Math Meth Oper Res (1999) 50:101–120

Mathematical Methods
of Operations Research

A class of extremum problems related to agency models with imperfect monitoring

Jörg Budde[1], **Norbert Gaffke**[2]

[1] Faculty of Economics and Management, University of Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany (e-mail: joerg.budde@wirtschafts-w.uni-magdeburg.de)
[2] Faculty of Mathematics, University of Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany (e-mail: norbert.gaffke@mathematik.uni-magdeburg.de)

Abstract. The cost minimization problem in an agency model with imperfect monitoring is considered. Under the first order approach, this can be stated as a convex minimization problem with linear inequality and equality constraints in a generally infinite dimensional function space. We apply the Fenchel Duality Theorem, and obtain as a dual problem a concave maximization problem of finite dimension. In particular, a Lagrange multiplier description of the optimal solution to the cost minimization problem is derived, justifying and extending thus the approach of Kim (1995). By the duality, the dependence of the minimum cost value on the information system used becomes particularly visible. The minimum cost value behaves monotonically w.r.t. the *convex ordering* of certain distributions induced by the competing information systems. Under the standard inequality constraint, one is led to the distributions of the score functions of the information systems and their convex order relation. It is shown that also for multivariate actions, Blackwell sufficiency implies the convex order relation of the score function distributions. A further result refers to a multi-agents model recently considered by Budde (1997), when the maximum of n independent and identically distributed (i.i.d.) univariate output variables is focussed. If two univariate information systems have monotone likelihood ratios, then the convex ordering between the two score function distributions implies the weaker *convex increasing ordering* between the distributions of the same score functions under the maximum distributions.

Key words: Convex order of distributions, Fenchel duality, information systems, moral hazard, principal-agent problem

Manuscript received: October 1998/final version received: March 1999

1. Introduction

We consider the following extremum problem. Let $(M, \mathscr{B}, P)$ be a probability space, Φ be a convex function on $[0, \infty)$ with values in $[0, \infty] = [0, \infty) \cup \{\infty\}$ and $\Phi(0) = 0$, where Φ is neither constantly zero nor constantly infinity on $(0, \infty)$. Let g_i, $i = 1, \ldots, m$, and h_j, $j = 1, \ldots, n$, be real valued measurable functions on M. Then, the problem is to

$$\text{minimize} \quad \mathrm{E}_P(\Phi \circ f) \tag{1.1a}$$

over all real valued measurable functions f on M (belonging to some suitable linear space of functions, see Section 2) which satisfy the constraints

$$f(x) \geq b \quad \text{for all } x \in M, \tag{1.1b}$$

$$\mathrm{E}_P(f \cdot g_i) \geq \alpha_i \quad \text{for all } i = 1, \ldots, m, \tag{1.1c}$$

$$\mathrm{E}_P(f \cdot h_j) = \beta_j \quad \text{for all } j = 1, \ldots, n, \tag{1.1d}$$

where $0 \leq b < \infty$, $\alpha_i \in \mathbb{R}$ (for $i = 1, \ldots, m$), and $\beta_j \in \mathbb{R}$ (for $j = 1, \ldots, n$) are given constants. By $\mathrm{E}_P(d)$ we have denoted the expectation of a real valued measurable function d on M w.r.t. the probability distribution P (provided that the expectation exists), and $\Phi \circ f$ denotes the composition of f and Φ, i.e., $(\Phi \circ f)(x) = \Phi(f(x))$ for all $x \in M$. Note that if Φ is not finite, i.e., $\Phi(u) = \infty$ for all $u > \bar{u}$, say, then the constraint $f \leq \bar{u}$ is implicitly present, since for any feasible f which exceeds $\bar{u}$ with positive probability we have $\mathrm{E}_P(\Phi \circ f) = \infty$. In fact, for reasons to be discussed below, it will be useful to admit functions Φ which are not finite. In that case we will assume that $b < \bar{u}$ to avoid trivialities.

Problems of this type arise in agency models of moral hazard, when the principal wishes to induce a prescribed agent's action a^0 at a minimum cost by application of a certain monitoring technology. The agent chooses an action a from some set $A \subset \mathbb{R}^n$ of possible actions, which is not directly observable by the principal. The observable signal is some $x \in M$, which is randomly disturbed and follows a distribution P_a on M (more precisely, P_a is a probability distribution on some suitable sigma-field $\mathscr{B}$ of subsets of M). The measurable space $(M, \mathscr{B})$ together with the family $(P_a)_{a \in A}$ is called an *information system* or a *statistical model*. Depending on the observed signal x, the agent receives a compensation $\sigma(x) \geq 0$ from the principal. Cost minimization reflects the principal's objective if he/she is risk-neutral, and thus his/her problem of choosing a utility maximizing pair (a, σ) can be separated into an action choice problem and a cost minimization problem (cf. Grossman & Hart (1983), p. 10 and Remark 3 on p. 17). Throughout we suppose that the action choice problem has been settled, i.e. the principal has decided a particular action a^0 to be most favourable. Then, the desired action $a^0 \in A$ is induced only if the principal provides a compensation function σ, such that a^0 maximizes the agent's expected utility, i.e.,

$$\mathrm{E}_{a^0}(U \circ \sigma) - V(a^0) = \max_{a \in A}\{\mathrm{E}_a(U \circ \sigma) - V(a)\}, \tag{1.2}$$

where E_a stands for taking expectation w.r.t. P_a, $U(s)$ is the utility caused by wealth $s \in [0, \infty)$, and $V(a)$ is the disutility caused by action $a \in A$. Condition (1.2) is called the *incentive compatibility constraint*. The principal wishes to minimize the expected cost, i.e., to choose a compensation cost function σ which minimizes $E_{a^0}(\sigma)$ subject to the constraint (1.2), and subject to some further constraints to be discussed later. Of course, constraint (1.2) is mathematically difficult to handle. The *first order approach* we will employ replaces (1.2) by the condition that a^0 is a stationary point of the agent's expected utility, where we assume that a^0 is an interior point of A and $E_a(U \circ \sigma)$ and $V(a)$ are smooth functions of a. That is, the weaker first order condition is employed,

$$\nabla_a E_a(U \circ \sigma)|_{a=a^0} = \nabla_a V(a)|_{a=a^0}, \tag{1.3}$$

where ∇_a stands for taking the gradient (vector of partial derivatives) w.r.t. a. In general, (1.3) is a relaxation of (1.2) which is used for mathematical tractability. *Equivalence* of (1.2) and (1.3) will require additional assumptions, which will essentially have to ensure that the expected utility on the right hand side of (1.2) is a concave function of a, if σ is the optimal solution to the relaxed problem. For example, in the standard agency model as in Kim (1995) this is met under the *monotone likelihood ratio* and the (very restrictive) *convexity of distribution function* conditions on the information system as in Rogerson (1985), and the monotonicity and convexity of the disutility function V (see our remark after Corollary 3.3 in Section 3). Alternative and less restrictive conditions have been proposed by Jewitt (1988).

Assume that the family P_a of distributions is smooth at a^0, in the sense that for all f from a suitable class of real valued measurable functions on M (to be specified later) we have

$$\nabla_a E_a(f)|_{a=a^0} = E_{a^0}(f \cdot S_{a^0}),$$

with a fixed $\mathbb{R}^n$-valued measurable function S_{a^0} on M, which is called the *score function* of the family $(P_a)_{a \in A}$ at a^0 (some authors call it the *likelihood ratio*). Usually the score function is given by

$$S_{a^0}(x) = \frac{1}{p_{a^0}(x)} \nabla_a p_a(x)|_{a=a^0}, \quad x \in M,$$

where p_a is a density of P_a for all $a \in A$ w.r.t. some fixed sigma-finite measure Λ on $(M, \mathscr{B})$. Now, substituting $f = U \circ \sigma$, the first order condition (1.3) rewrites as (1.1d), where

$$(h_1, \ldots, h_n) = S_{a^0}, \tag{1.4a}$$

$$(\beta_1, \ldots, \beta_n) = \nabla_a V(a)|_{a=a^0}, \quad \text{and} \quad P = P_{a^0}. \tag{1.4b}$$

Under the mild assumption that the utility function U is defined on $[0, \infty)$ with $U(0) = 0$, and as usual, that U is strictly increasing and concave, we consider the inverse function $\Phi = U^{-1}$, where in case that U is bounded, i.e., $\bar{u} = \lim_{s\to\infty} U(s) < \infty$, we define $\Phi(u) = \infty$ for all $u \geq \bar{u}$. Thus, Φ is a convex function on $[0, \infty)$ with values in $[0, \infty]$ and $\Phi(0) = 0$. The problem of minimizing the expected cost subject to the first order condition (1.3) now rewrites as (1.1a) subject to (1.1d).

