Attributes*

Diego Klabjan, Wojciech Olszewski, and Asher Wolinsky

November 3, 2010

Abstract

An agent makes the decision whether to acquire an object. Before making this decision, she can discover, at some cost, some attributes of the object (or equivalently, obtain some signals about the object's value).

We aim to characterize the optimal set of attributes to be discovered; and if the agent is allowed to discover attributes sequentially, the ordering in which the attributes should be discovered.

We provide such a characterization in some special cases. For example, when the attributes are ordered by second-order stochastic dominance, and the costs decrease either sufficiently slow or sufficiently fast with this ordering. The optimal sets (in these special cases) can be computed by pseudo polynomial-time algorithms.

Under sequential discovering, we provide a complete solution to the case of two attributes distributed symmetrically around their means, and the outside option no higher than the sum of means.

1 Introduction

Many economic decisions have the following form: An agent considers acquiring an object, which is characterized by several attributes. The agent knows the distribution of each attribute in the population, but not the particular realization of attributes for the object at hand. Before making the decision whether to acquire the object, the agent can at some cost discover the realization of each attribute; the cost of discovering can vary across different attributes. Examples abound, ranging

*We would like to thank Hector Chade and Rakesh Vohra for helpful comments and suggestions.

from everyday goods to marriage and labor markets. For instance, a potential spouse is characterized by intelligence, beauty, personality, etc.

Recent studies in decision theory provide additional support to the attribute approach. These studies argue that decision makers often use sequentially several rationales to discriminate among the available alternatives. (See, for example, Manzini and Mariotti (2007); and Tversky (1972) for the basic ideas.)

We study two versions of our decision problem: In one version of the model, the agent must decide up front which attributes to discover, and then, knowing the realizations of these attributes, decide to accept or reject the object. In the case of acceptance, the agent obtains utility $u(x_1, ..., x_n)$, which is a function of the realizations of all attributes. In the case of rejection, the agent obtains an exogenous reservation utility V.

When attributes are ordered by second-order stochastic dominance, and the cost of discovering is the same for all attributes, the optimal set consists of most dominated attributes. When attributes are ordered in a slightly stronger sense (second-order stochastic dominance away from the mean) and the differences in the benefits from discovering single attributes are equal to the differences in the costs of discovering them, the optimal set consists of most dominant attributes. As a conclusion, we derive the following principle:

An agent who considers obtaining signals from several sources should be inclined to obtain less informative but cheaper signals compared to the agent who obtains a signal just from one source.

We characterize the optimal set also in some other special cases. The optimal sets in these special cases can be computed by a pseudo polynomial-time algorithms. We show, however, that polynomial-time algorithms may not exist. In addition, it is an open question whether the optimal set can be computed by a pseudo polynomial-time algorithms in more general cases.

In the other version of the model, the agent is allowed to discover attributes sequentially: she is allowed to stop discovering, and to accept or reject the object at any time. In this case, we characterize the optimal decision rule in the case of two attributes which are distributed independently, and symmetrically around their means, and the reservation utility equal to the sum of these means. The solution resembles Gittins' indices (see Gittins and Jones (1974) and Gittins (1989)) and Weitzman's (1979) Pandora rule. Each attribute is attached an index, determined by this attribute only, and independent of the other attribute, and the agent is prescribed to discover first the attribute with the higher index.

Weitzman (1979) studied a problem similar to the sequential version of our multi-attribute model:

Suppose an agent opens sequentially boxes containing a reward of unknown value. It is costly to open each box and discover the value of the reward contained in the box. And the agent can take only one reward. Weitzman shows that the optimal strategy (*Pandora rule*) assigns to each box a reservation price, independent of the other boxes, and prescribes to search next the remaining box with the highest reservation price. The agent terminates search when the highest reward from already sampled boxes exceeds the highest reservation price across remaining, closed boxes.

Weitzman's model is almost identical with our multi-attribute model for the utility function $u(x_1, ..., x_n) = \max\{x_1, ..., x_n\}$. The only difference comes from the fact in Weitzman's model, the agent must take the reward from an open box, whereas in the multi-attribute model, the agent's utility is affected also be the realizations of the attributes she decided not to discover. This feature makes the problem studied in the present paper more complicated. In particular, a minor adjustment of the proof of Proposition 5 yields the Weitzman result.

The celebrated literature on multi-armed bandits is also closely related. This literature investigates situations in which an agent chooses which arm to pull in each single period. The reward obtained in this way depends on the state of the pulled arm, which then transits to another state. The state of all the other arms remain unaltered. Gittins and Jones (1974) (see Whittle (1980) and Weber (1992) for simpler proofs) showed that one can attach an index to each arm, which depends only on the state of that arm, and that pulling an arm with the largest index at any point in time is an optimal strategy.

Chade and Smith (2005) study a problem similar to the simultaneous version of our multi-attribute model: An agent selects a number of ranked stochastic options. The inclusion of each option to the selected set is costly. Only one option may be exercised from those that succeed. A leading example is a student applying to many colleges. They show that the celebrated greedy algorithm finds an optimal set. Chade and Smith's model is the simultaneous version of Weitzman's model, and so is almost identical with our multi-attribute model for the utility function $u(x_1, ..., x_n) = \max\{x_1, ..., x_n\}$, where each x_i takes value 1 (success) with probability p_i, and value 0 (failure) with the remaining probability. Again, the problem studied in the present paper is more complicated. In particular, it cannot be solved by applying simple schemes such as the greedy algorithm.

The first study of attributes in a context similar to ours is Neeman (1995). Neeman is interested in an optimal strategy in the following stopping problem: An agent faces a sequence of i.i.d. multi-attribute products, and can observe only one attribute of each product. At each stage the agent has to decide whether to stop, taking the best product so far, or to continue by observing an attribute of the next product. One of Neeman's main results resembles our Proposition 1. Namely, he shows that if the distribution of some attribute second-order stochastically dominates the distributions of all other attributes, then the agent should always choose to observe this attribute.

2 Model

The problem we are studying in this paper can be formulated in the following two equivalent ways:

2.1 Attribute version of the model

An agent must decide whether to accept or reject an object that has been offered to her. The object has n attributes, represented by random variables $x_1, ..., x_n$, whose realizations are initially unknown to the agent. These random variables are distributed according to some cumulative density function (in abbreviation, *cdf*). The distribution is known to the agent. The agent's utility of having an object with attributes $x_1, ..., x_n$ is denoted by $u(x_1, ..., x_n)$, and the agent's reservation utility, i.e., the utility the agent obtains when she rejects the object, is denoted by V. Before making a decision, the agent can discover the realization of each attribute $x_1, ..., x_n$ of the object at the cost $c_1, ..., c_n$, respectively.

We consider two scenarios. Under one of them, the agent discovers the attributes simultaneously. That is, she decides which set of attributes $S \subset \{1, ..., n\}$ to discover. After learning the realizations x_i, $i \in S$ (at the cost $\sum_{i \in S} c_i$), she makes the decision whether to accept or reject the object.

Under the other scenario, the agent may discover the attributes sequentially. That is, she decides which attribute (if any) she wants to discover first, and contingent on the realizations of the attributes she has already discovered, she decides which of the remaining attributes (if any) to discover next. After the discovering of each attribute, the agent may stop the process of learning the realizations of attributes, in which case she may accept or reject the object. She may also decide to continue the process, and discover the realization of one of the remaining attributes. Every time she decides to discover an attribute, she pays the cost associated with this attribute.

Under sequential discovering, one might also study the setting in which the agent decides first about the ordering in which she will discover attributes. She is allowed to interrupt this process, and accept or reject the object without learning the realizations of the remaining attributes, but she is not allowed to change the ordering. That is, the agent is allowed to discover attributes sequentially, but the ordering of attributes may not depend on the already discovered realizations. The junior recruiting in economics closely resembles this form of extracting information. However, we will not study this setting in the present paper.

2.2 Signal version of the model

Suppose that instead of having multiple attributes, the value of the object is uncertain. The agent has a prior over the value of the object offered to her, and before making the decision to accept or reject it, she can consult with several sources of information. Each available source $i = 1, ..., n$ provides a signal x_i at a cost c_i. The agent knows the value-dependent distributions of signals, and therefore for any set of signals x_i, $i \in S \subset \{1, ..., n\}$, she can compute the expected value of the object. Again, the agent may consult with the sources of information simultaneously or sequentially.

This model is formally equivalent with the model in which objects have attributes. However, some assumptions, or payoff functions that are easy to interpret in one version of the model may be less reasonable in the other version. In this paper, we will focus on the attribute version of the problem.

2.3 Assumptions

Throughout the paper, we assume that $x_1, ..., x_n$ are distributed *independently* with cdfs $F_1, ..., F_n$, respectively,

$$u(x_1, ..., x_n) = x_1 + ... + x_n$$

and

$$V = \int x_1 dF_1 = ... = \int x_n dF_n = 0.$$

The results of this paper easily extend to the utilities of the form $u(x_1, ..., x_n) = \alpha_1 x_1 + ... + \alpha_n x_n$, where $\alpha_1, ..., \alpha_n$ are some positive real numbers; in this case, the assumptions regarding the distributions of $x_1, ..., x_n$ must be replaced with analogous assumptions regarding the distributions of $x_1/\alpha_1, ..., x_n/\alpha_n$. The research on related topics (e.g., Gittins' indices, see Gittins and Jones (1974), Gittins (1989), or Weitzman's (1979) Pandora rule) suggests that the analysis would not generalize, and would be rather intractable without the independence assumption regarding the distributions of attributes.

