Detection of a Randomly Modulated Periodic Signal

Melvin J. Hinich and John W.C. Robinson

This draft December 25, 1999

Abstract

Detection of narrow-band signals in noise is often done by identifying high peaks in the periodogram (spectrogram) with the presence of a signal. Mathematically this can be motivated by Fisher's (1929) test (and relatives) in which a ratio between the largest value and the average of values of the periodogram in a certain frequency interval is compared to a threshold. However, the (normalized quadratic) structure of the test statistic then employed is postulated and not derived from underlying assumptions, which in particular implies loss of phase information. Moreover, the signal model implicit in this approach is for many applications unrealistic since it assumes perfect sinusoids (with no amplitude/frequency variation). Here we develop an alternative, Fourier-domain based, class of tests derived from a more realistic model for narrow-band signals utilizing so called Randomly Modulated Periodicites (RMPs). We show that the proposed class of tests can realize optimal detection according to variants of several of the standard criteria (e.g. Neyman-Pearson and deflection) and we suggest a frame-wise adaptive detector that also utilizes phase information.

1 Introduction

Detection of narrow-band signals in noise by Fourier spectral methods is a central problem in many applications of statistical time-series analysis. The mathematical treatment of the problem has its roots in the search for 'hidden periodicities' in time-series data initiated in the late 19:th century. One of the first to present a rigorously derived solution to this decision problem was R.F. Fisher [1] who formulated it in terms of a statistical test based on the periodogram (or more correctly, spectrogram) of the signal. The test, which later has become known as Fisher's test [2], [3]–[5], determines the presence of a narrow-band signal in Gaussian white noise by comparing the value of the periodogram of at a certain frequency with its average over all frequencies. If there is a significant deviation from the average at a certain frequency then the decision is made that there is a narrow-band component present at this frequency. The idea of basing the decision on a ratio between the value of the periodogram at a certain frequency and the average over neighboring frequencies has prevailed and been developed in various directions [6]–[9], in particular the whiteness and Gaussianity conditions on the noise were relaxed [10], [11] (c.f. also [4]). One important application is moving target detection/indication in radar [12].

Starting with Fishers seminal paper, a common model for what is called a narrow-band signal in time-series analysis and signal processing is a sinusoid whose amplitude and frequency are assumed to be fixed (but often unknown). A pure sinusoid should really be called a zero-band signal though, since it has no bandwidth; it is a 'line' in the spectrum. The simplistic zero-band model is a useful first approximation to a truly narrow-band signal and can serve to rationalize the periodogram as the statistic to be used in detecting a sinusoid in additive noise. A more realistic model for a narrow-band signal, however, requires a model of the modulation which produces a non zero bandwidth of the signal. [1] The 'line broadening' alluded to here is a general phenomenon which is always present in real life periodic signals as a consequence of the inevitable random modulation effects in the generation mechanism. For example, the rotational speed of a motor which is operating at a fixed load has some variation over time around the mean rpm of the shaft. Thus, a time-series over the acoustic pressure generated by its vibration will contain a signal which will have some positive bandwidth around the rotation frequency. In view of this it is clear that any efficient detection strategy for real life narrow-band signals must be based on a statistical model of the modulation of the signal, and this is the motivation behind the approach taken in the present work.

The class of models we employ to analyze the detection problem for narrow-band signals in noise is called randomly modulated periodicities (RMP). An RMP is defined in terms of the coefficients of a discrete Fourier transform of the signal, essentially as a signal-plus-noise model of the signal itself but in the Fourier domain. This makes it very easy to analyze the statistical properties of the narrow-band components and to address the associated hypothesis testing problem. A Fourier-transform based approach also enables us to look at the Fisher statistic in a new light and reexamine some of the conclusions about the optimality of quadratic test statistics often drawn (implicitly) from it. We shall show that there are indeed circumstances under which a Fourier-domain based test statistic should be quadratic but that in general the optimal detector is linear-quadratic. This qualitative conclusion holds for several different optimality criteria and formulations of the detection problem. It turns out that the structure is somewhat subtly connected to the type of a priori knowledge available about the underlying signal at the time of detection and we shall advocate a mixed

[1] As E.A. Robinson [13] puts it (on the concept of line spectra in optics): "Real 'lines' have finite with. This means that real lines behave like narrow-band noises and not like either single frequencies or a constant-amplitude lightly frequency-modulated signal."

linear-quadratic type of detector where the structure is allowed to change with time (even in a stationary scenario). Allowing the structure of the detector to change is very reasonable in Fourier-domain based detection since the signal processing in then normally done frame-wise, on data partitioned into (possibly overlapping) time frames. This means that the available a priori knowledge about the signal, like its mean phase relative to the frames, is likely to change from frame to frame, and by incorporating such information one can improve on a detector which depends only on the energy in a frequency band.

2 Randomly Modulated Periodicities

As a basic model for signals comprised of narrow-band components we will take (real-valued) signals $x(t), t \in \mathbb{R}$ of the form

$$x(t) = \frac{1}{2K+1} \sum_{k=-K}^{K} \big(s_k + u_k(t)\big) \exp(i2\pi f_k t), \quad f_k = \frac{k}{T}, \tag{1}$$

where s_k are (deterministic) complex numbers with the symmetries $s_{-k} = s_k^*$ and $u_k(t)$ are (possibly jointly dependent) zero mean fourth order random processes satisfying the symmetry condition $u_{-k}(t) = u_k^*(t)$, for $k = -K, \ldots, K$ and all times t. Time t is here naturally divided into *frames*, the m:th time frame F_m being the interval $[mT, mT+T-1)$, and the signals in (1) are assumed to posess two properties relating to this division; *periodic frame stationarity* and *finite dependence*. Periodic frame stationarity means that the distribution of $\big(u_{k_1}(t_1), \ldots, u_{k_n}(t_n)\big)$ is the same as that of $\big(u_{k_1}(t_1+T), \ldots, u_{k_n}(t_n+T)\big)$, for any indices $k_1, \ldots, k_n$ and times $t_1, \ldots, t_n$ in a frame F_m. Finite dependence concerns the joint distributions of $(u_{j_1}(t_1), \ldots, u_{j_m}(t_m))$ and $\big(u_{k_1}(\tau_1), \ldots, u_{k_n}(\tau_n)\big)$ which are required to be independent for any indices $j_1, \ldots, j_m, k_1, \ldots, k_n$ and time points $t_1 \leq \cdots \leq t_m < \tau_1 \leq \cdots \leq \tau_n$ such that $t_m + D < \tau_1$, for some maximal range of dependence D. Signals of the form (1) are called *randomly modulated periodicities* (RMP) (with period T) [14].

For a randomly modulated periodicity $x(t)$ we shall refer to the mean of $x(t)$ as the *periodic component* $s(t)$ and the deviation of $x(t)$ from its mean as the *modulation component* $u(t)$. In other words, $x(t)$ can be decomposed into its periodic and modulation components, respectively, as

$$x(t) = s(t) + u(t) \tag{2}$$

where

$$s(t) = \frac{1}{2K+1} \sum_{k=-K}^{K} s_k \exp(i2\pi f_k t), \; u(t) = \frac{1}{2K+1} \sum_{k=-K}^{K} u_k(t) \exp(i2\pi f_k t).$$

The RMP structure can provide a more realistic model in many real world scenarios containing signals with periodicities because it makes it possible to explicitly take into account the often present amplitude–phase 'jitter' of such signals in a straightforward way. Once the statistics of the modulation component are specified the RMP approach yields a *model* for the variation in amplitude and phase which is expressed additively, as a deviation from a nominal pure sinusoidal behaviour. By doing so the inherent structure of the associated likelihood ratio test becomes much clearer and easier to exploit when carried over into the Fourier domain. An RMP can be viewed as a special case of a general (strongly) cyclostationary process [15] where the variations in amplitude and phase are related to the division onto frames.

