The Annals of Statistics
1991, Vol. 19, No. 2, 982–998

ASYMPTOTICALLY OPTIMAL HYPOTHESIS TESTING WITH MEMORY CONSTRAINTS

BY J. A. BUCKLEW AND P. E. NEY

University of Wisconsin–Madison

The binary hypothesis testing problem of deciding between two Markov chains is formulated under memory constraints. The optimality criterion used is the exponential rate with which the probability of error approaches zero as the sample size tends to infinity. The optimal memory constrained test is shown to be the solution of a set of equations derived from suitable large deviation twistings of the transition matrices under the two hypotheses. A computational algorithm and some examples are given.

1. Introduction. The performance capabilities and cost of modern high speed electronic communication systems is often proportional to the memory or storage requirements of the task to be performed. A typical situation may be that one is sending digital information over a noisy communications channel. A logic i is communicated by transmitting the function $f_i(t)$, $0 \le t \le T$, $i \in \{0, 1\}$. Every T seconds, the communications receiver must perform a simple binary hypothesis test to determine whether a logic 1 or logic 0 was transmitted in the previous signaling interval. For a satellite or fiber optic system, T would typically be 10^{-7} to 10^{-9} [Lathi (1989)]. The electronic (thermal and man-made) noise over such a short signaling interval should be assumed to have some memory of its past (engineers would say the noise is colored or nonwhite). A useful model for the noise is that it be modeled as some sort of Markov process that is simply added to the transmitted waveform. If such a model is used, then the communication receiver design problem comes down to deciding which of two Markov processes is present in the previous signaling interval.

The classical solution to this problem is, of course, to implement a log-likelihood ratio test statistic which would give the best performance in the usual Bayes risk or Neyman–Pearson sense. Problems arise, however, when one considers how to go about implementing this test in electronic hardware. It is virtually impossible to build an arbitrary mapping from $C[0, T]$, the continuous functions on the interval $[0, T]$, into $\mathbb{R}$. One must first sample the process over the interval $[0, T]$. In many situations this sampled process will be a discrete time Markov chain with transition function P and we must test $P = P_0$ versus $P = P_1$ for given transition functions P_0, P_1. The log-likelihood ratio test would then necessitate computing

$$T_n \equiv \sum_1^n \log \frac{P_1(X_i, X_{i+1})}{P_0(X_i, X_{i+1})}. \tag{1.1}$$

Received July 1988; revised October 1989.
AMS 1980 *subject classifications*. Primary 62F05; secondary 60F10.
Key words and phrases. Testing hypotheses, memory constraints, large deviations.

Here a nonlinear operation must be performed every T/n seconds. Specifically, every T/n seconds, the value X_i must be recalled from a memory and the nonlinearity $\log[P_1(X_i, X_{i+1})/P_0(X_i, X_{i+1})]$ computed (a table lookup procedure would be fastest if there are only a few possible values). The very fastest random access memories currently available require about $5 \cdot 10^{-9}$ seconds to access a stored value [Wooley (1988)]. If T itself is around 10^{-9}, it is clear that an attempt via a straightforward optimal processing design is doomed to failure. The only other possibility would be to try to implement some sort of parallel computation architecture necessitating a large increase in complexity.

In problems of this sort, strict optimality is not the overriding concern. The probability of error in such systems is very low (10^{-6} is a typical number for a communications link before any error correction is implemented) [Haykin (1989)]. Developing fast (low memory) suboptimal structures is often a more desirable goal than implementing a slow optimal one. To this end, practitioners have frequently implemented a test statistic of the form

$$T_n \equiv \sum_1^n f(X_i), \tag{1.2}$$

where $f(\cdot)$ (called a memoryless nonlinearity) is chosen (in an ad hoc fashion) to guarantee that the mean value of T_n under the two hypotheses is different and thus will lead to a consistent test. A test of this form is called a *memoryless detector* in the engineering literature, since the present sample may be processed without knowledge of the previous samples. Memoryless nonlinearities may be computed (or at least staircase approximations) by relatively simple hardware at very high speed by so-called flash converters [Roden (1988)]. Typical functions to implement would be $f(x) = \operatorname{sgn}(x)$ [engineers would say $f(\cdot)$ is a hard limiter] or $f(x) = x$ for $x \in [-a, a]$ and $f(x) = a \cdot \operatorname{sgn}(x)$ otherwise (engineers would call $f(\cdot)$ a type of soft limiter [Haykin (1989)]). The nonlinearities are almost always bounded since it is desired that large noise spikes due to intentional or unintentional jamming and interference have limited impact on the test statistic.

In this paper we will be concerned with developing hypothesis testing designs under a memory constraint. Our performance criterion will be the asymptotic efficiency of the test, that is, the rate at which the maximum error probability goes to zero. Due to the large sample sizes encountered in many practical situations, an asymptotic criterion is not inappropriate. More precisely, we will show that:

1. A unique optimal g (in the previous sense) always exists.
2. This function is determined as the solution of a set of equations derived from suitable twistings of P_0 and P_1.
3. An algorithm is given for computing g.

We limit ourselves here to Markov chains on a finite state space and to tests of simple hypotheses versus simple alternatives. This makes the mathematics

very tractable and exposes many of the salient features of the problem. Some natural possible extensions are discussed in Section 6.

The problem we are considering has been studied extensively in the theoretical engineering literature usually via the locally optimal asymptotic relative efficiency (ARE) as a test fidelity criterion [Miller and Thomas (1972), Poor and Thomas (1979), Halverson and Wise (1984) and Sadowsky and Bucklew (1986)]. The ARE criterion requires that one let the two hypotheses collapse towards one another and at the same time increase the number of decision samples in order to maintain a given fixed power. Because of the assumption that one hypothesis converges to the other, the design problem may be handled relatively simply by central limit methods. In this paper, our performance criterion will be the Chernoff efficiency of the test, that is, the asymptotic rate at which the probability of error approaches zero as the sample size gets large. Thus (unlike the ARE), we allow the hypotheses to remain fixed and only the number of decision samples is changed to obtain the rate performance.

Our problem and the techniques presented here could allow for other interpretations and extensions. Our techniques can be viewed as providing a large sample approximation to the likelihood ratio for a fixed memory allowance. Extensions to processes with longer memory and test statistics with various other more complicated memory constraints would be of interest.

The organization of the rest of the paper is as follows: Section 2 contains a brief summary of the necessary mathematical background of matrix perturbation theory and large deviation theory for Markov chains. Section 3 is a statement of our results. Section 4 contains proofs of the results of Section 3. Section 5 discusses computational aspects of our algorithm. Section 6 contains a discussion of possible extensions of them.