The assumption that the means of $x_1, ..., x_n$ are all equal is of course only a convenient normalization. However, the assumption that the reservation utility V is equal to these means is binding. Propositions 1, 3, and 4 hold true for an arbitrary value of V, but the assumption that $V = 0$ is essential for Propositions 2 and 5. Moreover, it may be interesting to endogenize the value of V, e.g., by assuming that the agent searches a sequence of objects, and decides whether to accept the current object and stop searching, or to reject the current object and search for another one. However, we will not explore the model with endogenous V in the present paper.

2.4 Auxiliary concepts

We write that $y \precsim^{s.o.} z$ when random variable y that is *second-order stochastically dominated* by random variable z. We will often assume that $x_1 \precsim^{s.o.} ... \precsim^{s.o.} x_n$, i.e., that attributes are ordered by second-order stochastic dominance.

We will sometimes make even a slightly stronger assumption. Consider two random variables y and z with the mean zero and with cdfs G and H, respectively. We will say that y is *second-order stochastically dominated* by z *away from the mean realizations* if

$$G(x) \leq H(x)$$

for all $x \geq 0$, and

$$H(x) \leq G(x)$$

for all $x \leq 0$.[1] We will then write that $z \precsim^{a.s.o.} y$. Intuitively, the idea is that variable y is obtained by moving the probability assigned by variable z from the center to the tails of the distribution. Notice that y is positive first-order stochastically dominated by z, then it is also second-ordered stochastically dominated. Proposition 6 in Appendix contains a characterization of second-order stochastic dominance away from the mean analogous to the characterization second-order stochastic dominance of in terms of mean-preserving spreads and the integrals of cdfs

Finally, we say that a random variable y with mean zero and the cdf G is *symmetric around the mean* (or briefly, symmetric) if for every $y \geq 0$,

$$G(-y) = 1 - G(y).$$

2.5 Optimization problem

Under simultaneous discovering of attributes, the agent who checked the values of attributes from set S will accept the object if $\sum_{i \in S} x_i > 0$, and reject the object if $\sum_{i \in S} x_i < 0$. The agent's tie-breaking rule when $\sum_{i \in S} x_i = 0$ will be inessential for the analysis. Therefore, the agent's objective

[1] Since we assume that the mean of all attributes is equal to zero, we define the auxiliary concepts also only for distributions with zero mean.

function is

$$U(S) = \int ... \int \max\left(0, \sum_{i \in S} x_i\right) d \prod_{i \in S} F_i(x_i) - \sum_{i \in S} c_i. \tag{1}$$

The question we are interested in is whether there exists a "nice" way of describing the solution of the agent's maximization problem:

$$\max_{S \subset \{1,...,n\}} U(S). \tag{2}$$

More rigorous versions of this question includes: Can the optimization problem (1) be solved in polynomial, or pseudo-polynomial time[2]? Or, can we find an optimal set S by computing $U(S)$ only at $P(n)$ sets S, where $P(n)$ is a polynomial function of the number of attributes n?

Under sequential discovering of attributes, we seek simple characterizations of optimal strategies, i.e. rules that tell the agent the realization of which attribute to discover (or to stop the process of learning, and to accept or reject the object), contingent on any history of already discovered realizations.

3 Simultaneous Discovering of Attributes

3.1 Some "polar" cases

We provide two propositions regarding some special cases in which are able to find an optimal set S, solving problem (2), by computing $U(S)$ at only n sets S.

Proposition 1. Suppose the attributes $x_1, ..., x_n$ are ordered by the second-order stochastic dominance, i.e. $x_1 \precsim^{s.o.} ... \precsim^{s.o.} x_n$. Suppose further that $c_1 \leq ... \leq c_n$. Then a set solving problem (2) has the form $S = \{1, ..., k\}$; that is, it consists of a k most dominated attributes.

This proposition generalizes the observation that dominated distributions provide more valuable signals about the object. Therefore, when the cost of discovering the realization is higher for dominant distributions, the agent should discover the most dominated attributes. Proposition 1 follows immediately from the following lemma:

Lemma 1. Let G_0, G_1 and G_2 be the cdfs of random variables y_0, y_1 and y_2. Suppose that all three random variables y_0, y_1 and y_2 have mean 0, and y_2 second-order stochastically dominates y_1.

[2] Of course, we must restrict attention here to distributions with finite support.

That is, $Ey_0 = Ey_1 = Ey_2 = 0$ and $y_1 \precsim^{s.o.} y_2$. Then

$$\begin{aligned} &\int\int_{y_0+y_2\geq 0} (y_0 + y_2)\, dG_0(y_0) dG_2(y_2) \leq \\ \leq \quad &\int\int_{y_0+y_1\geq 0} (y_0 + y_1)\, dG_0(y_0) dG_1(y_1) \end{aligned}$$

This lemma says that the benefit from discovering an additional attribute is higher for a stochastically dominated attribute than for a stochastically dominant attribute. The rough intuition is very simple. Before making the decision regarding the object, the agent prefers to resolve the uncertainty involving high stakes to resolving the uncertainty which involves only low stakes.

Proposition 2. Suppose the distributions $x_1, ..., x_n$ are ordered by second-order dominance away from the mean, i.e. $x_1 \precsim^{a.s.o.} ... \precsim^{a.s.o.} x_n$. Suppose further that

$$c_i - c_{i+1} \geq \int_0^{+\infty} x_i dF_i(x_i) - \int_0^{+\infty} x_{i+1} dF_{i+1}(x_{i+1}). \tag{3}$$

Then a set solving problem (2) has the form $S = \{k, ..., n\}$; that is, it consists of a k most <u>dominant</u> attributes.

Condition (3) can be interpreted as follows: Suppose the agent considers discovering only one attribute. By Lemma 1, the benefit from discovering an attribute is higher for a stochastically dominated attribute than for a stochastically dominant attribute. Condition (3) says that discovering the stochastically dominant attributes is, however, less expensive, and that the differences in the benefits are offset by the differences in the costs of discovering attributes. Proposition 2 says that if this condition is satisfied, then the agent should discover the most dominant attributes.

Proposition 2 is a consequence of the following lemma. However, its proof is a little less immediate in this case than in the case of Proposition 1. Therefore we relegate it, together with the proofs of Lemmas 1 and 2, to Appendix.

Lemma 2. Let G_0, G_1 and G_2 be the cdfs of random variables y_0, y_1 and y_2. Suppose that all three random variables y_0, y_1 and y_2 have mean 0, i.e., $Ey_0 = Ey_1 = Ey_2 = 0$. Suppose further that y_1 is second-order stochastically dominated away from the mean by y_2, i.e., $y_1 \precsim^{a.s.o.} y_2$. Then

$$\begin{aligned} &\int\int_{y_0+y_1\geq 0} (y_0 + y_1)\, dG_0(y_0) dG_1(y_1) - \int_0^{+\infty} y_1 dG_1(y_1) \leq \\ \leq &\int\int_{y_0+y_2\geq 0} (y_0 + y_2)\, dG_0(y_0) dG_2(y_2) - \int_0^{+\infty} y_2 dG_2(y_2). \end{aligned}$$

Lemma 2 says that the marginal benefit from discovering any additional attribute together with a stochastically dominated (away from the mean) attribute is lower than the marginal benefit of discovering this additional attribute with a stochastically dominant attribute. One interpretation of this lemma is the following principle:

An agent who considers obtaining signals from several sources should be inclined to obtain less informative but cheaper signals compared to the agent who obtains a signal just from one source.

The following example illustrates Lemmas 1 and 2.

Example 1. Suppose y_a with cdf G_a is distributed uniformly on the interval $[-a, a]$, and y_b with cdf G_b is distributed uniformly on the interval $[-b, b]$, where $b < a$. That is, y_a is second-order stochastically dominated away from the mean, and so is second-order stochastically dominated by y_b. Then

$$\int_0^{+\infty} y_a dG_a(y_a) = \frac{a}{4}; \int_0^{+\infty} y_b dG_b(y_b) = \frac{b}{4},$$

and

$$\int\int_{y_a+y_b\geq 0} (y_a + y_b)\, dG_a(y_a) dG_b(y_b) = \frac{a}{4} + \frac{b^2}{12a}. \quad (4)$$

Suppose first that $y_0 = y_a$, and $y_i = y_{b_i}$ for some $b_2 < b_1 < a$. That is, $y_1 \precsim^{s.o.} y_2$. Lemma 1, in this case, says that (4) increases in b.

Suppose now that $y_0 = y_b$, and $y_i = y_{a_i}$ for some $b < a_2 < a_1$. That is, $y_1 \precsim^{a.s.o.} y_2$. Lemma 2, in this case, says that

$$\int\int_{y_a+y_b\geq 0} (y_a + y_b)\, dG_a(y_a) dG_b(y_b) - \int_0^{+\infty} y_a dG_a(y_a) = \frac{b^2}{12a}$$

decreases in a.