The signal model (1),(2) describes the noise-free case. In practice, however, all measurements are corrupted by noise. This is introduced in the data model in the standard fashion as additive noise. Thus, the noise corrupted randomly modulated periodicity $y(t)$ is defined as

$$y(t) = x(t) + n(t),$$

where $n(t)$ is a (real) noise process to be specified later. At this point it is convenient to pass over entirely to the Fourier domain and in doing so we will also discretize time, i.e. we will use the discrete Fourier transform (DFT) formulation. For simplicity we assume (uniform) sampling at the Nyqvist rate (sampling interval $T/2K$) which means that there are $2K$ discrete time samples in a time interval of length T. We may thus from now on assume that time is normalized so that $2K = T$ and that the sampling points are positioned so that one sample is taken exactly at time zero. The discrete time signal thereby obtained is likewise divided into time frames, each consituting T consecutive samples, where the m:th frame is given by the points $mT, mT+1, \ldots, mT+T-1$. By the assumptions the distribution of the samples is the same for each frame.

The DFT of $x(t)$ over the 0:th frame evaluated at frequency $f_r = r/T$ is denoted $X(r)$ and is given by

$$X(r) = \sum_{t=0}^{T-1} x(t) \exp(-i2\pi f_r t) = S(r) + U(r), \tag{3}$$

where $S(r)$ is the corresponding DFT of the periodic component $s(t)$ and $U(r)$ is the DFT of the random modulation $u(t)$, i.e.

$$S(r) = \sum_{t=0}^{T-1} s(t) \exp(-i2\pi f_r t), \quad U(r) = \frac{1}{T} \sum_{t=0}^{T-1} u(t) \exp(-i2\pi f_r t). \tag{4}$$

Similarly we will denote the DFT of the noise process $n(t)$ over frame 0 at frequency f_r by $N(r)$ and we have

$$Y(r) = X(r) + N(r), \tag{5}$$

where $Y(r)$ is the resulting DFT of the noise corrupted RMP. It is useful to note here that by the linearity of the DFT and the expectation operation the mean value of $X(r)$, which is $S(r)$, is the DFT of the mean value $s(t)$ of $x(t)$, and that the deviation $U(r)$ of $X(r)$ from its mean is the DFT of the deviation $u(t)$ of $x(t)$ from its mean. In the following we will use (3) as the representation of an RMP and (4),(5) for its noise corrupted counterpart.

3 Detection of an RMP

In order to address the problem of detecting an RMP in noise we need to impose some further statistical assumptions on data. The most important assumption will be that the RMP in the Fourier domain is Gaussian distributed. This can be motivated in several ways. First, in many applications it is reasonable to assume that the modulation components $u_k(t)$ for $k = -K, \ldots, K$ are jointly (complex) normal [2] which implies that the same holds for the different frequency bins of $U(r)$ (also after a possible averaging of frames) and thereby also for $X(r)$. Second, if data is not normal in the first place it can often be made so by frame averaging. An average of L frames of noise-free DFT data will be (pointwise) asymtotically (complex) normal as $L \to \infty$ because of the M-dependence [17, sec. 27] induced by the assumptions. Therefore, we will henceforth assume that $X(r)$ for all frequencies is multivariate Gaussian. [3] We shall also make the assumption that the noise $n(t)$ is multivariate Gaussian, so that this holds for all frequencies of $N(r)$, and that $N(r)$ is independent of $X(r)$ (across frequencies). When formulated in the Fourier domain like this, our problem of detecting presence of a given frequency component can therefore naturally be formulated as a test to discriminate between two Gaussian hypotheses. In order to simplify notation we shall denote by $\mathbf{X}(r)$ the two-dimensional random vector obtained by simply stacking the real and imaginary parts of $X(r)$ in a vector and similarly we shall denote by $\mathbf{N}(r)$ the two-dimensional vector obtained from $N(r)$.

3.1 Likelihood Ratio Test

To begin with we shall assume that the hypotheses testing problem is formulated as a test between two simple hypotheses for a single–frequency RMP.

[2] With complex normal we here mean simply that the real and imaginary parts are bivariate normal. A modulation component $u_k(t)$ with jointly complex real and imaginary parts has a Rayleigh distributed absolute value and uniformly distributed complex argument. This form of modulation is known as 'Swerling's case I' in radar detection [16], [12].

[3] It may then be that $X(r)$ is computed via frame averaging, and is thus really an average, but since it is not necessary for the detection theory to make the distinction we shall not do so.

In other words, we shall assume that the (mean) frequency and (mean) phase of the single sinusoid comprising the periodic component $s(t)$ is known. The null hypothesis H_0 is that the data observed consists only of stationary zero mean Gaussian noise and the alternative H_1 is that in addition to the noise there is a single–frequency RMP present in the observations. More precisely, in terms of $\mathbf{Z}(r)$, the two-dimensional vector formed of the real and imaginary parts of the DFT of data at frequency r, the two hypotheses are

$$\begin{array}{lrcl} H_0: & \mathbf{Z}(r) & = & \mathbf{N}(r) \sim N(\mathbf{0}, \mathbf{\Sigma}_N(r)) \\ H_1: & \mathbf{Z}(r) & = & \mathbf{X}(r) + \mathbf{N}(r) \sim N\big(\mathbf{S}(r), \mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r)\big), \end{array} \tag{6}$$

where $\mathbf{S}(r)$ and $\mathbf{\Sigma}_U(r)$ are, respectively, the mean vector and (nonsingular) covariance matrix of the random variable $\mathbf{X}(r)$ under H_1, and $\mathbf{\Sigma}_N(r)$ is similarly the (nonsingular) covariance matrix of $\mathbf{N}(r)$ under H_0, H_1. As a test statistic which is sufficient for the likelihood ratio we may thus consider $\ell(\mathbf{Z}(r))$ given by

$$\ell(\mathbf{Z}(r)) = \log\Big(\frac{p_1(\mathbf{Z}(r))}{p_0(\mathbf{Z}(r))}\Big), \tag{7}$$

where $p_0(\mathbf{z}), p_1(\mathbf{z})$ are Gaussian probability densities with mean $\mathbf{0}$ and $\mathbf{S}(r)$, respectively, and covariances $\mathbf{\Sigma}_N(r)$ and $\mathbf{\Sigma}_U(r)+\mathbf{\Sigma}_N(r)$, respectively. Written out, the expression for $\ell(\mathbf{Z}(r))$ takes the well-known form

$$\begin{aligned} \ell(\mathbf{Z}(r)) &= -\frac{1}{2}\big(\mathbf{Z}(r) - \mathbf{S}(r)\big)^T(\mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r))^{-1}(\mathbf{Z}(r) - \mathbf{S}(r)) \\ &\quad + \frac{1}{2}\mathbf{Z}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{Z}(r) - \det\big(\mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r)\big)^{1/2} + \det(\mathbf{\Sigma}_N(r))^{1/2} \\ &= \mathbf{v}(r)^T\mathbf{Z}(r) + \frac{1}{2}\mathbf{Z}^T(r)\mathbf{M}(r)\mathbf{Z}(r) + C(r), \end{aligned} \tag{8}$$

where the vector $\mathbf{v}(r)$ and matrix $\mathbf{M}(r)$ are given by

$$\mathbf{v}(r) = \mathbf{S}^T(r)\big(\mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r)\big)^{-1}, \quad \mathbf{M}(r) = \mathbf{\Sigma}_N^{-1}(r) - \big(\mathbf{\Sigma}_{U(r)} + \mathbf{\Sigma}_{N(r)}\big)^{-1}$$

and $C(r)$ is the constant

$$\begin{aligned} C(r) = -\frac{1}{2}\mathbf{S}^T(r)\big(\mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r)\big)^{-1}\mathbf{S}(r) \\ - \det\big(\mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r)\big)^{1/2} + \det\big(\mathbf{\Sigma}_N(r)\big)^{1/2}. \end{aligned}$$

An optimal test, or detector, in the Neyman-Pearson (N-P), Bayes or minimax sense is thus

$$\mathbf{v}^T(r)\mathbf{Z}(r) + \frac{1}{2}\mathbf{Z}^T(r)\mathbf{M}(r)\mathbf{Z}(r) \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma(r), \tag{9}$$

where $\gamma(r)$ is a suitable threshold.