2. Some mathematical background. The mathematics in this paper are based on a combination of large deviation theory for Markov chains and an elementary perturbation result for matrices. We summarize the main facts we will need.

Let $X_1, X_2, \ldots$ be an irreducible Markov chain on a finite state space $\mathscr{S} = \{x_1, x_2, \ldots, x_d\}$, with transition matrix $P = \{p(x, y);\ x, y \in \mathscr{S}\}$ and invariant measure $\pi = (\pi_1, \pi_2, \ldots, \pi_d)$. Let $g(\cdot)$ be a real valued function on $\mathscr{S}$, $\mu = E_\pi g(X_1)$ and $S_n = \sum_1^n g(X_i)$.

The objective of large deviation (LD) theory is to study the asymptotic behavior of the deviation of the sample mean, namely $P\{S_n/n > a\}$, for $a > \mu$ (if $a < \mu$, this quantity converges to one by the law of large numbers). Roughly speaking, LD theory tells us that these probabilities decay exponentially and the objective is to find the exponent. More precisely,

$$\lim_{n\to\infty} \frac{1}{n} \log P\left\{\frac{S_n}{n} > a\right\} = -I(g), \tag{2.1}$$

where $I(g)$ is called the *rate function*.

To describe $I(g)$, one considers the transform matrix $\hat{P}(\theta, g) = \{\exp[g(x)] \cdot p(x, y)\}$. This is a nonnegative irreducible matrix and hence has a maximal

eigenvalue $\lambda(\theta; g)$ with associated right and left eigenvectors $r(\theta, g) = (r(x_1; \theta, g), r(x_2; \theta, g), \ldots, r(x_d; \theta, g))$ and $l(\theta, g) = (l(x_1; \theta, g), l(x_2; \theta, g), \ldots, l(x_d; \theta, g))$. Let $\Lambda(\theta; g) = \log \lambda(\theta; g)$. The convex conjugate of $\Lambda(\theta; g)$ is defined to be

$$\Lambda^*(x; g) = \sup[\theta x - \Lambda(\theta; g); \theta \in \mathbb{R}]. \tag{2.2}$$

One of the basic results of LD theory says that the limit in (2.1) exists and that $I(g) = \Lambda^*(a; g)$. For a proof of this result see, for example, Miller (1961) or Ellis (1985) or in much greater generality, Donsker and Varadhan (1975a, b, 1976) or Ney and Nummelin (1987a, b). In the case when $\{X_n\}$ are iid, this result is just the classical Cramér/Chernoff theorem.

In the problems under study here, we will also refer to the so-called twisted transition function

$$\hat{Q}(\theta, g) = \left\{ \frac{\exp[\theta g(x)] p(x, y) r(x; \theta, g)}{\lambda(\theta) r(y; \theta, g)}; x, y \in \mathscr{E} \right\}. \tag{2.3}$$

It is easy to check that this is really a transition function and that its invariant probability measure is

$$\pi(\theta, g) = \{\pi(x; \theta, g)\} = \{l(x; \theta, g) r(x; \theta, g); x \in \mathscr{E}\}, \tag{2.4}$$

where r and l are normalized so that $\Sigma_x \pi(x; \theta, g) = 1$. The important property of $\hat{P}(\theta, g)$ is that if we take $\theta = \theta_a$ as the solution of $\Lambda'(\theta; g) = a$, then $E_{\pi(\theta_a, g)} g(X_1) = a$. Thus by going over to $\hat{P}$, one centers S_n/n at a.

The previous facts are basically all the LD theory we will need. We will also use a perturbation result for $\hat{P}(\theta, g)$. Namely, suppose f is any real valued function on $\mathscr{E}$ and take $\varepsilon > 0$ (thought of as small). Then we consider the perturbed matrix $\hat{P}(\theta, g + \varepsilon f)$. Clearly (as $\varepsilon \to 0$),

$$\hat{P}(\theta, g + \varepsilon f) = \{\exp[\theta g(x)] p(x, y) [1 + \varepsilon \theta f(x) + O(\varepsilon^2)]; x, y \in \mathscr{E}\}$$

and by a textbook result on perturbation of matrices [see, e.g., Chatelin (1983), page 13], the maximal eigenvalue of $\hat{P}(\theta, g + \varepsilon f)$ satisfies

$$\lambda(\theta; g + \varepsilon f) = \lambda(\theta; g) + \varepsilon \theta \tilde{\lambda}(\theta; g, f) + O(\varepsilon^2), \tag{2.5}$$

where

$$\tilde{\lambda}(\theta; g, f) = \lambda(\theta; g) \sum_x f(x) \pi(x; \theta, g). \tag{2.6}$$

Note that this says that the directional derivative of $\lambda(\theta; g)$ with respect to g in the direction f is

$$D_f \lambda(\theta; g) = \theta \lambda(\theta; g) \sum_x f(x) \pi(x; \theta, g). \tag{2.7}$$

3. Results. As in the previous section, $\{X_n, n = 0, 1, \ldots\}$ is an irreducible Markov chain on a finite state space $\mathscr{E} = \{x_1, \ldots, x_d\}$ with time homogeneous

transition function $P = \{p(x, y);\ x, y \in \mathscr{E}\}$. Let P^0 and P^1 be given irreducible stochastic matrices. We are to test

$$H_0\colon P = P^0 \quad \text{versus} \quad H_1\colon P = P^1.$$

Associated with each of P^0 and P^1 there will be all the objects defined in Section 2: invariant measures, transform matrices, eigenfunctions and eigenvalues, etc. We denote these by superscripts or subscripts $i = 0, 1$.

