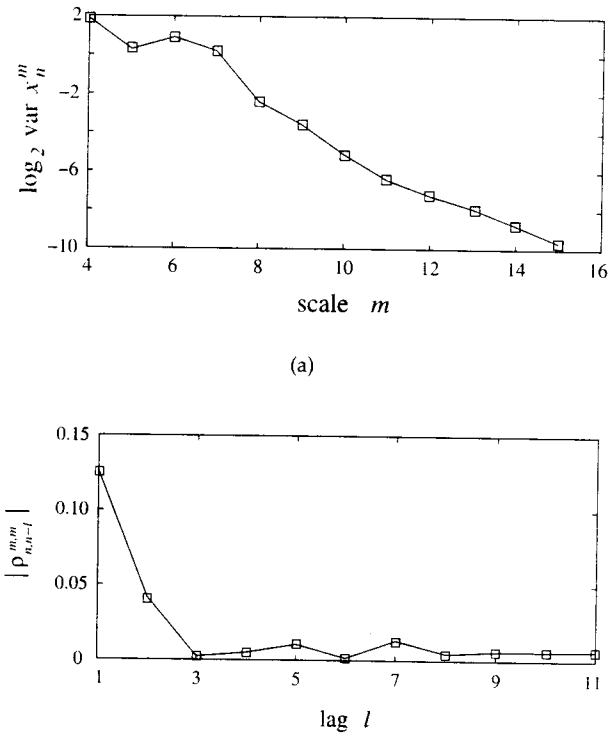


Figure 3.8. Wavelet-based analysis of the heartbeat interarrival times for a healthy patient. The time-series is analyzed using a 5th-order Daubechies wavelet basis. (a) Scale-to-scale wavelet coefficient sample-variance progression. (b) Average magnitude of the normalized along-scale sample-correlation between wavelet coefficients.

4

Detection and Estimation with Fractal Processes

4.1 INTRODUCTION

Given the ubiquity of physical signals exhibiting $1/f$-type behavior, there are many applications in which the need arises for efficient algorithms for processing such signals. For instance, one is frequently interested in problems of detection and classification; characterization and parametrization; prediction and interpolation; and separation of $1/f$ signals both from one another as well as from other types of known or partially known signals.

In some cases, the $1/f$ signal itself is of primary interest. An example would be the problem of modeling stock market data such as the Dow Jones Industrial Average as a $1/f$ process. In other cases, the $1/f$ signal represents a noise process obscuring some other signal of interest. This is more likely to be the case in optical and electronic systems, for example, where $1/f$ noise is a predominant form of background noise.

Even when the $1/f$ signal is of primary interest, one rarely has perfect access to such signals. Typically, our observations are incomplete. Indeed, they will generally be time-limited and resolution-limited. More generally, the observations may contain gaps, or there may be multiple observations. Additionally, any observations of $1/f$ signals will invariably be corrupted by some degree of broadband measurement noise.[1] It is important to both recognize and accommodate such measurement noise in any algorithms for

[1] Actually, the coexistence of $1/f$ and white noises in electronic and optical systems is well documented. In electronic systems, for instance, the predominant noise is $1/f$ noise at

processing such data. Indeed, because $1/f$ signals have a spectral density that vanishes for sufficiently high frequencies, there necessarily exists some frequency threshold beyond which the broadband noise is predominant. As a consequence, such noise can strongly affect the performance of signal processing algorithms.

In this chapter, we develop some optimal algorithms for addressing a number of basic signal processing problems involving detection and estimation with $1/f$-type signals. Our basic approach is to exploit the properties of wavelet-based representations of $1/f$-type processes. In particular, based upon the synthesis result of Theorem 3.4 and supported by the subsequent analysis results, our model for $1/f$ signals is signals that, when expanded into an orthonormal wavelet basis, yield coefficients that are effectively uncorrelated and obey the appropriate variance progression. That is, we exploit the role of the wavelet expansion as a Karhunen-Loève-like expansion for $1/f$-type processes. Because extremely efficient algorithms exist for computing orthonormal wavelet transformations as discussed in Section 2.3.3, this approach is not only analytically convenient for solving these signal processing problems, but leads to computationally highly efficient structures for implementing the resulting algorithms.

We routinely incorporate additive stationary white measurement noise to ensure the robustness of the algorithms we develop. Furthermore, most of these algorithms are designed specifically for the case of Gaussian $1/f$ processes and Gaussian measurement noises. While this requirement is principally motivated by tractability requirements, there are, in fact, many settings in which this assumption is physically quite reasonable. Furthermore, several of the algorithms we develop retain many of their important properties in the more general non-Gaussian case.

An important component of the algorithm development process is performance analysis, and this can take many forms. Accompanying each of the algorithms we develop is a set of basic though necessarily limited evaluations of its performance in various settings. These analyses both establish the essential viability of the algorithms and reveal some of their salient properties. Many of these performance studies involve Monte Carlo simulations with synthetic data. For these simulations, we generate $1/f$ processes using the Corsini-Saletti implementation of Keshner's synthesis [69]. Because this synthesis is fundamentally different from a wavelet-based synthesis, such simulations play an important role in verifying the robustness of the wavelet-based algorithms with respect to our particular model for $1/f$-type behavior. However, as a consequence, these simulations generally do not enable us to isolate the effects of modeling error alone.

frequencies below about 1 kHz, while at higher frequencies, it is white noise in the form of thermal (i.e., Johnson) and shot noise [75].

The wavelet-based implementation of each of the algorithms we develop requires that we select a suitable wavelet basis from among the large collection of candidates. However, given the apparent insensitivity of the wavelet-based model for $1/f$-type behavior to the choice of basis, for the simulations we choose somewhat arbitrarily to use the basis corresponding to Daubechies' 5th-order finite-extent maximally regular wavelet for which the corresponding conjugate quadrature filters have 10 non-zero coefficients. We remark that in addition to being realizable, this basis satisfies the conditions of the theorems of Section 3.3.2 concerning the synthesis and analysis of $1/f$-type behavior using wavelets. Specifically, the basis has more than enough vanishing moments to accommodate spectral parameters in our principal range of interest, $0 < \gamma < 2$.

Before beginning, we note that a basic understanding of the fundamentals of estimation and detection theory is assumed of the reader in this chapter. A sufficiently comprehensive treatment for our purposes can be found in, e.g., Van Trees [76].

4.2 $1/f$ SYNTHESIS AND WHITENING FILTERS

Many of the results on detection and estimation we derive in this chapter are conveniently interpreted in a canonical form through the concept of a reversible (or invertible) whitening filter for $1/f$ processes. In this section, we derive such whitening filters and their inverses for the particular wavelet-based model for $1/f$-type behavior which we intend to exploit in this chapter.

To begin, if $x(t)$ is a $1/f$ signal corresponding to some spectral exponent γ, we model the corresponding wavelet coefficients x_n^m as zero-mean random variables having negligible correlation and obeying a variance progression of the form

$$\operatorname{var} x_n^m = \sigma^2 \beta^{-m}$$

where, for notational convenience, we define

$$\beta = 2^\gamma. \tag{4.1}$$

In turn, we may express the x_n^m as

$$x_n^m = \sigma \beta^{-m/2} v_n^m$$

where the v_n^m are then zero-mean, unit-variance, uncorrelated random variables. Hence, the process $v(t)$ defined according to

$$v(t) = \sum_m \sum_n v_n^m \psi_n^m(t)$$

is a wide-sense stationary white noise process since the $\psi_n^m(t)$ constitute a complete orthonormal set. This suggests that we may model $x(t)$ as the output of a linear system driven by stationary white noise $v(t)$. In particular,

the system performs an orthonormal wavelet transform on the input $v(t)$, scales each of the resulting coefficients v_n^m by a factor

$$k_n^m = \sigma\beta^{-m/2},$$

then inverse wavelet transforms the resulting x_n^m to generate the output process $x(t)$. This $1/f$ *synthesis filter*, defined via

$$x(t) = \mathcal{W}_d^{-1}\left\{\sigma\beta^{-m/2}\mathcal{W}_d\left\{v(t)\right\}\right\}, \tag{4.2}$$

is a linear filter whose kernel[2] is

$$\kappa_s^\circ(t,\tau) = \sum_m\sum_n \psi_n^m(t)\sigma\beta^{-m/2}\psi_n^m(\tau). \tag{4.3a}$$

We emphasize that viewing $x(t)$ as the output of a linear system with kernel (4.3a) driven by stationary white noise $v(t)$ is especially useful in the Gaussian scenario, in which case $v(t)$ is a stationary white Gaussian process. Nevertheless, for non-Gaussian processes this characterization remains useful at least insofar as modeling the second-order properties of $x(t)$ is concerned.

From the wavelet-based characterization of the synthesis filter (4.2) we readily deduce that this filter is invertible, and that its inverse has kernel

$$\kappa_w^\circ(t,\tau) = \sum_m\sum_n \psi_n^m(t)\frac{1}{\sigma\beta^{-m/2}}\psi_n^m(\tau). \tag{4.3b}$$

This is, therefore, the corresponding *whitening filter* for our model of $1/f$-type behavior. Indeed, when this filter is driven by a process obtained as the output of our $1/f$ synthesis filter, the output is, evidently, a wide-sense stationary white process. When driven by an exactly-$1/f$ process, the properties of the output are readily described in terms of the analysis results of Section 3.3.2.

