Estimation of Fractal Signals from Noisy Measurements Using Wavelets

Gregory W. Wornell, *Member, IEEE,* and Alan V. Oppenheim, *Fellow, IEEE*

Abstract—The $1/f$ family of fractal processes are increasingly appealing candidates for data modeling in a variety of signal processing applications in light of the fact that such a wide range of phenomena are inherently well suited to these models. In contrast to the well-studied family of ARMA processes, $1/f$ processes are characterized by an inherent scale invariance and persistent long-term correlation structure. Despite their apparent applicability in many scenarios, they have received relatively little attention in the traditional signal processing literature. This has been due, at least in part, to the mathematical intractability of fractal processes. However, fractal signal representations in terms of orthonormal wavelet bases have recently been described that considerably simplify the analysis of these processes. We exploit the role of the wavelet transformation as a whitening filter for $1/f$ processes to address problems of parameter and signal estimation for $1/f$ processes embedded in white background noise. Robust, computationally efficient, and consistent iterative parameter estimation algorithms are derived based on the method of maximum likelihood, and Cramér-Rao bounds are obtained. Included among these algorithms are optimal fractal dimension estimators for noisy data. Algorithms for obtaining Bayesian minimum mean-square error signal estimates are also derived together with an explicit formula for the resulting error. These smoothing algorithms find application in signal enhancement and restoration. The parameter estimation algorithms, in addition to solving the spectrum estimation problem and to providing parameters for the smoothing process, are useful in problems of signal detection and classification. A variety of results from simulations are presented to demonstrate the viability of the algorithms.

I. Introduction

T
HE $1/f$ family of stochastic processes constitutes an increasingly important class of fractal signal models for a variety of signal processing applications due to the wide variety of data for which they are inherently well suited [9], [18], [15]. These intrinsically scale-invariant processes have a number of interesting characteristics, among which is a much more persistent long-term correlation structure than is present in, for example, the well-studied family of autoregressive moving average (ARMA) processes.[1]

The authors are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.
IEEE Log Number 9105655.

[1]Indeed, $1/f$ processes typically have correlation that decays polynomially in time rather than exponentially, as is the case for ARMA processes [9].

Traditionally, these processes have been mathematically awkward to manipulate. This has made the solution of many of the classical signal processing problems involving these processes rather difficult. In this paper, we make use of a new representation for $1/f$ processes [23], [25] to develop optimal estimation algorithms for problems involving $1/f$ signals corrupted by white noise. Specifically, we treat problems of parameter estimation, spectrum estimation, and signal estimation for $1/f$ processes in the presence of noise.[2]

Largely inspired by the seminal work on fractional Brownian motion by Mandelbrot and Van Ness [16], a considerable body of literature has evolved which explores these $1/f$ processes as a study in a self-similarity and long-range dependence. In addition to pursuing a complete characterization of all self-similar processes, there is considerable interest in assessing the implications of such persistent dependence in probability, statistics, and stochastic process theory. Taqqu, in [19], has compiled a substantial collection of references in this area.

An equally large body of literature is devoted to understanding both the physical origins and the ubiquity of such behavior in real data. Indeed, a tremendous range of physical phenomena are apparently well modeled as $1/f$ processes [9]. These include a large number of geophysical and economic time series, biological signals, noises in electronic devices, frequency variation in music, and burst error on communication channels [14]. More recently, two-dimensional extensions of $1/f$ processes have been considered for modeling of natural terrain and other textures [15], [18], [11].

While a number of useful models for $1/f$ processes has arisen, none has become universal. In the construction of Barnes and Allan [1] that was later refined by Mandelbrot and Van Ness [16], a class of Gaussian $1/f$ processes is modeled as filtered white noise, where the filtering is defined through what is essentially a convolution integral. Another construction described by van der Ziel [21] models $1/f$ processes through a superposition of first-order autoregressive (AR) processes, each of which is characterized by a single time constant. In [9], Keshner generates $1/f$ processes by driving white noise through an infinite cascade of pole-zero sections. More recently, Wornell [23], [25] has presented an orthonormal wavelet

[2]The interesting parallel work described in [4] considers some complementary estimation problems for a related family of multiresolution processes.

basis expansion for $1/f$ processes in terms of a collection of uncorrelated random variables. In this work we exploit this latter representation.

To jointly estimate the signal and noise process parameters from observations, we apply the method of maximum likelihood (ML). The parameter estimates, in addition to providing a solution to the associated 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. Robust estimation of the fractal dimension of $1/f$ processes is important in a number of applications such as in 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 [18], [11], fractal dimension can be of use in distinguishing among various man-made and natural objects. While several approaches to the fractal dimension estimation problem have been presented in the literature (see [11], [24], [8], and the reference therein), none has been able to adequately handle the presence of broad-band noise in the observation data. In fact, the quality of the estimates generally deteriorates dramatically in the presence of such noise [11]. Since noise is inherently present in any real data, this lack of robustness has limited the usefulness of these algorithms. In this paper, we obtain, indirectly, ML fractal dimension estimators for Gaussian $1/f$ processes that explicitly account for the presence of additive white Gaussian observation noise. The resulting iterative algorithms are computationally efficient, robust, and statistically consistent. Moreover, they retain many desirable properties in a range of non-Gaussian scenarios.

For the problem of estimating the underlying $1/f$ signal, we use a Bayesian framework to find signal estimates that minimize the mean-square estimation error. There are many potential problems involving signal enhancement and restoration to which these smoothing algorithms can be applied. Moreover, we note that these algorithms also make use of the signal and noise parameter estimates. In fact, it will become apparent in the ensuing development that the parameter and signal estimation components are actually quite closely coupled: smoothing is inherently involved in the parameter estimation process and vice versa.

In Sections II–IV, we briefly review $1/f$ processes, orthonormal wavelet bases, and prior work on wavelet-based expansions for $1/f$ processes. Sections V–VII then develop the use of this representation for parameter and signal estimation of $1/f$ processes embedded in white background noise. Finally, Section VIII contains some concluding remarks.

II. $1/f$ PROCESSES

The $1/f$ processes are generally defined [9] as processes whose empirical power spectra are of the form

$$S(\omega) \sim \frac{\sigma_x^2}{|\omega|^\gamma} \tag{1}$$

over several decades of frequency ω, where γ is some parameter in the range $0 < \gamma < 2$ and typically $\gamma \approx 1$. When this definition extends to all frequencies, (1) is no longer a valid power spectrum in the theory of stationary processes since such a spectrum is not integrable. Consequently, there have been a number of attempts to explain such spectra through nonstationary processes and notions of generalized spectra, e.g., [7], [13], [9], [16], [23]. For $1 \leq \gamma < 2$, the infinite-variance problem arises in the neighborhood of the spectral origin and is termed the infrared (IR) catastrophe. For many physical phenomena, measurements corresponding to very small frequencies show no low-frequency roll-off. In such cases, the underlying process is inherently nonstationary and (1) is interpreted as a generalized spectrum. Such is the case for Wiener processes ($\gamma \rightarrow 2$) with which we sometimes find it convenient to associate a $1/f^2$ spectrum for $f > 0$. For $0 < \gamma \leq 1$, the infinite-variance problem arises in the tails of the spectrum and is termed the ultraviolet (UV) catastrophe. It is avoided by reasoning that any physical process exhibits roll-off at sufficiently high frequencies. This is similar to the manner in which stationary white noise ($\gamma \rightarrow 0$) is interpreted in a physical setting.