Assume that $\pi^0 \neq \pi^1$. (Otherwise our test cannot discriminate between H_0 and H_1.) We constrain ourselves to tests of the following form, the so-called memoryless detectors:

$$\text{decide } H_0 \text{ if } S_n \geq na, \qquad H_1 \text{ if } S_n < na, \tag{3.1}$$

where

$$S_n = S_n(g) = \sum_{i=1}^{n} g(X_i),$$

g is a real valued function on $\mathscr{E}$ and $a \in \mathbb{R}$. We must choose the memoryless nonlinearity g and the threshold a. Clearly, it is no loss of generality to take $a = 0$ and we must thus choose g so as to maximize

$$I_0(g) \wedge I_1(g),$$

where

$$I_0(g) = -\lim_{n\to\infty} \frac{1}{n} \log P^{H_0}(S_n < 0)$$

and

$$I_1(g) = -\lim_{n\to\infty} \frac{1}{n} \log P^{H_1}(S_n \geq 0)$$

($a \wedge b = \min(a, b)$; $a \vee b = \max(a, b)$). These quantities are the error rates associated with H_0 and H_1. As indicated in Section 2, associated with each of P^0 and P^1 is a transform matrix $\hat{P}^i(\theta, g) = \{\exp[\theta g(x)]p^i(x, y)\}$ and a maximal eigenvalue $\lambda^i(\theta; g)$, $i = 0, 1$. By the definition (2.2) with $a = 0$,

$$I_i(g) = \Lambda^{i*}(0; g) = -\inf_{\theta} \log \lambda^i(\theta; g). \tag{3.2}$$

We will see that the infimum is actually achieved at a point θ_g^i (depending of course on g) and thus our problem reduces to finding g so as to minimize

$$\lambda^0\left(\theta_g^0; g\right) \vee \lambda^1\left(\theta_g^1; g\right). \tag{3.3}$$

We will call g an optimal solution if it minimizes (3.3). Let

$$\lambda\left(\theta^0, \theta^1, g\right) = \lambda^0\left(\theta^0; g\right) \vee \lambda^1\left(\theta^1; g\right). \tag{3.4}$$

We will see that there exist $(\theta^{0*}, \theta^{1*}, g^*)$ such that

$$\inf_{g, \theta^0, \theta^1} \lambda\left(\theta^0, \theta^1, g\right) = \lambda\left(\theta^{0*}, \theta^{1*}, g^*\right), \tag{3.5}$$

and thus g^* is an optimal solution. Let us call $(\theta^{0*}, \theta^{1*}, g^*)$ an optimal triple.

We will also see that this point is unique in a suitable sense and is determined as the solution of a set of equations.

We start by fixing (θ^0, θ^1) and finding an optimal related g; then we consider what happens as θ^0, θ^1 and g vary. The main result is Theorem 2.

THEOREM 1. *For every fixed $\theta^0 > 0 > \theta^1$, there exists a unique solution g^* of the pair of equations*

$$\lambda^0(\theta^0; g) = \lambda^1(\theta^1; g) \tag{3.6a}$$

and

$$\pi^0(x; \theta^0, g) = \pi^1(x; \theta^1, g), \qquad x \in \mathscr{E}. \tag{3.6b}$$

This solution is optimal with respect to (θ^0, θ^1) in the sense that it minimizes $\lambda(\theta^0, \theta^1, g)$. It is in fact the unique local optimum (i.e., the unique solution that minimizes λ in a neighborhood of g^*).

Thus for each fixed (θ^0, θ^1), there is a unique optimal g. (Later we give an algorithm for computing this g.) It remains to determine the particular pair (θ^0, θ^1) whose associated g is the global optimum we are seeking. To this end we will prove Theorem 2. Note that

$$\begin{aligned} \lambda^i\left(c\theta; \frac{g}{c}\right) &= \lambda^i(\theta; g), \qquad i = 0, 1 \\ \lambda\left(c\theta^0, c\theta^1, \frac{g}{c}\right) &= \lambda(\theta^0, \theta^1, g). \end{aligned} \tag{3.7}$$

Thus any uniqueness of an optimal triple will be valid only up to multiplicative constants as in (3.7).

THEOREM 2. *Let $(\theta^{0*}, \theta^{1*}, g^*)$ be an optimal solution. Then this triple satisfies*

$$\lambda^0(\theta^{0*}; g^*) = \lambda^1(\theta^{1*}; g^*), \tag{3.8a}$$

$$\pi^0(x; \theta^{0*}, g^*) = \pi^1(x; \theta^{1*}, g^*), \qquad x \in \mathscr{E}, \tag{3.8b}$$

$$\sum g^*(x)\pi^i(x; \theta^{i*}, g^*) = 0, \qquad i = 0, 1. \tag{3.8c}$$

Equations (3.8a, b, c) have a unique solution [up to multiplication by constants in the sense of (3.7)]. Thus $\lambda(\theta^0, \theta^1, g)$ achieves a minimium at $(\theta^{0*}, \theta^{1*}, g^*)$ and g^* is the sought-after optimal solution of the testing problem.

4. Proofs. Let $\pi^i = (\pi^i(x_1; \theta, g), \ldots, \pi^i(x_d; \theta, g))$, $i = 0, 1$, be the invariant probability measure of P^i and $\mu_i(g) = E_{\pi^i}[g(X_1)]$. Then $\lim(S_n/n) = \mu_i(g)$ exists a.s. $[P^i]$. Since we have assumed that $\pi^0 \neq \pi^1$, we may restrict ourselves to g's such that

$$\mu_0(g) < 0 < \mu_1(g), \tag{4.1}$$

since otherwise we would have $I_0(g) \wedge I_1(g) = 0$, whereas one can always find a g such that $I_0 \wedge I_1 > 0$.

Consider the transform matrices

$$\hat{P}^i(\theta, g) = \{e^{\theta g(x)} p^i(x, y); x, y \in \mathscr{S}\}, \qquad \theta \in \mathbb{R}, \tag{4.2}$$

with maximal eigenvalues $\lambda^i(\theta; g)$ and associated left and right eigenvectors $l^i(\cdot; \theta, g)$, $r^i(\cdot; \theta, g)$. Let $\Lambda^i(\theta; g) = \log \lambda^i(\theta; g)$ and $\Lambda^{i*}(\cdot; g)$ equals the convex conjugate of Λ^i. Then

$$I_i(g) = \Lambda^{i*}(0; g) = -\inf_{\theta} \log \lambda^i(\theta; g) \geq 0. \tag{4.3}$$

Let $\pi^i(x; \theta, g) = l^i(x; \theta, g) r^i(x; \theta, g)$ and normalize so that $\Sigma_x \pi^i(x; \theta, g) = 1$. Then $\pi^i(\theta, g) = (\pi^i(x_1; \theta, g), \ldots, \pi^i(x_d; \theta, g))$ is the invariant probability measure of the twisted stochastic matrix in (2.3), namely

$$\hat{Q}^i(x, y; \theta, g) = \hat{P}^i(x, y; \theta, g) \frac{r^i(x; \theta, g)}{\lambda^i(\theta; g) r^i(y; \theta, g)}.$$

Clearly the optimal g must satisfy

$$g(x_1) \wedge \cdots \wedge g(x_d) < 0 < g(x_1) \vee \cdots \vee g(x_d). \tag{4.4}$$