It may be necessary to take into account further restrictions like (1.1b) and (1.1c). The restriction (1.1b) may express e.g. a limited liability of the agent. Also, there might be reasons (other than boundedness of the utility function U) for bounding f from above by some prescribed constant $c > b$, e.g. due to a limitation of the principal's budget. In that case the constraint $f \leq c$ will be included implicitly by choosing Φ such that $\Phi(u) = \infty$ for all $u > c$. In addition, the restrictions (1.1c) may express a certain reservation level of utility causing the agent to sign the contract. In particular, for $m = 1$ and $g_1 = 1$ we have the standard inequality constraint, which gives a guaranteed expected utility value α_1 to the agent. Restrictions of this type are usually referred to as the *participation constraint* or *individual rationality constraint*.

In Kim (1995), Section 2, problem (1.1a)–(1.1d) has been studied for $m = n = 1$, $g_1 = 1$, and h_1 being the score function (1.4a). The optimal solution was claimed to have a description via Lagrange multipliers. However, the variable f is from some infinite dimensional function space, and the Lagrange multiplier approach may not be valid. In Section 2 we will embed the general extremum problem (1.1a)–(1.1d) into an appropriate function space. In Section 3 a duality result will be derived which is close to the Lagrangian approach (but not identical). In fact, the Lagrangian approach is valid only if the dual problem as well has an optimal solution. The duality result turns out to be particularly useful to study the dependence of the minimum expected cost value on the input distribution P and the input functions g_i and h_j. In Section 4, we will show that the minimum expected cost value decreases when the input is changed such that the joint distribution of $g_1, \ldots, g_m, h_1, \ldots, h_n$, gets larger w.r.t. the *convex order* of probability distributions on the multi-dimensional Euclidean space. A generalization of Kim's result (Kim (1995), Proposition 1) is thereby obtained. The general results allow to compare different possible information systems w.r.t. their resulting minimum expected cost values. In Section 5 it is shown that Blackwell sufficiency (of one information system for another one) implies the convex order relation of the corresponding score function distributions, extending thus a result for the one-dimensional case of Kim (1995), Proposition 4. A further result refers to a rank order tournament considered by Budde (1997), which settles a conjecture in his paper concerning the expectation of the score function under the distribution of the maximum of n i.i.d. real valued random variables.

2. Preliminaries: Orlicz spaces

A linear space which appears quite natural as an embedding space for problem (1.1a)–(1.1d) is the Orlicz space given by the so-called Young's function Φ. A function Φ is called a Young's function (cf. Krasnoselskii & Rutickii (1961), p. 3 ff.), if and only if

$\Phi : [0, \infty) \to [0, \infty],$

$\Phi(0) = 0,$

Φ is neither constantly zero nor constantly infinity on $(0, \infty)$,

Φ is convex,

Φ is left continuous at $\bar{u} = \sup\{u \geq 0 : \Phi(u) < \infty\}$, if $\bar{u} < \infty$.

The *conjugate* function of the Young's function Φ is defined by

$$\Psi(v) = \sup_{u \in [0,\infty)} (uv - \Phi(u)), \quad v \in [0, \infty), \tag{2.1}$$

which is a Young's function as well. The conjugate function of Ψ yields back Φ, i.e.,

$$\Phi(u) = \sup_{v \in [0,\infty)} (vu - \Psi(v)), \quad u \in [0, \infty). \tag{2.2}$$

For the following basic facts about Orlicz spaces the reader is referred to Krasnoselskii & Rutickii (1961). For a Young's function Φ and a probability space $(M, \mathscr{B}, P)$, the Orlicz space $L_\Phi(P)$ consists of all measurable real valued functions f on M such that $\mathrm{E}_P(\Phi \circ (r|f|)) < \infty$ for some positive real number r, where $|f|$ denotes the function $|f(x)|$, $x \in M$. As it is easily seen, $L_\Phi(P)$ is a real vector space, and

$$L^\infty(P) \subset L_\Phi(P) \subset L^1(P),$$

where $L^\infty(P)$ and $L^1(P)$ denote the space of all P-almost surely (P-a.s.) bounded measurable real valued functions on M and the space of all P-integrable real valued functions on M, respectively. Note that $L^\infty(P)$ and $L^1(P)$ are the Orlicz spaces corresponding to the Young's functions

$$\Phi_\infty(u) = \begin{cases} 0, & \text{if } u \leq 1 \\ \infty, & \text{if } u > 1 \end{cases}, \quad \Phi_1(u) = u, \quad u \in [0, \infty),$$

respectively, and Φ_∞ and Φ_1 are conjugate to each other. Moreover, for any non-finite Young's function Φ one has $L_\Phi(P) = L^\infty(P)$ and $L_\Psi(P) = L^1(P)$, where Ψ is the conjugate Young's function of Φ. Further well-known examples of Orlicz-spaces are the $L^p(P)$-spaces, for a given $p \in (1, \infty)$, corresponding to the Young's function $\Phi(u) = u^p$. The conjugate Young's function is given by $\Psi(v) = cv^q$, where $q = p/(p-1)$ and $c = (p-1)^{-q/p} q^{-q}$, and hence $L_\Psi(P) = L^q(P)$.

Let Φ be an arbitrary Young's function. A norm on $L_\Phi(P)$ is given by

$$\|f\|_\Phi = \inf\{k > 0 : \mathrm{E}_P(\Phi \circ (|f|/k)) \leq 1\}, \quad f \in L_\Phi(P), \tag{2.3}$$

where here and in the following two functions on M are viewed to be identical, if they coincide P-a.s., that is on a set of probability one. Together with this norm the space $L_\Phi(P)$ is a Banach space. Consider also the *conjugate* Orlicz space $L_\Psi(P)$, where Ψ is the conjugate function of Φ. Then,

$$f \in L_\Phi(P) \text{ and } g \in L_\Psi(P) \text{ imply } f \cdot g \in L^1(P), \quad \text{and}$$

$$|\mathrm{E}_P(f \cdot g)| \le \|f\|_\Phi \cdot \|g\|_\Psi.$$

Hence, any $f \in L_\Phi(P)$ defines a linear continuous real functional $f^\#$ on $L_\Psi(P)$ via

$$f^\# : \begin{cases} L_\Psi(P) & \to & \mathbb{R} \\ g & \to & \mathrm{E}_P(f \cdot g). \end{cases} \tag{2.4}$$

Thus, the Orlicz space $L_\Phi(P)$ is embedded into the dual space $L_\Psi(P)^\#$ of $L_\Psi(P)$ (consisting of all linear continuous real functionals on $L_\Psi(P)$).

An important question is whether *all* elements of the dual space $L_\Psi(P)^\#$ are obtained via (2.4), i.e., $L_\Psi(P)^\# \cong L_\Phi(P)$. The answer depends on the Young's functions Φ and Ψ, respectively. For example, if Ψ is not finite, or if Ψ is finite but increases too rapidly, then $L_\Psi(P)^\# \ncong L_\Phi(P)$. The Young's function Ψ is said to satisfy the Δ_2-condition, if and only if Ψ is finite and

$$\Psi(2v) \le k_0 \Psi(v) \quad \forall v \ge r_0, \tag{2.5}$$

for some real positive constants k_0 and r_0.

As it is easily seen, if Ψ satisfies (2.5), then $\mathrm{E}_P(\Psi \circ |g|) < \infty$ for all $g \in L_\Psi(P)$. The following result is due to Luxemburg & Zaanen (1956) and Rao (1968).

Theorem 2.1. *If Ψ satisfies the Δ_2-condition, then*

$$L_\Psi(P)^\# \cong L_\Phi(P),$$

i.e., any linear continuous real functional ℓ on $L_\Psi(P)$ is of the form

$$\ell(g) = \mathrm{E}_P(f \cdot g), g \in L_\Psi(P),$$

for some $f \in L_\Phi(P)$.