It follows immediately from the proof of Proposition 1 that this proposition would hold true even when the agent's reservation utility V was not assumed to be equal to 0, i.e., the mean of attributes. Whereas, this assumption is essential for Proposition 2. More precisely, it is essential to assume that the reservation utility is no higher than the sum of attributes' means.

3.2 Other cases, open questions

Propositions 1 and 2, together with Lemmas 1 and 2, deliver algorithms for finding an optimal set S which requires computing $U(S)$ only at a polynomial number of sets S also in some other cases:

Corollary. Suppose that all attributes have binary distributions, i.e., $x_i = a_i$ with probability p_i and $x_i = -b_i$, where $a_i, b_i > 0$, with the remaining probability,[3] and the costs of discovering them are equal ($c_1 = ... = c_n = c$). Suppose further that either values a_i and b_i or probabilities p_i belong to some finite, independent of n set. Then there exists an algorithms for finding an optimal set S which requires computing $U(S)$ only at $P(n)$ sets S, where $P(n)$ is a polynomial function of the number of attributes n.

Indeed, suppose for example that $p_i \in \{.1, .2, ..., .9\}$ for $i = 1, ..., n$. Then, for a given p the set S_p of attributes with $p_i = p$ is ordered by the second-order stochastic dominance. By Lemma 1, the optimal set S contains only most dominated attributes from this set. Thus, we have to compute $U(S)$ only for

$$\prod_p |S_p| \leq \left(\frac{n}{9}\right)^9$$

sets S in order to determine a set solving problem (2).

We conjecture that for general distribution of attributes and costs there is no algorithm for finding an optimal set S which requires computing $U(S)$ only at a polynomial number of sets S. A little more specific is the question regarding binary distribution of attributes:

Question. Suppose that the distribution of each attribute x_i is binary, i.e., x_i takes value a_i or $-b_i$. Does there exist an algorithm for finding an optimal set S which requires computing $U(S)$ only at a polynomial number of sets S?

One may also wonder whether for any $r < 1$, there exists an approximate algorithm of approximation ratio r which requires computing $U(S)$ only at a polynomial number of sets S; or such a pseudo-polynomial time algorithm. Recall that an *approximation algorithm A* with an *approximation ratio r* is an algorithm with the property that for any set of parameters of the model,

$$\frac{U^* - V}{U^*} \leq r,$$

where $U^* := \max_{S \subset \{1,...,n\}} U(S)$ and V denotes the returned value of algorithm A.

We do not know the answers to these questions even in the case of equal costs, $c_1 = ... = c_n = c$, when attributes are not ordered by second-order stochastic dominance.

[3] By the assumption that the mean of each attribute is equal to zero, we must have that

$$b_i = \frac{p_i a_i}{1 - p_i}.$$

3.3 Greedy algorithm

The *greedy algorithm* is a standard method of solving optimization problems in the discrete convex analysis, in which the objective function $U : P(N) \to R$ is defined on the set $P(N)$ of all subsets of some finite set $N = \{1, ..., n\}$. The greedy algorithm recursively defines a subset $S \subset N$ by the following procedure: Suppose we have already defined a subset S; we begin with $S = \varnothing$. We compare the value of U at the set S and all sets $S \cup \{k\}$ for all $k \notin S$. If

$$U(S) \geq U(S \cup \{k\})$$

for every $k \notin S$, then the algorithm stops and returns the set S as the output. Otherwise, we take an arbitrary

$$k \in \arg\max_{k \notin S} U(S \cup \{k\}),$$

replace S with $S \cup \{k\}$ and repeat the procedure.

The greedy algorithm requires computing the value of $U(S)$ at only $1 + n(n+1)/2$ sets S. It turns out that the greedy algorithm does not solve problem (2), and it seems instructive to discuss intuitively the reason. When the agent picks the first attributes to discover, she may prefer stochastically dominated attributes, even if discovering them is more costly, because discovering them resolves the uncertainty involving high stakes. (See Lemma 1.) However, when the agent keeps adding additional attributes to the set of attributes to discover, the benefit of resolving these higher stakes becomes offset not only by a higher cost of discovering, but also by a lower benefit from discovering the additional attributes. (See Lemma 2.) Therefore in the process of adding the additional attributes, the agent may reach the point in which she no longer wants the stochastically dominated attributes, which were initially included, into the set of attributes to discover. The following example illustrates this observation:

Example 2. Suppose x_1 takes values a and $-a$, each with probability $1/2$; and x_2, x_3, and x_4 take values $b < a$ and $-b$, each with probability $1/2$. Suppose further that the cost of discovering attribute x_1 is $c > 0$, and the cost of discovering each of the remaining attributes is $0 < d < c$. Finally, take the parameters such that

$$\frac{1}{2}a - c > \frac{1}{2}b - d. \tag{5}$$

Then the greedy algorithm returns $S = \{1\}$. Indeed, condition (5) guarantees that attribute 1 is included during the first step of running the algorithm. And the algorithm stops after the first step by the assumption that $b < a$ and $d > 0$, because even if the agent discovers an additional attribute (together with attribute 1), the optimal decision whether to accept or reject the object depends only on the value of attribute 1.

However, the payoff to discovering attributes x_2, x_3, and x_4 is

$$\frac{3}{4}b - 3d,$$

and can be higher that $a/2 - c$. Moreover, if a is in addition much higher b, the optimal set of attributes to discover consists of attributes 2, 3, and 4.

3.4 Polynomial-time algorithms

A naive algorithm solving problem (2) would be to compare the values of our objective function across all sets $S \subset \{1, ..., n\}$, and select the highest, but this would take 2^n time, and would be impractical except small values of n. We would prefer to have an algorithm whose time of running is polynomial in n.

The running time of an algorithm depends not only on the number of values $U(S)$ we have to compute, but also on the parameters of the model. These parameters include the number of values each random variable $x_1, ..., x_n$ is allowed to take, the size of these values, measured by the number of digits in their $0 - 1$ expansions, and the size costs $c_1, ..., c_n$. Suppose the size of the parameters is commonly bounded by a number M. Then, an algorithm that is polynomial in M and n is called *pseudo polynomial*. To be truly polynomial, an algorithm must have running time that is independent in M and polynomial in n (*strongly polynomial*), or polynomial in $\log M$ and n (*weakly polynomial*). Polynomial-time algorithms (strongly or weakly) are considered practical in discrete convex analysis, while pseudo polynomial time algorithm are practical only for restricted sets of parameters. We refer the reader to McCormick (2008) or Fujishige (2005) for a more-detailed discussion[4].

Proposition 3. There is no polynomial-time algorithm solving problem (2) even in the case in which each x_i has a binary distribution taking value a_i with probability 1/2 and value $-a_i$ with probability 1/2.

[4] These texts are to large extent devoted to the discussion of polynomial-time algorithms for minimization of submodular functions, or equivalently, the maximization of supermodular functions. In turn, there may not exist polynomial-time algorithms for maximization of submodular functions. Our objective function is (in the general case) neither supermodular nor submodular.

Note that binary distributions taking value a_i with probability 1/2 and value $-a_i$ with probability 1/2 are ordered by second-order stochastic dominance. Therefore by Proposition 1, there exist algorithms which deliver an optimal set S, solving problem (2), by computing $U(S)$ at a polynomial number of sets S. And Proposition 3 follows from the fact that the computation of $U(S)$ for some sets S cannot be performed in polynomial time.

More precisely, let $g(n)$ denote the expected value of

$$\max\left(0, \sum_{i=1}^{n} x_i\right).$$

In the proof of Proposition 3, we show that if there was a polynomial-time algorithm for evaluating $g(n)$ for every positive natural numbers $a_1, ..., a_n$, then it would be the case that P=NP. This result in turn follows from the fact that the following problem of partitioning is NP-complete (see Garey and Johnson (1979)):

Given positive natural numbers $a_1, ..., a_n$, is there an $S \subset \{1, ..., n\}$ such that

$$\sum_{i \in S} a_i = \frac{1}{2} \sum_{i=1}^{n} a_i.$$

We relegate the details of the proof of Proposition 3 to Appendix. In contrast, pseudo polynomial-time algorithms for solving problem (2) often exist.

Proposition 4. If there exists an algorithm solving problem (2) which requires computing $U(S)$ only for $P(n)$ sets S, where $P(n)$ is a polynomial function of n, then there also exists a pseudo polynomial-time algorithm solving problem (2).

This proposition follows immediately from the following lemma:

Lemma 3. There exists a pseudo polynomial-time algorithm for computing $g(n)$.

The proof of the lemma is again relegated to Appendix. Of course, any pseudo polynomial-time algorithm requires computing $U(S)$ only for a polynomial number of sets S. Therefore all the results and questions of sections 3.1 and 3.2 can be equivalently formulated in terms of pseudo polynomial-time algorithms.

4 Sequential Discovering of Attributes

First, note that Propositions 1 and 2 generalize to the sequential discovering of attributes. That is, if attributes are ordered by stochastic dominance, and costs are nondecreasing, then it is an optimal strategy to discover second-order stochastically dominated attributes earlier; and if the costs satisfy condition (3), then second-order stochastically dominant away from the mean attributes should be discovered first.

