_r \big[-c(-r) - g(x - r)\big]$ is increasing in x, which implies that $\arg\min_u f(x, u)$ is decreasing in x. In a similar fashion, letting $y \overset{\text{def}}{=} x + u$, it can be argued that the set $\arg\min_y f(x, y)$ is increasing in x, which concludes the proof. □

We remark that the monotonicity conditions derived above would hold even if constraints of the form $L \leq u \leq H$ (for $L < H \in \mathbb{R}$) were added. In fact, both can be simulated by adding suitable convex barriers to the cost function c.

7.2 Convex and Concave Envelopes

Our proofs make use of several known results concerning concave envelopes of functions, which are summarized below. The notation and statements follow quite closely those of Tardella [2008] and Tawarmalani et al. [2010], to which we refer the interested reader for a more comprehensive overview and references.

Definition 3. *Consider a function $f : S \to \mathbb{R}$, where S is a non-empty convex subset of $\mathbb{R}^n$. The function* $\text{conc}_S(f) : S \to \mathbb{R}$ *is said to be the* concave envelope *of f over S if and only if*

(i) $\text{conc}_S(f)$ *is concave over S*

(ii) $\text{conc}_S(f)(\boldsymbol{x}) \geq f(\boldsymbol{x}), \forall \boldsymbol{x} \in S$

(iii) $\text{conc}_S(f)(\boldsymbol{x}) \leq h(\boldsymbol{x})$, *for any concave $h(\boldsymbol{x})$ satisfying $h(\boldsymbol{x}) \geq f(\boldsymbol{x})$.*

In words, $\text{conc}_S(f)$ is the point-wise smallest concave function defined on S that over-estimates f. An example is included in Figure 3. In a similar fashion, one can define the *convex envelope* of f, denoted by $\text{conv}_S(f)$, as the point-wise largest convex under-estimator of f on S. For the rest of the exposition, we focus attention on concave envelopes, but all the concepts and results can be translated in a straightforward manner to convex envelopes, by recognizing that $\text{conv}_S(f) = -\text{conc}_S(-f)$.

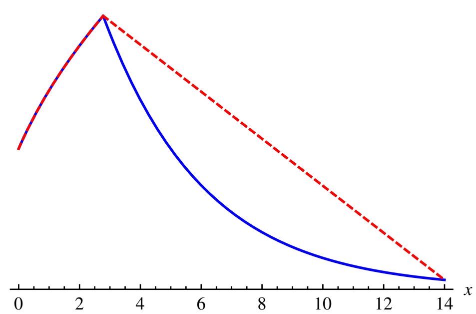


Figure 3: Example of a function $f : [0, 14] \to \mathbb{R}$ (solid line) and its concave envelope $\text{conc}_{[0,14]}(f)$ (dashed line).

One of the main reasons for the interest in concave envelopes is the fact that the set of global maxima of f is contained in the set of global maxima of $\text{conc}_S(f)$, and the two maximum values coincide. Expressing the concave envelope of a function is a difficult task in general, and even evaluating $\text{conc}_S(f)$ at a particular point $\boldsymbol{x}$ can be as hard as minimizing the function f [Tardella, 2008]. In some cases, however, concave envelopes can be constructed by restricting attention to a subset of the points in the domain S. One such instance, particularly relevant to the treatment in our paper, is summarized in the following definition.

Definition 4. *A function $f : P \to \mathbb{R}$, where P is a non-empty polytope, is said to be* concave-extendable from the set $S \subset P$ *if the concave envelope of f over P is the same as the concave envelope of the function $f|_S$ over P, where*

$$f|_S(\boldsymbol{x}) \stackrel{\text{def}}{=} \begin{cases} f(\boldsymbol{x}), & \boldsymbol{x} \in S \\ -\infty, & \textit{otherwise.} \end{cases}$$

When $S = \text{ext}(P)$, we say that f is concave-extendable from the vertices of P. Such functions are known to admit piece-wise affine concave envelopes, which further generate a relevant partition of the polytope P (this connection and other relevant results are included in Section 7.2.1). A natural question, in this context, is how to recognize a function that is concave-extendable from vertices. To the best of our knowledge, the most general characterization in the literature seems to be the following result from Tardella [2008].

Lemma 7 (Corollary 3 in Tardella [2008])**.** *Let $\mathscr{D}$ be a set of vectors in $\mathbb{R}^n$ parallel to some edges of the polytope P. Let f be a function that is convex[4] on P along all directions in $\mathscr{D}$, and let S denote the union of the faces of P (including the zero-dimensional faces $\text{ext}(P)$) that do not have any edge parallel to a direction in $\mathscr{D}$. Then, $\text{conc}_P(f) = \text{conc}_S(f)$. In particular, if f is edge-convex on P (i.e., $\mathscr{D}$ is maximal), then f is concave-extendable from $\text{ext}(P)$.*

This characterization yields several interesting functions. For instance, any f that is convex on P is concave-extendable; when P is a hypercube, any f that is component-wise convex is also concave-extendable (an example in the latter category often studied in the literature is the case of monomials). For more examples and references, the interested reader can check Tardella [2008] and Tawarmalani et al. [2010].

7.2.1 Concave Envelopes of Concave-Extendable Functions

Concave-extendable functions are known to admit polyhedral concave envelopes, i.e., concave envelopes that are given by the minimum of a finite collection of affine functions. The resulting concave envelopes also induce a polyhedral subdivision[5] of the domain P, which is relevant for several results in our treatment. To illustrate this connection, following Tawarmalani et al. [2010], let $V \in \mathbb{R}^{n \cdot |\text{ext}\, P|}$ denote a matrix with columns V_i given by the vertices of P, let $f(V) \stackrel{\text{def}}{=} \big(f(\boldsymbol{V}_1), f(\boldsymbol{V}_2), \ldots, f(\boldsymbol{V}_{|\text{ext}\, P|})\big)$, and consider the following primal-dual pair of linear programs

$$
\begin{aligned}
P(\boldsymbol{x}) \stackrel{\text{def}}{=} \min_{\boldsymbol{a},b} \;& \boldsymbol{a}^T\boldsymbol{x} + b \\
\text{s.t. } & \boldsymbol{a}^T V + \boldsymbol{e}^T b \geq f(V) \\
& \boldsymbol{a} \in \mathbb{R}^n,\ b \in \mathbb{R}
\end{aligned}
\qquad\qquad
\begin{aligned}
D(\boldsymbol{x}) \stackrel{\text{def}}{=} \max_{\boldsymbol{\lambda}} \;& f(V)^T\boldsymbol{\lambda} \\
\text{s.t. } & V\boldsymbol{\lambda} = \boldsymbol{w} \\
& \boldsymbol{e}^T\boldsymbol{\lambda} = 1 \\
& \boldsymbol{\lambda} \geq 0
\end{aligned}
\tag{25}
$$

It can be shown (see Rockafellar [1970] or Tawarmalani et al. [2010]) that the optimal values in both programs are finite, and equal to $\text{conc}_P(f)(\boldsymbol{x})$. Moreover, let $\mathcal{D}_{f,P}$ denote the feasible region of the primal program (which only depends on f and P, and is independent of $\boldsymbol{x}$), and, for a given $(\boldsymbol{a}, b) \in \mathcal{D}_{f,P}$, let $J(\boldsymbol{a}, b)$ be the index set of constraints of $\mathcal{D}_{f,P}$ that are tight at $(\boldsymbol{a}, b)$, let $V\big(J(\boldsymbol{a}, b)\big)$ be the matrix obtained from V by keeping the columns in $J(\boldsymbol{a}, b)$, and let $R(\boldsymbol{a}, b) \stackrel{\text{def}}{=} \text{conv}\big(V\big(J(\boldsymbol{a}, b)\big)\big)$. Then, the following theorem summarizes several relevant properties of the linear programs above, and their connection with $\text{conc}_P(f)$.

Theorem 6 (Theorem 2.4 in Tawarmalani et al. [2010])**.** *Consider a function $f : P \to \mathbb{R}$ which is concave-extendable from the vertices of P, where P is a full-dimensional polytope in $\mathbb{R}^n$. Then,*

[4] Tardella deals with convex envelopes, and his definitions are in terms of edge-concave functions. All of his results can be ported here by essentially switching *convex* with *concave*.

[5] For a polytope P, a set of n-dimensional polyhedra $P_1, \ldots, P_m \subseteq P$ is said to be a *polyhedral subdivision* of P if $P = \cup_{i=1}^m P_i$ and $P_i \cap P_j$ is a (possibly empty) face of both P_i and P_j.

1. *The optimal values in* $P(\boldsymbol{x})$ *and* $D(\boldsymbol{x})$ *are the same, and equal to* $\operatorname{conc}_P(f)(\boldsymbol{x})$.