3. A duality result

Consider the minimization problem (1.1a)–(1.1d) in the Orlicz space $L_\Phi(P)$, i.e., the variable f is restricted to this space. Let Ψ be the conjugate Young's function of Φ given by (2.1). We assume that g_i (for $i = 1, \dots, m$) and h_j (for $j = 1, \dots, n$) are from the conjugate Orlicz space $L_\Psi(P)$. Denote $G = (g_1, \dots, g_m)^t$ and $H = (h_1, \dots, h_n)^t$ the $\mathbb{R}^m$-valued and $\mathbb{R}^n$-valued functions formed by the real valued functions g_i and h_j, respectively, and $\alpha = (\alpha_1, \dots, \alpha_m)^t$ and $\beta = (\beta_1, \dots, \beta_n)^t$, where the superscript t denotes transposition. Recall that the lower bound b in (1.1b) is assumed to satisfy $0 \le b < \bar{u}$, where $\bar{u} = \sup\{u \geq 0 : \Phi(u) < \infty\}$. By Ψ_b we denote the function

$$\Psi_b(v) = \sup_{u \in [b,\infty)} (uv - \Phi(u)), \quad v \in \mathbb{R}, \tag{3.1}$$

which is a convex function on $\mathbb{R}$ with values in $\mathbb{R} \cup \{\infty\}$, and obviously,

$$bv - \Phi(b) \leq \Psi_b(v) \leq \Psi(\max\{v, 0\}) \quad \forall v \in \mathbb{R}. \tag{3.2}$$

Now we state as dual problem,

$$\text{maximize} \quad D(\lambda, \mu) = \lambda^t \alpha + \mu^t \beta - \mathrm{E}_P(\Psi_b \circ (\lambda^t G + \mu^t H)) \tag{3.3a}$$

$$\text{s.t.} \quad \lambda \in \mathbb{R}^m, \lambda \geq 0 \text{ (componentwise)}, \quad \mu \in \mathbb{R}^n. \tag{3.3b}$$

Note that the expectation on the right hand side of (3.3a) exists (but may be equal to ∞), since for any $g \in L_\Psi(P)$ we have by (3.2), $\Psi_b \circ g \geq bg - \Phi(b)$, and the lower bound is a P-integrable function. Thus, the objective function D in (3.3a) is a concave function on $\mathbb{R}^{m+n}$ with values in $\mathbb{R} \cup \{-\infty\}$.

We now state our first *weak duality* result.

Lemma 3.1. *Assume that* $g_i, h_j \in L_\Psi(P)$ *for all* $i = 1, \ldots, m$ *and* $j = 1, \ldots, n$. *If* $f \in L_\Phi(P)$ *is a feasible solution to* (1.1a)–(1.1d) *and* (λ, μ) *is a feasible solution to* (3.3a)–(3.3b), *then*

$$\mathrm{E}_P(\Phi \circ f) \geq D(\lambda, \mu).$$

Hence, denoting by ϕ^* *the infimum of the objective function in problem* (1.1a)–(1.1d) *(where the infimum over the empty set is defined to be infinity), and by* δ^* *the supremum of the objective function in problem* (3.3a)–(3.3b), *we have*

$$\phi^* \geq \delta^*.$$

Proof. By (1.1c), (1.1d), and (3.3b),

$$\lambda^t \alpha + \mu^t \beta \leq \mathrm{E}_P(f \cdot (\lambda^t G + \mu^t H)),$$

and hence

$$D(\lambda, \mu) \leq \mathrm{E}_P(f \cdot (\lambda^t G + \mu^t H) - \Psi_b \circ (\lambda^t G + \mu^t H)).$$

By (3.1), for all $u \in [b, \infty)$ and all $v \in \mathbb{R}$,

$$uv - \Psi_b(v) \leq \Phi(u),$$

and thus, by (1.1b), for all $x \in M$,

$$f(x) \cdot (\lambda^t G(x) + \mu^t H(x)) - \Psi_b(\lambda^t G(x) + \mu^t H(x)) \leq \Phi(f(x)).$$

From this we obtain

$$D(\lambda, \mu) \leq \mathrm{E}_P(\Phi \circ f).$$ □

Our *strong duality* result we state next requires Ψ to satisfy the Δ_2-condition (2.5). Recall the notations ϕ^* and δ^* for the optimal values of problem (1.1a)–(1.1d) and problem (3.3a)–(3.3b), respectively.

Theorem 3.2. *Assume that Ψ satisfies the Δ_2-condition, and $g_i, h_j \in L_\Psi(P)$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$. Then, $\phi^* = \delta^*$. If this value is finite, then there exists an optimal solution to problem* (1.1a)–(1.1d).

Proof. If $\delta^* = \infty$, then $\phi^* = \infty$ by Lemma 3.1. Now let $\delta^* < \infty$. Denote

$$\mathscr{D} = \{\lambda^t G + \mu^t H : \lambda \in \mathbb{R}^m, \lambda \geq 0, \mu \in \mathbb{R}^n\}, \tag{3.4a}$$

which is a convex subset of $L_\Psi(P)$, and for $d \in \mathscr{D}$ let

$$\ell(d) = \sup\{\lambda^t \alpha + \mu^t \beta : \lambda \in \mathbb{R}^m,\ \lambda \geq 0,\ \mu \in \mathbb{R}^n,\ \lambda^t G + \mu^t H = d\}. \tag{3.4b}$$

Clearly, $\ell(d) < \infty$, since otherwise we would have $\delta^* = \infty$, where it may be noted that $\mathrm{E}_P(\Psi_b \circ d) \leq \mathrm{E}_P(\Psi \circ |d|) < \infty$ by (3.2). Hence, (3.4b) defines a real valued function ℓ on $\mathscr{D}$. We will show that ℓ is concave. To this end let $d, \bar{d} \in \mathscr{D}$ and $\gamma \in (0,1)$. For a given $\varepsilon > 0$ one can find $\lambda, \bar{\lambda} \in \mathbb{R}^m$, $\lambda \geq 0$, $\bar{\lambda} \geq 0$, $\mu, \bar{\mu} \in \mathbb{R}^n$, such that

$$\lambda^t G + \mu^t H = d, \quad \lambda^t \alpha + \mu^t \beta \geq \ell(d) - \varepsilon,$$

$$\bar{\lambda}^t G + \bar{\mu}^t H = \bar{d}, \quad \bar{\lambda}^t \alpha + \bar{\mu}^t \beta \geq \ell(\bar{d}) - \varepsilon.$$

Hence, for $\lambda^* = \gamma\lambda + (1-\gamma)\bar{\lambda}$ and $\mu^* = \gamma\mu + (1-\gamma)\bar{\mu}$ we have $\lambda^* \geq 0$, $\mu^* \in \mathbb{R}^n$, $\lambda^{*t} G + \mu^{*t} H = \gamma d + (1-\gamma)\bar{d}$, and thus

$$\begin{aligned} \ell(\gamma d + (1-\gamma)\bar{d}) &\geq \lambda^{*t}\alpha + \mu^{*t}\beta \\ &= \gamma(\lambda^t\alpha + \mu^t\beta) + (1-\gamma)(\bar{\lambda}^t\alpha + \bar{\mu}^t\beta) \\ &\geq \gamma\ell(d) + (1-\gamma)\ell(\bar{d}) - \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this proves concavity of ℓ.

By F we denote the function on $L_\Psi(P)$,

$$F(g) = \mathrm{E}_P(\Psi_b \circ g), \quad g \in L_\Psi(P), \tag{3.5}$$

which is a real valued convex function, since by (3.2) $bg - \Phi(b) \leq \Psi_b \circ g \leq \Psi \circ |g|$ and hence $\Psi_b \circ g$ is P-integrable for all $g \in L_\Psi(P)$. Now problem (3.3a)–(3.3b) is equivalently stated as

$$\text{minimize} \quad F(d) - \ell(d) \tag{3.6a}$$

$$\text{s.t.} \quad d \in \mathscr{D}, \tag{3.6b}$$

Note that the infimum of (3.6a) subject to (3.6b) is equal to $-\delta^*$ which is finite. We will apply the Fenchel Duality Theorem from Luenberger (1969), Theorem 1, p. 201, which states that under certain regularity assumptions to be checked below we have

$$-\delta^* = \max\{\ell^*(f) - F^*(f) : f \in \mathscr{C}^* \cap \mathscr{D}^*\}, \tag{3.7}$$

and the maximum on the right hand side of (3.7) is attained. The sets $\mathscr{C}^*$, $\mathscr{D}^*$, and the real valued functions F^* on $\mathscr{C}^*$ and ℓ^* on $\mathscr{D}^*$ are defined by

$$\mathscr{C}^* = \left\{ f \in L_\Phi(P) : \sup_{g \in L_\Psi(P)} (\mathrm{E}_P(f \cdot g) - F(g)) < \infty \right\}, \tag{3.8a}$$

$$F^*(f) = \sup_{g \in L_\Psi(P)} (\mathrm{E}_P(f \cdot g) - F(g)) \quad \forall f \in \mathscr{C}^*, \tag{3.8b}$$

$$\mathscr{D}^* = \left\{ f \in L_\Phi(P) : \inf_{d \in \mathscr{D}} (\mathrm{E}_P(f \cdot d) - \ell(d)) > -\infty \right\}, \tag{3.8c}$$

$$\ell^*(f) = \inf_{d \in \mathscr{D}} (\mathrm{E}_P(f \cdot d) - \ell(d)) \quad \forall f \in \mathscr{D}^*. \tag{3.8d}$$