Hence, the optimal likelihood ratio detector for the basic problem (6) is *linear–quadratic* and it is worthwhile to make some observations pertaining to this fact. First, in the case of a vanishing modulation component $U(r)$ (for fixed nonzero mean $\mathbf{S}(r)$) the detector in (9) degenerates into a linear form, the same basic structure as in the so called *matched filter*, or correlator detector, for detection of a known deterministic signal in Gaussian noise ("coherent" detection [18]). Indeed, by writing out the defining relations for the matched filter detector for detection of a sinusoid with known parameters in Gaussian noise based on T samples of the time series with the 'in-phase' and 'quadrature' components separated one easily sees that the detector (9) really represents the matched filter detector for the underlying time series. Another, much more interesting, observation about the vanishing modulation component case is that the test then becomes a simple mean test. This is interesting because it indicates that we by Fourier transformation have moved the problem of detecting a (perfect) sinusoid into the 'right' domain since a test for a change in means is essentially the simplest one possible. Second, when $\mathbf{S}(r)$ approaches zero (for a given nonzero $\mathbf{\Sigma}_U(r)$) we obtain instead a purely quadratic test, a weighted variant of the *energy detector*, which is a structure that is optimal for discriminating between two Gaussian distributions with the same mean. Given this fact it is more or less immediate to realize that a quadratic test of this type in general must be far from optimal for detecting a pure sinusoid in noise in the Fourier domain, especially when the noise level is not very high. Still, as mentioned in the introduction, this is often what is done in ad hoc approaches to the detection problem based on searching for peaks in the periodogram. Taken together these observations make it intuitively clear why an optimal likelihood ratio test for the present signal and noise model in general must be of the mixed (location–dispersion) form in (9): The linear part detects the offset in mean at a single frequency in the DFT due to the periodic component and the quadratic part detects the difference in covariance at the frequency incurred by the Gaussian random modulation component.

3.2 Locally Most Powerful Test

The hypothesis testing situation considered above, with two simple hypotheses, is the most basic situation of hypotheses testing. It is, however, for many real world applications unrealistic. A natural question is therefore if any of the structural and qualitative properties of the detection problem above remain for other settings of the problem, such as variants with a composite alternative. For detection of weak, partially known, deterministic signals in noise the perhaps most useful formulation in the Neyman-Pearson framework is that of a locally most powerful (LMP) test [18]. A reasonable composite alternative hypothesis $\tilde{H}_1$ for detection of a weak partially known single-frequency RMP in noise is one where the amplitude of the periodic component and the covariance of the modulation component both have a linear parameter dependence. This amounts to replacing $\mathbf{S}(r)$ and $\mathbf{\Sigma}_N(r)$ in (6) by $\theta_1\mathbf{S}(r)$ and $\theta_2\mathbf{\Sigma}_N(r)$, respectively, for some $\theta_1, \theta_2 > 0$, which renders an alternative hypothesis $\tilde{H}_1$ of the form

$$\tilde{H}_1: \quad \mathbf{Z}(r) = \mathbf{X}_\Theta(r) + \mathbf{N}(r) \sim N\big(\theta_1\mathbf{S}(r), \theta_2\mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r)\big), \tag{10}$$

where $\mathbf{X}_\Theta(r)$ is the parametrized RMP and $\Theta = (\theta_1, \theta_2)$ is the parameter vector. The probability distribution of $\mathbf{Z}(r)$ under $\tilde{H}_1$ is denoted P_Θ, with density $p_\Theta(\mathbf{z})$, and, since we have $\tilde{H}_1 = H_0$ for $\Theta = \mathbf{0}$, the distribution of $\mathbf{Z}(r)$ under H_0 is accordingly denoted $P_\mathbf{0}$, with density $p_\mathbf{0}(\mathbf{z})$. The LPM detector for an alternative hypothesis of this type is obtained by the following variant of the standard technique.

Given a value $\Theta \neq \mathbf{0}$ and a test statistic $f(\mathbf{Z}(r))$ of data $\mathbf{Z}(r)$ (with prescribed threshold γ), denote the resulting false alarm (type I error) by P_F and the probability of detection (power of the test) by P_D. We then have $P_F = P_\mathbf{0}(\mathcal{A})$ and $P_D = P_\Theta(\mathcal{A})$, where $\mathcal{A} = \{\mathbf{z} \in \mathbb{R}^2 : f(\mathbf{z}) > \gamma\}$ is the acceptance region, and P_D can be expanded in a Taylor series near $\Theta = \mathbf{0}$, see appendix A, as

$$\begin{aligned} P_D &= P_\Theta(\mathcal{A}) \\ &= P_\mathbf{0}(\mathcal{A}) + \theta_1\frac{\partial}{\partial\theta_1}P_\Theta(\mathcal{A})\big|_{\Theta=\mathbf{0}} + \theta_2\frac{\partial}{\partial\theta_2}P_\Theta(\mathcal{A})\big|_{\Theta=\mathbf{0}} + \mathcal{O}(\|\Theta\|^2) \\ &= P_F + \theta_1\frac{\partial}{\partial\theta_1}P_\Theta(\mathcal{A})\big|_{\Theta=\mathbf{0}} + \theta_2\frac{\partial}{\partial\theta_2}P_\Theta(\mathcal{A})\big|_{\Theta=\mathbf{0}} + \mathcal{O}(\|\Theta\|^2). \end{aligned} \tag{11}$$

For a given significance level α and (small magnitude) parameter Θ, a locally optimal test statistic can be defined as one that maximizes the sum of the two middle terms on the right in (11) subject to $P_F \leq \alpha$, for some class of deviations from $\Theta = \mathbf{0}$. If we consider only local deviations of the form $\theta_2 = \eta\theta_1$, where η is some positive constant, the expansion in (11) can be written as (appendix A)

$$\begin{aligned} P_D &= P_F \\ &\quad + \theta_1\int_\mathcal{A}\big(\mathbf{S}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + c_1\big)\,p_0(\mathbf{z})\,d\mathbf{z} \\ &\qquad + \mathcal{O}(\|\Theta\|^2), \end{aligned} \tag{12}$$

where c_1 is a constant (not depending on $\mathcal{A}$). Therefore, we define here a *locally optimal detector* as a test that maximizes the integral on the right in (12) subject to $P_F \leq \alpha$. Now, the integrand in the integral in (12) can be renormalized to yield a probability density and this enables us to invoke the Neyman-Pearson lemma for the optimization problem. If we do this, a maximizing test statistic $\tilde{f}$ emerges as

$$\tilde{f}(\mathbf{z}) = \mathbf{S}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + c_1, \tag{13}$$

where the corresponding threshold $\tilde{\gamma}$ is choosen such that $P_F = \alpha$. By adjusting $\tilde{\gamma}$ the constant [4] c_1 may be dropped in (13) and sufficient statistic for a locally most powerful test for H_0 versus (any fixed member of the family) $\tilde{H}_1$ is therefore

$$\mathbf{S}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{Z}(r) + \frac{\eta}{2}\mathbf{Z}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{Z}(r), \tag{14}$$

with performance given by (12) with $f = \tilde{f}, \gamma = \tilde{\gamma}$.