Thus $\hat{P}^i(\theta, g)$ are nonnegative irreducible matrices having at least one entry $e^{\theta g(x)} p^i(x, y) \to \infty$ as $|\theta| \to \infty$ [for some (x, y)]. This forces $\lambda^i(\theta; g) \to \infty$ as $|\theta| \to \infty$ and since λ^i is a convex function of θ, the infimum in (4.3) is achieved at some point θ_g^i. Thus

$$I_i(g) = -\log \lambda^i\left(\theta_g^i; g\right)$$

and as indicated in Section 3, we must find g so as to minimize

$$\lambda^0\left(\theta_g^0; g\right) \vee \lambda^1\left(\theta_g^1; g\right).$$

LEMMA 1. *If g satisfies* (4.1), *then*

$$\theta_g^0 > 0 > \theta_g^1.$$

PROOF. $\lambda^0(\cdot; g)$ is differentiable and by the twisting property [see Ney and Nummelin (1987 a, b)]

$$\frac{d\lambda^0}{d\theta}\left(\theta_g^0\right) = 0.$$

Also $(d\lambda^0/d\theta)(0) = \mu_0(g) < 0$ by (4.1). But by convexity of λ^0, $d\lambda^0/d\theta$ is increasing and this forces $\theta_g^0 > 0$. Similarly, $\theta_g^1 < 0$. □

Turning to the proof of Theorem 1, we proceed via three lemmas.

LEMMA 2. *If g^* is locally optimal with respect to $\theta^0 > 0 > \theta^1$, then*

$$\lambda^0(\theta^0; g^*) = \lambda^1(\theta^1; g^*) \tag{4.5a}$$

and

$$\tilde{\lambda}^0(\theta^0; g^*, f) = \tilde{\lambda}^1(\theta^1; g^*, f) \quad \text{for all } f\colon \mathscr{E} \to \mathbb{R}. \tag{4.5b}$$

Conversely, if g^* satisfies (4.5) [*or* (3.6)], *then it is locally optimal with respect to (θ^0, θ^1). Furthermore* (3.6) *and* (4.5) *are equivalent.*

LEMMA 3. *There exists an optimal g with respect to $\theta^0 > 0 > \theta^1$, hence a solution of* (4.5).

LEMMA 4. *There exists at most one solution of* (4.5); *hence a unique locally optimal g.*

These lemmas together prove Theorem 1.

PROOF OF LEMMA 2. Suppose $\lambda^0(\theta^0; g^*) < \lambda^1(\theta^1; g^*)$. Then for $\varepsilon > 0$ ($\bar{\varepsilon} = (\varepsilon, \dots, \varepsilon) \in \mathbb{R}^d$) and small we could force

$$\max_{i=0,1} \lambda^i(\theta^i; g^* + \bar{\varepsilon}) = \max_{i=0,1} e^{\varepsilon\theta^i}\lambda^i(\theta^i; g^*) < \max_{i=0,1} \lambda^i(\theta^i; g^*),$$

contradicting local optimality of g^*. For the proof of (4.5b), note that from (2.6) for any constant vector $\bar{c} = (c, \dots, c) \in \mathbb{R}^d$,

$$\tilde{\lambda}^i(\theta; g, f + \bar{c}) = \tilde{\lambda}^i(\theta; g, f) + c\lambda^i(\theta; g) \tag{4.6}$$

and hence by (2.5),

$$\begin{aligned} &\lambda^i(\theta; g + \varepsilon(f + \bar{c})) \\ &\quad = \lambda^i(\theta; g) + \varepsilon\theta\left[\tilde{\lambda}^i(\theta; g, f) + c\lambda^i(\theta; g)\right] + O(\varepsilon^2). \end{aligned} \tag{4.7}$$

Consider the lines (as functions of c)

$$L^i(c) = \tilde{\lambda}^i(\theta^i; g, f) + c\lambda^i(\theta^i; g).$$

If g^* is locally optimal with respect to (θ^0, θ^1), then these lines (with $g = g^*$) must intersect somewhere on the c-axis.

Suppose there is no intersection. Then $L^0(c)L^1(c) < 0$ for some c and since also $\theta^0\theta^1 < 0$, we could find arbitrarily small ε such that [by (4.7)]

$$\max_{i=0,1} \lambda^i(\theta^i; g^* + \varepsilon(f + c)) < \max_{i=0,1} \lambda^i(\theta^i; g^*), \tag{4.8}$$

contradicting local optimality. Hence $L^0(c) = L^1(c) = 0$ for some c, namely

$$-c = \frac{\tilde{\lambda}^0(\theta^0; g^*, f)}{\lambda^0(\theta^0; g^*)} = \frac{\tilde{\lambda}^1(\theta^1; g^*, f)}{\lambda^1(\theta^1; g^*)}.$$

Since we already have (4.5a), this implies (4.5b).

Conversely, note that (4.5) and (2.5) imply that g^* is locally optimal for $\theta^0 > 0 > \theta^1$.

Finally, the equivalence of (4.5) and (3.6) follows from (2.6). This proves Lemma 2. □

PROOF OF LEMMA 3. Note that $\{e^{g(x)}p(x,y);\ x,y \in \mathscr{E}\}$ can be interpreted as the transform matrix in the parameter $(g(x_1),\ldots,g(x_d))$ of a d-dimensional Markov additive (MA) process, namely the occupation time of a Markov chain with transition function $\{p(x,y)\}$. Hence $\hat{P}(\theta) = \{e^{\theta g(x)}p(x,y)\}$ is the transform matrix (as a function of g) of an MA process and its maximal eigenvalue is a convex function of g [Miller (1961)]. Therefore $\lambda(\theta^0,\theta^1;g) \doteq \lambda^0(\theta^0;g) \vee \lambda^1(\theta^1;g)$ is also convex in g. Furthermore, $\lambda(\theta^0,\theta^1;g) \to \infty$ as $\|g\| \to \infty$, whenever $\theta_0 > 0$, $\theta_1 < 0$. Hence λ achieves its infimum for some g_{θ^0,θ^1}; that is, g_{θ^0,θ^1} is optimal with respect to (θ^0,θ^1). □

PROOF OF LEMMA 4. Suppose that there are two solutions g and g^1. By Lemma 3,

$$\tilde{\lambda}^0(\theta^0;g,f) = \tilde{\lambda}^1(\theta^1;g,f) \quad \text{and} \quad \tilde{\lambda}^0(\theta^0;g^1,f) = \tilde{\lambda}^1(\theta^1;g^1,f) \tag{4.9}$$

for all $f\colon \mathscr{E} \to \mathbb{R}$. Let $h(x) = g^1(x) - g(x)$ and for $\gamma \in \mathbb{R}$, let

$$\bar{P}^i(\gamma) = \{\bar{p}_i(x,y;\gamma);\ x,y \in \mathscr{E}\} = \{e^{\gamma\theta^i h(x)}\hat{p}^i(x,y;\theta^i,g)\}, \qquad i = 0,1,$$

where $\hat{p}^i(x,y;\theta^i,g) = e^{\theta^i g(x)}p^i(x,y)$. Thus $\bar{P}^i(\gamma)$ is itself a transform matrix (in the variable γ) of the (nonstochastic) matrix $\hat{P}^i(\theta^i,g)$.

