Adaptive Detection of Known Signals in Additive Noise by Means of Kernel Density Estimators

Rolf T. Gustafsson, Ola G. Hössjer, and Tommy Öberg, *Member, IEEE*

***Abstract*— We consider the problem of detecting known signals contaminated by additive noise with a completely unknown probability density function f. To this end, we propose a new adaptive detection rule. It is defined by plugging a kernel density estimator $\hat{f}$ of f into the maximum *a posteriori* (MAP) detector. The estimate $\hat{f}$ can either be computed off-line from a training sequence or on-line simultaneously with the detection. For the off-line detector, we prove that the (asymptotic) error probability for weak signals converges to the minimal error probability of the MAP detector as the number of training data tends to infinity, and we also establish rates of convergence and the optimal choice of bandwidth order for a certain class of noise densities. In a Monte Carlo study, the off-line plug-in MAP detectors are compared with the L^1- and L^2-detectors for various noise distributions. When the training sequence is long enough, the plug-in detectors have excellent performance for a wide range of distributions, whereas the L^2-detector breaks down for heavy-tailed distributions and the L^1-detector for distributions with little mass around the origin.**

***Index Terms*—Adaptive, additive noise, detection, kernel estimate, nonparametric, training sequence.**

I. Introduction

We consider the problem of detecting one out of a finite number of possible messages of known form, transmitted through a channel which is corrupted by additive noise.

In many situations, the noise is clearly non-Gaussian due to impulsive sources, especially when there are a few external interferring noise sources with high intensity. It is well known that even a small deviation from the normal distribution can drastically degrade the performance of the linear detectors that are optimal for the Gaussian environment. This has created interest in robust detectors that have nearly optimal performance for Gaussian noise *and* good performance when a fraction of the noise is impulsive, cf., e.g., [12] and [13]. Such detectors are based on the assumption that the majority of samples have a known nominal distribution, whereas a small fraction can have a more or less arbitrary distribution.

An alternative approach is to assume little or no *a priori* information about the noise statistics and then estimate the noise probability density function (pdf) $f \in \mathcal{F}$ either from a training sequence (off-line) or simultaneously with the detection (on-line). This can be done either parametrically or nonparametrically, depending on whether $\mathcal{F}$ is finite-dimensional or infinite-dimensional. A widely used physical parametric model is Middleton's Class A model [15]–[17]. Estimators of the parameters in this model have been considered in [27] and [28].

In this paper, we will follow the nonparametric approach, and only impose mild regularity conditions on f. We estimate f by means of a truncated kernel density estimator. The resulting estimate $\hat{f}$ is then considered as the true pdf and plugged into the maximum *a posteriori* detector (MAP). This "plug-in" detector belongs to the class of minimum distance M-detectors. As the performance criterion of the detector ϕ, we will use the asymptotic error probability (or risk) $\bar{P}(\phi)$ for weak signals, as in [10]. This criterion depends essentially only on the efficacy, as noted in [14] in the case of two signals. Since the efficacy is the inverse of the asymptotic variance of an M-estimator, our detection problem has analogies to M-estimation.

Statistically, our detection problem is semiparametric. The detection of the (weak) signal is a parametric problem, whereas the noise pdf can be regarded as an infinite-dimensional nuisance parameter f. The statistical study of semiparametric problems started with the fundamental paper of Stein [22]. He called an estimator of a Euclidean parameter that does not assume knowledge of f *adaptive* when it is asymptotically efficient for each $f \in \mathcal{F}$. This means that the estimator has the same asymptotic performance as the optimal estimator with f known. A recent, very extensive survey of semiparametric methods is given in [4].

Let ϕ_0 be the optimal (unknown) MAP detector associated with f, and $\hat{\phi}_n$ the plug-in MAP detector based on $\hat{f}_n$, which is computed from a training sequence of size n (the off-line case). Then adaptiveness means that $\overline{P}(\hat{\phi}_n) \xrightarrow{p} \overline{P}(\phi_0)$ as n tends to infinity for all $f \in \mathcal{F}$. We will refer to this property as *consistency* of $\hat{\phi}_n$, since $\overline{P}(\hat{\phi}_n)$ can be viewed as an estimator of $\overline{P}(\phi_0)$. In order to have a consistent detector, it is crucial to have a good estimate $\hat{\psi}_n$ of the optimal score function $\psi_0 = -f'/f$. In fact, we will show that $\overline{P}(\hat{\phi}_n) - \overline{P}(\phi_0)$ depends essentially on $\int (\hat{\psi}_n - \psi_0)^2 f\, dy$. This loss function is also important for the theory of adaptive M-estimation in the linear model (cf. [3], [4], [6], [11], [18], and [23]).

Under very weak conditions on $\mathcal{F}$, we prove consistency of the off-line detector. Related results for M-estimators can be found in [3] and [4, Sec. 7.8]. We also establish a rate of convergence of $n^{-4/9}$ for $\overline{P}(\hat{\phi}_n)$. Even though this requires a

Manuscript received April 10, 1995; revised June 8, 1996.

R. T. Gustafsson is with Responsor AB, Oddegatan 5, S-164 92 Kista, Sweden.

O. G. Hössjer is with the Department of Mathematical Statistics, Lund University, Box 118, S-221 00 Lund, Sweden.

T. Öberg is with the Signals and Systems Department, Uppsala University, Box 528, S-751 20 Uppsala, Sweden.

Publisher Item Identifier S 0018-9448(97)03445-7.

slightly smaller class $\mathcal{F}$ (e.g., f must have three derivatives), it still includes the Cauchy distribution, all t-distributions, the normal and logistic distributions, and finite mixtures of these. To our knowledge this is the first result on rates of convergence of a semiparametric procedure based on kernel estimates.

We have chosen M-detectors and kernel density estimators because of their simple structure. In fact, our density estimator only contains two parameters, a bandwidth and a truncation point (to avoid tail effects when estimating ψ_0), and is easy to compute using the so-called WARP techniques (Weighted Average of Rounded Points) described in [9].

Other adaptive techniques that have been studied include R-estimators and R-tests [1], [4], [24]; L-estimators [19], [20]; and estimators based on minimizing the Hellinger distance [2]. Kernel estimates have been used in [2], [3], [4], [18], and [23] for estimating f and ψ_0 in the semiparametric context, but other density estimation methods could also be used, such as orthogonal series [1] and splines [5], [6], [11]. In fact, Cox establishes a rate of convergence $n^{-2/3}$ of $\int (\hat{\psi}_n - \psi_0)^2 f\, dy$ towards 0 in [5], assuming a third derivative and periodicity of f. This is faster than our rate $n^{-4/9}$, but our regularity conditions seem to cover more standard densities.

As one referee noted, it may be more realistic to consider the error probability $P_t(\phi)$ for a fixed signal length t instead of the asymptotic limit $\overline{P}(\phi) = \lim_t P_t(\phi)$. The rate at which $P_t(\hat{\phi}_n)$ converges to $P_t(\phi_0)$ is then of interest. Alternatively, to get a simpler performance criterion, an exponential bound of $P_t(\phi)$ could be used. This is indeed an interesting topic for further research. Our limit $\overline{P}$ may be viewed as an explicit and relatively simple approximation of P_t.

The paper is organized as follows: We review the theory for detection of weak signals in Section II; the off-line detector and its asymptotic properties are considered in Sections III and IV. We present an on-line detector in Section V with recursive updating of $\hat{f}$. Finally, numerical results are given in Section VI, and the proofs are collected in the Appendices.

II. Detection Problem

A. Model and Optimal Detector

We consider a received vector of the form

$$\boldsymbol{X} = \theta \boldsymbol{s}_k + \boldsymbol{N} \tag{2.1}$$

where

$$\boldsymbol{X} = (X_1, \cdots, X_t)$$

and

$$\boldsymbol{s}_k = (s_{k1}, \cdots, s_{kt}), \qquad k = 1, \cdots, M$$

is one of M possible transmitted signals, θ is an amplitude factor, and $\boldsymbol{N} = (N_1, \cdots, N_t)$ is an additive noise vector. We assume that the N_i are independent and indentically distributed (i.i.d.) random variables with pdf f, and that the signals have *a priori* probabilities $\pi_1, \cdots, \pi_M$. The error probability of a nonrandomized decision rule $\phi: \mathbb{R}^t \to 1, \cdots, M$ is then given by

$$P_t(\phi) = \sum_{k=1}^{M} \pi_k P(\phi(\boldsymbol{X}) \neq k | \boldsymbol{s}_k).$$

It is well known [25, p. 48] that the error probability is minimized by the maximum *a posteriori* detector

$$\phi_0(\boldsymbol{X}) = \arg \min_{1 \le k \le M} \left(-\log \pi_k - \sum_{i=1}^{t} \log f(X_i - \theta s_{ki}) \right).$$

A minimum-distance detector is defined by first choosing a distance function $d: \mathbb{R}^t \to \mathbb{R}$ and then selecting the signal with shortest distance to the received signal

$$\phi(\boldsymbol{X}) = \arg \min_{1 \le k \le M} d(\boldsymbol{X} - \theta \boldsymbol{s}_k). \tag{2.2}$$

For a minimum-distance M-detector (MDM), the distance is given by

$$d(\boldsymbol{x}) = \sum_{i=1}^{t} \rho(x_i) \tag{2.3}$$

with $\boldsymbol{x} = (x_1, \cdots, x_t)$ and $\rho: \mathbb{R} \to \mathbb{R}$ a real-valued function. For equal *a priori* probabilities $\pi_k = 1/M$, the MAP detector ϕ_0 belongs to this class with $\rho_0 = -\log f$. The L^2-detector corresponds to $\rho(x) = x^2$ and Gaussian noise, whereas the L^1-detector with $\rho(x) = |x|$ corresponds to Laplacian noise $(f(x) = \exp(-|x|)/2)$.

B. Asymptotic Error Probability and Efficacy

In the weak signal approach, the amplitude depends on the signal length

$$\theta = \theta_t = \frac{C}{\sqrt{t}} \tag{2.4}$$

for some constant $C > 0$. For weak signals, the asymptotic error probability

$$\overline{P}(\phi) = \lim_{t \to \infty} P_t(\phi)$$

usually exists, and has a tractable expression for MDM detectors and equal *a priori* probabilities. (Strictly speaking, $\overline{P}$ depends on the whole sequence of decision rules as $t \to \infty$, but this will not be shown in the notation.) In practice, the signal amplitude θ is constant. The weak signal approach is merely a tool for finding a simple expression for P_t, which may be approximated by $\overline{P}$, with $C = \theta\sqrt{t}$.

