ON THE ASYMPTOTIC POWER OF THE TWO-SIDED KOLMOGOROV-SMIRNOV TEST

Hartmut Milbrodt and Helmut Strasser

November, 1988

Abstract

In this paper, the asymptotic power of the two-sided one-sample Kolmogorov-Smirnov test (KS) is investigated and compared with the power of some other goodness of fit tests (EDF tests). It is found that this test (similar to the ones of Anderson-Darling and Cramér-von Mises) does not behave like a well-balanced procedure for higher-dimensional alternatives: There are only few directions of deviations from the hypothesis for which it is of reasonable asymptotic power. These directions are determined as follows: Employing spectral theory of compact operators a principal components decomposition of the curvature of the asymptotic power function at the hypothesis is established. This decomposition is given in terms of an orthogonal series in the "tangent space" of directions of alternatives. It shows how the respective test distributes its power in the space of all alternatives. The beginning of the series is evaluated numerically for the KS test. Comparison with the curvature of the optimal power function for a given one-dimensional alternative yields local efficiencies of KS which are high for one direction only, and then rapidly decrease to zero. These findings are supported by global upper bounds of the asymptotic power function of EDF tests, which indicate that for directions with small curvature at the origin the whole power function must have bad efficiency properties.

[0] AMS 1980 Subject Classifications: 62G10, 62F05.
Keywords and phrases: Kolmogorov-Smirnov test, Cramér-von Mises test, Anderson-Darling test, Neyman's smooth test, tangent space, asymptotic power function, Brownian bridge, empirical distribution function.

1 Introduction: The Asymptotic Power Function

Let $P_0|\mathcal{B}(\mathbb{R})$ be a probability measure with a continuous distribution function F_0. We consider the problem of testing $P = P_0$ against $P \neq P_0$ based on n independent observations distributed according to P. For this "goodness of fit problem (case 0)" a variety of EDF tests, i.e. procedures based on comparison of the hypothetical distribution function F_0 with the empirical distribution function F_n are available: The two-sided tests of Cramér-von Mises (CM), Anderson-Darling (AD), Kolmogorov-Smirnov (KS) and many others (cf. Shorack and Wellner, [23], pp. 142–150).

The asymptotic behaviour of the first two of these is well-investigated (see Anderson and Darling, [1], Durbin and Knott, [8], and Neuhaus, [15]), whereas little seems to be known about the asymptotic power of the two-sided KS test. A survey of results concerning the distribution theory for KS tests is given in Durbin, [7]. Additionally, there are also Monte Carlo studies investigating the power of the KS test against some *specific* alternatives (notably Miller and Quesenberry, [20] and [21], and Solomon and Stephens, [24]). They indicate that the KS test might be inferior to AD and CM, and in particular to parametric tests of the χ^2-type, such as Neyman's, [18], smooth tests (NS).

The theoretical results known so far have been primarily of a qualitative nature. It is well known that the KS test (like AD and CM) is capable of finally detecting every deviation from P_0, if only the sample size is large enough. But such an assertion is only useful, if one knows which types of deviations can be detected with a reasonable sample size, and for which other alternatives its power is rather poor.

It is the aim of the present work to provide a complete and quantitative description of the asymptotic power of the two-sided KS test, which shows how this test distributes its power in the space of *all* alternatives when the sample size is large, and how it compares with the other tests mentioned above.

Our investigations are within the framework of the local asymptotic theory. They are based on the concept of a tangent vector of a sequence of local alternatives (P_n). The *full tangent space of* P_0 is

$$H(P_0) := \{h \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_0) : P_0(h) = 0\}.$$

The sequence (P_n) has a *tangent vector* $h \in H(P_0)$, if it approaches P_0 from direction h with rate $1/\sqrt{n}$:

$$\sqrt{\frac{dP_n}{dP_0}} = 1 + \frac{1}{2\sqrt{n}} h + \frac{1}{\sqrt{n}} r_n,$$

where $\lim_{n\to\infty} P_0(r_n^2) = 0$ and $\lim_{n\to\infty} nP_0\{dP_n/dP_0 = \infty\} = 0$. This concept, which is closely related to LeCam's condition of L^2-differentiability of square roots of likelihood ratios, was developed by Koshevnik and Levit, [12], and Pfanzagl and Wefelmeyer, [19].

For every sample size $n \in \mathbb{N}$ let φ_n be a test on $\mathbb{R}^n$. For reasonable sequences $\varphi := (\varphi_n)$

the power functions evaluated at local alternatives converge,

$$\lim_{n\to\infty} P_n^n(\varphi_n) =: g_\varphi(h, P_0, \alpha),$$

the limit being a function of the tangent vector $h \in H(P_0)$, the hypothesis P_0 and the *level*

$$\alpha := \lim_{n\to\infty} P_0^n(\varphi_n).$$

This function g_φ, which is characteristic for $\varphi = (\varphi_n)$, is the *asymptotic power function* of the sequence of tests. Note, that it depends on the alternatives only via the direction h from which they approach the hypothesis.

In the present paper we study this dependence in case of the two-sided Kolmogorov-Smirnov test (i.e. for $\varphi =$ KS).

(1.1) REMARK For tests φ based on comparison of the hypothetical and the empirical distribution function the usual quantile-transformation gives

$$g_\varphi(h, P_0, \alpha) = g_\varphi(h \circ F_0^{-1}, U(0,1), \alpha), \quad h \in H(P_0),$$

where $U(0,1)$ denotes the uniform distribution on the unit interval. Hence, in the following we shall assume w.l.g. that $P_0 = U(0,1)$, and drop P_0 from the notation wherever this is unambiguous (e.g. $H = H(U(0,1))$, $g_\varphi(\cdot,\cdot) = g_\varphi(\cdot, U(0,1), \cdot), \ldots$).

(1.2) DISCUSSION

1. As known from Hájek and Šidák ([10], Chapt. VI, Sect. 4.5) the asymptotic power function of KS tests can be expressed in terms of shifted Brownian bridges: Let

   $$Z_n \colon t \longmapsto \sqrt{n}(F_n(\cdot, t) - t), \quad 0 \le t \le 1, \quad n \in \mathbb{N},$$

   denote the normalized empirical processes, $Z = (Z_t)$ the canonical Brownian bridge on the Skorohod space $D[0,1]$, W_0 its distribution and define

   $$\tau(h) : t \mapsto \int_0^t h(s)ds, \quad t \in [0,1], \quad h \in H.$$

   If (P_n) is a sequence with tangent vector $h \in H$ then we have

   $$\mathcal{L}(Z_n | P_n^n) \to \mathcal{L}(\tau(h) + Z | W_0), \quad \text{weakly.}$$

   Hence, replacing the hypothesis by a sequence of such alternatives amounts asymptotically to shifting by the indefinite integral of their tangent vector.

2. For the asymptotic power function of the two-sided KS test of level $\alpha \in (0,1)$ this yields
$$g_{KS}(h,\alpha) = W_0\Big\{\sup_{0\le t\le 1} |\tau(h)(t) + Z_t| > c_\alpha\Big\}, \quad h \in H, \tag{1}$$
where c_α is the $(1-\alpha/2)$–quantile of the limiting distribution of the two-sided KS-test statistic:
$$4\sum_{j=1}^{\infty}(-1)^{j+1}\exp(-2j^2c_\alpha^2) = \alpha.$$

3. Similarly, also the asymptotic power of other EDF tests can be expressed in terms of shifts of Brownian bridges. Employing this representation and Fourier theory in L^2-spaces as in Anderson and Darling, [1], the asymptotic power function of the two-sided AD test and CM test were given by Durbin and Knott, [8], Sect. 5, and Neuhaus, [15], p. 99, in the following form:
$$h \mapsto W_0\Big\{\sum_{j=1}^{\infty}\mu_j(Y_j + \langle h, h_j\rangle)^2 \ge c_\alpha\Big\}, \quad h \in H, \tag{2}$$
where $\mu_j \searrow 0$ is summable, $(h_j) \subseteq H$ is an orthonormal system in the tangent space and the random variables (Y_j) are i.i.d. $\mathcal{N}(0,1)$. (For details cf. the beginning of section 2.)

To our knowledge a representation like (2) is not available for tests of the KS type. Hence, a direct analysis of the whole asymptotic power function seems to be much more difficult for these tests. Instead, we try to analyse it locally around the hypothesis, i.e. near the origin of the tangent space, hoping that the results thus obtained do also have implications for the whole asymptotic power function. A similar approach was first pursued by Hájek and Šidák ([10], VI 4.5 and VII 2.3) in case of one-sided KS tests. However, different to the situation met there, in the two-sided case we are not able to obtain the local asymptotic efficiencies in closed form; we rather have to evaluate them numerically.