We will solve the problem of sequential discovering even for attributes which are not ordered by stochastic dominance, but only in the case of two symmetrically distributed attributes x_1, x_2. The result generalizes to possibly asymmetric distributions, but the solution is much less elegant and harder to interpret.

Let

$$x_i^* = -c_i + \int_{-x_i^*}^{+\infty} (x_i^* + x_i) dF_i(x_i) \tag{6}$$

for $i = 1, 2$. This equation is interpreted as follows: Suppose that the agent discovers attribute j first and learn that its realization is x_i^*. Then, the agent is indifferent between accepting the object without discovering the realization of attribute i, and discovering the realization of attribute i and making then the decision regarding the object.

Notice that equation (6) is equivalent to the equation

$$0 = -c_i + \int_{x_i^*}^{+\infty} (-x_i^* + x_i) dF_i(x_i). \tag{7}$$

Indeed, (6) can be rewritten as

$$x_i^* F_i(-x_i^*) = -c_i + \int_{-x_i^*}^{+\infty} x_i dF_i(x_i), \tag{8}$$

and (7) can be rewritten as

$$x_i^* [1 - F_i(x_i^*)] = -c_i + \int_{x_i^*}^{+\infty} x_i dF_i(x_i). \tag{9}$$

By the symmetry of x_i, the left-hand sides of these two equations coincide. The right-hand sides coincide as well; indeed, since the mean of x_i is zero,

$$\int_{-x_i^*}^{+x_i^*} x_i dF_i(x_i) = 0$$

by the symmetry of x_i.

Equation (7) says that when the agent discovers that the realization of attribute j is $-x_i^*$, she is indifferent between rejecting the object without discovering the realization of attribute i, and discovering the realization of attribute i and making then the decision regarding the object.

Notice, if the distribution of x_i is continuous, that this equation has a unique solution $x_i^* > 0$, provided that the agent prefers discovering attribute i to accepting (or, equivalently, rejecting) the object without discovering the realization of attribute i.[5][6] Indeed, for $x_i^* = 0$ the left-hand side of (6) falls below the right-hand side when

$$0 < -c_i + \int_0^{+\infty} x_i dF_i(x_i), \tag{10}$$

and this condition means that the agent prefers discovering attribute i to accepting (rejecting) the object without discovering the realization of attribute i. For sufficiently large values of x_i^*, left-hand side of (6) exceeds the right-hand side. And the difference between the left-hand side and the right-hand side is increasing in x_i^*. Thus, condition (10) guarantees, that equation (6) has a unique solutions.

Notice finally that the value of x_i^* is determined only by attribute i (that is, by c_i and F_i), and is independent of attribute j.

Proposition 5. According to every optimal strategy, the agent should discover attribute i first whenever $x_j^* < x_i^*$. After discovering the realization x_i of attribute i, the agent should accept the object whenever $x_j^* < x_i$; she should reject the object whenever $x_i < -x_j^*$; otherwise, she should discover the realization x_j of attribute j, accept the object when $x_i + x_j > 0$ and reject the object when $x_i + x_j < 0$.

5 Appendix

5.1 Proofs of Lemmas 1 and 2, and Proposition 2

Proof of Lemma 1. We need to show that the integral

$$\int \int_{y_0 + y_i \geq 0} (y_0 + y_i)\, dG_0(y_0) dG_i(y_i)$$

increases when y_i gets replaced with a second-order stochastically dominated variable. This integral is exhibited in Figure 1(a).

Recall that any second-order stochastically dominated variable can be represented as a mean-preserving spread of y_i, that is, as the compound lottery such that in the first stage, we have a

[5] If this condition is violated, we may take $x_i^* = 0$.

[6] If the distribution of x_i is discrete, then there exist ${}^-x_i^*$ and ${}^+x_i^*$ such that the left-hand side of equation (7) is no higher that the right-hand side at ${}^-x_i^*$, and the left-hand side of equation (7) is no lower that the right-hand side at ${}^+x_i^*$.

The analysis of this section can be replicated for the discrete, and even for general distributions. However, we restrict attention to continuous distributions for the sake of simplicity.

lottery y_i; and in the second stage, we randomize each possible outcome of y_i further so the final outcome is $y_i + x$, where the distribution of x has mean zero, and may depend on the outcome of y_i.

It follows that if the integral was taken over the entire plane of outcomes (y_i, y_0), not only the outcomes such that $y_0 + y_i \geq 0$, replacing y_i with a second-order stochastically dominated variable would not affect the integral. Since we integrate only over the outcomes such that $y_0 + y_i \geq 0$, the integral may be affected. But the integral may only become higher. This can be easily seen from Figure 1(a). If for an outcome of (y_i, y_0), all outcomes of x are such that $y_0 + y_i + x \geq 0$ (an example of such a case is depicted in dashed arrows), then adding x to y_i does not affect the integral. And for the outcomes of (y_i, y_0) such that $y_0 + y_i + x < 0$ for some outcomes of x (an example of such a case is depicted in solid arrows), adding x to y_i may only make the integral higher, because the outcomes such that $y_0 + y_i + x < 0$ are located out of the region of integration.

In the proof of Lemma 2, we will need the following proposition, which is a straightforward generalization of the characterization of second-order stochastic dominance in terms of mean-preserving spreads and the integrals of cdfs.

Consider the following compound lottery: In the first stage, we have a lottery z, and in the second stage, we randomize each possible outcome of z further so the final outcome is $z + x$, where the distribution of x depends on the outcome of z. This randomization is required to satisfy the following two conditions: First, when the outcome of z is positive, x takes only positive values; and when the outcome of z is negative, x takes only negative values. Second, the randomization preserves the mean of z, i.e., the mean of the compound lottery is zero. When a lottery y can be obtained from lottery z in this manner for some x, we will say that y is a *(mean-preserving) spread* of z *away from the mean*.

Proposition 6. Consider random variables y and z with the mean zero and with cdfs G and H, respectively. Then the following statements are equivalent:

(i) y is second-order stochastically dominated away from the mean by z;

(ii) for any function $f : R \to R$ which is increasing on $(-\infty, 0)$ and decreasing on $(0, +\infty)$,

$$\int_{-\infty}^{+\infty} f(y) dG(y) \leq \int_{-\infty}^{+\infty} f(z) dH(z);$$

(iii) y is a spread of z away from the mean.

Proof of Lemma 2. Notice that

$$\int\int_{y_0+y_i\geq 0} (y_0 + y_i)\, dG_0(y_0) dG_i(y_i) - \int_0^{+\infty} y_i dG_i(y_i) =$$

$$= \int_0^{+\infty} \left(\int_{-y_0}^0 y_i dG_i(y_i) \right) dG_0(y_0) - \int_{-\infty}^0 \left(\int_0^{-y_0} y_i dG_i(y_i) \right) dG_0(y_0) +$$

$$+ \int \int_{y_0 + y_i \geq 0} y_0 dG_0(y_0) dG_i(y_i) =$$

$$= \int_0^{+\infty} \left(\int_{-y_0}^0 (y_0 + y_i) dG_i(y_i) \right) dG_0(y_0) + \int_{-\infty}^0 \left(\int_0^{-y_0} (-y_0 - y_i) dG_i(y_i) \right) dG_0(y_0).$$

This last two integrals are exhibited in Figure 1(b). We need to show that these integrals become lower when y_i gets replaced with a variable which is second-order stochastically dominated by y_i away from the mean. This follows from Proposition 6. Indeed, consider the integral

$$\int_0^{+\infty} \left(\int_{-y_0}^0 (y_0 + y_i) dG_i(y_i) \right) dG_0(y_0).$$

If a variable is a spread of y_i away from the mean, then the values of $y_0 + y_i + x$ are always no higher than the values of $y_0 + y_i$, because $y_i \leq 0$ over the entire region of integration. This means that any $y_0 + y_i$ from the region of integration is replaced either with a still positive but no higher $y_0 + y_i + x$ (an example of such a case is depicted in Figure 1(b) in dashed arrows), or with a negative $y_0 + y_i + x$, which is located out of the region of integration (an example of such a case is depicted in Figure 1(b) in solid arrows).

An analogous argument applies to the integral

$$\int_{-\infty}^0 \left(\int_0^{-y_0} (-y_0 - y_i) dG_i(y_i) \right) dG_0(y_0).$$

Proof of Proposition 2. Suppose that some set T solving problem (2) does not have the form $T = \{k, ..., n\}$. Consider the set R obtained from T by replacing any $i \in T$ with any $i < j \notin T$. By Lemma 2,

$$\int ... \int_{x_i, i \in R} \max \left(0, \sum_{i \in R} x_i \right) - \int ... \int_{x_i, i \in T} \max \left(0, \sum_{i \in T} x_i \right) \geq$$

$$\geq \int_0^{+\infty} x_j dF_j(x_j) - \int_0^{+\infty} x_i dF_i(x_i).$$

Indeed, apply Lemma 2 to

$$y_0 = \sum_{i \in T - \{i\}} x_i = \sum_{i \in R - \{j\}} x_i,$$

$y_1 = x_i$ and $y_2 = x_j$.

By assumption,

$$\int_0^{+\infty} x_j dF_j(x_j) - \int_0^{+\infty} x_i dF_i(x_i) \geq c_j - c_i.$$

Therefore, since the cost of the attributes in R and the cost of the attributes in T differ only by $c_j - c_i$, it follows that the value of (1) for R is no lower than the value of (1) for T.