2. *Let* $(\bar{\boldsymbol{a}}, \bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})$. *Then,* $(\bar{\boldsymbol{a}}, \bar{b})$ *is optimal for* $P(\boldsymbol{x})$ *if and only if* $\boldsymbol{x} \in R(\bar{\boldsymbol{a}}, \bar{b})$. *Further, the extreme points of* $\mathcal{D}_{f,P}$ *are in one-to-one correspondence with the non-vertical facets of* $\operatorname{conc}_P(f)$.

3. *For any* $(\bar{\boldsymbol{a}}, \bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})$, *the inequality* $\bar{\boldsymbol{a}}^T \boldsymbol{x} + \bar{b} \geq f(\boldsymbol{x})$ *defines a facet of* $\operatorname{conc}_P(f)$ *over* $R(\bar{\boldsymbol{a}}, \bar{b})$.

4. $\mathcal{R}_{f,P} \stackrel{\text{def}}{=} \left\{R(\bar{\boldsymbol{a}}, \bar{b}) : (\bar{\boldsymbol{a}}, \bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})\right\}$ *is a polyhedral subdivision of* $\operatorname{conv}(V)$, *and* $\operatorname{conc}_P(f)$ *can be computed by interpolating* f *affinely over each element of* $\mathcal{R}$.

Proof. This is a direct adaptation of Theorem 2.4 in Tawarmalani et al. [2010], to which we direct the reader for a proof and discussion. □

The previous theorem essentially states that $\operatorname{conc}_P(f)$ is given by affine interpolations of f over a particular polyhedral subdivision of P, given by the polytopes $R(\bar{\boldsymbol{a}}, \bar{b})$, for $(\bar{\boldsymbol{a}}, \bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})$ (also known as the *linearity domains* of $\operatorname{conc}_P(f)$ [Tardella, 2008]). From this result, utilizing the same notation as before, one can derive the following characterization concerning the problem of maximizing f over P.

Corollary 2. *For any full-dimensional polytope* P *and any concave-extendable function* $f : P \to \mathbb{R}$, *we have*

$$\begin{aligned}
\max_{\boldsymbol{x}\in P} f(\boldsymbol{x}) = \max_{\boldsymbol{x},t} \ & t \\
\text{s.t. } & t \leq \boldsymbol{a}^T \boldsymbol{x} + b, \ \forall\, (\boldsymbol{a}, b) \in \operatorname{ext}(\mathcal{D}_{f,P}). \\
& \boldsymbol{x} \in P.
\end{aligned}$$

Proof. For any function f, we have $\max_{\boldsymbol{x}\in P} f(\boldsymbol{x}) = \max_{\boldsymbol{x}\in P} \operatorname{conc}_P(f)(\boldsymbol{x})$. By Theorem 6, the latter function is exactly given by

$$\operatorname{conc}_P(f)(\boldsymbol{x}) = \min_{(\boldsymbol{a},b)\in \operatorname{ext}(\mathcal{D}_{f,P})} \boldsymbol{a}^T \boldsymbol{x} + b,$$

which immediately leads to the conclusion of the corollary. □

One particular case that is very relevant in our analysis is that of concave-extendable functions f defined on the unit hypercube, i.e., $P = \mathcal{H}_n = [0,1]^n$, which are also *supermodular*. It turns out that the concave envelope of any such function can be compactly described by the *Lovász extension* of the function f.

Definition 5 (Lovász [1982])**.** *Given any* $\boldsymbol{x} \in \mathcal{H}_n$, *find a permutation* $\pi \in \Pi(\{1, \ldots, n\})$ *such that* $x_{\pi(1)} \geq x_{\pi(2)} \geq \cdots \geq x_{\pi(n)}$. *Then, the* Lovász extension *of the function* $f : \mathcal{H}_n \to \mathbb{R}$ *at the point* $\boldsymbol{x}$ *is given by*

$$\begin{aligned}
f^{\mathcal{L}}(\boldsymbol{x}) &\stackrel{\text{def}}{=} \left(1 - x_{\pi(1)}\right) f(\boldsymbol{0}) + \sum_{j=1}^{n-1} (x_{\pi(j)} - x_{\pi(j+1)}) f\left(\sum_{r=1}^{j} \boldsymbol{1}_{\pi(r)}\right) + x_{\pi(n)} f\left(\sum_{r=1}^{n} \boldsymbol{1}_{\pi(r)}\right) \\
&= f(\boldsymbol{0}) + \sum_{i=1}^{n} \left[f\left(\sum_{j=1}^{i} \boldsymbol{1}_{\pi(j)}\right) - f\left(\sum_{j=1}^{i-1} \boldsymbol{1}_{\pi(j)}\right) \right] x_{\pi(i)}.
\end{aligned} \tag{26}$$

It can be seen from the definition that the Lovász extension of f is given by an affine interpolation of f on simplicies of the form $\Delta_\pi \stackrel{\text{def}}{=} \operatorname{conv}\Big(\big\{\mathbf{0}+\sum_{j=1}^{k} \mathbf{1}_{\pi(j)} \,:\, k=0,\dots,n\big\}\Big)$. The collection of corresponding simplicies $\{\Delta_\pi\}_{\pi\in\Pi(\{1,\dots,n\})}$ is known as the *Kuhn triangulation* of the hypercube. Using a result by Lovász [1982], one can show the following remarkable fact.

Theorem 7 (Theorem 3.3 in Tawarmalani et al. [2010])**.** *Consider a function $f : \mathcal{H}_n \to \mathbb{R}$. The concave envelope of f over $\mathcal{H}_n$ is given by $f^{\mathcal{L}}$ if and only if f is supermodular when restricted to $\{0,1\}^n$ and concave-extendable from $\{0,1\}^n$.*

In the context of Theorem 6, this result immediately yields the following corollary, which provides a full characterization of the concave envelope of supermodular and concave-extendable functions on hypercubes.

Corollary 3. *Consider a function $f : \mathcal{H}_n \to \mathbb{R}$ that is supermodular on $\{0,1\}^n$ and concave-extendable from $\{0,1\}^n$. Then,*

1. *The concave envelope of f on $\mathcal{H}_n$ is given by*

$$\operatorname{conc}_{\mathcal{H}_n}(f)(\boldsymbol{x}) = f(\mathbf{0}) + \min_{\pi\in\Pi(\{1,\dots,n\})} \sum_{i=1}^{n}\left[f\left(\sum_{j=1}^{i}\mathbf{1}_{\pi(j)}\right) - f\left(\sum_{j=1}^{i-1}\mathbf{1}_{\pi(j)}\right)\right] x_{\pi(i)}.$$

2. *The set of inequalities $\boldsymbol{a}^T\boldsymbol{x}+b \ge f(\boldsymbol{x})$ defining non-vertical facets of $\operatorname{conc}_{\mathcal{H}_n}(f)$ is given by*

$$\operatorname{ext}(\mathcal{D}_{f,P}) = \Bigg\{(\boldsymbol{a},b)\in\mathbb{R}^{n+1} \,:\, b=f(\mathbf{0}),\ \boldsymbol{a} = \sum_{i=1}^{n}\left[f\left(\sum_{j=1}^{i}\mathbf{1}_{\pi(j)}\right) - f\left(\sum_{j=1}^{i-1}\mathbf{1}_{\pi(j)}\right)\right]\mathbf{1}_{\pi(i)},\ \text{for } \pi\in\Pi(\{1,\dots,n\})\Bigg\}.$$

3. *The polyhedral subdivision $\mathcal{R}_{f,\mathcal{H}_n}$ of $\mathcal{H}_n$ yielding the concave envelope is exactly the Kuhn triangulation.*

Proof. The proof is a direct application of Theorem 6 and Theorem 7 and is omitted (see Tawarmalani et al. [2010] for complete details). ☐

In fact, the above results hold for the more general case of polytopes whose extreme points are *integer sub-lattices of* $\{0,1\}^n$. Any such polytope P is given by

$$P = \mathcal{H}_n \cap \{\boldsymbol{x} \,:\, x_i \ge x_j,\ \forall\,(i,j)\in E\} \cap \{\boldsymbol{x} \,:\, x_i = 0,\ \forall\, i\in\mathcal{I}_0\} \cap \{\boldsymbol{x} \,:\, x_i = 1,\ \forall\, i\in\mathcal{I}_1\},$$

for some $E \subseteq \{1,\dots,n\}^2$ and $\mathcal{I}_{0,1} \subseteq \{1,\dots,n\}$ (see Tawarmalani et al. [2010] and the original reference Grötschel et al. [1988] for details). For any such lattice, the Definition 5 of the Lovász extension is modified, by only including permutations that are compatible with the pre-order on P, i.e.,

$$\Pi^P = \big\{\pi\in\Pi(\{1,\dots,n\}) \,:\, \pi^{-1}(i) \le \pi^{-1}(j),\ \forall\,(i,j)\in E\big\}.$$

In other words, if $(i,j)\in E$, then i always appears before j in the permutations in Π^P.