As discussed at the outset of the chapter, any $1/f$-type process we consider are invariably accompanied by an additive stationary white observation noise component. Consequently, we frequently find the notion of synthesis and whitening filters for the combined $1/f$-plus-white processes convenient in interpreting our algorithms. These filters are, of course, closely related to the filters derived above. In fact, it is straightforward to establish that synthesis and whitening filters for $1/f$-plus-white processes are characterized by the respective kernels

$$\kappa_s(t,\tau) = \sum_m\sum_n \psi_n^m(t)\sigma_m\psi_n^m(\tau) \tag{4.4a}$$

[2]In our notation, the kernel $k(t,\tau)$ of a linear system defines the response of the system at time t to a unit impulse at time τ. Consequently the response of the system to a suitable input $x(t)$ is expressed as

$$y(t) = \int_{-\infty}^{\infty} x(\tau)\,k(t,\tau)\,d\tau.$$

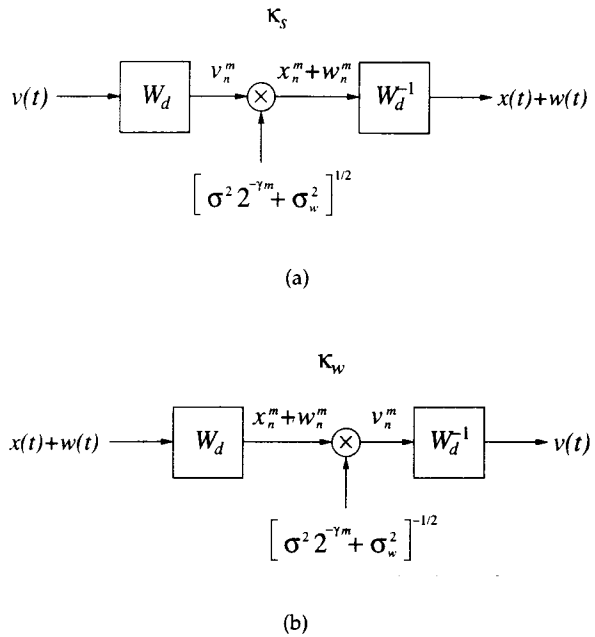


Figure 4.1. Canonical form realizations of synthesis and whitening filters for processes that are the superposition of $1/f$-type and white components, i.e., $1/f$-plus-white processes. (a) Synthesis filter, kernel $\kappa_s(t,\tau)$. (b) Whitening filter, kernel $\kappa_w(t,\tau)$.

$$\kappa_w(t,\tau) = \sum_m\sum_n \psi_n^m(t)\frac{1}{\sigma_m}\psi_n^m(\tau) \tag{4.4b}$$

where $\sigma_m > 0$ is defined by

$$\sigma_m^2 = \sigma^2\beta^{-m} + \sigma_w^2 \tag{4.5}$$

and σ_w^2 is the spectral density of the white noise component. The canonical wavelet-based realization of these filters is depicted in Fig. 4.1.

4.3 PARAMETER ESTIMATION FOR $1/f$ SIGNALS

In this section, we consider the problem of estimating the parameters of a Gaussian $1/f$ signal from observations corrupted by stationary white Gaussian noise [77]. Since we typically lack *a priori* knowledge of the spectral

density of the noise, we consider more specifically the problem of *jointly* estimating signal and noise parameters for this scenario.

Such parameter estimates, in addition to providing a solution to the associated $1/f$ spectrum estimation problem, are frequently of interest in their own right. Indeed, from the parameter estimates we can directly compute the fractal dimension of the underlying signal using the relationships developed in Chapter 3. Robust estimation of the fractal dimension of $1/f$ processes is important in a number of applications requiring signal detection and classification. For example, in image processing, where 2-D extensions of $1/f$ processes are used to model natural terrain and other patterns and textures [47] [58], fractal dimension can be useful in distinguishing among various man-made and natural objects.

While several approaches to the fractal dimension estimation problem can be found in the literature (see [58], [78], [79], and the references therein), a traditional problem with these approaches has been their inability to adequately handle the presence of broadband noise in the observation data. In fact, the quality of the estimates generally deteriorates dramatically in the presence of such noise even at high SNR [58]. Since noise is inherently present in any real data, this lack of robustness has limited the usefulness of these algorithms. In this section we describe fractal dimension estimators for Gaussian $1/f$ processes that explicitly take into account the presence of additive white Gaussian observation noise. The resulting iterative estimation algorithms are computationally efficient, robust, and statistically consistent.

Our basic approach is to apply the method of Maximum Likelihood (ML) estimation, exploiting the wavelet-based characterization of our $1/f$ model. While we specifically consider the case of Gaussian $1/f$ processes corrupted by additive stationary white Gaussian measurement noise in our formulation of the problem, we stress that the resulting estimators are, in fact, applicable to a broader class of non-Gaussian $1/f$ processes and measurement noise models, and retain many desirable properties.

We formulate our problem as follows. Suppose we have observations $r(t)$ of a zero-mean Gaussian $1/f$ process $x(t)$ embedded in zero-mean additive stationary white Gaussian noise $w(t)$ that is statistically independent of $x(t)$, so

$$r(t) = x(t) + w(t), \qquad -\infty < t < \infty. \tag{4.6}$$

From this continuous-time data, we assume we have extracted a number of wavelet coefficients r_n^m. In theory, we may assume these coefficients are obtained by projecting the wavelet basis functions onto the observed data, i.e., via

$$r_n^m = \int_{-\infty}^{\infty} \psi_n^m(t)\, r(t)\, dt.$$

However, in practice, these coefficients are more typically obtained by applying the computationally efficient DWT to the samples of a segment of data

that is both time-limited and resolution-limited, as described in Section 2.3.3. Let us assume that the finite set of available distinct scales, $\mathcal{M}$, is, in increasing order,

$$\mathcal{M} = \{m_1, m_2, \ldots, m_M\}, \tag{4.7a}$$

and that at each scale m the set of available coefficients $\mathcal{N}(m)$ is[3]

$$\mathcal{N}(m) = \{n_1(m), n_2(m), \ldots, n_{N(m)}(m)\}. \tag{4.7b}$$

Hence, the data available to the estimation algorithm are

$$\mathbf{r} = \{r_n^m \in \mathcal{R}\} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\}. \tag{4.8}$$

We remark before proceeding that, based on the discussion in Section 2.3.4, for an implementation via the DWT with $N = N_0 2^M$ samples of observed data, we have, typically,

$$\mathcal{M} = \{1, 2, \ldots, M\} \tag{4.9a}$$

$$\mathcal{N}(m) = \{1, 2, \ldots, N_0 2^{m-1}\}, \tag{4.9b}$$

where N_0 is a constant that depends on the length of the filter $h[n]$. While many of the results we derive are applicable to the more general scenario, we frequently specialize our results to this case.

Exploiting the Karhunen-Loève-like properties of the wavelet decomposition for $1/f$-type processes, and using the fact that the w_n^m are independent of the x_n^m and are decorrelated for any wavelet basis, the resulting observation coefficients

$$r_n^m = x_n^m + w_n^m$$

can be modeled as mutually independent zero-mean, Gaussian random variables with variance

$$\operatorname{var} r_n^m = \sigma_m^2 = \sigma^2 \beta^{-m} + \sigma_w^2$$

where β is defined in terms of the spectral exponent γ of the $1/f$ process according to (4.1). Hence, it is the parameter set

$$\Theta = (\beta, \sigma^2, \sigma_w^2)$$

we wish to estimate. As discussed at the outset, it is often the case that only β or some function of β such as the spectral exponent γ, the fractal dimension D, or the self-similarity parameter H, is of interest. Nevertheless, σ^2 and σ_w^2 still need to be estimated simultaneously as they are rarely known *a priori*. Furthermore, ML estimates of γ, D, H are readily derived from the ML estimate $\hat{\beta}_{\mathrm{ML}}$. Indeed, since each of these parameters is related to β through an invertible transformation, we have

$$\hat{\gamma}_{\mathrm{ML}} = \log_2 \hat{\beta}_{\mathrm{ML}} \tag{4.10a}$$

$$\hat{D}_{\mathrm{ML}} = (5 - \hat{\gamma}_{\mathrm{ML}})/2 \tag{4.10b}$$

$$\hat{H}_{\mathrm{ML}} = (\hat{\gamma}_{\mathrm{ML}} - 1)/2. \tag{4.10c}$$

[3]Note that without loss of generality we may assume $\mathcal{N}(m) \neq \emptyset$ for any m, or else the corresponding scale m could be deleted from $\mathcal{M}$.

Proceeding, we express the likelihood as a function of the parameters by

$$\mathcal{L}(\Theta) = p_{\mathbf{r}}(\mathbf{r};\Theta) = \prod_{m,n\in\mathcal{R}} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\sigma_m^2}\right]$$

for which the log-likelihood function is

$$L(\Theta) = \ln p_{\mathbf{r}}(\mathbf{r};\Theta) = -\frac{1}{2}\sum_{m,n\in\mathcal{R}}\left\{\frac{1}{\sigma_m^2}(r_n^m)^2 + \ln(2\pi\sigma_m^2)\right\}.$$

Equivalently,

$$L(\Theta) = -\frac{1}{2}\sum_{m\in\mathcal{M}} N(m)\left\{\frac{\hat{\sigma}_m^2}{\sigma_m^2} + \ln(2\pi\sigma_m^2)\right\} \tag{4.11}$$

where the M sample-variances

$$\hat{\sigma}_m^2 = \frac{1}{N(m)}\sum_{n\in\mathcal{N}(m)} (r_n^m)^2 \tag{4.12}$$

summarize the aspects of the data required in the estimation. It is straightforward to show that the likelihood function in this case is well behaved and bounded from above on

$$\beta \geq 0,\ \sigma^2 \geq 0,\ \sigma_w^2 \geq 0$$

so that, indeed, maximizing the likelihood function is reasonable.

While we assume that $\beta, \sigma^2, \sigma_w^2$ are all unknown, it will be appropriate during the development to also specialize results to the case in which σ_w^2 is known. Still more specific results will be described when $\sigma_w^2 = 0$, corresponding to the case of noise-free observations. We may also assume, where necessary, that all $m \in \mathcal{M}$ are positive without loss of generality. Indeed, if, for example, $m_1 < 0$, then we could define new parameters through the invertible transformation

$$\begin{aligned}\tilde{\sigma}_w^2 &= \sigma_w^2\\ \tilde{\sigma}^2 &= \sigma^2\beta^{m_1-1}\\ \tilde{\beta} &= \beta\end{aligned}$$

for which the observations correspond to positive scales

$$\tilde{\mathcal{M}} = \{1, m_2 - m_1 + 1, \ldots, m_M - m_1 + 1\}$$

and which lead to the same ML estimates for $\beta, \sigma^2, \sigma_w^2$.

4.3.1 Case I: β, σ^2, σ_w^2 Unknown

Differentiating $L(\Theta)$ with respect to σ_w^2, σ^2, and β, respectively, it follows that the stationary points of $L(\Theta)$ are given as the solutions to the equations

$$\begin{aligned}\sum_{m\in\mathcal{M}} T_m &= 0\\ \sum_{m\in\mathcal{M}} \beta^{-m} T_m &= 0\\ \sum_{m\in\mathcal{M}} m\beta^{-m} T_m &= 0\end{aligned}$$

where

$$T_m \triangleq \frac{N(m)}{\sigma_m^2}\left[1 - \frac{\hat{\sigma}_m^2}{\sigma_m^2}\right].$$

However, these equations are difficult to solve, except in special cases. Consequently, we utilize an iterative estimate-maximize (EM) algorithm [80].