As we shall show, all practically relevant sequences of EDF tests φ for the problem of goodness of fit of a simple hypothesis have the property
$$g_\varphi(th,\alpha) = \alpha + a_\varphi(h,\alpha)\frac{t^2}{2} + o(t^2) \quad \text{as } t \to 0, \quad h \in H. \tag{3}$$
The term $a_\varphi(h,\alpha)$ can be expanded as
$$a_\varphi(h,\alpha) = \sum_{j=1}^{\infty}\lambda_j\langle h, h_j\rangle^2, \tag{4}$$
where (h_j) is an orthonormal system in the tangent space H, and $\lambda_j \searrow 0$. For EDF tests of CM type this can be obtained as a consequence of the global representation (2) of the

asymptotic power function (see Sect. 2 and Neuhaus, [15], Lemma 2.5). However, the general result covering also KS tests is more involved.

The present paper contains two main results:

- The existence of a "principal components representation" (3) and (4) of the curvature of the asymptotic power function at the hypothesis.
- The numerical evaluation of the principal components decomposition (eigenvalues λ_j and directions h_j) for the two-sided KS test.

These results together with a detailed discussion are given in the next section. The numerical analysis for the KS test can be found in Sect. 3, and the proofs concerning the first item are contained in the final Sect. 4.

2 Results: Principle Components of the Asymptotic Power Function

All two-sided EDF tests $\varphi = (\varphi_n)$ discussed so far have the following simple structure: They arise from the composition of a semi-norm $|\cdot|_{D[0,1]}$ on the Skorohod-space with the paths of the standardized empirical processes Z_n, i.e.

$$\varphi_n = 1_{\{|Z_n| > c_\alpha\}}, \quad n \in \mathbb{N},$$

thus rejecting the hypothesis if some measure of discrepancy between the empirical distribution and the hypothetical distribution becomes too large.

(2.1) EXAMPLES

1. The two-sided KS test comes from the uniform norm $|z| := \sup_{0 \le t \le 1} |z(t)|$.
2. Tests of CM type (i.e. quadratic EDF tests) are based on semi-norms of the form
$$|z|^2 := \int z(t)^2 \, m(dt)$$
where $m|\mathcal{B}([0,1])$ is a measure. For $m = U(0,1)$ one obtains the usual CM test, and taking m as measure with the Anderson-Darling weight function $t \mapsto \big(t(1-t)\big)^{-1}$ as Lebesgue density gives the AD test.
3. Neyman's smooth tests reject the hypothesis of uniformity if for some prescribed finite orthonormal system $\mathcal{H} := \{h_1, \ldots, h_k\} \subset H$ in the tangent space the test statistics
$$\frac{1}{n} \sum_{j=1}^{k} \Big(\sum_{i=1}^{n} h_j(X_i) \Big)^2, \quad n \in \mathbb{N},$$

exceed some critical value c_α depending on the level α. Thus, they are defined by the semi-norm
$$|z|^2 := \sum_{j=1}^{k}\Big(\int_0^1 h_j(t)\, z(dt)\Big)^2.$$
(In Neyman's original paper [18] the set $\mathcal{H}$ was taken to consist of the first k Legendre polynomials on the unit interval).

(2.2) DISCUSSION Asymptotically, NS tests meet deviations in all directions belonging to span$\mathcal{H}$ with the same power, and under this constraint they are optimal among all sequences of tests with the same level. Since their asymptotic power function is

$$g_{\mathcal{H}}(h,\alpha) = 1 - \chi^2_{k,\sum_{j=1}^k \langle h,h_j\rangle^2}(c_\alpha), \quad h \in H, \tag{5}$$

NS tests are sometimes also called χ^2-tests. Note, that the asymptotically optimal power function given a single direction $\mathcal{H} = \{h_1\}$ is

$$g_{\{h_1\}}(h,\alpha) := \Phi\big(N_{\alpha/2} - \frac{\langle h,h_1\rangle}{\|h_1\|}\big) + \Phi\big(N_{\alpha/2} + \frac{\langle h,h_1\rangle}{\|h_1\|}\big).$$

(Φ denotes the standard normal distribution function and N_* its quantiles.)

More generally than (1) and (5), if the semi-norm defining the EDF test φ is W_0-a.e. continuous w.r.t. the uniform norm, then by the invariance principle for the standardized empirical distribution function, the asymptotic power function can be written as

$$g_\varphi(h,\alpha) = W_0\big\{|\tau(h) + Z| > c_\alpha\big\}, \quad h \in H, \tag{6}$$

where c_α is determined from $W_0\big\{|Z| > c_\alpha\big\} = \alpha$. Of course, this continuity assumption holds for KS and CM. But despite the fact that the AD norm is not always finite, it also holds for AD; using the law of the iterated logarithm this has been shown by Anderson and Darling ([1], pp. 196/197).

Let us mention some qualitative properties of the asymptotic power function $t \mapsto g_\varphi(th,\alpha)$ for fixed direction $h \in H$ obtainable from (6).

(2.3) REMARK From $g_\varphi(h,\alpha) = 1 - W_0\big\{\tau(h) + Z : |Z| \le c_\alpha\big\}$ and the tightness of W_0 it is clear that
$$\lim_{|t|\to\infty} g_\varphi(th,\alpha) = 1 \quad \text{if } |\tau(h)| \neq 0.$$
Moreover, $g_\varphi(h,\alpha) = g_\varphi(-h,\alpha)$. The Lemma of Anderson then yields that $t \mapsto g_\varphi(th,\alpha)$ is a non-decreasing function of $|t|$. Taking into account that this function is also analytic, we get $g_\varphi(h,\alpha) > \alpha$ for $|\tau(h)| \neq 0$. In particular, if $|\cdot|$ is even a norm (like for AD, CM and KS), the EDF test is strictly asymptotically unbiased.

By definition, the squared semi-norm of the NS test is a finite sum of random variables, which under W_0 are i.i.d. χ_1^2. Similarly, the squared semi-norm of tests of CM type are infinite weighted sums of random variables, which under W_0 are i.i.d. χ_1^2:

(2.4) THEOREM *Let $m|\mathcal{B}([0,1])$ be a measure satisfying*

$$\int_0^1 \tau(h)^2\,dm < \infty, \quad h \in H.$$

Then there are an orthonormal system $(h_j) \subset H$ and a summable sequence $\mu_j \searrow 0$ such that

$$\tau(h)^2 = \sum_{j=1}^{\infty} \mu_j \Big(\int_0^1 h \cdot h_j\,dm\Big)^2, \quad h \in H, \quad \textit{and}$$

$$|Z|^2 = \sum_{j=1}^{\infty} \mu_j \Big(\int_0^1 h_j\,dZ\Big)^2 \quad W_0\textit{-a.e.}$$

Essentially, this theorem is a slightly more general restatement of results of Anderson and Darling ([1], formula (4.5) on p. 197) and the well-known Kac-Siegert representation of the Brownian bridge. It shows that NS tests may be viewed as a particular kind of CM type tests, attributing equal weight to all of their finitely many principal components.

(2.5) EXAMPLES

1. As known from Anderson and Darling, ([1], p. 204) the principal components (h_j) of the semi-norm of the AD test are the Legendre polynomials on the unit interval, the weights being $\mu_j = \big(j(j+1)\big)^{-1}, j \in \mathbb{N}$.
2. The principal components of the ordinary CM test are $h_j : t \mapsto \sqrt{2}\,\cos(j\pi t), j \in \mathbb{N}$, and the weights are $\mu_j = (j\pi)^{-2}, j \in \mathbb{N}$.

For these examples as well as an account on other applications of the Kac-Siegert representation to the distribution theory of quadratic EDF tests see Shorack and Wellner ([23], Chapt. 5).

(2.6) REMARK The following results are due to Neuhaus, [15]. Due to orthonormality of the (h_j) the stochastic integrals $Y_j := \int_0^1 h_j\,dZ, j \in \mathbb{N}$, in Theorem (2.4) are i.i.d. $\mathcal{N}(0,1)$. Hence, the representation (2) of the asymptotic power of AD and CM follow immediately from (6) and this result. Moreover, it also gives the principal components decomposition (3) and (4) of the curvature of the asymptotic power function, the assumptions for that being the same as in the above theorem. It is the basic result of Neuhaus, [15], that the principal components (h_j) coincide in both representations. The weights are connected by

$$\lambda_j = 1 - \alpha - \int_{\left\{\sum_{j=1}^{\infty} \mu_j Y_j^2 \leq c_\alpha\right\}} Y_j^{\,2} \, dW_0, \quad j \in \mathbb{N}. \tag{7}$$

The local weights λ_j of CM type tests may be calculated from that numerically, using expansions of distributions of quadratic forms in Gaussian variables into series of central χ^2-distribution functions or Laguerre polynomials. (For details see Johnson and Kotz, [11], Chapt. VI, Sect.'s 5 and 6.)

