Weight Functions and Sign Regularity[1]

by

Per E. Manne[2] and Bertil Tungodden[3]

March 9, 1999.

Abstract: We examine the question of how the ranking between different distributions with respect to a one-parameter family of weight functions depend on the parameter. We argue that in this context sign regularity of the family of weight functions is a natural condition to consider. Several classical economical examples are shown to satisfy this condition. We use sign regularity to obtain results on the possible rankings similar to well-known bounds on the number of internal rates of return on an investment project, either in continuous or discrete time.

1. Introduction

Several problems in economics and social sciences can be regarded as trying to order a finite set of distributions. In the continuous case, we might have a number of functions $f_1, \ldots, f_N$ defined on some common interval, whereas in the discrete case the objects to order might be finite or infinite sequences $a^1, \ldots, a^N$ given by $a^j = (a_0^j, a_1^j, a_2^j, \ldots)$. In the case which initially motivated our investigations, there were given income distributions for a number of societies (or for the same society at different times), and the problem was to order them according to the prevalence of poverty. Another classical problem of the same kind is to order a number of investment projects according to their profitability.

Common to the problems above is that in most cases there is no unequivocal solution. How do we compare a small number of abject poor in one society against a larger number of moderately poor in another? Investment projects are often ranked according to their net present value, but it is well known that different levels of discounting can lead to different rankings.

A common "solution" to this enigma is to choose some weight function $w(t) \geq 0$, and to rank the functions $f_1, \ldots, f_N$ according to the value of the integral

$$\int_I f_j(t)w(t)\,\mathrm{d}t. \tag{1.1}$$

[1] A very preliminary version of this paper was presented at an internal department seminar March 11, 1998 at Geilo, and the authors would like to thank the participants, in particular Steinar Ekern, for valuable comments.

[2] Department of Finance and Management Science, Section of Mathematics and Statistics, Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway. E-mail: per.manne@nhh.no

[3] Norwegian School of Economics and Business Administration and Norwegian Research Centre in Organization and Management. Corresponding adress: LOS Centre, Rosenberggt. 39, N-5015 Bergen, Norway. E-mail: bertil.tungodden@nhh.no

The integral is taken over the common domain of definition I for the functions $f_1, \ldots, f_N$. In the discrete case, the weights are given by a sequence $c = (c_0, c_1, c_2, \ldots)$, and the sequences $a^1, \ldots, a^N$ are ranked according to the value of the sum $\sum_k c_k a_k^j$.

The main problem with this approach is that in general it is difficult to argue persuasively for one given weight function $w(t)$. In the investment case there might be different expectations about the interest rate, or there might be different preferences regarding how to distribute consumption between different periods of time in an optimal manner. This could lead to legitimate disagreement about which weight function to use. It is, however, often possible to argue that any reasonable weight function has to belong to some rather broad class of functions. The class of nonincreasing and nonnegative functions might be such a natural candidate, both in the investment problem and the poverty measurement problem. In the first case it reflects a weakly positive preference for time, whereas in the second case it corresponds to giving weak preference, *ceteris paribus,* to severely poor people before less poor people.

Narrowing down the choice to some large class $\mathcal{W}$ of weight functions might not seem to yield any substantial advantage. Nevertheless, in a sufficiently large number of cases to be of interest, it is possible to conclude that the ranking between two functions f, g will be the same for all choices of weight functions $w \in \mathcal{W}$. Such results are called dominance results. In this manner one obtains a partial order on any given family of functions $f_1, \ldots, f_N$. As this partial order is quite uncontroversial, much effort has been made to get the most out of these methods. See, e.g., Whitmore and Findlay (1978), Ekern (1981), and Ravallion (1994).

However, in many cases the dominance approach will not support any complete ranking of the alternatives, which implies that the choice of weight function will be of definite importance. In the study of investment projects, a response to this problem has been to report on internal rates of return, and we argue that this strategy can be adopted more generally. In the following, let us define a generalized internal rate of return to be a weight function which implies that two alternatives are considered equal along some dimension. For instance, two investment projects may be equally profitable, or two social states may have the same amount of poverty. As illustrated in Tungodden (1998), it is easy to provide an interpretation of the generalized internal rate of return in the context of poverty measurement.

The advantage of this approach is that we avoid taking a stand on an issue where there might be legitimate disagreement. For example, in the poverty measurement problem, the weight function specifies how to resolve conflicts within the poor group. This is a normative problem on which people differ in opinion, and thus by reporting on generalized internal rates of return we avoid incorporating any normative position in the analysis. We simply clarify the relationship between various normative positions and ordinal conclusions in poverty measurement.

However, it is well-known in corporate finance that there might be a complex landscape of internal rates of return. Thus we should not in general assume that this approach gives us a two-edged story, where the conclusion depends on whether you defend a position above or below a unique generalized internal rate of return. In many cases, there will be a number of generalized internal rates of return, and it is therefore of importance to get a better understanding of the bounds of the number of these rates of return. In corporate finance, there exist several results on bounding the number of internal rates of return of an investment project. Descartes' rule of signs (see, e.g., Borwein and Erdélyi, 1995) implies that this number cannot exceed the number of sign changes in the cash flow. Norstrøm's rule (see Norstrøm, 1972) considers the cumulative of the cash flow, i.e., the cash balance, and says that if the cash balance changes sign exactly once and does not end at zero, then there is a unique internal rate of return. Subsequent work by Pratt (1979) and Pratt and Hammond (1979) consider higher cumulatives and relate the number of internal rates of return to the number of sign changes in these cumulatives.

In this paper we generalize these results, and thereby make them relevant for a broader spectrum of problems in the social sciences. Consider a family of weight functions with one free parameter. We show that under the condition of sign regularity, we can give bounds on the number of generalized internal rates of return completely analogous to the classical bounds on the number of (ordinary) internal rates of return. This shows not only that the number of generalized internal rates of return is easily estimated for sign regular families of weight functions, but more importantly that the concept of generalized internal rate of return is robust and behaves well theoretically. Several classical examples of one-parameter families of weight functions are shown to be sign regular, and thus falling within the scope of this theory.

The plan of the present paper is as follows. In Section 2 we clarify the concept and give some examples of one-parameter families of weight functions. In Section 3 we argue that as the free parameter traverses its domain of definition, the weight should shift from one side of the spectrum to the other in a smooth and even manner, and we relate this to a certain variation diminishing property. This property is equivalent to the bounds on the number of generalized internal rates of return described above. In Section 4 we introduce sign regularity and show how this is related to the variation diminishing property. Finally, in Section 5 we consider numerous examples which show both that sign regularity is a natural context for studying the classical bounds on the number of internal rates of return, and that these methods are applicable in other contexts, previously unconnected with the concepts and methods traditionally used in the analysis of investment projects.