Replacing recursively the attributes in T by attributes that stochastically dominate them, we obtain a set of the form $S = \{k, ..., n\}$ for which the value of (1) is no lower than that for T.

5.2 Proof of Proposition 3

Suppose that $a_1, ..., a_n$ are positive natural numbers. Let $a = (a_1, ..., a_n)$, $\overline{a} = (a_1, ..., a_n, a_{n+1})$, where $a_{n+1} = 1$,

$$M := \frac{1}{2}\sum_{i=1}^{n} a_i,$$

and for any $\varnothing \neq S \subset \{1, ..., n\}$,

$$a(S) := \sum_{i \in S} a_i.$$

Define $\overline{M}$ and $\overline{a}(\overline{S})$, for any $\varnothing \neq \overline{S} \subset \{1, ..., n+1\}$, in a similar manner.

Let $u(a)$ stand for the number of sets S such that $a(S) = M$. The idea is to show that if we were able to compute expression $g(n)$ in polynomial time, then we would also be able to compute $u(a)$ in polynomial time. And then, by checking whether $u(a) = 0$, we would be able to solve the problem of partitioning.

Assume, without loss of generality, that M is an integer. (Indeed, if M is not an integer, there obviously exists no S such that $a(S) = M$.)

If S denotes set $\{i : x_i = -a_i\}$ and T denotes set $\{i : x_i = a_i\}$, then

$$\sum_{i=1}^{n} x_i = \sum_{i \in T} a_i - \sum_{i \in S} a_i = \left(\sum_{i=1}^{n} a_i - \sum_{i \in S} a_i\right) - \sum_{i \in S} a_i = -2\sum_{i \in S} a_i + \sum_{i=1}^{n} a_i,$$

and so

$$\sum_{i=1}^{n} x_i \geq 0 \;\Leftrightarrow\; \sum_{i \in S} a_i \leq \frac{1}{2}\sum_{i=1}^{n} a_i.$$

Thus,

$$g(n) = \frac{1}{2^n}\left\{\sum_{i=1}^{n} a_i + \sum\left[\left(-2\sum_{i \in S} a_i + \sum_{i=1}^{n} a_i\right) : \varnothing \neq S \subset \{1, ..., n\},\ \sum_{i \in S} a_i \leq \frac{1}{2}\sum_{i=1}^{n} a_i\right]\right\} =$$

$$= \frac{1}{2^{n-1}}\left\{\frac{1}{2}\sum_{i=1}^{n} a_i + \sum\left[\left(\frac{1}{2}\sum_{i=1}^{n} a_i - \sum_{i \in S} a_i\right) : \varnothing \neq S \subset \{1, ..., n\},\ \sum_{i \in S} a_i \leq \frac{1}{2}\sum_{i=1}^{n} a_i\right]\right\}.$$

Further, define

$$h(a) := 2^{n-1} \cdot g(n) - \frac{1}{2}\sum_{i=1}^{n} a_i = \sum_{\varnothing \neq S \subset \{1,...,n\}: a(S) \leq M} (M - a(S)),$$

$$f(a) := \sum_{\varnothing \neq S \subset \{1,...,n\}: a(S) \leq M} 1,$$

$$h(a,-1) = \sum_{\varnothing \neq S \subset \{1,...,n\}: a(S) \leq M-1} (M - 1 - a(S))$$

and

$$f(a,-1) := \sum_{\varnothing \neq S \subset \{1,...,n\}: a(S) \leq M-1} 1.$$

If we were able to compute $g(n)$ in polynomial time, then we would able to compute $h(a)$ and $h(\overline{a})$. We will now derive a nonsingular system of five linear equation with variables $f(a)$, $f(\overline{a})$, $f(a,-1)$, $h(a,-1)$, $u(a)$ and with constant terms including $h(a)$ and $h(\overline{a})$. Solving this system, we will find the value of $u(a)$.

Since all numbers $a(S)$ are integers, $M - a(S) = 0$ in the expression for $h(a)$ unless $a(S) \leq M - 1$. This yields that

$$h(a) = h(a,-1) + f(a,-1). \tag{11}$$

Further, observe that

$$-1 = f(a) + f(a,-1) - f(\overline{a}). \tag{12}$$

Indeed, $f(\overline{a})$ is the number of nonempty sets $\overline{S} \subset \{1,...,n+1\}$ such that $\overline{a}(\overline{S}) \leq \overline{M}$. There are three types of such sets $\overline{S}$: **(i)** If $n+1 \notin \overline{S} \subset \{1,...,n+1\}$, then $\overline{a}(\overline{S}) = a(\overline{S}) \leq \overline{M} = M + 1/2 \Leftrightarrow a(\overline{S}) \leq M$; **(ii)** If $n+1 \in \overline{S} \neq \{n+1\}$, then for $S := \overline{S} - \{n+1\}$, we have $\overline{a}(\overline{S}) = a(S) + 1 \leq \overline{M} = M + 1/2 \Leftrightarrow a(S) \leq M - 1$; **(iii)** If $\overline{S} = \{n+1\}$, then $\overline{a}(\overline{S}) = 1$. This yields that $f(\overline{a}) = f(a) + f(a,-1) + 1$.

Similarly,

$$h(\overline{a}) = \sum_{\varnothing \neq \overline{S} \subset \{1,...,n+1\}: \overline{a}(\overline{S}) \leq \overline{M}} (\overline{M} - \overline{a}(\overline{S})) =$$

$$= \sum_{\varnothing \neq \overline{S} \subset \{1,...,n\}: \overline{a}(\overline{S}) \leq \overline{M}} (\overline{M} - \overline{a}(\overline{S})) + \sum_{\{n+1\} \in \overline{S} \subset \{1,...,n+1\}: \overline{a}(\overline{S}) \leq \overline{M},\ \overline{S} \neq \{n+1\}} (\overline{M} - \overline{a}(\overline{S})) + \left(\overline{M} - 1\right) =$$

$$= \sum_{\varnothing \neq S \subset \{1,...,n\}: a(S) \leq M} (M + 1/2 - a(S)) + \sum_{\varnothing \neq S \subset \{1,...,n\}: a(S) \leq M-1} (M - 1/2 - a(S)) + (M - 1/2) =$$

$$= h(a) + \frac{1}{2} f(a) + h(a,-1) + \frac{1}{2} f(a,-1) + M - 1/2.$$

This is equivalent to

$$2\left(h(\overline{a})-h(a)+1/2-M\right)=f(a)+f(a,-1)+2h(a,-1). \tag{13}$$

Clearly,

$$u(a)=\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)=M}1=\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)\leq M}1-\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)\leq M-1}1=f(a)-f(a,-1);$$

that is,

$$0=-u(a)+f(a)-f(a,-1). \tag{14}$$

Finally,

$$\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)\leq M-1}1+\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)=M}1+\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)\geq M+1}1=2^n-1.$$

Since $a(S)\leq M-1\Leftrightarrow a(T)\geq M+1$, where T stands for the complement of S,

$$2\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)\leq M-1}1+\sum_{\varnothing\neq S\subset\{1,\dots,n\}:a(S)=M}1=2^n-1;$$

that is,

$$2^n-1=2f(a,-1)+u(a). \tag{15}$$

We can now solve the system of equation (11)-(15), and compute $f(a)$, $f(\overline{a})$, $f(a,-1)$, $h(a,-1)$, $u(a)$ as a function of $h(a)$ and $h(\overline{a})$. Indeed, the matrix

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 2 & 0 \\ 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 2 & 0 & 1 \end{bmatrix}$$

is nonsingular. Therefore, if we were able to compute $h(a)$ and $h(\overline{a})$ in polynomial time, then we would also be able to compute $u(a)$ in polynomial time.

5.3 Proof of Lemma 3

Suppose that x_i, $i=1,\dots,n$, takes values a_i in a finite set S_i.

For $t=1,\dots,n$ and $a^t=(a_1,\dots,a_t)$, define

$$p_t(a^t):=\prod_{i=1}^{t}\Pr\{x_i=a_i\},$$

and for any $t = 1, ..., n$, define

$$V_t(b) = \sum \left\{ p_t(a^t) \cdot \left(\sum_{i=1}^{t} a_i \right) : \sum_{i=1}^{t} a_i \geq b \right\}.$$

We will develop a recursive definition of $V_t(b)$:

$$V_t(b) = \sum \left\{ p_t(a^t) \cdot \left(a_t + \sum_{i=1}^{t-1} a_i \right) : \sum_{i=1}^{t} a_i \geq b \right\} =$$

$$= \sum \left\{ p_t(a^t) \cdot a_t + \Pr\{x_t = a_t\} p_{t-1}(a^{t-1}) \cdot \left(\sum_{i=1}^{t-1} a_i \right) : \sum_{i=1}^{t} a_i \geq b \right\} =$$

$$= \sum \left\{ p_t(a^t) \cdot a_t : \sum_{i=1}^{t} a_i \geq b \right\} + \sum_{a_t} \Pr\{x_t = a_t\} V_{t-1}(b - a_t).$$

Let

$$U_t(b) := \sum \left\{ p_t(a^t) \cdot a_t : \sum_{i=1}^{t} a_i \geq b \right\}.$$

Then

$$U_t(b) = \sum_{a_t} a_t \Pr\{x_t = a_t\} \cdot \left(\sum \left\{ p_{t-1}(a^{t-1}) : \sum_{i=1}^{t-1} a_i \geq b - a_t \right\} \right).$$

Let

$$z_t(b) := \sum \left\{ p_t(a^t) : \sum_{i=1}^{t} a_i \geq b \right\}.$$

Then

$$z_t(b) = \sum_{a_t} \Pr\{x_t = a_t\} \cdot \left(\sum \left\{ p_{t-1}(a^{t-1}) : \sum_{i=1}^{t-1} a_i \geq b - a_t \right\} \right) =$$

$$= \sum_{a_t} \Pr\{x_t = a_t\} \cdot z_{t-1}(b - a_t).$$

Therefore $z_t(b)$ can be computed recursively, and so $U_t(b)$ and $V_t(b)$ can be computed recursively.

