PREDICTION AND ESTIMATION FOR FRACTAL PROCESSES USING MULTISCALE STATE-SPACE ALGORITHMS

Alexander C. Wang and Gregory W. Wornell

Research Laboratory of Electronics MIT, Cambridge, MA 02139

ABSTRACT

The 1/f family of fractal processes provides useful models for the extraordinary variety of natural and man-made phenomena that exhibit long-term dependence. Using algorithms based on a multiscale state-space representation, we address the problems of parameter estimation of discrete 1/f signals in white noise, estimation of deterministic signals in 1/f noise, and prediction of discrete 1/f processes. Among other results, distant past data are shown to have a dramatically greater effect on these estimators than when ARMA processes are involved.

1. INTRODUCTION

The 1/f processes are empirically defined as having measured power spectral density of the form

$$S_s(\omega) \sim \frac{\sigma_s^2}{|\omega|^{\gamma}}$$

over several decades of frequency ω , where γ is a parameter in the range $0 < \gamma < 2$. As opposed to the traditional autoregressive moving-average (ARMA) models characterized by correlation functions with exponential decay, 1/f processes exhibit long-term dependence characterized by correlation functions with polynomial-type decay. As a result, these processes provide useful models for the extraordinary variety of natural and man-made phenomena that exhibit long-term dependence. A more general class of processes, called nearly 1/f, has measured power spectral density that is bounded according to

$$\frac{\sigma_L^2}{|\omega|^{\gamma}} \le S_s(\omega) \le \frac{\sigma_U^2}{|\omega|^{\gamma}}$$

where σ_L^2 and σ_U^2 satisfy $0 < \sigma_L^2 \leq \sigma_U^2 < \infty$. The data modeled as a 1/f process is generally represented as a discrete sequence. The discretization of the time axis limits the highest frequency at which 1/f spectral behavior can be observed. Therefore, of primary interest is the spectral behavior at low frequencies, which govern the longterm dependence of these processes. We define a discrete 1/f (or nearly-1/f) process as having 1/f (or nearly-1/f) spectral behavior in the neighborhood of the spectral origin. An example of a discrete 1/f process is the discrete fractionally differenced Gaussian noise (fdGn) process [5][6].

The power spectrum of a discrete or continuous 1/fprocess is not integrable in the neighborhood of the origin for $\gamma \geq 1$. This phenomenon, called the infrared catastrophe, has been interpreted as revealing that the process as inherently nonstationary [7][8]. In this paper, we assume that the 1/f process is stationary with the shape of the power spectrum changing from 1/f to flat below a certain frequency, although this low-frequency roll-off is not always observed in natural signals (see [7] and the references cited).

The need for efficient and robust signal processing algorithms involving fractals arises in many engineering contexts [2][4][11]. This paper develops a multiscale state-space representation for finite-length 1/f processes that is particularly well-suited for addressing several signal processing problems involving finite data lengths, such as prediction and signal estimation.

2. MULTISCALE STATE-SPACE REPRESENTATION

Van der Ziel [9] modeled continuous 1/f processes as the weighted superposition of a continuum of uncorrelated random processes. These models form the basis for the analogous discrete-time models used in this paper. Let s[n] be the superposition of uncorrelated first-order (single timeconstant) autoregressive processes

$$s[n] = \sum_{m=\underline{m}}^{\overline{m}} x_m[n]$$

where $x_m[n]$ has correlation function $R_m[k] = f_m \beta_m^{|k|}$. The spectrum of s[n] is the superposition of the spectra of the autoregressive processses:

$$S_s(\Omega) = \sum_{m=\underline{m}}^{\overline{m}} \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m \cos\Omega}.$$

When weights f_m and time-constants β_m are given by

$$\beta_m = \left(\frac{2}{\Delta^m + \sqrt{\Delta^{2m} + 4}}\right)^2$$
$$f_m = \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\beta_m^{-1} - \beta_m}$$

where Δ , σ^2 , and γ are parameters satisfying $1 < \Delta < \infty$, $\sigma^2 > 0$, and $0 < \gamma < 2$, then for $\underline{m} \to -\infty$ and $\overline{m} \to \infty$, $S_s(\Omega)$ has nearly-1/f spectral behavior, with parameter γ ,

This work has been supported in part by DARPA monitored by ONR under Contract No. N00014-93-1-0686, AFOSR under Grant No. F49620-96-1-0072, and an NSF Graduate Fellowship.

in some neighborhood of the spectral origin $0 < |\Omega| < \epsilon$. Therefore s[n] is a discrete nearly-1/f process.