Sign regularity was introduced by Schoenberg (1930), and has since been studied by many people, e.g., in the USSR by Gantmacher and Krein (1960), and in the west by Schoenberg and by Karlin (1968). It has found important applications in several different areas, including statistical decision theory, stochastic diffusion processes, and oscillating mechanical systems. Numerous references can be found in Karlin (1968). The mathematical results in this paper are thus not new, except possibly the formulation of Theorems 2 and 3, but we believe that the applications to rather well-known examples in economics might nevertheless be of some interest. We have not included the most general results possible, but have been satisfied with versions sufficiently strong to apply to examples of the kind that one typically encounters in economics. Proofs of most results are included, making the present paper almost self-contained. Our reasons for this were that these proofs are not so easily found in the literature, and that in the context of the present paper it was possible to give a somewhat simplified presentation.

2. One-Parameter Families of Weight Functions

We restrict our attention to parametrized families of weight functions $w_\alpha(t)$, defined on an interval I, with only one free parameter α, ranging over another interval J. In practise, as α traverses the interval J, one normally would like more and more weight to shift from one side of the interval I to the other. Typically, one extreme case will be where only one end of the distribution matters, e.g., if there is some $\epsilon > 0$ such that $f(t) < g(t)$ for all $0 \leq t < \epsilon$ then f is ranked lower than g with respect to w_α for all sufficiently large α. The other extreme, with the parameter at the other end of its domain of definition, might have been with all attention focused at the other end of the distribution; nevertheless, in practice it seems often to be a uniform weight function with constant unit weight. A typical case would be to let α range over the interval $J = [0, \infty)$ and to have $w_0(t) = 1$ for all t. As α increases from 0 towards ∞ one gradually shift emphasis towards one end of the interval I.

We can also consider, with practically no extra complications, and with completely analogous results, the case where one or both variables are discrete. In fact, assertions about the continuous case, such as Theorem 1 in Section 4 below, are frequently proved by reducing them to the discrete case by an approximation argument. In order to obtain a unified approach to the different cases, it is possible to use measure theoretic arguments (Karlin, 1968). We have chosen not to do this, and use only elementary arguments from analysis and linear algebra. Instead of giving detailed arguments in all cases, we have concentrated on the continuous case, but felt free to give examples where one or both variables are discrete when we believe such examples to be of interest.

We give some economical examples of one-paramter families of weight functions which behave in the manner described above.

Example 2.1. The net present value of a continuous income stream $f(t)$ is given by

$$\int_0^\infty f(t) \mathrm{e}^{-rt}\, \mathrm{d}t, \tag{2.1}$$

where r is the interest rate, often assumed to be constant. Here the weight function is $w_r(t) = \mathrm{e}^{-rt}$. With $r = 0$ all moments of time are given equal weight. As r increases, more and more weight is given to the immediate future.

In the discrete case, the income stream is given by the sequence $a = (a_0, a_1, \ldots)$, where a_k is the income in period k, and the net present value is given by $\sum_{k=0}^{\infty} a_k (1 + r)^{-k}$. The behavior of the weight function $(1 + r)^{-k}$ on the parameter r is similar to the continuous case.

Example 2.2. A common poverty measure, discussed in Foster et. al. (1984), is given by

$$\int_0^z f(x) \left(1 - \frac{x}{z}\right)^\alpha \mathrm{d}x. \tag{2.2}$$

Here $f(x)$ is an income distribution, so that $\int_a^b f(x)\, \mathrm{d}x$ gives the proportion of the population with income between a and b. The parameter z is called the poverty line (people are poor if and only if they have income not above z), and α is a parameter which indicates how to weigh interests between different groups of poor. The weight function is given by

$w_\alpha(x) = (1 - x/z)^\alpha$ for $x \le z$ and by $w_\alpha(x) = 0$ for $x > z$. We consider z to be more fixed than α, in the sense that one first decides upon a poverty line z and then let the parameter α determine how differences within the group of poor people should be weighted. With fixed poverty line z we have a one-parameter family of weight functions $w_\alpha(x)$. Often the condition $\alpha \ge 1$ is imposed to make each $w_\alpha(x)$ convex with respect to x, but we could just as well allow $\alpha \ge 0$. The choice $\alpha = 0$ would then give equal weight to all poor, and as α increases, more and more emphasis would be put on the abject poor relative to other groups of poor people.

Example 2.3. There is another method of poverty measurement which can be put under the same umbrella. The head count index simply adds up the proportion of the population under some poverty line z. This amounts to computing

$$\int_0^\infty f(x) w_z(x)\, \mathrm{d}x, \tag{2.3}$$

where

$$w_z(x) = \begin{cases} 1 & \text{if } x \le z, \\ 0 & \text{if } x > z. \end{cases} \tag{2.4}$$

Note, however, that here *small* values of the parameter $z \in (0, \infty)$ indicate more emphasis on the poorest segment of the population.

Example 2.4. For a similar example, which nevertheless does not belong to the same family as the examples above, we may consider Hannah–Kay indices, also known as generalized Herfindahl indices (see Hannah and Kay, 1977). Here one considers an economical sector consisting of k companies. The relative size of these companies is given by $s = (s_1, \ldots, s_k)$, where each $s_j \ge 0$ and $\sum s_j = 1$. The Hannah–Kay index is defined as

$$p_\alpha(s) = \left(\sum s_j^\alpha\right)^{1/(1-\alpha)}, \tag{2.5}$$

where α is a parameter satisfying $\alpha > 0$. (For $\alpha = 1$, the expression (2.5) is not well defined, and is replaced by $\lim_{\alpha \to 1} p_\alpha(s) = \exp\left(-\sum s_j \log s_j\right) = \prod s_j^{-s_j}$.) Small values of $p_\alpha(s)$ indicate a high degree of concentration with one or a few dominant companies in the given sector, and, conversely, large values of $p_\alpha(s)$ indicate a low degree of concentration with several companies of similar size sharing the market. In the continuous case, the discrete distribution s is replaced by a function $f(x) \ge 0$ such that $\int_{\mathbf{R}} f(x)\, \mathrm{d}x = 1$, and the index $\left(\sum s_j^\alpha\right)^{1/(1-\alpha)}$ by

$$\left(\int_{\mathbf{R}} f(x)^\alpha\, \mathrm{d}x\right)^{1/(1-\alpha)}. \tag{2.6}$$

Note, however, that this functional is not of the form $\int_{\mathbf{R}} f(x) w(x)\, \mathrm{d}x$, just as (2.5) is not of the form $\sum c_{\alpha,j} s_j$. In particular, the Hannah–Kay index is nonlinear as a functional of $f(x)$, and as such it is less tractable mathematically than the examples given above. The results in this paper does not pertain to indices of this kind, and we will not consider them here.