The test statistic in (14) is of the same linear–quadratic form as the one on the left of (9). Consequently, all the qualitative remarks made in the previous section carry over also to this case. In particular, the mixed linear–quadratic structure is a result of the more realistic assumption of uncertainty in both location and size of the deviation from the null, or nominal. This holds for any ratio η between the deviation in θ_2 and θ_1 near $(0, 0)$, and any value of $\mathbf{\Sigma}_N(r)$. In the important special case where $\mathbf{\Sigma}_U(r)$ is of the form $\sigma^2\mathbf{I}$, which amounts to a total lack of (statistical) syncronization between the modulation component $U(r)$ and the periodic component $S(r)$ (with all its spectral support at r), the quadratic part in (14) becomes

$$\eta\frac{\sigma^2}{2}\|\mathbf{\Sigma}_{N(r)}^{-1}\mathbf{Z}(r)\|^2. \tag{15}$$

A subsequent change of coordinates $\boldsymbol{\xi}(r) = \mathbf{\Sigma}_N^{-1}(r)\mathbf{Z}(r)$ shows that the detector obtained from (14) then becomes a weighted sum of a correlation detector and an energy detector. Moreover, in the equally important special case where the noise lacks the same kind of syncronization, so that also $\mathbf{\Sigma}_N(r)$ is a multiple of the identity, the detector collapses into correlator-energy detector form even without a change of coordinates. In practice one can interpret the value of σ^2 as a prior expressing how much deviation in size relative to the deviation in location is anticipated in the signal part of $\mathbf{Z}(r)$ under $\tilde{H}_1$ and therefore η can be absorbed into σ^2. This remark also generalizes to the case in (14) if we replace σ^2 by $\mathbf{\Sigma}_U(r)$.

3.3 Random Parameters

So far we have assumed that $S(r)$ (or equivalently its vector form $\mathbf{S}(r)$) is completely known, or known apart from a scaling factor as expressed by the parameter θ_1. In other words, we have assumed that the complex argument of $S(r)$ is known, i.e. that the phase of the periodic component is known.

[4]For future reference we note also that c_1 does not depend on $\mathbf{S}(r)$.

This is another assumption which in many applications may be questionable and thus is desirable to remove. A standard way to deal with this is to take a Bayesian view in which the phase of the periodic component is considered to be random. Then $\mathbf{S}(r)$ becomes a random vector with a probability distribution dG and the alternative hypothesis, denoted $\tilde{H}_1^Q$, can be written

$$\tilde{H}_1^Q: \quad \mathbf{Z}(r) = \mathbf{X}_\Theta(r) + \mathbf{N}(r) \sim Q_\Theta, \tag{16}$$

where the distribution Q of $\mathbf{Z}(r)$ under $\tilde{H}_1^Q$ has a density

$$q(\mathbf{z}) = \int_{\mathbb{R}^2} p_\Theta(\mathbf{z}|\mathbf{S}(r) = \mathbf{s})\, dG(\mathbf{s}). \tag{17}$$

Here $p_\Theta(\mathbf{z}|\mathbf{S}(r) = \mathbf{s})$ is the conditional probability density of $\mathbf{Z}(r)$ given $\mathbf{S}(r) = \mathbf{s}$, which is of the form (10). The probability of detection Q_D for a test statistic f (with threshold γ) and acceptance region $\mathcal{A}$ now becomes

$$\begin{aligned} Q_D &= Q(\mathcal{A}) \\ &= \int_{\mathcal{A}} q(\mathbf{z})\, d\mathbf{z} \\ &= \int_{\mathcal{A}} \int_{\mathbb{R}^2} p_\Theta(\mathbf{z}|\mathbf{S}(r) = \mathbf{s})\, dG(\mathbf{s}) d\mathbf{z} \\ &= \int_{\mathbb{R}^2} \int_{\mathcal{A}} p_\Theta(\mathbf{z}|\mathbf{S}(r) = \mathbf{s})\, d\mathbf{z} dG(\mathbf{s}) \\ &= \int_{\mathbb{R}^2} P_\Theta^{(\mathbf{s})}(\mathcal{A})\, dG(\mathbf{s}), \end{aligned} \tag{18}$$

where the interchange in order of integration in the third equality can be justified with Fubini's theorem and

$$P_\Theta^{(\mathbf{s})}(\mathcal{A}) = \int_{\mathcal{A}} p_\Theta(\mathbf{z}|\mathbf{S}(r) = \mathbf{s})\, d\mathbf{z}.$$

This last quantity we recognize as the conditional probability of detection given $\mathbf{S}(r) = \mathbf{s}$ and it follows that Q_D is a continuous convex combination (average) of detection probabilities $P_\Theta^{(\mathbf{s})}(\mathcal{A})$ of a type encountered in the previous section. Also for the present setting a locally most powerful test can be defined as one that maximizes Q_D near $\Theta = \mathbf{0}$ for deviations of the form $\theta_2 = \eta\theta_1$, $\eta > 0$, subject to a false alarm constraint. In order to find such a test we can exploit the close connection between Q_D and P_D and mimic the procedure used to find the statistic (13).

If we Taylor expand $P_\Theta^{(\mathbf{s})}(\mathcal{A})$ in Θ near $\mathbf{0}$ like in (11) and insert in (18) we obtain

$$\begin{aligned} Q_D &= P_{\mathbf{0}}(\mathcal{A}) \\ &\quad + \theta_1 \int_{\mathbb{R}^2} \Big(\frac{\partial}{\partial \theta_1} P_\Theta^{(\mathbf{s})}(\mathcal{A})\big|_{\Theta=\mathbf{0}} + \eta \frac{\partial}{\partial \theta_2} P_\Theta^{(\mathbf{s})}(\mathcal{A})\big|_{\Theta=\mathbf{0}}\Big)\, dG(\mathbf{s}) + \mathcal{R}_1(\Theta), \end{aligned} \tag{19}$$

where we have assumed that the rest term in the expansion of $P_\Theta^{(\mathbf{s})}$ is integrable dG so that $\mathcal{R}_1(\Theta)$ is well defined. [5] In this case $\mathcal{R}_1(\Theta)$ will be of order $\mathcal{O}(\|\Theta\|^2)$. Now, the probability of false alarm Q_F is here given by $Q_F = Q_\mathbf{0}(\mathcal{A}) = P_\mathbf{0}(\mathcal{A})$ so the steps in the previous section (and appendix A) can be traversed once more to obtain an expresssion for the integrand in (19), which upon insertion gives

$$\begin{aligned} Q_D &= Q_F \\ &+ \theta_1 \int_{\mathcal{A}} \left(E_G(\mathbf{S}^T(r))\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + c_1 \right) p_0(\mathbf{z})\, d\mathbf{z} \\ &\qquad + \mathcal{R}_1(\Theta), \end{aligned} \tag{20}$$

where $E_G(\mathbf{S}^T(r))$ is the mean of $\mathbf{S}^T(r)$ under dG and we have used Fubini's theorem once. The expression (20) is of exactly the same form as (12) and we can again invoke the Neyman-Pearson lemma to obtain a maximizing test statistic $\tilde{f}_Q$ as

$$\tilde{f}_Q(\mathbf{z}) = E_G(\mathbf{S}^T(r))\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + c_1, \tag{21}$$