In case of NS tests, however, the local weights can be calculated in a simple way: For any fixed normalized direction $h \in H$ the NS test of level α defined by an orthonormal subset $\mathcal{H} = \{h_1, \ldots, h_k\}$ of the tangent space has asymptotic power

$$\begin{aligned} g_{\mathcal{H}}(th, \alpha) &= 1 - \chi^2_{k, t^2 \sum_{j=1}^{k} \langle h, h_j \rangle^2}(c_\alpha) \\ &= \alpha + \lambda^{(k)} \sum_{j=1}^{k} \langle h, h_j \rangle^2 \cdot \frac{t^2}{2} + o(t^2) \quad \text{as } t \to 0. \end{aligned}$$

The coefficient

$$\lambda^{(k)} := \frac{d^2}{dt^2}\Big|_{t=0} \big(1 - \chi^2_{k,t^2}(c_\alpha)\big)$$

remains to be determined. The following simple formula, which will be proved in the last section, has been used for the computation of Table 2 below.

(2.7) Lemma *For every $k \in \mathbb{N}$ and every $c > 0$*

$$\frac{d^2}{dt^2}\Big|_{t=0} \big(1 - \chi^2_{k,t^2}(c)\big) = \chi^2_k(c) - \chi^2_{k+2}(c).$$

Although a principal components decomposition of the *semi-norm* seems to be restricted to quadratic EDF tests φ, the decomposition of the *curvature* $a_\varphi(\cdot, \alpha)$ according to (3) and (4) is available in much more general situations. This is one of the main result of this paper.

(2.8) Theorem *Let φ be a sequence of EDF tests of uniformity constructed from a semi-norm $|\cdot|$ on $D[0,1]$ which is W_0-a.e. continuous w.r.t. the sup-norm and finite on $\tau(H)$. If φ is asymptotically similar of level α, then there are an orthonormal system $(h_j) \subset H$ and a sequence $\lambda_j \searrow 0$ such that for every $h \in H$*

$$a_\varphi(h, \alpha) = \sum_{j=1}^{\infty} \lambda_j \langle h, h_j \rangle^2$$

and the remainders of this series can be estimated by

$$|a_\varphi(h,\alpha) - \sum_{j=1}^{k} \lambda_j \langle h, h_j\rangle^2| \le \lambda_{k+1} \|h\|^2 \,, \quad k \in \mathbb{N} \,.$$

Proof: From (5) and from [26], Example (70.7)(2) and Theorem (70.8), we get

$$g_\varphi(h,\alpha) = \int_{\{|z|>c_\alpha\}} \exp\Big(L(h) - \frac{\|h\|^2}{2}\Big)\, dW_0, \quad h \in H, \tag{8}$$

where

$$L(h)\colon z \mapsto \int_0^1 h(t)\, z(dt), \quad z \in D[0,1],$$

denotes the Wiener integral. Hence, considering the symmetric bilinear form

$$B_{\varphi,\alpha}\colon (h_1,h_2) \mapsto \int_{\{|z|>c_\alpha\}} L(h_1)L(h_2)\, dW_0 - \alpha\langle h_1, h_2\rangle,$$

it is clear that $a_\varphi(h,\alpha) = B_{\varphi,\alpha}(h,h),\ h \in H$. Apparently, $B_{\varphi,\alpha}$ is continuous. Since $t \mapsto g_\varphi(th,\alpha)$ has a minimum at $t=0$, we have $a_\varphi(h,\alpha) \ge 0$; so $B_{\varphi,\alpha}$ is positive semi-definite. Thus, there is a bounded, symmetric and positive semi-definite operator $T_{\varphi,\alpha}$ representing $a_\varphi(\cdot,\alpha)$:

$$a_\varphi(h,\alpha) = \langle T_{\varphi,\alpha}(h), h\rangle, \quad h \in H.$$

Moreover, by Lemma (4.1) $T_{\varphi,\alpha}$ is compact. Since the proof of that lemma is a bit more involved, it is postponed to the appendix. As a consequence there is a spectral representation

$$T_{\varphi,\alpha} = \sum_{j=1}^{\infty} \lambda_j \langle h_j, \cdot\rangle h_j$$

where $\lambda_j \searrow 0$ and (h_j) is an ON-system in the tangent space H. It follows that

$$a_\varphi(h,\alpha) = \sum_{j=1}^{\infty} \lambda_j \langle h, h_j\rangle^2.$$

Denoting $L_k := \operatorname{span}\{h_1,\ldots,h_k\}$, it is well-known that

$$\lambda_{k+1} = \sup_{h \perp L_k} \frac{\langle T_{\varphi,\alpha}(h), h\rangle}{\|h\|^2}, \quad k \in \mathbb{N},$$

which yields the estimate for the remainder:

$$|a_\varphi(h,\alpha) - \sum_{j=1}^{k} \lambda_j \langle h, h_j\rangle^2| = |\sum_{j=k+1}^{\infty} \lambda_j \langle h, h_j\rangle^2|$$

$$= |\langle T_{\varphi,\alpha}(h - p_{L_k}h), h - p_{L_k}h\rangle| \le \lambda_{k+1}\|h - p_{L_k}h\|^2.$$

□

From this theoretical result, in particular from the fact that the local weights converge to 0, it is already clear that all EDF tests will — at least from a local point of view — have reasonable asymptotic power only for a finite-dimensional space of directions of alternatives. Of course, this contradicts our intuitive understanding of omnibus tests. So the natural question suggesting itself is, how rapidly the principal directions loose their influence. In this respect, a major issue of the present paper are numerical results concerning the function $a_\varphi(\cdot,\alpha)$ in case φ is the sequence of two-sided KS tests.

The beginning of the sequence (λ_j) is tabulated in Table 1 for levels $\alpha = .05$, $\alpha = .01$, and $\alpha = .005$. Calculations have been performed along the lines explained in the next section. This table clearly exhibits the strong predominance of the first tangent vector h_1 over the influence of all other tangent vectors.

(2.9) REMARK Employing (7) one gets a similar but even more extreme behaviour of AD and CM, since their local weights decrease even more rapidly. E.g. for CM and the 5%–level (1%–level) we have up to three decimals $\lambda_1 = .226$, $\lambda_2 = .022$, $\lambda_3 = .008$, $\lambda_4 = .005$ and $\lambda_{10} = .000$ (.074, .004, .002, .001,. . . , .000).

It is also illuminating to compare the expansion of $a_{KS}(h,\alpha)$ with the expansion for parametric NS tests obtainable from Lemma (2.7). In the following we do so for $\alpha = .05$, $\alpha = .01$ and $\alpha = .005$, arriving at almost the same results for all three levels.

(2.10) DISCUSSION For all these levels, the local power of the KS test in its main direction h_1 comes close to the one of the respective asymptotically optimal unbiased test, which is the onedimensional NS test for $\mathcal{H} = \{h_1\}$ and level α. The “local efficiency” of KS for this optimal direction, i.e. the ratio of λ_1 and $\lambda^{(1)}$ is .86867 for the 5%–level, .87592 for the 1%–level and .87865 in the .5%–case. Comparing the KS test with NS tests for alternatives of dimension $k = 2$, we observe that the gain of efficiency in its main direction h_1 must be paid with a severe loss in the second direction h_2. This phenomenon becomes even more evident as the dimension of the alternative increases. (Observe that $\lambda^{(10)} > \lambda_2$!) It illustrates that the KS test (like AD and CM) behaves very much like a parametric test for a onedimensional alternative and not like a well-balanced test for higher-dimensional alternatives. Thus, at least from a local point of view the KS test (as well as AD and CM) does not have the omnibus property usually attributed to it. For multidimensional alternatives it is noticeably inferior to the corresponding NS tests.

Next, let us discuss the implications of local efficiencies for the global behaviour of the asymptotic power function.

