Multiscale State-Space Algorithms for Processing $1/f$ Signals

by

Alexander C. Wang

B.S., University of Illinois, Urbana-Champaign (1995)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1997

Author .....
Department of Electrical Engineering and Computer Science
May 16, 1997

Certified by .....
Gregory W. Wornell
Associate Professor of Electrical Engineering
Thesis Supervisor

Certified by .....
Arthur B. Baggeroer
Ford Professor of Ocean and Electrical Engineering
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

Multiscale State-Space Algorithms for Processing $1/f$ Signals

by

Alexander C. Wang

Submitted to the Department of Electrical Engineering and Computer Science
on May 16, 1997, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Natural landscapes, noise in electrical devices, and fluctuations in the stock market are among the extraordinary variety of phenomena that exhibit fractal structure. As a result, the need for efficient and robust algorithms for processing fractal signals arises in many engineering contexts.

This thesis develops a discrete multiscale state-space representation for an important class of fractal random processes referred to as $1/f$ processes. We show that this representation, which is based on a signal expansion in terms of simple first-order processes, satisfies a novel frequency-based characterization for discrete $1/f$ processes. Using this representation, we develop efficient algorithms for several signal processing problems that have previously had no practical solution, such as prediction.

Iterative algorithms are developed for maximum-likelihood parameter estimation of $1/f$ signals in white Gaussian noise. These algorithms exploit the computational efficiency of the Kalman smoother. Performance evaluation using simulations suggest robust performance in noise-corrupted scenarios. These algorithms are extended to address the problem of estimating a deterministic signal in $1/f$ and white noise. We evaluate the estimator for the special case of affine signals in $1/f$ noise, which have potential applications in ocean temperature and economic time series analysis. The estimation of the slope of an affine signal in $1/f$ noise is shown to have a polynomial dependence on the length of the data.

The thesis also develops efficient recursive algorithms for prediction and smoothing of $1/f$ signals in white Gaussian noise. The distant past for $1/f$ processes is shown to have a significant effect on predictions of future data samples, in contrast to the distant past for autoregressive and moving-average processes. The single-step prediction error covariance is shown to decrease to the minimum error covariance as a polynomial function of data length. The multi-step prediction error covariance is shown to increase to the signal variance as a polynomial function of the distance of the predicted point from the observed data.

Thesis Supervisor: Gregory W. Wornell
Title: Associate Professor of Electrical Engineering

Thesis Supervisor: Arthur B. Baggeroer
Title: Ford Professor of Ocean and Electrical Engineering

Acknowledgments

First and foremost, I would like to thank my thesis supervisor Professor Gregory Wornell for his patience and guidance. His knowledge and insights have been inspirational. Additionally, I have benefited greatly from his concern for my professional development. Thank you, Greg, for encouraging me to take "intellectual ownership" of this thesis.

I also would like to thank my counselor Professor Alan Oppenheim. I have found his experience and thoughtful advice to be extremely helpful.

I am grateful to my colleagues in the Digital Signal Processing Group for their friendship, humor, and advice. Working with this exceptional group of individuals has been extremely rewarding. I would especially like to thank Li Lee for her assistance and comments on this thesis, and her willingness to listen to and discuss my many research questions and conjectures.

I would like to give special thanks to Professor Hua Lee at UCSB for his guidance and advice. I would like to also thank Professor David Ruzic at the University of Illinois for providing my first experience in a research environment.

I would like to thank Linda Kim for her love and patience. I truly appreciate her unwavering support and industrious example. Finally, I dedicate this thesis to my parents, who have always supported and encouraged me in all things.

Contents

1 Introduction **9**

2 $1/f$ Processes **12**

2.1 Statistical Self-Similarity 13

2.2 Long-Term Dependence 16

2.3 Infrared Catastrophe 17

2.4 Existing Models 17

3 Multiscale State-Space Models for Discrete $1/f$ Processes **20**

3.1 Discrete $1/f$ Processes 21

3.2 Continuous ARMA Model 23

3.3 Discrete Infinite-Order ARMA Model 24

3.3.1 Properties of the Model 27

3.3.2 Properties of the Component Processes 28

3.4 Discrete Finite-Order ARMA Model 29

3.4.1 Error Analysis 34

3.4.2 Effective Signal Power 35

3.4.3 Notational Conventions 36

4 State Estimation with Discrete $1/f$ Processes **37**

4.1 First-Order Autogressive Process 38

4.2 Multiscale Model for $1/f$ Processes 39

4.3 Kalman Filter 41

4.4 Kalman Smoother 42

5 Parameter Estimation for $1/f$ Signals **45**

5.1 $1/f$ Signal in White Gaussian Noise 46

5.1.1 Special Case of Known Noise Statistics 48

5.1.2 Model Order Effects 50

5.1.3 Simulations 53

5.2 $1/f$ Signal and Deterministic Signal 55

5.2.1 Special Case of Known $1/f$ Parameters 59

6 Prediction and Smoothing of $1/f$ Signals **62**

6.1 Single-Step Prediction 63

6.2 Multi-Step Prediction 67

6.3 Smoothing 69

7 Conclusions **74**

A Proof of Proposition 3.1 **76**

B The EM Parameter Estimation Algorithm **78**

B.1 $1/f$ Signal in White Gaussian Noise 79

B.2 $1/f$ Signal and Deterministic Signal 84

List of Figures

2-1 Power spectrum $S(\omega)$ of $1/f$ process with parameter $\gamma = 1$ 13

2-2 $1/f$ process with parameter $\gamma = 1.67$ viewed on different time scales 14

2-3 Sample paths of $1/f$ processes with various values of γ 18

3-1 Power spectrum $S(\omega)$ of nearly-$1/f$ process using van der Ziel's continuous ARMA model 25

3-2 Power spectrum $S_{\beta_m}(\Omega)$ of weakly-correlated component processes 30

3-3 Power spectrum $S_{\beta_m}(\Omega)$ of strongly-correlated component processes 31

3-4 Finite-order nearly-$1/f$ process spectrum. 33

5-1 RMS error in estimates of γ as a function of order parameter M_H 52

5-2 RMS error in estimates of γ as a function of order parameter M_L 52

5-3 RMS error in estimates of γ and σ^2 as a function of data length N in the special case of no noise 54

5-4 RMS error in estimates of γ and σ^2 as a function of SNR with data length $N = 50$ 56

5-5 Error covariance in the estimate of the slope of a deterministic affine signal in $1/f$ noise of known parameters as a function of data length N 61

6-1 $R_s[N+1|N] - R_s[\infty|\infty]$ vs. N for a moving average process 66

6-2 $R_s[N+1|N] - R_s[\infty|\infty]$ vs. N for $1/f$ processes 66

6-3 $\sigma_s^2 - R_s[N+M|N]$ vs. M for autoregressive process 69

6-4 $\sigma_s^2 - R_s[N+M|N]$ vs. M for $1/f$ processes 70

6-5 SNR gain (dB) of the signal estimate as a function of SNR (dB) of the observations over relevant frequencies for various γ 72

6-6 Example of signal smoothing 73

List of Tables

3.1 Variances f_m of component processes 29

3.2 Finite component process choices satisfying error bound ϵ 35

4.1 Kalman Filter Equations 42

4.2 Kalman Smoothing Equations 44

5.1 EM algorithm for parameter estimation of $1/f$ signal in white noise of unknown statistics 49

5.2 EM algorithm for parameter estimation of $1/f$ signal in white noise of known statistics 50

5.3 EM algorithm for parameter estimation of deterministic signal obscured by $1/f$ and white noise 58

6.1 Single-step prediction algorithm 64

6.2 Multi-step prediction algorithm 67

6.3 Smoothing algorithm using finite-interval data 71

Chapter 1

Introduction

Natural landscapes, noise in electrical devices, and fluctuations in the stock market are among the extraordinary variety of phenomena that exhibit fractal structure. As a result, the need for efficient and robust algorithms for processing fractal signals arises in many engineering contexts.

The $1/f$ processes are an important class of fractal random processes. The fractal nature of these processes originates from self-similarity and scale invariance in statistical structure. As a consequence of this special structure, these processes demonstrate behavior qualitatively different from traditional random process models used in statistical signal processing. For example, in contrast with the widely used family of ARMA processes which have covariance functions that decay exponentially with lag, $1/f$ processes have strong long-term statistical dependence characterized by covariance functions with polynomial-type decay.