Under certain assumptions, $\overline{P}$ is closely related to the efficacy

$$\mathcal{E}(\psi) = \frac{\left(\int \psi' f\, dy \right)^2}{\int \psi^2 f\, dy} = \frac{\left(\int \psi \psi_0 f\, dy \right)^2}{\int \psi^2 f\, dy} := \frac{B(\psi)^2}{A(\psi)}.$$

Here, $\psi = \rho'$ (analogous to the score function for M-estimators) and $\psi_0 = \rho_0' = -f'/f$ is the optimal score function corresponding to the MAP detector. Before exploring the relation between the efficacy and the asymptotic error probability, let us state some regularity conditions:

Assume that

ia) The density f is absolutely continuous with $0<I(f)<\infty$, where $I(f)=\int (f')^2/f\,dy$ is the Fisher information.

ii) Let $\boldsymbol{S}_t=(\boldsymbol{s}_1',\cdots,\boldsymbol{s}_M')$ be a $t\times M$ signal matrix. Then $\boldsymbol{S}_t'\boldsymbol{S}_t/t\rightarrow\Sigma$ as $t\rightarrow\infty$, where $\Sigma=(\sigma_{ij})$ is a positive-semidefinite symmetric matrix of dimension $q\le M$.

iiia) $0<A(\psi)<\infty$.

iiib) ψ is discontinuous at most at a finite number of points and has a bounded derivative outside these points.

iva) $\int \psi f\,dy=0$.

ivb) For some subsequence $\mathbb{T}=\{t_j\}$ of $\mathbb{N}$ and for $k=1,\cdots,M$

$$\sum_{i=1}^{t_j} s_{ki}=0.$$

Condition iva) is satisfied when f is symmetric and ψ skew-symmetric. However, the symmetry is not necessary according to ivb), as long as all signals $\boldsymbol{s}_k$ have zero mean. In this case, we consider the asymptotic error probability as the limit along the subsequence $\mathbb{T}$

$$\overline{P}_{\mathbb{T}}(\phi)=\lim_{\mathbb{T}\ni t\rightarrow\infty}P_t(\phi).$$

The detection problem (2.1) can naturally be embedded into a multiple linear regression model [10], $\boldsymbol{X}=\boldsymbol{\theta}\boldsymbol{S}_t'+\boldsymbol{N}$, with the signal matrix $\boldsymbol{S}_t$ in ii) interpreted as the design matrix and $\boldsymbol{\theta}\in\mathbb{R}^M$ the unknown regression parameter. If $\theta\boldsymbol{s}_k$ is the transmitted signal, then $\boldsymbol{\theta}=\theta\boldsymbol{e}_k$, and $\boldsymbol{e}_k=(0,\cdots,0,1,0,\cdots,0)$, the unit vector with 1 in position k. When ivb) holds, $\boldsymbol{\theta}$ only contains slope parameters. It is well known that symmetry of f is not needed for estimating the slope parameters, whereas some condition like iva) is necessary when an intercept is included in the model.

Notice that ii) implies that the pairwise signal distances converge

$$||\theta_t\boldsymbol{s}_j-\theta_t\boldsymbol{s}_k||^2_{L^2(\mathbb{R}^t)}\rightarrow C^2(\sigma_{jj}+\sigma_{kk}-2\sigma_{jk})\text{ as } t\rightarrow\infty. \tag{2.5}$$

It is possible to find vectors $\boldsymbol{m}_1,\cdots,\boldsymbol{m}_M$ in $\mathbb{R}^q$ whose pairwise distances agree with those in (2.5) (cf. [10]). Define

$$\Omega_k=\{\boldsymbol{x}\in\mathbb{R}^q;||\boldsymbol{x}-\boldsymbol{m}_k||<||\boldsymbol{x}-\boldsymbol{m}_j||\;\forall j\neq k\}$$

as the set of points closest to $\boldsymbol{m}_k, k=1,\cdots,M$. Let $\boldsymbol{I}_q$ be the $q\times q$ identity matrix and $\boldsymbol{Z}(\mathcal{E})\sim\boldsymbol{N}_q(0,\mathcal{E}^{-1}\boldsymbol{I}_q)$ a normally distributed stochastic vector. Define the "error probability function"

$$G(\mathcal{E})=\frac{1}{M}\sum_{k=1}^{M}P(\boldsymbol{m}_k+\boldsymbol{Z}(\mathcal{E})\neq\Omega_k).$$

It is easily seen that G is differentiable and strictly monotone-decreasing. We then have the following theorem, which follows by combining [10, Lemma 1 and Propositions 1–2]:

Theorem 2.1: Assume equal *a priori* probabilities, ia), ii), iiia)–iiib), and iva). Then the asymptotic error probability exists

$$\overline{P}(\phi)=G(\mathcal{E}(\psi)) \tag{2.6}$$

or if ivb) holds instead of iva)

$$\overline{P}_{\mathbb{T}}(\phi)=G(\mathcal{E}(\psi)). \tag{2.7}$$

A more complicated asymptotic risk expression can also be derived for arbitrary *a priori* probabilities, cf. [10]. We state Theorem 2.1 here to stress the dependence of $\overline{P}$ on the efficacy. When we estimate ψ_0 (or ρ_0) in the following sections, it is thus important how well $\mathcal{E}(\psi_0)$ can be approximated. The function G (and hence also $\overline{P}$) depends not only on $\mathcal{E}$, but also on the matrix Σ and the amplitude factor C. Since we consider the signals s_k and C known, we have made only the dependence on $\mathcal{E}$ explicit. It may be seen by orthogonal transformations that G only depends on $\{\boldsymbol{m}_1,\cdots,\boldsymbol{m}_M\}$ through the pairwise distances $||\boldsymbol{m}_j-\boldsymbol{m}_k||$. This makes it possible to choose $\boldsymbol{m}_k$ in a convenient way.

Example 1($M=2$.): The case of two signals is thoroughly treated in the overview [13]. Suppose $\boldsymbol{s}_1=\boldsymbol{0}$ (the zero vector) and that $\boldsymbol{s}_2$ is arbitrary with $0<\sigma_{22}<\infty$ and

$$\sigma_{22}=\lim_{t\rightarrow\infty}\sum_{i=1}^{t}s_{2i}^2/t.$$

Then $q=1$, and we may choose $m_1=0$ and $m_2=C\sqrt{\sigma_{22}}$. This implies

$$G(\mathcal{E})=1-\Phi(C\sqrt{\mathcal{E}\sigma_{22}}/2)$$

with Φ the cumulative distribution function of a standard normal distribution. In particular, for the signal

$$s_{2i}=\frac{1}{\sqrt{1+\tau^2}}(1+\tau(-1)^i),\qquad i=1,\cdots,t \tag{2.8}$$

with $\tau\in\mathbb{R}$, we have $\sigma_{22}=1$ and hence

$$G(\mathcal{E})=1-\Phi(C\sqrt{\mathcal{E}}/2).$$

Example 2 (Sinusoidal Signals): Fix $t_0\ge 3$, let $\mathbb{T}=\{t_0,2t_0,3t_0,\cdots\}, M\ge 3$, and

$$s_{ki}=\sqrt{2}\sin\left(\frac{2\pi i}{t_0}+\frac{k2\pi}{M}\right),\qquad i=1,\cdots,t \tag{2.9}$$

for some $t\in\mathbb{T}$ (a sampled version of MPSK signals). Then $q=2$, and we may choose

$$\boldsymbol{m}_k=C(\cos(k2\pi/M),\sin(k2\pi/M)).$$

When $M=4$, G has a simple expression

$$G(\mathcal{E})=1-\Phi(C\sqrt{\mathcal{E}/2})^2.$$

III. Off-Line Detector

Suppose that we have an i.i.d. training sequence $\tilde{\boldsymbol{N}} = (\tilde{N}_1, \cdots, \tilde{N}_n)$ of noise samples with marginal distribution f. This means that the recipient knows the sent message corresponding to the first n samples or that no signal is sent ($\boldsymbol{s} = \boldsymbol{0}$) during this time. Our objective is to use $\tilde{\boldsymbol{N}}$ for estimation of f. This in turn produces an estimate of $\rho_0 = -\log f$ that is plugged into the MAP detector.

Let $h = h_n \to 0$ be a given sequence of numbers and K a kernel function. Define

$$\hat{f}_n(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - \tilde{N}_i}{h}\right)$$

as a kernel estimate of f with bandwidth h. Various properties of kernel estimates can be found in [21] and [26]. Next define estimators of ρ_0 and ψ_0 by

$$\hat{\rho}_n(y) = \begin{cases} -\log \hat{f}_n(y), & \text{if } |y| \le a_n \\ -\log \hat{f}_n(a_n), & \text{if } |y| > a_n \end{cases}$$

$$\hat{\psi}_n(y) = \hat{\rho}'_n(y) = \begin{cases} -\dfrac{\hat{f}'_n(y)}{\hat{f}_n(y)}, & \text{if } |y| \le a_n \\ 0, & \text{if } |y| > a_n. \end{cases}$$

Let also $\hat{\phi}_n$ be the "plug-in" MAP detector defined by putting $\rho = \hat{\rho}_n$ in (2.2) and (2.3). The truncation point $a = a_n$ may either be infinite or finite, where in the latter case $a_n \to \infty$ as n increases. It is actually preferable to choose a_n finite to avoid tail effects, as will be seen in Section IV. We will first check whether Theorem 2.1 holds with $\psi = \hat{\psi}_n$, when n and h are fixed. For this we need some regularity assumptions:

va) K is strictly positive

vb) K is differentiable with K'/K bounded.

vc) K' is continuous except in a finite set $\mathbb{D}_0$. Outside this set K' is differentiable with K''/K bounded.

These requirements are violated by the Gaussian kernel

$$K(u) = \exp(-u^2/2)/\sqrt{2\pi}$$

and kernels with compact support, such as the Epanechnikov kernel

$$K(u) = 3(1 - u^2)1_{|u| \le 1}/4.$$

Instead, we will use the logistic kernel

$$K(u) = \exp(-u)/(1 + \exp(-u))^2$$

in all of our simulations.