All the results in Theorem 7 and Corollary 3 then hold with the same modification for the set of permutations (see Tawarmalani et al. [2010] for details).

7.2.2 Summability of Concave Envelopes

For any two functions f, g defined on a polytope $P \subseteq \mathbb{R}^n$, it is always true that $\text{conc}_P(f+g) \leq \text{conc}_P(f) + \text{conc}_P(g)$, and equality holds if one of the two functions is affine [Tardella, 2008]. In practice, it is relevant to seek sufficient conditions on f, g and P that guarantee equality, since these would allow constructing the concave envelope of a (complex) sum of functions by characterizing the envelopes of individual components. The following result provides a general characterization of such conditions.

Theorem 8 (Theorem 3 in Tardella [2008])**.** *For a polytope $P \subset \mathbb{R}^n$, let f, g be two functions that are concave-extendable from* $\text{ext}(P)$*, and let $\mathcal{R}_{f,P} = \{F_i : i \in \mathcal{I}\}$, and $\mathcal{R}_{g,P} = \{G_j : j \in \mathcal{J}\}$ denote the polyhedral subdivisions of P that yield the linearity domains of* $\text{conc}_P(f)(\boldsymbol{w})$ *and* $\text{conc}_P(g)(\boldsymbol{w})$*, respectively. Then, the following conditions are equivalent:*

(i) $\text{conc}_P(f) + \text{conc}_P(g)$ *is concave-extendable from vertices*

(ii) $\text{conc}_P(f) + \text{conc}_P(g) = \text{conc}_P(f+g)$

(iii) $F_i \cap G_j$ *has all vertices in* $\text{ext}(P)$, $\forall\, i \in \mathcal{I}$, $\forall\, j \in \mathcal{J}$.

This theorem provides sufficient conditions for the concave envelope of a sum of two functions to be exactly given by the sum of the two separate concave envelopes. In view of the discussion in Section 7.2.1, the following corollary summarizes a particularly relevant class of functions that satisfy these requirements.

Corollary 4. *For any polytope $P \subseteq \mathcal{H}_n$ such that* $\text{ext}(P)$ *is a sublattice of $\{0,1\}^n$, and any finite collection of functions $h_i : P \to \mathbb{R}$, $i \in \mathcal{I}$, that are convex and supermodular on* $\text{ext}(P)$*,*

$$\text{conc}_P\left(\sum_{i\in\mathcal{I}} h_i\right) = \sum_{i\in\mathcal{I}} \text{conc}_P(h_i).$$

Proof. A sum of convex and supermodular functions is also convex and supermodular. By Theorem 7, the concave envelope of any convex and supermodular function is given by the Lovász extension, which is an affine interpolation of the function on the simplicies Δ_π in the Kuhn triangulation. Applying this result for each h_i and for the sum immediately yields the result. ☐

7.3 Technical Results

Lemma (Lemma 2)**.** *Suppose $f^\star : P \to \mathbb{R}$ is convex on P and supermodular on* $\text{ext}(P)$*. Consider an arbitrary $\hat{\boldsymbol{w}} \in \text{ext}(P) \cap \arg\max_{\boldsymbol{w}\in P} f^\star(\boldsymbol{w})$, and let $\boldsymbol{g}^\pi$ be given by (10). Then,*

1. *For any $\boldsymbol{w} \in P$, we have*

$$f^\star(\boldsymbol{w}) \leq f^\star(\hat{\boldsymbol{w}}) + (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^\pi,\ \forall \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}.$$

2. *There exists a set of convex weights $\{\lambda_\pi\}_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}}$ such that $\boldsymbol{g} = \sum_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \boldsymbol{g}^\pi$ satisfies*

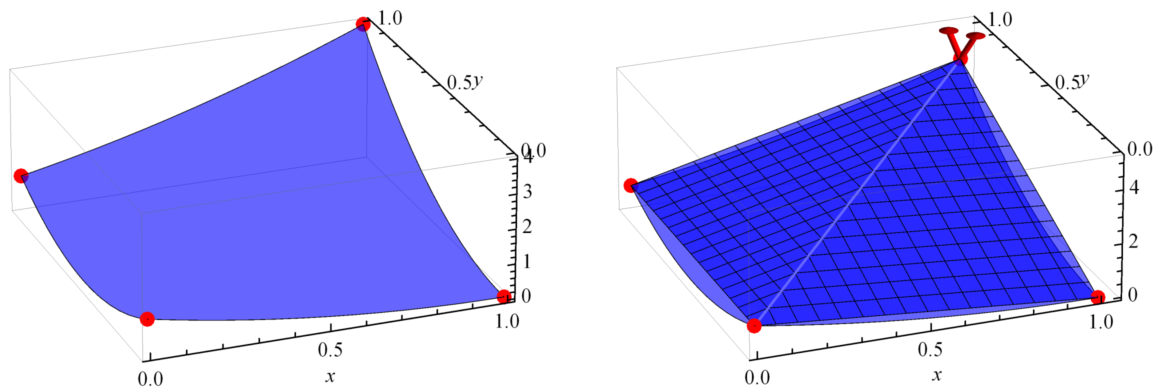
$$(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} \leq 0,\ \forall\, \boldsymbol{w} \in P.$$

Proof. Note first that the set $\text{ext}(P) \cap \arg\max_{\boldsymbol{w}\in P} f^\star(\boldsymbol{w})$ is nonempty, since $f^\star$ is convex. Therefore, since the vertices of $\text{ext}(P)$ are integral, $\hat{\boldsymbol{w}} = \mathbf{1}_S$ for some $S \subseteq \{1, \ldots, n\}$, and $\hat{\boldsymbol{w}}$ belongs to the intersection of all simplices Δ_π that correspond to permutations π in the set[6] $\mathscr{S}_{\hat{\boldsymbol{w}}} \stackrel{\text{def}}{=} \Pi^P(S) \times \Pi^P(S^C)$. Here, $\Pi^P(S)$ is any permutation of the elements in S that is consistent with the pre-order on P. For instance, if $\{i, j\} \subseteq S$ for some $(i, j) \in E$, then $\Pi^P(S)$ contains only permutations of S such that i appears before j.

[1] To argue the first claim, note that (by (9) and (10) in Lemma 1), the set $\{\boldsymbol{g}^\pi : \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}\}$ contains valid supergradients of the concave function $\text{conc}_{\mathcal{W}}(f^\star)$ at $\hat{\boldsymbol{w}}$. As such, the supergradient inequality applied to the concave function $\text{conc}_{\mathcal{W}}(f^\star)$ at $\hat{\boldsymbol{w}}$ yields

$$\text{conc}_{\mathcal{W}}(f^\star)(\boldsymbol{w}) \le \text{conc}_{\mathcal{W}}(f^\star)(\hat{\boldsymbol{w}}) + (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^\pi, \ \forall\, \boldsymbol{w} \in P, \ \forall\, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}.$$

The desired inequality follows since $f^\star(\hat{\boldsymbol{w}}) = \text{conc}_{\mathcal{W}}(f^\star)(\hat{\boldsymbol{w}})$, and $f^\star(\boldsymbol{w}) \le \text{conc}_{\mathcal{W}}(f^\star)(\boldsymbol{w})$, $\forall\, \boldsymbol{w} \in P$. For an example illustrating the relation, please refer to Figure 4.



(a) $f : \mathcal{H}_2 \to \mathbb{R}$.

(b) f, $\text{conc}_{\mathcal{H}_2}(f)$ and $\boldsymbol{g}^\pi$ at $\hat{\boldsymbol{w}} = (1, 1)$.

Figure 4: (a) Example of a convex and supermodular function $f : \mathcal{H}_2 \to \mathbb{R}$. (b) Its concave envelope and the supergradients $\boldsymbol{g}^{(1,2)}$ and $\boldsymbol{g}^{(2,1)}$ at the vertex $\hat{\boldsymbol{w}} = (1, 1)$.

[2] The intuition behind the second claim is geometric. We essentially seek to show that, at any vertex $\hat{\boldsymbol{w}}$ maximizing $f^\star$, there exists a supergradient of $f^\star$ (obtained as a convex combination of the supergradients corresponding to tight non-vertical facets of the concave envelope) that is a direction of decrease. In the example of Figure 4, this means that there is a convex combination $\boldsymbol{g}$ of $\boldsymbol{g}^{(1,2)}$ and $\boldsymbol{g}^{(2,1)}$ at $\hat{\boldsymbol{w}} = (1, 1)$, such that $\boldsymbol{g} \ge 0$.