Let $\bar{\lambda}^i(\gamma)$ be its maximal eigenvalue. If $h(x)$ is not identically equal to 0, then $\bar{\lambda}^i$ is a convex, differentiable function of γ, and by (2.5), for small γ,

$$\bar{\lambda}^i(\gamma) = \lambda^i(\theta^i;g+\gamma h) = \lambda^i(\theta^i;g) + \gamma\theta^i\tilde{\lambda}(\theta^i;g,h) + O(\gamma^2). \tag{4.10}$$

Hence

$$\frac{d\bar{\lambda}^i}{d\gamma}(0) = \theta^i\tilde{\lambda}(\theta^i;g,h).$$

Also by hypothesis,

$$\bar{\lambda}^0(0) = \lambda^0(\theta^0;g) = \lambda^1(\theta^1;g) = \bar{\lambda}^1(0).$$

Thus by (4.9) and (4.10), we have

(4.11) $\bar{\lambda}^i(\gamma)$, $i = 0,1$, intersect at 0 and their derivatives have opposite signs there, or are both equal to 0.

Similarly, for γ in a neighborhood of 1, rewrite $\bar{P}^i$ as

$$\bar{P}^i_i(\gamma) = \{e^{(\gamma-1)\theta^i h(x)}\hat{p}(x,y;\theta^i,g^1)\}.$$

Then

$$\begin{aligned}\bar{\lambda}^i(\gamma) &= \lambda^i(\theta^i;g^1 + (\gamma-1)h) \\ &= \lambda^i(\theta^i;g^1) + (\gamma-1)\theta^i\tilde{\lambda}^i(\theta^i;g,h) + O(1-\gamma)^2\end{aligned} \tag{4.12}$$

or

$$\frac{d\bar{\lambda}^i}{d\gamma}(1) = \theta^i\tilde{\lambda}^i(\theta^i;g,h).$$

Since $\bar{\lambda}^i(1) = \lambda^i(\theta^i; g^1)$, we have again

$$\bar{\lambda}^0(1) = \bar{\lambda}^1(1),$$

and by (4.9) and (4.12):

(4.13) $\bar{\lambda}^i(\gamma)$ intersect at 1 and their derivatives have opposite signs there, or are both equal to 0.

But (4.11) and (4.13) together can hold only if $\bar{\lambda}^i(\gamma)$ are constant, which implies $h(x) = 0$ or $g = g^1$. □

We proceed with the proof of Theorem 2 via the following lemmas.

LEMMA 5. *If* $(\theta^{0*}, \theta^{1*}, g^*)$ *is an optimal triple, then it satisfies* (3.8).

LEMMA 6. *An optimal triple exists.*

LEMMA 7. *There exists at most one optimal triple* [*up to multiplication by constants in the sense of* (3.7)].

PROOF OF LEMMA 5. We calculate the derivatives of $\lambda^i(\cdot; g)$. Namely, applying the perturbation formula (2.5), (2.6), we have

$$\lambda^i(\theta^i + \Delta\theta^i; g) = \lambda^i(\theta^i; g) + \lambda^i(\theta^i; g)\Delta\theta^i \sum_x g(x)\pi^i(x; \theta^i, g) + O(\Delta\theta^2)$$

and hence for $i = 0, 1$,

$$\frac{\partial\lambda^i}{\partial\theta}(\theta^i; g) = \lambda^i(\theta^i; g) \sum_x g(x)\pi^i(x; \theta^i, g). \tag{4.14}$$

If $(\theta^{0*}, \theta^{1*}, g^*)$ is an optimal triple, then clearly g^* is optimal with respect to $(\theta^{0*}, \theta^{1*})$ and hence by Theorem 1, (3.8a, b) hold. Furthermore, if (3.8c) failed, then by (4.14), we would have $(\partial\lambda^i/\partial\theta)(\theta^i; g) > 0$ or less than 0 for $i = 0, 1$, and we could make both eigenvalues smaller by perturbing (θ^0, θ^1), contradicting the definition of optimality. Hence (3.8) is satisfied. □

PROOF OF LEMMA 6. Let θ_g^0 and θ_g^1 be as in (3.2), which were shown to exist for each g in the proof of Theorem 1; and let g_{θ^0,θ^1} be as in the proof of Lemma 3. The difference between the present case and Lemma 3 is that now we do not know that $\lambda(\theta_g^0, \theta_g^1, g)$ is convex in g. One can show, however, that $\lambda(\theta^0, \theta^1, g_{\theta^0,\theta^1})$ is a continuous function of (θ^0, θ^1). Also $\inf_g \lambda(a\theta^0, a\theta^1, g) = \inf_g \lambda(\theta^0, \theta^1, ag) = \inf_g \lambda(\theta^0, \theta^1, g)$ and hence $\lambda(a\theta^0, a\theta^1, g_{a\theta^0,a\theta^1}) = \lambda(\theta^0, \theta^1, g_{\theta^0,\theta^1})$. Therefore,

$$\inf_{\theta^0,\theta^1} \lambda(\theta^0, \theta^1, g_{\theta^0,\theta^1}) = \inf_{(\theta^0,\theta^1)\in K} \lambda(\theta^0, \theta^1, g_{\theta^0,\theta^1}) \tag{4.15}$$

for some compact $K \subset \mathbb{R}^2$. Hence the infimum in (4.15) is acheived for some