The described models would require an infinite number of state variables to completely describe the discrete 1/fprocess over all frequencies. An appropriately selected *fi*nite subset of the infinite component processes is sufficient to generate nearly 1/f spectral behavior over a finite frequency range. We select \overline{m} so that component processes $\{x_m[n]\}, m > \overline{m}$ are effectively white, and replace these processes by a single white component process. For finitelength processes, the data length effectively constrains the lowest observable frequency of the power spectrum. We select <u>m</u> so that component processes $\{x_m[n]\}, m < \underline{m}$ have insignificant total power above this frequency. Discarding these processes creates a low-frequency roll-off where the spectrum is flat rather than 1/f below a certain frequency. When data length is increased by a factor of k, the lowest observable frequency decreases, and $\frac{2}{2-\gamma} \log_{\Delta} k$ additional component processes are required to generate nearly-1/f spectral behavior above this new frequency.

The state-space description for a measurement process z[n], composed of a finite-order 1/f process corrupted by additive white measurement noise w[n], is defined by

$$\begin{aligned} \mathbf{x}[n+1] &= A\mathbf{x}[n] + B\mathbf{u}[n] \\ z[n] &= C\mathbf{x}[n] + w[n] \end{aligned}$$

where $\mathbf{u}[n]$ is a $(\overline{m} - \underline{m} + 1)$ -dimensional driving vector of uncorrelated, zero-mean white Gaussian processes with unit variance, and w[n] is a zero-mean white Gaussian noise process with variance σ_w^2 . Block-diagonal matrices A and B

$$A = \begin{bmatrix} A_{\underline{m}} & 0 \\ & \ddots & \\ 0 & A_{\overline{m}} \end{bmatrix} \quad B = \begin{bmatrix} B_{\underline{m}} & 0 \\ & \ddots & \\ 0 & B_{\overline{m}} \end{bmatrix}$$

are composed of the state-space matrices for each component autoregressive process, and the block-row matrix $C = \begin{bmatrix} C_{\underline{m}} & \dots & C_{\overline{m}} \end{bmatrix}$ forms the superposition of component autoregressive processes. The initial condition $\mathbf{x}[0]$ is chosen so that the system begins in steady-state. We define the *standard system description* with state vector $\mathbf{x}[n] = [x_{\underline{m}}[n], \dots, x_{\overline{m}}[n]]^T$ composed of the present states of the component processes, and state-space matrices

$$A_m = \beta_m \quad B_m = \left(f_m (1 - \beta_m^2) \right)^{1/2} \quad C_m = 1.$$
 (1)

It will prove useful to define an equivalent $augmented \ system$ description with 2M-dimensional state vector

$$\mathbf{x}[n] = [x_{\underline{m}}[n], x_{\underline{m}}[n-1], \dots, x_{\overline{m}}[n], x_{\overline{m}}[n-1]]^T$$

composed of the present and most recent past states of the component processes, and state-space matrices

$$A_m = \begin{bmatrix} \beta_m & 0\\ 1 & 0 \end{bmatrix} \quad B_m = \begin{bmatrix} \left(f_m (1 - \beta_m^2) \right)^{1/2} \\ 0 \end{bmatrix}$$
$$C_m = \begin{bmatrix} 1 & 0 \end{bmatrix}. \tag{2}$$

3. PARAMETER ESTIMATION

We consider the problem in which we have observations z[n] of a discrete zero-mean Gaussian 1/f process s[n] with unknown parameters γ and σ^2 , corrupted by zero-mean independent identically distributed (i.i.d.) Gaussian noise w[n]

with unknown variance σ_w^2 , that is statistically independent of s[n], so

$$z[n] = s[n] + w[n], \quad 1 \le n \le N \tag{3}$$

where N is the length of the observed data. The observations take the form of an N-dimensional Gaussian random vector \mathbf{z} with probability density function

$$f_{\mathbf{Z}}(\mathbf{z};\theta) = \left[\det\left(2\pi\Lambda_{z}(\theta)\right)\right]^{-1/2} \exp\left[-\frac{1}{2}\mathbf{z}^{T}\Lambda_{z}^{-1}(\theta)\mathbf{z}\right]$$

where the covariance matrix $\Lambda_z(\theta)$ is indexed by the vector of unknown parameters $\theta = \{\gamma, \sigma^2, \sigma_w^2\}$.

