On the optimal stopping values induced by general dependence structures

Alfred Müller

Institut für Wirtschaftstheorie und Operations Research,

Universität Karlsruhe, Geb. 20.21, D-76128 Karlsruhe, Germany

email: mueller@wior.uni-karlsruhe.de

Ludger Rüschendorf

Institut für Mathematische Stochastik, Universität Freiburg

Eckerstrasse 1, D-79104 Freiburg, Germany

email: ruschen@stochastik.uni-freiburg.de

Abstract

The optimal stopping value of random variables $X_1, \ldots, X_n$ depends on the joint distribution function of the random variables and hence on their marginals as well as on their dependence structure. The maximal and minimal values of the optimal stopping problem is determined within the class of all joint distributions with fixed marginals $F_1, \ldots, F_n$. They correspond to some sort of strong positive respectively negative dependence of the random variables. Any value in between these two extremes is attained for some dependence structures. The determination of the minimal value is based on some new ordering results for probability measures in particular on lattice properties of stochastic orderings. Using an extension of Strassen's theorem on representation of the convex order which is of interest in its own we identify properties of dependence structures leading to the minimal optimal stopping value.

Keywords: Optimal stopping, distributions with given marginals, maximal dependence, stochastic ordering, Strassen theorem.

AMS 1991 subject classifications: 60 G40, 62 L15

1 Introduction

Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a sequence of n random variables with distribution functions $F_1, \ldots, F_n$. Let $V(\mathbf{X}) = \sup_{\tau \in \mathcal{T}} EX_\tau$ denote the optimal stopping value of $\mathbf{X}$, where $\mathcal{T}$ denotes the class of all stopping times. In general it is not known which kind of ordering of the dependence structure of the distributions of the random variables leads to a corresponding ordering of the optimal stopping value. Rinott and Samuel-Cahn (1987, 1991) have shown that a weak condition of negative dependence leads to an increase of the optimal stopping value compared to the case of independent components with the same marginals. A similar ordering result could be formulated more generally when restricting to threshold stopping times. This restriction, however, is justified only for independent sequences.

Generally one would believe that some form of 'negative dependence' should go with large optimal stopping values, while some form of 'positive dependence' should go with small optimal stopping values. This idea has been made precise in the different but related context of the random variable

$$M_n(\mathbf{X}) = \max_{1\leq i\leq n} X_i,$$

the value which a prophet could reach in an optimal stopping problem. Lai and Robbins (1976, 1978) constructed so called *maximally dependent* random variables $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ with distribution functions $F_1, \ldots, F_n$ such that $M_n(\mathbf{X}) \leq_{st} M_n(\tilde{\mathbf{X}})$ for any $\mathbf{X}$ with the same marginals. Here $\leq_{st}$ denotes the usual stochastic ordering.

The construction principle underlying maximally dependent random variables is as follows. Obviously, for any real number α

$$M_n(\mathbf{X}) \leq \alpha + \sum_{i=1}^{n} (X_i - \alpha)_+ \tag{1}$$

and equality holds in (1) if for some α^* the sets $A_i = \{X_i > \alpha^*\}$ are a disjoint partition of Ω. If all F_i are continuous (in α^*) then this is fulfilled for maximally dependent random variables which are constructed recursively and exhibit a strong form of negative dependence and

$$P(M_n(\tilde{\mathbf{X}}) \geq t) = \min\{1, \sum_{i=1}^{n} (1 - F_i(t))\}.$$

An extension of this construction to a general class of recursively defined functions $\kappa_n(\mathbf{X})$ was given in Rüschendorf (1980, Theorem 7), see also the discussion in Rachev and Rüschendorf (1998, pp. 155-56).

It was recently shown in Müller (2001) that maximally dependent random variables also maximize the optimal stopping value if the F_i are continuous. There a particular simple construction for maximally dependent random variables (which are not unique, as is obvious from (1)) is given. The main idea is as follows. For any stopping time τ and any X_i with d.f. F_i, $1 \leq i \leq n$ holds

$$X_\tau \leq M_n(\mathbf{X}) \leq_{st} M_n(\tilde{\mathbf{X}}).$$

Further, by (1) and the subsequent remarks, with

$$\tilde{\tau}_{\alpha^*} = \inf\{i : \tilde{X}_i \geq \alpha^*\}$$

one obtains $\tilde{\mathbf{X}}_{\tilde{\tau}_{\alpha^*}} = M_n(\tilde{\mathbf{X}})$. Therefore, $\tilde{\tau}_{\alpha^*}$ is an optimal stopping time for $\tilde{\mathbf{X}}$ and

$$V(\tilde{\mathbf{X}}) = E\tilde{\mathbf{X}}_{\tilde{\tau}_{\alpha^*}} = EM_n(\tilde{\mathbf{X}}).$$

Hence, with $\Gamma(F_1, \ldots, F_n)$ the Fréchet class of all n-dimensional distributions with marginals $F_1, \ldots, F_n$ the upper bound of the optimal stopping value is given by

$$V^+(F_1, \ldots, F_n) = \sup_{\mathbf{X} \in \Gamma(F_1, \ldots, F_n)} V(\mathbf{X}) = V(\tilde{\mathbf{X}}) = EM_n(\tilde{\mathbf{X}}).$$

In Müller (2001) there is also an interesting example showing that for discrete distributions $F_1, \ldots, F_n$, $V^+(F_1, \ldots, F_n) < EM_n(\tilde{\mathbf{X}})$ is possible.

In the present paper we determine the lower bound

$$V^-(F_1, \ldots, F_n) = \inf_{\mathbf{X} \in \Gamma(F_1, \ldots, F_n)} V(\mathbf{X})$$

and the 'positive dependence structure' of the random vector $\mathbf{X}$ with

$$V(\mathbf{X}) = V^-(F_1, \ldots, F_n).$$

The argument for the lower bound can not be based on a simple comparison with the maximum (or some related quantity) as in (1). Instead we use a special construction of a Snell envelope, which is based on an extension of the Strassen theorem on convex domination. We also introduce as main tool of the proof a new operation on distributions, related to lattice properties of stochastic orderings.