In order to construct the candidate vector $\boldsymbol{g}$, let us first consider the problem of maximizing $f^\star$ on P. By Corollary 2 in the Appendix (and also from Lemma 1), we have

$$\begin{aligned} \max_{\boldsymbol{w}\in P} f^\star(\boldsymbol{w}) = \max_{t,\,\boldsymbol{w}} \ & t \\ \text{s.t. } & t \le (\boldsymbol{g}^\pi)^T \boldsymbol{w} + g_0, \ \forall\, (\boldsymbol{g}^\pi, g_0) \in \text{ext}(\mathcal{D}_{f^\star,P}). \quad (*) \\ & \boldsymbol{w} \in P. \end{aligned}$$

If we denote the optimal value by $J^\star$, then $t = J^\star$ and $\boldsymbol{w} = \hat{\boldsymbol{w}}$ are optimal in the program on the right. Furthermore, the only constraints $(*)$ that are tight at $\hat{\boldsymbol{w}}$ are those corresponding to

[6]In other words, any such permutation π has in the first $|S|$ positions the elements $\{i : w_i = 1\}$, and in the remaining $|S^C|$ the elements $\{i : w_i = 0\}$.

$\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}$. As such, by adding and subtracting terms $(\boldsymbol{g}^\pi)^T \hat{\boldsymbol{w}}$, we have that the left program (in the following primal-dual pair) is equivalent to the problem above:

$$
\begin{aligned}
& \max_{t,\,\boldsymbol{w}} \quad t \\
\lambda_\pi \ &\rightarrow \ t \le J^\star + (\boldsymbol{w} - \hat{\boldsymbol{w}})'\,\boldsymbol{g}^\pi,\ \forall\, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}} \\
\boldsymbol{\eta} \ &\rightarrow \ \boldsymbol{w} \le \mathbf{1} \\
& \qquad \boldsymbol{w} \ge 0 \\
\mu_{i,j} \ &\rightarrow \ w_j - w_i \le 0
\end{aligned}
\quad = \quad
\begin{aligned}
\min_{\lambda_\pi, \boldsymbol{\eta}, \mu_{i,j}} \ & \mathbf{1}'\boldsymbol{\eta} + \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \left(J^\star - \hat{\boldsymbol{w}}'\boldsymbol{g}^\pi\right) \\
& -\sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \boldsymbol{g}^\pi + \boldsymbol{\eta} + \tilde{\boldsymbol{\mu}} \ge 0 \\
& \sum_{\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi = 1 \\
& \lambda_\pi, \boldsymbol{\eta}, \mu_{i,j} \ge 0
\end{aligned}
\tag{27}
$$

where $\tilde{\mu}_i \stackrel{\text{def}}{=} -\sum_{j\,:(i,j)\in E} \mu_{i,j} + \sum_{j\,:(j,i)\in E} \mu_{j,i}$. The primal and dual programs above have an optimal value $J^\star$, and, in any dual optimal solution, $\boldsymbol{\eta}^\star = \sum_\pi \lambda_\pi^\star \boldsymbol{g}^\pi - \tilde{\boldsymbol{\mu}}^\star$. Furthermore, by complementary slackness, there exists an optimal dual solution (corresponding to the primal optimal solution $J^\star$, $\hat{\boldsymbol{w}}$) satisfying $\eta_i^\star = 0,\ \forall i \in S^C$. This implies that $\boldsymbol{\eta}^\star$ satisfies $\boldsymbol{\eta}_{S^C}^\star = 0$, $\boldsymbol{\eta}_S^\star \ge 0$.

The candidate vector $\boldsymbol{g}$ we would like to consider is exactly $\boldsymbol{g} = \sum_\pi \lambda_\pi^\star \boldsymbol{g}^\pi = \boldsymbol{\eta}^\star + \tilde{\boldsymbol{\mu}}^\star$. To complete part **[2]**, we need to check that

$$
(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} \le 0,\ \forall\, \boldsymbol{w} \in P.
$$

However, note that $\boldsymbol{w} - \hat{\boldsymbol{w}}$ can be written (for any $\boldsymbol{w} \in P$) as a conic combination of the vectors $\boldsymbol{w}_a - \hat{\boldsymbol{w}}$, where $\boldsymbol{w}_a$ are vertices of P adjacent[7] to $\hat{\boldsymbol{w}}$. Therefore, the required condition holds at an arbitrary $\boldsymbol{w}$ if and only if it holds at all vertices of P adjacent to $\hat{\boldsymbol{w}}$.

To characterize the latter set, denoted by $\mathcal{A}(\hat{\boldsymbol{w}})$, we introduce the following sets of nodes:

$$
\mathcal{D}(T) \stackrel{\text{def}}{=} \{k \in \{1,\dots,n\} \,:\, \exists\, i \in T \text{ and a directed path in } G \text{ from } i \text{ to } k\} \tag{28a}
$$

$$
\mathcal{U}(T) \stackrel{\text{def}}{=} \{k \in \{1,\dots,n\} \,:\, \exists\, i \in T \text{ and a directed path in } G \text{ from } k \text{ to } i\}. \tag{28b}
$$

$\mathcal{D}(T)$ contains all the nodes "in the downstream" of nodes $i \in T$ (by definition, we automatically include in $\mathcal{D}(T)$ the set T itself). In particular, in any feasible $\boldsymbol{w} \in P$, we have $w_k \le w_i,\ \forall\, k \in \mathcal{D}(T),\ \forall\, i \in T$. Similarly, $\mathcal{U}(T)$ has all the nodes "in the upstream" of nodes $i \in T$, and any feasible $\boldsymbol{w} \in P$ satisfies $w_k \ge w_i,\ \forall\, k \in \mathcal{U}(T),\ \forall\, i \in T$. For an example, please refer to Figure 5.

With $\mathcal{D}(T)$ and $\mathcal{U}(T)$ as above, Lemma 8 in the Appendix provides the following inclusion relation for $\mathcal{A}(\hat{\boldsymbol{w}})$:

$$
\mathcal{A}(\hat{\boldsymbol{w}}) \subseteq \left\{\hat{\boldsymbol{w}} - \mathbf{1}_{\mathcal{D}(T)\cap S} \,:\, T \subseteq S\right\} \cup \left\{\hat{\boldsymbol{w}} + \mathbf{1}_{\mathcal{U}(T)\cap S^C} \,:\, T \subseteq S^C\right\}.
$$

To argue that $(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} \le 0,\ \forall\, \boldsymbol{w} \in \mathcal{A}(\hat{\boldsymbol{w}})$, it suffices to check the relation for the larger set on the right. Consider the following separate cases.

[C1] $\boldsymbol{w} - \hat{\boldsymbol{w}} = -\mathbf{1}_{\mathcal{D}(T)\cap S}$, for some $T \subseteq S$. Then,

$$
\begin{aligned}
(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} &= (\boldsymbol{w} - \hat{\boldsymbol{w}})^T (\boldsymbol{\eta}^\star + \tilde{\boldsymbol{\mu}}^\star) \\
&= -\sum_{i \in \mathcal{D}(T)\cap S} \eta_i^\star + \sum_{i \in \mathcal{D}(T)\cap S} \Big(\sum_{j\,:(i,j)\in E} \mu_{i,j}^\star - \sum_{j\,:(j,i)\in E} \mu_{j,i}^\star \Big)
\end{aligned}
$$

[7]The notion of adjacency used here is well established in polyhedral theory - two vertices of a polytope are said to be adjacent if there is an edge (i.e., a face of dimension 1) connecting them. We refer the interested reader to Schrijver [2000] for definitions and details.

Consider an arbitrary node i in the summation above. For any $j \in S^C$ such that $(i,j) \in E$, we must have $\mu^\star_{i,j} = 0$, by complementary slackness. For any $j \in S$ such that $(i,j) \in E$, we must have $j \in \mathcal{D}(T) \cap S$. Therefore, the dual variable $\mu^\star_{i,j}$ appears in the expression above twice, once with a "+" sign (for the edge (i,j) going out of node i), and once with a "-" sign (for the edge (i,j) going into node j). Since the two terms cancel out, the final expression above contains only terms in $\eta^\star_i$ or $\mu^\star_{i,j}$ with negative signs, hence must be non-positive. To better understand the relation, please refer to Figure 5 for an example.