Table 1: Local weights for the two-sided KS test

	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$
λ_1	.19901	.06525	.03828
λ_2	.04982	.01276	.00693
λ_3	.02280	.00554	.00295
λ_4	.01348	.00320	.00169
λ_5	.00906	.00212	.00112
λ_6	.00659	.00153	.00080
λ_7	.00504	.00116	.00062
λ_8	.00400	.00091	.00048
λ_9	.00326	.00074	.00039
λ_{10}	.00271	.00061	.00032
$\vdots$	$\vdots$	$\vdots$	$\vdots$
λ_{20}	.00076	.00017	.00009
$\vdots$	$\vdots$	$\vdots$	$\vdots$
λ_{30}	.00034	.00005	.00002
$\vdots$	$\vdots$	$\vdots$	$\vdots$
λ_{40}	.00002	.00000	.00000
$\vdots$	$\vdots$	$\vdots$	$\vdots$
λ_{49}	.00000	.00000	.00000

Table 2: Local weights for NS tests

	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$
$\lambda^{(1)}$	.22910	.07449	.04357
$\lambda^{(2)}$	.14982	.04606	.02649
$\lambda^{(3)}$	.11661	.03461	.01947
$\lambda^{(4)}$	.09795	.02885	.01637
$\lambda^{(5)}$	.08557	.02491	.01408
$\lambda^{(6)}$	.07670	.02213	.01247
$\lambda^{(7)}$	.06996	.02005	.01127
$\lambda^{(8)}$	.06464	.01841	.01034
$\lambda^{(9)}$	.06029	.01709	.00958
$\lambda^{(10)}$	.05668	.01600	.00896

Following an approach Hájek and Šidák ([10], VII 2.3) pursued in case of one-sided KS tests, the basic idea is, to show that local properties of the asymptotic power function have a strong impact on its global performance. This is important since, basically, its curvature at the origin is *not* a quantitative measure of the efficiency of the test: (Pitman) Efficiency statements are based on critical values, i.e. on level points of the power function. The power function of a more efficient test reaches the 50%–level or an even higher level earlier than the power function of a less efficient test.

In the following we apply upper bounds for the asymptotic power function obtained by Strasser, [28]. Upper bounds for the power function give us lower bounds for its level points. These bounds show that for tangent vectors h with small values of $a_{KS}(h, \alpha)$ the whole asymptotic power function has bad efficiency properties.

(2.11) REMARK Let us briefly summarize the main assertion of Strasser, [28], as far as it is of importance for us here.

Given a normalized direction $h \in H$ in the tangent space, and a sequence φ of EDF tests satisfying the assumptions of Theorem (2.8), we know three things about the asymptotic power function $t \mapsto g_\varphi(th, \alpha)$: The value and the first two derivatives at 0, namely

$$g_\varphi(0, \alpha) = \alpha, \quad \frac{d}{dt}\Big|_{t=0} g_\varphi(th, \alpha) = 0 \quad \text{and}$$

$$\frac{d^2}{dt^2}\Big|_{t=0} g_\varphi(th, \alpha) = a_\varphi(h, \alpha) = \sum_{j=1}^{\infty} \lambda_j \langle h, h_j \rangle^2 =: \gamma.$$

Then by Theorem (2.7) of Strasser, [28],

$$g_\varphi(th, \alpha) \leq 1 - b\big(\Phi(a - t) - \Phi(-a - t)\big), \quad t \in \mathbb{R},$$

where the constants a, b are chosen according to

$$\alpha = 1 - b\big(2\Phi(a) - 1\big), \quad \text{and} \quad \gamma = 2ab\dot{\Phi}(a).$$

These upper bounds cannot be improved since there are omnibus tests attaining them.

In Table 3 we summarize some numerical results for KS obtained from these bounds. The table contains lower bounds for the 50% and 90% level points of the asymptotic power function for the first ten principal directions in the tangent space. These bounds are compared with the exact level points of the corresponding NS tests. It is evident that the verdict obtained on the basis of local properties of the asymptotic power function is also valid globally: Directions of alternatives for which the asymptotic power function has small curvature at the origin must also be bad from a global point of view. For reasons of fairness, however, it has to be admitted that the bounds do not indicate such drastic losses of global efficiency as the local point of view suggests.. Nevertheless, given a minimal power NS

Table 3: Lower bounds for the level points of KS, $\alpha = .05$

	50%		90%	
dim	KS	NS	KS	NS
1	2.03467	1.95	3.32302	3.24
2	2.67499	2.22	3.98847	3.55
3	2.97616	2.40	4.29252	3.76
4	3.16112	2.53	4.47831	3.92
5	3.29330	2.64	4.61085	4.05
6	3.39505	2.74	4.71278	4.17
7	3.47819	2.82	4.79603	4.28
8	3.54808	2.90	4.86599	4.37
9	3.60874	2.97	4.92667	4.46
10	3.66247	3.03	4.98046	4.53

tests are able to detect alternatives of considerably higher dimension than KS tests are. (Of course, a similar assertion is valid for CM and AD.)

So far we have seen that the KS test — like the other nonparametric EDF tests discussed here — is suitable for detecting deviations from the hypothesis in very few directions only. It remains to identify and describe these tangent vectors. This has been done numerically in the next section.

(2.12) DISCUSSION In Figure 1 plots of the first four tangent vectors are provided in case of the 5%–level (solid lines); the principal components for the 1%–level and the .5%–level are qualitatively the same. For comparison the well-known principal components of CM (broken lines) and AD (broken and dotted lines), which do not depend on the level at all, are also graphed.

Evidently, the tangent vectors for AD are the steepest and those for KS the flattest near the ends of the unit interval, indicating that among these procedures the AD test is most sensitive to tail alternatives and the KS test is the one which is least sensitive to tail alternatives.

Apart from this obvious difference the shape of the principal components is similar for all three procedures. They are by turns antisymmetric and symmetric about $t = 1/2$, and they oscillate faster and faster. Calculating squared "correlations" between the components of these tests exhibits their great similarity: One obtains doubly stochastic infinite matrices with strongly dominating entries in the main diagonal.

For illustration we give the matrix of squared inner products between the first four principal components of CM and KS (5%–level):

$$\begin{pmatrix} .98473 & .00000 & .00100 & .00000 \\ .00000 & .88333 & .00000 & .06331 \\ .01488 & .00000 & .68416 & .00000 \\ .00000 & .11391 & .00000 & .40984 \end{pmatrix}.$$

Since the Legendre polynomials, which are the components of AD, arise from the orthogonalization of powers, it is clear that in some sense for all these tests the first four principal components describe deviations in location, scale, skewness and curtosis, respectively, from the hypothesis.

As we have seen, for EDF tests the curvature of the asymptotic power function at the origin has a strong influence on its global behaviour. Moreover, for AD, CM and KS the eigenvalues in the principal components decomposition of the curvature rapidly decrease. Hence, knowledge of the first few principal directions of the tests (for KS and $\alpha = .05$, $\alpha = .01$ and $\alpha = .005$ the first ones are provided in Table 4 below) also renders information of their performance for *fixed* (non-local) alternatives, when the sample size is large.

Finally, let us mention some attempts to remove the defects of the EDF tests discussed here.

(2.13) REMARKS

1. In order to improve the sensitivity of KS to tail alternatives one may use weighted KS tests based on semi-norms of the type

$$|z| = \sup_{t \in [0,1]} |z(t)\,\psi(t)|$$

where ψ is a suitable weight function on the unit interval. The drawbacks of this suggestion of Anderson and Darling ([1], pp. 204–211) were already noticed by these authors: Computational difficulties with the resulting boundary crossing problems, and the semi-norm thus obtained from the AD weight function is not continuous W_0-a.e.

2. Mason and Schuenemeyer, [14], obtained a modified KS test sensitive to tail alternatives by taking as test statistics a weighted maximum of the KS test statistic and Rényi-type statistics.

3. A very interesting result is due to Neuhaus, [17]. This author suggests the use of the onedimensional NS test with direction h_1 estimated from the data by kernel methods. As shown by Neuhaus, this yields an omnibus test with the same principal directions as CM. The advantage of this approach is that one can influence the weights by the choice of the kernel and the bandwidth. Roughly, a large bandwidth leads to a strong

Figure 1: The first four principal components for KS $(\alpha = .05)$, CM and AD

predominance of the first principal direction whereas a small bandwidth leads to slowly decreasing weights and a test which spreads its power more evenly over the various directions.

4. From the results of the present paper it is clear that we recommend specifying a *finite* set of interesting directions of alternatives (this is what the EDF test does for you anyhow!), and then using the corresponding NS test.

3 The Curvature of the Asymptotic Power Function at the Origin: Numerical Investigations

Our goal in this section is the numerical evaluation of the principal components decomposition of the curvature $a_{KS}(\cdot, \alpha)$ for the KS test, partly given in Table 4 and Figure 1. For this we employ a representation of $a_{KS}(\cdot, \alpha)$ which is inspired by Hájek and Šidák, [10], VI.4. Let $\|Z\|_{u,v} := \max_{u \le r \le v} |Z_r|$ $(u \le v)$.