The paper is organized as follows. Section 2 contains the main results, including the determination of the lower bound, a result on the dependence structure of the corresponding random variables, and the proof of the fact that the set of all stopping values is an interval. In section 3 we describe the explicit construction of the distribution $F \nabla G$, which is the main tool in section 2. This makes it possible to determine the minimal value explicitly. Section 4 contains the extension of the Strassen theorem on convex domination, which allows to determine the strong positive dependence property of the random variables delivering that minimal value.

2 Determination of the lower bound

The Hoeffding-Fréchet class $\Gamma(F_1, \dots, F_n)$ has a 'unique' maximally positive dependent element $\mathbf{X}^* = (F_1^{-1}(U), \dots, F_n^{-1}(U))$, U uniformly distributed on $[0,1]$, with distribution function

$$F(\mathbf{x}) = \min_{1 \le i \le n} F_i(x_i).$$

This distribution minimizes the maximum in the stochastic order (cf. Rüschendorf (1980, 1981), i.e.

$$M_n(\mathbf{X}^*) \le_{st} M_n(\mathbf{X}) \quad \text{for all } \mathbf{X} \in \Gamma(F_1, \dots, F_n).$$

Therefore, it is natural to conjecture that $\mathbf{X}^*$ also yields the lower bound for the optimal stopping value. The following example, however, shows that this is not true in general, i.e.

$$V(\mathbf{X}^*) > V^-(F_1, \dots, F_n).$$

is possible.

Example 2.1. Assume that X_1 and X_2 have uniform distributions with $X_1 \sim U(-1, 0)$, $X_2 \sim U(-2, 2)$ and denote by F_1 and F_2 the corresponding distribution functions. Consider the optimal stopping problem of $\mathbf{X}^* = (F_1^{-1}(U), F_2^{-1}(U))$. A simple calculation shows that the optimal stopping time is given by

$$\tau^* = \begin{cases} 1, & \text{if } X_1 < -\frac{2}{3}, \\ 2, & \text{otherwise} \end{cases}$$

with optimal stopping value $V(\mathbf{X}^*) = EX_{\tau^*} = \frac{1}{6}$. In the independent case $\tilde{\mathbf{X}} = (\tilde{X}_1, \tilde{X}_2)$ with $\tilde{X}_1 \sim U(-1, 0)$, $\tilde{X}_2 \sim U(-2, 2)$ the optimal stopping time obviously is $\tilde{\tau} \equiv 2$ and

$$V(\tilde{\mathbf{X}}) = EX_2 = 0 < \frac{1}{6}.$$

So in this example the strongest positive dependence does not yield the smallest optimal stopping value. Indeed, here the independence structure yields the lower bound, since EX_2 is an obvious lower bound of the optimal stopping value irrespective of the dependence structure.

For the determination of the lower bound $V^-(F_1,\ldots,F_n)$ of the optimal stopping values we remind some properties of stochastic orderings and the corresponding lattices. For a comprehensive treatment of stochastic orders see Stoyan (1983) or Shaked and Shanthikumar (1994). Recall that for distributions F and G the usual stochastic order $F \leq_{st} G$ holds if $\int f \mathrm{d}F \leq \int f \mathrm{d}G$ for all increasing functions f for which the integrals are defined. It is well known that $F \leq_{st} G$ holds iff $F(x) \geq G(x)$ for all real x. Using this characterization it is easy to see that the set of all distributions endowed with the order $\leq_{st}$ forms a lattice, where the least upper bound $F \vee_{st} G$ has the distribution function $\min\{F,G\}$ and the greatest lower bound $F \wedge_{st} G$ has the distribution function $\max\{F,G\}$.

For distributions with a finite mean the *increasing concave order* $F \leq_{icv} G$ holds if $\int f \mathrm{d}F \leq \int f \mathrm{d}G$ for all increasing concave functions f for which the integrals are defined. Equivalently with the so-called *integrated distribution function*

$$\Phi_F(x) = \int \max\{0, x-t\} F(\mathrm{d}t) = \int_{-\infty}^{x} F(t)\mathrm{d}t$$

$F \leq_{icv} G$ holds iff $\Phi_F(x) \geq \Phi_G(x)$ for all real x, see e.g. Stoyan (1983, p. 11). Since F is the right derivative of Φ_F, the integrated distribution function characterizes the distribution, and moreover, a function Φ is an integrated distribution function iff it is increasing convex with

$$\lim_{x\to-\infty} \Phi_F(x) = 0 \quad \text{and} \lim_{x\to\infty} (\Phi_F(x) - x) = -\int x F(\mathrm{d}x),$$

see e.g. Müller (1998). The lattice property of the increasing convex order is an easy consequence of this observation. The greatest lower bound $F \wedge_{icv} G$ has the integrated distribution function $\max\{\Phi_F, \Phi_G\}$, and the least upper bound $F \vee_{icv} G$ has the integrated distribution function $conv\{\Phi_F, \Phi_G\}$, where *conv* denotes the convex hull operator (see Kellerer (1972), Kertz and Rösler (1992) and Müller (1996), where the corresponding results were obtained for the increasing convex order which is in some sense dual to $\leq_{icv}$). Note that the usual stochastic order $\leq_{st}$ can be characterized via the integrated distribution function, too. $F \leq_{st} G$ holds iff $\Phi_F - \Phi_G$ is increasing.

In the sequel we will need an upper bound $F \nabla G$ of F and G that combines the lattice properties of $\leq_{st}$ and $\leq_{icv}$. For distributions F and G we are looking for the smallest distribution H with respect to $\leq_{icv}$ such that $F \leq_{st} H$ and $G \leq_{icv} H$.

Let $L(F,G) = \{H : F \leq_{st} H \text{ and } G \leq_{icv} H\}$. For any element $H \in L(F,G)$ the function Φ_H has the property that $\Phi_G \geq \Phi_H$ and $\Phi_F - \Phi_H$ is increasing. Therefore, the function

$$\Psi(x) = \sup\{\Phi_H(x) : H \in L(F,G)\}$$

shares these properties, too. Moreover, it is easy to see that Ψ is an integrated distribution function. We denote the corresponding distribution function(which is the right derivative of Ψ) by $F\nabla G$. Notice that this operation is not commutative; in general $F\nabla G \neq G\nabla F$. $F\nabla G$ can be determined explicitly under some regularity conditions, see section 3.