A detailed development of the EM algorithm for our problem is given in Appendix C. The essential steps of the algorithm are summarized below, where we denote the estimates of the parameters β, σ^2, and σ_w^2 generated on the lth iteration by $\hat{\beta}^{[l]}$, $\hat{\sigma}^{2[l]}$, and $\hat{\sigma}_w^{2[l]}$, respectively.

E step: As shown in Appendix C, this step reduces to estimating the noise and signal portions of the wavelet coefficient variances at each scale $m \in \mathcal{M}$ using current estimates of the parameters $\hat{\beta}^{[l]}$, $\hat{\sigma}^{2[l]}$, and $\hat{\sigma}_w^{2[l]}$:

$$S_m^w(\hat{\Theta}^{[l]}) = A_m(\hat{\Theta}^{[l]}) + B_m^w(\hat{\Theta}^{[l]})\hat{\sigma}_m^2 \tag{4.13a}$$

$$S_m^x(\hat{\Theta}^{[l]}) = A_m(\hat{\Theta}^{[l]}) + B_m^x(\hat{\Theta}^{[l]})\hat{\sigma}_m^2 \tag{4.13b}$$

where

$$A_m(\hat{\Theta}^{[l]}) = \frac{\hat{\sigma}_w^{2[l]}\cdot\hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}}{\hat{\sigma}_w^{2[l]} + \hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}} \tag{4.14a}$$

$$B_m^w(\hat{\Theta}^{[l]}) = \left(\frac{\hat{\sigma}_w^{2[l]}}{\hat{\sigma}_w^{2[l]} + \hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}}\right)^2 \tag{4.14b}$$

$$B_m^x(\hat{\Theta}^{[l]}) = \left(\frac{\hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}}{\hat{\sigma}_w^{2[l]} + \hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}}\right)^2. \tag{4.14c}$$

M step: This step reduces to using these signal and noise variance estimates to obtain the new parameter estimates $\hat{\beta}^{[l+1]}$, $\hat{\sigma}^{2[l+1]}$, and $\hat{\sigma}_w^{2[l+1]}$:

$$\hat{\beta}^{[l+1]} \leftarrow \sum_{m\in\mathcal{M}} C_m N(m) S_m^x(\hat{\Theta}^{[l]})\beta^m = 0 \tag{4.15a}$$

$$\hat{\sigma}^{2[l+1]} = \frac{\sum_{m\in\mathcal{M}} N(m) S_m^x(\hat{\Theta}^{[l]})\left[\hat{\beta}^{[l+1]}\right]^m}{\sum_{m\in\mathcal{M}} N(m)} \tag{4.15b}$$

$$\hat{\sigma}_w^{2[l+1]} = \frac{\sum_{m\in\mathcal{M}} N(m) S_m^w(\hat{\Theta}^{[l]})}{\sum_{m\in\mathcal{M}} N(m)} \tag{4.15c}$$

where

$$C_m \triangleq \frac{m}{\sum_{m\in\mathcal{M}} mN(m)} - \frac{1}{\sum_{m\in\mathcal{M}} N(m)}. \tag{4.16}$$

4.3.2 Case II: β, σ^2 Unknown; σ_w^2 Known

If σ_w^2 is known, the above algorithm simplifies somewhat. In particular, we may omit the estimation (4.15c) and replace occurrences of $\hat{\sigma}_w^{2[l]}$ in the algorithm with the true value σ_w^2. This eliminates the need to compute $S_m^w(\hat{\Theta}^{[l]})$ and, hence, $B_m^w(\hat{\Theta}^{[l]})$. The resulting algorithm is as follows.

E step: Estimate the signal portion of the wavelet coefficient variances at each scale $m \in \mathcal{M}$ using current estimates of the parameters $\hat{\beta}^{[l]}$ and $\hat{\sigma}^{2[l]}$:

$$S_m^x(\hat{\Theta}^{[l]}) = A_m(\hat{\Theta}^{[l]}) + B_m^x(\hat{\Theta}^{[l]})\hat{\sigma}_m^2 \tag{4.17}$$

where

$$A_m(\hat{\Theta}^{[l]}) = \frac{\sigma_w^2 \cdot \hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}}{\sigma_w^2 + \hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}} \tag{4.18a}$$

$$B_m^x(\hat{\Theta}^{[l]}) = \left(\frac{\hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}}{\sigma_w^2 + \hat{\sigma}^{2[l]}\left[\hat{\beta}^{[l]}\right]^{-m}}\right)^2. \tag{4.18b}$$

M step: Use these signal variance estimates to obtain the new parameter estimates $\hat{\beta}^{[l+1]}$ and $\hat{\sigma}^{2[l+1]}$:

$$\hat{\beta}^{[l+1]} \leftarrow \sum_{m\in\mathcal{M}} C_m N(m) S_m^x(\hat{\Theta}^{[l]})\beta^m = 0 \tag{4.19a}$$

$$\hat{\sigma}^{2[l+1]} = \frac{\sum_{m\in\mathcal{M}} N(m) S_m^x(\hat{\Theta}^{[l]})\left[\hat{\beta}^{[l+1]}\right]^m}{\sum_{m\in\mathcal{M}} N(m)} \tag{4.19b}$$

where C_m is as in (4.16).

4.3.3 Case III: β, σ^2 Unknown; $\sigma_w^2 = 0$

If σ_w^2 is known (or assumed) to be zero, the EM algorithm becomes unnecessary as the likelihood may be maximized directly. Specifically, with $\sigma_w^2 = 0$, the signal variance estimates are available directly as $\hat{\sigma}_m^2$. Hence the estimation simplifies to the following:

$$\hat{\beta}_{\mathrm{ML}} \leftarrow \sum_{m\in\mathcal{M}} C_m N(m)\hat{\sigma}_m^2\beta^m = 0 \tag{4.20a}$$

$$\hat{\sigma}_{\mathrm{ML}}^2 = \frac{\sum_{m\in\mathcal{M}} N(m)\hat{\sigma}_m^2\left[\hat{\beta}_{\mathrm{ML}}\right]^m}{\sum_{m\in\mathcal{M}} N(m)} \tag{4.20b}$$

with C_m still as in (4.16).

It is worth discussing this special case in more detail not only for its own sake, but also because it characterizes one of the components of each iteration of the EM algorithm. The derivation of the parameter estimates in this case is essentially the same as the derivation of the M step in Appendix C. We begin by differentiating the likelihood function to find equations for its stationary points. This leads to a pair of equations in terms of σ^2 and β. Eliminating σ^2 from these equations is straightforward and gives (4.20a) directly. Having determined $\hat{\beta}_{\mathrm{ML}}$ as the solution to this polynomial equation, $\hat{\sigma}_{\mathrm{ML}}^2$ is obtained by back substitution.

From Lemma C.1 in Appendix C, it is apparent that (4.20a) has exactly one positive real solution, which is the ML estimate $\hat{\beta}_{\mathrm{ML}}$. Hence, L has a unique local and hence global maximum. Moreover, we may use bisection as a method to find the solution to this equation, provided we start with an initial interval containing $\hat{\beta}_{\mathrm{ML}}$. For instance, when we expect $0 < \gamma < 2$, an appropriate initial interval is $1 < \beta < 4$. Naturally, with some caution, Newton iterations [81] may be used to accelerate convergence.

Again, since solving equations of the form of (4.20) constitutes the M step of the iterative algorithm for the more general problem, the above remarks are equally applicable in those contexts.

4.3.4 Properties of the Estimators

In this section, we consider two principal issues: how the parameter estimates of the EM algorithm converge to the ML parameter estimates; and how the ML parameter estimates converge to the true parameter values.

Regarding the first of these issues, we are assured that the EM algorithm always adjusts the parameter estimates at each iteration so as to increase the likelihood function until a stationary point is reached. It can be shown that in our problem, the likelihood function has multiple stationary points, one of which corresponds to the desired ML parameter estimates. Others correspond to rather pathological saddle points of the likelihood function at the boundaries of the parameter space:

$$\hat{\beta} = \hat{\beta}_{\mathrm{ML}}\big|_{\sigma_w^2=0}$$

$$\hat{\sigma}^2 = \hat{\sigma}_{\mathrm{ML}}^2\big|_{\sigma_w^2=0}$$

$$\hat{\sigma}_w^2 = 0$$

and

$$\begin{aligned} \hat{\beta} &: \text{arbitrary} \\ \hat{\sigma}^2 &= 0 \\ \hat{\sigma}_w^2 &= \frac{\sum_{m\in\mathcal{M}} N(m)\hat{\sigma}_m^2}{\sum_{m\in\mathcal{M}} N(m)}. \end{aligned}$$

That they are saddle points is rather fortunate, for the only way they are reached is if the starting value for any one of $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{\sigma}_w^2$ is chosen to be exactly zero. Given arbitrarily small positive choices for these initial parameters, the algorithm iterates towards the ML parameters.

The preceding discussion suggests that the EM algorithm is fundamentally rather robust in this application. However, the selection of the initial parameter values naturally affects the rate of convergence of the algorithm. Moreover, it should be noted that the EM algorithm converges substantially faster for the case in which σ_w^2 is known. In essence, for the general algorithm much of the iteration is spent locating the noise threshold in the data.

Turning now to a discussion of the properties of the ML estimates themselves, it is well known that ML estimates are generally asymptotically efficient and consistent. This, specifically, turns out to be the case here [82]. It is also the case that at least in some higher signal-to-noise ratio (SNR) scenarios, the Cramér-Rao bounds closely approximate the true estimation error variances.

To compute the Cramér-Rao bounds for the estimates of γ, σ^2, and σ_w^2, we construct the corresponding Fisher matrix

$$\mathbf{I} = \sum_{m\in\mathcal{M}} \frac{N(m)}{2(\sigma_m^2)^2} \begin{bmatrix} [\ln 2^m\sigma^2\beta^{-m}]^2 & -\ln 2^m\sigma^2[\beta^{-m}]^2 & -\ln 2^m\sigma^2\beta^{-m} \\ -\ln 2^m\sigma^2[\beta^{-m}]^2 & [\beta^{-m}]^2 & \beta^{-m} \\ -\ln 2^m\sigma^2\beta^{-m} & \beta^{-m} & 1 \end{bmatrix} \tag{4.21}$$

from which we get

$$\begin{aligned} \operatorname{var}\hat{\gamma} &\geq I^{11} \\ \operatorname{var}\hat{\sigma}^2 &\geq I^{22} \\ \operatorname{var}\hat{\sigma}_w^2 &\geq I^{33} \end{aligned}$$

for any unbiased estimates $\hat{\gamma}, \hat{\sigma}^2, \hat{\sigma}_w^2$, and where I^{kk} is the kth element on the diagonal of $\mathbf{I}^{-1}$. However, local bounds such as these are of limited value in general both because our estimates are biased and because the bounds involve the true parameter values, which are unknown.