Through a suitable interpretation of (1) as a generalized spectrum [7], it can be reasoned that if $x(t)$ is a $1/f$ process with parameters γ and σ_x^2, so is $a^{-H} x(at - b)$ for any $a > 0$ and any b, where

$$H = \frac{\gamma - 1}{2} \tag{2}$$

is the self-similarity parameter. In fact, this apparent statistical self-similarity relation is frequently used to define scale-invariant processes. Nevertheless, regardless of the chosen definition, the parameter γ, or some affine function of γ, turns out to be the essential parameter of interest. Indeed, via Mandelbrot's work we can express the so-called "fractal dimension" of sample functions of $1/f$ processes by

$$D = \frac{5 - \gamma}{2} \tag{3}$$

when $\gamma > 1$. Mandelbrot introduced the notion of fractal dimension to describe, in a rigorous and quantitative sense, the roughness of geometrical objects [15]. Subsequently, it has become an important tool in the classification of fractals. Consistent with this notion of fractal dimension, as γ is increased from 0 to 2, we find the sample functions of the corresponding $1/f$ processes to be increasingly smooth in appearance.

It is generally convenient to extend the notion of $1/f$ processes to include *nearly* $1/f$ processes that are defined [23] as having power spectra bounded according to

$$\frac{k_1}{|\omega|^\gamma} \leq S(\omega) \leq \frac{k_2}{|\omega|^\gamma} \tag{4}$$

where k_1 and k_2 satisfy $0 < k_1 \leq k_2 < \infty$ but are otherwise arbitrary. These processes exhibit only a constant
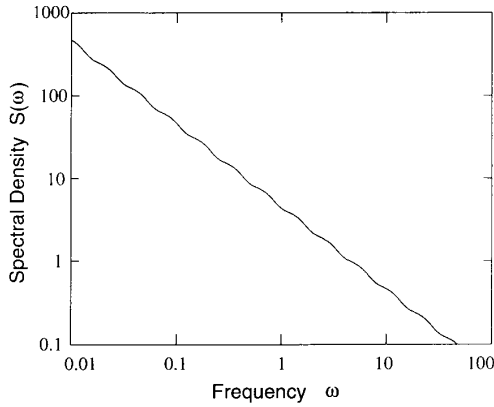
Fig. 1. The nearly $1/f$ spectrum of a typical dyadic fractal process. The parameters of the nearly $1/f$ spectrum are $\gamma = 1$ and $k_2/k_1 = 1.103$ in this case.

percentage deviation from the nominal $1/f$ characteristic, and, hence, effectively retain the characteristics of scale invariance and long-term dependence typically associated with $1/f$ phenomena.

There are many classes of such nearly $1/f$ processes. Examples include the stochastic processes that are statistically invariant only to changes of scale by factors of two. We have found it convenient to refer to these as dyadic fractal processes. Fig. 1 depicts the spectrum of a typical dyadic fractal process corresponding to $\gamma = 1$ and $k_2/k_1 = 1.103$. Note that for such processes the ripples superimposed on the $1/f$ spectrum are necessarily octave spaced.

III. ORTHONORMAL WAVELET BASES

In this section, we review some results from orthonormal wavelet theory that we require for subsequent developments in the paper. For a more complete discussion of orthonormal wavelets, see, for example, [12], [6].

An orthonormal wavelet transformation of a signal $x(t)$

$$x(t) \leftrightarrow x_n^m$$

is defined through the synthesis/analysis equations

$$x(t) = \sum_m \sum_n x_n^m \psi_n^m(t) \tag{5a}$$

$$x_n^m = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) \, dt \tag{5b}$$

and has the special property that all basis functions are dilations and translations of a single function:

$$\psi_n^m(t) = 2^{m/2} \psi(2^m t - n) \tag{6}$$

where m and n are the dilation and translation indices, respectively. Since the basic wavelet, $\psi(t)$, typically has an essentially bandpass Fourier transform, the wavelet transformation can often be interpreted in terms of a generalized constant-Q or octave-band filter bank.

It is also frequently possible to view the wavelet transformation in the context of a multiresolution analysis

framework [17], [12]. In our discussion, a resolution-limited approximation of a signal $x(t)$ in which details on scales smaller than 2^M are discarded is defined as[3]

$$A_M x(t) = \sum_{m < M} \sum_n x_n^m \psi_n^m(t). \tag{7}$$

In turn, for each M, $A_M x(t)$ has an orthonormal expansion of the form

$$A_M x(t) = \sum_n a_n^M \phi_n^M(t) \tag{8}$$

with the coefficients a_n^M obtained by projection:

$$a_n^M = \int_{-\infty}^{\infty} x(t) \phi_n^M(t) \, dt. \tag{9}$$

The basis functions $\phi_n^m(t)$ also have the property that they are dilations and translations of a single scaling function $\phi(t)$ having an essentially low-pass Fourier transform. From this perspective, we can view the mapping

$$D_m x(t) = \sum_n x_n^m \psi_n^m(t) \tag{10}$$

as the additional information or detail in going from a resolution-2^m approximation $A_m x(t)$ to a resolution-2^{m+1} approximation $A_{m+1} x(t)$. Accumulating this information over all scales m leads to the synthesis formula (5a).

As one example, consider a Haar-based multiresolution analysis in which the approximations are piecewise constant on intervals of length 2^m. Here the scaling function is given by

$$\phi(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

and the corresponding wavelet is

$$\psi(t) = \begin{cases} 1 & 0 < t < 1/2 \\ -1 & 1/2 \le t < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

This analysis exhibits excellent time localization but very poor frequency localization due to the abrupt time-domain transitions of the approximations. At the other extreme, consider the example of a sinc-based multiresolution analysis based upon band-limited approximations. In this case, the Fourier transform of the scaling function is

$$\Phi(\omega) = \begin{cases} 1 & |\omega| \le \pi \\ 0 & |\omega| > \pi \end{cases} \tag{13}$$

and corresponding wavelet is given by

$$\Psi(\omega) = \begin{cases} 1 & \pi \le |\omega| \le 2\pi \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

In contrast, this analysis exhibits excellent frequency localization but very poor time localization. More generally, we can consider the family of Battle–Lemarie mul-

[3]Note that, in our notation, m increasing corresponds to increasingly smaller (i.e., finer) scales of detail.

tiresolution analyses based upon orthogonalized Kth-order spline functions [12], [6]. In this context, our examples above correspond to the cases $K = 0$ and $K \rightarrow \infty$, respectively.

The development of other families of wavelet-based multiresolution analyses continues to receive considerable attention in the literature. As an example, Daubechies has designed a class of compactly supported wavelet bases [6]. In addition to fulfilling a practical requirement of having finite extent basis functions, these bases exhibit good localization in both time and frequency. Moreover, the basis functions are maximally regular, in the sense of having the maximum number of vanishing moments. Such regularity is considered desirable in a number of applications of wavelet bases (see, e.g., [3]).

The notion of an underlying multiresolution analysis leads rather conveniently to a discrete-time implementation of the wavelet-transformation in terms of conjugate quadrature filters.[4] In fact, it is from this perspective that wavelet transformations are associated with a class of multirate filter banks. Indeed, if we define the conjugate quadrature filter pair for a particular choice of wavelet basis by

$$h[n] = \int_{-\infty}^{\infty} \phi_0^{-1}(t) \phi_n^0(t) \, dt \qquad (15a)$$

$$g[n] = \int_{-\infty}^{\infty} \psi_0^{-1}(t) \phi_n^0(t) \, dt \qquad (15b)$$

the approximation coefficients a_n^m and detail coefficients x_n^m are related through the filter-downsample and upsample-filter relations

$$a_n^m = \sum_k h[k - 2n] a_k^{m+1} \qquad (16a)$$

$$x_n^m = \sum_k g[k - 2n] a_k^{m+1} \qquad (16b)$$

$$a_n^{m+1} = \sum_k \{h[n - 2k] a_k^m + g[n - 2k] x_k^m\}. \qquad (16c)$$

Hence, if the discrete-time observations of a process correspond to samples a_n^0 of a resolution-limited approximation to the discrete-time process, the analysis/synthesis equations (16) give a computationally efficient algorithm for extracting the wavelet coefficients at scales $m < 0$.