Lemma 2.2.
$F\nabla G$ is an element of $L(F,G)$ and
a) $F\nabla G \leq_{icv} H$ for all $H \in L(F,G)$.
b) If $G_1 \leq_{icv} G_2$ then $F\nabla G_1 \leq_{icv} F\nabla G_2$.

Proof. By definition $\Phi_{F\nabla G} = \Psi = \sup\{\Phi_H; H \in L(F,G)\} \leq \Phi_G$ and $\Phi_F - \Phi_{F\nabla G} = \inf\{\Phi_F - \Phi_H; H \in L(F,G)\}$ is increasing. Therefore, $F\nabla G \in L(F,G)$.

a) is obvious from the definition of $F\nabla G$.

b) follows from the fact that $G_1 \leq_{icv} G_2$ implies $L(F,G_2) \subset L(F,G_1)$.

The following proposition will be useful for an explicit construction of $F\nabla G$ in Section 3. Let C_G denote the set of all continuity points of G. □

Proposition 2.3. *a)* $\min\{F,G\} \leq F\nabla G \leq F$;
In particular: $F(x) \leq G(x)$ implies $F\nabla G(x) = F(x)$.
b) For $F\nabla G$ almost all $x \in C_G$ it is $F\nabla G(x) \in \{F(x), G(x)\}$.

Proof. a) is obvious from the definition of $F\nabla G$.
b) Let $H^* = F\nabla G$ and assume that for some $x \in C_G$, $G(x) < H^*(x) < F(x)$ and $\Phi_{H^*}(x) < \Phi_G(x)$. Then we obtain

$$\inf_{y \geq x} [\Phi_G(y) - (\Phi_{H^*}(x) + (y-x)H^*(x))] = 0. \tag{2}$$

Otherwise, it would be possible to increase the value of $H^*(y)$ and thus also that of $\Phi_{H^*}(y)$ in a neighbourhood of x without violating the admissibility of H^* in contradiction to the maximality of Φ_{H^*}.

If the infimum in (2) is attained at a finite value y_0, then $G(z) < F(z)$ on $[x, y_0)$ while $\Phi_G(y_0) = \Phi_{H^*}(x) + (y_0 - x)H^*(x)$ implies that $H^*(z) = H^*(x)$ for all $z \in (x, y_0)$ i.e. $H^*([x, y_0)) = 0$. Furthermore, $H^*(y_0) = G(y_0)$ if $y_0 \in C_G$. If $y_0 = \infty$, then $H^*(z) = H^*(x)$ for all $z \geq x$ and so $H^*(x) = 1$.
If $G(y) < F(y)$ for $y \in [x, x_0]$ and $\Phi_{H^*}(x) = \Phi_G(x)$, then maximality of Φ_{H^*} implies that

$H^*(y) = G(y)$ for all $y \in [x, x_0]$. This is the case in particular if $G(x) < F(x)$, $x \in C_G$ and $\Phi_{H^*}(x) = \Phi_G(x)$. Thus the statement of the Proposition holds in all possible cases. $\square$

As a consequence of Proposition 2.3 b) we obtain

Corollary 2.4. *If F and G are continuous then $F\nabla G$ is continuous, too.*

An iterated application of the ∇-operator leads to smallest majorizing sequences with respect to the increasing concave ordering.

Proposition 2.5. *Let $F_1, \dots, F_n$ be any distribution functions on the real line. Define by backward induction*

$$G_n^* = F_n,$$

$$G_i^* = F_i \nabla G_{i+1}^*, \quad i = n-1, n-2, \dots, 1.$$

and denote

$$A(F_1, \dots, F_n) = \{G = (G_1, \dots, G_n) : G_i \geq_{st} F_i,\ G_1 \geq_{icv} \dots \geq_{icv} G_n\}.$$

Then $G^* = (G_1^*, \dots, G_n^*)$ is the smallest sequence in $A(F_1, \dots, F_n)$ w.r.t. $\geq_{icv}$ in the sense that for any $G \in A(F_1, \dots, F_n)$ and any $i = 1, \dots, n$

$$G_i^* \leq_{icv} G_i.$$

Proof. By definition $G_i^* = F_i \nabla G_{i+1}^* \geq_{st} F_i$ and $G_i^* \geq_{icv} G_{i+1}^*$. Hence $G^* \in A(F_1, \dots, F_n)$. Moreover, for any $G \in A(F_1, \dots, F_n)$ we have $G_n^* = F_n \leq_{st} G_n$ and hence $G_n^* \leq_{icv} G_n$. Since $G_{n-1} \geq_{st} F_{n-1}$ and $G_{n-1} \geq_{icv} G_n$ it holds that $G_{n-1} \in L(F_{n-1}, G_n)$. Therefore, Lemma 2.2 yields

$$G_{n-1} \geq_{icv} F_{n-1} \nabla G_n \geq_{icv} F_{n-1} \nabla G_n^* = G_{n-1}^*.$$

By induction the result follows. $\square$

Theorem 2.6. (Lower stopping value)
Let $F_1, \dots, F_n$ be arbitrary distribution functions and let $G_1^*, \dots, G_n^*$ be the majorizing sequence constructed in Proposition 2.5. Then

$$V^-(F_1, \dots, F_n) = \inf_{\mathbf{X} \in \Gamma(F_1, \dots, F_n)} V(\mathbf{X}) = \int x G_1^*(\mathrm{d}x).$$

Proof. By the well-known backward induction principle for any $\mathbf{X} \in \Gamma(F_1, \dots, F_n)$ it holds that $V(\mathbf{X}) = EZ_1$ where $\mathbf{Z} = (Z_1, \dots, Z_n)$ is the least majorizing supermartingale of $\mathbf{X}$ (the so-called Snell envelope), see e.g. Chow et al. (1971). Let G_i be the distribution function of Z_i. The supermartingale property implies $G_1 \geq_{icv} \dots \geq_{icv} G_n$ and further $Z_i \geq X_i$ implies $G_i \geq_{st} F_i$, i.e. $G \in A(F_1, \dots, F_n)$ and therefore by Proposition 2.5

$$EZ_1 = \int xG_1(\mathrm{d}x) \geq \int xG_1^*(\mathrm{d}x),$$

i.e.

$$V(\mathbf{X}) \geq \int xG_1^*(\mathrm{d}x) \tag{3}$$

for all $\mathbf{X} \in \Gamma(F_1, \dots, F_n)$.