$(\theta^0, \theta^1) = \bar{\theta}$. But then (by the continuity)

$$\inf_{(\theta^0,\theta^1,g)} \lambda(\theta^0,\theta^1,g) = \inf_{\substack{\theta^0>0\\ \theta^1<0}} \inf_g \lambda(\theta^0,\theta^1,g) = \inf_{\substack{\theta^0>0\\ \theta^1<0}} \lambda\big(\theta^0,\theta^1,g_{\theta^0,\theta^1}\big) = \lambda\big(\bar{\theta};g_{\bar{\theta}}\big);$$

and we have the required existence. □

PROOF OF LEMMA 7. Suppose (θ^0, θ^1, g) and $(\hat{\theta}^0, \hat{\theta}^1, \hat{g})$ are two solutions. Consider the matrix

$$(4.16)\quad \{\bar{p}_i(x,y;\mu)\} = \big\{\exp\big[\mu\big[\theta^i g(x) - \hat{\theta}^i \hat{g}(x)\big] + \hat{\theta}^i \hat{g}(x)\big] p_i(x,y)\big\}$$

and let $\lambda_i(\mu)$ denote it maximal eigenvalue. For μ near 0, this is a perturbation of the maximal eigenvalue of $\{\exp[\hat{\theta}^i \hat{g}(x)]p_i(x,y)\}$, namely of $\lambda^i(\hat{\theta}^i; g)$. Let $f_i(x) = \theta^i g(x) - \hat{\theta}^i \hat{g}(x)$. Then

$$\begin{aligned}\{\bar{p}_i(x,y;\mu)\} &= \big\{\exp\big[\mu f_i(x)\big]\exp\big[\hat{\theta}^i \hat{g}(x)\big] p_i(x,y)\big\}\\ &= \big\{\exp\big[\hat{\theta}^i g(x)\big] p_i(x,y)\big[1 + \mu f_i(x) + O(\mu^2)\big]\big\}.\end{aligned}$$

By (2.5) and (2.6),

$$\begin{aligned}\lambda_i(\mu) &= \lambda^i(\hat{\theta}^i;\hat{g}) + \mu\lambda^i(\hat{\theta}^i;\hat{g})\sum_x f_i(x)\pi^i(x;\hat{\theta}^i,\hat{g})\\ &= \lambda^i(\hat{\theta}^i;\hat{g})\Big[1 + \mu\sum_x \big[\theta^i g(x) - \hat{\theta}^i \hat{g}(x)\big]\pi^i(x;\hat{\theta}^i,\hat{g})\Big]\end{aligned}$$

and by (3.8c) this becomes

$$= \lambda^i(\hat{\theta}^i;\hat{g})\Big[1 + \mu\theta^i\sum_x g(x)\pi^i(x;\hat{\theta}^i,\hat{g})\Big].$$

But note that since $(\hat{\theta}^0, \hat{\theta}^1, \hat{g})$ satisfies (3.8b),

$$\sum_x g(x)\pi^0(x;\hat{\theta}^0,\hat{g}) = \sum_x g(x)\pi^1(x;\hat{\theta}^1,\hat{g}).$$

But $\lambda^i(\hat{\theta}^i;\hat{g}) = \lambda_i(0)$ and hence by (4.17),

$$\begin{aligned}\lambda_i'(0) &= \lambda^1(\hat{\theta}^1;g)\theta^1\sum_x g(x)\pi^1(x;\hat{\theta}^1,\hat{g})\\ &= \lambda^0(\hat{\theta}^0;g)\theta^0\sum_x g(x)\pi^0(x;\hat{\theta}^0,\hat{g}).\end{aligned}$$

Now since θ^0 and θ^1 have opposite signs (by Lemma 1), $\lambda^i > 0$, $\theta^0\Sigma g(x)\pi^0(x;\hat{\theta},g)$ and $\theta^1\Sigma g(x)\pi^1(x;\hat{\theta}^1,g)$ also have opposite signs and hence $\lambda_0'(0)$ and $\lambda_1'(0)$ have opposite signs. Similarly, $\lambda_0'(1)$ and $\lambda_1'(1)$ have opposite

signs. But by convexity, this is impossible unless $\lambda_i(\mu)$ is constant and this forces $f_i(x) = 0$ for $i = 0, 1$. This in turn implies the lemma. □

Finally, Lemmas 5, 6, and 7 imply Theorem 2.

5. Computational aspects. Fix (θ^0, θ^1). Let $\bar{c} = (c, c, \ldots, c) \in \mathbb{R}^d$ and note that

$$l^i(x; g + \bar{c}) = l^i(x; g), \tag{5.1a}$$

$$r^i(x; g + \bar{c}) = r^i(x; g), \tag{5.1b}$$

$$\lambda^i(\theta; g + \bar{c}) = \exp[c\theta]\lambda^i(\theta; g). \tag{5.1c}$$

Note that one can always find a c such that $\lambda^0(\theta^0; g + \bar{c}) = \lambda^1(\theta^1; g + \bar{c})$ since with

$$c(g) = \frac{1}{\theta^0 - \theta^1} \log \frac{\lambda^1(\theta^1; g)}{\lambda^0(\theta^0; g)}, \tag{5.2}$$

we have $\lambda^0(\theta^0; g + \overline{c(g)}) = \exp[c(g)\theta^0]\lambda^0(\theta^0; g) = \exp[c(g)\theta^1]\lambda^1(\theta^1; g) = \lambda^1(\theta^1; g + \overline{c(g)})$. Write $\bar{g} = g + \overline{c(g)}$ and let g^* denote the unique solution of (3.6) in Theorem 1.

Let $L(g) = \{g + \bar{c}: c \in \mathbb{R}\}$ be the diagonal lines through g. Then

$$\bar{g} \text{ is a solution of (3.6a) for all } g \tag{5.3a}$$

and

$$\text{the solutions of (3.6b) are precisely the points of } L(g^*). \tag{5.3b}$$

(Note that $\overline{g^*} = g^*$). Assertion (5.3a) is obvious. That all points of $L(g^*)$ are solutions of (3.6b) follows from (5.1). Conversely, if $g \notin L(g^*)$, then $\bar{g} \neq \overline{g^*} = g^*$, while both $\bar{g}$ and g^* are solutions of (3.6a). Hence if $\bar{g}$ were also a solution of (3.6b), the uniqueness part of Theorem 1 would be contradicted.

Also note that $\lambda^i(\theta; cg) = \lambda^i(c\theta; g)$ and hence we may rescale so that

$$\theta^0 - \theta^1 = 1. \tag{5.4}$$

Assume from now on that this has been done. Denote by $\bar{\lambda}(g)$ the common value

$$\bar{\lambda}(g) = \lambda^0(\theta^0; \bar{g}) = \lambda^1(\theta^1; \bar{g}). \tag{5.5}$$

We can now state the computing algorithm.