While the discrete-time nature of the observation limits access to the finer scales of detail, the length of the observation limits access to the coarser scales of detail. Hence, in practice we typically have access to wavelet coefficients over a finite range of scales for a given signal. Moreover, because the effective width of the wavelet basis functions halves at each finer scale, we expect roughly a doubling of the number of available coefficients at each successively finer scale consistent with what would be obtained through the discrete implementation (16). This finite-length, discrete implementation of the wavelet transform is termed the discrete wavelet transformation

(DWT), and has asymptotic complexity $\mathcal{O}(N \log_2 L)$ for data of length N and filters of length L [22]. Further, we note that while there are a number of ways to handle the unusual data windowing problem inherent in the decomposition, an assumption that the data is periodic outside the observation window leads to a computationally convenient implementation.

IV. WAVELET REPRESENTATIONS OF $1/f$ PROCESSES

In this section, we review the wavelet-based Karhunen-Loève-like expansions for $1/f$ processes described in [23], which have since been developed further in [25]. The work [23] specifically provides a synthesis result, i.e., that one can construct a class of nearly $1/f$ processes using wavelet expansions in terms of uncorrelated transform coefficients having the variance progression

$$\text{var } x_n^m = \sigma^2 2^{-\gamma m} \qquad (17)$$

where γ is the exponent of the nearly $1/f$ spectrum, and σ^2 is a positive constant proportional[5] to σ_x^2. In fact, the nearly $1/f$ processes obtained in this way are precisely examples of the dyadic fractal processes defined earlier. Moreover, the spectrum of Fig. 1 corresponds to a dyadic fractal process obtained by an orthonormal wavelet expansion using the first-order (i.e., $K = 1$) Battle-Lemarie wavelet basis.

It is somewhat remarkable that, at least empirically, there appears to be a corresponding analysis result, i.e., that wavelet expansions appear to be robust, nearly optimal representations for all $1/f$ processes, including exactly-$1/f$ processes. Specifically, for a reasonably arbitrary choice of wavelet, there is strong empirical evidence that the wavelet coefficients from these processes not only obey the variance progression (17), but turn out to be weakly correlated both along and across scales as well. Recent work [20], [25] suggests that this result can be made rigorous. Certainly, the wavelet transform is effective in removing strong, long-range dependence from the process. Consequently, in this and related work, we exploit the wavelet transformation's apparent role as a whitening filter for $1/f$ processes.[6] Ultimately, this renders signal processing problems involving $1/f$ processes considerably more tractable.

V. THE PARAMETER ESTIMATION PROBLEM

In this section, we derive ML parameter estimation algorithms for $1/f$ processes by exploiting the wavelet representation. Although, we specifically consider the case of Gaussian $1/f$ processes corrupted by additive stationary white Gaussian measurement noise, the resulting estimators are, in fact, applicable to a broader class of non-Gaussian $1/f$ processes and measurement noise models.

---

[4]In fact, this has become a convenient domain in which to design wavelet bases.

[5]The exact relationship depends on both the choice of wavelet and the absolute labeling of the scales, but is otherwise unimportant.

[6]It is interesting to note that infinite-interval whitening filters for $1/f$ processes have also been derived recently by Barton and Poor [2] using an approach based on reproducing kernel Hilbert spaces.

Let us 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

$$r(t) = x(t) + w(t), \quad -\infty < t < \infty. \quad (18)$$

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:

$$r_n^m = \int_{-\infty}^{\infty} \psi_n^m(t) r(t) \, dt. \quad (19)$$

However, in practice, these coefficients can be obtained by applying the computationally efficient DWT to the samples of a segment of data which is both time limited and resolution limited. Let us assume that the finite set of available distinct scales $\mathfrak{M}$ is, in increasing order,

$$\mathfrak{M} = \{m_1, m_2, \cdots, m_M\} \quad (20a)$$

and that at each scale m the set of available coefficients $\mathfrak{N}(m)$ is[7]

$$\mathfrak{N}(m) = \{n_1(m), n_2(m), \cdots, n_{N(m)}(m)\}. \quad (20b)$$

Hence, the data available to the estimation algorithm is

$$r = \{r_n^m \in \mathfrak{R}\} = \{r_n^m, m \in \mathfrak{M}, n \in \mathfrak{N}(m)\}. \quad (21)$$

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

$$\mathfrak{M} = \{1, 2, \cdots, M\} \quad (22a)$$

$$\mathfrak{N}(m) = \{1, 2, \cdots, N_0 2^{m-1}\} \quad (22b)$$

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

Exploiting the wavelet decomposition's role as a whitening filter for $1/f$ 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

$$r_n^m = x_n^m + w_n^m \quad (23)$$

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

$$\text{var } r_n^m = \sigma_m^2 = \sigma^2 \beta^{-m} + \sigma_w^2 \quad (24)$$

where we have defined

$$\beta = 2^\gamma \quad (25)$$

for future convenience. Hence, it is the parameter set

$$\Theta = (\beta, \sigma^2, \sigma_w^2) \quad (26)$$

we wish to estimate. As discussed in Section II, 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 will still need to be estimated jointly as they are rarely known *a priori*. Furthermore, ML estimates of γ, D, H are readily derived from the ML estimate $\hat{\beta}_{\text{ML}}$. Indeed, since each of these parameters is related to β through an invertible transformation, we have

$$\hat{\gamma}_{\text{ML}} = \log_2 \hat{\beta}_{\text{ML}} \quad (27a)$$

$$\hat{D}_{\text{ML}} = (5 - \hat{\gamma}_{\text{ML}})/2 \quad (27b)$$

$$\hat{H}_{\text{ML}} = (\hat{\gamma}_{\text{ML}} - 1)/2. \quad (27c)$$

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

$$\mathcal{L}(\Theta) = p_r(r; \Theta) = \prod_{m,n \in \mathfrak{R}} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(r_n^m)^2}{2\sigma_m^2}\right] \quad (28)$$

for which the log-likelihood function is

$$L(\Theta) = \ln p_r(r; \Theta) = -\frac{1}{2} \sum_{m,n \in \mathfrak{R}} \cdot \left\{\frac{1}{\sigma_m^2}(r_n^m)^2 + \ln(2\pi\sigma_m^2)\right\}. \quad (29)$$

Equivalently,

$$L(\Theta) = -\frac{1}{2} \sum_{m \in \mathfrak{M}} N(m) \left\{\frac{\hat{\sigma}_m^2}{\sigma_m^2} + \ln(2\pi\sigma_m^2)\right\} \quad (30)$$

where the M sample variances

$$\hat{\sigma}_m^2 = \frac{1}{N(m)} \sum_{n \in \mathfrak{N}(m)} (r_n^m)^2 \quad (31)$$

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

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

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

While we shall assume that β, σ^2, σ_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

---

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

[8]For example, if $m_1 < 0$, the we could define new parameters through the invertible transformation

$$\tilde{\sigma}_w^2 = \sigma_w^2$$

$$\tilde{\sigma}^2 = \sigma^2 \beta^{m_1 - 1}$$

$$\tilde{\beta} = \beta$$

for which the observations correspond to positive scales

$$\tilde{\mathfrak{M}} = \{1, m_2 - m_1 + 1, \cdots, m_M - m_1 + 1\}$$

and which lead to the same ML estimates for β, σ^2, σ_w^2.