Conversely, since $G_1^* \geq_{icv} \dots \geq_{icv} G_n^*$, there exists a supermartingale $\mathbf{Z}^* = (Z_1^*, \dots, Z_n^*)$ with $Z_i^* \sim G_i^*$, $1 \leq i \leq n$, see Strassen (1965) or Theorem 4.4 below. Let $V_1, \dots, V_n$ be i.i.d. uniformly $U(0,1)$-distributed random variables independent of $\mathbf{Z}^*$ and define for a distribution function G the generalized inverse

$$\tau_G(z, v) = P(Z < z) + vP(Z = z),$$

where $Z \sim G$. Then

$$X_i^* := F_i^{-1}(\tau_{G_i^*}(Z_i^*, V_i)) \sim F_i, \quad 1 \leq i \leq n.$$

Furthermore, $F_i \leq_{st} G_i^*$ implies $F_i^{-1} \leq (G_i^*)^{-1}$, and hence

$$X_i^* \leq (G_i^*)^{-1}(\tau_{G_i^*}(Z_i^*, V_i)) = Z_i^* \quad \text{a.s.},$$

see Rüschendorf (1981). Thus $\mathbf{Z}^* = (Z_1^*, \dots, Z_n^*)$ is a majorizing supermartingale of $\mathbf{X}^* = (X_1^*, \dots, X_n^*)$ and $\mathbf{X}^* \in \Gamma(F_1, \dots, F_n)$. Therefore,

$$V(\mathbf{X}^*) \leq EZ_1^* = \int xG_1^*(\mathrm{d}x)$$

and this yields

$$V^-(F_1, \dots, F_n) \leq V(\mathbf{X}^*) \leq \int xG_1^*(\mathrm{d}x). \tag{4}$$

From (3) and (4) the assertion of the theorem follows. $\square$

In the following proposition we determine properties of a dependence structure, for which the lowest optimal stopping value is attained. Recall that a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is said to be *conditionally increasing in sequence* (CIS), if $P(X_{i+1} > t | X_i = x_i, \ldots, X_1 = x_1)$ is an increasing function of $x_1, \ldots, x_i$ for all t and all i. When $\mathbf{X} = (X_1, \ldots, X_n)$ is in addition Markovian, then it is called a *stochastically increasing Markov chain.*

Proposition 2.7. *In the problem*

$$V^-(F_1, \ldots, F_n) = \inf_{\mathbf{X} \in \Gamma(F_1, \ldots, F_n)} V(\mathbf{X})$$

the infimum is attained for a vector $\mathbf{X} = (X_1, \ldots, X_n)$ *which is CIS. Moreover, if* $F_1, \ldots, F_n$ *are continuous then it is attained for a stochastically increasing Markov chain.*

Proof. It is well known that the vector $\mathbf{Z}^* = (Z_1^*, \ldots, Z_n^*)$ constructed in the proof of Theorem 2.6 via Strassen's theorem can be chosen to be Markovian. It will be shown in Theorem 4.5 below that the transition probabilities in addition can be chosen stochastically increasing. The stochastic monotonicity is preserved under the monotone transformation to $\mathbf{X}^*$. If the distributions $F_1, \ldots, F_n$ are continuous then by Proposition 2.3 also the distribution functions $G_1^*, ..., G_n^*$ are continuous. Therefore the transformation from the Z_i^* to the X_i^* is one-to-one, and hence also the Markov property is preserved under the transformation. $\square$

Notice that CIS is a strong notion of positive dependence, stronger than e.g. association or positive quadrant dependence, and hence implying that all correlations within the vector $\mathbf{X}$ are non-negative, see e.g. Barlow and Proschan (1975) for details.

The following proposition states that any value between $V^-(F_1, ..., F_n)$ and $V^+(F_1, ..., F_n)$ is attained as optimal stopping value for some dependence structure.

Proposition 2.8. *For any value* ν *between the upper stopping value* $V^+(F_1, \ldots, F_n)$ *and the lower stopping value* $V^-(F_1, \ldots, F_n)$ *there is some* $\mathbf{X} \in \Gamma(F_1, \ldots, F_n)$ *with* $V(\mathbf{X}) = \nu$.

Proof. For any $\mathbf{X}, \mathbf{Y} \in \Gamma(F_1, \ldots, F_n)$ and $\alpha \in [0, 1]$ define the mixture

$$P_\alpha := \alpha P_{\mathbf{X}} + (1 - \alpha) P_{\mathbf{Y}}.$$

Then $P_\alpha \in \Gamma(F_1, \ldots, F_n)$ and

$$h(\alpha) := V(P_\alpha) = \sup_\tau (\alpha E X_\tau + (1 - \alpha) E Y_\tau)$$

is a convex, lower semicontinuous function of α, since it is a supremum of linear functions. This implies that h is even continuous. Thus any value between $V(\mathbf{X})$ and $V(\mathbf{Y})$ is attained for some α. Hence the set $\{V(\mathbf{X}) : \mathbf{X} \in \Gamma(F_1, \ldots, F_n)\}$ is an interval. $\square$

In general, we do not know whether or not the interval in Proposition 2.8 is closed. It follows from Theorem 2.6 that the lower bound $V^-(F_1, \ldots, F_n)$ is always attained, and it has been shown in Müller (2001) that the upper bound $V^+(F_1, \ldots, F_n)$ is attained if $F_1, ..., F_n$ are continuous. If at least one of the distributions $F_1, ..., F_n$ is not continuous, however, then it is an open problem whether or not there is a dependence structure for which the upper bound is attained.