(3.1) THEOREM *There is a measure defining function* $G_\alpha \colon [0,1]^2 \to [0,\infty)$ *such that*

$$a_{KS}(h,\alpha) = (1-\alpha)\,\|h\|^2 - \int_0^1 \int_0^1 h(s)h(t)\,G_\alpha(ds,dt), \quad h \in H.$$

For $0 \le s,\, t \le 1$,

$$G_\alpha(s,t) = \int_0^1 \int_0^1 \xi\eta W_0\Big(\|Z\|_{0,1} \le c_\alpha \Big| Z_s = \xi,\, Z_t = \eta\Big)\Big)\, \mathcal{L}(Z_s, Z_t|W_0)(d\xi, d\eta).$$

Proof: In view of

$$a_{KS}(h,\alpha) = (1-\alpha)\|h\|^2 - \int_{\|Z\|_{0,1} \le c_\alpha} L(h)^2, dW_0,$$

it suffices to establish

$$\int_{\|Z\|_{0,1} \le c_\alpha} L(h)^2 dW_0 = \int_0^1 \int_0^1 h(s)h(t)G_\alpha(ds,dt), \quad h \in H,$$

which is obvious for indicator functions of intervals. The rest is routine. □

Of course, such a representation of the curvature of the asymptotic power function also holds for all other EDF tests coming from a semi-norm which is W_0-a.e. continuous w.r.t. the sup norm. The measure defining function G_α is fundamental for our results concerning the two-sided KS test. Its values can be calculated numerically, which is the basis for the numerical approximation of eigenvalues and eigenvectors of $T_{KS,\alpha}$ (compare the proof of Theorem (2.8)). The following lemma serves this purpose.

(3.2) LEMMA

1. G_α *vanishes on the boundary of the unit square,*

$$G_\alpha(s,\cdot\,) \equiv G_\alpha(\,\cdot, t) \equiv 0, \quad 0 \le s, t \le 1,$$

and is symmetric about both of its diagonals:

$$G_\alpha(s,t) = G_\alpha(t,s) \quad \text{and} \quad G_\alpha(1-s, 1-t) = G_\alpha(t,s), \quad 0 \le s,\, t \le 1.$$

2. $G_\alpha(s,s) =$

$$\int_{-c_\alpha}^{c_\alpha} \xi^2\, W_0(\|Z\|_{0,s} \le c_\alpha | Z_s = \xi)\, W_0(\|Z\|_{s,1} \le c_\alpha | Z_s = \xi)\, \mathcal{N}(0, s(1-s))(d\xi)$$

and $G_\alpha(s,t)\ =$

$$\int_{-c_\alpha}^{c_\alpha} \int_{-c_\alpha}^{c_\alpha} \xi\, \eta\, W_0(\|Z\|_{0,s} \le c_\alpha | Z_s = \xi)\, W_0(\|Z\|_{s,t} \le c_\alpha | Z_s = \xi,\, Z_t = \eta)$$

$$\cdot W_0(h\|Z\|_{t,1} \le c_\alpha | Z_t = \eta) \mathcal{N}(0, \begin{pmatrix} s(1-s) & s(1-t) \\ s(1-t) & t(1-t) \end{pmatrix})(d\xi, d\eta) \quad (0 < s < t < 1).$$

3. $W_0(\|Z\|_{u,v} \le c \,|\, Z_u = \xi,\, Z_v = \eta) =$

$$1 - \sum_{m=1}^{\infty} \big(\exp(\frac{A_m}{u-v}) + \exp(\frac{B_m}{u-v}) \exp(\frac{C_m}{u-v}) + \exp(\frac{D_m}{u-v}) \big)$$

$(0 \le u \le v \le 1,\ c > 0,\ \xi, \eta \in [-c, c])$, *where*

$$\begin{aligned} A_m &= 2\big((2m-1)c - \xi\big)\big((2m-1)c - \eta\big), \\ B_m &= 2\big((2m-1)c + \xi\big)\big((2m-1)c + \eta\big), \\ C_m &= 2(2mc - \xi)(2mc + \eta) + \xi\eta, \\ D_m &= 2(2mc + \xi)(2mc - \eta) + \xi\eta. \end{aligned}$$

Proof: Part 1 is obvious. A generalization of Part 3 is posed as an exercise in Hájek and Šidák, [10], V.3, p. 199. For the proof of Part 2 let $0 < s < t < 1$. Since $(Z_r)_{0 \le r \le 1}$ and $(Z_{1-r})_{0 \le r \le 1}$ are both Markov processes w.r.t. their natural filtrations

$$\mathcal{F}_r := \sigma(Z_u \mid 0 \le u \le r) \quad \text{and} \quad \mathcal{G}_r := \sigma(Z_u \mid r \le u \le 1), \quad 0 \le r \le 1,$$

and since for Markov processes the past and the future are conditionally independent given the presence, we have for every $\xi, \eta \in \mathbb{R}$

$$\begin{aligned} & W_0(\{\|Z\|_{0,1} \le c_\alpha\}|Z_s = \xi, Z_t = \eta) \\ = \;& W_0(1_{\{\|Z\|_{0,t} \le c_\alpha\}} \cdot W_0(\|Z\|_{t,1} \le c_\alpha|\mathcal{F}_t) \,\big|\, Z_s = \xi, Z_t = \eta) \\ = \;& W_0(\|Z\|_{t,1} \le c_\alpha|Z_t = \eta) \cdot W_0(\|Z\|_{0,t} \le c_\alpha|Z_s = \xi, Z_t = \eta) \\ = \;& W_0(\|Z\|_{t,1} \le c_\alpha|Z_t = \eta) \cdot W_0(1_{\{\|Z\|_{s,t} \le c_\alpha\}} \\ & \cdot W_0(\|Z\|_{0,s}|\mathcal{G}_s) \,\big|\, Z_s = \xi, Z_t = \eta) \\ = \;& W_0(\|Z\|_{t,1} \le c_\alpha|Z_t = \eta) \cdot W_0(\|Z\|_{s,t} \le c_\alpha|Z_s = \xi, Z_t = \eta) \\ & \cdot W_0(\|Z\|_{0,s} \le c_\alpha|Z_s = \xi). \end{aligned}$$

Together with Theorem (3.1) this leads to the expression for $G_\alpha(s,t)$. □

The formulae in Parts 2 and 3 of Lemma (3.2) yield the values of G_α in terms of (iterated) integrals of smooth integrands over compact intervals. Thus, these values are easily available by standard methods of numerical integration such as (iterated) Romberg quadrature. For levels $\alpha = .05,\ \alpha = .01$ and $\alpha = .005$ G_α has been evaluated with absolute accuracy of at least eight decimals on a quadratic lattice with width .02 in the unit square. In view of the symmetry properties (3.2) (1) it suffices to calculate the function G_α for arguments belonging to the "discrete triangle" $\{(\frac{n}{50}, \frac{m}{50}) \mid 0 \le n \le m \le 50\}$. It is found that the shape of G_α does not depend much on the level α: In all three cases the graph has negatively parabola-shaped sections parallel to the main diagonal of the unit square and biparabolic sections with maximum on the main diagonal parallel to the second diagonal. Keeping in mind that

$$G_\alpha(s,t) = \int_{\|Z\|_{0,1} \le c_\alpha} Z_s Z_t \, dW_0, \qquad 0 \le s,\, t \le 1,$$

with $W_0\{\|Z\|_{0,1} \le c_\alpha\} = 1 - \alpha \approx 1$, it is not surprising that the shape of G_α is similar to the one of the covariance structure of the standard Brownian bridge. Figures 2 show threedimensional plots of the graph of $G_{.05}$.