Note that we have used that by Theorem 2.1 the dual space of $L_\Psi(P)$ coincides with $L_\Phi(P)$. To describe the domains and functions in (3.8a)–(3.8d) more explicitly, we firstly observe that for any $f \in L_\Phi(P)$,

$$\sup_{g \in L_\Psi(P)} [\mathrm{E}_P(f \cdot g) - \mathrm{E}_P(\Psi_b \circ g)] = \begin{cases} \mathrm{E}_P(\Phi \circ f), & \text{if } f \geq b \\ \infty, & \text{otherwise} \end{cases}, \tag{3.9}$$

the proof of which will be given later. Hence, by (3.5), (3.8a), and (3.8b),

$$\mathscr{C}^* = \{f \in L_\Phi(P) : f \geq b, \mathrm{E}_P(\Phi \circ f) < \infty\},$$

$$F^*(f) = \mathrm{E}_P(\Phi \circ f), \quad f \in \mathscr{C}^*.$$

By (3.4a)–(3.4b), we have for any $f \in L_\Phi(P)$,

$$\begin{aligned} &\inf_{d \in \mathscr{D}} (\mathrm{E}_P(f \cdot d) - \ell(d)) \\ &\quad = \inf \left\{ \sum_{i=1}^m \lambda_i (\mathrm{E}_P(f \cdot g_i) - \alpha_i) + \sum_{j=1}^n \mu_j (\mathrm{E}_P(f \cdot h_j) - \beta_j) : \lambda_i \geq 0, \mu_j \in \mathbb{R} \right\} \\ &\quad = \begin{cases} 0, & \text{if } \mathrm{E}_P(f \cdot g_i) \geq \alpha_i \ \forall i \text{ and } \mathrm{E}_P(f \cdot h_j) = \beta_j \forall j, \\ -\infty, & \text{otherwise} \end{cases}. \end{aligned}$$

We have thus obtained,

$$\mathscr{D}^* = \{f \in L_\Phi(P) : \mathrm{E}_P(f \cdot g_i) \geq \alpha_i \ \forall i \text{ and } \mathrm{E}_P(f \cdot h_j) = \beta_j \ \forall j\},$$

$$\ell^*(f) = 0, \quad f \in \mathscr{D}^*.$$

Thus, (3.7) says that

$$-\delta^* = \max\{ - \mathrm{E}_P(\Phi \circ f) : f \in L_\Phi(P),\ f \geq b,$$
$$\mathrm{E}_P(f \cdot g_i) \geq \alpha_i\ \forall i,\ \mathrm{E}_P(f \cdot h_j) = \beta_j\ \forall j\},$$

which is obviously the same as asserted by the theorem.

We will check the regularity assumptions for the duality theorem, namely that $L_\Psi(P) \cap \mathscr{D}$ contains points in the relative interior of $L_\Psi(P)$ and $\mathscr{D}$ (which is trivially true by $\mathscr{D} \subset L_\Psi(P)$), and that the epigraph of F,

$$\{(g, r) \in L_\Psi(P) \times \mathbb{R} : F(g) \leq r\}$$

has nonempty interior. In fact, an interior point of the epigraph is given by $g_0 = 0$, $r_0 = 2$, since

$$F(0) = \Psi_b(0) \leq 0,$$

and for any $g \in L_\Psi(P)$ with $\|g\|_\Psi < 1$ and any $r \in [1, \infty)$ we have by (3.2)

$$F(g) = \mathrm{E}_P(\Psi_b \circ g) \leq \mathrm{E}_P(\Psi \circ |g|) \leq 1 \leq r.$$

It remains to prove (3.9). For a given $f \in L_\Phi(P)$ we denote by $\gamma(f)$ the supremum on the left hand side of (3.9).

Case 1: $f \geq b$.
By (3.1), for any $g \in L_\Psi(P)$, we have $f \cdot g - \Psi_b \circ g \leq \Phi \circ f$, and hence

$$\mathrm{E}_P(f \cdot g) - \mathrm{E}_P(\Psi_b \circ g) \leq \mathrm{E}_P(\Phi \circ f),$$

showing that $\gamma(f) \leq \mathrm{E}_P(\Phi \circ f)$. To prove the reverse inequality we observe that by (3.2), $\Psi_b(v) \leq \Psi(|v|)$ for all $v \in \mathbb{R}$, and hence

$$\gamma(f) \geq \sup_{g \in L_\Psi(P)} \{\mathrm{E}_P(f \cdot g) - \mathrm{E}_P(\Psi \circ |g|)\}. \tag{3.10}$$

Since f is nonnegative and $|g| \in L_\Psi(P)$ whenever $g \in L_\Psi(P)$, the supremum on the right hand side of (3.10) is not changed when we restrict to all nonnegative $g \in L_\Psi(P)$. As it is well known from measure theory, f can be written as a pointwise limit of an increasing sequence f_ν, $\nu \in \mathbb{N}$, of nonnegative measurable step functions. In particular, $0 \leq f_\nu \leq f$ for all ν, and hence the right hand side of (3.10) does not get larger when f is replaced by f_ν. Thus, for all ν,

$$\gamma(f) \geq \sup_{g \in L_\Psi(P), g \geq 0} \{\mathrm{E}_P(f_\nu \cdot g) - \mathrm{E}_P(\Psi \circ g)\}. \tag{3.11}$$

For any fixed ν, let $u_1, \ldots, u_k$ be the distinct values of f_ν attained on pairwise disjoint measurable subsets $B_1, \ldots, B_k$ of M. Of course, k, $u_1, \ldots, u_k$, and $B_1, \ldots, B_k$ will depend on ν, which is dropped for notational simplicity. Clearly, the right hand side of (3.11) does not get larger when we restrict to nonnegative step functions g which are constant on each B_i, $i = 1, \ldots, k$. We thus obtain

$$\gamma(f) \geq \sup_{v_1,\ldots,v_k \geq 0} \left\{ \sum_{i=1}^{k} u_i v_i P(B_i) - \sum_{i=1}^{k} \Psi(v_i) P(B_i) \right\}$$

$$= \sum_{i=1}^{k} \sup_{v_i \geq 0} \{u_i v_i - \Psi(v_i)\} P(B_i) = \sum_{i=1}^{k} \Phi(u_i) P(B_i) = \mathrm{E}_P(\Phi \circ f_\nu).$$

By the monotone convergence theorem, $\lim_{\nu\to\infty} \mathrm{E}_P(\Phi \circ f_\nu) = \mathrm{E}_P(\Phi \circ f)$, and hence $\gamma(f) \geq \mathrm{E}_P(\Phi \circ f)$.

Case 2: $P(\{f < b\}) > 0$.

Choose an $\varepsilon > 0$ such that $P(\{f \leq b - \varepsilon\}) > 0$. For an arbitrary $v < 0$ consider the step function g_v which is constantly equal to v on the set $B = \{f \leq b - \varepsilon\}$ and zero outside B. By (3.1), $\Psi_b(v) \leq bv$, and we obtain

$$\gamma(f) \geq \mathrm{E}_P(f \cdot g_v) - \mathrm{E}_P(\Psi_b \circ g_v) \geq (b - \varepsilon) v P(B) - bvP(B) = -\varepsilon v P(B),$$

and the last lower bound tends to infinity when $v \to -\infty$. Hence $\gamma(f) = \infty$. □

The interplay of an optimal solution f^* to problem (1.1a)–(1.1d) and an optimal solution (λ^*, μ^*) to problem (3.3a)–(3.3b) (if there exists any) is illucidated by the following corollary. A Lagrange multiplier description of f^* is thereby obtained, as employed in Kim (1995), Section 2, for the case $m = n = 1$, $g_1 = 1$. Recall that $0 \leq b < \bar{u} = \sup\{u \geq 0 : \Phi(u) < \infty\}$.