Due to the wide range of phenomena modeled as $1/f$ processes, there is increasing demand for applications for processing these signals. Algorithms for predicting future values of a $1/f$ signal based on observations of the process over a finite time interval could have applications in economic forecasting, for instance. This thesis develops algorithms for processing fractal signals based on a multiscale state-space representation. Since measured data is typically represented as a discrete sequence, the thesis focuses on discrete-time $1/f$ processes observed over a finite data length. This state-space representation is particularly well-suited for several signal processing problems, such as prediction, that have previously had no practical solution. In addition, the algorithms in this thesis are robust in the presence

of broadband measurement noise.

The thesis begins with a brief background for continuous $1/f$ processes in Chapter 2. These processes have several important empirical properties including statistical self-similarity and scale invariance. We review several existing models for describing, generating, and analyzing $1/f$ processes.

This thesis focuses on discrete $1/f$ processes which have different self-similarity characteristics than continuous $1/f$ processes. Chapter 3 examines these differences and presents a frequency-domain characterization for discrete-time $1/f$ processes based on a restricted notion of self-similarity. We present a infinite-order multiscale state-space representation for nearly-$1/f$ processes that satisfies this characterization. This representation, analogous to a continuous framework due to van der Ziel [17], is based on the superposition of first-order autoregressive processes with different characteristic time-constants. We show that a properly chosen finite-order approximation to this infinite-order representation exhibits $1/f$ spectral behavior over a finite frequency range. Using this approximation, a state-space system description for $1/f$ signals in a background of additive white Gaussian noise is presented in Chapter 4. This description allows computationally efficient optimal estimation algorithms based on noisy observations of $1/f$ signals, using the Kalman filtering and smoothing equations.

In signal processing applications involving $1/f$ signals, it is frequently necessary to characterize and parameterize the signal. Chapter 5 addresses the problem of parameter estimation with $1/f$ signals. Joint signal and noise estimation of the parameters of a $1/f$ signal process obscured by additive white Gaussian noise is performed using an Estimate-Maximize (EM) algorithm. The Estimation step in the algorithm exploits the state-space system description and filtering equations from Chapter 4. The properties of the estimator are analyzed with Monte-Carlo simulations. This algorithm is extended to include joint signal and noise estimation of the parameters of an arbitrary parameterized deterministic signal obscured by a combination of additive $1/f$ noise and white Gaussian noise. This algorithm is potentially applicable in processing signals such as economic time series and oceanographic temperature data. The special case of affine deterministic signals corrupted by $1/f$ noise of known parameters is analyzed.

Chapter 6 presents computationally efficient algorithms for single-step and multi-step prediction and smoothing of $1/f$ processes. These algorithms exploit the state-space de-

scription of Chapter 4 and are computationally efficient. The distant past for $1/f$ processes is shown to have a significant effect on predictions of future data samples, in contrast to the distant past for autoregressive and moving-average processes.

Finally, Chapter 7 reviews the contributions of the thesis and suggests directions for future research.

Chapter 2

$1/f$ Processes

The $1/f$ family of fractal random processes provides useful models for an extraordinary variety of natural and man-made phenomena that exhibit long-term dependence. Examples of such processes include voltages and currents in electrical devices such as diodes and transistors; phase noise in oscillators; the amplitude and pitch of music; and fluctuations in stock market indices.

A $1/f$ process is empirically defined as having measured power spectral density of the form

$$S_x(\omega) \sim \frac{\sigma_x^2}{|\omega|^\gamma} \tag{2.1}$$

over several decades of frequency ω, where γ is a parameter in the range $0 \leq \gamma \leq 2$. Two particular cases are important and well-known. White noise, with a constant power spectrum, corresponds to $\gamma = 0$, while Brownian motion (also known as the Wiener process) corresponds to $\gamma = 2$. In many examples of $1/f$ behavior in nature, $\gamma \approx 1$. Figure 2-1 gives an example of a $1/f$ spectrum with $\gamma = 1$.

A more general class of processes has spectral behavior that is approximately $1/f$. *Nearly*-$1/f$ processes [19] have a measured power spectral density that is bounded according to

$$\frac{\sigma_L^2}{|\omega|^\gamma} \leq S(\omega) \leq \frac{\sigma_U^2}{|\omega|^\gamma} \tag{2.2}$$

where σ_L^2 and σ_U^2 are arbitrary constants satisfying $0 < \sigma_L^2 \leq \sigma_U^2 < \infty$.

The $1/f$ processes are characterized by statistical self-similarity and long-term statistical dependence. The first part of the chapter discusses the empirical properties of $1/f$ processes

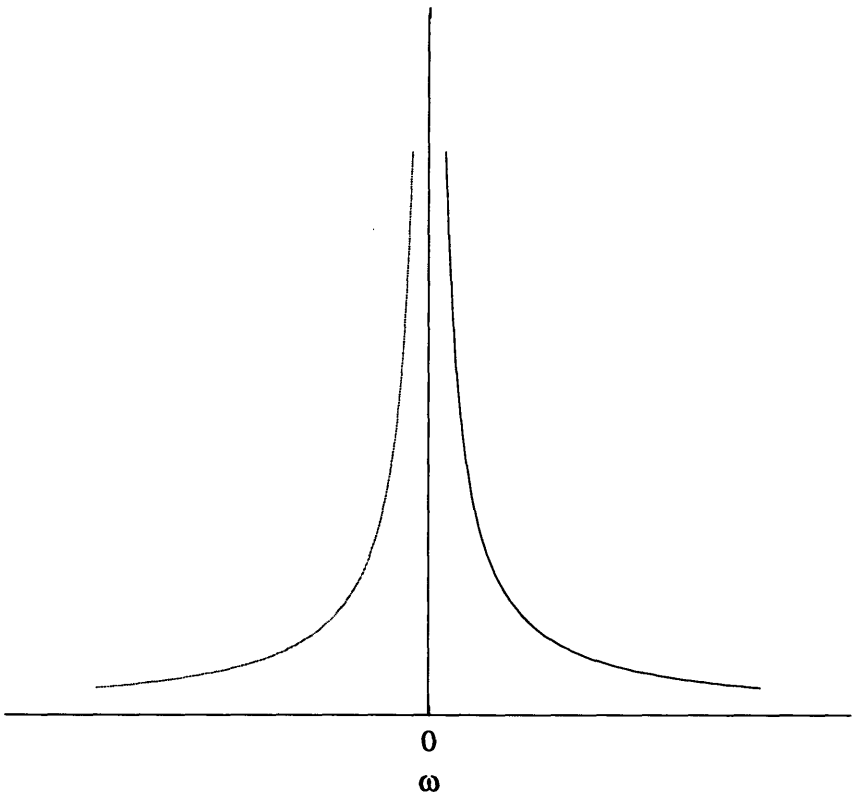


Figure 2-1: Power spectrum $S(\omega)$ of $1/f$ process with parameter $\gamma = 1$

and reviews a frequency-based characterization for continuous $1/f$ processes. A similar characterization for discrete $1/f$ processes will be presented in the Chapter 3.

Due to the wide variety of phenomena that are well modeled by $1/f$ processes, it is often desirable to perform signal processing with $1/f$ signals. Consequently, mathematical models for generating and analyzing $1/f$ processes are necessary. The second part of this chapter gives a brief overview of some existing models for $1/f$ and nearly-$1/f$ processes. In Chapter 3, a multiscale state-space model for nearly-$1/f$ behavior will be developed which addresses several signal processing problems that are effectively intractable using these models.

2.1 Statistical Self-Similarity

The $1/f$ processes are inherently self-similar: the statistical structure of these processes does not change when observed on different time scales. Figure 2-2 illustrates a $1/f$ process with parameter $\gamma = 1.67$ observed on different scales. A formal definition of this property is given below.