3 Explicit Construction of $F\nabla G$

For the calculation of the optimal stopping interval it is of interest to have an explicit method for the construction of $F\nabla G$. Such a construction can be given if F and G have an at most countable discrete set of crossing points $-\infty \leq s_1 < t_1 < s_2 < t_2 < \ldots$ such that $F(x) \leq G(x)$ iff $x \in (s_i, t_i)$ for some i. It is closely related to the idea of the proof of Proposition 2.3. Define:

$$\begin{aligned}
s_1 &:= \inf\{x : G(x) > F(x)\} \\
t_1 &:= \inf\{x > s_1 : F(x) > G(x)\} \\
&\ \vdots \\
s_i &:= \inf\{x > t_{i-1} : G(x) > F(x)\} \\
t_i &:= \inf\{x > s_i : F(x) > G(x)\} \\
&\ \vdots
\end{aligned}$$

In the case $s_1 = -\infty$ we have that $F \leq G$ on $(-\infty, t_1)$ while for $s_1 > -\infty$ we have $G \leq F$ on $(-\infty, s_1)$. For the construction of $F\nabla G$ our aim is to find a distribution function H such that $H \leq F$ and $\Phi_H \leq \Phi_G$ and Φ_H is maximal in this class. Therefore, we have to define H recursively between two crossing points $s_i \leq x \leq t_i$ such that $H(x) \leq F(x)$ and $\Phi_H \leq \Phi_G$ is maximal. For this we have to observe that since H has to be chosen monotonically nondecreasing, the inequality $\Phi_H \leq \Phi_G$ also has to hold in $G^{-1}(x)$ if H is chosen constant on the interval $[x, G^{-1}(x)]$. For this argument compare also the proof of Proposition 2.3.

So we have to choose $H(x) = F(x)$ on an interval $[s_i, y_i), y_i \leq s_{i+1}$, $H(x)$ constant equal

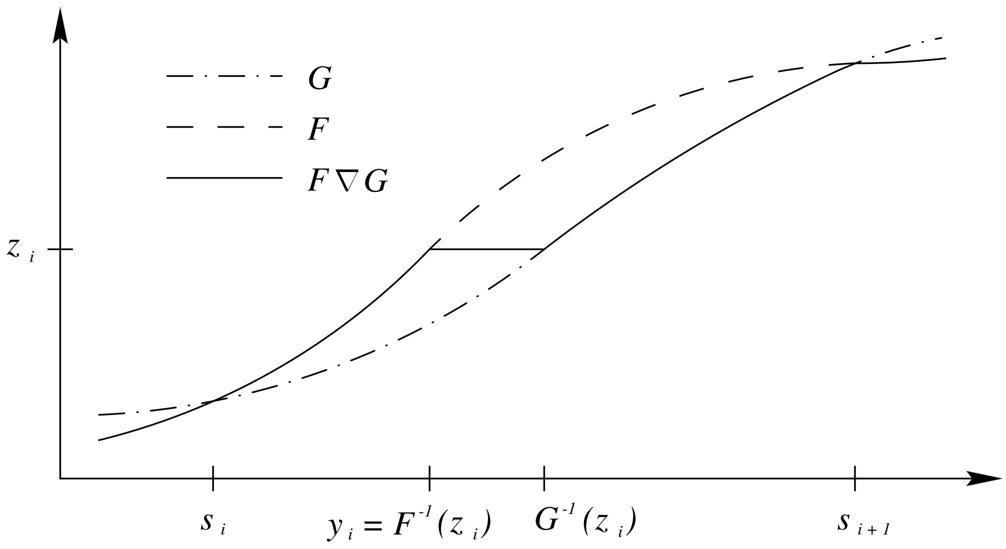

Figure 1: Construction of $F\nabla G$

to z_i on the interval $[y_i, G^{-1}(z_i))$ and $H(x) = G(x)$ on the interval $[G^{-1}(z_i), s_{i+1})$. Here

$$\begin{aligned} z_i &= \sup\left\{ z > G(t_i) : \Phi_F(F^{-1}(z)) + z(G^{-1}(z) - F^{-1}(z)) \;\leq\; \Phi_G(G^{-1}(z))\right\} \\ &= \sup\left\{ z > G(t_i) : z \;\leq\; \frac{\Phi_G(G^{-1}(z)) - \Phi_F(F^{-1}(z))}{\Phi_G(G^{-1}(z)) - \Phi_F(F^{-1}(z)}\right\} \end{aligned}$$

and $y_i = \min\{F^{-1}(z_i), s_{i+1}\}$.

The distribution function $F\nabla G$ is obtained as follows:

$$F\nabla G(x) = \begin{cases} G(x), & x < s_1 \\ \vdots & \\ F(x), & s_i \leq x < y_i \\ z_i, & y_i \leq x < G^{-1}(z_i) \\ G(x), & G^{-1}(z_i) \leq x < s_{i+1} \\ \vdots & \end{cases}$$

To see, why this construction indeed yields $F\nabla G$, notice that by Proposition 2.3 on its support $F\nabla G$ equals either F or G. Moreover, $F\nabla G(x)$ equals $F(x)$ whenever $F(x) \leq G(x)$, and when $F\nabla G(x) = G(x)$ then also $\Phi_{F\nabla G}(x) = \Phi_G(x)$. However, there are gaps in the support of $F\nabla G$ between $G^{-1}(z_i)$ and $F^{-1}(z_i)$, where $F\nabla G$ switches from F to G (see Figure 1).