When σ_w^2 is known, the Fisher information matrix simplifies to the upper submatrix

$$\mathbf{I} = \sum_{m\in\mathcal{M}} \frac{N(m)[\beta^{-m}]^2}{2(\sigma_m^2)^2} \begin{bmatrix} [\ln 2^m\sigma^2]^2 & -\ln 2^m\sigma^2 \\ -\ln 2^m\sigma^2 & 1 \end{bmatrix} \tag{4.22}$$

from which we get

$$\begin{aligned} \operatorname{var}\hat{\gamma} &\geq I^{11} \\ \operatorname{var}\hat{\sigma}^2 &\geq I^{22}. \end{aligned}$$

As one would expect, both the actual error variances and the Cramér-Rao bounds are smaller for this case. Note that because the bounds are still a function of the parameters in this case, their usefulness remains limited. Nevertheless, except in very low SNR settings, the estimate biases are small in a relative sense and the estimation error variance is reasonably well approximated by these bounds. Hence, the bounds are at least useful in reflecting the quality of estimation that can be expected in various scenarios.

When $\sigma_w^2 = 0$, we get still further simplification, and we can write

$$\mathbf{I} = \begin{bmatrix} (\ln 2)^2/2 \sum_{m\in\mathcal{M}} m^2N(m) & -(\ln 2)/(2\sigma^2) \sum_{m\in\mathcal{M}} mN(m) \\ -(\ln 2)/(2\sigma^2) \sum_{m\in\mathcal{M}} mN(m) & 1/(2\sigma^4) \sum_{m\in\mathcal{M}} N(m) \end{bmatrix} \tag{4.23}$$

from which we get

$$\begin{aligned} \operatorname{var}\hat{\gamma} &\geq 2/\left[(\ln 2)^2 J\right] \sum_{m\in\mathcal{M}} N(m) \\ \operatorname{var}(\hat{\sigma}^2/\sigma^2) &\geq 2/J \sum_{m\in\mathcal{M}} m^2N(m) \end{aligned}$$

where

$$J = \left[\sum_{m\in\mathcal{M}} m^2N(m)\right]\left[\sum_{m\in\mathcal{M}} N(m)\right] - \left[\sum_{m\in\mathcal{M}} mN(m)\right]^2.$$

In this case, the bounds no longer depend on the parameters. Moreover, in practice, these expressions give an excellent approximation to the variances of the ML estimates. Evaluating the Cramér-Rao bounds asymptotically for the usual implementation scenario described by (4.9), we get

$$\operatorname{var}\hat{\gamma}_{\mathrm{ML}} \sim 2/\left[(\ln 2)^2 N\right] \tag{4.24a}$$

$$\operatorname{var}(\hat{\sigma}_{\mathrm{ML}}^2/\sigma^2) \sim 2(\log_2 N)^2/N \tag{4.24b}$$

where N is the number of observation samples.

4.3.5 Simulations

For the Monte Carlo simulations of this section, we synthesize discrete samples of resolution-limited Gaussian $1/f$ processes embedded in stationary white Gaussian noise. In general, we vary the length N and SNR of the observations sequence as well as the spectral exponent γ of the underlying $1/f$ processes. We then perform parameter estimation using algorithms for

the most general scenario, corresponding to the case in which all signal and noise parameters $\beta, \sigma^2, \sigma_w^2$ are unknown.

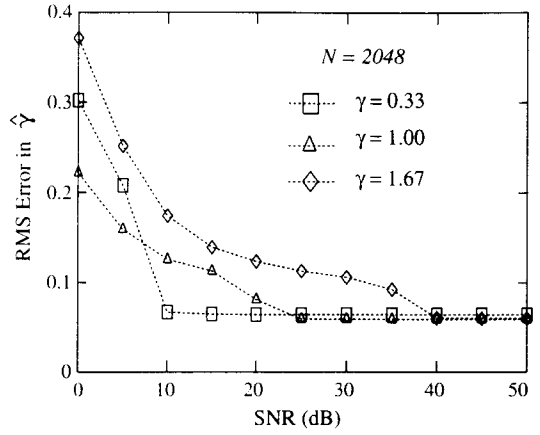
In Fig. 4.2, the RMS error in the estimates of γ and σ^2 is plotted for various values of γ as a function of SNR where the observation sequence length is fixed to $N = 2048$. The results from 64 trials were averaged to obtain the error estimates shown. As the results suggest, the quality of the estimates of both parameters is bounded as a consequence of the finite length of the observations. Moreover, the bounds are virtually independent of the value of γ and are achieved asymptotically. For increasing values of γ, the results suggest that the bounds are attained at increasing SNR thresholds.

In Fig. 4.3, the RMS error in the estimates of γ and σ^2 is plotted for various values of γ as a function of observation sequence length N where the SNR is fixed to 20 dB. Again, results from 64 trials were averaged to obtain the error estimates shown. While the results show that the estimation error decreases with data length as expected, they also suggest, particularly for the case of σ^2, that the convergence toward the true parameters can be rather slow. Note, too, that a rather large amount of data is required before the relative estimation error in σ^2 can be made reasonably small.

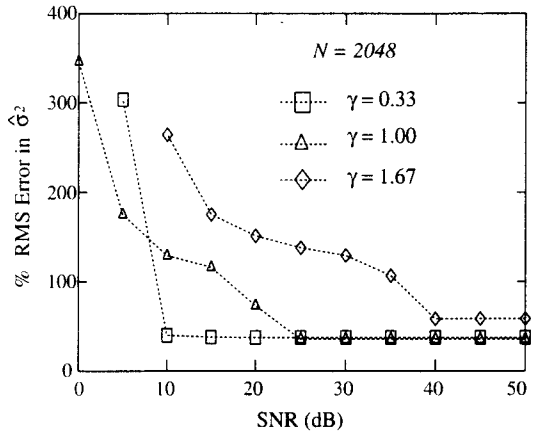
We conclude this section with a demonstration of the tracking capabilities of the parameter estimation algorithm. Specifically, Fig. 4.4 illustrates the performance of the parameter estimation in tracking a step-change in the spectral exponent γ of a noise-free $1/f$ signal. The signal was constructed such that the left and right halves of the signal correspond to $\gamma = 0.90$ and $\gamma = 1.10$, respectively, but identical variances. Local estimates of γ are computed by applying the Case III parameter estimation algorithm to the signal under a sliding window of length 16,384 centered about the point of interest. Note that the algorithm not only accurately resolves the appropriate spectral exponents, but accurately locates the point of transition as well. It is useful to point out that, as is the case with most tracking algorithms, using a wider estimation window would reduce the variance in the parameter estimates within each half of the waveform, but at the expense of an increase in the width of the transition zone.

4.4 SMOOTHING OF $1/f$ SIGNALS

In this section, we consider the problem of extracting a $1/f$ signal from a background of additive stationary white noise [77]. There are many potential problems involving signal enhancement and restoration to which the resulting smoothing algorithms can be applied. For this signal estimation problem, we use a Bayesian framework to derive algorithms that are optimal with respect to a mean-square error criterion. We specifically consider the Gaussian case, for which the resulting algorithms not only yield estimates

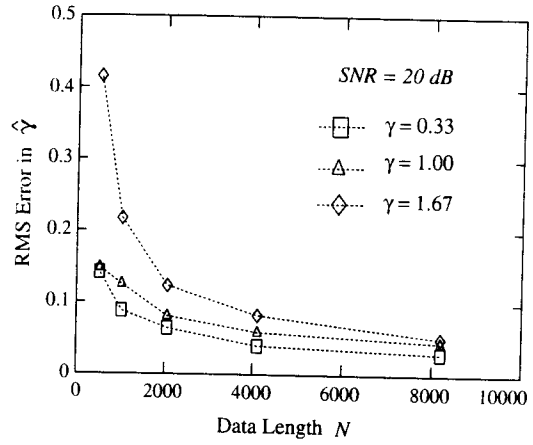


(a)

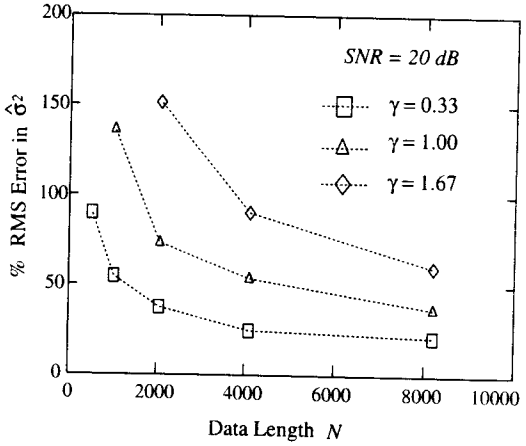


(b)

Figure 4.2. RMS Errors in the estimates of the signal parameters as a function of the SNR of the observations. The symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only. (a) Absolute RMS error in $\hat{\gamma}_{ML}$. (b) Percentage RMS error in $\hat{\sigma}^2_{ML}$.

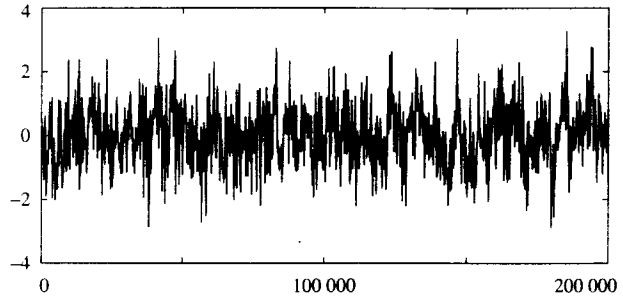


(a)

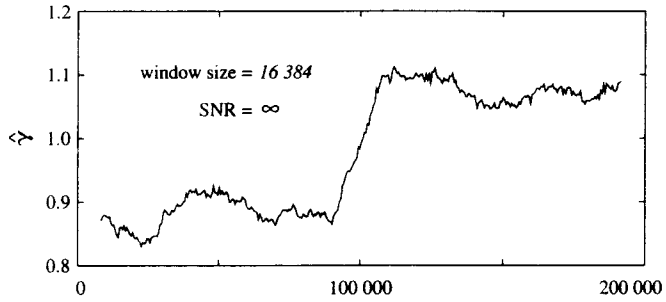


(b)

Figure 4.3. RMS Errors in the estimates of the signal parameters as a function of the data length N of the observations. Again, the symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only. (a) Absolute RMS error in $\hat{\gamma}_{\mathrm{ML}}$. (b) Percentage RMS error in $\hat{\sigma}^2_{\mathrm{ML}}$.