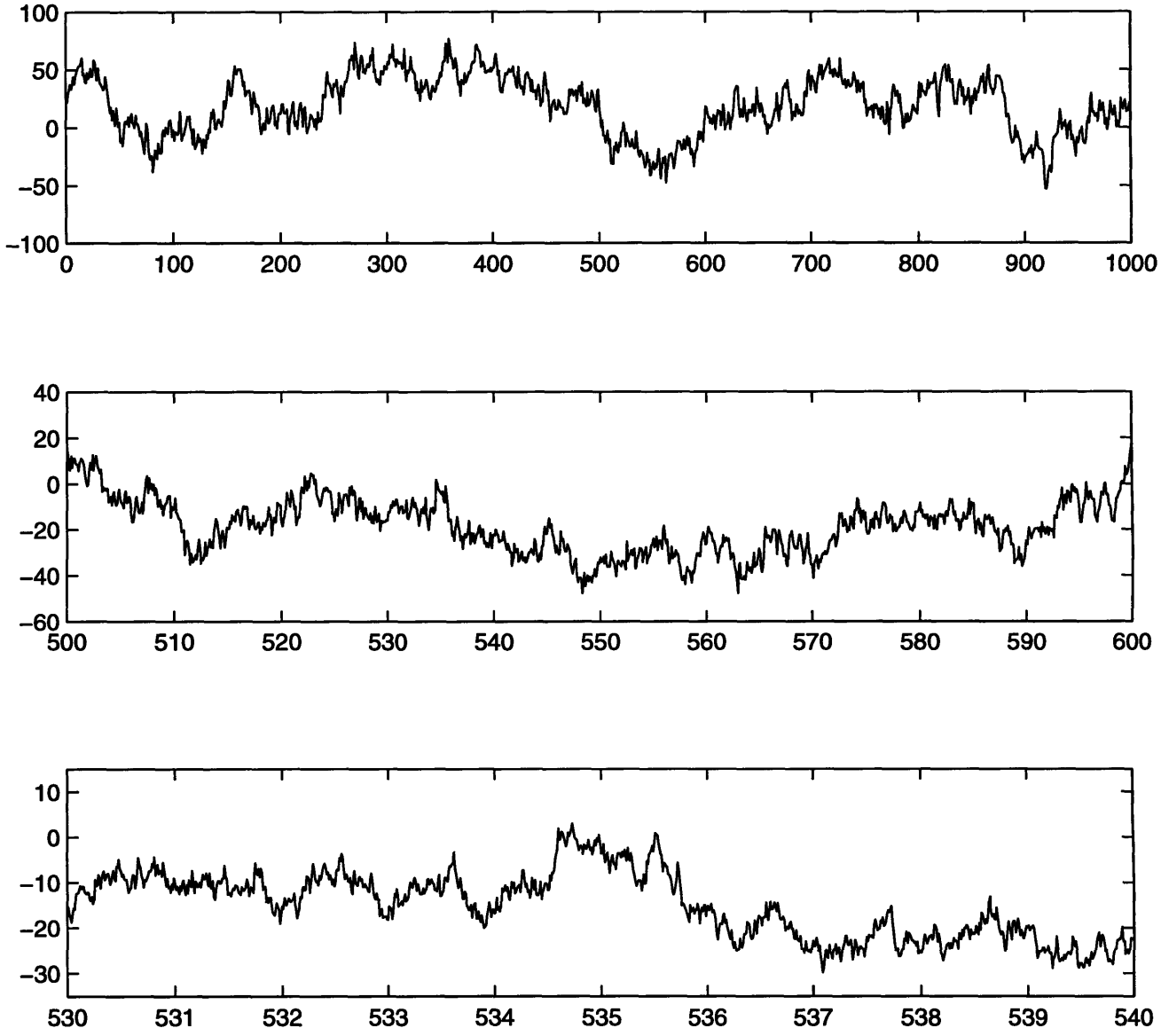


Figure 2-2: $1/f$ process with parameter $\gamma = 1.67$ viewed on different time scales

Definition 2.1 *A random process $x(t)$ defined on $-\infty < t < \infty$ is said to be* statistically self-similar [20] *with parameter H if, for any real $a > 0$,*

$$x(t) \stackrel{P}{=} a^{-H} x(at) \tag{2.3}$$

where $\stackrel{P}{=}$ denotes equality in a statistical sense.

This equality holds for all finite-dimensional joint probability distributions in the case of *strict-sense* self-similar processes. For *wide-sense* self-similar processes, the equality holds for second-order statistics; the self-similarity property (2.3) may be expressed as

$$M_x(t) \equiv E[x(t)] = a^{-H} M_x(at) \tag{2.4}$$

$$R_x(t,s) \equiv E[x(t)x(s)] = a^{-2H} R_x(at, as). \tag{2.5}$$

In this thesis, we consider zero-mean Gaussian processes, for which these definitions are equivalent.

A $1/f$ process with parameter γ is statistically self-similar with parameter H, where

$$\gamma = 2H + 1.$$

A mathematical characterization for $1/f$ processes in the frequency domain based on this property is given in [20]:

Definition 2.2 *A wide-sense statistically self-similar zero-mean random process $x(t)$ is said to be a* $1/f$ process *if there exist ω_0 and ω_1 satisfying $0 < \omega_0 < \omega_1 < \infty$ such that when $x(t)$ is filtered by an ideal bandpass filter with frequency response*

$$B_1(\omega) = \begin{cases} 1 & \omega_0 < |\omega| < \omega_1 \\ 0 & otherwise \end{cases} \tag{2.6}$$

the resulting process $y(t)$ is wide-sense stationary and has finite variance.

This definition is justified by the following theorem [20]:

Theorem 2.1 *A $1/f$ process $x(t)$, when filtered by an ideal bandpass filter with frequency response*

$$B(\omega) = \begin{cases} 1 & \omega_L < |\omega| \leq \omega_U \\ 0 & \textit{otherwise} \end{cases} \tag{2.7}$$

for arbitrary $0 < \omega_L < \omega_U < \infty$, yields a wide-sense stationary random process $y(t)$ with finite variance and having power spectrum

$$S_y(\omega) = \begin{cases} \sigma_x^2/|\omega|^\gamma & \omega_L < |\omega| \leq \omega_U \\ 0 & \textit{otherwise} \end{cases} \tag{2.8}$$

for some $\sigma_x^2 > 0$, and where the spectral exponent γ is related to the self-similarity parameter H according to $\gamma = 2H + 1$.

In Section 3.1, we give a mathematical characterization for *discrete* random processes in the frequency domain.

2.2 Long-Term Dependence

As shown above, the statistical structure of $1/f$ processes at long time scales is similar to the statistical structure on short time scales. As a result, even distant samples of the process exhibit relatively strong correlation. As opposed to the traditional autoregressive moving-average (ARMA) models characterized by correlation functions with exponential decay, $1/f$ processes exhibit strong long-term dependence characterized by correlation functions with polynomial-type decay. The generalized Fourier pair [3]

$$\frac{|\tau|^{\gamma-1}}{2\Gamma(\gamma)\cos(\gamma\pi/2)} \leftrightarrow \frac{1}{|\omega|^\gamma} \tag{2.9}$$

valid for $\gamma > 0$ but $\gamma \neq 1, 2, 3, \ldots$, suggests that for $0 < \gamma < 1$, the autocorrelation $R_x(\tau)$ associated with $S_x(\omega)$ is characterized by polynomial-type decay of the form

$$R_x(\tau) \sim |\tau|^{\gamma-1}. \tag{2.10}$$

The strength of this dependence varies with the parameter γ. White noise, corresponding to $\gamma = 0$, exhibits no long-term dependence. As γ increases, the strength of the long-term

dependence increases. Brownian motion, corresponding to $\gamma = 2$, represents a process with extremely strong long-term correlation. Figure 2-3 illustrates sample paths of $1/f$ processes for various values of γ.

2.3 Infrared Catastrophe

For $\gamma \geq 1$, the integral of the power spectrum of the $1/f$ process is infinite. This infinite low-frequency energy suggests that a stationary $1/f$ process would have infinite variance. This phenomenon is termed the infrared catastrophe. This problem has been interpreted as revealing that the process as inherently nonstationary [11] [14]. In this thesis, we assume that the $1/f$ process is stationary with a power spectrum that changes from $1/f$ to flat below a certain frequency, although this low-frequency roll-off is not always observed in natural signals (see [11] and the references cited).

For $\gamma \leq 1$, the integral of the power spectrum of the $1/f$ process is infinite due to the high-frequency energy of the signal. However, it is shown in Chapter 3 that this phenomenon, termed the ultraviolet catastrophe, does not arise in the context of discrete $1/f$ processes.

2.4 Existing Models

This section briefly reviews the fractional Brownian motion, discrete fractionally differenced Gaussian noise, and the wavelet-based model for $1/f$ processes.