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 $m \in \mathfrak{M}$ are positive without loss of generality.[8]

A. 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

$$\sum_{m \in \mathfrak{M}} T_m = 0 \tag{32a}$$

$$\sum_{m \in \mathfrak{M}} \beta^{-m} T_m = 0 \tag{32b}$$

$$\sum_{m \in \mathfrak{M}} m\beta^{-m} T_m = 0 \tag{32c}$$

where

$$T_m \triangleq \frac{N(m)}{\sigma_m^2} \left[1 - \frac{\hat{\sigma}_m^2}{\sigma_m^2} \right]. \tag{32d}$$

However, these equations are difficult to solve, except in special cases. Consequently, we utilize an estimate-maximize (EM) algorithm [10].

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

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

$$S_m^w(\hat{\Theta}^{(l)}) = A_m(\hat{\Theta}^{(l)}) + B_m^w(\hat{\Theta}^{(l)})\hat{\sigma}_m^2 \tag{33a}$$

$$S_m^x(\hat{\Theta}^{(l)}) = A_m(\hat{\Theta}^{(l)}) + B_m^x(\hat{\Theta}^{(l)})\hat{\sigma}_m^2 \tag{33b}$$

where

$$A_m(\hat{\Theta}^{(l)}) = \frac{\hat{\sigma}_w^{2(l)} \cdot \hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}}{\hat{\sigma}_w^{2(l)} + \hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}} \tag{34a}$$

$$B_m^w(\hat{\Theta}^{(l)}) = \left(\frac{\hat{\sigma}_w^{2(l)}}{\hat{\sigma}_w^{2(l)} + \hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}} \right)^2 \tag{34b}$$

$$B_m^x(\hat{\Theta}^{(l)}) = \left(\frac{\hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}}{\hat{\sigma}_w^{2(l)} + \hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}} \right)^2. \tag{34c}$$

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

$$\hat{\beta}^{(l+1)} \leftarrow \sum_{m \in \mathfrak{M}} C_m N(m) S_m^x(\hat{\Theta}^{(l)})\beta^m = 0 \tag{35a}$$

$$\hat{\sigma}^{2(l+1)} = \frac{\sum_{m \in \mathfrak{M}} N(m) S_m^x(\hat{\Theta}^{(l)}) [\hat{\beta}^{(l+1)}]^m}{\sum_{m \in \mathfrak{M}} N(m)} \tag{35b}$$

$$\hat{\sigma}_w^{2(l+1)} = \frac{\sum_{m \in \mathfrak{M}} N(m) S_m^w(\hat{\Theta}^{(l)})}{\sum_{m \in \mathfrak{M}} N(m)} \tag{35c}$$

where

$$C_m \triangleq \frac{m}{\sum_{m \in \mathfrak{M}} mN(m)} - \frac{1}{\sum_{m \in \mathfrak{M}} N(m)}. \tag{36}$$

B. 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 (35c) 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 $m \in \mathfrak{M}$ using current estimates of the parameters $\hat{\beta}^{(l)}$, $\hat{\sigma}^{2(l)}$:

$$S_m^x(\hat{\Theta}^{(l)}) = A_m(\hat{\Theta}^{(l)}) + B_m^x(\hat{\Theta}^{(l)})\hat{\sigma}_m^2 \tag{37}$$

where

$$A_m(\hat{\Theta}^{(l)}) = \frac{\sigma_w^2 \cdot \hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}}{\sigma_w^2 + \hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}} \tag{38a}$$

$$B_m^x(\hat{\Theta}^{(l)}) = \left(\frac{\hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}}{\hat{\sigma}_w^2 + \hat{\sigma}^{2(l)}[\hat{\beta}^{(l)}]^{-m}} \right)^2. \tag{38b}$$

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

$$\hat{\beta}^{(l+1)} \leftarrow \sum_{m \in \mathfrak{M}} C_m N(m) S_m^x(\hat{\Theta}^{(l)})\beta^m = 0 \tag{39a}$$

$$\hat{\sigma}^{2(l+1)} = \frac{\sum_{m \in \mathfrak{M}} N(m) S_m^x(\hat{\Theta}^{(l)}) [\hat{\beta}^{(l+1)}]^m}{\sum_{m \in \mathfrak{M}} N(m)} \tag{39b}$$

where C_m is as in (36).

C. 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:

$$\hat{\beta}_{ML} \leftarrow \sum_{m \in \mathfrak{M}} C_m N(m) \hat{\sigma}_m^2 \beta^m = 0 \tag{40a}$$

$$\sigma_{ML}^2 = \frac{\sum_{m \in \mathfrak{M}} N(m) \hat{\sigma}_m^2 [\hat{\beta}_{ML}]^m}{\sum_{m \in \mathfrak{M}} N(m)} \tag{40b}$$

with C_m still as in (36). Let us discuss 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 these latter estimates is essentially the same as the derivation of the M step in the Appendix. 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 (40a) directly. Having determined $\hat{\beta}_{ML}$ as the solution to this polynomial equation, $\hat{\sigma}^2_{ML}$ is obtained by back substitution.

From Lemma 1 in Appendix A, it is apparent that (40a) has exactly one positive real solution, which is the ML estimate $\hat{\beta}_{ML}$. Hence, L has a unique local and thus 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 $\hat{\beta}_{ML}$. Since we expect $0 < \gamma < 2$, an appropriate initial interval is $1 < \beta < 4$. Naturally, with some caution, Newton iterations may be used to accelerate convergence.

Again, since solving equations of the form of (40) constitutes the M step of the iterative algorithm for the more general problem, the above remarks are equally applicable in those contexts.

$$I = \sum_{m \in \mathfrak{M}} \frac{N(m)}{2(\sigma^2_m)^2} \begin{bmatrix} [\ln 2^m \sigma^2 \beta^{-m}]^2 & -\ln 2^m \sigma^2 [\beta^{-m}]^2 & -\ln 2^m \sigma^2 \beta^{-m} \\ -\ln 2^m \sigma^2 [\beta^{-m}]^2 & [\beta^{-m}]^2 & \beta^{-m} \\ -\ln 2^m \sigma^2 \beta^{-m} & \beta^{-m} & 1 \end{bmatrix} \tag{43}$$

D. Properties of the Estimators

In this section, we consider two principal issues: 1) how the parameter estimates of the EM algorithm converge to the ML parameter estimates; and 2) 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:

$$\hat{\beta} = \hat{\beta}_{ML}\big|_{\sigma^2_w = 0} \tag{41a}$$

$$\hat{\sigma} = \hat{\sigma}^2_{ML}\big|_{\sigma^2_w = 0} \tag{41b}$$

$$\hat{\sigma}^2_w = 0 \tag{41c}$$

and

$$\hat{\beta}: \quad \text{arbitrary} \tag{42a}$$

$$\hat{\sigma}^2 = 0 \tag{42b}$$

$$\hat{\sigma}^2_w = \frac{\sum\limits_{m \in \mathfrak{M}} N(m) \hat{\sigma}^2_m}{\sum\limits_{m \in \mathfrak{M}} N(m)}. \tag{42c}$$

That they are saddle points is rather fortunate, for the only way they are reached is if the starting value for any one of $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{\sigma}^2_w$ is chosen to be exactly zero. Given arbitrarily small positive choices for these initial parameters, the algorithms will iterate 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 will naturally affect 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^2_w is known. In essence, for the general algorithm much of the iteration is spent locating the noise threshold in the data.