(a)



(b)

Figure 4.4. Tracking the time-varying spectral exponent γ of a noise-free $1/f$-type signal. For the left half of the signal, $\gamma = 0.90$, while for right half, $\gamma = 1.10$. (a) $1/f$ signal with step change in γ. (b) Estimate of γ under a time-limited window.

having the minimum possible mean-square error, but correspond to *linear* data processors as well. However, more generally, for non-Gaussian scenarios the estimators we derive are optimal in a *linear least-squares* sense, i.e., no other linear data processor is capable of yielding signal estimates with a smaller mean-square error [76].

Our basic formulation is to consider the estimation of a $1/f$ signal $x(t)$ from noisy observations $r(t)$ of the form (4.6), viz.,

$$r(t) = x(t) + w(t)$$

where $w(t)$ is stationary white noise, and where we still consider zero-mean processes. We assume in our derivation that the signal and noise parameters $\beta, \sigma^2, \sigma_w^2$ are all known, though, in practice they are typically estimated using the parameter estimation algorithms of the last section. In fact, the parameter and signal estimation problems are quite closely coupled. Indeed it will become apparent in our subsequent development that smoothing was inherently involved in the parameter estimation process as well.

We, again, exploit the wavelet decomposition to obtain our results. Specifically, we begin with the set of wavelet coefficients (4.8). Then, since

$$r_n^m = x_n^m + w_n^m$$

where the x_n^m and w_n^m are all mutually independent with variances $\sigma^2\beta^{-m}$ and σ_w^2 respectively, it follows immediately using classical estimation theory that the estimate of x_n^m that minimizes the mean-square estimation error is given by

$$\hat{x}_n^m = E\left[x_n^m|\mathbf{r}\right] = \begin{cases} E\left[x_n^m|r_n^m\right] & m,n \in \mathcal{R} \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, when x_n^m and r_n^m are jointly Gaussian, it is straightforward to establish that the least-squares estimates are linear and given by

$$E\left[x_n^m|r_n^m\right] = \left[\frac{\sigma^2\beta^{-m}}{\sigma^2\beta^{-m} + \sigma_w^2}\right] r_n^m. \tag{4.25}$$

From these estimates, we can express our optimal estimate of the $1/f$ signal as

$$\hat{x}(t) = \sum_{m,n} \hat{x}_n^m \psi_n^m(t) = \sum_{m,n\in\mathcal{R}} \left[\frac{\sigma^2\beta^{-m}}{\sigma^2\beta^{-m} + \sigma_w^2}\right] r_n^m \psi_n^m(t). \tag{4.26}$$

Note that, consistent with our earlier discussion of Wiener filtering for this problem, the smoothing factor

$$\frac{\sigma^2\beta^{-m}}{\sigma^2\beta^{-m} + \sigma_w^2}$$

in (4.26) has a thresholding role: at coarser scales where the signal predominates the coefficients are retained, while at finer scales where noise predominates, the coefficients are discarded. Note, too, that this factor appears in (4.14c), which allows us to interpret (4.13b) in terms of sample-variance estimates of the *smoothed* data. Evidently smoothing is inherently involved in the parameter estimation problem.

Interpreting the optimal estimator (4.26) in terms of the whitening filters of Section 4.2 leads to a conceptually convenient and familiar realization. In particular, as depicted in Fig. 4.5, the optimal linear processor consists of two stages. In the first stage, the noisy observations $r(t)$ are processed by a

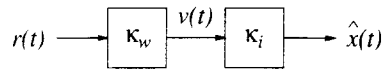


Figure 4.5. A canonic form implementation of the optimal linear filter for estimating a $1/f$ signal $x(t)$ from noisy observations $r(t)$. The linear least-squares filter is the cascade of a whitening filter followed by an innovations filter. The intermediate innovations process $v(t)$ is stationary and white.

whitening filter with kernel $\kappa_w(t,\tau)$ given by (4.4b) to generate an intermediate white *innovations* process $v(t)$ whose wavelet coefficients are

$$v_n^m = \frac{r_n^m}{\sigma_m}.$$

In the second stage, $v(t)$ is processed by an *innovations filter* with kernel

$$\kappa_i(t,\tau) = \sum_m \sum_n \psi_n^m(t) \left[\frac{\sigma^2\beta^{-m}}{\sigma_m}\right] \psi_n^m(\tau) \tag{4.27}$$

to generate the optimal estimate $\hat{x}(t)$ with wavelet coefficients given by (4.25). This innovations-based implementation is a classical estimation structure [40].

In practice, good performance is achieved by these estimators even in very poor SNR scenarios. This is not surprising given the preponderance of energy at low frequencies (coarse scales) in $1/f$-type processes. Let us then turn to a quantitative analysis of the estimation error. First, we note that because our set of observations is finite the total mean-square estimation error

$$\int_{-\infty}^{\infty} E\left[(\hat{x}(t) - x(t))^2\right] dt$$

is infinite. Nevertheless, when we define

$$x_{\mathcal{R}}(t) = \sum_{m,n\in\mathcal{R}} x_n^m \psi_n^m(t)$$

as the best possible approximation to $x(t)$ from the finite data set, we can express the total mean-square error in our estimate with respect to $x_{\mathcal{R}}(t)$ as

$$\begin{aligned} \varepsilon &= \int_{-\infty}^{\infty} E\left[(\hat{x}(t) - x_{\mathcal{R}}(t))^2\right] dt \\ &= \sum_{m,n\in\mathcal{R}} E\left[(\hat{x}_n^m - x_n^m)^2\right] \\ &= \sum_{m,n\in\mathcal{R}} E\left[\text{var}(x_n^m|r_n^m)\right] \end{aligned}$$

which, through routine manipulation, reduces to

$$\varepsilon = \sum_{m\in\mathcal{M}} N(m) \left[\frac{\sigma^2\beta^{-m}\cdot\sigma_w^2}{\sigma^2\beta^{-m} + \sigma_w^2}\right]. \tag{4.28}$$

As a final comment, note that while we do not develop the signal estimation in terms of Wiener filtering in the frequency domain, interpretations in this domain provide useful insight. In particular, it is clear that at high frequencies the white noise spectrum dominates, while at low frequencies the $1/f$ signal spectrum dominates. In fact, at sufficiently low frequencies, there is always arbitrarily high SNR regardless of the noise threshold. Consequently, Wiener filtering for this problem involves a form of low-pass filtering, where the exact filter shape and "cut-off" are governed by the particular parameters of the noise and signal spectra. Moreover, this low-pass filtering is effectively implemented on a logarithmic frequency scale—the scale which is most natural for these processes.

4.4.1 Simulations

For the simulations of this section, discrete samples of resolution-limited Gaussian $1/f$ processes embedded in Gaussian white noise are synthesized. In general, the SNR of the observations sequence is varied as well as the spectral exponent γ of the underlying $1/f$ processes. Parameter estimation is then performed, followed by signal estimation, using algorithms for the most general scenario, corresponding to the case in which all signal and noise parameters $\beta, \sigma^2, \sigma_w^2$ are unknown. Note that by using the estimated parameters in the signal estimation algorithm, these experiments do not allow us to distinguish between those components of signal estimation error due to errors in the estimated parameter values and those due to the smoothing process itself. This limitation is not serious, however, since the quality of the signal estimation is generally rather insensitive to errors in the parameter estimates used.

In Fig. 4.6, the SNR gain of the smoothed signal estimates is plotted for various values of γ as a function of the SNR of the observations where the sequence length is fixed to $N = 2048$. Again, results from 64 trials were averaged to obtain the error estimates shown. The SNR gains predicted by the total mean-square error formula (4.28) are also superimposed on each plot. As the results indicate, the actual SNR gain is typically no more than 1 dB below the predicted gain, as would be expected. However, under some circumstances the deviation can be more than 3 dB. Worse, the SNR gain can be negative, i.e., the net effect of smoothing can be to increase the overall distortion in the signal. Such degradations in performance are due primarily to limitations on the accuracy to which the wavelet coefficients at coarser scales can be extracted via the DWT. In particular, they arise as a result of undesired effects introduced by modeling the data outside the observation interval as periodic to accommodate the inherent data-windowing problem. By contrast, error in the parameter estimates is a much less significant factor in these degradations at reasonably high SNR. The plots also indicate that

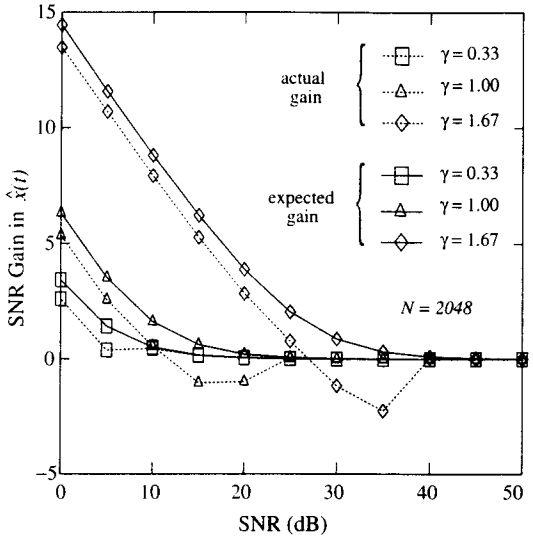


Figure 4.6. SNR gain (dB) of the signal estimate as a function of the SNR of the observations. Both the gains predicted by eq. (4.28) and gains actually obtained are indicated.

better gains are achieved for larger values of γ for a given SNR. This is to be expected since for larger values of γ there is more signal energy at coarser scales and correspondingly less at finer scales where the noise predominates and the most attenuation takes place.

We conclude this section with a signal estimation example. Fig. 4.7 shows a segment of a 65,536-sample $1/f$ signal, the same signal embedded in noise, and the signal estimate. In this example, the spectral exponent is $\gamma = 1.67$, and the SNR in the observations is 0 dB. The estimated spectral exponent is $\hat{\gamma}_{\mathrm{ML}} = 1.66$, and the SNR gain of the signal estimate is 13.9 dB. As anticipated, the signal estimate effectively preserves detail at the coarse scales where the SNR was high, while detail at fine scales is lost where the SNR was low.