An iterative estimate-maximize (EM) algorithm [3] finds the maximum likelihood (ML) estimate for the parameter vector θ . The complete data is defined as the observed signal \mathbf{z} together with the samples $\{x_m[n]\}_{n=0}^{N-1}$ of each component autoregressive process of the 1/f signal, each of which may be viewed as a column vector \mathbf{x}_m .

The EM algorithm begins with initial parameter estimates $\theta^{[1]} = \{\gamma^{[1]}, \sigma^{2[1]}, \sigma^{2[1]}_w\}$ and iterates between an estimation and maximization step, until it converges to a stationary point of the likelihood function. On the *l*th iteration, the estimation step efficiently calculates the following statistics of the complete data by applying the fixed-interval Kalman smoothing equations [1] to the observed data, using the augmented system description (2) of a 1/f signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$ and white noise with parameter $\sigma^{2[l]}_w$:

$$\hat{x}_m^{[l]}[n] = \mathbf{E} \left[x_m[n] \mid \mathbf{z}; \theta^{[l]} \right]$$
(4)

$$\hat{x}_{mm}^{[l]}[n, n-1] = \mathbf{E} \begin{bmatrix} x_m[n]x_m[n-1] \mid \mathbf{z}; \theta^{[l]} \end{bmatrix}$$
(5)

$$\hat{x}_{mk}^{[l]}[n] = \mathbf{E} \left[x_m[n] x_k[n] \mid \mathbf{z}; \theta^{[l]} \right]$$
(6)

for $0 \leq n \leq N-1$ and $\underline{m} \leq m, k \leq \overline{m}$. Collecting the estimates $\{\hat{x}_m^{[l]}[n]\}$ into a column vector $\hat{\mathbf{x}}_m^{[l]}$, note that (4) and (5) generate the main diagonal and adjacent diagonals of

$$\widehat{\mathbf{x}_m \mathbf{x}_m^T} = \mathbf{E} \left[\mathbf{x}_m \mathbf{x}_m^T \mid \mathbf{z}; \theta^{[l]} \right],$$

which are sufficient to compute $t_m = \operatorname{tr}(H_m^{-1} \widehat{\mathbf{x}_m \mathbf{x}_m^T})$ for

$$H_m^{-1} = \begin{bmatrix} 1 & -\beta_m & & 0 \\ -\beta_m & 1 + \beta_m^2 & -\beta_m & & \\ & \ddots & \ddots & \ddots & \\ & & -\beta_m & 1 + \beta_m^2 & -\beta_m \\ 0 & & & -\beta_m & 1 \end{bmatrix}$$
(7)

since this matrix is tridiagonal.

The maximization step then generates the parameter estimates for the subsequent iteration. The estimate for the variance of the white noise

$$\sigma_w^{2[l+1]} = \frac{1}{N} \left(\mathbf{z}^T \mathbf{z} - 2\mathbf{z}^T \sum_{m=\underline{m}}^{\overline{m}} \hat{\mathbf{x}}_m^{[l]} + \sum_{n=1}^N \sum_{m=\underline{m}}^{\overline{m}} \sum_{k=\underline{m}}^{\overline{m}} \hat{x}_{mk}^{[l]}[n] \right)$$

is straightforward. We solve the equation

$$\sum_{m=\underline{m}}^{\overline{m}} (2m+1-\underline{m}-\overline{m}) \frac{\Delta^{\gamma^{[l+1]}m}}{2\beta_m \Delta^{2m}} t_m = 0$$

	$\Delta = 4$		RMS error in γ		% RMS error in σ^2	
γ	\underline{m}	\overline{m}	N=50	N=100	N=50	N = 100
0.33	-5	7	0.2075	0.1541	43.31%	34.37%
1.00	-7	2	0.1567	0.1257	15.82%	10.93%
1.67	-11	1	0.1362	0.1182	41.06%	31.88%

Table 1: RMS error in estimates of γ and σ^2 as a function of data length N for the special case of no noise.

using a root-finding algorithm to obtain $\gamma^{[l+1]}$. Finally, this value is used to find

$$\sigma^{2[l+1]} = \frac{1}{NM} \sum_{m=\underline{m}}^{\overline{m}} \frac{\Delta^{\gamma^{[l+1]}m}}{\beta_m \Delta^{2m}} t_m$$

Table 1 illustrates the performance of the parameter estimation algorithm based on Monte Carlo simulations for the noise-free case $\sigma_w^2 = 0$.