Lemma 3.1: Assume va)–vc) and $0 < a \le \infty$, then

$$||\hat{\psi}_n||_{L^\infty(\mathbb{R})} \le ||K'/K||_{L^\infty(\mathbb{R})} h^{-1}$$

the set $\mathbb{D}_n$ of discontinuities of $\hat{\psi}_n$ is finite (with a cardinality that might depend on n) and[1]

$$||\hat{\psi}'_n||_{L^\infty(\mathbb{R} \setminus \mathbb{D}_n)} \le (||K''/K||_{L^\infty(\mathbb{R}\setminus\mathbb{D}_0)} + ||K'/K||^2_{L^\infty(\mathbb{R})}) h^{-2}.$$

[1] Given $A \subset \mathbb{R}$ and $\psi \in L^\infty(\mathbb{R})$, we define

$$\psi_{L^\infty(A)} = \operatorname*{ess\,sup}_{x \in A} |\psi(\chi)|.$$

Lemma 3.1 implies iiia)–iiib) for fixed n and h. In conjunction with Theorem 2.1 this gives:

Corollary 3.2: Assume ia), ii), ivb), and va)–vc). Then the asymptotic error probability of $\hat{\phi}_n$ exists and is given by

$$\overline{P}_{\mathbb{T}}(\hat{\phi}_n) = G(\mathcal{E}(\hat{\psi}_n)). \tag{3.1}$$

When the signals have nonzero mean, we need iva), which is typically not satisfied for $\psi = \hat{\psi}_n$. However, iva) holds if

ib) f is symmetric

and ψ skew-symmetric. It is possible to modify $\hat{\psi}_n$ to obtain a skew-symmetric score function. Define

$$\begin{aligned} \overline{\rho}_n(y) &= \tfrac{1}{2}\left(\hat{\rho}_n(y) + \hat{\rho}_n(-y)\right) \\ \overline{\psi}_n(y) &= \overline{\rho}'_n(y) = \tfrac{1}{2}\left(\hat{\rho}'{}_n(y) - \hat{\rho}'{}_n(-y)\right) \end{aligned} \tag{3.2}$$

and let $\overline{\phi}_n$ be the corresponding detector. Clearly, $\overline{\rho}_n$ is symmetric and $\overline{\psi}_n$ skew-symmetric. Analogously to Proposition 3.2 we have:

Proposition 3.3: Assume ia)–ib), ii), and va)–vc). Then the asymptotic error probability of $\overline{\phi}_n$ exists

$$\overline{P}(\overline{\phi}_n) = G(\mathcal{E}(\overline{\psi}_n)). \tag{3.3}$$

IV. Asymptotics of the Off-Line Detector

A. Convergence of the Efficacy

In order to see how well $\overline{P}_{\mathbb{T}}(\hat{\phi}_n)$ and $\overline{P}(\overline{\phi}_n)$ approximate $\overline{P}(\phi_0)$, Theorem 2.1 tells us that we should investigate how well $\mathcal{E}(\hat{\psi}_n)$ and $\mathcal{E}(\overline{\psi}_n)$ approximate $\mathcal{E}(\psi_0)$. It is easy to see, using the Cauchy–Schwarz inequality, that $\psi = \psi_0$ maximizes $\mathcal{E}(\psi)$. This means that $\mathcal{E}(\psi_0) - \mathcal{E}(\hat{\psi}_n)$ and $\mathcal{E}(\psi_0) - \mathcal{E}(\overline{\psi}_n)$ will always be nonnegative, and we can regard these quantities as loss functions when we estimate ψ_0. Since G is a strictly decreasing function, $\overline{P}(\hat{\phi}_n) - \overline{P}(\phi_0)$ and $\overline{P}(\overline{\phi}_n) - \overline{P}(\phi_0)$ will also be nonnegative.

Define the inner product

$$\langle \psi_1, \psi_2 \rangle = \int \psi_1 \psi_2 f \, dy$$

the (squared) norm

$$\begin{aligned} ||\psi||^2 &= \langle \psi, \psi \rangle \\ \hat{\varepsilon}_n &= \hat{\psi}_n - \psi \end{aligned}$$

and

$$\overline{\varepsilon}_n = \overline{\psi}_n - \psi.$$

Proposition 4.1: If $||\hat{\varepsilon}_n|| \to 0$ as $n \to \infty$

$$\begin{aligned} \mathcal{E}(\psi_0) - \mathcal{E}(\hat{\psi}_n) &= ||\hat{\varepsilon}_n||^2 - \left\langle \hat{\varepsilon}_n, \frac{\psi_0}{||\psi_0||} \right\rangle^2 + O(||\hat{\varepsilon}_n||^3) \\ &= \int \hat{\varepsilon}_n^2 f \, dy - \frac{\left(\displaystyle\int \hat{\varepsilon}_n f' \, dy\right)^2}{I(f)} + O(||\hat{\varepsilon}_n||^3). \end{aligned} \tag{4.1}$$

In particular, if $||\hat{\varepsilon}_n|| \xrightarrow{p} 0$, then[2]

$$\mathcal{E}(\psi_0) - \mathcal{E}(\hat{\psi}_n) \le ||\hat{\varepsilon}_n||^2(1 + o_p(1)) \tag{4.2}$$

and if (3.1) holds, then

$$\overline{P}_{\mathbb{T}}(\hat{\phi}_n) - \overline{P}_{\mathbb{T}}(\phi_0) \le |G'(\mathcal{E}(\psi_0))|^2 ||\hat{\varepsilon}_n||^2(1 + o_p(1)). \tag{4.3}$$

If (3.3) is satisfied, the same conclusions hold with $\hat{\psi}_n, \hat{\varepsilon}_n$ and $\overline{P}_{\mathbb{T}}$ replaced by $\overline{\psi}_n, \overline{\varepsilon}_n$ and $\overline{P}$ in (4.1)–(4.3).

Remark: If we ignore the remainder terms, Pythagoras' Theorem implies that $\mathcal{E}(\psi_0) - \mathcal{E}(\hat{\psi}_n)$ equals the distance between $\hat{\psi}_n$ and the line $H = \{c\psi_0; c \in \mathbb{R}\}$. This is natural, since $\mathcal{E}(c\psi_0) = \mathcal{E}(\psi_0)$ for all c.

B. Consistency

In this section, we will prove convergence of the asymptotic error probability. According to Proposition 4.1, we should first investigate the quantity

$$||\hat{\varepsilon}_n||^2 = \int (\hat{\psi}_n - \psi_0)^2 f\, dy.$$

We will need the following additional regularity assumptions.

vd) The kernel K satisfies $\mu_1(K) = 0$ and $\mu_2(K) < \infty$, where

$$\mu_j(K) = \textstyle\int u^j K(u)\, du.$$

via) $h_n \to 0$.

vib) $a_n \to \infty$.

vic) $a_n h_n^{-3} n^{-1} \to 0$.

Theorem 4.2: Assume ia), va)–vb), vd), and via)–vic). Then

$$\int (\hat{\psi}_n - \psi_0)^2 f\, dy \xrightarrow{p} 0$$

as $n \to \infty$.

Convergence of

$$||\overline{\varepsilon}_n||^2 = \int (\overline{\psi}_n - \psi_0)^2 f\, dy$$

is implied by Theorem 4.2 and the following result, which may be found in, e.g., [3] and [4]:

Proposition 4.3: If f is symmetric (Condition ib)), then

$$\int (\overline{\psi}_n - \psi_0)^2 f\, dy \le \int (\hat{\psi}_n - \psi_0)^2 f\, dy.$$

Combining Propositions 3.2, 3.3, 4.1, 4.3, and Theorem 4.2, we obtain consistency of $\overline{\phi}_n$ and $\hat{\phi}_n$:

Corollary 4.4: Assume ia), ii), ivb), va)–vd), and via)–vic). Then $\hat{\phi}_n$ is consistent, i.e.,

$$\overline{P}_{\mathbb{T}}(\hat{\phi}_n) \xrightarrow{p} \overline{P}(\phi_0)$$

as $n \to \infty$. If ia)–ib), ii), va)–vd), and via)–vic) hold, then $\overline{\phi}_n$ is consistent, i.e.,

$$\overline{P}(\overline{\phi}_n) \xrightarrow{p} \overline{P}(\phi_0).$$

We do not need $\int K\, du = 1$, since multiplicative factors of $\hat{f}_n$ will not affect the decision rules. Conditions vib)–vic) imply that the truncation points a_n tend to infinity, but with a slower rate than nh_n^3 (and hence $h_n \gg n^{-1/3}$).

[2] $Y_n = o_p(1)$ means that $P(|Y_n| > \varepsilon) \to 0$ for any $\varepsilon > 0$.

C. Rates of Convergence and Bandwidth Selection

By adopting more assumptions on K, f, a_n, and h_n, it is possible to establish rates of convergence of $\overline{P}_{\mathbb{T}}(\hat{\phi}_n)$ and $\overline{P}(\overline{\phi}_n)$. These assumptions are:

ic) f'' is absolutely continuous

$$\int (f'/f)^6 f\, dy < \infty \qquad \int |f''/f|^3 f\, dy < \infty$$

and

$$\int (f^{(3)}/f)^2 f\, dy < \infty.$$

id)

$$\int_{|y|>x} (f'/f)^2 f\, dy = O(x^{-2})$$

as $x \to \infty$.

ie) Let

$$\overline{f}_\delta(y) = \int_{y-\delta}^{y+\delta} f\, dx/(2\delta).$$

Then there exists $\delta_0 > 0$ and $0 < c_0 < 1$ such that

$$\overline{f}_\delta(y) \ge c_0 \overline{f}_{\delta'}(y), \quad \forall 0 < \delta' < \delta \le \delta_0$$

and $\forall y \in \mathbb{R}$.

ve) There exists a number $u_0 > 0$ such that $-u^2 K'(u)$ is nonincreasing for $u \ge u_0$.

vid) $a_n = C_1 h_n^{-2}$ for some constant $C_1 > 0$.

vie) $h_n = C_2 n^{-1/9}$ for some constant $C_2 > 0$.