3. The Variation Diminishing Property

We now wish to be more precise about the manner in which the weight is supposed to shift from one extreme to the other as the parameter α traverses the interval J. Consider the ranking of two distributions $f(t)$ and $g(t)$ with respect to the weight function $w_\alpha(t)$ for various values of α. Let $h(t) = f(t) - g(t)$, then this ranking depends only on the sign of $\int_I h(t)w_\alpha(t)\,dt$. If this integral is positive then $f(t)$ is ranked above $g(t)$, and if it is negative then the ranking is reversed. If the integral is equal to zero then $f(t)$ and $g(t)$ are given the same ranking, and the weight function $w_\alpha(t)$ with this particular value of α is then called a generalized internal rate of return for $h(t)$.

Assume now that weight in some sense shifts from right to left on the interval I as α increases, and that large values of α indicate that almost all weight is concentrated on the left end of I. Let $h(t) = f(t) - g(t)$ be as described in Figure 1. In this case $\int_I h(t)w_\alpha(t)\,dt > 0$ for large values of α. As α decreases, the area A_1 will receive relatively less weight and A_2 will receive more. It might be the case that that A_1 will dominate over A_2 for all α, or that A_2 will dominate over A_1 for sufficiently small α. In any case, we expect the integral $\int_I h(t)w_\alpha(t)\,dt$ to change sign *at most once* as α traverses its domain of definition, and hence to get at most one generalized internal rate of return in this case.

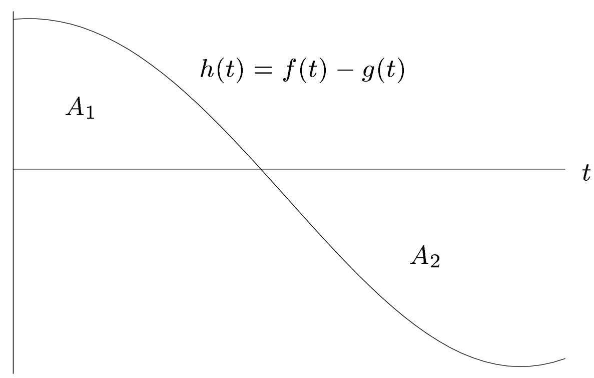

Figure 1

Consider now the case where $h = f - g$ changes sign twice, as in Figure 2. The area A_3 will receive largest relative weight for small α, and it might well be that for such α sufficient weight will be concentrated on A_1 and A_3 to cause f to be ranked above g. As now α increases from these small values, weight will shift from A_2 to A_1, and from A_3 to A_2. It could be the case that we would first see a significant transfer of weight from A_3 to A_2, tipping the balance in favor of g above f, followed by a shift of weight from A_2 to A_1, leaving f on top again. If we would now see a new shift from A_3 to A_2 with little change for A_1, the ranking between f and g could change for the third time (with a fourth change coming when A_1 receives almost all weight). This would be a rather uneven shift of weight, however, where the region seeing the largest shifts is moving back and forth several times. Inversely, more than two changes in the ranking between f and g as α traverses the interval J would seem to indicate an uneven shift of weight.

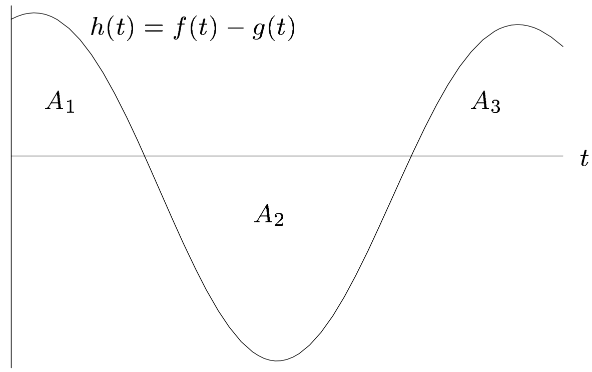

Figure 2

Generalizing, if the weight shifts nicely and if $h = f - g$ changes sign n times on the interval I, we would expect at most n changes in the ranking between f and g as α traverses J, and hence at most n generalized internal rates of return. In the literature this is called the variation diminishing property of the transformation

$$f - g \mapsto \int (f - g) w_\alpha \, \mathrm{d}t, \tag{3.1}$$

as the right hand side of (3.1) has no more "variation about zero," i.e., changes of sign, than the left hand side (see Schoenberg (1930) and Karlin (1968)). Observe that in (3.1) the right hand side is a function of α, whereas the left hand side is a function of t.

Before considering in the next section how to verify whether a one-parameter family satisfies the variation diminishing property, we give some examples which do *not* have this property.

Example 3.1. The following trivial example shows that a one-parameter family of weight functions may have several desirable properties without being variation diminishing. Let three weight functions $w_1(t)$, $w_2(t)$, and $w_3(t)$ be given as in Figure 3, and let $h(t) = f(t) - g(t)$ be as in Figure 4. Consider now how the choice of weight function will affect the ranking of f and g. Both w_1 and w_3 give somewhat more weight to the area A_1 where $f > g$ than to the area A_2, but the difference is not sufficient to compensate for the smaller size of A_1. Hence they will rank g above f. On the other hand, w_2 gives much more weight to the area A_1 than to the area A_2, and, as a result, use of w_2 will lead to ranking f above g.[1] The result is that the ranking between f and g changes twice, and hence that any family $w_\alpha(t)$ which interpolates continuously between $w_1(t)$, $w_2(t)$, and $w_3(t)$ will have at least two generalized internal rates of revenue, even though $f - g$ changes sign only once.

We observe that in this example, each weight function is positive, decreasing, and convex. Also, each $w_i(t)$ has the same fixed end points $w_i(0) = 1$ and $w_i(T) = 0$. Nevertheless,

[1] Numerical integration gives $\int_0^1 h(t) w_i(t) \, \mathrm{d}t = -0.007, 0.013, -0.001$ for $i = 1, 2, 3$, respectively.