4.5 COHERENT DETECTION IN $1/f$ NOISE

In this section the problem of detecting a known signal of finite energy in a background of Gaussian $1/f$ and white noise is considered. The detection algorithms we develop may be used in a variety of applications; they may be exploited, for example, in communication and pattern recognition systems.

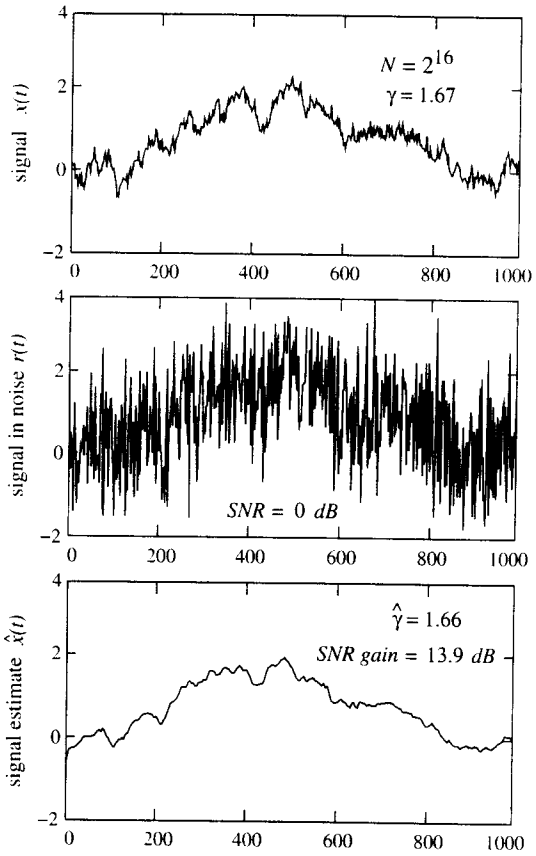


Figure 4.7. Optimal smoothing of a noisy $1/f$ signal.

In our approach, we explicitly include stationary white Gaussian measurement noise in our model. In general, this refinement improves the robustness properties of the resulting algorithms and, in particular, it precludes certain singular detection scenarios. As will become apparent, the wavelet-based approach we develop is not only analytically and conceptually convenient, but leads to practical implementation structures as well.

Let us pose our detection problem in terms of a binary hypothesis test with a Neyman-Pearson optimality criterion [76]. Specifically, given noisy observations $r(t)$, we wish to determine a rule for deciding whether or not a known signal is present in the observations. For our test formulation, under

hypothesis H_1 we observe a signal of energy $E_0 > 0$ against a background of Gaussian $1/f$ and white noise, while under hypothesis H_0 we observe only the background noise, i.e.,

$$\begin{aligned} H_1 : r(t) &= \sqrt{E_0}s(t) + x(t) + w(t) \\ H_0 : r(t) &= x(t) + w(t) \end{aligned}$$

where $w(t)$ is stationary white Gaussian noise and $x(t)$ is Gaussian $1/f$-type noise, and $s(t)$ is a unit energy signal:

$$\int_{-\infty}^{\infty} s^2(t)\,dt = 1.$$

We assume that $w(t)$ and $x(t)$ are statistically independent processes under either hypothesis, and also that our observations generally extend over the infinite interval $-\infty < t < \infty$. We then seek to design a decision rule that maximizes the probability of detecting $s(t)$

$$P_D = \Pr(\text{decide } H_1 \mid H_1 \text{ true})$$

subject to a constraint on the maximum allowable false alarm probability

$$P_F = \Pr(\text{decide } H_1 \mid H_0 \text{ true}).$$

As is well known, the solution to this problem takes the form of a likelihood ratio test [76].

An equivalent hypothesis test can be constructed in terms of observations of the respective wavelet coefficients

$$\mathbf{r} = \{r_n^m\}$$

as

$$\begin{aligned} H_1 : r_n^m &= \sqrt{E_0}s_n^m + x_n^m + w_n^m \\ H_0 : r_n^m &= x_n^m + w_n^m \end{aligned}$$

for $-\infty < m < \infty$ and $-\infty < n < \infty$. According to our model, under each hypothesis, the coefficients w_n^m and x_n^m are all statistically independent, and have variances σ_w^2 and $\sigma^2\beta^{-m}$, respectively.

In this case, since joint distributions of the observations under the respective hypotheses are

$$\begin{aligned} p_{\mathbf{r}|H_1}(\mathbf{r}) &= \prod_{m,n} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(r_n^m - \sqrt{E_0}s_n^m)^2}{2\sigma_m^2}\right] \\ p_{\mathbf{r}|H_0}(\mathbf{r}) &= \prod_{m,n} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\sigma_m^2}\right], \end{aligned}$$

the likelihood ratio

$$\frac{p_{\mathbf{r}|H_1}(\mathbf{r})}{p_{\mathbf{r}|H_0}(\mathbf{r})}$$

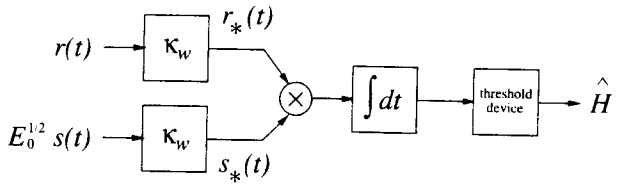


Figure 4.8. Canonical prewhitening implementation of the optimal receiver for detection of a known $s(t)$ in the presence of both Gaussian $1/f$ and stationary white Gaussian noise, where $\kappa_w(t,\tau)$ is the kernel of the whitening filter for $1/f$-plus-white noise.

can be simplified substantially to yield a test of the form

$$\ell = \sum_m \sum_n \frac{r_n^m s_n^m}{\sigma_m^2} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \tag{4.29}$$

where α is the threshold of the test.

This optimal detector may be realized using a whitening-filter-based implementation as shown in Fig. 4.8. The statistic ℓ is constructed by processing both $r(t)$ and $\sqrt{E_0}s(t)$ with a *prewhitening filter* whose kernel is given by (4.4b), and then correlating the respective outputs $r_*(t)$ and $s_*(t)$. It is straightforward to verify this implementation: since the prewhitened signals $r_*(t)$ and $s_*(t)$ have wavelet coefficients r_n^m/σ_m and $\sqrt{E_0}s_n^m/\sigma_m$, respectively, it suffices to recognize the expression for ℓ in (4.29) as the inner product between $s_*(t)/\sqrt{E_0}$ and $r_*(t)$, which allows us to rewrite (4.29) as

$$\ell = \int_{-\infty}^{\infty} r_*(t)\, s_*(t)/\sqrt{E_0}\, dt \underset{H_0}{\overset{H_1}{\gtrless}} \alpha.$$

This is, of course, a canonical form receiver for optimal detection in the presence of colored noise as described in [76].

Let us turn now to a discussion of the performance of this optimal receiver. Via the implementation of this receiver in terms of the whitened observations $r_*(t)$, we note that the performance is necessarily equivalent to that of an optimal detector for $s_*(t)$ in the presence of stationary white Gaussian noise of unit variance. Indeed, if we define the performance index d according to

$$d^2 \triangleq \int_{-\infty}^{\infty} s_*^2(t)\, dt = E_0 \sum_m \sum_n \frac{(s_n^m)^2}{\sigma_m^2} \tag{4.30}$$

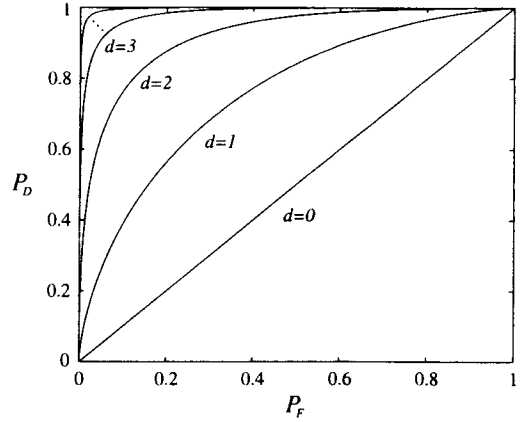


Figure 4.9. The receiver operating characteristic for the detection of a known signal in a background of $1/f$ plus white noise for various thresholds determined via the performance index d.

then

$$\begin{aligned} E[\ell|H_0] &= 0 \\ E[\ell|H_1] &= d^2/\sqrt{E_0} \\ \text{var}\{\ell|H_0\} = \text{var}\{\ell|H_1\} &= d^2/E_0. \end{aligned}$$

Hence, expressing our arbitrary threshold in the form

$$\alpha = \frac{d}{\sqrt{E_0}}\left[\frac{\ln\eta}{d} + \frac{d}{2}\right]$$

for some $0 < \eta < \infty$, the performance of the test can be described in terms of the detection and false alarm probabilities, respectively

$$P_D = Q\left(\frac{\ln\eta}{d} - \frac{d}{2}\right) \tag{4.31a}$$

$$P_F = Q\left(\frac{\ln\eta}{d} + \frac{d}{2}\right) \tag{4.31b}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-v^2/2}\, dv. \tag{4.32}$$

The familiar receiver operating characteristic (ROC) associated with such Gaussian detection problems is as shown in Fig. 4.9 for various values of d.

In concluding this section, we make some brief remarks on the problem of optimum signal design for use in $1/f$-plus-white backgrounds. Based on our analysis, it is apparent that we can optimize performance if we choose $s(t)$, or equivalently s_n^m, to maximize d^2 in (4.30) subject to the energy constraint

$$\int_{-\infty}^{\infty} s^2(t)\,dt = \sum_m \sum_n (s_n^m)^2 = 1.$$

However, this signal optimization problem is not well posed. Indeed, because of the spectral distribution of the background noise, the optimization attempts to construct a signal whose energy is at frequencies sufficiently high that the $1/f$ noise is negligible compared to the white component. Consequently, to preclude the generation of an arbitrarily high frequency signal, generally some form of bandwidth constraint is necessary. An example of how this can be accommodated in a communications scenario is described in Wornell [83].

4.6 DISCRIMINATING BETWEEN $1/f$ SIGNALS

In this section, we consider the ability of an optimal Bayesian detector to discriminate between Gaussian $1/f$ processes of distinct parameters in a background of stationary white Gaussian noise. The signal classification algorithms we derive are useful in a variety of potential applications. The problem of distinguishing $1/f$ processes is, of course, very closely related to the parameter estimation problem treated in Section 4.3. Indeed, parameter estimation can be viewed as distinguishing among an arbitrarily large number of $1/f$ processes with incrementally different parameters. Nevertheless, as we will see, approaching the problem from a detection perspective affords a number of new and useful insights.