Corollary 3.3. *As in* Theorem 3.2 *assume that* Ψ *satisfies the* Δ_2*-condition, and* $g_i, h_j \in L_\Psi(P)$ *for all* $i = 1, \ldots, m$ *and* $j = 1, \ldots, n$. *Assume further that* Φ *restricted to the interval* $[b, \bar{u})$ *is differentiable with strictly increasing derivative* $\varphi(u)$, $u \in [b, \bar{u})$, *and denote* $\varphi(\bar{u}) = \lim_{u \uparrow \bar{u}} \varphi(u)$. *If* (λ^*, μ^*) *is an optimal solution to problem* (3.3a)–(3.3b), *then the optimal solution* f^* *to problem* (1.1a)–(1.1d) *is given by*

$$f^*(x) = \begin{cases} b, & \text{if } d^*(x) \leq \varphi(b) \\ \varphi^{-1}(d^*(x)), & \text{if } \varphi(b) < d^*(x) < \varphi(\bar{u}), \\ \bar{u}, & \text{if } d^*(x) \geq \varphi(\bar{u}) \end{cases} \quad (x \in M),$$

where $d^* = \sum_{i=1}^{m} \lambda_i^* g_i + \sum_{j=1}^{n} \mu_j^* h_j$.

Moreover, for all $i \in \{1, \ldots, m\}$ *with* $\lambda_i^* > 0$, *we have* $\mathrm{E}_P(f^* \cdot g_i) = \alpha_i$.

Proof. Clearly, $\delta^* = D(\mu^*, \lambda^*) < \infty$. By Theorem 3.2, an optimal solution f^* to problem (1.1a)–(1.1d) (which exists by the theorem) satisfies

$$\mathrm{E}_P(\Phi \circ f^*) = D(\lambda^*, \mu^*).$$

Hence all the inequalities in the proof of Lemma 3.1 have to be equalities for the pair f^* and (λ^*, μ^*), from which we see that

$$\mathrm{E}_P(f^* \cdot g_i) = \alpha_i \quad \text{whenever } \lambda_i^* > 0,$$

and for P-almost every $x \in M$,

$$f^*(x) \cdot d^*(x) - \Psi_b(d^*(x)) = \Phi(f^*(x)).$$

The latter implies, in view of (3.1), that for P-almost every $x \in M$ the value $u^* = f^*(x)$ attains

$$\max_{u \in [b,\infty)} (d^*(x)u - \Phi(u)),$$

from which the result is easily obtained, observing that φ, as the derivative of a convex function, is continuous. □

Remark. Consider the standard agency model with one-dimensional actions as in Kim (1995), i.e., $m = n = 1$, $g_1 = 1$, h_1 the score function of the information system at a^0 from (1.4a), $\beta = \beta_1$ given by (1.4b), $P = P_{a^0}$, and the set A of possible actions is an interval of the real line (not necessarily bounded). Consider the optimal solution f^* given by Corollary 3.3 (under the assumptions of that corollary). We have $\mu^* \geq 0$, provided that $\beta > 0$ (i.e., the derivative of the disutility function V at a^0 is positive which we assume in the following). To see this, suppose that $\mu^* < 0$. Then, trivially $\lambda^*\alpha + \mu^*\beta < \lambda^*\alpha$. By $\mathrm{E}_P(h_1) = 0$ and by Jensen's inequality we have

$$\mathrm{E}_P(\Psi_b \circ (\lambda^* + \mu^* h_1)) \geq \Psi_b(\lambda^*).$$

We thus obtain $D(\lambda^*, \mu^*) < D(\lambda^*, 0)$, contradicting the optimality of (λ^*, μ^*).

Now assume that the signals x are real valued (and we choose $M = \mathbb{R}$ and $\mathscr{B}$ the Borel sigma field in $\mathbb{R}$), and that the score function h_1 is an increasing function, which is met if the distribution family P_a, $a \in A$, has the monotone likelihood ratio (MLR) property. Then, by $\mu^* \geq 0$, f^* is an increasing function.

Assume further that the distribution family satisfies the *convexity of distribution function* (CDF) condition, i.e., the distribution functions F_a of P_a, $a \in A$, are such that $F_a(x)$ is a convex function of $a \in A$ for any fixed $x \in \mathbb{R}$. This is the same as saying that the distribution family is *concave in the usual stochastic order*. Recall that the usual *stochastic order* relation of probability distributions on the real line is defined by

$$\tilde{Q} \leq_{\mathrm{st}} Q \iff \tilde{F}(x) \geq F(x) \ \forall x \in \mathbb{R},$$

where $\tilde{Q}$ and Q are any two probability distributions on the real line, and $\tilde{F}$ and F denote their distribution functions, respectively. Also, $\tilde{Q} \leq_{\mathrm{st}} Q$ holds true if and only if

$$\mathrm{E}_{\tilde{Q}}(f) \leq \mathrm{E}_Q(f)$$

for all increasing real valued functions f on $\mathbb{R}$ for which the expectations exist (cf., e.g., Shaked & Shantikumar (1994), pp. 3–4). Hence, the CDF condition says that for any $a_1, a_2 \in A$ and any $\gamma \in (0,1)$,

$$\gamma F_{a_1}(x) + (1-\gamma)F_{a_2}(x) \geq F_{\gamma a_1+(1-\gamma)a_2}(x) \quad \forall x \in \mathbb{R}, \text{ i.e.,}$$

$$\gamma P_{a_1} + (1-\gamma)P_{a_2} \leq_{\mathrm{st}} P_{\gamma a_1+(1-\gamma)a_2},$$

that is, the mapping $a \to P_a$ is concave w.r.t. the stochastic order of probability distributions. The CDF condition is extremely restrictive, and is not met by the popular distribution families, as pointed out previously by Jewitt (1988). Anyway, since the optimal solution f^* is an increasing function, the CDF condition implies that the agent's expected utility $\mathrm{E}_{P_a}(f^*)$ is a concave function of $a \in A$, and hence, assuming convexity of the disutility function V, we conclude that the expected utility

$$\mathrm{E}_{P_a}(f^*) - V(a)$$

is a concave function of $a \in A$. Hence we have obtained as in Rogerson (1985), that if the information system satisfies the MLR and the CDF conditions, then the first order approach is valid.

4. Monotonicity of the optimum value w.r.t. the convex order of distributions

Let Q and $\tilde{Q}$ be two probability distributions on (the Borel sigma-field of) $\mathbb{R}^N$. We say that $\tilde{Q}$ is smaller than Q w.r.t. the convex order, abbreviated as

$$\tilde{Q} \leq_{\mathrm{cx}} Q,$$

if and only if

$$\mathrm{E}_{\tilde{Q}}(C) \leq \mathrm{E}_Q(C)$$

for all real valued convex functions C on $\mathbb{R}^N$ for which the expectations w.r.t. $\tilde{Q}$ and Q exist (cf., e.g., Shaked & Shantikumar (1994), p. 154, Eq. 5.A.4). In the one-dimensional case $N = 1$, an equivalent definition of $\tilde{Q} \leq_{\mathrm{cx}} Q$ is that

$$\int_{-\infty}^{z} \tilde{F}(u)\,\mathrm{d}u \leq \int_{-\infty}^{z} F(u)\,\mathrm{d}u \quad \text{for all } z \in \mathbb{R},$$

$$\text{and} \quad \mathrm{E}_{\tilde{Q}}(I_{\mathbb{R}}) = \mathrm{E}_Q(I_{\mathbb{R}}) \quad \text{if the expectations exist,}$$

where F and $\tilde{F}$ denote the distribution functions of Q and $\tilde{Q}$, respectively, and $I_{\mathbb{R}}$ denotes the identity function on $\mathbb{R}$, i.e., $I_{\mathbb{R}}(u) = u$ for all $u \in \mathbb{R}$, (cf., e.g., Shaked & Shantikumar (1994), p. 57, Theorem 2.A.1). In this case ($N = 1$), the convex order relation $\tilde{Q} \leq_{\mathrm{cx}} Q$ is also referred to as Q being a *mean preserving spread of* $\tilde{Q}$ (cf. Kim (1995), p. 93, or Rothschild & Stiglitz (1970), p. 230f.).