The theory of of the fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) models was developed by Mandelbrot and Van Ness [15]. Let H be the self-similarity parameter of the fBm. Mandelbrot and Van Ness define fractional Brownian motion

$$x(t) \equiv \frac{1}{\Gamma(H+1/2)} \left[\int_{-\infty}^{0} \left(|t-\tau|^{H-1/2} - |\tau|^{H-1/2} \right) w(\tau) d\tau + \int_{0}^{t} |t-\tau|^{H-1/2} w(\tau) d\tau \right] \tag{2.11}$$

for $0 < H < 1$, where $w(t)$ is a zero-mean, stationary white Gaussian noise process with unit spectral density. The generalized derivative of fractional Brownian motion [2]

$$x'(t) = \frac{d}{dt} x(t) = \frac{1}{\Gamma(H+1/2)} \int_{-\infty}^{t} |t-\tau|^{H'-1/2} w(\tau) d\tau \tag{2.12}$$

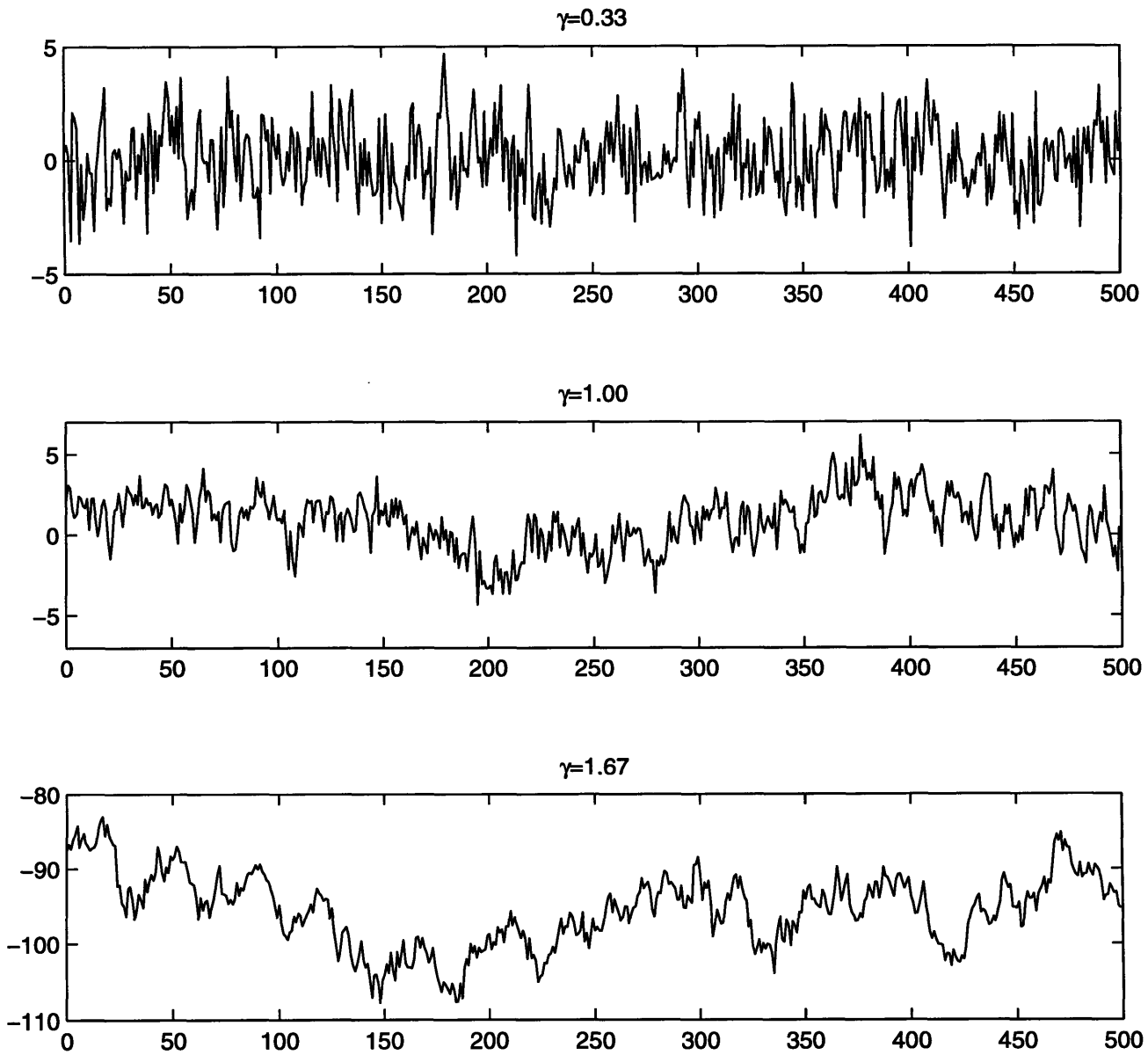


Figure 2-3: Sample paths of $1/f$ processes with various values of γ

is termed fractional Gaussian noise, where $H' = H-1$. This process is zero-mean, stationary and statistically self-similar with parameter H'. Additionally, fGn has $1/f$-type spectral behavior. The problem of detecting a deterministic signal obscured by additive fGn over a finite sequence is addressed by Barton and Poor [2]. A discrete-time extension to fractional Brownian motion and fractional Gaussian noise via sampling is given in [13].

The fractional Brownian motion may be viewed as the fractional integral of white noise of order $H + 0.5$. By analogy, Granger and Joyeux [8] and Hosking [10] define a discrete fractionally differenced Gaussian noise (fdGn) process parameterized by d as the fractional difference of discrete white Gaussian noise

$$w_d[n] = \sum_{k=0}^{\infty} \binom{-d}{k} (-1)^k w[n-k] \tag{2.13}$$

$$= \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!(d-1)!} w[n-k] \tag{2.14}$$

where w[n] is a discrete zero-mean white Gaussian noise process with variance σ_w^2. The discrete spectral density of dfGn

$$S(\Omega) = \frac{2^{-2d}\sigma_w^2}{\left(\sin\left(\frac{\Omega}{2}\right)\right)^{2d}} \approx \sigma_w^2 \Omega^{-2d} \tag{2.15}$$

approaches $1/f$-type behavior at low frequencies, with parameter $\gamma = 2d$. Deriche and Tewfik [5] give a procedure for estimating the parameters of dfGn sequences.

The scale-invariance characteristic of $1/f$ processes suggests that natural models for these processes might be based on representing the model as the sum of many component processes, each with a characteristic scale. Although not the focus on this thesis, wavelet expansions, with their characteristic scale-invariance, represent examples of a natural approach to modeling $1/f$-type behavior. Wavelet-based models for $1/f$-like behavior are presented in [20].

Chapter 3

Multiscale State-Space Models for Discrete $1/f$ Processes

This chapter introduces a multiscale state-space representation for $1/f$ processes, based on the superposition of first-order autoregressive processes. We focus on the typical case in which observations of the $1/f$ signal are available only as a discrete data sequence over a finite time interval, limiting the range of frequencies over which $1/f$ spectral behavior can be observed. In Chapters 5 and 6, we see that this state-space representation is particularly well suited for several signal processing problems that existing models have not addressed, such as prediction of $1/f$ signals based on noisy observations.

We begin the chapter with an examination of the differences between discrete and continuous $1/f$ processes. In a discrete data sequence, the highest observable frequency is constrained by limited time-resolution due to sampling of the data. This consideration motivates a frequency-based characterization for discrete $1/f$ processes which requires $1/f$ spectral behavior asymptotically at low frequencies.

We continue with a description of the discrete multiscale state-space models for nearly-$1/f$ processes that are the focus of this chapter. We review a generalized autoregressive moving-average (ARMA) framework for continuous $1/f$ signals which was originally described by van der Ziel [17]. This framework describes nearly-$1/f$ processes composed of many underlying processes, each with a single characteristic scale. We develop a discrete-time adaptation of this framework and show that the resulting models describe processes that are nearly-$1/f$.

The infinite order of these models suggest that an infinite number of state variables are required to represent the memory of a $1/f$ process. Fortunately, since the finite data length constrains the lowest frequency observable in the process, only a finite number of state variables which generate $1/f$ behavior above this frequency are required for the applications considered in this thesis. We examine how particular choices of state variables affect the accuracy of these finite-order models over various frequencies. Using these models, $1/f$ signals in a background of white Gaussian noise are represented in Chapter 4 with a finite state-space system, allowing the use of the Kalman filter and smoother for applications involving $1/f$ signals of finite data length.