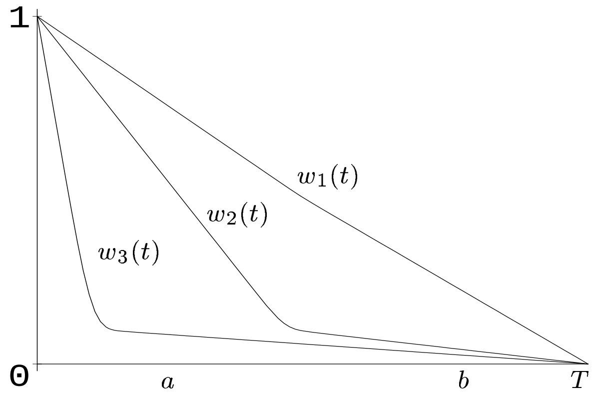

Figure 3

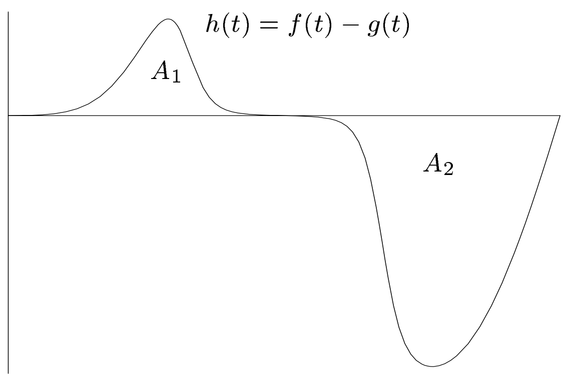

Figure 4

it is not difficult to find faults with the family $w_i(t)$. Let $a < b$ be as in Figure 3, and consider the fraction $\frac{w_i(a)}{w_i(b)}$. This expression measures to what degree we give more weight to a than to b. If weight shifts from right to left in a nice manner as i increases, we would expect $\frac{w_i(a)}{w_i(b)}$ to increase as i increases. But this is not the case in this example, which shows that $w_3(t)$ gives *less* emphasis on some small values than $w_2(t)$. In the terminology of Section 4, it follows that this family of weight functions is not sign regular of order 2.

Example 3.2. In this example the irregularities of the shifts in relative weight is not so easily observed. It is cast in the language of an investment project, but could of course just as easily be applied to, e.g., poverty measurement. We have also chosen a discrete framework, simplifying the necessary computations.

Consider an investment project spanning over three periods of time, with cash flow

$$a_0 = -8, \qquad a_1 = 22, \qquad a_2 = -13 \tag{3.2}$$

in period 0, 1, and 2, respectively. Let the following four different discounting schemes be given:

	Period 0	Period 1	Period 2
Alternative 1	1	1	1
Alternative 2	1	1/2	1/4
Alternative 3	1	4/9	1/9
Alternative 4	1	0	0

(3.3)

The number in each cell gives the present value under a given discounting scheme of one unit obtained in a given period. Here Alternative 1 involves no discounting at all, and Alternative 4 is given by infinite discounting, putting all emphasis on the present moment. Alternative 2 and 3 are in between, with Alternative 3 giving a heavier discounting, reflecting higher rates of interest than Alternative 2. The weights in the table above come from quite explicit and elementary functions, drawn in Figure 5 below. We note that each weight function is decreasing and convex. We assume that there are other weight functions of similar nature varying continuously and spanning the gaps between the four given weight functions. To construct Figure 5, we have chosen the one-parameter family $w_r(t)$ given by $w_r(t) = r^{-t}$ for $1 \leq r \leq 2$, $w_r(t) = (1 - t/3)^r$ for $r \geq 3$, and $w_r(t) = (3 - r)w_2(t) + (r - 2)w_3(t)$ for $2 < r < 3$. The argument is independent of these choices, however, as we will only use the numbers in the table above.

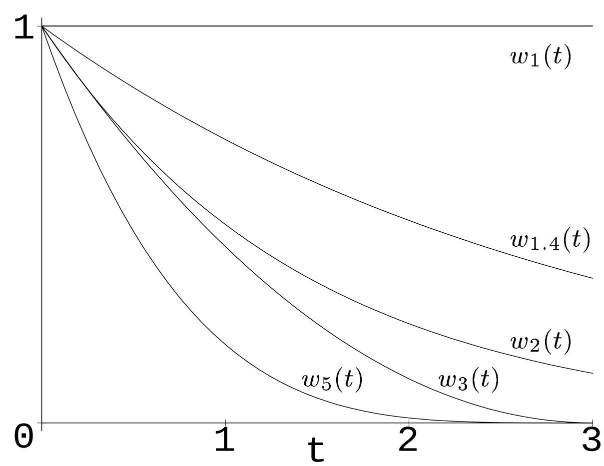

Figure 5

Computing the net present value $c_0a_0 + c_1a_1 + c_2a_2$ of the project given above for each of the four different discounting schemes, we get the results

$$1, \qquad -\frac{1}{4}, \qquad \frac{1}{3}, \qquad -8, \tag{3.4}$$

respectively. In each case the net present value changes sign when we pass from one alternative to the next. This implies that between for instance the first and the second weight function there must be at least one generalized internal rate of return. Since there are three changes of signs in the list of net present values above, we must have at least *three* generalized internal rates of return in this example, instead of the two solutions we would normally expect to find.

It will be clear from the subsequent discussion that there are no fundamental problems *per se* with the concept of generalized internal rate of return with respect to a one-parameter family of discounting functions. However, such a generalized internal rate of return may behave badly unless the family of discounting functions satisfy the condition of sign regularity, introduced in the following section. The present example is a demonstration of what "badly" means in this context.

We notice that the weights in (3.3) satisfy the condition that $\frac{w_i(a)}{w_i(b)}$ is increasing in i whenever $a < b$, showing that this desirable property introduced in Example 3.1 is nevertheless not sufficient for a nicely behaved generalized internal rate of return.

4. Sign Regularity

There is an explicit determinant criterion which is very closely related to the variation diminishing property.

Definition. A function $K(x, y)$ defined on a rectangle $I \times J$ is called sign regular of order n if the following condition is satisfied. For any choice of $x_1 < \ldots < x_k$ and $y_1 < \ldots < y_k$ with $1 \le k \le n$, the sign of the determinant

$$\begin{vmatrix} K(x_1, y_1) & \ldots & K(x_1, y_k) \\ \vdots & \ddots & \vdots \\ K(x_k, y_1) & \ldots & K(x_k, y_k) \end{vmatrix} \tag{4.1}$$

should depend only on k, and not on the choice of the points x_i and y_j. Here we allow the determinant (4.1) to be equal to zero. If $K(x, y)$ is sign regular of order n and the determinant (4.1) is never equal to zero, then $K(x, y)$ is called strictly sign regular of order n.

A function which is (strictly) sign regular of order n for all $n = 1, 2, \ldots$ is called (strictly) sign regular.

A special case of sign regularity occurs when each determinant (4.1) is nonnegative. The function $K(x, y)$ is then called totally positive. If each determinant (4.1) is strictly positive then $K(x, y)$ is called strictly totally positive.

In order to discuss sign regularity of a family $w_\alpha(x)$ of weight functions, we make the formal definition $K(x, \alpha) = w_\alpha(x)$. The meaning of total positivity, etc., of the family $w_\alpha(x)$ is then clear.