Condition ic) is equivalent to

ic}' ψ_0' is absolutely continuous

$$\int \psi_0^6 f\, dy < \infty \qquad \int |\psi_0'|^3 f\, dy < \infty$$

and

$$\int (\psi_0'')^2 f\, dy < \infty.$$

Theorem 4.5: Assume ia), ic), ie), va)–vb), vd)–ve), and via). Then

$$E\left(\int (\hat{\psi}_n - \psi_0)^2 f\, dy\right) = O\left(h_n^4 + n^{-1}h_n^{-3}a_n + \int_{|y|>a_n} (f'/f)^2 f\, dy\right). \tag{4.4}$$

If also id) and vid) are satisfied, then

$$E\left(\int (\hat{\psi}_n - \psi_0)^2 f\, dy\right) = O(h_n^4 + n^{-1}h_n^{-5}) \tag{4.5}$$

and then choosing the bandwidths according to vie) gives the optimal rate of convergence

$$E\left(\int (\hat{\psi}_n - \psi_0)^2 f\, dy\right) = O(n^{-4/9}). \tag{4.6}$$

Here h_n^4 is essentially the integrated squared bias of $\hat{\psi}_n$. It has the same order of magnitude as the squared integrated bias when estimating f or f'. The second term of (4.4),

$n^{-1}h_n^{-3}a_n$, is an integrated variance term of $\hat{\psi}_n$ over the interval $[-a_n, a_n]$. Compare this with the pointwise variance of $\hat{\psi}_n(y)$, which is $O(n^{-1}h_n^{-3})$ for each $y \in \mathbb{R}$, a *smaller* order as $a_n \to \infty$. It is the presence of f in the denominator of ψ_0 that makes the integrated variance larger. There is no such discrepancy when f or f' are estimated, then the local and integrated variances are of the same order ($O(n^{-1}h_n^{-1})$ and $O(n^{-1}h_n^{-3})$, respectively). The third term of (4.4) is the error induced by the truncation of $\hat{\rho}_n$ at $\pm a_n$. The magnitude of this terms depends on the behaviour of the tails of f. We could have required a faster rate of convergence than $O(x^{-2})$ in id). This would make $\int_{|y|>a_n} (f'/f)^2 f\, dy$ smaller, and then we could have chosen smaller values of a_n and h_n, giving a faster rate than $O(n^{-4/9})$ in (4.6). However, a fast rate of convergence in id) would exclude heavy-tailed densities like the Cauchy distribution and t-distributions with few degrees of freedom. Therefore, apart from the assumption that $f^{(3)}$ exists, the regularity assumptions ia) and ic)–ie) imposed on f are rather weak. For instance, they are satisfied by the normal, logistic, Cauchy and all t-distributions and also by finite mixtures of these distributions.

The bandwidth choice vie) is optimal for our chosen class of densities. It is of *larger* order than the typical ones for estimation of f ($h_n \sim n^{-1/5}$) and f' ($h_n \sim n^{-1/7}$). As mentioned above, if we restricted ourselves to lighter tailed densities, the optimal bandwith would be of a smaller order than $n^{-1/9}$.

Combining Theorem 4.5 with Propositions 3.2, 3.3, 4.1, and 4.3 we obtain

Corollary 4.6: Assume ia), ic)–ie), ii), ivb), va)–ve), and vid)-vie). Then

$$\overline{P}_{\mathbb{T}}(\hat{\phi}_n) - \overline{P}(\phi_0) = O_p(n^{-4/9}).$$

If ia)–ie), ii), va)–ve), and vid)–vie) hold, then

$$\overline{P}(\overline{\phi}_n) - \overline{P}(\phi_0) = O_p(n^{-4/9}).$$

V. On-Line Detector

Suppose now that the transmission procedure (2.1) is repeated L times

$$\boldsymbol{X}_m = \theta \boldsymbol{s}_{k_m} + \boldsymbol{N}_m, \qquad m = 1, 2, \cdots, L \tag{5.1}$$

with $\boldsymbol{X}_m = (X_{m1}, \cdots, X_{mt})$, and all noise vectors $\boldsymbol{N}_m = (N_{m1}, \cdots, N_{mt})$ are independently distributed, each having i.i.d. components with marginal density f. The numbers $k_1, \cdots, k_L \in \{1, \cdots, M\}$ determine the true signals during the L transmission intervals.

The detector will be updated recursively and *simultaneously* with the detection as follows: To choose the signal, use the MDM detector $\tilde{\phi}_{m-1}$ described in (2.2) and (2.3) with $\rho = \tilde{\rho}_{m-1}$. This produces a signal estimate $\boldsymbol{s}_{\hat{k}_m}$ which can be used to estimate $\boldsymbol{N}_m$ according to

$$\hat{\boldsymbol{N}}_m = \boldsymbol{X}_m - \theta \boldsymbol{s}_{\hat{k}_m} = \boldsymbol{N}_m + \theta(\boldsymbol{s}_{k_m} - \boldsymbol{s}_{\hat{k}_m}) := (\hat{N}_{m1}, \cdots, \hat{N}_{mt}).$$

Since we now have more information about the noise, we will use $\hat{\boldsymbol{N}}_m$ to recursively update $\tilde{\rho}_{m-1}$. Analogously to the off-line case, $\tilde{\rho}_m$ is computed from a density estimate $\tilde{f}_m$ according to

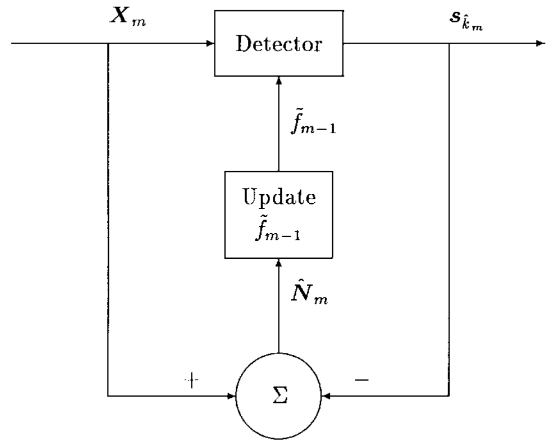

Fig. 1. Flowchart of the on-line detector.

$$\tilde{\rho}_m(y) = \begin{cases} -\log \tilde{f}_m(y), & \text{if } |y| \le \tilde{a}_m \\ -\log \tilde{f}_m(\tilde{a}_m), & \text{if } |y| > \tilde{a}_m \end{cases}$$

where $\tilde{a}_m$ is an appropriate sequence of truncation numbers. The density estimate $\tilde{f}_m$ is computed from $\tilde{f}_{m-1}$ according to

$$\tilde{f}_m(y) = (1 - \gamma_m)\tilde{f}_{m-1}(y) + \gamma_m \check{f}_m(y) \tag{5.2}$$

where

$$\check{f}_m(y) = \frac{1}{t\tilde{h}_m} \sum_{i=1}^{t} K\left(\frac{y - \hat{N}_{mi}}{\tilde{h}_m}\right)$$

$\tilde{h}_m$ is the bandwidth in step m, and $0 < \gamma_m \le 1$ measures how much faith we put in the latest noise estimates relative to the past information. The structure of the on-line detector is shown in Fig. 1.

If we have n training samples $(\tilde{N}_1, \cdots, \tilde{N}_n)$ available, it is natural to use

$$\tilde{f}_0(y) = \hat{f}_n(y)$$

and

$$\gamma_m = \frac{t}{tm + n}. \tag{5.3}$$

This means that we put equal emphasis on all noise samples (both from the training sequence and the estimated ones), since (5.2) may now be expanded as

$$\begin{aligned} \tilde{f}_m(y) &= \frac{n}{tm+n}\tilde{f}_0(y) + \frac{t}{tm+n}\sum_{j=1}^{m}\check{f}_j(y) \\ &= \frac{1}{tm+n}\left(\sum_{i=1}^{n}\frac{1}{h_n}K\left(\frac{y-\tilde{N}_i}{h_n}\right)\right. \\ &\quad \left. + \sum_{j=1}^{m}\sum_{i=1}^{t}\frac{1}{\tilde{h}_j}K\left(\frac{y-\hat{N}_{ji}}{\tilde{h}_j}\right)\right). \end{aligned}$$

Hence, $\tilde{f}_m$ is based on a noise vector of size $tm+n$. Using the asymptotic theory of Section IV-C, it is natural to choose

$$\tilde{h}_m = C_2(tm+n)^{-1/9} \quad \text{and} \quad \tilde{a}_m = C_1\tilde{h}_m^{-2}$$

TABLE I

WITH SIGNAL CONFIGURATION S1 AND VARIOUS NOISE DISTRIBUTIONS F, ESTIMATES OF $P_t(\phi)$ ARE GIVEN FOR THE L_1-DETECTOR ϕ_1, THE L_2-DETECTOR ϕ_2, AND THE PLUG-IN MAP DETECTOR $\overline{\phi}_n$, WITH n THE LENGTH OF THE TRAINING SEQUENCE (NO TRAINING SEQUENCE IS NEEDED FOR ϕ_1 AND ϕ_2). THREE REPLICATES OF EACH PLUG-IN MAP DETECTOR ARE CONSIDERED. THE ASYMPTOTIC LIMITS $\overline{P}(\cdot)$ ARE GIVEN IN BRACKETS (WHICH CORRESPOND TO $t=\infty$ FOR ϕ_1 AND ϕ_2 AND $n,\ t=\infty$ FOR THE PLUG-IN MAP DETECTOR, i.e., $\overline{P}(\phi_0)$)

F	Error probabilities in %													
	ϕ_1	$\bar{P}(\phi_1)$	ϕ_2	$\bar{P}(\phi_2)$	$\bar{\phi}_{100}$	$\bar{\phi}_{100}$	$\bar{\phi}_{100}$	$\bar{\phi}_{1000}$	$\bar{\phi}_{1000}$	$\bar{\phi}_{1000}$	$\bar{\phi}_{10000}$	$\bar{\phi}_{10000}$	$\bar{\phi}_{10000}$	$\bar{P}(\phi_0)$
F1	4.40	(5.53)	2.30	(2.28)	3.19	3.58	3.01	2.48	3.00	2.74	2.37	2.31	2.31	(2.28)
F2	6.08	(7.10)	13.39	(13.90)	6.14	6.00	6.74	4.79	4.85	4.34	4.01	3.88	3.86	(3.70)
F3	7.24	(8.64)	2.56	(2.55)	2.57	3.03	3.62	2.36	2.44	2.38	2.29	2.25	2.27	(2.13)
F4	33.18	(-)	44.57	(-)	43.71	37.94	38.54	43.47	46.03	45.07	43.28	43.78	43.79	(-)
F5	14.52	(15.87)	4.18	(4.16)	0.13	0.06	0.10	0.03	0.04	0.03	0.02	0.01	0.02	(0.00)
F6	3.18	(2.28)	7.74	(7.86)	11.34	13.94	4.58	4.34	4.12	3.66	3.44	3.61	3.42	(2.28)
F7	9.75	(10.15)	43.19	(50.00)	14.62	11.04	19.67	10.16	9.58	10.75	8.61	8.65	8.70	(7.86)
F8	6.19	(7.08)	11.32	(12.41)	11.34	12.82	10.04	6.66	7.66	6.99	5.62	5.63	5.64	(5.12)

for some positive constants C_1 and C_2. (When the error probability is small, $\hat{\boldsymbol{N}}_m = \boldsymbol{N}_m$ with high probability, and the situation is then very similar to the one described in Sections III and IV.)