3.1 Discrete $1/f$ Processes

The data modeled as a $1/f$ process is generally represented with a discrete sequence. The data may be samples of a continuous-time $1/f$ process, such as in the case of ocean temperature data taken at fixed time intervals. In other cases, such as the fluctuations in the stock market, the data is fundamentally discrete. The discretization of the time axis limits the shortest time-scale – or equivalently the highest frequency – in which $1/f$ spectral behavior can be observed.

Useful insights regarding the effects of discretization can be gained by considering the power spectrum of a hypothetical discrete process with parameter γ satisfying

$$S_x(\Omega) \sim \frac{\sigma_x^2}{|\Omega|^\gamma}, \quad -\pi < \Omega < \pi. \tag{3.1}$$

While the spectral density of continuous $1/f$ signals

$$S_x(\omega) \sim \frac{\sigma_x^2}{|\omega|^\gamma}, \quad -\infty < \omega < \infty \tag{3.2}$$

vanishes for sufficiently high frequencies, the discrete process described by (3.1) will have significant power at the highest frequency $\Omega = \pi$. In addition, the discrete process with parameter $\gamma \leq 1$ will not have infinite high-frequency energy, as opposed to continuous $1/f$ signals with parameter $\gamma \leq 1$. However, the discrete process with parameter $\gamma \geq 1$ will have infinite low-frequency energy, as in the case of continuous $1/f$ signals with parameter $\gamma \geq 1$ described in Section 2.3.

This analysis suggests that the low-frequency behavior of discrete $1/f$ processes is of primary interest which motivates the following frequency-based characterization for discrete $1/f$ processes.

Definition 3.1 *Let $x[n]$ by a discrete-time zero-mean wide-sense stationary random process. Then $x[n]$ is said to be a* discrete $1/f$ process *with parameter γ if*

$$\lim_{k \longrightarrow \infty} \int_0^{\pi/k} \left(S_x(\Omega) - k^{-\gamma} S_x(\Omega/k)\right)^2 d\Omega = 0 \tag{3.3}$$

where $S_x(\Omega)$ is the power spectrum of $x[n]$.

Essentially, we define a process to be a discrete $1/f$ process if its spectral behavior is self-similar at arbitrarily low frequencies. We can interpret this property in the time domain by viewing $S_x(\Omega/k)$ in terms of a downsampling operation. Let $\tilde{x}[n]$ be the output when $x[n]$ is passed through an ideal low-pass filter with frequency response

$$B(\Omega) = \begin{cases} 1 & |\Omega| \leq \pi/k \\ 0 & \pi/k < |\Omega| \leq \pi \end{cases} \tag{3.4}$$

and let

$$\tilde{x}_k[n] = \tilde{x}[kn] \tag{3.5}$$

representing $x[n]$ downsampled by a factor of k. Then the spectrum of $\tilde{x}_k[n]$ is [16]

$$S_{\tilde{x}_k}(\Omega) = \tfrac{1}{k} S_x(\Omega/k), \quad |\Omega| < \pi. \tag{3.6}$$

Heuristically, a discrete $1/f$ process can be viewed as being self-similar under downsampling; *i.e.*, the downsampled $1/f$ process has a similar long-term statistical structure as the original process. Compare this to continuous $1/f$ processes which can be viewed as being self-similar for both expansion and contraction of the time axis. We define discrete *nearly*-$1/f$ processes that approximate this behavior.

Definition 3.2 *A discrete-time zero-mean random process $x[n]$ is a* discrete nearly-$1/f$ process *with parameter γ if there exists a frequency Ω_0 such that for all $\Omega < \Omega_0$,*

$$\sigma_L^2 S_y(\Omega) \leq S_x(\Omega) \leq \sigma_U^2 S_y(\Omega) \tag{3.7}$$

for some $0 < \sigma_L^2 \leq \sigma_U^2 < \infty$, *where* $S_y(\Omega)$ *is the power spectrum of some discrete* $1/f$ *process with parameter* γ *satisfying Definition 3.1.*

3.2 Continuous ARMA Model

In this section, we review a model for continuous processes exhibiting nearly-$1/f$ spectral behavior. This model forms the basis for an analogous discrete model in the following section. Although we focus on van der Ziel's model in this thesis, another ARMA-based model for nearly-$1/f$ processes developed by Keshner [11] gives useful insight into the long-term memory of $1/f$ processes.

Van der Ziel [17] modeled $1/f$ processes as the weighted superposition of a continuum of uncorrelated random processes. Each component process is governed by a distinct characteristic time-constant $1/\alpha$, with correlation function

$$R_\alpha(\tau) = e^{-\alpha|\tau|} \tag{3.8}$$

and corresponding Lorenzian spectrum

$$S_\alpha(\omega) = \frac{2\alpha}{\alpha^2 + \omega^2}. \tag{3.9}$$

Choosing a scale-invariant weighting function of the form

$$f(\alpha) = \alpha^{-\gamma} \tag{3.10}$$

for $0 < \gamma < 2$, the weighted superposition of the continuum of these processes has an effective spectrum

$$S_x(\omega) = \int_0^\infty S_\alpha(\omega) f(\alpha) d\alpha \tag{3.11}$$

which is $1/f$, *i.e.*,

$$S_x(\omega) \propto \frac{1}{|\omega|^\gamma}. \tag{3.12}$$

In fact, the superposition of a countably infinite collection of appropriately distributed single time-constants processes is sufficient to model nearly-$1/f$ behavior as described by (2.2). In particular, consider a distribution of exponentially-spaced poles described accord-

ing to

$$\alpha_m = \Delta^m, \quad -\infty < m < \infty \tag{3.13}$$

for some $1 < \Delta < \infty$, and weighting function

$$g_m = \sigma^2 \alpha_m^{1-\gamma} \tag{3.14}$$

where σ^2 is a parameter governing the amplitude of the $1/f$ process. The weighted superposition of these processes has the corresponding spectrum

$$S_x(\omega) = \sum_m \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\omega^2 + \Delta^{2m}} \tag{3.15}$$

which is bounded according to

$$\frac{\sigma_L^2}{|\omega|^\gamma} \leq S_x(\omega) \leq \frac{\sigma_U^2}{|\omega|^\gamma} \tag{3.16}$$

for some $0 < \sigma_L^2 \leq \sigma_U^2 < \infty$ with ripple such that

$$|\omega|^\gamma S_x(\omega) = |\Delta^k \omega|^\gamma S_x(\Delta^k \omega) \tag{3.17}$$

for all integers k [20]. On a log-log frequency plot, the process has a spectrum that is $1/f$ with a superimposed ripple with uniform spacing and amplitude. Both the amplitude and the spacing of the ripple decrease as $\Delta \to 1$. Figure 3-1 is an example of several nearly-$1/f$ spectra for both $\Delta = 10$ and $\Delta = 1000$ and $\gamma = 1$.