It is, again, convenient to formulate our problem in terms of a binary hypothesis test in which under each hypothesis we have noisy observations $r(t)$ of distinct $1/f$ signals. Specifically, we have as our two hypotheses

$$H_0 : r(t) = \dot{x}(t) + w(t) \tag{4.33a}$$
$$H_1 : r(t) = \ddot{x}(t) + w(t) \tag{4.33b}$$

where $\dot{x}(t)$ and $\ddot{x}(t)$ are Gaussian $1/f$ processes[4] with distinct parameters and $w(t)$ is a white measurement noise, statistically independent of $\dot{x}(t)$ or $\ddot{x}(t)$, whose variance is the same under both hypotheses. For this test we develop a minimum probability of error [$\Pr(\varepsilon)$] decision rule under the assumption of equally likely hypotheses.

[4] In this section, the notations $\dot{}$ and $\ddot{}$ are used to distinguish the $1/f$ processes and their respective parameters under the two hypotheses. These symbols should not be confused with differentiation operators, for which we have generally reserved the notation $'$ and $''$.

Once again, the optimum receiver is best developed and analyzed in the wavelet domain. Rewriting the hypothesis test in terms of the corresponding wavelet coefficients as

$$H_0 : r_n^m = \dot{x}_n^m + w_n^m$$
$$H_1 : r_n^m = \ddot{x}_n^m + w_n^m,$$

we model the r_n^m under each hypothesis as a collection of zero-mean statistically independent Gaussian random variables with variances

$$\mathrm{var}\{r_n^m|H_0\} = \dot{\sigma}_m^2 = \dot{\sigma}^2\dot{\beta}^{-m} + \sigma_w^2 \tag{4.34a}$$
$$\mathrm{var}\{r_n^m|H_1\} = \ddot{\sigma}_m^2 = \ddot{\sigma}^2\ddot{\beta}^{-m} + \sigma_w^2 \tag{4.34b}$$

where

$$\dot{\beta} = 2^{\dot{\gamma}}$$
$$\ddot{\beta} = 2^{\ddot{\gamma}}.$$

In our derivation we assume that, in general, only a *finite* collection of observation coefficients of the form

$$\mathbf{r} = \{r_n^m \in \mathcal{R}\} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\},$$

where $\mathcal{M}$ and $\mathcal{N}(m)$ are as defined in (4.7), are available. In fact, as we will see, the problem turns out to be singular (i.e., perfect detection is achievable) if complete observations over the infinite interval are available. In our simulations, we assume the observation set $\mathcal{R}$ to be of the particular form (4.9), which corresponds to the collection of coefficients generally available from time- and resolution-limited observations of $r(t)$ via a DWT algorithm as discussed in Section 2.3.4.

The likelihood ratio test for this problem can, through straightforward manipulation, be simplified to a test of the form

$$\ell = \frac{1}{2}\sum_{m\in\mathcal{M}} N(m)\left\{\left[\frac{1}{\dot{\sigma}_m^2} - \frac{1}{\ddot{\sigma}_m^2}\right](r_n^m)^2 - \ln\frac{\ddot{\sigma}_m^2}{\dot{\sigma}_m^2}\right\} \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{4.35}$$

It is straightforward to show that this test can be implemented in the canonical form shown in Fig. 4.10. Here the observations $r(t)$ are processed by $1/f$-plus-white whitening filters corresponding to each hypothesis, for which the respective kernels are

$$\dot{\kappa}_w(t,\tau) = \sum_m \sum_n \psi_n^m(t)\frac{1}{\dot{\sigma}_m}\psi_n^m(\tau)$$
$$\ddot{\kappa}_w(t,\tau) = \sum_m \sum_n \psi_n^m(t)\frac{1}{\ddot{\sigma}_m}\psi_n^m(\tau).$$

Consequently, only one of the residual processes $\dot{v}(t)$ and $\ddot{v}(t)$ is white, depending on which hypothesis is true. To decide between the two hypotheses,

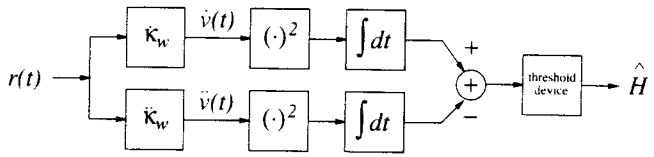


Figure 4.10. A canonical form implementation of the optimal receiver for discriminating between $1/f$ models with distinct parameters based on noisy observations $r(t)$.

the receiver computes the difference in energy in the two residuals and compares it to the appropriate threshold.

Although the detection problem is Gaussian, it is apparent that the the log-likelihood ℓ is not conditionally Gaussian under either hypothesis. Consequently, evaluating the performance of such receivers is rather difficult in general. Nevertheless, it is possible to obtain approximate performance results by exploiting a procedure described in Van Trees [76] based upon the use of the Chernoff bound. Specifically, defining

$$\mu(s) \triangleq \ln E\left[e^{s\ell}|H_0\right],$$

for an arbitrary real parameter s, we can bound the performance of our optimal detector according to

$$\Pr(\varepsilon) \leq \frac{1}{2} e^{\mu(s_*)}, \tag{4.36}$$

where s_* is the parameter value yielding the best possible bound, i.e.,

$$s_* = \arg\min_s \mu(s).$$

When there are sufficiently many observations to justify modeling ℓ as Gaussian via a central limit theorem (CLT) argument, we can also obtain the following asymptotic expression for the error probability

$$\Pr(\varepsilon) \approx \frac{1}{2s_*(1-s_*)\sqrt{2\pi\mu''(s_*)}} e^{\mu(s_*)}, \tag{4.37}$$

which is a more optimistic and accurate estimate of the achievable performance [76].

In our case, $\mu(s)$ and its first two derivatives are given by

$$\mu(s) = \frac{1}{2}\sum_{m\in\mathcal{M}} N(m)\left\{s\ln\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - \ln\left[s\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} + (1-s)\right]\right\} \tag{4.38a}$$

$$\mu'(s) = \frac{1}{2}\sum_{m\in\mathcal{M}} N(m)\left\{\ln\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - \left[\frac{\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - 1}{s\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} + (1-s)}\right]\right\} \tag{4.38b}$$

$$\mu''(s) = \frac{1}{2}\sum_{m\in\mathcal{M}} N(m)\left\{\frac{\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} - 1}{s\frac{\dot{\sigma}_m^2}{\ddot{\sigma}_m^2} + (1-s)}\right\}^2. \tag{4.38c}$$

It is generally not possible to derive a closed form expression for the minimum value of $\mu(s)$ via either (4.38a) or (4.38b) for this asymmetric detection problem. Fortunately though, a numerical optimization is reasonable: it suffices to consider a numerical search over values of s within a limited range. Indeed, since

$$\mu(0) = \mu(1) = 0,$$

and since from (4.38c) we have that $\mu(s)$ is convex

$$\mu''(s) \geq 0,$$

it follows that the minimum of $\mu(s)$ can be found in the range $0 \leq s \leq 1$.

4.6.1 Simulations

In this section, we obtain, via (4.36) and (4.37), numerical estimates of the probability of error performance. As our scenario, we assume that coefficients r_n^m are available in the range (4.9) consistent with what could be extracted from $N = 2^M$ samples of time- and resolution-limited observations via a DWT algorithm. In the simulations, we consider the ability of the optimal receiver to distinguish between $1/f$ processes of different spectral exponents γ (or, equivalently, fractal dimensions D, or self-similarity parameters H). In particular, we do not consider the capabilities of the algorithms to discriminate on the basis of variance differences. Consequently, in all our tests, we choose the variance parameters $\dot{\sigma}^2$ and $\ddot{\sigma}^2$ such that the variance of the observations is identical under either hypothesis.

In the first set of simulations, the bound (4.36) is used as an estimate of the probability of error performance of an optimal detector in discriminating between two equal-variance $1/f$ processes whose spectral exponents differ by $\Delta\gamma$ based on noisy observations of length N corresponding to a prescribed SNR. In the tests, three different spectral exponent regimes are considered, corresponding to $\gamma = 0.33$, $\gamma = 1.00$, and $\gamma = 1.67$.

In Fig. 4.11, performance is measured as a function of SNR for noisy observations of length $N = 128$ and a parameter separation $\Delta\gamma = 0.1$. Note that there is a threshold phenomenon: above a certain γ-dependent SNR,

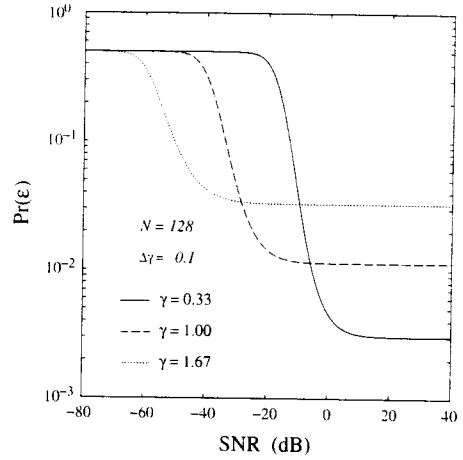


Figure 4.11. Optimal discriminator performance as a function of SNR, as estimated via the Chernoff bound.

$\Pr(\varepsilon)$ drops dramatically. Moreover, the threshold is lower for larger values of γ. This is to be expected since larger values of γ correspond to an effective $1/f$ power spectrum that is increasingly peaked at the $\omega = 0$, so that a correspondingly greater proportion of the total signal power is not masked by the white observation noise. Beyond this threshold performance saturates as the data is essentially noise-free. However, note that there is crossover behavior: at SNR values above the thresholds, better performance is obtained for smaller values of γ. In subsequent tests, we restrict our attention to performance in this high SNR regime.

In Fig. 4.12, performance is plotted as a function of the number of samples N of observed data corresponding to an SNR of 20 dB and hypotheses whose parameter separation is $\Delta\gamma = 0.1$. In this case, there is thresholding behavior as well. For data lengths beyond a critical order-of-magnitude we get strongly increasing performance as a function of data length. Again, because we are in the high SNR regime, observe that the best performance is achieved for the smallest values of γ.