Example 4.1. The family $w_i(t)$ of weight functions described in Example 3.1 is not sign regular of order 2. In this case,

$$\begin{vmatrix} w_1(a) & w_1(b) \\ w_2(a) & w_2(b) \end{vmatrix} < 0, \qquad \begin{vmatrix} w_2(a) & w_2(b) \\ w_3(a) & w_3(b) \end{vmatrix} > 0, \tag{4.2}$$

contradicting the requirement that such matrices should have the same sign. In fact, $\begin{vmatrix} w_i(a) & w_i(b) \\ w_j(a) & w_j(b) \end{vmatrix} < 0$ precisely when $\frac{w_i(a)}{w_i(b)} < \frac{w_j(a)}{w_j(b)}$, which was the property considered in Example 3.1.

Proposition 1. If $K(x,y)$ is sign regular on $I \times J$, and if ϕ, ψ are increasing or decreasing functions of one variable taking values in I and J, respectively, then $K(\phi(x), \psi(y))$ is sign regular on $\phi^{-1}(I) \times \psi^{-1}(J)$.

Proposition 2. If $K(x,y)$ is sign regular and $\sigma(x)$, $\tau(y)$ are any functions of one variable which do not change sign on their domains of definition, then

$$K_1(x,y) = \sigma(x)\tau(y)K(x,y) \tag{4.3}$$

is sign regular.

The proof of Proposition 1 is trivial. To prove Proposition 2, observe that

$$\begin{vmatrix} K_1(x_1,y_1) & \dots & K_1(x_1,y_k) \\ \vdots & \ddots & \vdots \\ K_1(x_k,y_1) & \dots & K_1(x_k,y_k) \end{vmatrix} \\ = \sigma(x_1)\dots\sigma(x_k)\tau(y_1)\dots\tau(y_k) \begin{vmatrix} K(x_1,y_1) & \dots & K(x_1,y_k) \\ \vdots & \ddots & \vdots \\ K(x_k,y_1) & \dots & K(x_k,y_k) \end{vmatrix}, \tag{4.4}$$

whence the difference in sign between the two determinants only depend on k, and not on the choice of $x_1, \dots, x_k$ and $y_1, \dots, y_k$.

If $K(x,0)$ and $K(0,y)$ never are equal to zero, we may in Proposition 2 choose $\sigma(x) = \frac{1}{K(x,0)}$ and $\tau(y) = \frac{1}{K(0,y)}$. We then get $K_1(x,0) = K_1(0,y) = 1$ for all x, y. This normalization is often used in practice.

In order to describe the relation between sign regularity and the variation diminishing property, we must first be more specific about how to count the number of sign changes of a sequence or a function. Counting the number of sign changes of a sequence $(a_1, \dots, a_m)$, we first eliminate all zeros. We further define the number of sign changes of a function $h(t)$ defined on some interval to be the maximal value, assuming that it exists, of the number of sign changes of the sequence $(h(t_1), \dots, h(t_m))$, where $m \geq 1$ and $t_1 < \dots < t_m$ are all arbitrary.

Example 4.2. The sequence $(-1, 0, 1, 0, 1)$ and the function $h(t) = t(t-1)^2$ both have one change of sign.

Theorem 1. If $K(x,y)$ is continuous and sign regular on the rectangle $I \times J$, $u(x)$ is any continuous function defined on the interval I, and $v(y)$ is defined by

$$v(y) = \int_I u(x)K(x,y)\,\mathrm{d}x, \tag{4.5}$$

then v has no more sign changes on J than u has on I.

If the interval I is unbounded, we only consider functions $u(x)$ such that the integral in (4.5) is absolutely convergent for each y in J.

In other words, for a sign regular function $K(x, y)$, the transformation (4.5) has the variation diminishing property.

In the discrete case, the function $K(x, y)$ is replaced by a rectangular matrix $C = (c_{ij})$, and instead of (4.5) one considers the linear transformation $v = Cu$. If C is sign regular, meaning that the sign of any subdeterminant only depends on the dimension of the subdeterminant, then the vector v cannot have more sign changes than the vector u. If questions of convergence are handled properly, one could let the matrix C be infinite.

We are also interested in the mixed case, where one variable varies over a countable set and the other over an interval. For an example of this kind, consider an investment project with discrete time and continuous interest rate. Another example is given by Descartes' rule, which compares the number of positive zeros of a polynomial to the number of sign changes among the coefficients of the polynomial.

For a full discussion of Theorem 1, together with full proofs of several variants, see Karlin (1968). We have included an outline of a proof in an Appendix to the present paper.

The bound on the number of sign changes in $v(y)$ given by Theorem 1 is quite weak, and in many particular cases it is possible to improve this bound. This is usually done by cumulating the function $u(x)$. In the present context we get quite easily the following theorem:

Theorem 2. Let $u(x)$ be defined on the interval $I = [a, b]$. Let $u_0(x) = u(x)$ and define inductively $u_{k+1}(x) = \int_a^x u_k(\xi)\,\mathrm{d}\xi$, so that $u_k(x)$ is the kth cumulative of $u(x)$. Assume that $K(x, y)$ is sufficiently differentiable on some rectangle $I \times J'$ (meaning that all expressions which appear are well defined and continuous), that $\frac{\partial^n K}{\partial x^n}(x, y)$ is sign regular on $I \times J'$, and that $\frac{\partial^k K}{\partial x^k}(b, y) = 0$ for $k = 0, 1, \ldots, n$. Let

$$v(y) = \int_a^b u(x)K(x, y)\,\mathrm{d}x. \tag{4.6}$$

Then $v(y)$ has no more sign changes on J' than $u_n(x)$ has on I.

Note that cumulating a function will never increase the number of sign changes, but it can often decrease this number. In typical applications, $K(x, y)$ will be sign regular on some rectangle $I \times J$, and the interval J' will be contained in J. Often $J' = J$, but J' may be strictly smaller than J, as in Example 5.6 below. In practise, $\frac{\partial^k K}{\partial x^k}(x, y)$ will usually be sign regular on $I \times J'$ for all $k = 0, 1, \ldots, n$, but this is actually not required by Theorem 2.

The proof of Theorem 2 is by an n-fold integration by parts, obtaining

$$v(y) = (-1)^n \int_a^b u_n(x)\frac{\partial^n K}{\partial x^n}(x, y)\,\mathrm{d}x, \tag{4.7}$$

followed by appealing to Theorem 1.

If the interval I is given as $[0, \infty)$, as it often is in applications, we only consider functions $u(x)$ with the property that $u_k(x)\frac{\partial^k K}{\partial x^k}(x, y)$ is absolutely integrable with respect to x and approaches zero as $x \to \infty$ for all $y \in J'$ and all $k = 0, 1, \ldots, n$. The same result then follows as above.

Discrete versions of Theorem 2 follow in the same manner as the discrete versions of Theorem 1. We state one such version explicitly, which seems sufficient to cover most cases of interest.