[C2] $\boldsymbol{w} - \hat{\boldsymbol{w}} = \mathbf{1}_{\mathcal{U}(T)\cap S^C}$, for some $T \subseteq S^C$. Then, the complementary slackness conditions at $\hat{\boldsymbol{w}}$ imply that $\eta_i = 0$, $\forall\, i \in \mathcal{U}(T) \cap S^C$. We have

$$\begin{aligned}(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} &= (\boldsymbol{w} - \hat{\boldsymbol{w}})^T (\boldsymbol{\eta}^\star + \tilde{\boldsymbol{\mu}}^\star) \\ &= \sum_{i \in \mathcal{U}(T)\cap S^C} \Big(- \sum_{j\,:(i,j)\in E} \mu^\star_{i,j} + \sum_{j\,:(j,i)\in E} \mu^\star_{j,i} \Big).\end{aligned}$$

By a similar argument as above, consider an arbitrary i in the summation. For any $j \in S$ such that $(j,i) \in E$, we must have $\mu^\star_{j,i} = 0$. For any $j \in S^C$ such that $(j,i) \in E$, we must have $j \in \mathcal{U}(T) \cap S^C$. Therefore, $\mu^\star_{i,j}$ again appears twice, once with a "+" sign (for the edge (j,i) going into i), and once with a "-" sign (for the edge (j,i) going out of j). Since the two terms cancel out, the final expression again contains only terms with negative signs.

$\square$

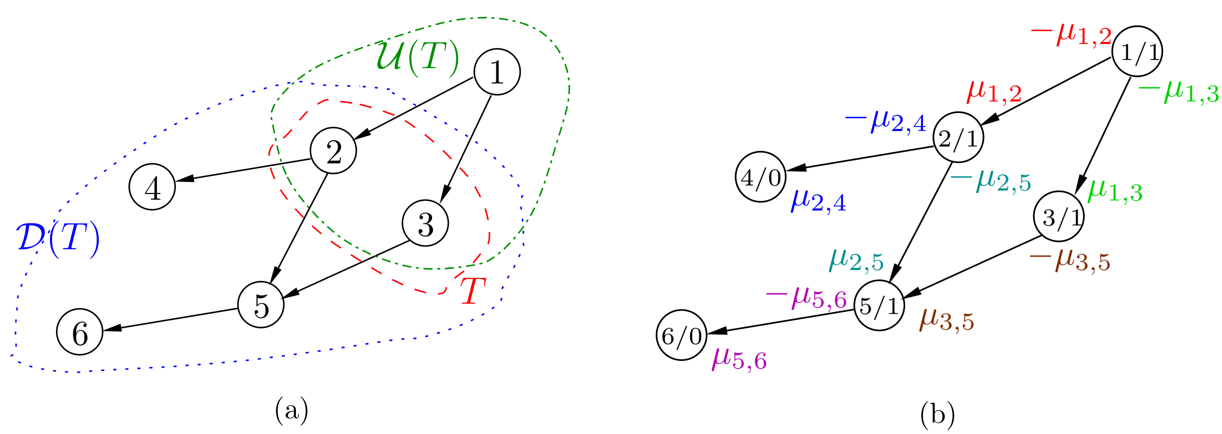


Figure 5: Example of a preorder graph with downstream and upstream nodes (a), and a vertex in the corresponding uncertainty set $\mathcal{W}$, with the dual variables (b). Here, $G = (V,E)$, where $V = \{1,\dots,5\}$ and $E = \{(1,2),(1,3),(2,4),(2,5),(3,5),(5,6)\}$. In (a), $T = \{2,3\}$, so that $\mathcal{D}(T) = \{2,3,4,5,6\}$, and $\mathcal{U}(T) = \{1,2,3\}$. In (b), the relevant vertex in $\mathcal{W}$ is $\hat{\boldsymbol{w}} = (1,1,1,0,1,0)$, so that $S_y = \{1,2,3,5\}$, and $\mu_{24} = \mu_{56} = 0$.

The following lemma provides a (partial) characterization for the set of adjacent points in a sublattice polytope of the form (3).

Lemma 8. *Consider a polytope $P = \{\boldsymbol{w} \in \mathcal{H}_n : x_i \geq x_j,\ \forall\,(i,j) \in E\}$, where $E \subseteq \{1,\dots,n\}^2$ is any set of directed edges. Let $\boldsymbol{y} \equiv \mathbf{1}_{S_y}$ denote any vertex of P, where $S_y \subseteq \{1,\dots,n\}$. Then, all* *the vertices of P adjacent to $\boldsymbol{y}$ are contained in the set*

$$\{\boldsymbol{y} - \mathbf{1}_{\mathcal{D}(T)\cap S_y} \,:\, T \subseteq S_y\} \cup \{\boldsymbol{y} + \mathbf{1}_{\mathcal{U}(T)\cap S_y^c} \,:\, T \subseteq S_y^c\}, \tag{29}$$

where $\mathcal{D}(T)$ and $\mathcal{U}(T)$ are given by:

$$\begin{aligned}
\mathcal{D}(T) &\stackrel{\text{def}}{=} \{k \in \{1,\dots,n\} \,:\, \exists\, i \in T \text{ and a directed path in } G \text{ from } i \text{ to } k\} \\
\mathcal{U}(T) &\stackrel{\text{def}}{=} \{k \in \{1,\dots,n\} \,:\, \exists\, i \in T \text{ and a directed path in } G \text{ from } k \text{ to } i\}.
\end{aligned}$$

Proof. Consider any vertex $\boldsymbol{x}$ adjacent to $\boldsymbol{y}$, and let $\boldsymbol{x} = \mathbf{1}_{S_x}$ for some $S_x \subseteq \{1,\dots,n\}$. We claim that $\boldsymbol{x} \le \boldsymbol{y}$ or $\boldsymbol{x} \ge \boldsymbol{y}$. Otherwise, since $\mathbf{1}_{S_x\cup S_y}$ and $\mathbf{1}_{S_x\cap S_y}$ are also valid vertices of P satisfying $\mathbf{1}_{S_x\cup S_y} + \mathbf{1}_{S_x\cap S_y} = \mathbf{1}_{S_x} + \mathbf{1}_{S_y}$, we would obtain two distinct convex representations for $(\boldsymbol{x}+\boldsymbol{y})/2$ in terms of vertices of P, which can never be the case if $\boldsymbol{x}$ and $\boldsymbol{y}$ are adjacent (see, e.g., Lemma 1 in Gurgel and Wakabayashi [1997]).

To complete the proof, we claim that

$$\{\boldsymbol{y} - \mathbf{1}_{\mathcal{D}(T)\cap S_y},\, T \subseteq S_y\} = \{\boldsymbol{x} \in \operatorname{ext}(P) \,:\, \boldsymbol{x} \le \boldsymbol{y}\}, \tag{30a}$$

$$\{\boldsymbol{y} + \mathbf{1}_{\mathcal{U}(T)\cap S_y^c},\, T \subseteq S_y^c\} = \{\boldsymbol{x} \in \operatorname{ext}(P) \,:\, \boldsymbol{x} \ge \boldsymbol{y}\}. \tag{30b}$$

We argue (30a) by double inclusion, and (30b) follows by an analogous argument.

Note that "$\subseteq$" follows trivially, since all the points in the set on the left of (30a) are valid extreme points of P and are $\le \boldsymbol{y}$.

To argue "$\supseteq$", consider any $\boldsymbol{x} \le \boldsymbol{y}$ and note that $\boldsymbol{x} = \boldsymbol{y} - \mathbf{1}_{S_y\setminus S_x}$. By definition, $S_y \setminus S_x \subseteq \mathcal{D}(S_y\setminus S_x)\cap S_y$, and we claim the reverse inclusion also holds. To this end, note that S_x cannot contain any elements in $\mathcal{D}(S_y\setminus S_x)$, since the components corresponding to the latter indices are always set to zero when the components corresponding to $S_y \setminus S_x$ are set to zero. Therefore, $\mathcal{D}(S_y\setminus S_x)\cap S_y = S_y\setminus S_x$, so that $\boldsymbol{x} = \boldsymbol{y} - \mathbf{1}_{\mathcal{D}(S_y\setminus S_x)\cap S_y}$, which completes the reverse inclusion. □

We note that the set of adjacent vertices in a binary sublattice does not seem to have a trivial characterization. In particular, it is easy to construct examples showing that the inclusion of the former set in the set in (29) can be strict. For instance, when $P = \mathcal{H}_n$, and $\boldsymbol{y} = \mathbf{1}$, the set in (29) is actually $\operatorname{ext}(P)$, i.e., *all* the extreme points of P.

A natural conjecture would be that the former set can be reached by changing *a single* coordinate i at a time, together with all the relevant corresponding coordinates in $\mathcal{U}(\{i\})$ or $\mathcal{D}(\{i\})$ (in other words, that we can restrict (29) to sets T with $|T| = 1$). Unfortunately, this characterization turns out to be incomplete. To see this, consider the simple example in Figure 1, where $P = \{\boldsymbol{x} \in \mathbb{R}^3 \,:\, x_1 \ge x_2,\, x_1 \ge x_3\}$. Here, vertex $(0,0,0)$ is adjacent to *all* the vertices of P - in particular, $(1,1,1)$ - which cannot be reached by changing only one coordinate at a time.