We now turn to a discussion of the properties of the ML estimate themselves. It is well known that ML estimates are generally asymptotically efficient and consistent. This, specifically, turns out to be the case here. 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.

The compute the Cramér–Rao bounds corresponding to the estimates of γ, σ^2, and σ^2_w, we construct the Fisher matrix[9]

from which we get

$$\text{var } \hat{\gamma} \geq I^{11} \tag{44a}$$

$$\text{var } \hat{\sigma}^2 \geq I^{22} \tag{44b}$$

$$\text{var } \hat{\sigma}^2_w \geq I^{33} \tag{44c}$$

for any unbiased estimates $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\sigma}^2_w$, and where I^{kk} is the kth element on the diagonal of 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^2_w is unknown, the Fisher information matrix simplifies to the upper submatrix

$$I = \sum_{m \in \mathfrak{M}} \frac{N(m)[\beta^{-m}]^2}{2(\sigma^2_m)^2} \begin{bmatrix} [\ln 2^m \sigma^2]^2 & -\ln 2^m \sigma^2 \\ -\ln 2^m \sigma^2 & 1 \end{bmatrix} \tag{45}$$

from which we get

$$\text{var } \hat{\gamma} \geq I^{11} \tag{46a}$$

$$\text{var } \hat{\sigma}^2 \geq I^{22}. \tag{46b}$$

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.

[9]In examining the variance of our estimates, we revert back to considering γ rather than β, recalling that it is usually some affine function of γ rather than β that is actually of interest.

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

$$I = \begin{bmatrix} (\ln 2)^2/2 \sum_{m \in \mathfrak{M}} m^2 N(m) & -(\ln 2)/(2\sigma^2) \sum_{m \in \mathfrak{M}} mN(m) \\ -(\ln 2)/(2\sigma^2) \sum_{m \in \mathfrak{M}} mN(m) & 1/(2\sigma^4) \sum_{m \in \mathfrak{M}} N(m) \end{bmatrix} \quad (47)$$

from which we get

$$\text{var } \hat{\gamma} \geq 2/[(\ln 2)^2 J] \sum_{m \in \mathfrak{M}} N(m) \quad (48a)$$

$$\text{var } (\hat{\sigma}^2/\sigma^2) \geq 2/J \sum_{m \in \mathfrak{M}} m^2 N(m) \quad (48b)$$

where

$$J = \left[\sum_{m \in \mathfrak{M}} m^2 N(m) \right] \left[\sum_{m \in \mathfrak{M}} N(m) \right] - \left[\sum_{m \in \mathfrak{M}} mN(m) \right]^2. \quad (49)$$

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 (22), we get

$$\text{var } \hat{\gamma}_{\text{ML}} \sim 2/[(\ln 2)^2 N] \quad (50a)$$

$$\text{var } (\hat{\sigma}_M^2/\sigma^2) \sim 2 (\log_2 N)^2)/N \quad (50b)$$

where N is the number of observation samples.

VI. THE SIGNAL ESTIMATION PROBLEM

In this section, we derive an algorithm for the optimal estimation of a $1/f$ signal embedded in additive stationary white noise. For the case of Gaussian processes, these estimators are optimal in the sense that they minimize the mean-square estimation error. As is well known from classical estimation theory, for non-Gaussian scenarios the estimates are the best linear estimates.

While we do not specifically derive our 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 will dominate, while at low frequencies the $1/f$ signal spectrum will dominate.[10] Consequently, Wiener filtering for this problem involves a form of low-pass filtering, where the exact filter shape and "cutoff" are governed by the particular parameters of the noise and signal spectra.

Turning now to our specific statement of the problem, let us consider the estimation of a $1/f$ signal $x(t)$ from noisy observations $r(t)$ given by (18), where we still consider zero-mean processes. In our derivation, we shall assume that the signal and noise parameters β, σ^2, σ_w^2 are all known, though in practice they are estimated using the parameter algorithms of the last section. In fact, as we

[10] In fact, at sufficiently low frequencies, there will always be arbitrarily high SNR regardless of the noise threshold.

shall see, smoothing is inherently involved in the parameter estimation algorithm we derived earlier.

Again we exploit the wavelet decomposition to obtain our results. Specifically, we begin with the set of wavelet coefficients (21). Then, since

$$r_n^m = x_n^m + w_n^m \quad (51)$$

where x_n^m and w_n^m are 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

$$\hat{x}_n^m = E[x_n^m | r] = \begin{cases} E[x_n^m | r_n^m] & m, n \in \mathfrak{R} \\ 0 & \text{otherwise.} \end{cases} \quad (52)$$

Furthermore, since 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

$$E[x_n^m | r_n^m] = \left[\frac{\sigma^2 \beta^{-m}}{\sigma^2 \beta^{-m} + \sigma_w^2} \right] r_n^m. \quad (53)$$

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

$$\hat{x}(t) = \sum_{m,n} \hat{x}_n^m \psi_n^m(t) = \sum_{m,n \in \mathfrak{R}} \left[\frac{\sigma^2 \beta^{-m}}{\sigma^2 \beta^{-m} + \sigma_w^2} \right] r_n^m \psi_n^m(t). \quad (54)$$

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 (54) 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 (34c), which allows us to interpret (33b) in terms of sample-variance estimates of the smoothed data. Evidently, smoothing is inherently involved in the parameter estimation problem.

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 $1/f$ 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 mean-square estimation error

$$E[(\hat{x}(t) - x(t))^2]$$

is infinite. Nevertheless, when we define

$$\tilde{x}(t) = \sum_{m,n \in \mathfrak{R}} x_n^m \psi_n^m(t) \quad (55)$$

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 $\tilde{x}(t)$ as

$$\epsilon = \int_{-\infty}^{\infty} E[(\hat{x}(t) - \tilde{x}(t))^2] \, dt = \sum_{m,n \in \Re} E[(\hat{x}_n^m - x_n^m)^2]$$

$$= \sum_{m,n \in \Re} E[\text{var } (x_n^m | r_n^m)] \qquad (56)$$

which, through routine manipulation, reduces to

$$\epsilon = \sum_{m \in \mathfrak{M}} N(m) \left[\frac{\sigma^2 \beta^{-m} \cdot \sigma_w^2}{\sigma^2 \beta^{-m} + \sigma_w^2} \right]. \qquad (57)$$

VII. SIMULATIONS

In this section, we present some examples demonstrating the viability and basic functionality of the derived algorithms. Although a number of properties of the algorithms are suggested by Monte Carlo simulations, we are not suggesting that they are in any way comprehensive. The work of this section represents only a very preliminary investigation of these algorithms.

There are a variety of simulations and simulation scenarios that could be considered for this work. For our basic scenario, we have chosen to synthesize discrete samples of resolution-limited Gaussian $1/f$ processes embedded in Gaussian white 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, followed by signal estimation, using algorithms for the most general scenario, corresponding to the case in which all signal and noise parameters β, σ^2, σ_w^2 are unknown. Note that by using the estimated parameters in the signal estimation algorithm, our 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.[11]

There are a number of methods for simulating $1/f$ processes available in the literature. Indeed, the wavelet-based expansion for $1/f$ processes in [23] suggests one more method for synthesizing $1/f$ processes. However, for the simulations of this section, we construct Gaussian $1/f$ processes using the discrete-time implementation of Keshner's model for $1/f$ processes described in [5]. By choosing a synthesis algorithm fundamentally different from wavelet-based synthesis, modeling sensitivity and robustness issues are not bypassed in the simulations with our inherently wavelet-based algorithms. However, as a result, our experiments do not allow us to distinguish between errors inherent in the modeling and errors inherent in the estimation process.