We will consider the dependence of the minimum value ϕ^* of problem (1.1a)–(1.1d) on the joint distribution of the input functions $g_1, \ldots, g_m$, $h_1, \ldots, h_n$,

$$Q = (g_1, \ldots, g_m, h_1, \ldots, h_n)(P), \tag{4.1}$$

which is a probability distribution on $\mathbb{R}^{m+n}$. Note that the input constants $b, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n$ are kept fixed (as well as the function Φ). We will restrict to situations when the strong duality $\phi^* = \delta^*$ holds true as it is ensured by Theorem 3.2. Thus we are led to study the dependence of the optimum value δ^* of problem (3.3a)–(3.3b) on the distribution Q from (4.1). In fact, δ^* is a function of Q (we will thus write $\delta^*(Q)$), since

$$\delta^*(Q) = \sup\{\lambda^t\alpha + \mu^t\beta - \mathrm{E}_Q(\Psi_b \circ \ell_{\lambda,\mu}) : \lambda \in \mathbb{R}^m, \lambda \geq 0, \mu \in \mathbb{R}^n\}, \quad (4.2a)$$

where for any given $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^n$ we have denoted by $\ell_{\lambda,\mu}$ the linear function on $\mathbb{R}^{m+n}$ given by

$$\ell_{\lambda,\mu}(z) = \lambda^t z^{(1)} + \mu^t z^{(2)},$$
$$\text{for all } z = \begin{pmatrix} z^{(1)} \\ z^{(2)} \end{pmatrix} \in \mathbb{R}^{m+n}, \text{ where } z^{(1)} \in \mathbb{R}^m, z^{(2)} \in \mathbb{R}^n. \quad (4.2b)$$

Theorem 4.1. *Consider two probability spaces $(M, \mathscr{B}, P)$ and $(\tilde{M}, \tilde{\mathscr{B}}, \tilde{P})$ with measurable real valued functions $g_1, \ldots, g_m, h_1, \ldots, h_n$ on M and $\tilde{g}_1, \ldots, \tilde{g}_m, \tilde{h}_1, \ldots, \tilde{h}_n$ on $\tilde{M}$. Let Q be given by* (4.1) *and let $\tilde{Q}$ be defined analogously (as the joint distribution of the functions $\tilde{g}_i, \tilde{h}_j$, $1 \leq i \leq m, 1 \leq j \leq n$, under $\tilde{P}$). Assume that Ψ is finite, and $g_i, h_j \in L_\Psi(P)$ for all $i = 1, \ldots, m$, $j = 1, \ldots, n$. If $\tilde{Q} \leq_{\mathrm{cx}} Q$, then $\tilde{g}_i, \tilde{h}_j \in L_\Psi(\tilde{P})$ for all $i = 1, \ldots, m$, $j = 1, \ldots, n$, and*

$$\delta^*(\tilde{Q}) \geq \delta^*(Q).$$

Proof. For any given $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^n$, consider the linear function $\ell_{\lambda,\mu}$ on $\mathbb{R}^{m+n}$ given by (4.2b). Since $\Psi \circ |\ell_{\lambda,\mu}|$ is a real valued nonnegative convex function on $\mathbb{R}^{m+n}$, the assumption $\tilde{Q} \leq_{\mathrm{cx}} Q$ implies

$$\begin{aligned}\mathrm{E}_{\tilde{P}}(\Psi \circ |\lambda^t\tilde{G} + \mu^t\tilde{H}|) &= \mathrm{E}_{\tilde{Q}}(\Psi \circ |\ell_{\lambda,\mu}|) \\ &\leq \mathrm{E}_Q(\Psi \circ |\ell_{\lambda,\mu}|) = \mathrm{E}_P(\Psi \circ |\lambda^t G + \mu^t H|),\end{aligned}$$

where we have denoted $G = (g_1, \ldots, g_m)^t$, $H = (h_1, \ldots, h_n)^t$, $\tilde{G} = (\tilde{g}_1, \ldots, \tilde{g}_m)^t$, and $\tilde{H} = (\tilde{h}_1, \ldots, \tilde{h}_n)^t$. By choosing suitable scalar multiples of unit vectors for λ and μ, we see that $\tilde{g}_i, \tilde{h}_j \in L_\Psi(\tilde{P})$ for all $i = 1, \ldots, m$, $j = 1, \ldots, n$. Similarly, for any given λ and μ, the function $\Psi_b \circ \ell_{\lambda,\mu}$ is a real valued convex function on $\mathbb{R}^{m+n}$, and thus $\tilde{Q} \leq_{\mathrm{cx}} Q$ implies

$$\mathrm{E}_{\tilde{Q}}(\Psi_b \circ \ell_{\lambda,\mu}) \leq \mathrm{E}_Q(\Psi_b \circ \ell_{\lambda,\mu}).$$

From (4.2a) and its analogue for $\delta^*(\tilde{Q})$ we conclude that $\delta^*(\tilde{Q}) \geq \delta^*(Q)$. □

In Kim (1995), Proposition 1, the case $m = n = 1$ with $g_1 = \tilde{g}_1 = 1$ was considered, and the result $\phi^*(Q) \leq \phi^*(\tilde{Q})$ was derived (in our notation). However, the proof is incomplete since the existence of Lagrange multipliers is tacitely assumed. We note that if $m = 1$ and $g_1 = \tilde{g}_1 = 1$, then the condition

$\tilde{Q} \leq_{\mathrm{cx}} Q$ is equivalent to

$$(\tilde{h}_1, \ldots, \tilde{h}_n)(\tilde{P}) \leq_{\mathrm{cx}} (h_1, \ldots, h_n)(P),$$

as it is easily seen.

5. Blackwell sufficiency and convex order of score function distributions

Let there be given two 'information systems' (or 'statistical models'),

$$(M, \mathscr{B}, (P_a)_{a \in A}) \quad \text{and} \quad (\tilde{M}, \tilde{\mathscr{B}}, (\tilde{P}_a)_{a \in A}),$$

where P_a and $\tilde{P}_a$ are probability distributions on the measurable spaces $(M, \mathscr{B})$ and $(\tilde{M}, \tilde{\mathscr{B}})$, respectively, and A is a nonempty subset of $\mathbb{R}^n$. Consider a fixed a^0 in the interior of A, and assume that the information systems are smooth at a^0 in the following sense. For any real valued, measurable, bounded functions f and $\tilde{f}$ on M and $\tilde{M}$, respectively, the expectations $\mathrm{E}_{P_a}(f)$ and $\mathrm{E}_{\tilde{P}_a}(\tilde{f})$ as functions of a are differentiable at a^0, and

$$\nabla_a \mathrm{E}_{P_a}(f)|_{a=a^0} = \mathrm{E}_{P_{a^0}}(f \cdot S_{a^0}), \quad \nabla_a \mathrm{E}_{\tilde{P}_a}(\tilde{f})|_{a=a^0} = \mathrm{E}_{\tilde{P}_{a^0}}(\tilde{f} \cdot \tilde{S}_{a^0}), \tag{5.1}$$

where S_{a^0} and $\tilde{S}_{a^0}$ are fixed P_{a^0}- and $\tilde{P}_{a^0}$-integrable $\mathbb{R}^n$-valued functions on M amd $\tilde{M}$, respectively (the *score functions* of the information systems at a^0). In particular, choosing $f = 1$ and $\tilde{f} = 1$, (5.1) implies that the score functions S_{a^0} and $\tilde{S}_{a^0}$ have expectations zero w.r.t. P_{a^0} and $\tilde{P}_{a^0}$, respectively. Usually, the score functions are given by

$$S_{a^0}(x) = \frac{1}{p_{a^0}(x)} \nabla_a p_a(x)|_{a=a^0}, \quad x \in M,$$

$$\tilde{S}_{a^0}(\tilde{x}) = \frac{1}{\tilde{p}_{a^0}(\tilde{x})} \nabla_a \tilde{p}_a(\tilde{x})|_{a=a^0}, \quad \tilde{x} \in \tilde{M},$$

where p_a and $\tilde{p}_a$ are densities of P_a and $\tilde{P}_a$, for all $a \in A$, w.r.t. some fixed sigma-finite measures Λ on $(M, \mathscr{B})$ and $\tilde{\Lambda}$ on $(\tilde{M}, \tilde{\mathscr{B}})$, respectively.

The information system $(M, \mathscr{B}, (P_a)_{a \in A})$ is said to be *Blackwell sufficient* for the information system $(\tilde{M}, \tilde{\mathscr{B}}, (\tilde{P}_a)_{a \in A})$, if and only if there exists a transition kernel (also called a stochastic kernel or a Markov kernel) $K = K(x, \tilde{B})$, $x \in M$, $\tilde{B} \in \tilde{\mathscr{B}}$, such that

$$\begin{aligned} \tilde{P}_a &= P_a K \quad \text{for all } a \in A, \text{ i.e.,} \\ \tilde{P}_a(\tilde{B}) &= \mathrm{E}_{P_a}(K(\cdot, \tilde{B})) \quad \text{for all } \tilde{B} \in \tilde{\mathscr{B}} \text{ and } a \in A. \end{aligned} \tag{5.2}$$

Note that by definition, a transition kernel K has the properties that for any fixed $x \in M$ the function $\tilde{B} \to K(x, \tilde{B})$ is a probability distribution on $(\tilde{M}, \tilde{\mathscr{B}})$, and for any fixed $\tilde{B} \in \tilde{\mathscr{B}}$ the function $x \to K(x, \tilde{B})$ is a measurable function on M (with values in the interval $[0, 1]$).