Theorem 3. Given a family $c_i(r)$ of weights, where $i = 0, 1, \ldots$, and where r may be either a discrete or a continuous parameter. Consider the transformation

$$v(r) = \sum_{i=0}^{\infty} c_i(r) u_i. \tag{4.8}$$

Let $u_i^0 = u_i$ for each i, and define inductively $u_i^{k+1} = u_0^k + \ldots + u_i^k$ for each i and k, so that the sequence $(u_0^k, u_1^k, \ldots)$ is the kth cumulative of $(u_0, u_1, \ldots)$. Let $c_i^0(r) = c_i(r)$ for each i, and define inductively $c_i^{k+1}(r) = c_{i+1}^k(r) - c_i^k(r)$ for each i and k, so that the sequence $(c_0^k(r), c_1^k(r), \ldots)$ is the kth difference of the sequence $(c_0(r), c_1(r), \ldots)$. Assume that there is some n such that $c_i^n(r)$ is sign regular with respect to i and r, and that $c_N^k(r) u_N^{k+1} \to 0$ as $N \to \infty$ for each $k = 0, 1, \ldots, n-1$ and for each r. Then $v(r)$ has no more sign changes with respect to r than the sequence $(u_0^n, u_1^n, \ldots)$.

The proof is this time by an n-fold summation by parts, using each time the summation formula

$$\sum_{i=0}^{N} c_i^k(r) u_i^k = c_N^k(r) u_N^{k+1} - \sum_{i=0}^{N-1} c_i^{k+1}(r) u_i^{k+1}. \tag{4.9}$$

Letting $N \to \infty$, we obtain

$$v(r) = (-1)^n \sum_{i=0}^{\infty} c_i^n(r) u_i^n, \tag{4.10}$$

and the result follows from the discrete version of Theorem 1.

Note that even if only finitely many u_i are nonzero, all u_i^k with $k \geq 1$ will typically be nonzero. Hence we need the conditions on the behavior at infinity, even in the cases with finite sequences. See Example 5.2 below.

5. Examples

We now consider again the examples given earlier in this paper, and we give some other constructions of sign regular one-parameter families of weight functions.

Example 5.1. The family $\{e^{rt}\}$ is strictly totally positive. To show this, we must demonstrate that for any choice of $r_1 < \ldots < r_k$ and $t_1 < \ldots < t_k$, the inequality

$$\begin{vmatrix} \exp(r_1 t_1) & \ldots & \exp(r_1 t_k) \\ \vdots & \ddots & \vdots \\ \exp(r_k t_1) & \ldots & \exp(r_k t_k) \end{vmatrix} > 0 \tag{5.1}$$

is satisfied. This is a classical result in mathematical analysis, but for the convenience of the reader we have included a proof. We proceed by induction on k. For $k = 1$ we observe that $e^{r_1 t_1} > 0$. Assume that the claim (5.1) is correct whenever $k = \ell - 1$. Let $k = \ell$, and consider the determinant as a function of t_ℓ. Expanding this determinant by the last column, we obtain

$$g(t_\ell) = c_1 \mathrm{e}^{r_1 t_\ell} + \ldots + c_\ell \mathrm{e}^{r_\ell t_\ell}. \tag{5.2}$$

Here c_ℓ is positive by the inductive hypothesis, and since $r_\ell > r_1, \ldots, r_{\ell-1}$, it follows that the determinant will be positive for sufficiently large t_ℓ. It thus suffices to show that it is nonzero for all $t_\ell > t_{\ell-1}$. An inductive argument, using Rolle's theorem on $g(t_\ell)\mathrm{e}^{-r_\ell t_\ell}$, shows that $g(t_\ell)$ has at most $\ell - 1$ zeros when t_ℓ is allowed to vary over the whole real line. Since $g(t_\ell) = 0$ for $t_\ell = t_1, \ldots, t_{\ell-1}$, it follows that there cannot be any zeros when $t_\ell > t_{\ell-1}$. This shows that if (5.1) is correct for $k = \ell - 1$, then (5.1) is correct also for $k = \ell$. Induction now gives that the claim (5.1) is correct for all k.

It now follows from Proposition 1 that the family $\{\mathrm{e}^{-rt}\}$ is strictly sign regular. As a consequence we obtain the familiar upper bound on the number of internal rates of return of a continuous time investment project given by the number of sign changes in the cash flow. We note that $\frac{\partial^n}{\partial t^n}\mathrm{e}^{-rt} = (-1)^n r^n \mathrm{e}^{-rt}$, and by Proposition 2 this function is strictly sign regular for $r > 0$. We also have $\lim_{t\to\infty} r^n \mathrm{e}^{-rt} = 0$ for $r > 0$, and hence Theorem 2 applies. We can thus obtain better bounds on the number of internal rates of return of a continuous time investment project by considering cumulatives of the cash flow. See also Pratt and Hammond (1979, p. 1238) for a brief description of similar results.

Example 5.2. Observe that $\frac{1}{(1+r)^t} = \mathrm{e}^{-\phi(r)t}$ with $\phi(r) = \log(1+r)$. Since ϕ is strictly increasing, it follows from the preceeding example and Proposition 1 that $K(r,t) = \frac{1}{(1+r)^t}$ is strictly sign regular. By the discrete version of Theorem 1, we obtain the classical rule that the number of internal rates of return of a discrete time investment project cannot exceed the number of sign changes in the cash flow.

We now consider differences of this family of weight functions. Using the notation from Theorem 3, we let $c_i(r) = c_i^0(r) = \frac{1}{(1+r)^i}$. Then $c_i^1(r) = c_{i+1}(r) - c_i(r) = \frac{r}{(1+r)^i}$, and inductively we obtain that the kth differences are given by $c_i^k(r) = \frac{r^k}{(1+r)^i}$. Since $\frac{1}{(1+r)^i}$ is sign regular, it follows from Proposition 2 that $\frac{r^n}{(1+r)^i}$ is sign regular with respect to i and r for $r > 0$. Note also that if u_i is nonzero for only finitely many i, as is the case in a finite-time investment project, then the kth cumulatives u_N^k are bounded by a constant multiple of N^k as $N \to \infty$, whereas the kth differences $c_N^k(r)$ tends to zero exponentially with respect to N as $N \to \infty$. Hence $c_N^k(r)u_N^{k+1}(r) \to 0$ as $N \to \infty$ for each k and r. Theorem 3 now applies, and it follows that the number of internal rates of return of a finite discrete time investment project is bounded by the number of sign changes in the nth cumulative of the income stream. Norstrøm's rule (Norstrøm, 1972) now follow directly as a special case of $n = 1$, whereas the case $n \geq 1$ is the main rule given by Pratt (1979) and Pratt and Hammond (1979). We believe it should not be difficult to obtain the other versions of this rule given in Pratt (1979) and Pratt and Hammond (1979) by the methods of the present paper.