4. SIGNAL ESTIMATION IN 1/f NOISE

The 1/f process is often a noise process obscuring another signal of interest. Suppose we have observations z[n] of a deterministic signal y[n] obscured by a discrete zero-mean Gaussian 1/f noise process s[n] with unknown parameters γ and σ^2 , so

$$z[n] = y[n] + s[n], \quad 0 \le n \le N - 1$$

where N is the length of the observed data. The signal is parameterized as a linear combination of a finite set of known basis signals \mathbf{b}_p ,

$$y[n] = \sum_{p=1}^{P} \lambda_p b_p[n]$$

for unknown real parameters $\lambda_1, \ldots, \lambda_P$. Again, an EM algorithm finds the ML estimate for the parameter vector $\theta = \{\lambda_1, \ldots, \lambda_P, \gamma, \sigma^2\}$. The complete data is defined as the observed signal z together with the samples \mathbf{x}_m of each component autoregressive process of the 1/f process.

To calculate the statistics of the complete data for the estimation step, on the lth iteration we form the modified observations sequence

$$\mathbf{z}^{\prime \left[l\right]} = \mathbf{z} - \sum_{p=1}^{P} \lambda_{p}^{\left[l\right]}$$

The fixed-interval Kalman smoothing equations [1] are applied to the modified observations $\mathbf{z}'^{[l]}$, using the augmented system description (2) of a 1/f signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$, to find:

$$\hat{x}_m^{[l]}[n] = \mathbf{E}\left[x_m[n] \mid \mathbf{z'}^{[l]}; \theta^{[l]}\right]$$
(8)

$$\hat{x}_{mm}^{[l]}[n, n-1] = \mathbf{E} \left[x_m[n] x_m[n-1] \mid \mathbf{z'}^{[l]}; \theta^{[l]} \right]$$
(9)

for $0 \le n \le N - 1$ and $\underline{m} \le m \le \overline{m}$. Note that (8) and (9) generate the main diagonal and adjacent diagonals of

$$\widehat{\mathbf{x}_m \mathbf{x}_m^T} = \mathbf{E}\left[\mathbf{x}_m \mathbf{x}_m^T \mid \mathbf{z}'^{[l]}; \theta^{[l]}\right]$$



Figure 1: Normalized error covariance in the estimate of the slope of a deterministic affine signal in 1/f noise of known parameters for several values of γ , as a function of data length N.

which are sufficient to compute $t_m = \operatorname{tr}(H_m^{-1} \widehat{\mathbf{x}_m \mathbf{x}_m^T})$ where H_m^{-1} is defined in (7).

The maximization step generates the parameter estimates for the subsequent iteration:

$$\gamma^{[l+1]} \leftarrow \sum_{m=\underline{m}}^{\overline{m}} (2m+1-\underline{m}-\overline{m}) \frac{\Delta^{\gamma^{[l+1]}m}}{2\beta_m \Delta^{2m}} t_m = 0$$

$$\sigma^{2[l+1]} = \frac{1}{NM} \sum_{m=\underline{m}}^{\overline{m}} \frac{\Delta^{\gamma^{[l+1]}m}}{\beta_m \Delta^{2m}} t_m$$

$$\lambda_p^{[l+1]} = (\mathbf{b}_p^T \mathbf{b}_p)^{-1} \mathbf{b}_p^T \left(\mathbf{z} - \sum_{k\neq p} \lambda_k^{[l]} \mathbf{b}_k - \sum_{m=\underline{m}}^{\overline{m}} \mathbf{x}_m^{[l]} \right).$$

We consider the special case of estimating a deterministic affine signal $y[n] = \lambda_1 + \lambda_2 n$ in 1/f noise of known parameters. Fig. 1 illustrates the normalized error covariance in the estimate of the slope λ_2 of the signal as a function of data length. For Brownian motion ($\gamma = 2$), it is well-known that the error covariance is proportional to 1/N, whereas for white noise ($\gamma = 0$), the error covariance is asymptotically proportional to $1/N^3$. For intermediate values of γ , the error covariance is asymptotically proportional to $1/N^{(3-\gamma)}$. As γ increases, it becomes increasingly difficult to estimate the slope of an affine signal in 1/f noise.