If no training sequence is available we have to use a deterministic $\tilde{f}_0$, e.g., the logistic or the double exponential distribution. We can still use (5.3), and then $n \geq 0$ can be interpreted as the *a priori* confidence we have in $\tilde{f}_0$.

For symmetric noise and signals with nonzero mean, we can symmetrize $\tilde{\rho}_m$, as was done in Section III for the off-line case.

In situations where the noise is nonstationary, it is advisable to use

$$\gamma_m \equiv \gamma$$

which produces a detector that is more flexible to fast changes in f. On the other hand, the resulting detector is not consistent, when the data is stationary. The reason is that the effective number of samples used for calculating $\tilde{f}_m$ is of the order t/γ for all large m, which does not increase with m.

VI. NUMERICAL RESULTS

In this section, we perform a Monte Carlo study, which compares the error probability of the off-line detector $\hat{\phi}_n$ with the L^1- and L^2-detectors, for different noise distributions and signals. Let us first give a more detailed description of the simulation study:

Detectors: L^1-detector $(\phi_1), L^2$-detector $(\phi_2), \hat{\phi}_{100}, \hat{\phi}_{1000}, \hat{\phi}_{10000}, \overline{\phi}_{100}, \overline{\phi}_{1000}$, and $\overline{\phi}_{10000}$.

Noise Distributions: The following noise distributions will be used:

F1: Normal: $N(0,1)$.

F2: Normal mixture with heavy tails:

$$0.9N(0,1) + 0.1N(0,5^2).$$

F3: Normal mixture with several modes:

$$0.8N(0,1) + 0.1N(-1,0.5^2) + 0.1N(1,0.5^2).$$

F4: Normal non-symmetric mixture: $0.8N(0,1) + 0.2N(2,1)$.

F5: Uniform: $U[-2,2], f(y) = 1_{|y|\leq 2}/4$.

F6: Laplace: $f(y) = \exp(-|y|)/2$.

F7: Cauchy distribution.

F8: t-distribution with three degrees of freedom.

Signals: Two sets of signals are included:

S1: The signals from Example 1 $(M = 2)$, (2.8), with $\tau = 0.5, t = 100$, and $C = 4$.

S2: The signals from Example 2, (2.9), with $M = 4, t = 100$, and $C = 4$.

Number of Monte Carlo Iterations: 100 000 iterations for each combination of detector, signal, and noise. For each Monte Carlo replicate, the true signal was chosen randomly with equal probability $1/M$ among all M possible signals.

Calculation of $\hat{\phi}_n$: In practice, it is advisable to standardize the data and replace $\tilde{\boldsymbol{N}}$ by $(\tilde{\boldsymbol{N}} - \hat{\boldsymbol{\mu}})/\hat{s}$, with $\hat{\boldsymbol{\mu}} = (\hat{\mu}, \cdots, \hat{\mu})$. Here $\hat{\mu}$ and $\hat{s}$ are robust measures of location and scale. We have used the median and normalized interquartile range. That is, if $\tilde{N}_{(1)} \leq \cdots \leq \tilde{N}_{(n)}$ are the ordered training samples, we put $\hat{\mu} = \tilde{N}_{([0.5n])}$, and

$$\hat{s} = (\tilde{N}_{([0.75n])} - \tilde{N}_{([0.25n])})/(2\Phi(0.75) - 1).$$

When we have Gaussian noise, the normalization ensures that $\hat{s}$ is a consistent estimator of the standard deviation for a normal distribution. The next step is to compute $\hat{f}_{n,st}$ and $\hat{\rho}_{n,st}$ from the standardized noise samples, with $C_1 = 1/20$ and $C_2 = 1/5$ in vid)–vie) and a logistic kernel. (The subscript "*st*" indicates "standardized data.") Then transform back and define $\hat{\rho}_n(\cdot) = \hat{\rho}_{n,st}((\cdot - \hat{\mu})/\hat{s})$. The symmetrized version $\overline{\rho}_n$ is then computed from $\hat{\rho}_n$ according to (3.2).

The results of the Monte Carlo simulations are shown in Tables I–III. Notice that the symmetrized detector $\overline{\phi}_n$ is used for S1, for which ivb) does not hold. We have also included asymptotic error probabilities for various signals, detectors, and noise distributions. (Apart from the combination F4/S1, where neither ib), iva), nor ivb) hold.)

The agreement between the Monte Carlo results and the asymptotic limits is quite good for $\phi_2, \overline{\phi}_{10000}$ and $\hat{\phi}_{10000}$. On the other hand, for ϕ_1 the asymptotic error probability is somewhat higher than the simulated one. This discrepancy was also noted in [10] and is probably due to the discontinuity of the ψ-function, which makes the convergence of P_t towards $\overline{P}$ rather slow in Theorem 2.1. The detectors $\overline{\phi}_{1000}, \overline{\phi}_{10000}, \hat{\phi}_{1000}$, and $\hat{\phi}_{10000}$ show excellent performance for all noise distributions. However, it is clear from the tables that a training sequence

TABLE II

WITH SIGNAL CONFIGURATION S2 AND VARIOUS NOISE DISTRIBUTIONS F, ESTIMATES OF $P_t(\phi)$ ARE GIVEN FOR THE L_1-DETECTOR ϕ_1, THE L_2-DETECTOR ϕ_2, AND THE PLUG-IN MAP DETECTOR $\hat{\phi}_n$, WITH n THE LENGTH OF THE TRAINING SEQUENCE (NO TRAINING SEQUENCE IS NEEDED FOR ϕ_1 AND ϕ_2). THREE REPLICATES OF EACH PLUG-IN MAP DETECTOR ARE CONSIDERED. THE ASYMPTOTIC LIMITS $\overline{P}(\cdot)$ ARE GIVEN IN BRACKETS (WHICH CORRESPOND TO $t=\infty$ FOR ϕ_1 AND ϕ_2 AND $n,\ t=\infty$ FOR THE PLUG-IN MAP DETECTOR, i.e., $\overline{P}(\phi_0)$)

F	Error probabilities in %												
	ϕ_1	$\bar{P}(\phi_1)$	ϕ_2	$\bar{P}(\phi_2)$	$\hat{\phi}_{100}$	$\hat{\phi}_{100}$	$\hat{\phi}_{100}$	$\hat{\phi}_{1000}$	$\hat{\phi}_{1000}$	$\hat{\phi}_{1000}$	$\hat{\phi}_{10000}$	$\hat{\phi}_{10000}$	$\hat{\phi}_{10000}$	$\bar{P}(\phi_0)$
F1	1.47	(2.39)	0.48	(0.47)	1.89	2.65	2.72	0.72	0.69	0.80	0.51	0.48	0.48	(0.47)
F2	2.56	(3.75)	11.85	(12.11)	11.71	7.40	5.09	1.93	1.79	2.29	1.43	1.33	1.34	(1.15)
F3	3.41	(5.32)	0.58	(0.58)	0.69	0.95	1.32	0.81	0.77	0.70	0.52	0.46	0.46	(0.41)
F4	4.24	(6.10)	2.64	(3.47)	3.88	4.50	5.64	2.53	2.84	2.97	2.24	2.18	2.13	(2.02)
F5	12.43	(15.11)	1.39	(1.43)	0.01	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.00	(0.00)
F6	1.07	(0.47)	4.46	(4.50)	3.00	6.42	4.54	1.56	1.74	2.12	1.31	1.48	1.24	(0.47)
F7	6.71	(7.05)	63.44	(75.00)	20.30	21.90	10.41	6.52	7.46	9.46	5.67	5.62	5.64	(4.50)
F8	2.88	(3.72)	8.94	(9.98)	8.26	10.64	11.37	3.74	3.40	3.77	2.41	2.69	2.64	(2.08)

TABLE III

WITH SIGNAL CONFIGURATION S1 AND VARIOUS NOISE DISTRIBUTIONS F, ESTIMATES OF $P_t(\phi)$ ARE GIVEN FOR THE L_1-DETECTOR $\phi_{\mu 1}$, THE L_2-DETECTOR $\phi_{\mu 2}$, AND THE PLUG-IN MAP DETECTOR $\overline{\phi}_{\mu n}$, WITH n THE LENGTH OF THE TRAINING SEQUENCE ($n = 1000$ FOR ϕ_1 AND ϕ_2). THREE REPLICATES OF EACH PLUG-IN MAP DETECTOR ARE CONSIDERED

F	Error probabilities in %										
	$\phi_{\mu 1}$	$\phi_{\mu 2}$	$\bar{\phi}_{\mu 100}$	$\bar{\phi}_{\mu 100}$	$\bar{\phi}_{\mu 100}$	$\bar{\phi}_{\mu 1000}$	$\bar{\phi}_{\mu 1000}$	$\bar{\phi}_{\mu 1000}$	$\bar{\phi}_{\mu 10000}$	$\bar{\phi}_{\mu 10000}$	$\bar{\phi}_{\mu 10000}$
F1	4.05	2.30	21.30	4.74	11.95	3.37	4.25	3.05	2.46	2.46	2.39
F2	20.28	13.54	5.45	9.68	6.67	4.75	4.45	5.61	3.97	3.98	3.86
F3	3.91	2.75	16.39	18.18	8.98	3.72	3.84	4.17	2.20	2.21	2.28
F4	20.04	20.31	11.21	30.93	12.84	14.13	11.38	9.16	11.31	11.99	10.84
F5	12.96	6.69	32.31	13.31	26.28	1.01	0.04	0.33	0.01	0.07	0.02
F6	12.44	10.04	9.32	6.88	8.59	5.16	4.78	3.71	3.47	3.53	3.58
F7	44.29	43.08	33.71	19.17	19.07	9.89	9.77	11.16	8.45	8.92	8.72
F8	15.90	11.48	18.49	11.60	7.51	7.16	6.44	7.11	5.86	5.91	5.69

of 100 is too small for the plug-in MAP detector when the noise distribution has heavy tails (F2, F7, F8). The L^2-detector deteriorates for heavy-tailed distributions and the L^1-detector for distributions with little mass around the origin (F3–F5).