The next result is a complete proof of the main theorem of Section 4 in the paper.

Theorem (Theorem 3)**.** *Consider an optimization problem of the form*

$$\max_{\boldsymbol{w}\in P}\Big[\boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} h_i(\boldsymbol{w})\Big], \tag{31}$$

where $P \subset \mathbb{R}^k$ is any polytope, $\boldsymbol{a} \in \mathbb{R}^n$ is an arbitrary vector, $\mathcal{I}$ is a finite index set, and $h_i : \mathbb{R}^n \to \mathbb{R}$ are functions satisfying the following properties

[P1] *h_i are concave extendable from* $\operatorname{ext}(P)$, $\forall\, i \in \mathcal{I}$,

[P2] $\operatorname{conc}_P(h_i + h_j) = \operatorname{conc}_P(h_i) + \operatorname{conc}_P(h_j)$, *for any* $i \neq j \in \mathcal{I}$.

Then there exists a set of affine functions $z_i(\boldsymbol{w})$, $i \in \mathcal{I}$, *satisfying* $z_i(\boldsymbol{w}) \geq h_i(\boldsymbol{w})$, $\forall\, \boldsymbol{w} \in P$, $\forall\, i \in \mathcal{I}$, *such that*

$$\max_{\boldsymbol{w}\in P}\Big[\,\boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} z_i(\boldsymbol{w})\,\Big] \;=\max_{\boldsymbol{w}\in P}\Big[\,\boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} h_i(\boldsymbol{w})\,\Big]. \tag{32}$$

Proof. We prove the result for a case with $|\mathcal{I}| = 2$. The general result follows by induction on $|\mathcal{I}|$, and by noting that properties **[P1]** and **[P2]** are preserved under addition of functions. Furthermore, to avoid technicalities, we consider the case when the optimal value in (31), denoted by $J^\star$, is finite[8].

When $|\mathcal{I}| = 2$, note that the affine function $z(\boldsymbol{w}) = J^\star - \boldsymbol{a}^T\boldsymbol{w}$ trivially satisfies the constraints

$$\begin{aligned} z(\boldsymbol{w}) &\geq h_1(\boldsymbol{w}) + h_2(\boldsymbol{w}),\ \forall\, \boldsymbol{w} \in P,\\ J^\star &= \max_{\boldsymbol{w}\in P}\big[\,\boldsymbol{a}^T\boldsymbol{w} + z(\boldsymbol{w})\big]. \end{aligned}$$

Therefore, to prove our claim, it suffices to find two affine functions $z_{1,2}(\boldsymbol{w})$, satisfying

$$\begin{aligned} &z_1(\boldsymbol{w}) + z_2(\boldsymbol{w}) = z(\boldsymbol{w})\\ &\qquad\qquad z_i(\boldsymbol{w}) \geq h_i(\boldsymbol{w}),\ \forall\, \boldsymbol{w} \in P,\ \forall\, i \in \mathcal{I}. \end{aligned}$$

With $z_2 = z - z_1$, this is equivalent to finding a single affine function z_1 satisfying

$$h_1(\boldsymbol{w}) \leq z_1(\boldsymbol{w}) \leq z(\boldsymbol{w}) - h_2(\boldsymbol{w}),\ \forall\, \boldsymbol{w} \in P. \tag{33}$$

To this end, let us consider the functions $f \overset{\text{def}}{=} h_1$ and $g \overset{\text{def}}{=} z - h_2$. By Property **[P1]**, since z is affine, both f and $-g$ are concave-extendable from $\operatorname{ext}(P)$ (see Section 7.2 of the Appendix or Proposition 2 in Tardella [2008]). Also, $f \leq g$ on P. We claim that

$$\operatorname{conc}_P(f)(\boldsymbol{w}) \leq \operatorname{conv}_P(g)(\boldsymbol{w}),\ \forall\, \boldsymbol{w} \in P. \tag{34}$$

To see this, consider the function $f - g$. Since both f and $-g$ are concave-extendable from $\operatorname{ext}(P)$, so is $f - g$. By Property **[P2]**, we also have that $\operatorname{conc}_P(f-g) = \operatorname{conc}_P(f) + \operatorname{conc}_P(-g) = \operatorname{conc}_P(f) - \operatorname{conv}_P(g)$. Therefore,

$$\begin{aligned} \max_{\boldsymbol{w}\in P}\big[\operatorname{conc}_P(f)(\boldsymbol{w}) - \operatorname{conv}_P(g)(\boldsymbol{w})\big] &= \max_{\boldsymbol{w}\in P} \operatorname{conc}_P(f-g)(\boldsymbol{w})\\ &= \max_{\boldsymbol{w}\in P}(f-g)(\boldsymbol{w})\\ &\leq 0. \end{aligned}$$

If the maximum is actually 0 in the above expression, then $\operatorname{conc}_P(f) = \operatorname{conv}_P(g)$, so that both are affine functions on P, and $z_1 = \operatorname{conc}_P(f) = \operatorname{conv}_P(g)$ would satisfy the requirement in (33). Therefore, we assume throughout that there exists $\boldsymbol{w} \in P\ :\ \operatorname{conc}_P(f)(w) < \operatorname{conv}_P(g)(\boldsymbol{w})$. We can now introduce the following two sets:

$$\begin{aligned} \mathcal{H}_f &\equiv \operatorname{hypo}(\operatorname{conc}_P(f)) \overset{\text{def}}{=} \big\{(\boldsymbol{w}, t) \in \mathbb{R}^{k+1}\ :\ \boldsymbol{w} \in P,\ t \leq f^+(\boldsymbol{w})\big\} && \text{(hypograph of } f^+\text{)}, && \text{(35a)}\\ \mathcal{E}_g &\equiv \operatorname{epi}(\operatorname{conv}_P(g)) \overset{\text{def}}{=} \big\{(\boldsymbol{w}, t) \in \mathbb{R}^{k+1}\ :\ \boldsymbol{w} \in P,\ t \geq g^-(\boldsymbol{w})\big\} && \text{(epigraph of } g^-\text{)}. && \text{(35b)} \end{aligned}$$

[8] The arguments are extendable to a case when $J^\star = +\infty$, by allowing extended-real convex functions h_i to be used.

Note that $\mathcal{H}_f$ ($\mathcal{E}_g$) is a convex, closed set, since it the hypograph (epigraph) of a proper, closed, concave (convex) function. Furthermore, both $\mathcal{H}_f$ and $\mathcal{E}_g$ are *polyhedral* sets, since the concave envelopes of concave-extendable functions are polyhedral (see Section 7.2.1 of the Appendix). With $\mathrm{ri}(K)$ denoting the relative interior of a convex set K, we claim that[9]

$$\mathcal{H}_f \cap \mathrm{ri}(\mathcal{E}_g) = \emptyset.$$

This follows because any $(\boldsymbol{w}, t) \in \mathcal{H}_f \cap \mathrm{ri}(\mathcal{E}_g)$ would satisfy $\mathrm{conv}_P\,(g)\,(\boldsymbol{w}) < t \leq \mathrm{conc}_P\,(f)\,(\boldsymbol{w})$, in direct contradiction with (34). Therefore, we have two polyhedral sets, $\mathcal{H}_f$ and $\mathcal{E}_g$, such that $\mathcal{H}_f \cap \mathrm{ri}(\mathcal{E}_g) = \emptyset$. By Theorem 20.2 in Rockafellar [1970], there exists a hyperplane separating $\mathcal{H}_f$ and $\mathcal{E}_g$ *properly* (i.e., not both sets belonging to the hyperplane). In particular, there exist $\boldsymbol{z}_1 \in \mathbb{R}^k$, $z_{1,0}$, $\beta \in \mathbb{R}$ such that $(\boldsymbol{z}_1, z_{1,0}) \neq \mathbf{0}$, and

$$\forall\, \boldsymbol{w} \in P,\, \boldsymbol{z}_1'\boldsymbol{w} + z_{1,0}\, f(\boldsymbol{w}) \leq \beta \leq \boldsymbol{z}_1'\boldsymbol{w} + z_{1,0}\, g(\boldsymbol{w}),$$

By proper separability, $z_{1,0} \neq 0$. If $z_{1,0} < 0$, then $f \geq g$ on P, which would contradict our standing assumption that $\exists\, \boldsymbol{w} \in P\, :\, \mathrm{conc}_P\,(f)\,(\boldsymbol{w}) < \mathrm{conv}_P\,(g)\,(\boldsymbol{w})$. Therefore, we are left with $z_{1,0} > 0$, which implies that $z_1(\boldsymbol{w}) \stackrel{\text{def}}{=} (\beta - \boldsymbol{z}_1'\boldsymbol{w})/z_{1,0}$ checks equation (33), and hence completes the construction and the proof. $\square$

An example outlining the role of requirement **[P2]** is presented in Figure 6.