Having tabulated G_α on the lattice $\{(\frac{n}{50}, \frac{m}{50}) \mid 0 \le n,\, m \le 50\}$, we may use Theorem (2.8) to calculate the curvature $a_{KS}(h, \alpha)$ exactly for directions h in

$$L := \{h \in H : h = \sum_{n=1}^{50} h(\frac{n}{50}) \cdot 1_{\left(\frac{n-1}{50}, \frac{n}{50}\right]}\},$$

the 49–dimensional subspace of the tangent space consisting of those step functions which are constant on all intervalls $\left(\frac{n-1}{50}, \frac{n}{50}\right],\ 1 \le n \le 50$. For other directions h the curvature may be approximated by discretizing via $D : H \to L$ defined by

$$D\colon h \mapsto \sum_{n=1}^{50} h\big(\frac{2n-1}{100}\big) \cdot 1_{\left(\frac{n-1}{50}, \frac{n}{50}\right]} \in L,$$

Figure 2: The measure defining function $G_{.05}$

leading to the approximation $a_{KS}\big(\frac{D(h)}{\|D(h)\|}, \alpha\big)$. (Other discretizations such as taking conditional expectations

$$50 \sum_{n=1}^{50} \int_{\frac{n-1}{50}}^{\frac{n}{50}} h(x)\, dx \cdot 1_{\left(\frac{n-1}{50}, \frac{n}{50}\right]}, \quad h \in H,$$

are of course also possible and yield essentially the same.)

Our next aim is the principal components decomposition for the KS test, partly provided in Table 4 and Figures 1. The following considerations, which lead to these results, are based on a finite-dimensional approximation of the operator $T_{KS,\alpha}$ defined in the proof of Theorem (2.8). We restrict the bilinear form $B_{KS,\alpha}$ to L, thus considering the operator $p_L \circ T_{KS,\alpha}|_L \colon L \to L$ instead of $T_{KS,\alpha}$ ($p_L : H \to L$ the orthogonal projection):

$$B_{KS,\alpha}(\cdot,\cdot)|_{L\times L} = \langle p_L \circ T_{KS,\alpha}, \cdot\rangle|_{L\times L}.$$

In view of the compactness of the operator $T_{KS,\alpha}$ established in Lemma (4.1), such an approximation seems to be reasonable since for “sufficiently large” L the norm of $T_{KS,\alpha}|_{L^\perp}$ must be small.

Let $A_\alpha := \big(a_{ij}^{(\alpha)}\big)_{1\le i,\, j\le 50}$ be the matrix whose values are the G_α-measures of the squares $(\frac{i-1}{50}, \frac{i}{50}] \times (\frac{j-1}{50}, \frac{j}{50}]$:

$$a_{ij}^{(\alpha)} := G_\alpha\big(\frac{i}{50}, \frac{j}{50}\big) - G_\alpha\big(\frac{i-1}{50}, \frac{j}{50}\big) - G_\alpha\big(\frac{i}{50}, \frac{j-1}{50}\big) + G_\alpha\big(\frac{i-1}{50}, \frac{j-1}{50}\big).$$

Denote

$$I \colon L \to \mathbb{R}_*^{50} : h \mapsto 50^{-1/2}\big(h\big(\frac{i}{50}\big)_{1\le i\le 50}\big).$$

Then I is an isometry, and we have

(3.3) LEMMA

1. $p_L \circ T_{KS,\alpha}|_L = I^{-1} \circ ((1-\alpha)\, Id - A_\alpha) \circ I$. *Hence,* $(1-\alpha)Id - A_\alpha$ *is positive semi-definite.* $\lambda \in \mathbb{R}$ *is an eigenvalue of* $p_L \circ T_{KS,\alpha}$ *with eigenvector* $h \in L$ *iff it is an eigenvalue of* $(1-\alpha)Id - A_\alpha$ *with eigenvector* $I(h) \in \mathbb{R}_*^{50}$.
2. $1-\alpha$ *is the largest eigenvalue of* $(1-\alpha)Id - A_\alpha$, $I(1)$ *a corresponding normalized eigenvector and* $\mathbb{R}_*^{50}$ *an invariant subspace.*
3. *If* $h \in L$ *solves*
$$a_{KS}(h,\alpha) \overset{!}{=} Max \quad \text{under} \quad \|h\| = 1, \tag{9}$$
then $h = I^{-1}(x)$ *where* $(1-\alpha)\, x - A_\alpha x = \lambda_{max}\, x$ *and* λ_{max} *is the second largest eigenvalue of* $(1-\alpha)Id - A_\alpha$.

Table 4: Values of h_1 for KS on $(\frac{n-1}{50}, \frac{n}{50}]$, $1 \leq n \leq 25$.

n	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$
1	1.230412	1.192409	1.180689
2	1.230412	1.192408	1.180689
3	1.230412	1.192408	1.180689
4	1.230406	1.192408	1.180689
5	1.230326	1.192408	1.180688
6	1.229844	1.192363	1.180674
7	1.228181	1.192121	1.180569
8	1.224130	1.191228	1.180108
9	1.216256	1.188847	1.178702
10	1.203112	1.183758	1.175359
11	1.183394	1.174463	1.168711
12	1.156028	1.159333	1.157115
13	1.120204	1.136755	1.138797
14	1.075385	1.105255	1.112004
15	1.021296	1.063596	1.075134
16	.957906	1.010846	1.026858
17	.885413	.946429	.966208
18	.804218	.870153	.892642
19	.714908	.782219	.806096
20	.618233	.683223	.076995
21	.515085	.574135	.596259
22	.406474	.456271	.475272
23	.293513	.331245	.345835
24	.177388	.200919	.210098
25	.059344	.067339	.070471

Proof: Part 1 is a direct consequence of Theorems (2.8) and (3.1).

Part 2: By the proof of Theorem (3.1) we have

$$\begin{aligned}\langle A_\alpha x, x\rangle &= \langle I^{-1}\circ A_\alpha \circ I(I^{-1}(x)), I^{-1}(x)\rangle \\ &= \int_0^1\int_0^1 I^{-1}(x)(s)\, I^{-1}(x)(t)\, G_\alpha(ds,dt) \\ &= \int\limits_{\{z \| \|z\|_{0,1}\le c_\alpha\}} L(I^{-1}(x))^2\, dW_0 \ge 0\,, \quad x\in \mathbb{R}^{50}_*.\end{aligned}$$

Hence, A_α is positive semi-definite and $(1-\alpha)Id - A_\alpha$ cannot have eigenvalues larger than $1-\alpha$. The rest of (2) follows immediately from the fact that

$$\sum_{j=1}^{50} a_{ij}^{(\alpha)} = 0, \qquad 1\le i\le 50.$$

Part 3: Let $h\in L$ be a solution of (9). Then $I(h)$ minimizes $\langle A_\alpha x, x\rangle,\ \ x\in\mathbb{R}^{50}_*$, under the constraints $\sum_{i=1}^{50} x_i = 0$ and $\sum_{i=1}^{50} x_i^2 = 1$. Hence, there are Lagrangian multipliers $\nu_1,\, \nu_2 \in \mathbb{R}$ satisfying

$$A_\alpha\circ I(h) = \mu_1 I(1) + \mu_2 I(h).$$

In view of Part 2 we must have $\nu_1 = 0$, which of course implies $\nu_2 > 0$. (Otherwise,

$$a_{KS}(h,\alpha) = \langle p_L\circ T_{KS}\alpha(h), h\rangle = \langle (1-\alpha)I(h) - A_\alpha\circ I(h), I(h)\rangle$$

$$\ge 1-\alpha > \sqrt{\frac{2}{\pi \mathrm{e}}} \ge -\sqrt{\frac{2}{\pi}}\, N_{\alpha/2}\exp(-\frac{1}{2}\, N_{\alpha/2}^2) = \lambda^{(1)}\,,$$

contradicting the optimality of the power function $g_{\{h\}}(th,\alpha)$ of the onedimensional NS test for $\mathcal{H} = \{h\}$.) Apparently, ν_2 must be chosen as small as possible. □

Generalizing Part 3 of Lemma (3.3) to $m\ge 2$, a normalized eigenvector for the m-th largest eigenvalue of $(1-\alpha)Id - A_\alpha$ maximizes $a_{KS}(h,\alpha)$ under the constraint $\|h\| = 1$ on the subspace of L which is orthogonal to the eigenspaces of all larger eigenvalues. In view of this Lemma our problem is now reduced to determining the spectral decomposition of A_α. Employing first Householder's transformation and then the *QL* Algorithm for symmetric tridiagonal matrices (see Wilkinson and Reinsch, [29], II/3) this has been done for levels $\alpha = .05$, $\alpha = .01$ and $\alpha = .005$ with an accuracy of 32 decimals. (So nothing of the 7 decimal-accuracy of the entries of A_α is lost.) Calculations have been performed using *FORTRAN V* translations of the *ALGOL 60* Procedures *tred1, tql1* and *trbak1* in Wilkinson and Reinsch, [29] (pp. 216–218 and 233/234). The eigenvalues thus obtained are (partly) given in Table 1, and the eigenvectors corresponding to the second largest eigenvalues (λ_1) — transformed back by I^{-1} — have been tabulated in 4. (Note that the functions are antisymmetric about $t = 1/2$, so the table may terminate with $n = 25$.) Figure 1 (solid lines) exhibits smoothed plots of eigenvectors belonging to λ_1, λ_2, λ_3 and λ_4

for the 5%–level. Plots of the eigenvectors for the 1%– and the .5%–level show that these are similar to those in the 5%–case: The eigensystem does not depend much on the level. This is, of course, not surprising since the shape of the measure defining function G_α was, for small level α, found to be nearly independent of α.