Finally, there are a large number of wavelet bases from which to select for our analysis. However, given the empirical insensitivity of the algorithms to the choice of basis, for our simulations we somewhat arbitrarily chose to

---

[11]However, it turns out that the quality of the signal estimation is rather insensitive to errors in the parameter estimates used.
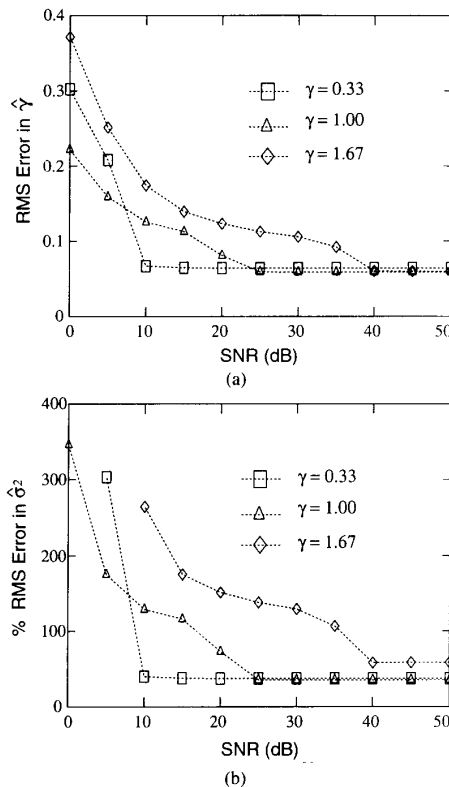

Fig. 2. Rms errors in the estimates of the signal parameters as a function of the SNR of the observations. The data length of the observations is fixed to $N = 2048$ samples. (The symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only.) (a) absolute rms error in $\hat{\gamma}_{ML}$. (b) Percentage rms error in $\hat{\sigma}_{ML}^2$.

use Daubechies fifth-order finite-extent "maximally regular" wavelet basis for which the corresponding conjugate quadrature filters have 10 nonzero coefficients.

The first simulations demonstrate some aspects of the performance of the parameter estimation algorithms. In Fig. 2, we plot the rms error of the estimates of γ and σ^2 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 bounds would appear to be attained at increasing SNR thresholds.

In Fig. 3, we plot the rms error of the estimates of γ and σ^2 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
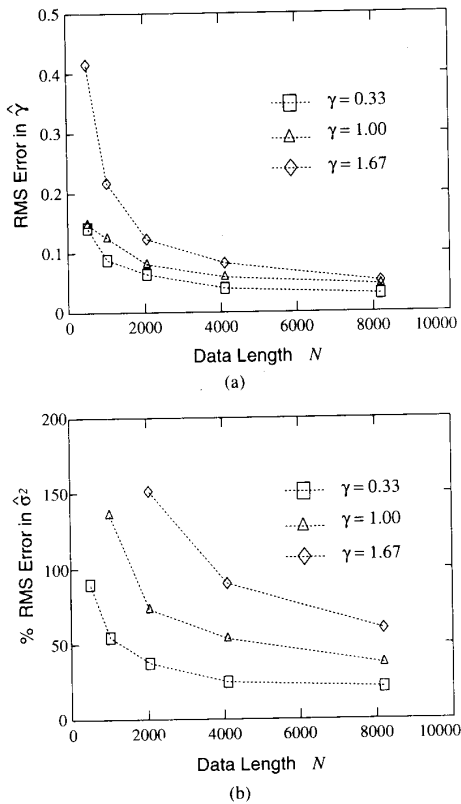
Fig. 3. Rms errors in the estimates of the signal parameters as a function of the data length N of the observations. The SNR of the observations is fixed to 20 dB. (Again, the symbols associated with each γ mark the actual empirical measurements; dashed lines are provided as visual aides only.) (a) Absolute rms error in $\hat{\gamma}_{ML}$. (b) Percentage rms error in $\hat{\sigma}^2_{ML}$.

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.

The second set of simulations demonstrate some aspects of the performance of the signal estimation. In Fig. 4, we plot the SNR gain of our smoothed signal estimates 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 mean-square error formula (57) are also superimposed on each plot. As the results indicated, 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 pa-
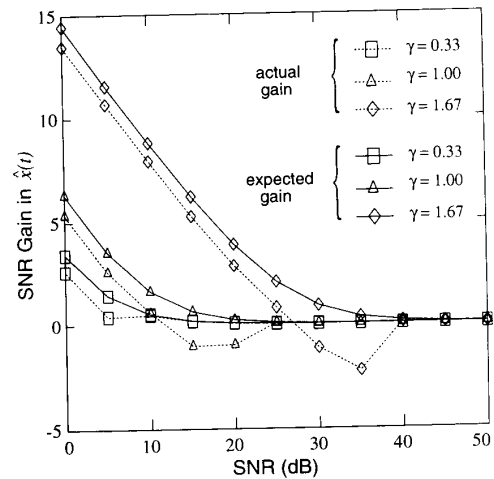

Fig. 4. SNR gain (decibels) of the signal estimate as a function of the SNR of the observations. The data length of the observations is $N = 2048$ samples. Both the gains predicted by (57) and gains actually obtained are indicated.
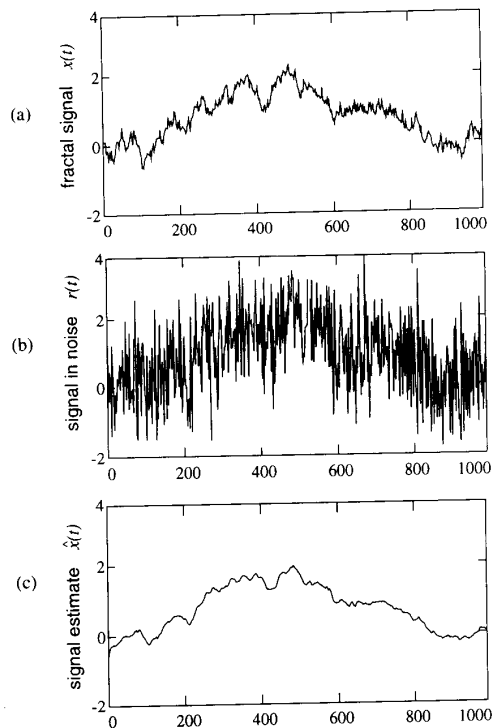

Fig. 5. A Smoothing example. A 65 536-sample $1/f$ signal with $\gamma = 1.67$ is embedded in noise corresponding to 0 dB SNR. An estimate of the signal is generated from the noisy data using parameter estimates obtained from the data; specifically, $\hat{\gamma}_{ML} = 1.66$. The SNR gain in the estimate is 13.9 dB. (a) A 1000 sample segment of the $1/f$ signal. (b) The corresponding segment of the same signal embedded in noise. (c) The corresponding segment of the smoothed estimate of the signal.

rameter estimates is a much less significant factor in these degradations. The plots also indicate that better gains are achieved for larger values of γ for a given SNR. This is

to be expected since for larger vaues of γ there is more signal power at coarser scales and correspondingly less at finer scales where the noise predominates and the most attenuation takes place.