Example 5.3. It should not come as a surprise that Descartes' rule of signs follows from Theorem 1. Let $f(x) = a_0 + a_1 x + \ldots + a_n x^n$, and let $0 < x_1 < \ldots < x_m$ be given.

Then

$$\begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix} = \begin{pmatrix} 1 & x_1 & \dots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \dots & x_m^n \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_n \end{pmatrix}. \tag{5.3}$$

Since the x_i are all strictly positive, any $k \times k$ subdeterminant of the matrix in (5.3) can be written on the form (5.1). Hence the transformation (5.3) has the variation diminishing property, which means that the number of positive zeros of $f(x)$ is bounded above by the number of sign changes in $(a_0, \dots, a_n)$.

Example 5.4. If $\phi(t)$ is any positive function, then we can construct a one-parameter family of weight functions $\phi_r(t) = \mathrm{e}^{-rt}\phi(t)$. Example 5.1 and Proposition 2 show that this family is strictly sign regular.

Example 5.5. Given $w(t)$ with $w(t) > 0$ for all t. Define $w_r(t) = w(t)^r$. We claim that if $w(t)$ is (strictly) monotone then the family $\{w_r(t)\}$ is (strictly) sign regular. Define $\phi(t) = \log w(t)$. Then ϕ is (strictly) increasing or decreasing as w is (strictly) increasing or decreasing, and $w_r(t) = \mathrm{e}^{r\phi(t)}$. The claim now follows from the previous example and Proposition 1.

Example 5.6. As a special case of the previous example, we get that the family $\{(1 - x/z)^\alpha\}$ with z fixed is sign regular with respect to x and α, with $0 \leq x \leq z$ and $-\infty < \alpha < \infty$. This family will of course also be sign regular if α is restricted to any smaller interval, such as $\alpha \geq 0$ or $\alpha \geq 1$.

With $K(x, \alpha) = (1 - x/z)^\alpha$, we get

$$\frac{\partial^k K}{\partial x^k}(x, \alpha) = \frac{(-1)^k}{z^k}\alpha(\alpha - 1)\dots(\alpha - k + 1)\left(1 - \frac{x}{z}\right)^{\alpha - k}, \tag{5.4}$$

which by Proposition 2 is sign regular for $\alpha > k - 1$. Note that here each cumulation reduces the interval on which we can use Theorem 2 to get information on the number of sign changes. See also Tungodden (1998).

Example 5.7. Given

$$K(x, y) = \begin{cases} 1 & \text{if } x \leq y, \\ 0 & \text{if } x > y, \end{cases} \tag{5.5}$$

and choose $x_1 < \dots < x_k$ and $y_1 < \dots < y_k$. Consider the matrix

$$\begin{pmatrix} K(x_1, y_1) & \dots & K(x_1, y_k) \\ \vdots & \ddots & \vdots \\ K(x_k, y_1) & \dots & K(x_k, y_k) \end{pmatrix}. \tag{5.6}$$

Counting from the left, in each row there will be an initial number of zeros (maybe none) followed by only ones. The number of ones cannot increase as we pass to a lower row. The only way for such a matrix to be nonsingular is to be upper triangular with only ones on and above the main diagonal. This happens iff

$$x_1 \leq y_1 < x_2 \leq y_2 < \dots < x_k \leq y_k. \tag{5.7}$$

In particular, the determinant is either 1 or 0, and hence never negative. It follows that $K(x, y)$ is sign regular. A direct consequence of this is that the head count index (Example 2.3) is sign regular, and hence satisfies the variation diminishing property.

As $K(x, y)$ is not everywhere differentiable, Theorem 2 cannot be applied here.

Example 5.8. Given two weight functions $w_0(t)$ and $w_1(t)$ with $w_0(t) \geq w_1(t)$ for all t. We might be interested in interpolating between $w_0(t)$ and $w_1(t)$, and it is natural to do so by convex combinations. We therefore define $w_r(t) = (1-r)w_0(t) + rw_1(t)$ for $0 \leq r \leq 1$ (or even for all real r). A computation shows that if $\lambda < \mu$ then

$$\begin{vmatrix} (1-\lambda)w_0(s) + \lambda w_1(s) & (1-\lambda)w_0(t) + \lambda w_1(t) \\ (1-\mu)w_0(s) + \mu w_1(s) & (1-\mu)w_0(t) + \mu w_1(t) \end{vmatrix} = (\mu - \lambda) \begin{vmatrix} w_0(s) & w_0(t) \\ w_1(s) & w_1(t) \end{vmatrix}. \tag{5.8}$$

It follows that if the two-element family $\{w_0, w_1\}$ is (strictly) sign regular of order two, then the whole family $\{w_r\}$ will also be (strictly) sign regular of order two. Of course, any $n \times n$-matrix (4.1) with $n > 2$ will have zero determinant by linear dependence. In particular, $w_r(t)$ cannot be *strictly* sign regular of any order $k > 2$.

We observe that Example 3.2 was constructed by choosing a sign regular family $\{w_r(t)\}$ for $r \leq 1$, another sign regular family $\{w_r(t)\}$ for $r \geq 2$, and joining them by taking convex combinations of $w_1(t)$ and $w_2(t)$. Apparently, the idea was not good in that case, at least not if we expected to keep the properties ensured by sign regularity.

Example 5.9. Another method for interpolation between two positive functions is to use geometric averages. Let again $w_0(t)$ and $w_1(t)$ be given with $w_0(t) > w_1(t) > 0$ for all t, and assume that the two-element family $\{w_0(t), w_1(t)\}$ is strictly sign regular. Define $w_r(t) = w_0(t)^r w_1(t)^{1-r}$ for $0 \leq r \leq 1$. We claim that $w_r(t)$ is strictly sign regular for all r. Note that the strictly sign regular property of $\{w_0, w_1\}$ implies that $\frac{w_0(t)}{w_1(t)}$ is a strictly monotone function of t. By Example 5.5, $(w_0/w_1)^r$ is strictly sign regular, and by Proposition 2 we now get that $w_r(t) = w_1(t)(w_0(t)/w_1(t))^r$ is strictly sign regular.

Example 5.10. The family $w_r(t)$ given in Example 3.2 will violate the criterion of sign regularity. This family does not shift weight smoothly to the left with increasing r, at least not if we consider "second order" effects. This can be seen by computing

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 1/2 & 1/4 \\ 1 & 4/9 & 1/9 \end{vmatrix} = \frac{1}{36} > 0 \tag{5.9}$$

and

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 1/2 & 1/4 \\ 1 & 0 & 0 \end{vmatrix} = -\frac{1}{4} < 0, \tag{5.10}$$

which shows that $w_r(t)$ is not sign regular of order three.[1]

[1] In fact, when we pass to continuous time this family is not even sign regular of order two. A computation shows that for small positive t the value of $w_2(t)$ is in fact larger than $w_1(t)$. The difference is very minute, however, and hardly visible from the graphs of w_1 and w_2. This defect could easily be removed at the cost of slightly more complicated expressions.