Because of Lemma (3.3), Part 3, the first eigenvector gives the locally optimal direction in the tangent space for the two-sided KS test: It is the direction in which the curvature of its power function is maximal.

4 Appendix: Proofs

Finally, we turn to the proofs of the results stated in Section 2.

Proof of Lemma (2.7) We restrict ourselves to indicating the essential steps of the manipulations which constitute the proof. Let $X_1, \ldots, X_k$ be i.i.d. random variables distributed as $\mathcal{N}(t/sqrtk, 1)$. From

$$\chi^2_{k,t^2}(c) = Prob\{X_1^2 + \ldots + X_k^2 \leq c\}$$

we obtain

$$\frac{d^2}{dt^2}\Big|_{t=0} \chi^2_{k,t^2}(c) = \frac{1}{(2\pi)^{k/2}k} \int\limits_{x_1^2+\ldots+x_k^2 \leq c} (\sum_{i=1}^k x_i)^2 \exp(-\frac{1}{2}\sum_{i=1}^k x_i^2)\, d\underline{x} \;-\; \chi^2_k(c)\,.$$

To evaluate the multiple integral, one first rotates the coordinate system so that the vector $(1, 1, \ldots, 1)^T$ moves to $(\sqrt{k}, 0, \ldots, 0)^T$ and then transforms to spherical coordinates. Thus, the first summand becomes

$$(2\pi)^{-k/2} \int_{r^2 \leq c} r^{k+1} \exp(-\frac{r^2}{2})\, dr \int_0^\pi \cos^2\vartheta \sin^{k+2}\vartheta\, d\vartheta \int_0^\pi \sin^{k+3}\vartheta\, d\vartheta \ldots \int_0^\pi \sin\vartheta\, d\vartheta$$

$$= \chi^2_{k+2}(c) \cdot \frac{2\pi \int_0^\pi \cos^2\vartheta \sin^{k-2}\vartheta\, d\vartheta}{\int_0^\pi \sin^k\vartheta d\vartheta \int_0^\pi \sin^{k-1}\vartheta d\vartheta \int_0^\pi \sin^{k-2}\vartheta\, d\vartheta}\ .$$

Evaluating these integrals one gets that the ratio is 1. □

Next,we will prove that the operator representing the curvature of the asymptotic power function in the sense of Theorem (2.8) is compact, provided the acceptance region of the test is based on a measurable semi-norm.

Let H be a Hilbert space, B be a separable Banach space and $\tau\colon H \to B$ be a bounded linear injection with dense range. By $\mathcal{N}_H$ we denote the standard normal distribution, which is a cylindrical measure on H. We assume that the image of $\mathcal{N}_H$ under τ is a probability measure P_0 on the Borel-σ-field $\mathcal{B}(B)$. Then (H, B, τ) is called an *Abstract Wiener space.*

There is a uniquely determined linear Gaussian process $(L(h))_{h\in H}$ on $(B,\mathcal{B}(B),P_0)$ satisfying $\mathcal{L}(L(h)|P_0)=\mathcal{N}(0,\|h\|^2)$ and $L(h)\circ\tau=\langle h,\cdot\rangle,\ h\in H$. We consider the bilinear form

$$B^{(c)}\colon (h_1,h_2)\longmapsto \int\limits_{\|\cdot\|_B>c} L(h_1)\,L(h_2)\,dP_0-\langle h_1,h_2\rangle\,P_0\{\|\cdot\|_B>c\}$$

$(c>0$ fixed), for which we state the following

(4.1) LEMMA *Let $(h_n)_{n\in\mathbf{N}}\subseteq H$ be a sequence converging to 0, weakly. Then*

$$\lim_{n\to\infty} B^{(c)}(h_n,h_n)=0\,.$$

It follows that the operator $T\colon H\to H$ defined by $B^{(c)}(h,\cdot\,)=:\langle Th,\cdot\rangle,\ h\in H$, is compact (cf. Dunford and Schwartz, [6], X, Ex. 2.5), which completes the proof of Theorem (2.8).

The proof of Lemma (4.1) is divided into several steps, in the first of which the integral is rewritten. Let $\mathcal{L}$ denote the set of finite-dimensional subsets of H, and $p_L\colon H\to L$, $p_{L,L_0}\colon L\to L_0$ ($\{L_0\subseteq L\}\subset\mathcal{L}$) the orthogonal projections.

(4.2) LEMMA *For every $h\in H$*

$$\int_{\|\cdot\|_B>c}(L(h))^2\,dP_0=\sup_{L\in\mathcal{L}}\int_{\|\tau|_L\|>c}\langle h,\cdot\rangle^2\,d\mathcal{N}_L\,.$$

Proof: It suffices to give a proof for $h\in\tau^*(B^*)$. Then $L(h)=x^*\in B^*$ is a linear map. From Strasser, [26], Theorem 73.5, it follows that

$$\int_{\|\cdot\|_B>c}(L(h))^2\,dP_0=\sup_{L\in\mathcal{L}}\int_{\|\tau|_L\|>c}(x^*\circ\tau)^2\,d\mathcal{N}_L\,.$$

Regarding $x^*\circ\tau=L(h)\circ\tau=\langle h,\cdot\rangle$, the proof is complete. □

The following considerations are based on the known fact that $h\mapsto\|\tau(h)\|_B$, $h\in H$, is a *measurable semi-norm*, meaning that for every $\varepsilon>0$ there is a $L_\varepsilon\in\mathcal{L}$ such that $L\perp L_\varepsilon$ implies

$$\mathcal{N}_H\{\|\tau\circ p_L\|_B>\varepsilon\}<\varepsilon$$

(which follows e.g. from Dudley et al., [5], Theorems 2 and 3). This can be used to derive lower and upper bounds for the integrals on the right hand side in Lemma (4.2).

(4.3) LEMMA *For every $\varepsilon>0$ and every $L\in\mathcal{L}$ such that $L\supset L_\varepsilon$,*

$$\int\limits_{\|\tau\circ p_{L,L_\varepsilon}\|_B>c+\varepsilon}\langle h,\cdot\rangle^2\,d\mathcal{N}_L-\sqrt{3\varepsilon}\,\|h\|^2\le\int\limits_{\|\tau|_L\|_B>c}\langle h,\cdot\rangle^2\,d\mathcal{N}_L$$

$$\leq \int_{\|\tau\circ p_{L,L_\varepsilon}\|_B > c-\varepsilon} \langle h,\cdot\rangle^2 \, d\mathcal{N}_L + \sqrt{3\varepsilon}\,\|h\|^2$$

Proof: Decomposing $\tau|_L = \tau\circ p_{L,L_\varepsilon} + \tau\circ p_{L,L\cap L_\varepsilon^\perp}$ it follows that

$$\{\|\tau|_L\|_B > c\} \subset \{\|\tau\circ p_{L,L_\varepsilon}\|_B > c-\varepsilon\} \cup \{\|\tau\circ p_{L,L\cap L_\varepsilon^\perp}\|_B > \varepsilon\}$$

which together with the preceding remarks and the Cauchy-Schwartz inequality implies

$$\int_{\|\tau|_L\|_B > c} \langle h,\cdot\rangle^2 \, d\mathcal{N}_L \leq \int_{\|\tau\circ p_{L,L_\varepsilon}\|_B > c-\varepsilon} \langle h,\cdot\rangle^2 \, d\mathcal{N}_L + \Big(\varepsilon \int x^4 \, \mathcal{N}(0,\|h\|^2)(dx)\Big)^{1/2} .$$

Analogously, the lower bound is established. □

(4.4) LEMMA *Let* $L_0 \in \mathcal{L},\ L\in\mathcal{L},\ L\subset L_0$ *and* $d>0$. *Then*

$$\Big| \int_{\|\tau\circ p_{L,L_0}\|_B > d} \langle h,\cdot\rangle^2 \, d\mathcal{N}_L - \mathcal{N}_H\{\|\tau\circ p_{L_0}\|_B > d\}\cdot \|p_L h\|^2 \Big| \leq 4\,\|h\|\,\|p_{L_0}h\| .$$