The recursive definition of $V_t(b)$ provides an algorithm for computing $g(n) = V_n(0)$ whose running time is polynomial in n,

$$s := \max_{i=1,...,n} |S_i|,$$

and the common bound on the size of $a_i \in S_i$, $i = 1, ..., n$.

5.4 Proof of Proposition 5

The only nonobvious part of Theorem 1 is the rule regarding which attribute to discover first. Suppose that $x_1^* < x_2^*$. The payoffs contingent on any possible pair of realizations of the two attributes are exhibited in Figure 2(a). The top row in each area is the payoff from discovering the realization of attribute 1 first, and playing the optimal continuation strategy contingent on any realization of this attribute. Recall that according to this strategy, the agent should accept the object whenever $x_2^* < x_1$; she should reject the object whenever $x_1 < -x_2^*$; and when $-x_2^* < x_1 < x_2^*$, she should discover the realization x_2 of attribute 2, accept the object when $x_1 + x_2 > 0$ and reject the object when $x_1 + x_2 < 0$.

The bottom row is the payoff from discovering the realization of attribute 2 first, but then playing the suboptimal continuation strategy, according to which the agent should accept the object whenever $x_2^* < x_2$; she should reject the object whenever $x_2 < -x_2^*$; and when $-x_2^* < x_2 < x_2^*$, she should discover the realization x_1 of attribute 1, accept the object when $x_1 + x_2 > 0$ and reject the object when $x_1 + x_2 < 0$. (The optimal continuation strategy would have threshold x_1^* not x_2^*.)

We obtain Figure 2(b) from Figure 2(a) by deleting the common components of corresponding top and bottom payoffs. Finally, we obtain Figure 2(c) from Figure 2(b) by means of (6) and (7). Notice that for any given value of x_1, the component $-c_2$ appears in the bottom row of Figure 2(b) either for every single value of x_2 or for no value of x_2. The component $-c_2$ appears for every single value of x_2 when $x_1 > x_2^*$ or when $x_1 < -x_2^*$. In the former case, we can replace $-c_2$ with $x_2^* + x_2$ for $x_2 < -x_2^*$ and with 0 for $x_2 > -x_2^*$. This does not affect the integral of the bottom row payoff across all pairs of realizations of the two attributes, because by virtue of (8) and the fact that

$$\int_{-x_2^*}^{+\infty} x_2 dF_2(x_2) = \int_{-x_2^*}^{+x_2^*} x_2 dF_2(x_2) + \int_{+x_2^*}^{+\infty} x_2 dF_2(x_2) = -\int_{-\infty}^{-x_2^*} x_2 dF_2(x_2),$$

which follows from the symmetry of x_2. Similarly, we can replace $-c_2$ with $x_2^* - x_2$ for $x_2 > x_2^*$ and with 0 for $x_2 < x_2^*$ in the latter case, without affecting the integral of the bottom row across all pairs of realizations of the two attributes, by virtue of (9).

We replace $-c_1$ in the top row of Figure 2(b) in a similar fashion. More precisely, the component $-c_1$ appears for every single value of x_1 when $x_2 > x_2^*$ or when $x_2 < -x_2^*$. In the former case, we replace $-c_1$ with $x_2^* + x_1$ for $x_1 < -x_2^*$ and with 0 for $x_1 > -x_2^*$. And in the latter case, we replace $-c_1$ with $x_2^* - x_1$ for $x_1 < -x_2^*$ and with 0 for $x_1 > -x_2^*$. This will increase the integral of the top row across all pairs of realizations of the two attributes by (8) and (9). Indeed, we would not affect the integral by replacing $-c_1$ with $x_1^* + x_1$ for $x_1 < -x_1^*$ and with 0 for $x_1 > -x_1^*$ in the former case, and replacing $-c_1$ with $x_1^* - x_1$ for $x_1 < -x_1^*$ and with 0 for $x_1 > -x_1^*$ in the latter case. However, $x_1^* < x_2^*$, so the left-hand sides of (8) and (9) exceed the right-hand sides of the two equations, respectively.

The entries in the top and bottom rows of Figure 2(c) coincide. This implies that the expected payoff to discovering attribute 2 first, followed by playing a suboptimal continuation strategy is no lower than the expected payoff to discovering attribute 1 first, followed by playing the optimal continuation strategy.

6 References

Chade H. and L. Smith (2005): "Simultaneous Search," *Econometrica* **74**, 1293–1307.

Fujishige S. (2005). *Submodular Functions and Optimization.* Elsevier, Amsterdam.

Garey M.R. and D.S. Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H. Freeman & Co., New York.

Gittins, J.C. and D.M. Jones (1974): "A dynamic allocation index for the sequential design of experiments," in *Progress in Statistics (J. Gani et al., eds.)*, 241–266, North-Holland, Amsterdam.

Gittins, J.C. (1989). *Multi-armed bandit allocation indices.* John Wiley & Sons, Ltd., New York.

Manzini, P. and M. Mariotti (2007): "Sequentialy Rationalizable Choice," *American Economic Review*, **97**, 1824-1839.

McCormick S.T. (2008): "Submodular Function Optimization," Mimeo.

Neeman, Z. (1995): "On Determining the Importance of Attributes with a Stopping Problem," *Mathematical Social Sciences*, **29**, 195-212.

Tversky, A. (1972): "Elimination by Aspects: A Theory of Choice," *Psychological Review*, **79**, 281–99.

Weber, R. (1992): "On the Gittins Index for Multiarmed Bandits," *Annals of Applied Probability*, **2**, 1024-1033.

Weitzman, M.L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, **47**, 641-654.

Whittle, P. (1980): "Multi-armed Bandits and the Gittins Index," *Journal of the Royal Statistical Society Ser. B (Methodology),* **42**, 143–149.