References

A. Bemporad, F. Borrelli, and M. Morari. Min-Max Control of Constrained Uncertain Discrete-Time Linear Systems. *IEEE Transactions on Automatic Control*, 48(9):1600–1606, 2003. ISSN 0018-9286.

A. Ben-Tal and A. Nemirovski. Robust Solutions of Uncertain Linear Programs. *Operations Research Letters*, 25:1–13, 1999.

A. Ben-Tal and A. Nemirovski. Robust optimization - methodology and applications. *Math. Program.*, 92(3):453–480, 2002. ISSN 0025-5610. doi: http://dx.doi.org/10.1007/s101070100286.

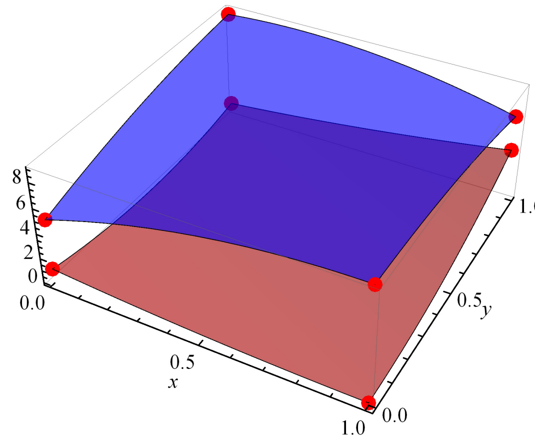
A. Ben-Tal, A. Nemirovski, and C. Roos. Robust Solutions of Uncertain Quadratic and Conic-Quadratic Problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002. ISSN 1052-6234. doi: http://dx.doi.org/10.1137/S1052623401392354.

A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Math. Program.*, 99(2):351–376, 2004. ISSN 0025-5610. doi: http://dx.doi.org/10.1007/s10107-003-0454-y.

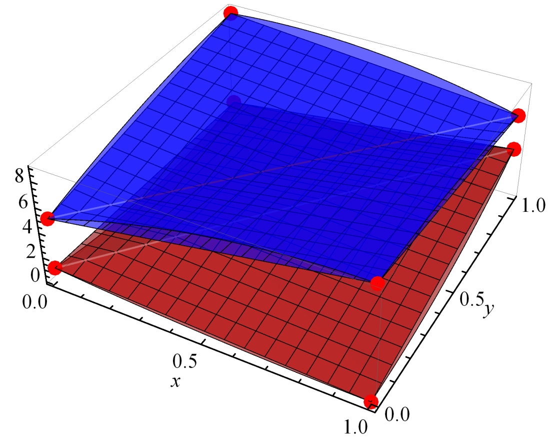
A. Ben-Tal, S. Boyd, and A. Nemirovski. Control of Uncertainty-Affected Discrete Time Linear Systems via Convex Programming. Working paper, 2005a.

A. Ben-Tal, B. Golany, A. Nemirovski, and J.-P. Vial. Retailer-Supplier Flexible Commitments Contracts: A Robust Optimization Approach. *Manufacturing & Service Operations Management*, 7(3):248–271, 2005b. ISSN 1526-5498. doi: http://dx.doi.org/10.1287/msom.1050.0081.

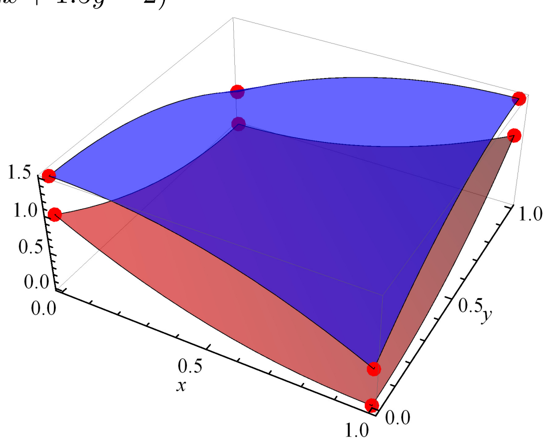
[9] We can equivalently prove that $\mathcal{E}_g \cap \mathrm{ri}(\mathcal{H}_f) = \emptyset$.



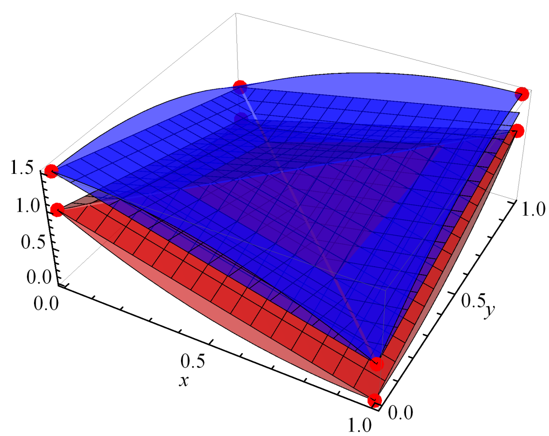
(a) $f(x,y) \stackrel{\text{def}}{=} (x+2y-1)^2 \le g(x,y) \stackrel{\text{def}}{=} 9 - (2x+1.5y-2)^2$



(b) f and $\text{conc}_P(f)$; g and $\text{conv}_P(g)$



(c) $f(x,y) \stackrel{\text{def}}{=} (x+y-1)^2 \le g(x,y) \stackrel{\text{def}}{=} 1.5 - (x-y)^2$



(d) f and $\text{conc}_P(f)$; g and $\text{conv}_P(g)$

Figure 6: The role of requirement **[P2]**. In both (a) and (c), $h_{1,2}$ are convex (cvx). In (a), $h_{1,2}$ are also supermodular (spm), so that f and $-g$ are cvx, spm, $\text{conc}_P(f-g) = \text{conc}_P(f) - \text{conc}_P(g)$, and $\text{conc}_P(f) \le \text{conv}_P(g)$ in (b). In (c), h_1 is spm, but h_2 is not, so that $-g$ is not spm, and $\text{conc}_P(f-g) \le \text{conc}_P(f) - \text{conc}_P(g)$. Note that $\text{conc}_P(f) \not\le \text{conv}_P(g)$ in (d).

A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending Scope of Robust Optimization: Comprehensive Robust Counterparts of Uncertain Problems. *Mathematical Programming*, 107(1):63–89, 2006. ISSN 0025-5610. doi: http://dx.doi.org/10.1007/s10107-005-0679-z.

Aharon Ben-Tal, Laurent El-Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

Alain Bensoussan, M Crourty, and J. M. Proth. *Mathematical Theory of Production Planning*. Elsevier Science Ltd., 1983.

D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA, 2001. ISBN 1886529272.

D. P. Bertsekas and I.B. Rhodes. On the minmax reachability of target tubes. *Automatica*, 7:233–247, 1971. doi: http://dx.doi.org/10.1007/s10107-005-0679-z.

D. P. Bertsekas and J. N. Tsitsiklis. *Neurodynamic Programming*. Athena Scientific, Belmont, MA, 1996.

Dimitris Bertsimas and Vineet Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming*, pages 1–41, 2010. ISSN 0025-5610. URL http://dx.doi.org/10.1007/s10107-011-0444-4. 10.1007/s10107-011-0444-4.

Dimitris Bertsimas and Melvyn Sim. Robust Discrete Optimization and Network Flows. *Mathematical Programming*, 98(1-3):49–71, 2003.

Dimitris Bertsimas and Melvyn Sim. The Price of Robustness. *Operations Research*, 52(1):35–53, 2004. URL http://or.journal.informs.org/cgi/content/abstract/52/1/35.

Dimitris Bertsimas, Dan A. Iancu, and Pablo A. Parrilo. Optimality of Affine Policies in Multistage Robust Optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010. doi: 10.1287/moor.1100.0444. URL http://mor.journal.informs.org/cgi/content/abstract/35/2/363.

Dimitris Bertsimas, Constantine Caramanis, and David B. Brown. Theory and Applications of Robust Optimization. *SIAM Review*, 53(3):464–501, 2011. Submitted for publication.

Robert L. Bray and Haim Mendelson. Deriving Supply Chain Metrics From Financial Statements. Working paper., 2012.

A. Charnes, W. W. Cooper, and G. H. Symonds. Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil. *Management Science*, 4(3):235–263, April 1958.

Li Chen and Hau L. Lee. Information sharing and order variability control under a generalized demand model. *Management Science*, 55(5):781–797, 2009. doi: 10.1287/mnsc.1080.0983.