Proof: Denoting $L_1 := L\cap L_0^\perp$, elementary manipulations yield

$$\int \big|\langle h,\cdot\rangle^2 - \langle p_{L_1}h,\cdot\rangle^2\big| \, d\mathcal{N}_L$$

$$\leq \int \langle p_{L_0}h,\cdot\rangle^2 \, d\mathcal{N}_L + 2\int \big|\langle p_{L_0}h,\cdot\rangle\langle p_{L_1}h,\cdot\rangle\big| \, d\mathcal{N}_L \leq 3\,\|h\|\,\|p_{L_0}h\| .$$

By Fubini's theorem we obtain

$$\begin{aligned} &\int_{\|\tau\circ p_{L,L_0}\|_B > d} \langle p_{L_1}h,\cdot\rangle^2 \, d\mathcal{N}_L \\ &\quad= \mathcal{N}_{L_0}\{\|\tau|_{L_0}\|_B > d\}\cdot \int_{L_1} \langle p_{L_1}h,\cdot\rangle^2 \, d\mathcal{N}_{L_1} \\ &\quad= \mathcal{N}_H\{\|\tau\circ p_{L_0}\|_B > d\}\cdot \big(\|p_L h\|^2 - \|p_{L_0}h\|^2\big) \end{aligned}$$

which, combined with the above inequality, yields the assertion. □

Proof of Lemma (4.1) Fix $\varepsilon > 0$ and $L,\ L_0$ in $\mathcal{L}$ such that $L_\varepsilon \subseteq L_0 \subseteq L$. Combining Lemmas (4.3) and (4.4) we obtain

$$\int\limits_{\|\tau|_L\|_B > c} \langle h, \cdot \rangle^2 \, d\mathcal{N}_L \leq \mathcal{N}_H\{\|\tau|_{L_0}\|_B > c - \varepsilon\} \cdot \|p_L h\|^2 + \sqrt{3\varepsilon}\,\|h\|^2 + 4\,\|h\|\,\|p_{L_0} h\|\,.$$

Passing to the sup over $L \in \mathcal{L}$ on both sides, Lemma (4.2) then yields

$$\begin{aligned} & \int\limits_{\|\cdot\|_B > c} (L(h))^2 \, dP_0 \\ \leq\ & \mathcal{N}_H\{\|\tau|_{L_0}\|_B > c - \varepsilon\} \cdot \|h\|^2 + \sqrt{3\varepsilon}\,\|h\|^2 + 4\,\|h\|\,\|p_{L_0} h\| \\ \leq\ & P_0\{\|\cdot\|_B > c - \varepsilon\} \cdot \|h\|^2 + \sqrt{3\varepsilon}\,\|h\|^2 + 4\,\|h\|\,\|p_{L_0} h\|\,. \end{aligned}$$

Inserting a sequence (h_n) which converges weakly to zero, we obtain

$$\limsup_{n\to\infty} \Big(\int\limits_{\|\cdot\|_B > c} L(h_n))^2 \, dP_0 - P_0\{\|\cdot\|_B > c - \varepsilon\}\,\|h_n\|^2 \Big) \leq \sqrt{3\varepsilon}\, \sup_{n\in\mathbb{N}} \|h_n\|^2\,.$$

Similarly, we obtain

$$\liminf_{n\to\infty} \Big(\int\limits_{\|\cdot\|_B > c} L(h_n))^2 \, dP_0 - P_0\{\|\cdot\|_B > c + \varepsilon\}\,\|h_n\|^2 \Big) \geq -\sqrt{3\varepsilon}\, \sup_{n\in\mathbb{N}} \|h_n\|^2\,.$$

Letting $\varepsilon \to 0$, completes the proof. □

References

[1] Anderson, T.W. and D.A. Darling (1952): *Asymptotic theory of certain "Goodness of fit" criteria based on stochastic processes.* Ann. Math. Statist. **23**, 193–212.

[2] Anderson, T.W. and D.A. Darling (1954): *A test of goodness of fit.* J. Amer. Statist. Assoc. **49**, 765–769.

[3] Donsker, M.D. (1952): *Justification and extension of Doob's heuristic approach to the Kolmogorov- Smirnov theorems.* Ann. Math. Statist. **23**, 277–281.

[4] Doob, J.L. (1949): *Heuristic approach to the Kolmogorov-Smirnov theorems.* Ann. Math. Statist. **20**, 393–403.

[5] Dudley, R.M., J. Feldman and L. Le Cam (1971): *On seminorms and probabilities, and abstract Wiener spaces.* Ann. Math. Ser. 2, **93**, 372–390.

[6] Dunford, N. and J.T. Schwartz (1957): *Linear Operators, Part II.* Interscience, New York.

[7] Durbin, J. (1973): *Distribution Theory for Tests Based on the Sample Distribution function.* SIAM, Philadelphia.

[8] Durbin, J. and M. Knott (1972): *Components of Cramér-von Mises Statistics I.* J. Roy. Statist. Soc. Ser. B, **34**, 290–307.

[9] Durbin, J., M. Knott and C.C. Taylor (1975): *Components of Cramér-von Mises Statistics II.* J. Roy. Statist. Soc. Ser. B, **37**, 216–237.

[10] Hájek, J. and Z. Šidák (1967): *Theory of Rank Tests.* Academic Press, New York.

[11] Johnson, N.L. and S. Kotz (1970): *Continuous univariate distributions – 2.* Houghton Mifflin, Boston.

[12] Koshevnik, Yu.A. and B.Ya. Levit (1976): *On a nonparametric analogue of the information matrix.* Theor. Prob. Appl. **21**, 738–753.

[13] Locke C. and J.D. Spurrier (1978): *On tests of uniformity.* Commun. Statist.–Theor. Meth. **A7**, 241–258.

[14] Mason, J.H. and D.M. Schuenemeyer (1983): *A modified Kolmogorov-Smirnov test sensitive to tail alternatives.* Ann. Statist. **11**, 933–946.

[15] Neuhaus, G. (1976): *Asymptotic Power Properties of the Cramér-von Mises Test Under Contiguous Alternatives.* J. Multiv. Anal. **6**, 95–110.

[16] Neuhaus, G. (1979): *Asymptotic Theory of Goodness of Fit Tests when Parameters are Present: A Survey.* Math. Operationsforschung Statist., Ser. Statist. **10**, 479–494.

[17] Neuhaus, G. (1986): *A Class of Quadratic Goodness of Fit Tests.* Preprint **86–11**, Institut für Mathematische Stochastik der Universität Hamburg.

[18] Neyman, J. (1937): *"Smooth test" for goodness of fit.* Skand. Aktuarietidskr. **20**, 149–199.

[19] Pfanzagl, J. with the Assistance of W. Wefelmeyer (1982): *Contributions to a General Asymptotic Statistical Theory.* Lecture Notes Stat. **13**, Springer, New York.

[20] Quesenberry, C.P. and F.L. Miller Jr. (1977): *Power Studies of Some Tests for Uniformity.* J. Statist. Comput. Simul. **5**, 169–191.

[21] Quesenberry, C.P. and F.L. Miller Jr. (1979): *Power Studies of Some Tests for Uniformity II.* Commun. Statist.–Simula. Computa. **B8**, 271–290.

[22] Schoenfeld, D.A. (1977): *Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramér-von Mises Statistic.* Ann. Statist. **5**, 1017–1026.

[23] Shorack, G.R. and J.A. Wellner (1986): *Empirical processes with applications to statistics.* Wiley, New York.

[24] Solomon, H. and M.A. Stephens (1983): *On Neyman's Statistic for Testing Uniformity.* Commun. Statist.–Simula. Computa. **12**, 127–134.

[25] Stephens, M.A. (1974): *EDF Statistics for Goodness of Fit and Some Comparisons.* J. Amer. Statist. Assoc. **69**, 730–737.

[26] Strasser, H. (1985): *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory.* de Gruyter, Berlin.

[27] Strasser, H. (1988): *Lokale Zerlegung der Gütefunktionen von Anpassungstests und ihre Interpretation.* Forschungsbericht, Universität Bayreuth.

[28] Strasser, H. (1988): *Global Extrapolation of Local Efficiency.* Submitted for publication.

[29] Wilkinson, J. and C. Reinsch (1971): *Handbook for Automatic Computation II, Linear Algebra.* Springer, Berlin.

Hartmut Milbrodt
Fachbereich Mathematik
Universität Dortmund
Postfach 50 05 00
D-4600 Dortmund 50
West Germany

Helmut Strasser
Institut für Mathematik
Universität Bayreuth
Postfach 10 12 51
D-8580 Bayreuth
West Germany