where the corresponding threshold $\tilde{\gamma}_Q$ is chosen such that $Q_F = \alpha$. Also here we can drop the constant c_1 in (21) by adjusting the threshold $\tilde{\gamma}_Q$ and it follows that an optimal test statistic of data $\mathbf{Z}(r)$ exists in the present setting which is identical with the one in (14) if $\mathbf{S}(r)$ replaced by $E_G(\mathbf{S}(r))$ and has performance given by (20) for $f = \tilde{f}_Q, \gamma = \tilde{\gamma}_Q$. Moreover, the qualitative remarks made in connection with (15) regarding the quadratic part carry over also to this case. If $E_G(\mathbf{S}(r)) = \mathbf{0}$ the first term in (21) vanishes and this happens for instance if the distribution of $\mathbf{S}(r)$ is radially symmetric, i.e. a uniformly distributed phase. In this case a locally optimal test statistic $\tilde{f}_Q$ thus exists which is purely quadratic. Uniform phase is often a realistic modeling assumption since it corresponds to a total lack of (statistical) synchronization beteen the periodic component of the RMP and the clock governing the sampling. [6] Finally, since one can model uncertainty in the magnitude of $\mathbf{S}(r)$ by variations in θ_1 we can without much loss of generality assume that the distribution dG has support in a bounded region in $\mathbb{R}^2$. In particular, in the uniform phase case we can assume that it is concentrated on the unit circle in $\mathbb{R}^2$. The rest term $\mathcal{R}_1(\Theta)$ in (20) will then be of order $\mathcal{O}(\|\Theta\|^2)$ as assumed.

3.4 Deflection Index

The classical formulations of hypothesis testing (N-P, Bayes, minimax) require that the probability distributions of data under the various hyptheses

[5] This is a mild assumption and is for instance true if the the measure dG has bounded support in $\mathbb{R}^2$, which is both reasonable and natural to assume as we shall see shortly.

[6] In Bayesian terms this corresponds to no prior information about the phase.

are completely known. In many practical applications this is a severe obstacle in the design and evaluation of detectors/tests and therefore a number of alternative criteria of optimality with corresponding detection strategies have been developed. Among the easiest to apply in practice are (second-order) moment based criteria, of which the *deflection* is central [19]. The deflection $D(g)$ of a (sufficiently integrable) test statistic g of data is defined as

$$D(g) = \frac{\big(E_1(g) - E_0(g)\big)^2}{V_0(g)}$$

where $E_0(g), E_1(g)$ are, respectively, the mean of g under the null and alternative hypothesis, and $V_0(g)$ is the variance under the null hypothesis. It is easy to see that $D(\mu g+\nu) = D(g)$ for any constants μ, ν such that $\mu \neq 0$ and we shall therefore henceforth assume that all test statistics g have shifted means so that $E_0(g) = 0$. The deflection criterion represents an alternative performance index of a detector which is not based on probabilities but on a signal-to-noise ratio, or mean square error [20]. However, optimal detectors based on deflection are closely related to those obtained by the classical theory. Indeed, it is well-known [19] that unconstrained maximization of $D(g)$ yields the likelihood ratio as the optimal solution. Since the linear–quadratic statistic in (9) is sufficient for (8) it follows that the test in (9) is equivalent to a test which is optimal for deflection for the RMP detection problem in (6).

As mentioned before however, it is often more realistic to consider composite alternative hypotheses and in particular it is desirable to find detectors that are are locally optimal for alternatives in a neighboorhood of the null hypothesis. Such detectors can be developed also under the deflection criterion of optimality but one must proceed somewhat differently than in the N-P setting. The fact that the likelihood ratio is an optimal test statistic also for deflection and the intuitive idea that a smooth map (like the logarithm) should not affect the structure of a locally optimal detector much leads one to suspect that there might be a close connection between the LMP test statistic in (14) and a statistic which is locally optimal with respect to deflection. As we shall see next, when the two problems are formulated properly such a close connection does indeed exist.

The two hypotheses we shall consider first are the same as before in the composite N-P setting, namely the null H_0 in (6) and the composite alternative $\tilde{H}_1$ in (10) (wihout any prior on Θ). The expectation under $\tilde{H}_1$ of a test statistic $g(\mathbf{Z}(r))$ of data $\mathbf{Z}(r)$ is denoted $E_\Theta(g)$ and the corresponding deflection is

$$D_\Theta(g) = \frac{\big(E_\Theta(g)\big)^2}{V_0(g)}.$$

Then, under some (weak) regularity assumptions on g, see appendix B, the deflection $D_\Theta(g)$ can be expanded in a Taylor series near $\Theta = \mathbf{0}$ as

$$
\begin{aligned}
D_\Theta(g) = D_{\mathbf{0}}(g) &+ \theta_1\big(\frac{\partial}{\partial\theta_1}D_\Theta(g)\big)\big|_{\Theta=\mathbf{0}} + \theta_2\big(\frac{\partial}{\partial\theta_2}D_\Theta(g)\big)\big|_{\Theta=\mathbf{0}} \\
&+ \theta_1\theta_2\big(\frac{\partial^2}{\partial\theta_1\partial\theta_2}D_\Theta(g)\big)\big|_{\Theta=\mathbf{0}} \\
&+ \frac{\theta_1^2}{2}\big(\frac{\partial^2}{\partial\theta_1^2}D_\Theta(g)\big)\big|_{\Theta=\mathbf{0}} + \frac{\theta_2^2}{2}\big(\frac{\partial^2}{\partial\theta_2^2}D_\Theta(g)\big)\big|_{\Theta=\mathbf{0}} + \mathcal{O}(\|\Theta\|^3). \quad (22)
\end{aligned}
$$

Evaluation of the terms on the right hand side (appendix B) leads to

$$
\begin{aligned}
D_\Theta(g) &= \\
&\frac{1}{V_0(g)}\Big(\theta_1\big(\frac{\partial}{\partial\theta_1}E_\Theta(g)\big)\big|_{\Theta=\mathbf{0}} + \theta_2\big(\frac{\partial}{\partial\theta_2}E_\Theta(g)\big)\big|_{\Theta=\mathbf{0}}\Big)^2 + \mathcal{O}(\|\Theta\|^3) \\
&= \frac{1}{V_0(g)}\Big(\int_{\mathbb{R}^2} g(\mathbf{z})\big(\theta_1\frac{\partial}{\partial\theta_1}p_\Theta(\mathbf{z}) + \theta_2\frac{\partial}{\partial\theta_2}p_\Theta(\mathbf{z})\big)\big|_{\Theta=\mathbf{0}}\,d\mathbf{z}\Big)^2 + \mathcal{O}(\|\Theta\|^3). \quad (23)
\end{aligned}
$$

Also here we shall consider only deviations from $\Theta = \mathbf{0}$ of the form $\theta_2 = \eta\theta_1$, for some positive η, and we see then that maximizing $D_\Theta(g)$ near $\Theta = \mathbf{0}$ is the same as maximizing the first term on the right in (23). This is a straightforward L^2-optimization problem and a maximizer $\hat{g}$ in the class of slowly increasing functions with zero mean under H_0 is easily found using the Cauchy-Schwarz inequality as (appendix B)

$$
\hat{g}(\mathbf{z}) = \mathbf{S}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + c_2, \quad (24)
$$

for some constant c_2. In other words, a test statistic of data $\mathbf{Z}(r)$ which maximizes deflection locally is

$$
\mathbf{S}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{Z}(r) + \frac{\eta}{2}\mathbf{Z}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{Z}(r), \quad (25)
$$

which is the same statistic as in (14).