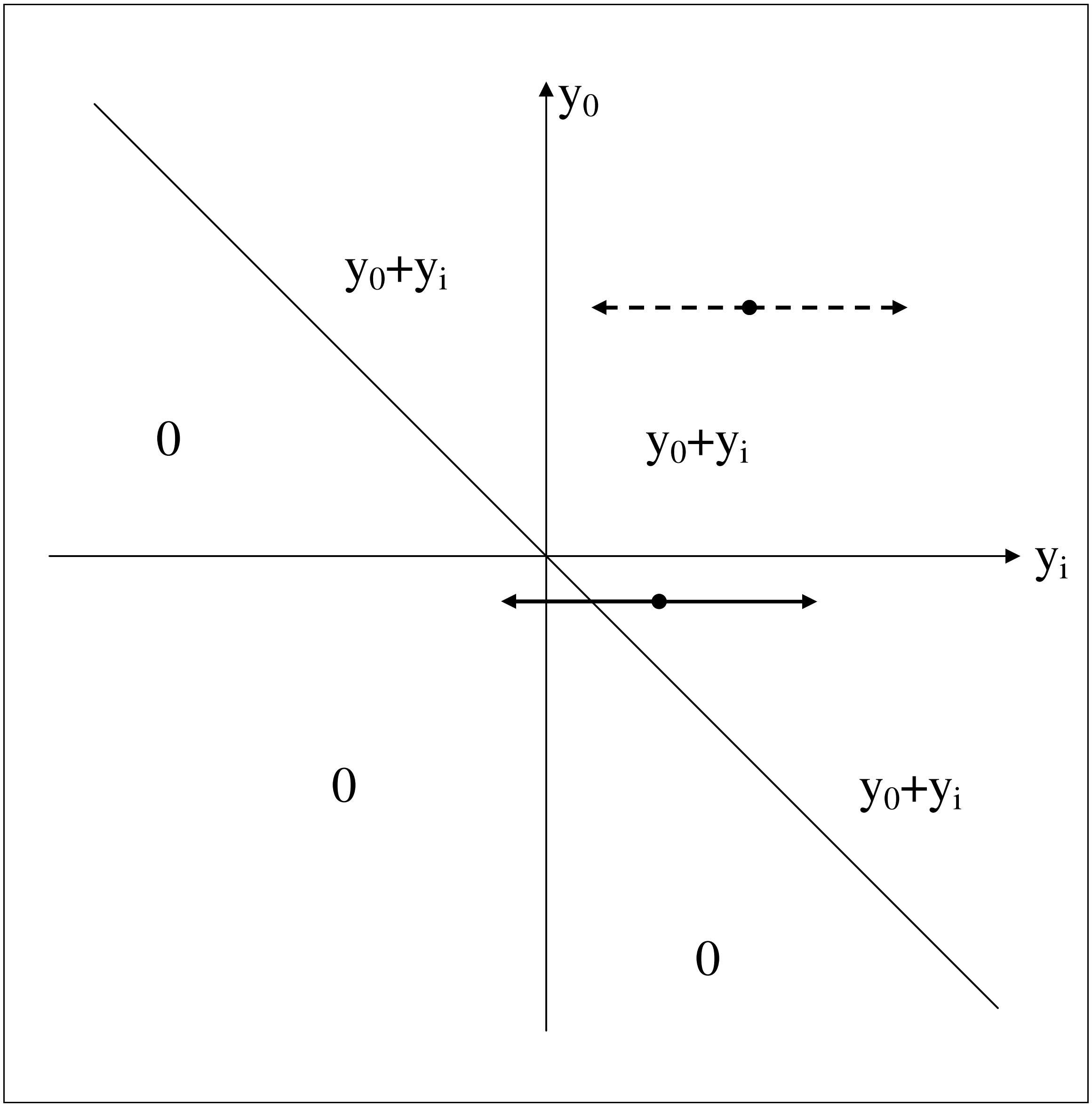


Figure 1 (a)

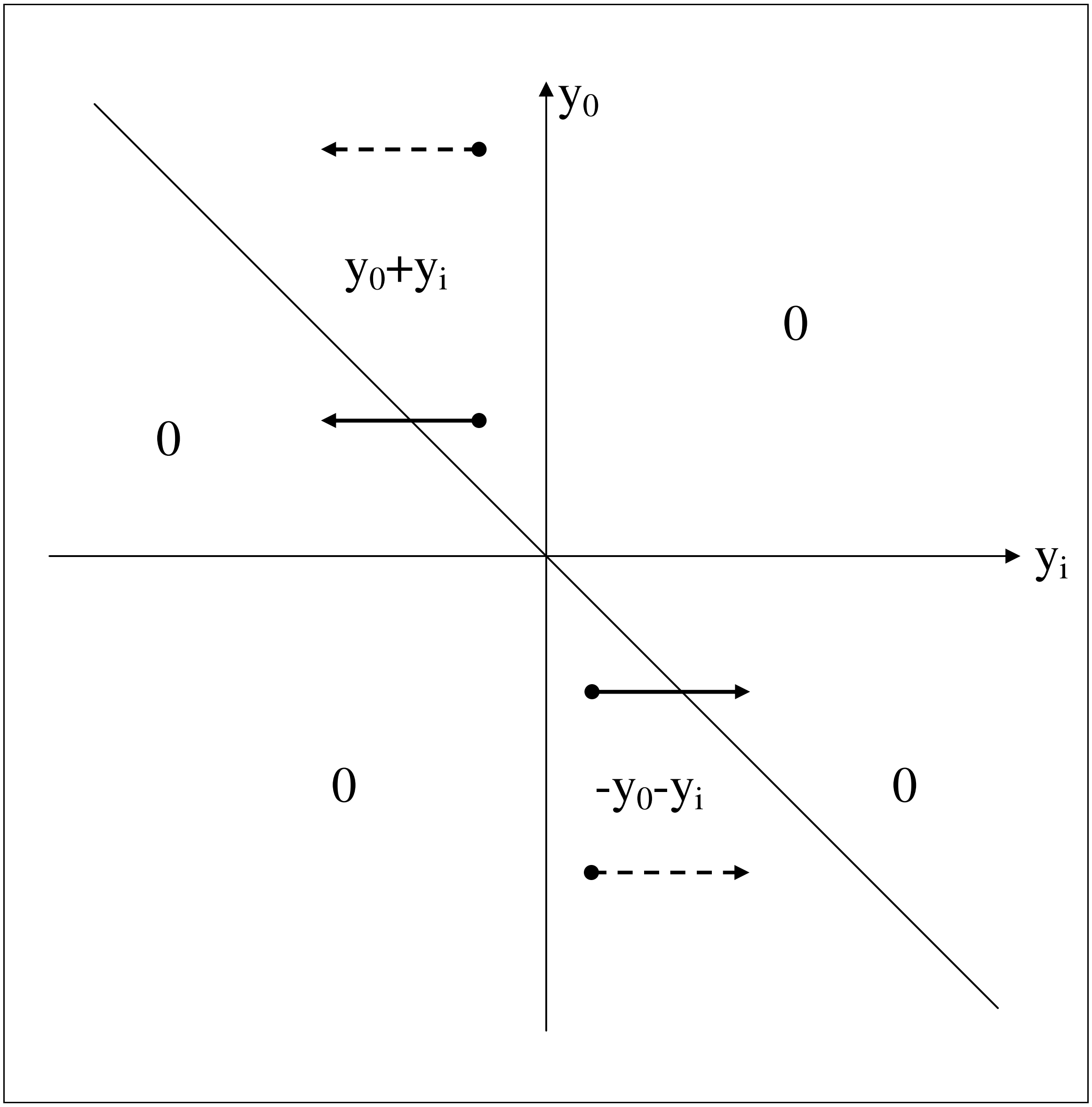


Figure 1 (b)

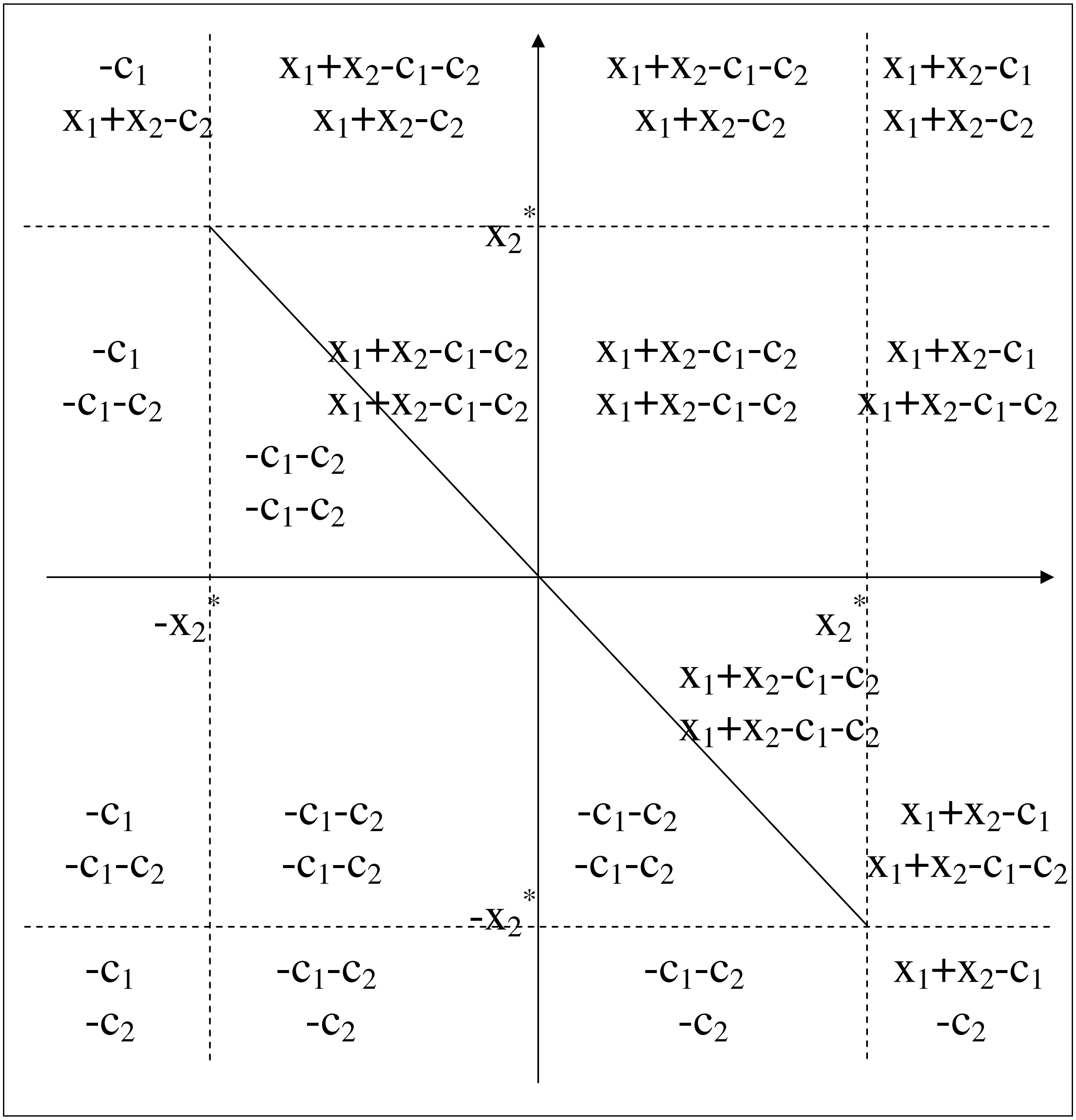
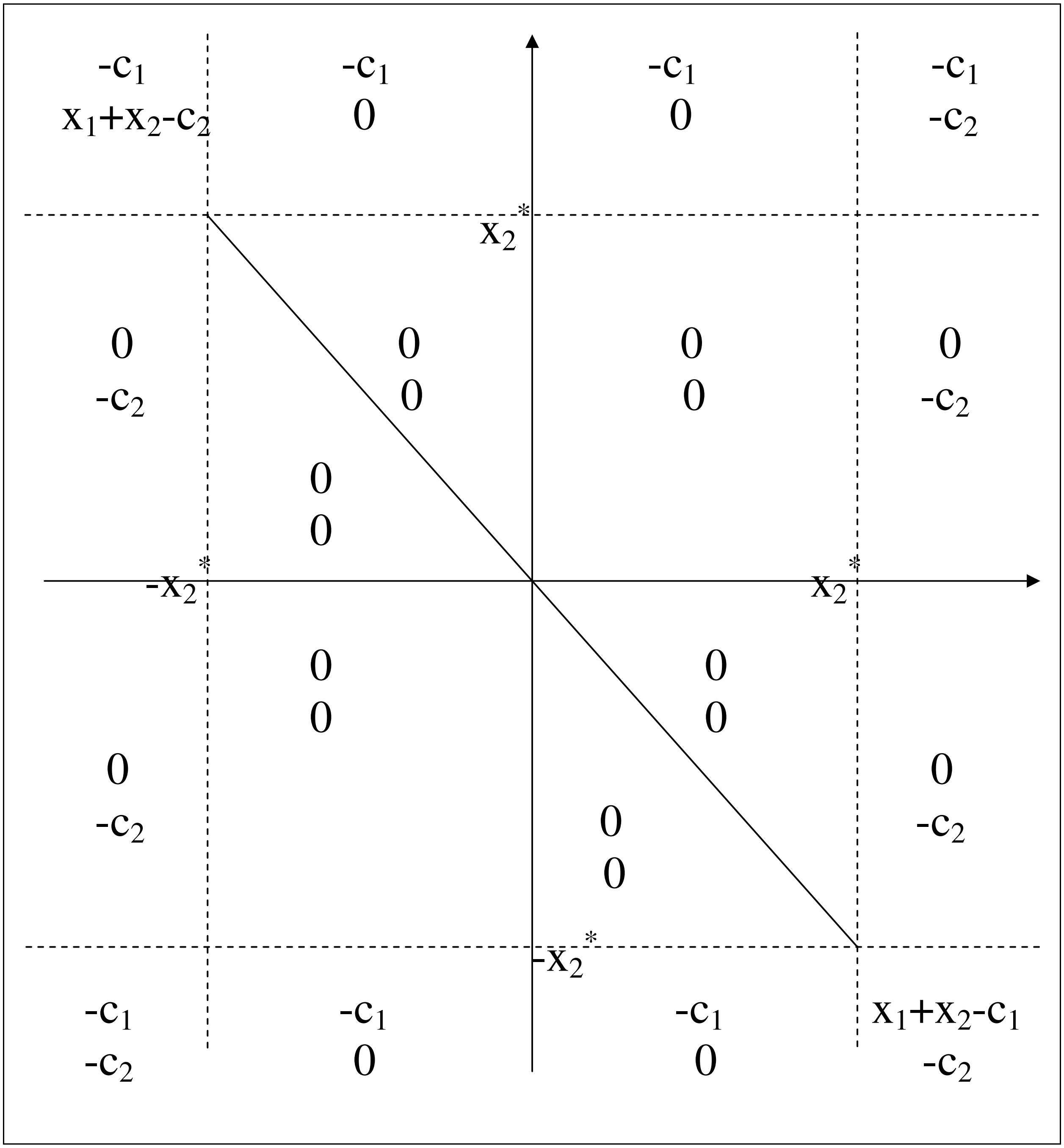