Example 3.1. (Continuation of 2.1) Assume $F_1 = U(-1,0)$ and $F_2 = U(-2,2)$. These distribution functions are piecewise linear, and hence the integrated distribution functions are piecewise quadratic. An easy calculation yields

$$Z_1^* \sim F_1 \nabla F_2 = \frac{2}{3} \cdot U(-1, -\frac{1}{3}) + \frac{1}{3} \cdot U(\frac{2}{3}, 2).$$

The supermartingale (Z_1^*, Z_2^*) constructed via Strassen's theorem has the transition probability measure

$$P(z_1, \mathrm{d}z_2) = \begin{cases} -\frac{3}{4}z_1 \cdot U(-2, -\frac{2}{3}) + (1 + \frac{3}{4}z_1) \cdot U(-\frac{2}{3}, \frac{2}{3}), & -1 \le z_1 < -\frac{1}{3} \\ \delta_{z_1}, & z_1 \ge \frac{2}{3} \end{cases},$$

where as usual δ_x denotes the point mass in x.

The monotone transformation described in the proof of Theorem 2.6 yields the worst case dependence structure $\mathbf{X}^* = (X_1^*, X_2^*) \in \Gamma(F_1, F_2)$ with transition probability measure

$$Q(x_1, \mathrm{d}x_2) = \begin{cases} -\frac{3}{4}x_1 \cdot U(-2, -\frac{2}{3}) + (1 + \frac{3}{4}x_1) \cdot U(-\frac{2}{3}, \frac{2}{3}), & -1 \le x_1 < -\frac{1}{3} \\ \delta_{4x_1+2}, & -\frac{1}{3} \le x_1 \le 0 \end{cases}.$$

Any stopping time τ^* with $\tau^* = 2$ if $X_1^* \ge -1/3$ is optimal and the corresponding optimal stopping value is

$$V^-(F_1, F_2) = V(\mathbf{X}^*) = EZ_1^* = 0.$$

There are many other random vectors $\mathbf{X} \in \Gamma(F_1, F_2)$, which also yield the worst case optimal stopping value $V^-(F_1, F_2) = V(\mathbf{X}) = 0$. As we have already observed in Example 2.1 this even holds for the case of independent components.

Remark 3.2. Let $F = \delta_0, G = \frac{1}{2}(\delta_{-1} + \delta_2)$ then $F \nabla G = \frac{3}{4}\delta_0 + \frac{1}{4}\delta_2$ (see Figure 2).

In this case we obtain

$$H^*(x) = \sup\{H(x); H \le F, \Phi_H \le \Phi_G\} = 1 \text{ for } x \ge \frac{1}{2}$$

since $\delta_x \in L(F,G)$ for $x \ge \frac{1}{2}$. This shows that the sup of $L(F,G)$ in the increasing concave order $\le_{icv}$ is strictly smaller than the sup of $L(F,G)$ in the stochastic order $\le_{st}$.

4 Monotone representation of the convex order

This section deals with Strassen's theorem stating that two probability measures P and Q can be compared in convex order, if and only if there is a martingale (X,Y) with marginals

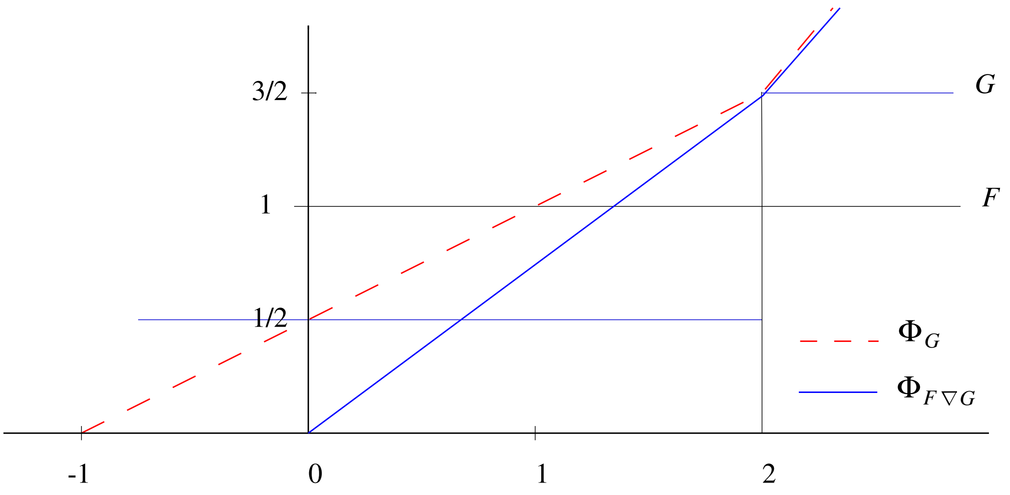

Figure 2: Φ_G and $\Phi_{F\nabla G}$

P and Q, respectively. Our proof will reveal the new result that in the real case the martingale can be chosen such that Y is stochastically increasing in X. Strassen (1965) gives a nonconstructive functional analytic proof of the representation result in a general context.

For distributions on the real line, we give a simple algorithmic proof of Strassens result, which delivers additional insight into the problem. Though we need in this paper the version of the theorem for the increasing concave order $\leq_{icv}$, we give the proof in the usual formulation for the dual convex order $\leq_{cx}$, respectively for $\leq_{icx}$. Two probability measures P and Q are said to be in convex order (written $P_1 \leq_{cx} P_2$), if $\int f \, \mathrm{d}P_1 \leq \int f \, \mathrm{d}P_2$ holds for all convex functions f, such that the integrals exist. Strassen's theorem states that in this case there are random variables X and Y with distributions P_1 and P_2, such that (X, Y) is a martingale, i.e. $X = E[Y|X]$, or in other words, there is a Markov kernel $Q(\cdot, \cdot)$ with $\int yQ(x, \mathrm{d}y) = x$ and $P_2(A) = \int Q(x, A)P_1(\mathrm{d}x)$.

The idea of our proof is based on the representation of convex order by so called *mean preserving spreads*, a concept that is well known in economics, see e.g. Machina and Pratt (1997), Müller (1998) and Rothschild and Stiglitz (1970).

We do not claim that this is the first constructive proof of Strassen's theorem. Indeed, even the classic paper of Blackwell (1953) contains a constructive proof. Our proof, however, is much simpler and can easily be transformed into an efficient algorithm. This is not the case for Blackwell's proof. The same holds true for the proof based on the concept of fusions introduced by Elton and Hill (1992). A constructive proof for distributions with finite support in arbitrary Euclidean spaces can be found in Elton and Hill (1998).