Finally, in Fig. 5 we show 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 of 0 dB. The estimated spectral exponent is $\hat{\gamma}_{ML} = 1.66$, 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 on fine scales is lost where the SNR was low.

VIII. CONCLUSIONS

In this work we derive computationally efficient and robust wavelet-based algorithms for performing ML parameter estimation and minimum mean-square error smoothing of $1/f$ processes embedded in additive white observation noise. Moreover, somewhat indirectly, our algorithms allow us to compute ML estimates of the fractal dimension of $1/f$ signals from noisy observations, which are of considerable interest in their own right for many applications.

While a considerable amount of additional analysis and testing of these algorithms remains to be done, the preliminary study presented here suggests they are highly practicable and widely applicable in a number of signal processing applications.

A number of outstanding issues associated with this work remain to be addressed. One relates to the wavelet-based representation of $1/f$ processes. In particular, we have relied on strong empirical evidence that there exists a matching analysis result corresponding to the synthesis result of [23]. However, a strong theoretical justification that the orthonormal wavelet decomposition of $1/f$ processes yields virtually independent transform coefficients would be an important result.

Another issue pertains to the unusual data-windowing problem inherent in the wavelet decomposition. In this work, we avoid the problem by modeling the data as periodic outside the finite observation interval in computing the DWT. However, this leads to a number of rather undesirable effects, some of which manifested themselves in the smoothing simulations. More effective approaches to accommodating observations on the finite interval need to be developed.

Finally, a number of interesting, straightforward and useful extensions to this work are suggested by the approaches described here. Specifically, the problem of distinguishing and isolating two superimposed fractal signals is, in principle, readily solved by the methods of this work. In addition, the separable extension of the results presented herein to two and higher dimensions is likewise straightforward. In each case, we anticipate that a number of powerful yet practical algorithms can be developed.

APPENDIX
AN EM ALGORITHM

In order to derive an EM algorithm for estimation of $\Theta = (\beta, \sigma^2, \sigma_w^2)$, we first define our observed (incomplete) data to be

$$r = \{r_n^m, m, n \in \mathfrak{R}\} \tag{58}$$

and our complete data to be (x, r) where

$$x = \{x_n^m, m, n \in \mathfrak{R}\}. \tag{59}$$

Consequently, the EM algorithm for the problem is defined as [10]

E step: Compute

$$U(\Theta, \hat{\Theta}^{(l)}) \tag{60}$$

M step:

$$\max_{\Theta} U(\Theta, \hat{\Theta}^{(l)}) \rightarrow \hat{\Theta}^{(l+1)} \tag{61}$$

where

$$U(\Theta, \tilde{\Theta}) \triangleq E[\ln p_{r,x}(r, x; \Theta)|r; \tilde{\Theta}]. \tag{62}$$

For our case, U is obtained conveniently via

$$U(\Theta, \tilde{\Theta}) = E[\ln p_{r|x}(r|x; \Theta) + \ln p_x(x; \Theta)|r; \tilde{\Theta}] \tag{63}$$

with

$$p_{r|x}(r|x; \Theta) = \prod_{m,n \in \mathfrak{R}} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp - \frac{(r_n^m - x_n^m)^2}{2\sigma_w^2} \tag{64}$$

and

$$p_x(x; \Theta) = \prod_{m,n \in \mathfrak{R}} \frac{1}{\sqrt{2\pi\sigma^2\beta^{-m}}} \exp - \frac{(x_n^m)^2}{2\sigma^2\beta^{-m}}. \tag{65}$$

Then

$$U(\Theta, \tilde{\Theta}) = -\frac{1}{2} \sum_{m \in \mathfrak{M}} N(m) \left\{ \frac{1}{\sigma_w^2} S_m^w(\tilde{\Theta}) + \ln 2\pi\sigma_w^2 \right. $$
$$\left. + \frac{1}{\sigma^2\beta^{-m}} S_m^x(\tilde{\Theta}) + \ln 2\pi\sigma^2\beta^{-m} \right\} \tag{66}$$

where

$$S_m^w(\Theta) = \frac{1}{N(m)} \sum_{n \in \mathfrak{N}(m)} E[(w_n^m)^2|r_n^m; \Theta] \tag{67a}$$

$$S_m^x(\Theta) = \frac{1}{N(m)} \sum_{n \in \mathfrak{N}(m)} E[(x_n^m)^2|r_n^m; \Theta] \tag{67b}$$

are (quasi) conditional sample-variance estimates from the data based upon the model parameters Θ. Evaluating the expectations we get

$$S_m^w(\Theta) = A_m(\Theta) + B_m^w(\Theta)\hat{\sigma}_m^2 \tag{68a}$$

$$S_m^x(\Theta) = A_m(\Theta) + B_m^x(\Theta)\hat{\sigma}_m^2 \tag{68b}$$

where

$$A_m(\Theta) = \frac{\sigma_w^2 \cdot \sigma^2\beta^{-m}}{\sigma_w^2 + \sigma^2\beta^{-m}} \tag{69a}$$

$$B_m^w(\Theta) = \Sigma \left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma^2 \beta^{-m}} \right)^2 \qquad (69b)$$

$$B_m^x(\Theta) = \left(\frac{\sigma^2 \beta^{-m}}{\sigma_w^2 + \sigma^2 \beta^{-m}} \right)^2 \qquad (69c)$$

which completes our derivation of the E step.

To derive the structure of the M step, we maximize $U(\Theta, \tilde{\Theta})$ as given by (66). This maximization is always well defined as $U(\Theta, \tilde{\Theta}) \leq L(\Theta)$ for any $\Theta, \tilde{\Theta}$.

The local extrema are obtained by differentiating $U(\Theta, \tilde{\Theta})$ with respect to each of the parameters of Θ. Since (66) expresses $U(\Theta, \tilde{\Theta})$ as the sum of two terms, one of which depends only on σ_w^2 and the other of which depends only on β and σ^2, the maximization can be broken down into two independent parts.

Considering first our maximization over σ_w^2, we readily obtain the maximizing $\hat{\sigma}_w^2$ as the sample average

$$\hat{\sigma}_w^2 = \frac{\sum\limits_{m \in \mathfrak{M}} N(m) S_m^w(\tilde{\Theta})}{\sum\limits_{m \in \mathfrak{M}} N(m)}. \qquad (70)$$

Turning next to β and σ^2, we find that the maximizing parameters $\hat{\beta}$ and $\hat{\sigma}^2$ satisfy

$$\sum\limits_{m \in \mathfrak{M}} N(m) S_m^x(\tilde{\Theta}) \beta^m = \sigma^2 \sum\limits_{m \in \mathfrak{M}} N(m) \qquad (71a)$$

$$\sum\limits_{m \in \mathfrak{M}} m N(m) S_m^x(\tilde{\Theta}) \beta^m = \sigma^2 \sum\limits_{m \in \mathfrak{M}} m N(m). \qquad (71b)$$

Eliminating σ^2 we obtain that $\hat{\beta}$ is the solution of the polynomial equation

$$\sum\limits_{m \in \mathfrak{M}} C_m N(m) S_m^x(\tilde{\Theta}) \beta^m = 0 \qquad (72)$$

where C_m is as defined in (36). The eliminated variable $\hat{\sigma}^2$ is trivially obtained by back-substitution:

$$\hat{\sigma}^2 = \frac{\sum\limits_{m \in \mathfrak{M}} N(m) S_m^x(\tilde{\Theta}) \hat{\beta}^m}{\sum\limits_{m \in \mathfrak{M}} N(m)}. \qquad (73)$$

Finally, to show that the maximizing parameters are the only solution to (71) it suffices to show that the solution to (72) is unique, which we establish via the following lemma.