3.3 Discrete Infinite-Order ARMA Model

By analogy to the continuous-time model based on exponentially-spaced poles described in Section 3.2, we develop a model for discrete-time $1/f$ processes based on the superposition of a countable collection of uncorrelated processes. For each continuous component process, we find a discrete component process whose spectrum is asymptotically equivalent to the spectrum of the continuous component process at low frequencies. The superposition of the resulting component processes has a spectrum that asymptotically approaches the spectrum of (3.15) at frequencies near zero. Consequently, the discrete superposition is nearly-$1/f$

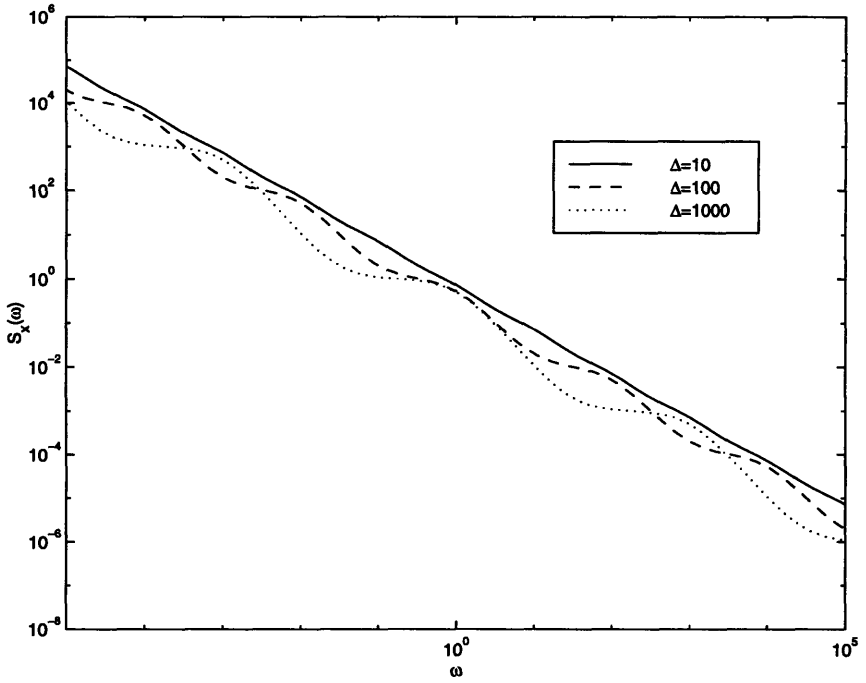


Figure 3-1: Power spectrum $S(\omega)$ of nearly-$1/f$ process using van der Ziel's continuous ARMA model

according to Definition 3.2.

A discrete first-order autoregressive process governed by scaling parameter $f_m > 0$ and characteristic time-constant $1/\beta_m$ satisfying $0 < \beta_m < 1$ has correlation function

$$R_{\beta_m, f_m}[k] = f_m \beta_m^{|k|} \tag{3.18}$$

corresponding to a spectral density [9]

$$S_{\beta_m, f_m}(\Omega) = \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m \cos\Omega} \tag{3.19}$$

which can be modeled as the output of a causal LTI function with system function

$$H_{\beta_m, f_m}(z) = \frac{\sqrt{f_m(1-\beta_m^2)}}{1-\beta_m z^{-1}} \tag{3.20}$$

driven by an independent stationary white noise source. Substituting the Taylor series expansion at $\Omega = 0$,

$$\cos\Omega = 1 - \frac{\Omega^2}{2!} + \frac{\Omega^4}{4!} - \ldots \approx 1 - \frac{\Omega^2}{2} \tag{3.21}$$

gives an approximate spectrum

$$S_{\beta_m,f_m}(\Omega) \approx \frac{f_m(\beta_m^{-1} - \beta_m)}{\Omega^2 + \left(\beta_m^{-1/2} - \beta_m^{1/2}\right)^2}, \quad |\Omega| << 1. \tag{3.22}$$

We wish to find β_m and f_m for all integer m such that

$$S_x(\Omega) = \sum_{m=-\infty}^{\infty} S_{\beta_m,f_m}(\Omega) \tag{3.23}$$

has $1/f$ behavior over low frequencies.

Therefore, we choose β_m and f_m such that $S_x(\Omega)$ has approximately the same frequency behavior as the continuous-time spectrum $S_x(\omega)$ in (3.15) as Ω approaches 0, *i.e.*,

$$S_x(\Omega) \approx \sum_{m=-\infty}^{\infty} \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\Omega^2 + \Delta^{2m}} \tag{3.24}$$

where $1 < \Delta < \infty$. For the approximation (3.22) to satisfy the desired low-frequency behavior (3.24), we simultaneously solve

$$\left(\beta_m^{-1/2} - \beta_m^{1/2}\right)^2 = \Delta^{2m} \tag{3.25}$$

$$f_m(\beta_m^{-1} - \beta_m) = \sigma^2 \Delta^{(2-\gamma)m} \tag{3.26}$$

for β_m and f_m. Although (3.25) has multiple solutions for β_m, only

$$\beta_m = \left(\frac{2}{\Delta^m + \sqrt{\Delta^{2m} + 4}}\right)^2 \tag{3.27}$$

satisfies $0 < \beta_m < 1$. Note that the time-constants β_m are independent of the parameters γ and σ^2 of the $1/f$ process. Substituting (3.27) into (3.26), we have that

$$f_m = \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\beta_m^{-1} - \beta_m}. \tag{3.28}$$

The weights f_m depend on the $1/f$ process parameters γ and σ^2.

3.3.1 Properties of the Model

We first show that this model describes a discrete nearly-$1/f$ process. The discrete-time spectrum satisfies

$$S_x(\Omega) = \sum_{m=-\infty}^{\infty} S_{\beta_m, f_m}(\Omega) \tag{3.29}$$

$$\approx \sum_{m=-\infty}^{\infty} \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\Omega^2 + \Delta^{2m}} \tag{3.30}$$

where this approximation is valid for low frequencies. We have seen in Section 3.2 that (3.30) represents the spectrum of a nearly-$1/f$ process. The following proposition gives bounds for the difference between the actual discrete-time spectrum (3.29) and the approximation (3.30) below a given frequency Ω_0.

Proposition 3.1 *Let $S_x(\Omega)$ be the spectrum of the process with parameters γ and σ^2 described by the model in this section,* i.e.,

$$S_x(\Omega) = \sum_{m=-\infty}^{\infty} \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m \cos\Omega} \tag{3.31}$$

with

$$\beta_m = \left(\frac{2}{\Delta^m + \sqrt{\Delta^{2m}+4}} \right)^2 \tag{3.32}$$

$$f_m = \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\beta_m^{-1} - \beta_m} \tag{3.33}$$

where $1 < \Delta < \infty$. Let

$$S_y(\Omega) = \sum_{m=-\infty}^{\infty} \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\Omega^2 + \Delta^{2m}}, \quad \pi < \Omega < \pi. \tag{3.34}$$

Then for any Ω_0 satisfying $0 < \Omega_0 < \pi$,

$$S_y(\Omega) < S_x(\Omega) < \frac{12}{12 - \Omega_0^2} S_y(\Omega) \tag{3.35}$$

for all Ω satisfying $0 < \Omega < \Omega_0$.

The proof of Proposition 3.1 is given in Appendix A. The following proposition is a direct result of Proposition 3.1 and (3.16).

Proposition 3.2 *Let $S_x(\Omega)$ be the spectrum of the process with parameters γ and σ^2 described by the model in this section. Then $S_x(\Omega)$ describes a discrete nearly-$1/f$ process as defined in Definition 3.2.*

3.3.2 Properties of the Component Processes

In this section, we analyze the component process parameters β_m and f_m. We observe that

$$\lim_{m\to-\infty} \beta_m = 1 \tag{3.36}$$

$$\lim_{m\to\infty} \beta_m = 0 \tag{3.37}$$

illustrating that the model contains processes whose behavior ranges from being similar to Brownian motion (low-frequency) to being similar to white noise (high-frequency). It is straightforward to show that for large positive values of m corresponding to high-frequency component processes,

$$f_m \approx \sigma^2 \Delta^{-\gamma m} \tag{3.38}$$

and therefore

$$\lim_{m\to\infty} f_m = \begin{cases} \sigma^2, & \gamma = 0 \\ 0, & \gamma > 0 \end{cases} \tag{3.39}$$

implying that all discrete $1/f$ processes with $\gamma > 0$ have finite high frequency power. At $\gamma = 0$, corresponding to the case of white noise, the multiscale model breaks down. We also see that as γ is increased, there is decreasing power at high frequencies. For extremely negative values of m corresponding to low-frequency component processes,

$$f_m \approx \frac{\sigma^2 \Delta^{(2-\gamma)m}}{2\Delta^{2m} + \sqrt{\Delta^{2m} + 4}} \tag{3.40}$$

and therefore

$$\lim_{m\to-\infty} f_m = \begin{cases} 0, & \gamma < 1 \\ \sigma^2/2, & \gamma = 1 \\ \infty, & \gamma > 1 \end{cases} \tag{3.41}$$

β_m	$\gamma = 0.33$	$\gamma = 1.00$	$\gamma = 1.67$
$1 - 10^{-5}$	0.0002	0.5000	1119.3606
$1 - 10^{-4}$	0.0010	0.5000	239.3150
$1 - 10^{-3}$	0.0049	0.5000	51.1646
$1 - 10^{-2}$	0.0229	0.5000	10.9387
0.9049	0.1068	0.4994	2.3358
0.3820	0.4472	0.4472	0.4472
9.805×10^{-3}	0.4587	0.0981	0.0210
9.998×10^{-5}	0.2187	0.0100	0.0005
1.000×10^{-6}	0.1023	0.0010	0.0000
1.000×10^{-8}	0.0479	0.0001	0.0000
1.000×10^{-10}	0.0224	0.0000	0.0000

Table 3.1: Variances f_m of component processes

implying that discrete $1/f$ processes with parameter $\gamma \geq 1$ have infinite low-frequency power. This infinite low-frequency power suggests that these processes are inherently non-stationary. Additionally, we see that as γ is increased, there is increasing power at low frequencies. Table 3.1 presents a subset of the variances f_m for various values of γ, with $\Delta = 10$.