As expected, all detectors significantly degrade in performance for the nonsymmetric noise distribution F4 and signal S1 (which has a nonzero sum). To reduce this effect we symmetrized the ρ-functions around $\hat{\mu}$ instead of 0. This means

$$\begin{aligned} \rho_{\mu L^1}(y) &= |y-\hat{\mu}| \\ \rho_{\mu L^2}(y) &= (y-\hat{\mu})^2 \\ \overline{\rho}_{\mu n}(y) &= \frac{1}{2}\left(\hat{\rho}_{n,st}\left(\frac{y-\hat{\mu}}{\hat{s}}\right) + \hat{\rho}_{n,st}\left(\frac{\hat{\mu}-y}{\hat{s}}\right)\right) \end{aligned}$$

where the subscript μ indicates symmetrization around $\hat{\mu}$. We denote the corresponding detectors by $\phi_{\mu 1}, \phi_{\mu 2}$, and $\overline{\phi}_{\mu n}$. The results are shown in Table III. The performance with F4 is now improved, especially for $\overline{\phi}_{\mu n}$. The price for this is increased failure rate for many of the other distributions. This is especially apparent for $\phi_{\mu 1}$ and $\overline{\phi}_{\mu 100}$, whereas for $\phi_{\mu 2}, \overline{\phi}_{\mu 1000}$, and $\overline{\phi}_{\mu 10000}$ the difference is smaller.

To summarize, the plug-in MAP detectors have good performance for a wide range of distributions when the training sequence is long enough, whereas ϕ_1 and ϕ_2 have a more variable performance. It is also worthwile to adjust ϕ_2 and the plug-in MAP detectors for possible asymmetry in the noise when the training sequence is large and at least one signal has nonzero mean.

APPENDIX I
PROOFS FROM SECTIONS II AND III

Proof of Theorem 2.1

Equation (2.6) is proved in [10], assuming ia), ii), iiia), iva), and the additional requirements

$$\begin{aligned} &\int \psi(y+\delta) f(y)\, dy = \delta B(\psi) + o(\delta) \\ &\int (\psi(y+\delta) - \psi(y))^2 f(y)\, dy = o(1) \\ &\int (\rho(y+\delta+\eta) - \rho(y+\delta-\eta) - 2\eta\psi(y+\delta))^2 f(y)\, dy \\ &\qquad = o(\delta^2+\eta^2) \\ &\int (\rho(y+\delta+\eta) - \rho(y+\delta-\eta) - 2\eta\psi(y+\delta)) f(y)\, dy \\ &\qquad = o(\delta^2+\eta^2) \end{aligned} \tag{A1}$$

as $\delta, \eta \to 0$. By inspecting the proof of [10, Lemma 1], one sees that iva) may be replaced by ivb). Therefore, it suffices to prove that iiib) implies (A1). We will concentrate on last equation, the other three are easier to establish. Let $\mathbb{D} = \{d_1, \cdots, d_r\}$ be the finite set of discontinuities of ψ. Then ivb) implies

$$\psi(y) = \psi_1(y) + \sum_{i=1}^{r} b_i \operatorname{sgn}(y-d_i)$$

with ψ_1 having a bounded derivative. By linearity, it suffices to establish the last equation of (A1) for ψ_1 and $\psi_2 = \operatorname{sgn}(\cdot - d)$ separately. Let ρ_i be a primitive function of ψ_i (take, e.g., $\rho_2 = |\cdot - d|$), and $L(\lambda) = \operatorname{sgn}(\lambda)(1-|\lambda|)_+$—a function supported on $[-1,1]$. Taylor expansion of ρ gives

$$\begin{aligned}\Bigg|\int_{\mathbb{R}} &(\rho_1(y+\delta+\eta) - \rho_1(y+\delta-\eta) - 2\eta\psi_1(y+\delta))\\ &\cdot f(y)\,dy\Bigg|\\ &= \left|\int_{\mathbb{R}}\int_{-1}^{1} L(\lambda)(\psi_1'(y+\delta+\lambda\eta) - \psi_1'(y+\delta))\right.\\ &\left.\cdot f(y)\,d\lambda\,dy\right|\eta^2\\ &\le \int_{-1}^{1}|L(\lambda)|\int_{\mathbb{R}}|\psi_1'(y+\delta+\lambda\eta) - \psi_1'(y+\delta)|f(y)\,dy\,d\lambda\\ &\cdot\eta^2 = o(\eta^2)\end{aligned}$$

where the last relation holds provided we show that

$$\begin{aligned}\int_{\mathbb{R}} &|\psi_1'(y+\delta+\lambda\eta) - \psi_1'(y+\delta)|f(y)\,dy\\ &\le \int_{\mathbb{R}}|\psi_1'(y+\lambda\eta) - \psi_1'(y)|f(y)\,dy\\ &\quad + \int_{\mathbb{R}}|\psi_1'(y+\delta+\lambda\eta) - \psi_1'(y+\delta)|\\ &\quad \cdot|f(y+\delta) - f(y)|\,dy := I_1 + I_2\end{aligned}$$

tends to zero uniformly in λ as δ, η tend to zero. This is true for I_1 because of L^1-continuity. For the second term, let C_3 be a large constant. Then

$$\begin{aligned}I_2 &\le 2\|\psi_1'\|_{L^\infty(\mathbb{R})}\left(\int_{|y|\le C_3} + \int_{|y|>C_3}\right)\\ &\quad\cdot|f(y+\delta) - f(y)|\,dy\\ &\le 4\|\psi_1'\|_{L^\infty(\mathbb{R})}\left(C_3\omega_f(\delta) + \int_{|y|>C_3} f(y)\,dy\right)\end{aligned}$$

with ω_f the modulus of continuity of f. The last expression can be made arbitrarily small by first choosing C_3 large enough and then δ small enough. (Remember that f is absolutely continuous.)

For ρ_2 we obtain

$$\begin{aligned}\Bigg|\int_{\mathbb{R}} &(|y-d+\delta+\eta| - |y-d+\delta-\eta|\\ &- 2\eta\operatorname{sgn}(y-d+\delta))f(y)\,dy\Bigg|\\ &= 2\eta^2\left|\int_{-1}^{1} L(\lambda)(f(d-\delta+\eta\lambda) - f(d-\delta))\,d\lambda\right|\\ &= o(\eta^2)\end{aligned}$$

because of the continuity of f. □

Proof of Lemma 3.1

Notice that

$$\begin{aligned}\hat\psi_n &= -\frac{\hat f_n'}{\hat f_n}\\ \hat\psi_n' &= \left(\frac{\hat f_n'}{\hat f_n}\right)^2 - \frac{\hat f_n''}{\hat f_n}\end{aligned}$$

where $\hat f_n''$ may contain Dirac delta functions, and

$$\frac{\hat f_n^{(j)}(y)}{\hat f_n(y)} = h^{-(j+1)}\sum_{i=1}^{n} w_i\frac{K^{(j)}}{K}\left(\frac{y-\tilde N_i}{h}\right)$$

for $j = 1, 2$, with

$$w_i = K((y-\tilde N_i)/h)\Big/\sum_{k=1}^{n} K((y-\tilde N_k)/h)$$

weights satisfying $\Sigma_i w_i = 1$. This proves the lemma, with

$$\mathbb{D}_n = \bigcup_{1\le i\le n}(\tilde N_i + h\mathbb{D}_0)\cup\{-a, a\}. \qquad \square$$

Appendix II
Proofs from Subsections IV-A and IV-B

Proof of Proposition 4.1

Let $\tilde\psi = \psi_0 + \tilde\varepsilon$ be a function "close to" ψ_0. Then

$$\begin{aligned}\mathcal{E}(\tilde\psi) &= \frac{\langle\tilde\psi,\psi_0\rangle^2}{\|\tilde\psi\|^2} = \frac{\langle\psi_0+\tilde\varepsilon,\psi_0\rangle^2}{\|\psi_0+\tilde\varepsilon\|^2}\\ &= \|\psi_0\|^2 - \frac{\|\tilde\varepsilon\|^2\|\psi_0\|^2 - \langle\tilde\varepsilon,\psi_0\rangle^2}{\|\psi_0\|^2 + 2\langle\tilde\varepsilon,\psi_0\rangle + \|\tilde\varepsilon\|^2}\\ &= \mathcal{E}(\psi_0) - \frac{\|\tilde\varepsilon\|^2\|\psi_0\|^2 - \langle\tilde\varepsilon,\psi_0\rangle^2}{\|\psi_0\|^2} + O(\|\tilde\varepsilon\|^3).\end{aligned}$$

Apply this with either $\tilde\varepsilon = \hat\varepsilon_n$ or $\tilde\varepsilon = \overline\varepsilon_n$. □

Proof of Theorem 4.2

Assume without loss of generality (w.l.o.g.) that $\mu_0(K) = 1$. Introduce

$$f_n(y) = E\hat f_n(y) = \int K(u)f(y-uh)\,du \qquad \text{(A2)}$$

and

$$\psi_n = -\frac{f_n'}{f_n}. \qquad \text{(A3)}$$

Following [3] and [4, Sec. 7.8], we make the decomposition

$$\begin{aligned}(\hat\psi_n - \psi_0)\sqrt{f} &= \hat\psi_n(\sqrt{f} - \sqrt{f_n}) + (\hat\psi_n - \psi_n)\sqrt{f_n}\\ &\quad + (\psi_n\sqrt{f_n} - \psi_0\sqrt{f}).\end{aligned}$$

Hence

$$\begin{aligned}\int &(\hat\psi_n - \psi_0)^2 f\,dy\\ &\le 3\left(\int\hat\psi_n^2(\sqrt{f_n} - \sqrt{f})^2\,dy + \int(\hat\psi_n - \psi_n)^2 f_n\,dy\right.\\ &\quad\left. + \int\left(\frac{f_n'}{\sqrt{f_n}} - \frac{f'}{\sqrt{f}}\right)^2 dy\right) := 3(\mathrm{I} + \mathrm{II} + \mathrm{III}).\end{aligned}$$