Finally, in Fig. 4.13, performance is plotted as a function of the separation between the two hypotheses—specifically, the difference between the spectral parameters for noisy observations of length $N = 128$ corresponding to an SNR of 20 dB. As one would expect, the results illustrate that the larger the distinction between the hypotheses, the better the performance achievable

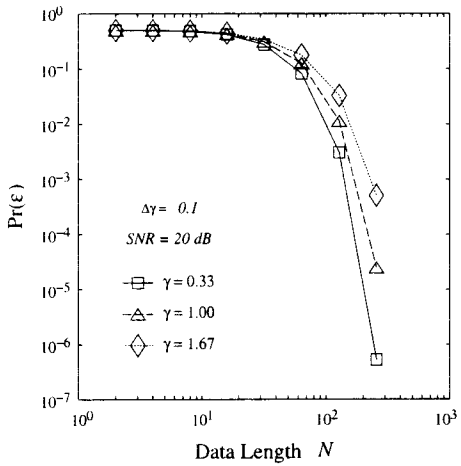


Figure 4.12. Optimal discriminator performance as a function of the number of samples N of noisy observations, as estimated by the Chernoff bound. The symbols □, △, and ◇ correspond to actual estimates; the lines are provided as visual aides only in this case.

by the receiver. Again, as we are in the high SNR regime, better performance is achieved for smaller values of γ.

Whenever the probability of error is low—i.e., either when the SNR is high, large data lengths are involved, or the hypotheses are well separated—it turns out that the CLT-based approximation (4.37) represents a more optimistic estimate of performance than does (4.36). However, in high $\Pr(\varepsilon)$ scenarios, (4.36) constitutes a more useful measure of system performance than does (4.37). This behavior is illustrated in Figs. 4.14, 4.15, and 4.16 for hypotheses in the $\gamma = 1$ regime. Note that only at sufficiently high SNR, data lengths, and parameter separations does the CLT-based approximation actually yield a $\Pr(\varepsilon)$ estimate that is below the bound. From these plots we cannot, of course, assess whether the CLT-based approximation is overly optimistic in the high SNR regime. In general, we can only expect the estimate to be asymptotically accurate as $N \to \infty$. Nevertheless, the fact that the rate of change of $\Pr(\varepsilon)$ with respect to SNR, data length N, and parameter separation $\Delta\gamma$ has a similar form for both the bound and the approximation suggests that the Chernoff bound provides a meaningful estimate of achievable performance.

Before concluding this section, we consider a potentially useful and practical refinement of the optimal discrimination problem. There are a number of application contexts in which we would be more interested in

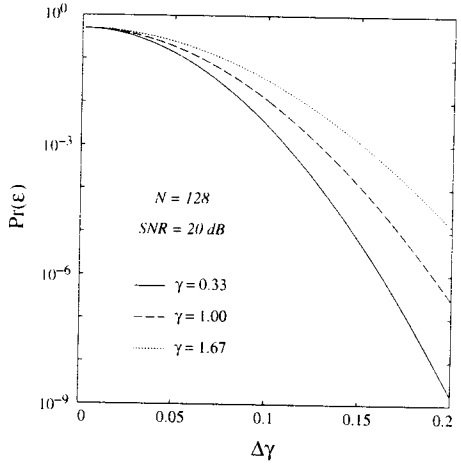


Figure 4.13. Optimal discriminator performance as a function of the parameter separation $\Delta\gamma$ between the two hypotheses, as estimated via the Chernoff bound.

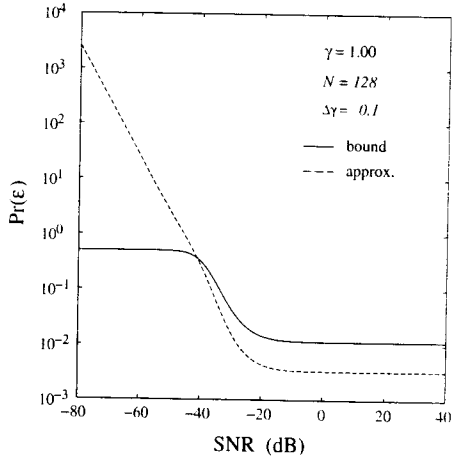


Figure 4.14. Optimal discriminator performance as a function of SNR, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37).

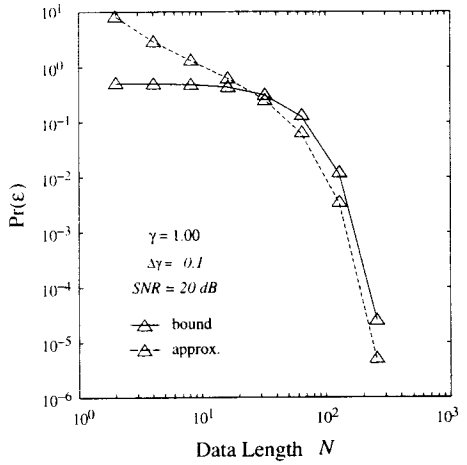


Figure 4.15. Optimal discriminator performance as a function of the number of samples of observed data, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37). The Δ symbols correspond to actual estimates; the lines are provided as visual aides only.

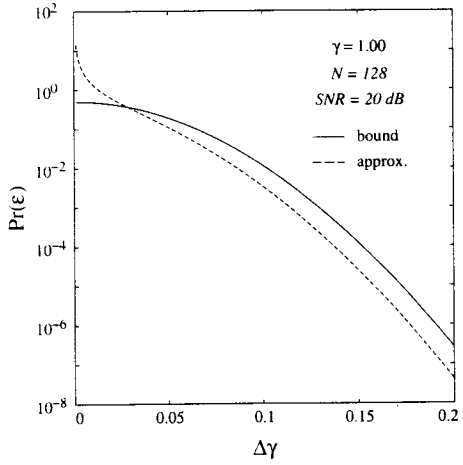


Figure 4.16. Optimal discriminator performance as a function of the parameter separation $\Delta\gamma$, as estimated via both the Chernoff-bound (4.36) and the CLT-based approximation (4.37).

distinguishing $1/f$ processes strictly on the basis of their spectral exponents, fractal dimensions, or self-similarity parameters. This would correspond to a hypothesis test (4.33) in which $\dot{\sigma}^2$, $\ddot{\sigma}^2$ and σ_w^2 would be *unwanted* parameters of the problem. In this case, a solution could be obtained using a *generalized* likelihood ratio test [76] of the form

$$\frac{\max_{\dot{\sigma}^2,\ddot{\sigma}^2,\sigma_w^2} \prod_{m,n\in\mathcal{R}} \frac{1}{\sqrt{2\pi\ddot{\sigma}_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\ddot{\sigma}_m^2}\right]}{\max_{\dot{\sigma}^2,\sigma_w^2} \prod_{m,n\in\mathcal{R}} \frac{1}{\sqrt{2\pi\dot{\sigma}_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\dot{\sigma}_m^2}\right]} \begin{array}{c} H_1 \\ \gtrless \\ H_0 \end{array} 1. \tag{4.39}$$

In general, expressions for the maxima involved in the construction of the likelihood function of (4.39) cannot be obtained in closed form. However, a practical implementation of this receiver could potentially exploit an EM algorithm of the general type developed in Section 4.3. In terms of performance, we would anticipate that, in general, it would only be possible to adequately evaluate such a receiver through Monte Carlo simulations.

4.7 ALTERNATIVE APPROACHES AND RELATED DEVELOPMENTS

Until relatively recently, problems of detection and estimation involving $1/f$-type processes received relatively little attention in the literature. However, there has been strongly increasing interest in the topic and a number of interesting and useful related results have been developed. This section contains a summary of at least some of these results, though the list here is certainly not comprehensive.

One example is the work described in Barton and Poor [57], which considers problems of detection in the presence of fractional Gaussian noise using reproducing kernel Hilbert space theory. Using this framework, both infinite- and finite-interval whitening filters are developed for this class of $1/f$ noises, which, in turn, yields some important results on the detection of deterministic and Gaussian signals in the presence of such noise.

There is also a substantial and growing body of recent literature on the general topic of multiresolution stochastic processes, systems, and signal processing. A broad overview of work in this area is contained in Basseville et al. [84] and the references therein. In that work, the authors develop a tree/lattice-based framework for modeling multiscale processes and problems, and introduce some novel notions of "stationarity in scale" for such processes. Treating multiscale processes as "dynamical systems in scale," leads to several highly efficient algorithms for addressing a variety of problems involving parameter estimation, signal smoothing and interpolation, and data fusion. The $1/f$-type models we exploit in this chapter constitute a special class of the multiresolution stochastic processes developed by these authors. In particular, they are examples of processes characterized by a "Markov scale-to-scale" property. As a consequence, many of the multiresolution signal processing algorithms developed using this framework are directly applicable to $1/f$ processes as shown in, e.g., Chou [85].

There has also been progress in exploiting a discrete-time analog of fractional Brownian motion to obtain useful parameter and signal estimation algorithms. Results in this area are described in, e.g., Deriche and Tewfik [86] [87]. In addition, possible refinements to the estimation algorithms developed in this chapter are described in Kaplan and Kuo [88]. Finally, some extensions and generalizations of the algorithms in this chapter are developed in Lam [89].

4.8 SUMMARY

In this chapter, we exploited the efficiency of wavelet basis expansions for the $1/f$ family of fractal random processes to develop solutions to some fundamental problems of optimal detection and estimation involving $1/f$-type signals. As a foundation, we first derived wavelet-based synthesis and whitening filters for $1/f$ processes that formed the basis for essentially all the algorithms derived in the chapter.

We then proceeded to develop maximum likelihood algorithms for estimating the parameters of $1/f$ processes from noise-corrupted observations given various degrees of *a priori* knowledge. These included algorithms for robust fractal dimension estimation, which are useful in a wide range of applications. Next, we developed complementary minimum mean-square error signal smoothing algorithms for separating and extracting $1/f$ processes from noisy observations. By exploiting our wavelet-based framework, we obtained algorithms that are extremely computationally efficient and highly practicable.

In many applications, it is the noise that has the $1/f$ characteristics rather than the signal of interest. Such is the case, for example, in a variety of communications applications where information-bearing waveforms are transmitted in a combination of $1/f$ and white backgrounds. Motivated by these kinds of scenarios, we developed efficient wavelet-based algorithms for optimally detecting known signals in such backgrounds.

Finally, we addressed a common signal classification problem involving $1/f$ processes—specifically, discriminating between $1/f$ signals with different parameters. In this case, too, efficient minimum probability of error decision strategies were developed via a wavelet domain formulation.