Theorem 5.1. *Let the information system* $(M, \mathscr{B}, (P_a)_{a \in A})$ *be Blackwell sufficient for the information system* $(\tilde{M}, \tilde{\mathscr{B}}, (\tilde{P}_a)_{a \in A})$. *Assume the smoothness condition* (5.1) *and consider the distributions of the score functions at* a^0 *under* P_{a^0} *and* $\tilde{P}_{a^0}$, *respectively, i.e.,*

$$Q_{a^0} = S_{a^0}(P_{a^0}) \quad \textit{and} \quad \tilde{Q}_{a^0} = \tilde{S}_{a^0}(\tilde{P}_{a^0}).$$

Then: $\tilde{Q}_{a^0} \leq_{\mathrm{cx}} Q_{a^0}$.

Proof. Consider the probability space $(M^*, \mathscr{B}^*, P^*)$, where $M^* = M \times \tilde{M}$ (the cartesian product), $\mathscr{B}^* = \mathscr{B} \otimes \tilde{\mathscr{B}}$ (the product sigma-field), and

$$P^* = P_{a^0} \otimes K, \quad \text{i.e.}$$

$$P^*(B^*) = \int_M K(x, B_2^*(x))\, \mathrm{d}P_{a^0}(x), \quad B^* \in \mathscr{B}^*, \tag{5.3}$$

$$\text{where } B_2^*(x) = \{\tilde{x} \in \tilde{M} : (x, \tilde{x}) \in B^*\} \quad \forall x \in M,$$

with a transition kernel K according to (5.2). From (5.3) and (5.2) it is easily seen that the marginal distributions of P^* are P_{a^0} and $\tilde{P}_{a^0}$. We consider the score functions as functions on M^*, i.e., we consider

$$S_{a^0}^*(x, \tilde{x}) = S_{a^0}(x), \quad \tilde{S}_{a^0}^*(x, \tilde{x}) = \tilde{S}_{a^0}(\tilde{x}), \quad \forall (x, \tilde{x}) \in M^*.$$

Obviously, we have

$$S_{a^0}^*(P^*) = Q_{a^0} \quad \text{and} \quad \tilde{S}_{a^0}^*(P^*) = \tilde{Q}_{a^0}.$$

Thus, by Rüschendorf (1981), Theorem 6, part (c), (or by Shaked & Shantikumar (1994), Theorem 5.A.1, p. 154), it suffices to prove that

$$\mathrm{E}_{P^*}(S_{a^0}^* | \tilde{S}_{a^0}^*) = \tilde{S}_{a^0}^* \quad P^*\text{-a.s..} \tag{5.4}$$

To this end we firstly note that (5.3) extends to any P^*-integrable function f^* on M^*,

$$\mathrm{E}_{P^*}(f^*) = \int_M \left(\int_{\tilde{M}} f^*(x, \cdot\,)\, \mathrm{d}K(x, \cdot\,) \right) \mathrm{d}P_{a^0}(x), \tag{5.5}$$

where $f^*(x, \cdot\,)$ denotes the function $\tilde{x} \to f^*(x, \tilde{x})$, and $K(x, \cdot\,)$ denotes the probability distribution $\tilde{B} \to K(x, \tilde{B})$, for any fixed $x \in M$. By definition, (5.4) means that

$$\mathrm{E}_{P^*}(S_{a^0}^* \cdot (\mathbb{1}_R \circ \tilde{S}_{a^0}^*)) = \mathrm{E}_{P^*}(\tilde{S}_{a^0}^* \cdot (\mathbb{1}_R \circ \tilde{S}_{a^0}^*),)$$

$$\text{for all Borel subsets } R \text{ of } \mathbb{R}^n, \tag{5.6}$$

where $\mathbb{1}_R$ denotes the function on $\mathbb{R}^n$ being one on R and zero outside R. By (5.5), the left hand side of (5.6) is equal to

$$\mathrm{E}_{P_{a^0}}(S_{a^0} \cdot K(\,\cdot\,, \tilde{B})), \quad \text{where } \tilde{B} = \{\tilde{x} \in \tilde{M} : \tilde{S}_{a^0}(\tilde{x}) \in R\}.$$

Since the second marginal of P^* is given by $\tilde{P}_{a^0}$, the right hand side of (5.6) is equal to

$$\mathrm{E}_{\tilde{P}_{a^0}}(\tilde{S}_{a^0} \cdot (\mathbb{1}_R \circ \tilde{S}_{a^0})) = \mathrm{E}_{\tilde{P}_{a^0}}(\tilde{S}_{a^0} \cdot \mathbb{1}_{\tilde{B}}),$$

with $\tilde{B}$ as defined above (for a given Borel subset R). Hence it suffices to show that

$$\mathrm{E}_{P_{a^0}}(S_{a^0} \cdot K(\,\cdot\,, \tilde{B})) = \mathrm{E}_{\tilde{P}_{a^0}}(\tilde{S}_{a^0} \cdot \mathbb{1}_{\tilde{B}}) \quad \text{for all } \tilde{B} \in \tilde{\mathscr{B}}. \tag{5.7}$$

Now, using the smoothness assumption (5.1) and the Blackwell condition (5.2), we conclude

$$\begin{aligned}
\mathrm{E}_{P_{a^0}}(S_{a^0} \cdot K(\,\cdot\,, \tilde{B})) &= \nabla_a \mathrm{E}_{P_a}(K(\,\cdot\,, \tilde{B}))|_{a=a^0} \\
&= \nabla_a \tilde{P}_a(\tilde{B})|_{a=a^0} \\
&= \nabla_a \mathrm{E}_{\tilde{P}_a}(\mathbb{1}_{\tilde{B}})|_{a=a^0} \\
&= \mathrm{E}_{\tilde{P}_{a^0}}(\mathbb{1}_{\tilde{B}} \cdot \tilde{S}_{a^0}),
\end{aligned}$$

and thus (5.7). □

In Budde (1997), the problem was studied of implementing a certain action a as a symmetric Nash-equilibrium in a multi-agent setting of moral hazard by a simple rank-order tournament at minimal cost. Adapting a model used by Green & Stokey (1983), the impact of different performance measures on total payments to the agents was analyzed. If only the best performing agent receives a bonus, the real valued observations $x_1, \ldots, x_n$ on the agents' actions are comprised to $x_{n:n} = \max_{i=1,\ldots,n} x_i$. Under the assumption that all agents choose the same action $a \in A \subset \mathbb{R}$ and the observations $x_1, \ldots, x_n$ are independent and identically distributed according to P_a, the distribution of $x_{n:n}$ is given by

$$P_{a,n:n} = \mathrm{Max}_n(P_a^{\otimes n}),$$

where Max_n denotes the function $\mathrm{Max}_n(x_1, \ldots, x_n) = x_{n:n}$ on $\mathbb{R}^n$, and $P_a^{\otimes n}$ is the n-fold product distribution of P_a. The comparison of the efficiencies of two competing information systems

$$(\mathbb{R}, \mathscr{B}, (P_a)_{a \in A}) \quad \text{and} \quad (\mathbb{R}, \mathscr{B}, (\tilde{P}_a)_{a \in A}), \tag{5.8}$$

(where $\mathscr{B}$ denotes the Borel sigma-field in $\mathbb{R}$) led under the first order approach to comparing

$$\mathrm{E}_{P_{a^0,n:n}}(S_{a^0}) \quad \text{and} \quad \mathrm{E}_{\tilde{P}_{a^0,n:n}}(\tilde{S}_{a^0}),$$

where S_{a^0} and $\tilde{S}_{a^0}$ denote the score functions at a^0 of the information systems (5.8). It was conjectured that if the Blackwell sufficiency (5.2) holds for the information systems (5.8) and the score functions S_{a^0} and $\tilde{S}_{a^0}$ (satisfying (5.1)) are increasing functions on some sets of P_{a^0}- and $\tilde{P}_{a^0}$-probability one, respectively, then

$$\mathrm{E}_{\tilde{P}_{a^0,n:n}}(\tilde{S}_{a^0}) \le \mathrm{E}_{P_{a^0,n:n}}(S_{a^0}), \tag{5.9}$$