-c1
x1+x2-c2
x1+x2-c1-c2
x1+x2-c2
x1+x2-c1-c2
x1+x2-c2
x1+x2-c1
x1+x2-c2
x2*
-c1
-c1-c2
x1+x2-c1-c2
x1+x2-c1-c2
x1+x2-c1-c2
x1+x2-c1-c2
x1+x2-c1
x1+x2-c1-c2
-c1-c2
-c1-c2
-x2*
x2*
x1+x2-c1-c2
x1+x2-c1-c2
-c1
-c1-c2
-c1-c2
-c1-c2
-c1-c2
-c1-c2
x1+x2-c1
x1+x2-c1-c2
-x2*
-c1
-c2
-c1-c2
-c2
-c1-c2
-c2
x1+x2-c1
-c2

Figure 2 (a)



-c1
x1+x2-c2
-c1
0
-c1
0
-c1
-c2
x2*
0
-c2
0
0
0
0
0
-c2
0
0
-x2*
x2*
0
0
0
0
0
-c2
0
0
0
-c2
-x2*
-c1
-c2
-c1
0
-c1
0
x1+x2-c1
-c2

Figure 2 (b)

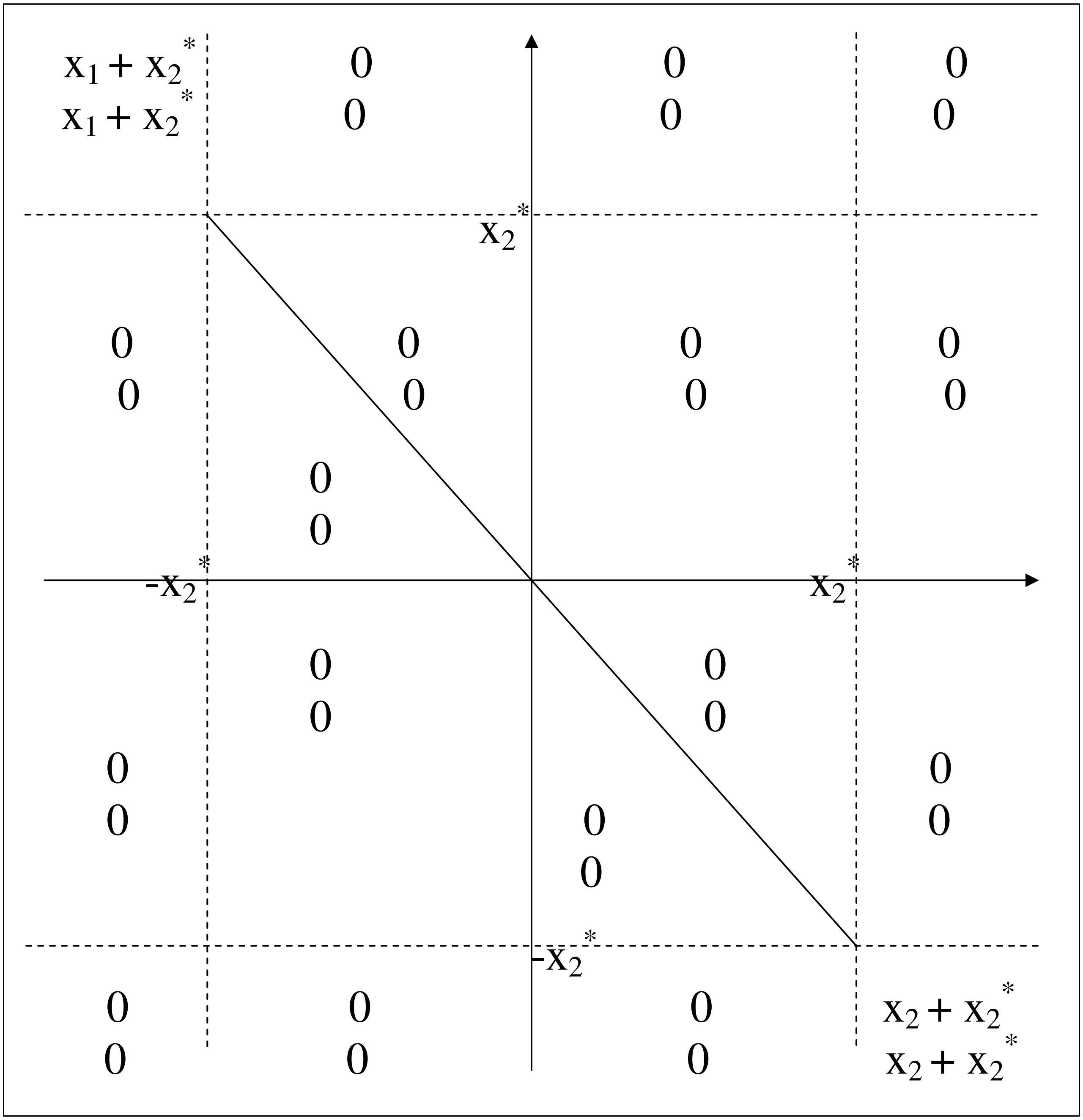

$x_1 + x_2^*$
$x_1 + x_2^*$
0
0
0
0
0
0
x_2^*
0
0
0
0
0
0
0
0
0
0
$-x_2^*$
x_2^*
0
0
0
0
0
0
0
0
0
0
$-x_2^*$
0
0
0
0
0
0
$x_2 + x_2^*$
$x_2 + x_2^*$

Figure 2 (c)