The main tool for the proof is the so called *integrated survival function*

$$\Psi_X(a) := E(X-a)^+ = \int_a^\infty (1-F_X(t))\,\mathrm{d}t, \quad a \in \mathbb{R}.$$

This is the counterpart to the integrated distribution function considered in section 2. It is straightforward to show that the integrated survival function can be characterized as follows.

Proposition 4.1.
a) Ψ_X *has the following properties:*

(i) Ψ_X *is decreasing and convex;*

(ii) the right derivative $D^+\Psi_X$ *exists and* $-1 \leq D^+\Psi_X \leq 0$*;*

(iii) $\lim_{t\to\infty}\Psi_X(t) = 0,\ \lim_{t\to-\infty}(\Psi_X(t)-t) = EX.$

b) To every function $\Psi : \mathbb{R} \to \mathbb{R}$*, which fulfills (i) - (iii) in a) there is a random variable* X*, such that* Ψ *is the integrated survival function of* X*. The distribution function of* X *is given by* $F_X(t) = D^+\Psi_X(t) + 1,\ t \in \mathbb{R}$.

The following result is an immediate consequence.

Corollary 4.2. *The following conditions are equivalent:*

1. $X \leq_{cx} Y$*;*

2. $\Psi_X(a) \leq \Psi_Y(a)$ *for all* $a \in \mathbb{R}$ *and* $\lim_{a\to-\infty}(\Psi_Y(a) - \Psi_X(a)) = 0$.

This characterization will be crucial for the proof of our main result. We first consider a special case.

Lemma 4.3. *Let* Ψ_1 *be an integrated survival function and let* ℓ *be some affine function with* $\ell(t) = at+b$ *for some* $a \in (-1,0)$ *and some* $b \in \mathbb{R}$*. Define* $\Psi_2(t) = \max\{\Psi_1(t), \ell(t)\}$*. Then there are random variables* X_1, X_2 *with integrated survival functions* Ψ_1, Ψ_2 *such that* $E[X_2|X_1] = X_1$*. Moreover, the conditional law* $[Y|X=x]$ *is stochastically increasing in* x*.*

Proof. If $\ell(t) \leq \Psi_1(t)$ for all $t \in \mathbb{R}$, then $\Psi_1 = \Psi_2$ and hence the result is trivial. Therefore let us assume that this is not the case. Then there are two points $t_1, t_2 \in \mathbb{R}$ where ℓ and Ψ_1 coincide.

Now we first have to show that Ψ_2 is an integrated survival function, but that follows easily from Proposition 4.1. Next observe that Ψ_1 and Ψ_2 fulfill condition 2 of Corollary 4.2. Hence the corresponding distributions P_1 and P_2 are ordered with respect to convex order. From convexity of Ψ_1 and linearity of ℓ we conclude $\Psi_2(t) = \ell(t)$ if $t \in [t_1, t_2]$ and $\Psi_2(t) = \Psi_1(t)$ else. This means that $P_2(A) = 0$, if $A \subset (t_1, t_2)$, and $P_2(A) = P_1(A)$, if $A \subset (-\infty, t_1) \cup (t_2, \infty)$, i.e. P_2 is obtained from P_1 by removing all mass from the interval (t_1, t_2) and moving it to the endpoints in such a way that the mean is preserved.

It is easy to see that in this case there is a Markov kernel with the desired properties. In fact, define $Q(x, \mathrm{d}y)$ as follows:

$$Q(x, \cdot) := \begin{cases} \delta_x & , \quad x \notin (t_1, t_2) \\ \frac{x-t_1}{t_2-t_1}\delta_{t_2} + \frac{t_2-x}{t_2-t_1}\delta_{t_1} & , \quad t_1 \leq x \leq t_2 \end{cases}, \tag{5}$$

where δ_x is the degenerated probability measure with point mass in x. It is obvious that $\int yQ(x, \mathrm{d}y) = x$. Moreover, $\int Q(x, A)P_1(\mathrm{d}x) = 0$, if $A \subset (t_1, t_2)$ and $\int Q(x, A)P_1(\mathrm{d}x) = P_1(A)$, if $A \subset (-\infty, t_1) \cup (t_2, \infty)$. Hence

$$\int Q(x, A)P_1(\mathrm{d}x) = P_2(A) \quad \text{for all measurable } A.$$

Moreover, it is clear from the definition in (5) that Q is stochastically monotone (see Keilson (1979) or Stoyan (1983) for a definition). $\square$

From the special case treated in Lemma 4.3 it will now be easy to derive the general result.

Theorem 4.4. (An extension of Strassen's theorem.)
For two probability measures P and Q on the real line the following conditions are equivalent:

1. *$P \leq_{cx} Q$;*
2. *There are random variables X, Y with distributions P, Q, such that $E[Y|X] = X$ and the conditional law $[Y|X = x]$ is stochastically increasing in x.*

Proof. Let Ψ_1 and Ψ_2 be the integrated survival functions corresponding to P and Q. Since Ψ_2 is convex, it can be written as the supremum of a countable family $\ell_1, \ell_2, \ldots$ of affine functions (e.g. the lines of support in all rational points). Now define recursively the functions ϕ_n, $n \in \mathbb{N}$, by $\phi_1 := \Psi_1$, and $\phi_{n+1} := \max\{\phi_n, \ell_n\}$, $n \in \mathbb{N}$. Then ϕ_n, $n \in \mathbb{N}$, is an increasing sequence of integrated survival functions converging to Ψ_2. Moreover, all pairs (ϕ_n, ϕ_{n+1}) fulfill the assumptions of Lemma 4.3.