ALGORITHM. *For fixed* (θ^0, θ^1), *define the sequence* $\{g_n(x) \in \mathbb{R}^d,\ n = 0, 1, \ldots\}$ *by*

$$g_{n+1}(x) = g_n(x) + \varepsilon\theta^0(1 - \theta^0)\bar{\lambda}(g_n)\left[\pi^0\left(x; \theta^0, g_n\right) - \pi^1\left(x; \theta^1, g_n\right)\right]. \tag{5.6}$$

Let g^* denote the unique solution of (3.6). *Then given any $\varepsilon' > 0$, we can choose $\varepsilon > 0$ such that*

$$|\bar{g}_n - g^*| < \varepsilon' \tag{5.7}$$

for sufficiently large n.

PROOF OF ALGORITHM. The previous conclusion will follow from the following facts:

$$D_f\bar{\lambda}(g) = \theta^0(1-\theta^0)\bar{\lambda}(g)\sum_x f(x)\left[\pi^0(x;\theta^0,g) - \pi^1(x;\theta^1,g)\right] \tag{5.8}$$

and

(5.9) the critical points of $\bar{\lambda}(\cdot)$ are all global minima.

To see that (5.7) follows from (5.8) and (5.9), note that (5.6) can be rewritten as

$$g_{n+1}(x) = g_n(x) + \varepsilon \operatorname{grad}\left(\bar{\lambda}(g_n)\right). \tag{5.10}$$

Thus by the method of steepest decent, $\{g_n\}$ will be within ε' of a global minimum of $\bar{\lambda}$ for large n and suitable ε. But by Theorem 1, (5.2) and the definition of $\bar{\lambda}$, the global minima of $\bar{\lambda}$ are precisely the points on the line $L(g^*)$. Thus for all sufficiently large n, g_n will be in an ε'-neighborhood of $L(g^*)$, which is equivalent to (5.7).

To prove (5.8), note first that by (5.2) and (5.4), $c(g) = \log[\lambda^1(\theta^1;g)/\lambda^0(\theta^0;g)]$ and hence by (2.5) and (2.6) and some calculating, one has

$$\begin{aligned} c(g+\epsilon f) &= c(g) + \varepsilon\theta^1\sum_x f(x)\pi^1(x;\theta^1,g) \\ &\quad - \varepsilon\theta^0\sum_x f(x)\pi^0(x;\theta^0,g) + O(\varepsilon^2). \end{aligned} \tag{5.11}$$

Now

$$\bar{\lambda}(g+\varepsilon f) = \lambda^0\left(\theta^0, g+\varepsilon f + \overline{c(g+\varepsilon f)}\right) = e^{c(g+\varepsilon f)\theta^0}\lambda^0(\theta^0, g+\varepsilon f),$$

and hence by (2.5), (2.7) and (5.11), we get after some further calculations, that

$$\begin{aligned} \bar{\lambda}(g+\varepsilon f) &= \bar{\lambda}(g) + \varepsilon\theta^0(1-\theta^0)\bar{\lambda}(g) \\ &\quad \times \sum_x f(x)\left[\pi^0(x;\theta^0,g) - \pi^1(x;\theta^1,g)\right] + O(\varepsilon^2), \end{aligned}$$

which implies (5.8).

Finally, we prove (5.9). The critical points g are those satisfying $D_f\lambda(g) = 0$ for all f, and hence for such points,

$$\pi^0(x;\theta^0,g) = \pi^1(x;\theta^1,g).$$

Hence by (5.1) also,

$$\pi^0(x;\theta^0,\bar{g}) = \pi^1(x;\theta^1,\bar{g}).$$

But $\bar{g}$ also satisfies (3.6) and by the converse part of Lemma 2, $\bar{g}$ is locally optimal with respect to (θ^0,θ^1).

Now g is a local minimum of $\bar{\lambda}(\cdot)$, for if not, there would be points g^1 arbitrarily close to g such that $\bar{\lambda}(g^1) > \bar{\lambda}(g)$, and hence by continuity, points $\bar{g}^1$ close to $\bar{g}$ such that

$$\bar{\lambda}(\bar{g}^1) = \bar{\lambda}(g^1) > \bar{\lambda}(g) = \bar{\lambda}(\bar{g}),$$

contradicting local optimality of $\bar{g}$.

Finally, we note that all the local minima of $\bar{\lambda}(\cdot)$ are in fact global minima. For suppose $\hat{g}$ is a local minimum. Then $\bar{\lambda}(g) = \bar{\lambda}(\hat{g})$ for $g \in L(\hat{g}) = \{\hat{g} + \bar{c};\ \bar{c} \in \mathbb{R}^d\}$ and all such g's are local minima. Suppose $g' \notin L(\hat{g})$ were another local minima. Then $\bar{g}' \neq \bar{g}$ will both satisfy (3.6), contradicting the uniqueness in Theorem 1. Thus all local minima are in the line $L(\hat{g})$ and are hence global minima. This proves (5.9) and the algorithm. □

Note that we must now perform a search in the (θ^0,θ^1) parameter set. We have available, in equation (4.14), an explicit representation of the derivative of the eigenvalues with respect to θ. In the computation of the examples we performed a steepest descent search in the (θ^0,θ^1) parameters. Note that the overall method is not a steepest descent since we do not go choose the largest gradient with respect to all parameters. Our method falls into the class of algorithms known as zig zag methods wherein one descends in each variable sequentially.

EXAMPLE 1 (Comparison with an iid approximation). The state transition matrices under the two hypotheses are

$$P^0 = \begin{bmatrix} 0.3 & 0.7 & 0.0 \\ 0.0 & 0.3 & 0.7 \\ 0.7 & 0.0 & 0.3 \end{bmatrix},$$

with stationary distribution $[\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}]$ and

$$P^1 = \begin{bmatrix} 0.0 & 0.2 & 0.8 \\ 0.0 & 0.2 & 0.8 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

with stationary distribution $[\frac{2}{25} \quad \frac{3}{25} \quad \frac{20}{25}]$. The memoryless detector based upon assuming the data is iid and distributed as the stationary distributions would be $g_{\text{iid}} = [\log(25/6) \quad \log(25/9) \quad \log(25/60)]$. If we add a constant shift $c(g_{\text{iid}})$ to optimize this test's performance, we find it to have an error rate of -0.176.