Lemma 3.1 implies

$$\hat{\psi}_n(y) = \begin{cases} -\hat{f}_n'(y)/\hat{f}_n(y), & \text{if } |y| \le a_n \text{ and} \\ & |\hat{f}_n'(y)| \le c_n \hat{f}_n(y) \\ 0, & \text{otherwise} \end{cases}$$

with $c_n = ||K'/K||_{L^\infty(\mathbb{R})} h_n^{-1}$. It is proved in [4, Sec. 7.8] that I, II, III $\xrightarrow{p} 0$ under the assumption

$$h_n c_n \to 0. \tag{A4}$$

However, inspection of the proof reveals that II, III $\xrightarrow{p} 0$ under the weaker assumption

$$h_n c_n = O(1) \tag{A5}$$

which is satisfied in our case. (Notice that c_n is not involved in III, so (A5) is trivially sufficient, given that (A4) is.) It remains to prove that (A5) also implies I $\xrightarrow{p} 0$. Observe first that

$$I \le (c_n h)^2 \int \left(\frac{\sqrt{f_n} - \sqrt{f}}{h} \right)^2 dy$$

so it suffices to prove

$$\int \left(\frac{\sqrt{f_n} - \sqrt{f}}{h} \right)^2 dy \to 0 \text{ as } h \to 0.$$

Let

$$A_\varepsilon = \{y; f(y) \ge \varepsilon\}$$
$$I_1 = \int_{A_\varepsilon} \left((\sqrt{f_n} - \sqrt{f})/h \right)^2 dy$$

and

$$I_2 = \int_{A_\varepsilon^c} \left((\sqrt{f_n} - \sqrt{f})/h \right)^2 dy.$$

Observe that

$$\frac{\sqrt{f_n(y)} - \sqrt{f(y)}}{h} = -\frac{1}{2} \int_0^1 \frac{\int_{-\infty}^{\infty} u f'(y - \lambda h u) K(u)\, du}{\left(\int f(y - \lambda h u) K(u)\, du \right)^{1/2}} d\lambda.$$

The relation

$$\frac{E(U)^2}{E(V)} \le E\left(\frac{U^2}{V} \right) \tag{A6}$$

which follows from Cauchy–Schwarz inequality, implies

$$\left(\frac{\sqrt{f_n(y)} - \sqrt{f(y)}}{h} \right)^2 \le \frac{1}{4} \int_0^1 \frac{\left(\int_{-\infty}^{\infty} u f'(y - \lambda h u) K(u)\, du \right)^2}{\int f(y - \lambda h u) K(u)\, du} d\lambda.$$
$$\le \frac{1}{4} \int_0^1 \int_{-\infty}^{\infty} u^2 \frac{f'(y - \lambda h u)^2}{f(y - \lambda h u)} K(u)\, du\, d\lambda.$$

Insertion into the definition of I_2 gives

$$I_2 \le \frac{1}{4} \int_0^1 \int_{-\infty}^{\infty} J_2(\lambda, u, h) u^2 K(u)\, du\, d\lambda$$

with

$$J_2(\lambda, u, h) = \int_{A_\varepsilon^c} \frac{f'(y - \lambda h u)^2}{f(y - \lambda h u)} dy.$$

Now $J_2(\lambda, u, h) \le I(f)$ and

$$\lim_{h \to 0} J_2(\lambda, u, h) \to \int_{A_\varepsilon^c} (f')^2/f\, dy$$

for all (λ, u) (dominated convergence). Another application of the dominated convergence theorem gives

$$\lim_{h \to 0} I_2 \le \frac{\mu_2(K)}{4} \int_{A_\varepsilon^c} \frac{(f')^2}{f} dy. \tag{A7}$$

If $y \in A_\varepsilon$

$$\left| \frac{\sqrt{f_n(y)} - \sqrt{f(y)}}{h} \right| \le \left| \frac{f_n(y) - f(y)}{h\sqrt{\varepsilon}} \right| = \frac{1}{\sqrt{\varepsilon}} \left| \int_0^1 \int_{-\infty}^{\infty} (f'(y - \lambda h u) - f'(y)) u K(u)\, du\, d\lambda \right|. \tag{A8}$$

Since $I(f) < \infty, ||f||_{L^\infty(\mathbb{R})} < \infty$ and hence

$$\int (f')^2\, dy \le ||f||_{L^\infty(\mathbb{R})} I(f) < \infty.$$

Minkowski's integral inequality and (A8) imply

$$I_1^{1/2} \le \frac{1}{\sqrt{\varepsilon}} \int_0^1 \int_{-\infty}^{\infty} J_1(\lambda, u, h) |u| K(u)\, du\, d\lambda$$

where

$$J_1(\lambda, u, h) = ||f'(\cdot - \lambda h u) - f'(\cdot)||_{L^2(\mathbb{R})}.$$

But

$$J_1(\lambda, u, h) \le 2\sqrt{||f||_{L^\infty(\mathbb{R})} I(f)} \quad \text{and} \quad J_1(\lambda, u, h) \to 0$$

as $h \to 0$ for all (λ, u) (L^2-continuity). Dominated convergence gives

$$\lim_{h \to 0} I_1 = 0. \tag{A9}$$

To finish the proof, combine (A7) with (A9) and let $\varepsilon \to 0$. □

Proof of Proposition 4.3

We include this proof for completeness. Condition ib) implies that ψ_0 is skew-symmetric. Hence

$$\begin{aligned}\int(\overline{\psi}_n-\psi_0)^2 f\,dy &= \int\left(\frac{1}{2}(\hat{\psi}_n(y)-\psi_0(y))+\frac{1}{2}(\hat{\psi}_n(-y)-\psi_0(-y))\right)^2 f(y)\,dy\\ &\le \frac{1}{2}\int(\hat{\psi}_n-\psi_0)^2(y)f(y)\,dy+\frac{1}{2}\int(\hat{\psi}_n-\psi_0)^2(-y)\,f(y)\,dy\\ &= \int(\hat{\psi}_n-\psi_0)^2(y)f(y)\,dy.\end{aligned}$$

Appendix III
Proof of Theorem 4.5

It suffices to establish (4.4), then (4.5) and (4.6) follow easily. Assume w.l.o.g. $\mu_0(K)=1$. We will let C denote a positive constant, whose value may change from line to line. Let $A=\{y;|y|\le C_1h^{-2}\}$. Then

$$E\int(\hat{\psi}_n-\psi_0)^2 f\,dy = E\int_A(\hat{\psi}_n-\psi_0)^2 f\,dy+\int_{A^c}\psi_0^2 f\,dy. \tag{A10}$$

The first term of (A10) is decomposed as

$$E\int_A(\hat{\psi}_n-\psi_0)^2 f\,dy \le 2E\int_A(\hat{\psi}_n-\psi_n)^2 f\,dy+2\int(\psi_n-\psi_0)^2 f\,dy \tag{A11}$$

Notice that

$$\hat{\psi}_n-\psi_n=\hat{\psi}_n\frac{\hat{f}_n-f_n}{f_n}-\frac{\hat{f}_n'-f_n'}{f_n}.$$

We will use the estimates

$$E(\hat{f}_n^{(i)}(y)-f_n^{(i)}(y))^2\le Cf_n(y)n^{-1}h^{-1-2i}, \qquad \text{for } i=1,2$$

derived in [23]. The first term of (A11) may be estimated according to

$$\begin{aligned}E\int_A(\hat{\psi}_n-\psi_n)^2 f\,dy &\le 2\|\hat{\psi}_n\|^2_{L^\infty(\mathbb{R})}\int_A\frac{E(\hat{f}_n-f_n)^2}{f_n^2}f\,dy\\ &\quad+2\int_A\frac{E(\hat{f}_n'-f_n')^2}{f_n^2}f\,dy\\ &\le Ch^{-2}\int_A n^{-1}h^{-1}\frac{f}{f_n}\,dy+C\int_A n^{-1}h^{-3}\frac{f}{f_n}\,dy\\ &= Cn^{-1}h^{-3}\int_A\frac{f}{f_n}\,dy\le Cn^{-1}h^{-3}a_n\end{aligned} \tag{A12}$$

where we made use of Lemma 3.1 as well as of Lemma III.1 below. Consider now the second term of (A11). Notice that

$$\psi_n-\psi_0=\frac{f'}{f}\frac{f_n-f}{f_n}-\frac{f_n'-f'}{f_n}$$

and

$$\begin{aligned}(\psi_n-\psi_0)^2 f &\le 2\left(\left(\frac{f'}{f^{5/6}}\right)^2\left(\frac{f_n-f}{f_n^{2/3}}\right)^2\left(\frac{f}{f_n}\right)^{2/3}+\frac{(f_n'-f')^2}{f_n}\frac{f}{f_n}\right)\\ &\le C\left(\left(\frac{f'}{f^{5/6}}\right)^2\left(\frac{f_n-f}{f_n^{2/3}}\right)^2+\frac{(f_n'-f')^2}{f_n}\right)\end{aligned} \tag{A13}$$

because of Lemma III.1. Now $\mu_0(K)=1$ and $\mu_1(K)=0$ imply

$$f_n'(y)-f'(y) = \frac{h^2}{2}\int_0^1\int_{-\infty}^{\infty}2(1-\lambda)K(u)f^{(3)}(y-\lambda uh)\,du\,d\lambda$$

and

$$\begin{aligned}f_n(y) &\ge \frac{c_1}{2}\int_0^1 2(1-\lambda)f_n(y;\lambda h)\,d\lambda\\ &= \frac{c_1}{2}\int_0^1\int_{-\infty}^{\infty}2(1-\lambda)K(u)f(y-\lambda uh)\,du\,d\lambda\end{aligned} \tag{A14}$$

because of Lemma III.1. Therefore,

$$\begin{aligned}\frac{(f_n'-f')^2}{f_n}(y) &\le Ch^4\frac{\left(\int_0^1\int_{-\infty}^{\infty}2(1-\lambda)K(u)f^{(3)}(y-\lambda uh)\,du\,d\lambda\right)^2}{\int_0^1\int_{-\infty}^{\infty}2(1-\lambda)K(u)f(y-\lambda uh)\,du\,d\lambda}\\ &\le Ch^4\int_0^1\int_{-\infty}^{\infty}\frac{f^{(3)}(y-\lambda uh)^2}{f(y-\lambda uh)}\,du\,d\lambda\end{aligned}$$

where we used (A6) in the last step. Integration with respect to y gives

$$\int\frac{(f_n'-f')^2}{f_n}\,dy\le Ch^4\int\frac{(f^{(3)})^2}{f}\,dy=Ch^4, \tag{A15}$$

because of ic). Consider now the first term of (A13). Hölder's inequality implies

$$\int\left(\frac{f'}{f^{5/6}}\right)^2\left(\frac{f_n-f}{f_n^{2/3}}\right)^2 dy \le \left(\int\left(\frac{f'}{f}\right)^6 f\,dy\right)^{1/3}\left(\int\frac{|f_n-f|^3}{f_n^2}\,dy\right)^{2/3} \tag{A16}$$

$$= C\left(\int\frac{|f_n-f|^3}{f_n^2}\,dy\right)^{2/3} \tag{A17}$$

because of ic). Combination of

$$f_n(y)-f(y)=\frac{h^2}{2}\int_0^1\int_{-\infty}^{\infty}2(1-\lambda)K(u)f^{(2)}(y-\lambda uh)\,du\,d\lambda$$

and (A14) implies

$$\frac{|f_n - f|^3}{f_n^2}(y) \le Ch^6 \frac{\left|\int_0^1 \int_{-\infty}^{\infty} 2(1-\lambda)K(u)f^{(2)}(y-\lambda uh)\,du\,d\lambda\right|^3}{\left(\int_0^1 \int_{-\infty}^{\infty} 2(1-\lambda)K(u)f(y-\lambda uh)\,du\,d\lambda\right)^2}.$$