Lemma 1: Any polynomial equation of the form

$$\sum\limits_{m \in \mathfrak{M}} C_m K_m \beta^m = 0 \qquad (74)$$

where C_m is given by (36) and $K_m \geq 0$ has a unique positive real solution provided $M \geq 2$ and not all K_m are zero.

Proof: Let

$$m_* = \frac{\sum\limits_{m \in \mathfrak{M}} m N(m)}{\sum\limits_{m \in \mathfrak{M}} N(m)} \qquad (75)$$

be a weighted average of the $m \in \mathfrak{M}$, so $m_1 < m_* < m_M$.

Then, from (36), for $m > m_*$, $C_m > 0$, while for $m < m_*$, $C_m < 0$. Hence, $C_m(m - m_*) \geq 0$ with strict inequality for at least two values of $m \in \mathfrak{M}$ from our hypothesis. Now let $f(\beta)$ be the left-hand side of (74), and observe that

$$\tilde{f}(\beta) \triangleq f(\beta) \beta^{-m_*} \qquad (76)$$

is increasing for $\beta > 0$, i.e.,

$$\tilde{f}'(\beta) = \sum\limits_{m \in \mathfrak{M}} C_m(m - m_*) N(m) \hat{\sigma}_m^2 \beta^{m - m_* - 1} > 0. \qquad (77)$$

Then, since $\tilde{f}(0) = -\infty$ and $\tilde{f}(\infty) = \infty$, we see $\tilde{f}(\beta)$ has a single real root on $\beta > 0$. Since $f(\beta)$ shares the same roots on $\beta > 0$, we have the desired result. ∎

This completes our derivation for the M step. The complete algorithm follows directly.

REFERENCES

[1] J. A. Barnes and D. W. Allan, "A statistical model of flicker noise," *Proc. IEEE*, vol. 54, pp. 176-178, Feb. 1966.
[2] R. J. Barton and V. H. Poor, "Signal detection in fractional Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 943-959, Sept. 1988.
[3] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms," *Commun. Pure Appl. Math.*, to be published.
[4] K. C. Chou, A. S. Willsky, A. Benveniste, and M. Basseville, "Recursive and iterative estimation algorithms for multiresolution stochastic processes," in *Proc. 28th IEEE Conf. Decision Contr.* (Tampa, FL), Dec. 1989.
[5] G. Corsini and R. Saletti, "Design of a digital 1/f^{α} noise simulator," in *Proc. Noise Physical Syst. and 1/f Noise*, Université de Montréal, Canada, 1987, pp. 82-86.
[6] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909-996, Nov. 1988.
[7] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 197-199, Jan. 1989.
[8] R. Fox and M. S. Taqqu, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Ann. Stat.*, vol. 14, no. 2, pp. 517-532, 1986.
[9] M. S. Keshner, "1/f noise," *Proc. IEEE*, vol. 70, pp. 212-218, Mar. 1982.
[10] N. M. Laird, A. P. Dempster, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Ann. Roy. Stat. Soc.*, pp. 1-38, Dec. 1977.
[11] T. Lundahl, W. J. Ohley, S. M. Kay, and R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. Med. Imaging*, vol. MI-5, pp. 152-161, Sept. 1986.
[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, pp. 674-693, July 1989.
[13] B. Mandelbrot, "Some noises with 1/f spectrum, a bridge between direct current and white noise," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 289-298, Apr. 1967.
[14] B. B. Mandelbrot, "Self-similar error clusters in communication systems and the concept of conditional stationarity," *IEEE Trans. Commun. Technol.*, vol. COM-13, pp. 71-90, Mar. 1965.
[15] B. B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco, CA: Freeman, 1982.
[16] B. B. Mandelbrot and H. W. Van Ness, "Fractional Brownian motions, fractional noises, and applications," *SIAM Rev.*, vol. 10, p. 422-436, Oct. 1968.

[17] Y. Meyer, "Ondelettes et functions splines," Séminaire EDP, Ecole Polytechnique, Paris, France, Dec. 1986.

[18] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-6, pp. 661-674, Nov. 1984.

[19] M. S. Taqqu, "A bibliographic guide to self-similar processes and long-range dependence," in *Dependence in Probability and Statistics*, E. Eberlein and M. S. Taqqu, Eds. Boston, MA: Birkhauser, 1986.

[20] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motions," preprint, Dep. Elec. Eng., Univ. Minnesota, Oct. 1990.

[21] A. van der Ziel, "On the noise spectra of semiconductor noise and of flicker effect," *Physica*, vol. 16, no. 4, pp. 359-372, 1950.

[22] M. Vetterli and C. Herley, "Wavelets and filter banks: Relationships and new results," in *Proc. ICASSP* (Albuquerque, NM), 1990.

[23] G. W. Wornell, "A Karhunen-Loève-like expansion for $1/f$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, pp. 859-861, July 1990.

[24] Y. Yajima, "On estimation of long-memory time series models," *Austral. J. Stat.*, vol. 27, no. 3, pp. 321-325, 1985.

[25] G. W. Wornell, "Synthesis, analysis, and processing of fractal signals," RLE Tech. Rep. 566, M.I.T., Cambridge, MA, Oct. 1991.

Gregory W. Wornell (S'83-M'85) was born in Montréal, Canada, in 1962. He received the B.A.Sc. degree (with honors) from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering, in 1985, 1987, and 1991, respectively.

Since 1991 he has been on the faculty at M.I.T., where he is currently Assistant Professor in the Department of Electrical Engineering and Computer Science. His current research interests include applications of fractal geometry and nonlinear dynamical system theory in signal processing and communications. He has held several research staff appointments with MacDonald Dettwiler and Associates, Canada, between 1985 and 1989, and he was a Visiting Investigator at the Woods Hole Oceanographic Institution, Woods Hole, MA, during 1990. He is currently coinventor on a pending patent pertaining to the processing of remotely sensed imagery. From 1985 to 1989 he held a 1967 Science and Engineering Scholarship from the Natural Sciences and Engineering Research Council of Canada.

Dr. Wornell received the M.I.T. Goodwin Medal in 1991 for "conspicuously effective teaching." He is a member of Tau Beta Pi and Sigma Xi.
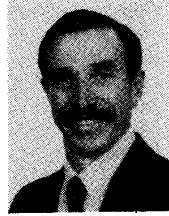
Alan V. Oppenheim (S'57-M'65-SM'71-F'77) received the S.B. and S. M. degrees in 1961 and the Sc.D. degree in 1964, all in electrical engineering, from the Massachusetts Institute of Technology, Cambridge.

In 1964 he joined the faculty at M.I.T. where he is currently Distinguished Professor of Electrical Engineering. From 1978 to 1980 he was Associate Head of the Data Systems Division at M.I.T. Lincoln Laboratory. Since 1977, he has also been a Guest Investigator at the Woods Hole Oceanographic Institution, Woods Hole, MA. He is the author of several widely used textbooks on digital signal processing.

Dr. Oppenheim has been a Guggenheim Fellow, a Sackler Fellow, and has held the Cecil H. Green Distinguished Chair in Electrical Engineering and Computer Science. He has also received a number of awards for outstanding research and teaching, including the 1988 IEEE Education Medal, and is an elected member of the National Academy of Engineering. He is also a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.