The close connection between the locally optimal detector with respect to deflection and the LMP detector persists if we consider the type of alternative hypotheses $\tilde{H}_1^Q$ as in (16). What is changed, basically, is that the probability density of the observations under $\tilde{H}_1^Q$ is instead given by $q_\Theta(\mathbf{z})$ defined in (17). This means in particular that the expectation of a statistic g under the alternative $\tilde{H}_1^Q$ reads

$$
\begin{aligned}
E_\Theta(g) &= \int_{\mathbb{R}^2} g(\mathbf{z})\,q_\Theta(\mathbf{z})\,d\mathbf{z} \\
&= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} g(\mathbf{z})\,p_\Theta(\mathbf{z}|\mathbf{S}(r)=\mathbf{s})\,dG(\mathbf{s})d\mathbf{z} \\
&= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} g(\mathbf{z})\,p_\Theta(\mathbf{z}|\mathbf{S}(r)=\mathbf{s})\,d\mathbf{z}dG(\mathbf{s}) \\
&= \int_{\mathbb{R}^2} E_\Theta^{(\mathbf{s})}(g)\,dG(\mathbf{s}),
\end{aligned}
$$

where the interchange in the orders of integration can be justified as before and

$$E_{\Theta}^{(\mathbf{s})}(g) = \int_{\mathbb{R}^2} g(\mathbf{z})\, p_{\Theta}(\mathbf{z}|\mathbf{S}(r) = \mathbf{s})\, d\mathbf{z}.$$

The last quantity is the conditional expectation of g under $\tilde{H}_1$ in (10) for a vector $\mathbf{S}(r)$ with the value $\mathbf{s}$. Hence, also here the performance is determined by a convex combination of the performances in the family of cases corresponding to the composite alternative hypothesis $\tilde{H}_1$ and we can again hope to obtain an optimal solution by mimicking previous developments. Indeed, using totally analogous arguments as in appendix B it is easy to show that the deflection also in the present case can be expanded near $\Theta = \mathbf{0}$ in a Taylor series of the form (22) and that $D_{\Theta}(g)$ now locally is given by an expression as in (23) but with $p_{\Theta}(\mathbf{z})$ replaced by $q_{\Theta}(\mathbf{z})$. The procedure outlined in appendix B can then be carried out once more and it is straightforward to see that one will obtain the same result as before but with $\mathbf{S}(r)$ now replaced by $E_G(\mathbf{S}(r))$. Consequently, a statistic $\tilde{g}(\mathbf{Z}(r))$ of data $\mathbf{Z}(r)$ that is locally optimal with respect to deflection in the present setting is given by

$$\tilde{g}(\mathbf{z}) = E_G(\mathbf{S}^T(r))\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z},$$

which is the statistic obtained from (21) when applied to data.

4 Discussion

Despite its widespread use, the optimality of the periodogram as a test statistic is seldom questioned. Indeed, most analyses of periodogram based detection schemes focus on the asymtotic distribution of the periodogram, and properties of detectors based on the periodogram, and do not address the more fundamental question of determining an optimal test statistic in the Fourier domain. A key issue therefore is to determine, under realistic modeling conditions, if there is a fundamental reason to use an inherently quadratic statistic and, if this is the case, to explain it. It is known [5, Thm. 4.6.3] that Fisher's test is the uniformly most powerful symmetric invariant test for detecting a single sinusoid signal in white Gaussian noise of unknown variance, if the frequency of the signal is at one of the DFT frequency bins. However, in practice the parameters of the noise background are often slowly varying and can be estimated 'online' so that the biggest uncertainty is in the parameters of the narrow-band signal (amplitude, phase etc) to be detected. Moreover, there is in general no reason to impose a symmetry restriction on the test statistic. Guided by these considerations we have here developed various optimal tests based on the strong intutitive appeal of a Fourier domain setting and the power of a model-based approach utilizing the structure of the RMPs. It has turned out that for several realistic formulations the resulting optimal test statistics in this framework are not quadratic but linear–quadratic and a few (additional) remarks about this fact are in order.

First, we note that for detection of an RMP with one frequency component of unknown (mean) amplitude $S(r)$ and dispersion $U(r)$ the locally optimal statistic in (14) is virtually never purely quadratic. A look at the Taylor expansion (11) shows that it becomes quadratic only in the degenerate case where one only considers (co-) parameter variations of the form $\theta_1 = \theta_2^2$ locally near $(0,0)$. Likewise, it becomes linear only when $\theta_2 = \theta_1^2$ near $(0,0)$. In the corresponding case of locally optimal detection with random $\mathbf{S}(r)$ it becomes quadratic precisely when $E_G(\mathbf{S}(r)) = 0$, which in practice occurs essentially only if $s(t)$ has uniformly distributed phase, i.e. no a priori knowledge of the phase. In many applications there can however indeed be a priori knowledge of the mean phase, such as efter the first frame in multi-frame detection where $\mathbf{S}(r)$ is constant between frames (but random). It may then be that there is no a priori knowledge about the phase of $s(t)$, that is, the components of $\mathbf{S}(r)$, at the beginning of the first frame but after the first frame the a posteriori distribution of $\mathbf{S}(r)$ does no longer correspond to a uniformly distributed phase. [7] The optimal detector must then change structure accordingly; from purely quadratic to linear–quadratic (a case of *semi-coherent* detection). In the limit of an infinite number of frames the detector becomes linear; it degenerates into a matched filter. One could also concieve cases where not only the real and imaginary parts of $S(r)$ are random but where $(S(r), U(r)$ are random with a joint distribution. However, our formulation offers reasonable power in modeling while retaining simplicity since the only 'tuning' parameter in the optimal test statistic (14) is the ratio η.

5 Conclusions

We have devised a descision theoretic framework for detection of narrow-band signals in noise that preserves the intuitive appeal of Fisher's original idea of using a Fourier-domain based statistic with a natural and more realistic model for the narrow-band signal. The resulting optimal tests have been shown to be linear–quadratic rather than quadratic for several important formulations of the detection problem, including locally most poweful formaulations as well as formulation based on second order statistics such as deflection. The two parts of the detector, the linear and the quadratic, can be interpreted as serving two different purposes; the linear part makes the detector responsive to the sinusoidal structure of the signal and the quadratic

[7] Also, note than when the phase has a distribution other than uniform it is only its mean that enters into the detector, a quantity that is easy to estimate/assign values to.

part 'robustifies' the detector to be better adapted to the inevitable variations of the signal around the ideal sinusoidal shape. (This section needs to be laid out somewhat better.)

Appendix

A LMP Test

We shall here provide some details that were omitted in the development of the LMP test above. To begin with we show that the probability of detection $P_\Theta(f)$ can indeed be Taylor expanded to any degree around $\Theta = \mathbf{0}$.

Lemma A.1. *Let* $\mathbf{A}, \mathbf{B}$ *be two square matrices and* δ_0 *a real number such that* $(\delta_0\mathbf{A}+\mathbf{B})^{-1}$ *exists. Then* $-(\delta_0\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}(\delta_0\mathbf{A}+\mathbf{B})^{-1}$ *is the (Frechét) differential of* $(\delta\mathbf{A}+\mathbf{B})^{-1}$ *with respect to* δ *at* $\delta = \delta_0$.