3.4 Discrete Finite-Order ARMA Model

The model for discrete $1/f$ processes described in the previous section has infinite order, suggesting that an infinite number of state variables would be necessary to completely describe the process over all frequencies. Fortunately, we are generally interested in the process over a finite data length. The spectrum of a finite-length discrete $1/f$ process is effectively bandlimited – the sampling of the process constrains the highest frequency of the spectrum, while the finite data length constrains the lowest frequency of the spectrum. Therefore, an appropriately selected *finite* subset of the infinite component processes is sufficient to model $1/f$ spectral behavior over the frequencies relevant to the data length. The resulting model can then be described by a finite number of state variables.

Generally, we wish to choose the minimum number of retained component processes that exhibit $1/f$ spectral behavior, within a certain error tolerance, over relevant frequencies. The error can be divided into four parts. First, discrete infinite-order model is an approximation to the continuous infinite-order model; the error for this approximation is bounded in Proposition 3.1. Second, the continuous infinite-order model has only nearly-$1/f$ spec-

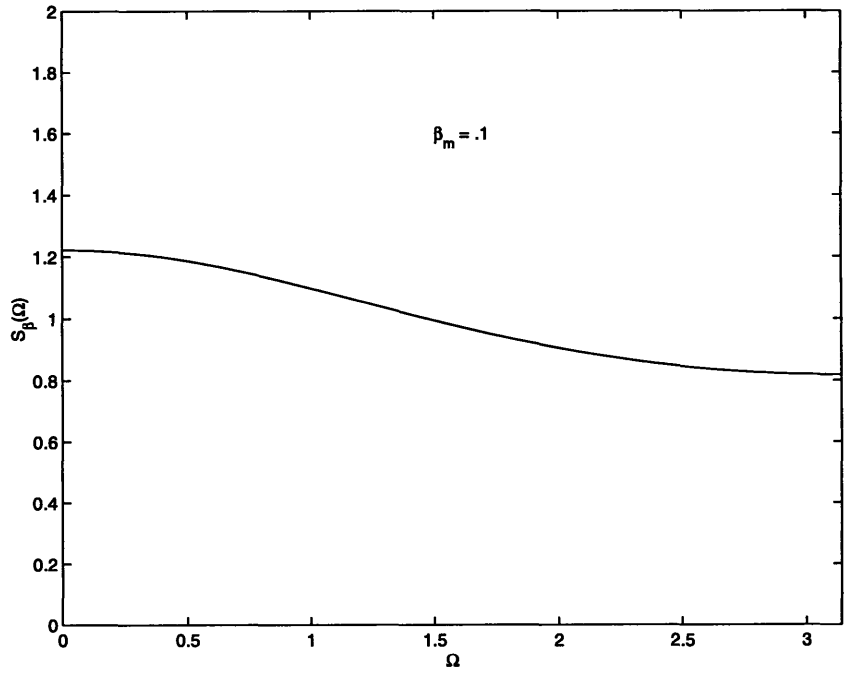


Figure 3-2: Power spectrum $S_{\beta_m}(\Omega)$ of weakly-correlated component processes

tral behavior. This is briefly discussed in Section 3.2. This error is a function of the model parameter Δ. This section discusses the last two parts of the error, which result from the approximation of the infinite-order model with a finite number of component processes.

Weakly-Correlated Component Processes

We refer to a component process as *weakly-correlated* if the long-term correlation of the process is weak, *i.e.* the parameter β_m of the process is near zero. The power of these processes is distributed roughly evenly across all frequencies, as shown in Figure 3-2, in contrast to processes with strong long-term correlation (β_m near 1) in which case the power of the process is predominantly low-frequency, as shown in Figure 3-3.

As described in Section 3.3, a component process with parameters f_m and β_m has a spectral density

$$S_{\beta_m}(\Omega) = \frac{f_m(1-\beta_m^2)}{1+\beta_m^2-2\beta_m\cos\Omega} \tag{3.42}$$

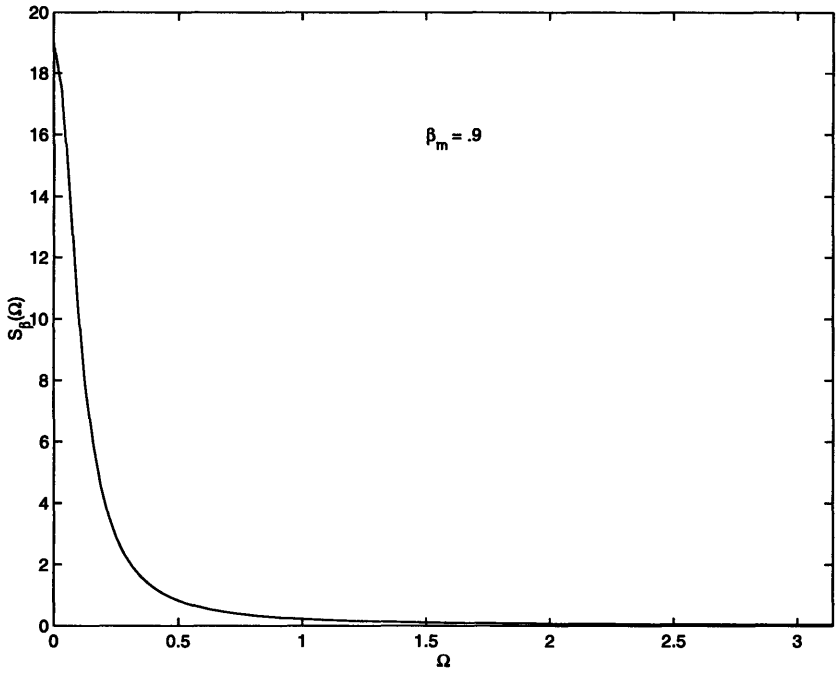


Figure 3-3: Power spectrum $S_{\beta_m}(\Omega)$ of strongly-correlated component processes

which reaches a maximum at $\Omega = 0$ of

$$S_{\beta_m}(0) = \frac{f_m(1-\beta_m^2)}{1+\beta_m^2-2\beta_m} = \sigma^2\Delta^{-\gamma m}. \tag{3.43}$$

We wish to discard all component processes with correlations weaker than a certain threshold, *i.e.* processes corresponding to $m > M_H$, where $m = M_H$ corresponds to the most weakly-correlated component process that is retained. The total power $S_H(\Omega)$ of the discarded weakly-correlated component processes at any frequency Ω is bounded by

$$\begin{aligned} S_H(\Omega) &= \sum_{m=M_H+1}^{\infty} S_{\beta_m}(\Omega) && (3.44)\\ &\leq \sum_{m=M_H+1}^{\infty} \sigma^2\Delta^{-\gamma m} && (3.45)\\ &= \sigma^2\frac{\Delta^{-\gamma(M_H+1)}}{1-\Delta^{-\gamma}} && (3.46) \end{aligned}$$

This suggests that by discarding an additional process, we increase the discarded power $S_H(\Omega)$ by a factor of Δ^{γ}. We can also see that as γ approaches zero corresponding to a

$1/f$ process with weak long-term correlation, the power of the discarded weakly-correlated component processes increases. Generally, as γ increases, we can discard additional weakly-correlated component processes.