6. Appendix

For the convenience of the reader, we here present a proof of Theorem 1 based on the expositions in Borwein and Erdélyi (1995) and Karlin (1968).

Proof of Theorem 1. (Outline.) We only need to consider the case where u has a finite number of sign changes. Choose $y_1 < \ldots < y_n$ arbitrarily with $v(y_j) \neq 0$ for each j. Choose $\epsilon > 0$ so small that for any partition $a = x_0 < \ldots < x_N = b$ of $I = [a, b]$ with mesh size $\max_{1 \leq i \leq N}(x_i - x_{i-1}) < \epsilon$, $v(y_j)$ and

$$c_j = \sum_{i=1}^{N} u(x_i) K(x_i, y_j)(x_i - x_{i-1}) \tag{6.1}$$

will have the same sign for each $j = 1, \ldots, n$. This is possible since each such sum converges to $v(y_j)$ as the mesh size of the partition approaches zero, and since we have only a finite number of points $y_1, \ldots, y_n$.

Take any such partition, and let

$$A = \begin{pmatrix} K(x_1, y_1) & \ldots & K(x_N, y_1) \\ \vdots & & \vdots \\ K(x_1, y_n) & \ldots & K(x_N, y_n) \end{pmatrix}, \qquad b = \begin{pmatrix} u(x_1)(x_1 - x_0) \\ \vdots \\ u(x_N)(x_N - x_{N-1}) \end{pmatrix}. \tag{6.2}$$

Define $c = Ab$. By assumption, the matrix A is sign regular. By refining the partition of I, if necessary, we may assume that b has as many sign changes as $u(x)$. It thus suffices to show that c does not have more sign changes than b.

Let b have p sign changes. Partition $b = (b_1, \ldots, b_N)^T$, where the superscript T denotes transpose, into $p + 1$ segments

$$(b_1, \ldots, b_{i_1}), (b_{i_1+1}, \ldots, b_{i_2}), \ldots, (b_{i_p+1}, \ldots, b_{i_{p+1}}), \tag{6.3}$$

with $i_{p+1} = N$, where nonzero coefficients of consecutive segments have opposite sign, and some but not all coefficients of a segment may be equal to zero. Without loss of generality, we may assume that the coefficients of the first segment are all ≥ 0. Let

$$d_k = |b_{i_{k-1}+1}| A_{i_{k-1}+1} + \ldots + |b_{i_k}| A_{i_k}, \qquad k = 1, \ldots, p+1, \tag{6.4}$$

where A_j is the j-th column of the matrix A. We then have

$$c = d_1 - d_2 + \ldots + (-1)^p d_{p+1} = De, \tag{6.5}$$

where D is the matrix with columns $d_1, \ldots, d_{p+1}$ and where e is the column vector $(1, -1, \ldots, (-1)^p)^T$.

The matrix D is sign regular, since one can show that any $k \times k$ subdeterminant of D can be written as a linear combination with positive coefficients of $k \times k$ subdeterminants of A. Sign regularity is thus inherited from A to D, and if A is strictly sign regular then D will be strictly sign regular as well.

We consider first the case where the matrix D is strictly sign regular. Assume that c has more than p sign changes. Choose $i_1 < \ldots < i_{p+1}$ such that $c_{i_1}, \ldots, c_{i_{p+1}}$ are nonzero and alternate in sign. Consider the determinant

$$\begin{vmatrix} d_{i_1 1} & \ldots & d_{i_1 p} & c_{i_1} \\ \vdots & \ddots & \vdots & \vdots \\ d_{i_{p+1} 1} & \ldots & d_{i_{p+1} p} & c_{i_{p+1}} \end{vmatrix}. \tag{6.6}$$

The last column is a linear combination of the p first columns, so the determinant (6.6) is equal to zero. Expanding the determinant by the last column, we obtain

$$0 = \sum_{j=1}^{p+1} (-1)^{p+j} c_{i_j} \Delta_j, \tag{6.7}$$

where Δ_j is the subdeterminant obtained from (6.6) by deleting row j and the last column. Each term in the sum has the same nonzero sign, and thus cannot sum to zero. This contradiction shows that c cannot have more than p sign changes. The theorem is thus proven in the case where D is strictly sign regular.

If D is sign regular and of full rank, but not strictly sign regular, we let $F_\epsilon = (f_{ij})$ be the matrix given by $f_{ij} = \mathrm{e}^{(i-j)^2/\epsilon}$, and approximate $D = (d_{ij})$ by $D_\epsilon = F_\epsilon D$. This matrix can be shown to be strictly sign regular for all $\epsilon > 0$, and $D_\epsilon \to D$ as $\epsilon \to 0$. Let $c_\epsilon = D_\epsilon e$, then c_ϵ has at most p sign changes, and $c_\epsilon \to c$ as $\epsilon \to 0$. It follows that c cannot have more than p sign changes in this case.

The case where D does not have full rank is a little more complicated, and we refer to Karlin (1968, p. 221).

References

P. Borwein and T. Erdélyi: *Polynomials and Polynomial Inequalities.* Springer, New York, 1995.

S. Ekern: *Time Dominance Efficiency Analysis.* Journal of Finance **36** (1981), 1023–1034.

J. Foster, J. Greer, and E. Thorbecke: *A Class of Decomposable Poverty Measures.* Econometrica **52** (1984), 761–766.

F. R. Gantmacher and M. G. Krein: *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme.* Akademie-Verlag, Berlin, 1960.

L. Hannah and J. A. Kay: *Concentration in Modern Industry: Theory, Measurement and the U.K. Experience.* Macmillan, London, 1977.

S. Karlin: *Total Positivity.* Stanford University Press, Stanford, 1968.

C. J. Norstrøm: *A Sufficient Condition for a Unique Nonnegative Internal Rate of Return.* Journal of Financial and Quantitative Analysis **7** (1972), 1835–1839.

J. W. Pratt: *Finding How Many Roots a Polynomial Has in* $(0, 1)$ *or* $(0, \infty)$. American Mathematical Monthly **86** (1979), 630–637.

J. W. Pratt and J. S. Hammond, III: *Evaluating and Comparing Projects: Simple Detection of False Alarms.* Journal of Finance **34** (1979), 1231–1242.

M. Ravallion: *Poverty Comparisons.* Harwood, Chur, 1994.

I. J. Schoenberg: *Über variationsvermindernde lineare Transformationen.* Mathematische Zeitschrift **32** (1930), 321–328.

B. Tungodden: *Poverty Measurement: The Critical Comparison Value.* NHH, Bergen, 1998.

G. A. Whitmore and M. C. Findlay (eds.): *Stochastic Dominance.* Lexington Books, Lexington, 1978.