Proof. First we note that $\big((\delta_0+\varepsilon)\mathbf{A}+\mathbf{B}\big)^{-1}$ is clearly continuous in ε near $\varepsilon = 0$ and by direct multiplication it is easy to verify that

$$\big((\delta_0+\varepsilon)\mathbf{A}+\mathbf{B}\big)^{-1} - (\delta_0\mathbf{A}+\mathbf{B})^{-1} = -\varepsilon\big((\delta_0+\varepsilon)\mathbf{A}+\mathbf{B}\big)^{-1}\mathbf{A}(\delta_0\mathbf{A}+\mathbf{B})^{-1},$$

from which we obtain the growth estimate

$$\|\big((\delta_0+\varepsilon)\mathbf{A}+\mathbf{B}\big)^{-1} - (\delta_0\mathbf{A}+\mathbf{B})^{-1}\| \leq \mathcal{O}(|\varepsilon|).$$

Thus,

$$\begin{aligned}
&\|\big((\delta_0+\varepsilon)\mathbf{A}+\mathbf{B}\big)^{-1} - (\delta_0\mathbf{A}+\mathbf{B})^{-1} + \varepsilon(\delta_0\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}(\delta_0\mathbf{A}+\mathbf{B})^{-1}\| \\
&= |\varepsilon|\|\big((\delta_0+\varepsilon)\mathbf{A}+\mathbf{B}\big)^{-1}\mathbf{A}(\delta_0\mathbf{A}+\mathbf{B})^{-1} - (\delta_0\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}(\delta_0\mathbf{A}+\mathbf{B})^{-1}\| \\
&\leq |\varepsilon|\|\big((\delta_0+\varepsilon)\mathbf{A}+\mathbf{B}\big)^{-1} - (\delta_0\mathbf{A}+\mathbf{B})^{-1}\|\|\mathbf{A}\|\|(\delta_0\mathbf{A}+\mathbf{B})^{-1}\| \\
&= o(|\varepsilon|).
\end{aligned}$$

□

If we put

$$\mathcal{E}(\Theta,\mathbf{z}) = \exp\big(-\frac{1}{2}(\mathbf{z}-\theta_1\mathbf{S}(r))^T(\theta_2\mathbf{\Sigma}_U(r)+\mathbf{\Sigma}_N(r))^{-1}(\mathbf{z}-\theta_1\mathbf{S}(r))\big)$$

then straightforward calculation gives

$$\frac{\partial}{\partial\theta_1}\mathcal{E}(\Theta,\mathbf{z}) = \big(\mathbf{S}^T(r)(\theta_2\mathbf{\Sigma}_U(r)+\mathbf{\Sigma}_N(r))^{-1}(\mathbf{z}-\theta_1\mathbf{S}(r))\big)\mathcal{E}(\Theta,\mathbf{z}) \tag{26}$$

and by using the lemma above likewise

$$\begin{aligned}
\frac{\partial}{\partial\theta_2}\mathcal{E}(\Theta,\mathbf{z}) = \frac{1}{2}(\mathbf{z}-\theta_1\mathbf{S}(r))^T(\theta_2\mathbf{\Sigma}_U(r)+\mathbf{\Sigma}_N(r))^{-1}\mathbf{\Sigma}_U(r)\,\cdot \\
\cdot\,(\theta_2\mathbf{\Sigma}_U(r)+\mathbf{\Sigma}_N(r))^{-1}(\mathbf{z}-\theta_1\mathbf{S}(r))\mathcal{E}(\Theta,\mathbf{z}).
\end{aligned} \tag{27}$$

This means that

$$\frac{\partial}{\partial \theta_j}\mathcal{E}(\Theta,\mathbf{z}) = h_j(\Theta,\mathbf{z})\mathcal{E}(\Theta,\mathbf{z}), \quad j=1,2,$$

where $h_j(\Theta,\mathbf{z})$ is a polynomial in $\mathbf{z}$ with C^∞-coefficients in Θ. Similarly, by induction it is easy to see that any higher order partial derivative $\partial_\Theta^m \mathcal{E}(\Theta,\mathbf{z})$ of $\mathcal{E}(\Theta,\mathbf{z})$ with respect to components of Θ, where m is a multi-index [21, p. 266], is of the form

$$\partial_\Theta^m \mathcal{E}(\Theta,\mathbf{z}) = h_m(\Theta,\mathbf{z})\mathcal{E}(\Theta,\mathbf{z}),$$

where $h_m(\Theta,\mathbf{z})$ is a polynomial in $\mathbf{z}$ with C^∞-coefficients in Θ. Therefore, any higher order order derivative $\partial_\Theta^m p_\Theta(\mathbf{z})$ of the density

$$p_\Theta(\mathbf{z}) = c(\theta_1)\mathcal{E}(\Theta,\mathbf{z}),$$

where

$$c(\theta_2) = \frac{1}{2\pi \det(\theta_2 \mathbf{\Sigma}_U(r) + \mathbf{\Sigma}_N(r))^{1/2}},$$

is of the form

$$\partial_\Theta^m p_\Theta(\mathbf{z}) = \tilde{h}_m(\Theta,\mathbf{z})\mathcal{E}(\Theta,\mathbf{z})$$

where $\tilde{h}_m(\Theta,\mathbf{z})$ is a polynomial in $\mathbf{z}$ with C^∞-coefficients in Θ. By the exponential decay of $\mathcal{E}(\Theta,\mathbf{z})$ as $\|\mathbf{z}\| \to \infty$ it follows that for any sufficiently small neighborhood $\mathcal{N}$ of $\mathbf{0}$ and fixed positive integer k the majorant

$$\sup_{|m|\le k, \Theta\in\mathcal{N}} |\partial_\Theta^{\mathbf{m}} p_\Theta(\mathbf{z})|$$

is an integrable function. Using this fact together with a standard result from integration theory ([21, p. 54]) we can justify differentiation directly on the integrand in the representation

$$P_\Theta(\mathcal{A}) = \int_{\mathcal{A}} p_\Theta(\mathbf{z})\, d\mathbf{z},$$

where $\mathcal{A} = \{\mathbf{z} \in \mathbb{R}^2 : f(\mathbf{z}) > \gamma\}$ is the decision region as before, and obtain the terms in the Taylor expansion (11) as well as the representation

$$\frac{\partial}{\partial \theta_1} P_\Theta(\mathcal{A})\big|_{\Theta=\mathbf{0}} + \eta \frac{\partial}{\partial \theta_2} P_\Theta(\mathcal{A})\big|_{\Theta=\mathbf{0}} = \int_{\mathcal{A}} \big(\frac{\partial}{\partial \theta_1} p_\Theta(\mathbf{z}) + \eta \frac{\partial}{\partial \theta_2} p_\Theta(\mathbf{z})\big)\big|_{\Theta=\mathbf{0}}\, d\mathbf{z}. \tag{28}$$

The integrand on the right in (28) can be expressed as

$$\begin{aligned} &\big(\frac{\partial}{\partial \theta_1} p_\Theta(\mathbf{z}) + \eta \frac{\partial}{\partial \theta_2} p_\Theta(\mathbf{z})\big)\big|_{\Theta=\mathbf{0}} = \\ &\quad \big(\eta \frac{\partial}{\partial \theta_2} c(\theta_2)\big)\big|_{\Theta=\mathbf{0}}\, \mathcal{E}(\mathbf{0},\mathbf{z}) + c(0)\big(\frac{\partial}{\partial \theta_1}\mathcal{E}(\Theta,\mathbf{z}) + \eta \frac{\partial}{\partial \theta_2}\mathcal{E}(\Theta,\mathbf{z})\big)\big|_{\Theta=\mathbf{0}}, \end{aligned} \tag{29}$$

where we note that $p_{\mathbf{0}}(\mathbf{z}) = c(0)\mathcal{E}(\mathbf{0},\mathbf{z})$. This taken together with (26),(27) and (28) yields the expansion in (12).