Xin Chen, Melvyn Sim, Peng Sun, and Jiawei Zhang. A Linear Decision-Based Approximation Approach to Stochastic Programming. *Operations Research*, 56(2):344–357, 2008. doi: 10.1287/opre.1070.0457.

Patrick Cheridito, Freddy Delbaen, and Michael Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11:57–106, 2006.

Gustave Choquet. Theory of Capacities. *Annales de l'Institut Fourier*, 5:131–295, 1954.

A. J. Clark and H. Scarf. Optimal policies for a multi-echelon inventory problem. *Manage. Sci.*, 50(12 Supplement):475–490, July 1960. doi: http://dx.doi.org/10.1287/mnsc.1040.0265.

Daniela Pucci de Farias and Benjamin Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6):850–865, November-December 2003. doi: 10.1287/opre.51.6.850.24925.

G. E. Dullerud and F. Paganini. *A Course in Robust Control Theory*. Springer, 2005. ISBN 0387989455.

L. El-Ghaoui, F. Oustry, and H. Lebret. Robust Solutions to Uncertain Semidefinite Programs. *SIAM J. Optimization*, 9(1):33–52, 1998.

Larry G. Epstein and Martin Schneider. Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1 – 31, 2003. ISSN 0022-0531. doi: DOI:10.1016/S0022-0531(03)00097-8. URL http://www.sciencedirect.com/science/article/B6WJ3-48N2RT8-8/2/a76f81dfefcee4d3bddcb0247ea5d08e.

M.K.H. Fan, A. Tits, and J. Doyle. Robustness in the Presence of Mixed Parametric Uncertainty and Unmodeled Dynamics. *IEEE Transactions on Automatic Control*, 36(1):25–38, 1991. ISSN 0018-9286.

Hans Föllmer and Alexander Schied. *Stochastic Finance in Discrete Time*. de Gruyter Studies in Mathematics. Walter de Gruyter, Walter de Gruyter GmbH & Co., KG, 10785, Berlin, Germany, second edition edition, 2004.

Satoru Fujishige. *Submodular Functions and Optimization*, volume 58 of *Annals of Discrete Mathematics*. Elsevier, second edition, 2005.

Stanley J. Garstka and Roger J.-B. Wets. On Decision Rules in Stochastic Programming. *Mathematical Programming*, 7(1):117–143, 1974. ISSN 0025-5610. doi: 10.1007/BF01585511.

Stephen C. Graves, David B. Kletter, and William B. Hetzel. A dynamic model for requirements planning with application to supply chain optimization. *Operations Research*, 46(3-Supplement-3):S35–S49, May/June 1998. doi: 10.1287/opre.46.3.S35.

Martin Grötschel, Lázló Lovácz, and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, Berlin, New York, 1988.

Maria Angela C.M. Gurgel and Yoshiko Wakabayashi. Adjacency of vertices of the complete pre-order polytope. *Discrete Mathematics*, 175(1-3):163 – 172, 1997. ISSN 0012-365X. doi: 10.1016/S0012-365X(96)00143-4. URL http://www.sciencedirect.com/science/article/pii/S0012365X96001434.

Lars Peter Hansen and Thomas J. Sargent. Robust control and model uncertainty. *The American Economic Review*, 91(2):pp. 60–66, May 2001. URL http://www.jstor.org/stable/2677734.

Lars Peter Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press, 2008.

Warren H. Hausman. Sequential decision problems: A model to exploit existing forecasters. *Management Science*, 16(2):B93–B111, 1969. ISSN 00251909. URL http://www.jstor.org/stable/2628493.

David C. Heath and Peter L. Jackson. Modeling the Evolution of Demand Forecasts With Application to Safety Stock Analysis in Production/Distribution Systems. *IIE Transactions*, 26(3):17–30, 1994.

Daniel P. Heyman and Matthew J. Sobel. *Stochastic Models in Operations Research*. Dover Publications, Inc., 1984.

Woonghee Tim Huh and Ganesh Janakiraman. On the optimal policy structure in serial inventory systems with lost sales. *Operations Research*, 58(2):486–491, March/April 2010. doi: 10.1287/opre.1090.0716. URL http://or.journal.informs.org/content/58/2/486.abstract.

E.C. Kerrigan and J.M. Maciejowski. On robust optimization and the optimal control of constrained linear systems with bounded state disturbances. In *Proceedings of the 2003 European Control Conference*, Cambridge, UK, September 2003.

E.C. Kerrigan and J.M. Maciejowski. Properties of a New Parameterization for the Control of Constrained Systems with Disturbances. *Proceedings of the 2004 American Control Conference*, 5:4669–4674 vol.5, June-July 2004. ISSN 0743-1619. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01384049.

Daniel Kuhn, Wolfram Wiesemann, and Angelos Georghiou. Primal and Dual Linear Decision Rules in Stochastic and Robust Optimization. *Mathematical Programming*, pages 1–33, December 2009. doi: 10.1007/s10107-009-0331-4.

J. Löfberg. Approximations of closed-loop minimax MPC. *Proceedings of the 42nd IEEE Conference on Decision and Control*, 2:1438–1442 Vol.2, Dec. 2003. ISSN 0191-2216. doi: 10.1109/CDC.2003.1272813.

Lázló Lovász. Submodular functions and convexity. In *Mathematical Programming - The State of the Art*, pages 235–257. Springer, 1982.

Paul Milgrom and Chris Shannon. Monotone Comparative Statics. *Econometrica*, 62(1):pp. 157–180, 1994. ISSN 00129682. URL http://www.jstor.org/stable/2951479.

Arkadi Nemirovski and Alexander Shapiro. *Continuous Optimization: Current Trends and Applications*, volume 99 of *Applied Optimization*, chapter On Complexity of Stochastic Programming Problems, pages 111–146. Springer US, 2005. doi: 10.1007/b137941.

P. M. Pardalos and J. B. Rosen. Methods for global concave minimization: A bibliographic survey. *SIAM Review*, 28(3):pp. 367–379, 1986. ISSN 00361445. URL http://www.jstor.org/stable/2031354.

Evan L. Porteus. *Foundations of Stochastic Inventory Management*. Stanford Business Books, 2002.

Warren Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley Series in Probability and Statistics. Wiley-Interscience, 2007.

T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970. ISBN 0691080690.

H. Scarf. *The Optimality of (s, S) Policies in the Dynamic Inventory Problem.*, pages 196–202. Stanford University Press, Stanford, CA, USA, 1960.

David Schmeidler. Integral Representation Without Additivity. *Proceedings of the Americal Mathematical Society*, 97:255–261, 1986.

Alexander Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, second edition, 2000.

Alexander Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, 1 edition, 2003. ISBN 9783540443896.

David Simchi-Levi, Xin Chen, and Julien Bramel. *The Logic of Logistics: Theory, Algorithms, and Applications for Logistics and Supply Chain Management*. Springer Series in Operations Research and Financial Engineering. Springer, 2 edition, 2004.

J. Skaf and S. Boyd. Design of Affine Controllers Via Convex Optimization. *IEEE Transactions on Automatic Control*, 55(11):2476–2487, October 2010.

Kalyan T. Talluri and Garrett J. van Ryzin. *The Theory and Practice of Revenue Management*. Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA, paperback edition, 2005.

Fabio Tardella. Existence and sum decomposition of vertex polyhedral convex envelopes. *Optimization Letters*, 2:363–375, 2008. ISSN 1862-4472. URL http://dx.doi.org/10.1007/s11590-007-0065-2. 10.1007/s11590-007-0065-2.

Mohit Tawarmalani, Jean-Philippe P. Richard, and Cuanhui Xiong. Explicit convex and concave envelopes through polyhedral subdivisions. Submitted for publication, June 2010.

Donald M. Topkis. *Supermodularity and Complementarity*. Princeton University Press, 1998.

Jr. Veinott, Arthur F. The Status of Mathematical Inventory Theory. *Management Science*, 12(11):745–777, 1966. doi: 10.1287/mnsc.12.11.745. URL http://mansci.journal.informs.org/cgi/content/abstract/12/11/745.

Jr. Veinott, Arthur F. "lattice programming". Unpublished notes from lectures delivered at Johns Hopkins University, 1989.

K. Zhou and J. C. Doyle. *Essentials of Robust Control*. Prentice Hall, 1998. ISBN 0135258332.

P. Zipkin. *Foundations of Inventory Management*. McGraw Hill, 2000. ISBN 0256113793.

Paul Zipkin. On the structure of lost-sales inventory models. *Operations Research*, 56(4):937–944, July/August 2008. doi: 10.1287/opre.1070.0482. URL http://or.journal.informs.org/content/56/4/937.abstract.