TABLE 1

μ_0	μ_1	Optimal rate	Memoryless rate
0.7	0.3	−0.919	−0.916
0.95	0.05	−0.454	−0.435
0.8	0.2	−0.807	−0.800
0.6	0.4	−0.981	−0.979
0.51	0.49	−0.9998	−0.9997
0.9	0.1	−0.614	−0.600
0.95	0.3	−0.721	−0.717
0.5	0.49	−0.99995	−0.99994
0.5	0.1	−0.895	−0.894

Using the steepest descent search technique, we find the optimal choice is $g^* = [2.6 \quad 0.79 \quad -1.3]$ with associated error rate of -0.189.

EXAMPLE 2 (Comparison with optimal detector). Consider the five state Markov chain transition matrix

$$\begin{bmatrix} \mu_i & 1-\mu_i & 0 & 0 & 0 \\ \mu_i & 0 & 1-\mu_i & 0 & 0 \\ 0 & \mu_i & 0 & 1-\mu_i & 0 \\ 0 & 0 & \mu_i & 0 & 1-\mu_i \\ 0 & 0 & 0 & \mu_i & 1-\mu_i \end{bmatrix}.$$

We consider the testing problem between two Markov chains labeled 0 and 1. Chain i has the previous transition matrix parameter μ_i, $i = 0, 1$. Table 1 compares the performance of the optimal test statistic versus the memoryless test statistic for various choices of μ_0 and μ_1. One can see that there appears to be little loss in using a memoryless detector for this particular example over a wide range of parameter values. If the previous five state example is replaced by an even state symmetric ($\mu_0 = 1 - \mu_1$) case, then one can show by a regeneration argument that the rate performance of the memoryless and optimal tests are identical even though the tests themselves are not.

REMARK. We computed the principal eigenvalues and associated eigenvectors of the twisted matrices by a modified power method. Since our matrices were irreducible and aperiodic, we were guaranteed that there was a strictly largest (in modulus) positive eigenvalue. However, we found that the iterative nature of the program tended to push the next largest (in modulus) eigenvalue out until machine accuracy failed to be able to distinguish between the relative sizes. In such a situation, a simple powering up of matrices must fail since essentially two eigenvalues of equal size are being excited. We found that shifting (i.e., adding a positive constant k times an identity matrix) was necessary. This operation has the effect of separating the largest positive

eigenvalue away from the others, has no effect on the eigenvectors and merely adds the constant k to the computed principal eigenvalue.

6. Discussion and other problems. One should not try to read too much from the previous examples, since every hypothesis testing problem is different. It is quite easy to generate examples where memoryless structures perform arbitrarily close to or far from the performance of the optimal structure. Clearly, if two Markov chains have the same stationary distribution, then the memoryless test cannot distinguish between them at all. On the other hand in the degenerate case of iid testing, the memoryless structure is optimum.

There are a number of variants of the basic problem that would be interesting and useful to consider. The authors have in the previous framework considered a Neyman–Pearson type of formulation where the type I error rate is fixed and one desires to minimize the type II error rate. We have also considered our problem when the observations are assumed to be continuous time Markov processes with a finite state space. The case of more general state spaces in discrete and continuous time remains unresolved at this time. An anonymous reviewer has suggested a generalization of the problem considered in the text. Suppose $\{X_i\}$ is an m-Markov chain. Suppose we are free to choose a real-valued function operating on any k $(k < m)$ past samples of the chain. One then sums the function values to obtain the test statistic. What is the optimal function to choose in the sense of maximizing the exponential error rate? It appears that many of the techniques we have used in the previous sections may be used here but technical problems remain with the existence and uniqueness properties. We are currently working on these. This problem would be of particular practical interest as Markov-m models are a very traditional class of noise models in communication systems.

Other questions hereto not investigated are the questions of robustness or sensitivity of the designed detector to the underlying noise models. Another question of interest is how to handle various classes of composite hypotheses in this setting. We feel that this is a rich problem area with a solid practical foundation for researchers interested in hypothesis testing.

REFERENCES

CHATELIN, F. (1983). *Spectral Approximation of Linear Operators*. Academic, New York.

DONSKER, M. and VARADHAN, S. (1975a). Asymptotic evaluation of certain Markov process expectations for large time. Part I. *Comm. Pure. Appl. Math.* **28** 1–47.

DONSKER, M. and VARADHAN, S. (1975b). Asymptotic evaluation of certain Markov process expectations for large time. Part II. *Comm. Pure Appl. Math.* **28** 279–301.

DONSKER, M. and VARADHAN, S. (1976). Asymptotic evaluation of certain Markov process expectations for large time. Part III. *Comm. Pure. Appl. Math.* **29** 389–461.

ELLIS, R. (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York.

HALVERSON, D. and WISE, G. (1984). Asymptotic memoryless discrete-time detection of φ-mixing signals in φ-mixing noise. *IEEE Trans. Inform. Theory* **IT-30** 189–198.

HAYKIN, S. (1989). *An Introduction to Analog and Digital Communications*. Wiley, New York.

LATHI, B. P. (1989). *Modern Digital and Analog Communication Systems*, 2nd ed. Holt, Rinehart and Winston, Philadelphia.

MILLER, H. (1961). A convexity property in the theory of random variables on a finite Markov chain. *Ann. Math. Statist.* **32** 1260–1270.

MILLER, J. and THOMAS, J. (1972). Detectors for discrete-time signals in non-Gaussian noise. *IEEE Trans. Inform. Theory* **IT-18** 241–250.

NEY, P. and NUMMELIN, E. (1987a). Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Probab.* **15** 561–592.

NEY, P. and NUMMELIN, E. (1987b). Markov additive processes. II. Large deviations. *Ann. Probab.* **15** 593–609.

POOR, H. and THOMAS, J. (1979). Memoryless discrete-time detection of a constant signal in m-dependent noise. *IEEE Trans. Inform. Theory* **IT-25** 54–61.

RODEN, M. (1988). *Digital Communications System Design*. Prentice-Hall, Englewood Cliffs, N.J.

SADOWSKY, J. and BUCKLEW, J. (1986). A nonlocal approach for asymptotic memoryless detection. *IEEE Trans. Inform. Theory* **IT-32** 115–120.

WOOLEY, B. (1988). Special issue on logic and memory. *IEEE J. Solid-State Circuits* **23**.

DEPARTMENT OF ELECTRICAL AND
COMPUTER ENGINEERING
UNIVERSITY OF WISCONSIN–MADISON
MADISON, WISCONSIN 53706

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN–MADISON
MADISON, WISCONSIN 53706