Let P_n be the probability measures corresponding to ϕ_n. According to Lemma 4.3 there is a Markov kernel $Q_n(x, \mathrm{d}y)$ with $\int yQ_n(x, \mathrm{d}y) = x$ that links P_n to P_{n+1}. Hence there is a Markovian martingale $X \equiv X_1, X_2, \ldots$ with $X_n \sim P_n$, $n \in \mathbb{N}$. Since

$$\begin{aligned} E|X_n| &= 2EX_n^+ - EX_n \\ &= 2\phi_n(0) - \lim_{a\to-\infty}(\phi_n(a) - a) \\ &\leq 2\Psi_2(0) - \lim_{a\to-\infty}(\Psi_1(a) - a) \\ &= 2EY^+ - EX \end{aligned}$$

the sequence $X_1, X_2, \ldots$ is an L_1-bounded martingale, and hence it converges almost surely to a random variable Y with $E[Y|X] = X$, and obviously Y has the integrated survival function Ψ_2 and hence the distribution Q.

Moreover, we know from Lemma 4.3 that all Markov kernels $Q_n,\ n \in \mathbb{N}$ are stochastically monotone. Hence the sequence $X_1, X_2, \ldots$ is a stochastically increasing non-homogeneous Markov chain, converging a.s. to Y. Therefore it follows from results in Stoyan (1983) that $[Y|X = x]$ is stochastically increasing in x. □

It is now easy to extend the result of Theorem 4.3 to the orderings $\leq_{icx}$ (defined as $P \leq_{icx} Q$, if $\int f\mathrm{d}P \leq \int f\mathrm{d}Q$ for all increasing convex functions) and $\leq_{icv}$.

Theorem 4.5. *For distributions on the real line*
a) $P \leq_{icx} Q$ holds iff there are random variables X, Y with distributions P, Q, such that $E[Y|X] \geq X$ and the conditional law $[Y|X = x]$ is stochastically increasing in x.
b) $P \geq_{icv} Q$ holds iff there are random variables X, Y with distributions P, Q, such that $E[Y|X] \leq X$ and the conditional law $[Y|X = x]$ is stochastically increasing in x.

Proof. a) It is well known that $P \leq_{icx} Q$ holds iff there is a P_1 such that $P \leq_{st} P_1 \leq_{cx} Q$, see e.g. Makowski (1994) or Müller (1996). It is easy to find random variables X and X_1 with distributions P and P_1 such that $X \leq X_1$ with probability 1 and with the property that $[X_1|X = x]$ is stochastically increasing in x. (They can simply be obtained by applying the inverses of the distribution functions to the same uniformly distributed random variable.) The assertion then follows from Theorem 4.3.
b) is obtained from a) since $X \geq_{icv} Y$ holds iff $-X \leq_{icx} -Y$. □

References

[1] Barlow, R. E., F. Proschan (1975). *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart and Winston, New York.

[2] Blackwell, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Stat.* **24,** 265 – 272.

[3] Chow, Y.S., H. Robbins, and D. Siegmund (1971). *Great Expectations: Theory of Optimal Stopping.* Houghton Mifflin Company, Boston.

[4] Elton, J. and Hill, T. P. (1992). Fusions of a probability distribution. *Ann. Probab.* **20,** 421 – 454.

[5] Elton, J. and Hill, T. P. (1998). On the basic representation theorem for convex domination of measures. *J. Math. Anal. Appl.* **228,** 449–466.

[6] Keilson, J. (1979). *Markov Chain Models – Rarity and Exponentiality.* Springer, New York.

[7] Kellerer, H. G. (1972). Markov-Komposition und eine Anwendung auf Martingale. *Math. Ann.* **198**, 99–122.

[8] Kertz, R.P. and U. Rösler (1992). Martingales with given maxima and terminal distributions. *Isr. J. Math.* **69**, 173 - 192.

[9] Lai, T.L. and H. Robbins (1976). Maximally dependent random variables. *Proc. Nat. Acad. Sci., USA* **73**, 286 - 288.

[10] Lai, T.L. and H. Robbins (1978). A class of dependent random variables and their maxima. *Z. Wahrscheinlichkeitstheor. verw. Geb.* **42**, 89 - 112.

[11] Machina, M. and Pratt, J. (1997). Increasing risk: Some direct constructions. *Journal of Risk and Uncertainty* **14,** 103 – 127.

[12] Makowski A.M. (1994). On an elementary characterization of the increasing convex ordering, with an application. *J. Appl. Probab.* **31**, 834 - 840.

[13] Müller, A. (1996). Orderings of risks: A comparative study via stop-loss transforms. *Insur. Math. Econ.* **17**, 215 - 222.

[14] MÜLLER, A. (1998). Comparing risks with unbounded distributions. *J. Math. Econom.* **30,** 229–239.

[15] MÜLLER, A. (2001). Bounds for optimal stopping values of dependent random variables with given marginals. *Statist. Probab. Lett.* **52**, 73 - 78.

[16] RACHEV, S.T. AND L. RÜSCHENDORF (1998). *Mass Transportation Problems. Vol. 1: Theory. Vol. 2: Applications.* Springer, New York.

[17] RINOTT, Y. AND E. SAMUEL-CAHN (1987). Comparisons of optimal stopping values and prophet inequalities for negatively dependent random variables. *Ann. Stat.* **15**, 1482 - 1490.

[18] RINOTT, Y. AND E. SAMUEL-CAHN (1991). Orderings of optimal stopping values and prophet inequalities for certain multivariate distributions. *J. Multivariate Anal.* **37**, 104 - 114.

[19] ROTHSCHILD, M. AND STIGLITZ, J. E. (1970). Increasing risk. I. A definition. *J. Econom. Theory* **2,** 225–243.

[20] RÜSCHENDORF, L. (1980). Inequalities for the expectation of Δ-monotone functions. *Z. Wahrscheinlichkeitstheor. verw. Geb.* **54**, 341 - 349.

[21] RÜSCHENDORF, L. (1981). Stochastically ordered distributions and monotonicity of the OC-function of sequential probability ratio test. *Math. Operationsforsch. Stat., Ser. Optimization* **12** 327 - 338.

[22] SHAKED, M. AND J.G. SHANTHIKUMAR (1994). *Stochastic Orders and their Applications.* Academic Press, London.

[23] STOYAN, D. (1983). *Comparison Methods for Queues and Other Stochastic Models.* Wiley, Chichester.

[24] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* **36**, 423 - 439.