Application of the relation $|E(U)|^3/(EV)^2 \le E(|U|^3/V^2)$ (which follows from Hölder's inequality) yields

$$\frac{|f_n - f|^3}{f_n^2}(y) \le Ch^6 \int_0^1 \int_{-\infty}^{\infty} 2(1-\lambda) \cdot K(u) \frac{|f^{(2)}(y-\lambda uh)|^3}{f(y-\lambda uh)^2}\,du\,d\lambda.$$

Integrate with respect to y and apply ic).

$$\int \frac{|f_n - f|^3}{f_n^2}\,dy \le Ch^6 \int \left|\frac{f^{(2)}}{f}\right|^3 f\,dy = Ch^6.$$

Combine this with (A17) to obtain

$$\int \left(\frac{f'}{f^{5/6}}\right)^2 \left(\frac{f_n - f}{f_n^{2/3}}\right)^2 dy \le Ch^4. \tag{A18}$$

The theorem now follows from (A10)–(A13), (A15), and (A18). □

Lemma III.1: Let

$$f_n(y;h) = \int f(y-uh)K(u)\,du.$$

Then there exist constants $h_0 > 0$ and $0 < c_1 < 1$ such that

$$f_n(y;h) \ge c_1 f_n(y;h') \quad \forall 0 \le h' < h \le h_0, y \in \mathbb{R}.$$

Proof: Let $L(u) = -2uK'(u)$ and $\overline{f}_\delta$ as defined in ie). Then notice that

$$f_n(y;h) = \int_0^\infty \overline{f}_{sh}(y)L(s)\,ds.$$

When $h' = 0$, the lemma then follows from ie). When $h' > 0$, we will prove that $c_1 = (c_0^{-2} I_2/I_1 + 1)^{-1}$ suffices, with c_0 as defined in ie).

$$I_1 := -\int_{u_0}^{2u_0} u^2 K'(u)\,du$$

and

$$I_2 := -\int_{u_0}^{\infty} u^2 K'(u)\,du.$$

(It follows from vb), vd), and ve) that $I_1 > 0$ and $I_2 < \infty$.) Put $h' = \lambda h, 0 < \lambda < 1$. Write

$$\begin{aligned} f_n(y;\lambda h) &= \int_0^\infty \overline{f}_{\lambda sh}(y)L(s)\,ds \\ &= \int_0^{\lambda u_0 h} \overline{f}_\delta(y) L\left(\frac{\delta}{\lambda h}\right)\frac{d\delta}{\lambda h} \\ &\quad + \int_{\lambda u_0 h}^{u_0 h} \overline{f}_\delta(y) L\left(\frac{\delta}{\lambda h}\right)\frac{d\delta}{\lambda h} \\ &\quad + \int_{u_0 h}^{\infty} \overline{f}_\delta(y) L\left(\frac{\delta}{\lambda h}\right)\frac{d\delta}{\lambda h} \\ &:= f_{n1}(y;\lambda h) + f_{n2}(y;\lambda h) + f_{n3}(y;\lambda h). \end{aligned}$$

For h sufficiently small

$$\begin{aligned} f_{n1}(y;\lambda h) &= \int_0^{u_0 h} \overline{f}_{\delta\lambda}(y) L\left(\frac{\delta}{h}\right)\frac{d\delta}{h} \\ &\le c_0^{-1} \int_0^{u_0 h} \overline{f}_\delta(y) L\left(\frac{\delta}{h}\right)\frac{d\delta}{h} \\ &= c_0^{-1} f_{n1}(y;h), \\ f_{n2}(y;\lambda h) &= \int_{u_0 h}^{u_0 h/\lambda} \overline{f}_{\delta\lambda}(y) L\left(\frac{\delta}{h}\right)\frac{d\delta}{h} \\ &\le c_0^{-1} \int_{u_0 h}^{u_0 h/\lambda} \overline{f}_{u_0 h}(y) L\left(\frac{\delta}{h}\right)\frac{d\delta}{h} \\ &\le c_0^{-1} \overline{f}_{u_0 h}(y) I_2 \\ &\le c_0^{-2} \frac{I_2}{I_1} \int_{u_0 h}^{2u_0 h} f_\delta(y) L\left(\frac{\delta}{h}\right)\frac{d\delta}{h} \\ &\le c_0^{-2} \frac{I_2}{I_1} f_{n3}(y;h) \end{aligned}$$

and

$$\begin{aligned} f_{n3}(y;\lambda h) &= \int_{u_0 h}^{\infty} \overline{f}_\delta(y) L\left(\frac{\delta}{\lambda h}\right)\frac{d\delta}{\lambda h} \\ &\le \int_{u_0 h}^{\infty} \overline{f}_\delta(y) L\left(\frac{\delta}{h}\right)\frac{d\delta}{h} \end{aligned}$$

where we used the fact that $uL(u)$ is nonincreasing for $u \ge u_0$ in the last inequality. Combining the last three estimates, we see that we can choose the value of c_1 mentioned above.

ACKNOWLEDGMENT

The authors wish to thank the Associate Editor and two referees for comments on the manuscript that improved the presentation of the paper.

REFERENCES

[1] R. Beran, "Asymptotically efficient adaptive rank estimates in location models," *Ann. Statist.*, vol. 2, pp. 63–74, 1974.
[2] R. Beran, "An efficient and robust adaptive estimator of location," *Ann. Statist.*, vol. 6, pp. 292–313, 1978.
[3] P. J. Bickel, "On adaptive estimation," *Ann. Statist.*, vol. 10, pp. 647–671, 1982.
[4] P. J. Bickel, C. A. Klaasen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: John Hopkins Univ. Press, 1993.
[5] D. D. Cox, "A penalty method for nonparametric estimation of the logarithmic derivative of a density function," *Ann. Inst. Statist. Math.*, vol. 37, pp. 271–288, 1985.

[6] J. J. Faraday, "Smoothing in adaptive estimation," *Ann. Statist.*, vol. 20, pp. 414–427, 1992.
[7] L. M. Garth and H. V. Poor, "Detection of non-Gaussian signals: A paradigm for modern statistical signal processing," *Proc. IEEE*, vol. 82, no. 7, pp. 1061–1095, July 1994.
[8] J. Hájek, "Asymptotically most powerful rank order tests," *Ann. Math. Statist.*, vol. 33, pp. 1124–1147, 1962.
[9] W. Härdle and D. W. Scott, "Smoothing by weighted averaging of rounded points," *Comp. Statist.*, vol. 7, pp. 97–128, 1992.
[10] O. Hössjer and M. Mettiji, "Robust multiple classification of known signals in additive noise—An asymptotic weak signal approach," *IEEE Trans. Inform. Theory*, vol. 39, pp. 594–608, Mar. 1993.
[11] K. Jin, "Empirical smoothing parameter selection in adaptive estimation," *Ann. Statist.*, vol. 20, pp. 1844–1874, 1992.
[12] S. A. Kassam, *Signal-Detection in Non-Gaussian Noise*. Berlin, Germany: Springer-Verlag, 1988.
[13] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, July 1985.
[14] R. D. Martin and S. C. Schwartz, "Robust detection of a known signal in nearly Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50–56, Jan. 1971.
[15] D. Middleton, "Statistical-physical models of urban radio-noise environments-Part I: Foundations," *IEEE Trans. Electromagn. Compat.*, vol. EMC-14, no. 2, pp. 38–56, 1972.
[16] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, pp. 106–127, Aug. 1977.
[17] ——, "Canonical and quasi-canonical probability models of Class A interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-25, 76–106, May 1983.
[18] B. U. Park, "Efficient estimation in the two-sample semiparametric location-scale models," *Probab. Theory Rel. Fields*, vol. 86, pp. 21–39, 1990.
[19] S. Portnoy and R. Koenker, "Adaptive L-estimation for linear models," *Ann. Statist.*, vol. 17, pp. 362–381, 1989.
[20] J. Sacks, "An asymptotically efficient sequence of estimators of a location parameter," *Ann. Statist.*, vol. 3, pp. 285–298, 1975.
[21] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman and Hall, 1986.
[22] C. Stein, "Efficient nonparametric testing and estimation," in *Proc. 3rd Berkeley Symp. on Statistical Probability*, vol. 1. Berkeley, CA: Univ. California Press, 195, pp. 187–195.
[23] C. J. Stone, "Adaptive maximum likelihood estimators of a location parameter," *Ann. Statist.*, vol. 3, pp. 267–284, 1975.
[24] C. van Eeden, "Efficient-robust estimation of location," *Ann. Math. Statist.*, vol. 41, pp. 172–181, 1970.
[25] H. L. van Trees, *Detection, Estimation and Modulation Theory, 1*. New York: Wiley, 1968.
[26] M. P. Wand and M. C. Jones, *An Introduction to Kernel Smoothing*. London, U.K.: Chapman and Hall, 1994.
[27] S. M. Zabin and H. V. Poor, "Parameter estimation for Middleton Class A interference processes," *IEEE Trans. Commun.*, vol. 37, pp. 1042–1051, Oct. 1989.
[28] ——, "Recursive algorithms for identification of impulsive noise channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 559–578, May 1990.