(cf. Budde (1997), Conjecture 1). The following result settles that conjecture and moreover extends inequality (5.9) to

$$\mathrm{E}_{\tilde{P}_{a^0,n:n}}(C \circ \tilde{S}_{a^0}) \le \mathrm{E}_{P_{a^0,n:n}}(C \circ S_{a^0}), \tag{5.10}$$

for any increasing convex real valued function C on $\mathbb{R}$ for which the expectations in (5.10) exist. Note that (5.10) can be stated in terms of the distributions of S_{a^0} and $\tilde{S}_{a^0}$ under $P_{a^0,n:n}$ and $\tilde{P}_{a^0,n:n}$, respectively, as

$$\tilde{S}_{a^0}(\tilde{P}_{a^0,n:n}) \le_{\mathrm{icx}} S_{a^0}(P_{a^0,n:n}), \tag{5.11}$$

where $\le_{\mathrm{icx}}$ abbreviates the *increasing convex order* of probability distributions on $\mathbb{R}$. That is, for any two probability distributions Q and $\tilde{Q}$ on $\mathbb{R}$, we write $\tilde{Q} \le_{\mathrm{icx}} Q$, if and only if $\mathrm{E}_{\tilde{Q}}(C) \le \mathrm{E}_Q(C)$ for all increasing convex real valued functions C on $\mathbb{R}$ for which the expectations exist (cf., e.g., Shaked & Shantikumar (1994), p. 83). In fact, (5.11) (or, equivalently, (5.10)) is a consequence from $\tilde{S}_{a^0}(\tilde{P}_{a^0}) \le_{\mathrm{cx}} S_{a^0}(P_{a^0})$ (implied by Theorem 5.1) and the monotonicity of S_{a^0} and $\tilde{S}_{a^0}$, as the following lemma shows.

Lemma 5.2. *Let P and $\tilde{P}$ be two probability distributions on the Borel sigma-field of* $\mathbb{R}$, *and let S and $\tilde{S}$ be two real valued measurable functions on* $\mathbb{R}$, *such that S is increasing on some set of P-probability one, $\tilde{S}$ is increasing on some set of $\tilde{P}$-probability one, and $\tilde{S}(\tilde{P}) \le_{\mathrm{cx}} S(P)$. Denote*

$$P_{n:n} = \mathrm{Max}_n(P^{\otimes n}) \quad \textit{and} \quad \tilde{P}_{n:n} = \mathrm{Max}_n(\tilde{P}^{\otimes n}),$$

Then:

$$\tilde{S}(\tilde{P}_{n:n}) \le_{\mathrm{icx}} S(P_{n:n}).$$

Proof. We abbreviate $Q = S(P)$ and $\tilde{Q} = \tilde{S}(\tilde{P})$. The assumption $\tilde{Q} \le_{\mathrm{cx}} Q$ carries over to the n-fold product distributions (cf., e.g., Shaked & Shantikumar (1994), p. 155, Theorem 5.A.3), i.e.,

$$\tilde{Q}^{\otimes n} \le_{\mathrm{cx}} Q^{\otimes n}.$$

Let C be any increasing convex real valued function on $\mathbb{R}$. Then the composition $C \circ \mathrm{Max}_n$ is a convex real valued function on $\mathbb{R}^n$, and hence (provided that the following expectations exist),

$$\mathrm{E}_{\tilde{Q}^{\otimes n}}(C \circ \mathrm{Max}_n) \leq \mathrm{E}_{Q^{\otimes n}}(C \circ \mathrm{Max}_n). \tag{5.12a}$$

Now, $Q^{\otimes n}$ and $\tilde{Q}^{\otimes n}$ are the joint distributions of $S(x_1), \ldots, S(x_n)$ and of $\tilde{S}(x_1), \ldots, \tilde{S}(x_n)$ under $P^{\otimes n}$ and $\tilde{P}^{\otimes n}$, respectively. Thus, (5.12a) is the same as

$$\mathrm{E}_{\tilde{P}^{\otimes n}}(C \circ \tilde{S}_{n:n}) \leq \mathrm{E}_{P^{\otimes n}}(C \circ S_{n:n}), \tag{5.12b}$$

where we have denoted by $S_{n:n}$ and $\tilde{S}_{n:n}$ the functions on $\mathbb{R}^n$ defined by

$$S_{n:n}(x_1, \ldots, x_n) = \max_{i=1,\ldots,n} S(x_i), \quad \tilde{S}_{n:n}(x_1, \ldots, x_n) = \max_{i=1,\ldots,n} \tilde{S}(x_i).$$

By the monotonicity of S and $\tilde{S}$ on some sets of P- and $\tilde{P}$-probability one, we have

$$S_{n:n} = S \circ \mathrm{Max}_n \quad P^{\otimes n}\text{-a.s.}, \qquad \tilde{S}_{n:n} = \tilde{S} \circ \mathrm{Max}_n \quad \tilde{P}^{\otimes n}\text{-a.s.}.$$

Thus,

$$S_{n:n}(P^{\otimes n}) = S(P_{n:n}), \quad \tilde{S}_{n:n}(\widetilde{P}^{\otimes n}) = \tilde{S}(\widetilde{P}_{n:n}),$$

and (5.12b) rewrites as

$$\mathrm{E}_{\tilde{S}(\tilde{P}_{n:n})}(C) \leq \mathrm{E}_{S(P_{n:n})}(C). \tag{5.12c}$$

The proof is completed by observing that the expectations in (5.12c) exist if and only if the expectations in (5.12a) exist. □

6. Conclusions

The contribution of this paper is twofold. On the one hand, it embeds the principal's optimization problem in an agency model with imperfect monitoring into a more general cost minimization problem. Thereby the mere structure of the problem is emphasized. It becomes obvious that the findings could be applied to a wider range of situations where a decision maker has to determine a particular action choice depending on some observed signal.

On the other hand, regarding the standard agency setting, the paper generalizes a previous result by Kim (1995) concerning the ranking of information systems as monitoring devices. At this, a different characterization of the solution to the principal's optimization problem is given. As an immediate consequence, the relation of the minimum cost value and the applied information system becomes visible from the dual optimization problem. Furthermore, the generalized approach enables to make prediction about the efficiency of information systems even if one has to account for more constraints than has been done in previous studies. An obvious application of this extension is the case where the agent's action is multi-dimensional. So far, such multi-task agency models mainly have been analyzed in a specific framework where signals are normally distributed and the agent's preferences are described by an exponential utility function (e.g. Holmström & Milgrom (1991) and Feltham & Xie (1994)). Under certain assumptions about the timing of the agent's provision of effort, the optimal compensation $\sigma(x)$ is linear and closed form solutions to the principal's problem can be obtained (cf. Holmström & Milgrom (1987)). In contrast, the model presented in this paper allows for predictions on the ranking of information systems in a more general multi-task agency setting, provided the first order approach is valid. Limitations to these predictions can arise from the dimension of the performance measure. If an information system provides less linearly independent signals than the agent chooses actions, the feasible region of the cost minimization problem presented in this paper is likely to be empty. Therefore, different information systems may not be comparable in terms of compensation costs because they are not capable of inducing the same actions. However, global criteria like Blackwell sufficiency are not affected by this, since they hold for all relevant actions.

Acknowledgement: The authors wish to thank Ludger Rüschendorf for giving valuable hints to references.

References

Budde J (1997) Variance analysis as an incentive device when payments are based on rank order. Working paper, Faculty of Economics and Management, University of Magdeburg

Feltham GA, Xie J (1994) Performance measurement congruity and diversity in multi-task principal/agent relations. The Accounting Review 69:429–453

Green JR, Stokey NL (1983) A comparison of tournaments and contracts. Journal of Political Economy 91:349–364

Grossman SJ, Hart O (1983) An analysis of the principal-agent problem. Econometrica 51:7–45

Holmström B, Milgrom P (1987) Aggegation and linearity in the provision of intertemporal incentives. Econometrica 55:303–328

Holmström B, Milgrom P (1991) Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. Journal of Law, Economics and Organization 7:24–52

Jewitt I (1988) Justifying the first-order approach to principal-agent problems. Econometrica 56:1177–1190

Kim, S K (1995) Efficiency of an information system in an agency model. Econometrica 63:89–102

Krasnoselskii, Rutikii (1961) Convex functions and Orlicz spaces. Groningen

Luenberger DG (1969) Optimization by vector space methods. Wiley, New York

Luxemburg WAJ, Zaanen (1956) Conjugate spaces of Orlicz spaces. Indag. Math. 18:217–228

Rao MM (1968) Linear functionals on Orlicz spaces: General theory. Pacific J. Mathematics 25:553–585

Rogerson WP (1985) The first-order approach to principal-agent problems. Econometrica 53:1357–1367

Rothschild M, Stiglitz JE (1970) Increasing risk: a definition. Journal of Economic Theory 2:225–243

Rüschendorf L (1981) Orderings of distributions and rearrangement of functions. The Annals of Probability 9:276–283

Shaked M, Shantikumar JG (1994) Stochastic orders and their applications. Academic, London