5. PREDICTION

Given observations of the form (3) (expressed as an *N*-length column vector **z**) with known parameters γ , σ^2 , and σ_w^2 , we consider estimates of s[n] for $n \geq N$. For the single-step prediction problem, $\hat{s}[N] = \mathrm{E}[s[N] \mid \mathbf{z}]$ is obtained by applying the Kalman filter [1] to the observed data, using the standard system description (1). This algorithm also produces the prediction error covariance $R_s[N] =$



Figure 2: Relative normalized single-step prediction error covariance $(R_s[N]-R_s[\infty])/\sigma_s^2$ vs. data length N for several random processes.

 $\mathbb{E}\left[(s[N] - \hat{s}[N])^2 \mid \mathbf{z}\right]$, which depends only on the number of observed samples N and the parameters γ , σ^2 , and σ_w^2 . We examine how $R_s[N]$ decreases with N relative to its minimum value

$$R_s[\infty] = \lim_{N \to \infty} R_s[N] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_s(\Omega) d\Omega.$$

for the noise-free case $\sigma_w^2 = 0$. In general, when $R_s[N]$ converges to $R_s[\infty]$ quickly, additional observations of past samples have limited value in prediction, implying that the memory of the process is short. On the other hand, when $R_s[N]$ converges to $R_s[\infty]$ slowly, the memory of the process is long. Fig. 2 shows $(R_s[N] - R_s[\infty])/\sigma_s^2$ as a function of N for several representative ARMA and 1/f processes. The ARMA processes have exponential convergence while the 1/f processes have much more persistent memory than ARMA processes.

Multi-step predictions $\hat{s}[N+M]$ for $M \geq 1$ given observations \mathbf{z} are obtained by again applying the Kalman filter. The algorithm provides the prediction error covariance $R_s^M[N] = \mathbb{E}\left[(s[N+M] - \hat{s}[N+M])^2 \mid \mathbf{z}\right]$ which depends on the number of observed samples N, the prediction distance M, and the parameters of the observations. We analyze how the multi-step predicticion error covariance increases as a function of M relative to its maximum value

$$R_s^{\infty}[N] = \lim_{M \to \infty} R_s^M[N] = \operatorname{var}(s[n])$$

for the special case of $\sigma_w^2 = 0$. Rapid covergence of $R_s^M[N]$ to $R_s^\infty[N]$ indicates a process with short memory. Fig. 3 shows $(R_s^\infty[N] - R_s^M[N])/\sigma_s^2$ as a function of M for several representative ARMA and 1/f processes. Again, the autoregressive and moving-average processes have exponential convergence while the 1/f processes have polynomial convergence, reflecting that 1/f processes have much more persistent memory than ARMA processes.



Figure 3: Relative normalized multi-step prediction error covariance $(R_s^{\infty}[N] - R_s^M[N])/\sigma_s^2$ vs. prediction distance M for several random processes, with observed data length $N = 10^4$.

6. REFERENCES

- B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.
- [2] R.J. Barton and V.H. Poor, "Signal detection in fractional Gaussian noise", *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 943-959, 1988.
- [3] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. Royal Statistical Society, pp. 1-38, 1977.
- [4] M. Deriche and A.H. Tewfik, "Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process", *IEEE Trans. Signal Processing*, vol. 41, no. 10, pp. 2977-2989, 1993.
- [5] C.W. Granger and R. Joyeux, "An introduction to long memory time series models and fractional differencing", J. Time Series Anal., vol. 1, no. 1, 1980.
- [6] J.R.M. Hosking, "Fractional differencing", Biometrika, vol. 68, no. 1, pp. 165-176, 1981.
- [7] M.S. Keshner, "1/f noise", Proc. of the IEEE, vol. 70, no. 3, pp. 212-218, 1982.
- [8] B.B. Mandelbrot, "Some noises with 1/f spectrum, a bridge between direct current and white noise", *IEEE Trans. Inform. Theory*, vol. 13, pp. 289-298, 1967.
- [9] A. van der Ziel, "On the noise spectra of semiconductor noise and of flicker effect", *Physica*, vol. 16, no. 4, pp. 359-372, 1950.
- [10] A.C. Wang, Multiscale state-space algorithms for processing 1/f signals, SM Thesis, MIT, May 1997.
- [11] G.W. Wornell, Signal Processing with Fractals: A Wavelet-Based Approach, Prentice-Hall, 1995.