B Locally Optimal Deflection

The class of admissible test statistics we consider consist of functions $g(\mathbf{z})$ that are locally (Lebesgue) integrable and of 'slow', i.e. at most polynomial, growth as $\|\mathbf{z}\| \to \infty$. For such functions it is easy, using arguments from appendix A, to justify the interchange in differentation and integration in the definition

$$E_\Theta(g) = \int_{\mathbb{R}^2} g(\mathbf{z}) p_\Theta(\mathbf{z})\, d\mathbf{z}. \tag{30}$$

Hence, the Taylor series in (22) is well-defined for test statistics in this class. Further, by the assumptions $E_{\mathbf{0}}(g) = 0$ and if we use this fact together with the definition (30) and its differentiability properties it is easy to see that the first three terms on the right hand side of (22) are zero. The middle three terms can be written

$$\begin{aligned}
\theta_1\theta_2\Big(\frac{\partial^2}{\partial\theta_1\partial\theta_2} D_\Theta(g)\Big)\Big|_{\Theta=\mathbf{0}} &= \frac{2\theta_1\theta_2}{V_0(g)}\Big(\frac{\partial}{\partial\theta_1}E_\Theta(g)\frac{\partial}{\partial\theta_2}E_\Theta(g)\Big)\Big|_{\Theta=\mathbf{0}},\\
\frac{\theta_1^2}{2}\Big(\frac{\partial^2}{\partial\theta_1^2} D_\Theta(g)\Big)\Big|_{\Theta=\mathbf{0}} &= \frac{\theta_1^2}{V_0(g)}\Big(\frac{\partial}{\partial\theta_1}E_\Theta(g)\Big)^2\Big|_{\Theta=\mathbf{0}},\\
\frac{\theta_2^2}{2}\Big(\frac{\partial^2}{\partial\theta_2^2} D_\Theta(g)\Big)\Big|_{\Theta=\mathbf{0}} &= \frac{\theta_2^2}{V_0(g)}\Big(\frac{\partial}{\partial\theta_2}E_\Theta(g)\Big)^2\Big|_{\Theta=\mathbf{0}},
\end{aligned}$$

and it follows that the Taylor series of $D_\Theta(g)$ near $\Theta = \mathbf{0}$ can be expressed as in (23). If we only consider local deviations of the form $\theta_2 = \eta\theta_1$, $\eta > 0$, we see that maximizing $D_\Theta(g)$ locally is the same as maximizing the first term on the right hand side of (23) with respect to slowly increasing functions g such that $E_0(g) = 0$, i.e. maximizing the quantity

$$\begin{aligned}
&\frac{1}{V_0(g)}\Big(\int_{\mathbb{R}^2} g(\mathbf{z})\Big(\frac{\partial}{\partial\theta_1}p_\Theta(\mathbf{z}) + \eta\frac{\partial}{\partial\theta_2}p_\Theta(\mathbf{z})\Big)\Big|_{\Theta=\mathbf{0}}\, d\mathbf{z}\Big)^2 = \\
&\frac{1}{V_0(g)}\Big(\int_{\mathbb{R}^2}\big(\mathbf{S}^T(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\boldsymbol{\Sigma}_N^{-1}(r)\boldsymbol{\Sigma}_U(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z}\big)g(\mathbf{z})p_0(\mathbf{z})\, d\mathbf{z}\Big)^2,
\end{aligned} \tag{31}$$

where we have used (26),(27) and (29) together with the definition (30) and its properties. The assumption $E_0(g) = 0$ moreover implies that

$$\begin{aligned}
&\Big(\int_{\mathbb{R}^2}\big(\mathbf{S}^T(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\boldsymbol{\Sigma}_N^{-1}(r)\boldsymbol{\Sigma}_U(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z}\big)g(\mathbf{z})p_0(\mathbf{z})\, d\mathbf{z}\Big)^2 = \\
&\Big(\int_{\mathbb{R}^2}\big(\mathbf{S}^T(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\boldsymbol{\Sigma}_N^{-1}(r)\boldsymbol{\Sigma}_U(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z} - c_2\big)g(\mathbf{z})p_0(\mathbf{z})\, d\mathbf{z}\Big)^2
\end{aligned}$$

for any constant c_2, in particular for the choice

$$c_2 = \int_{\mathbb{R}^2}\big(\mathbf{S}^T(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\boldsymbol{\Sigma}_N^{-1}(r)\boldsymbol{\Sigma}_U(r)\boldsymbol{\Sigma}_N^{-1}(r)\mathbf{z}\big)p_0(\mathbf{z})\, d\mathbf{z}. \tag{32}$$

Now, by the Cauchy-Schwarz inequality the quantity in (31) is bounded by

$$\int_{\mathbb{R}^2} \left(\mathbf{S}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} - c_2\right)^2 p_0(\mathbf{z})\, d\mathbf{z}$$

with equality if

$$\hat{g}(\mathbf{z}) = \mathbf{S}^T(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} + \frac{\eta}{2}\mathbf{z}^T\mathbf{\Sigma}_N^{-1}(r)\mathbf{\Sigma}_U(r)\mathbf{\Sigma}_N^{-1}(r)\mathbf{z} - c_2.$$

However, this choice of g is clearly admissible since by (32) we have $E_0(\hat{g}) = 0$. This renders the statistic (25) as a locally optimal statistic on data $\mathbf{Z}(r)$.

References

[1] R.A. Fisher, "Tests of Significance in Harmonic Analysis," *Proc. Roy. Soc. London Ser. A,* Vol. 125, pp. 54–59, 1929.

[2] M.B. Priestley, *Spectral Analysis and Time Series,* Academic Press, London, 1981.

[3] U. Grenander and M. Rosenblatt, *Statistical Analysis of Sationary Time Series,* Almqvist & Wiksell, Uppsala, 1956.

[4] E.J. Hannan, *Multiple Time Series,* Wiley, New York, 1970.

[5] T.W. Anderson, *The Statistical Analysis of Time Series,* Wiley, New York, 1971.

[6] A.F. Siegel, "Testing for Periodicity in a Time Series," *J. Am. Stat. Assoc.,* Vol. 75, No. 370, pp. 345–348, June 1980.

[7] E. Bølviken, "New Tests of Significance in Periodogram Analysis," *Scand. J. Stat.,* Vol. 10, pp. 1–9, 1983.

[8] E. Bølviken, "The Distribution on Certain Rational Functions of Order Statistics From Exponential Distributions," *Scand. J. Stat.,* Vol. 10, pp. 117–123, 1983.

[9] L.G. Tatum, "Control Charts for the Detection of a Periodic Component," *Technometrics,* Vol. 38, No. 2, pp. 152–160, May 1996.

[10] P. Whittle, *Hypothesis Testing in Time Series Analysis,* Thesis, Uppsala University, Amqvist & Wiksell, Uppsala, Hafner, New York, 1951.

[11] A.M. Walker, "Some Asumptotic Results for the Periodogram of a Stationary Time Series," J. Aust. Math. Soc., Vol. 5, pp. 107–128, 1965.

[12] D.C. Schleher, "MTI and Pulsed Doppler Radar," *Artech House,* Norwood, MA, 1991.

[13] E.A. Robinson, "A Historical Perspective of Spectrum Estimation," *Proc. IEEE,* Vol. 70, No. 9, pp. 885–907, Sept. 1982.

[14] M.J. Hinich, "A Statistical Theory of Signal Coherence," in review, *IEEE Trans. Signal Processing*

[15] W.A. Gardner and L.E. Franks, "Characterization of Cyclostationary Random Signal Processes," *IEEE Trans. Inform. Theory,* Vol. IT-21, pp. 4–14, 1975.

[16] J.I. Marcum and P. Swerling, "A Statistical Theory of Target Detection by Pulsed Radar," *IRE Trans. Inform. Theory,* Vol. 6, pp. 59–145, April 1960.

[17] P. Billingsley, *Probability and Measure,* 2:nd ed., Wiley, New York, 1986.

[18] H.V. Poor, "An Introduction to Signal Detection and Estimation," Springer, New York, 1988.

[19] B. Picinbono and P. Duvaut, "Optimal Linear-Quadratic Systems for Detection and Estimation," *IEEE Trans. Inform. Theory.,* Vol. 34, No. 2, pp. 304–311, March 1988.

[20] W.A. Gardner, "A Unifying View of Second-Order Measures of Quality for Signal Classification," *IEEE Trans. Comm.,* Vol. 28, No. 6, pp. 807–816, June 1980.

[21] G.B. Folland, *Real Analysis,* Wiley, New York, 1984.