The discarded weakly-correlated component processes are each nearly white. Therefore, to further reduce the error from $1/f$ behavior, the discarded processes can be replaced by a single white process $w_H[n]$ with power spectrum

$$S_{w_H}(\Omega) = \sigma^2 \frac{\Delta^{-\gamma(M_H+1)}}{1-\Delta^{-\gamma}}. \tag{3.47}$$

Strongly-Correlated Component Processes

The finite data length constrains the lowest frequency observable in the spectrum of the $1/f$ process. As described by Keshner [11], the finite data length effectively smoothes the frequency response by averaging over a bandwidth inversely proportional to the data length:

$$\Omega_L \sim \frac{2\pi}{N}. \tag{3.48}$$

We are concerned only with frequency behavior above Ω_L due to this smoothing. To minimize the number of component processes necessary to model $1/f$ behavior at low frequencies, we discard any component processes that do not have significant power above Ω_L. We refer to a component process as *strongly-correlated* if the long-term correlation of the process is strong, *i.e.* the parameter β_m of the process is near one. Strongly-correlated processes have predominantly low-frequency power and effectively appear constant over the observation time N. Discarding strongly-correlated processes has the effect of changing the apparent steady value of the process. In the frequency domain, discarding strongly-correlated component process has the effect of generating a flat (instead of $1/f$) spectrum below a certain frequency. Figure 3-4 illustrates the frequency spectrum of a finite-order nearly-$1/f$ process with parameter $\gamma = 1$. Essentially, the goal is to guarantee that the roll-off between $1/f$ behavior and flat low-frequency behavior occurs below the observable frequency Ω_L.

We first examine the low-frequency behavior of a single component process with characteristic time-constants $1/\beta_m$. As described in the previous section, this process will have

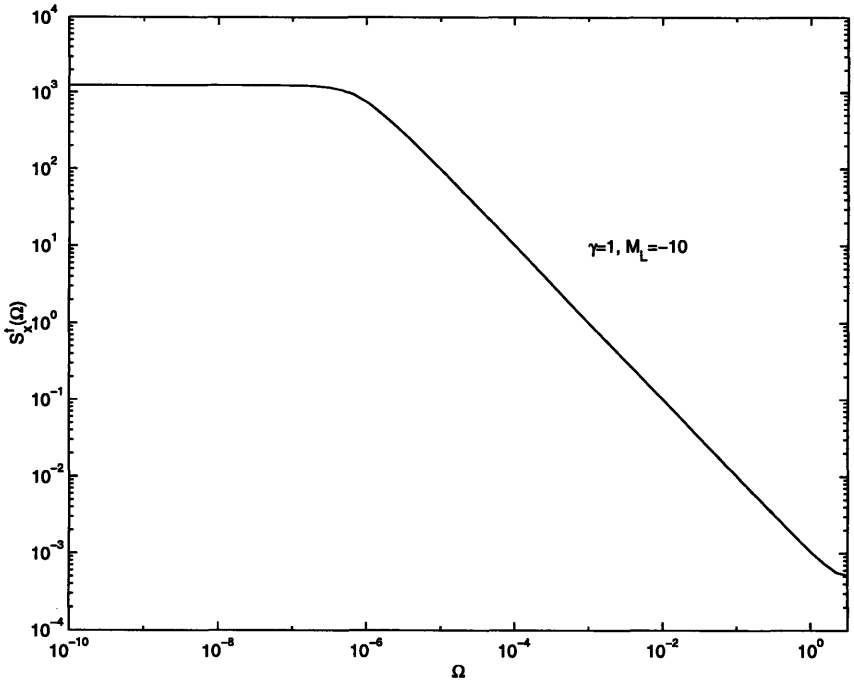


Figure 3-4: Finite-order nearly-$1/f$ process spectrum.

a spectral density of the form

$$S_{\beta_m}(\Omega) = \frac{f_m(1-\beta_m^2)}{1+\beta_m^2-2\beta_m\cos\Omega} \approx \frac{\sigma^2\Delta^{(2-\gamma)m}}{\Omega^2+\Delta^{2m}} \tag{3.49}$$

where the approximation is valid at low frequencies. For this section, we assume that this approximation holds for the frequencies of interest. The spectral density of the process is bounded by the inequality

$$S_{\beta_m}(\Omega) < \frac{\sigma^2\Delta^{(2-\gamma)m}}{\Omega^2} \tag{3.50}$$

over the relevant frequencies. We note that as m decreases, the corresponding process will generally have decreasing power in this frequency region. We wish to discard all component processes corresponding to $m < M_L$, where $m = M_L$ corresponds the lowest-frequency process that we retain in our finite-order model. The total power $S_L(\Omega_L)$ of the discarded component processes at frequency Ω_L is

$$S_L(\Omega_L) = \sum_{m=-\infty}^{M_L-1} S_{\beta_m}(\Omega_L) \tag{3.51}$$

$$< \frac{\sigma^2}{\Omega_L^2} \cdot \frac{\Delta^{(2-\gamma)(M_L-1)}}{1-\Delta^{-(2-\gamma)}}. \tag{3.52}$$

This suggests that discarding an additional component process will increase the discarded power $S_L(\Omega_L)$ at low frequencies by a factor of $\Delta^{2-\gamma}$. We can also see that as γ approaches 2 corresponding to a $1/f$ process with strong long-term correlation, the power of the discarded strongly-correlated component processes increases. Generally, as γ decreases, we can discard additional strongly-correlated processes.

Additionally, if we increase the data length by a factor of k, the discarded power $S_L(\Omega_L)$ at low frequencies will increase by a factor of k^2, approximately. In order to maintain the same discarded power while increasing data length by k, we should retain $\frac{2}{2-\gamma}\log_\Delta k$ additional component processes.

3.4.1 Error Analysis

The bounds given by (3.46) and (3.52) can provide a measure of the error of a finite-order model with component processes corresponding to $M_L \leq m \leq M_H$ from the infinite-order model. Denote the spectrum of the finite-order model

$$S_x^f(\Omega) = \sum_{m=M_L}^{M_H} S_{\beta_m}(\Omega) \tag{3.53}$$

which will always have less power than the infinite-order model $S_x(\Omega)$.

We establish an approximate bound on the worst-case percentage error in the spectrum due to discarding the high-frequency processes by examining the power of the finite-order spectrum at frequency $\Omega = \pi$ and comparing to the upper bound of the power $S_H(\Omega)$ of the discarded high-frequency processes given in (3.46). We can further reduce this error by replacing the discarded high-frequency processes with a single white ($\beta_m = 0$) component process with power equivalent to the discarded processes.

Similarly, we establish an approximate bound on the worst-case percentage error in the spectrum (over relevant frequencies) due to discarding the low-frequency processes by examining the power of the finite-order spectrum at frequency $\Omega = \Omega_L$ and comparing to the upper bound of the power $S_L(\Omega_L)$ of the discarded low-frequency processes given in (3.52).

We give several examples of finite-order models for $1/f$ processes for various values of γ.

	$\epsilon = 0.05$		$\epsilon = 0.01$		$\epsilon = 0.001$	
γ	M_L	M_H	M_L	M_H	M_L	M_H
0.33	-5	7	-6	10	-7	15
1.00	-7	2	-8	3	-10	5
1.67	-11	1	-15	2	-20	3

Table 3.2: Finite component process choices satisfying error bound ϵ

We measure the worst-case percentage error at high and low frequencies from the discrete infinite-order model. For this section, we choose $\Delta = 4$ and $\Omega_L = 10^{-3}$ for the purposes of illustration.

We wish to find a choice of finite component processes so that the worst-case percentage errors are each below a certain error tolerance ϵ, *i.e.*,

$$\frac{S_H(\Omega)}{S_x^f(\pi) + S_H(\Omega)} < \epsilon \tag{3.54}$$

$$\frac{S_L(\Omega_L)}{S_x^f(\Omega_L) + S_L(\Omega_L)} < \epsilon \tag{3.55}$$

where $S_H(\Omega)$ and $S_L(\Omega_L)$ are bounded by (3.46) and (3.52). Table 3.2 gives examples of appropriate choices of component processes for several values of γ and ϵ.

3.4.2 Effective Signal Power

The variance of a $1/f$ signal is not an ideal measure of the signal's power over finite data lengths. For instance, $1/f$ signals with parameter $\gamma > 1$ have infinite variance, although over a finite data length the process varies only within a comparatively small range relative to its time average. Additionally, the variance of a process described with the finite model is strongly dependent on the number of scales used in the description.

The preceding sections have suggested that only frequencies above a certain frequency Ω_L are relevant when the data has finite length. Therefore, we are primarily interested in the signal's power over the relevant frequencies. We define

$$\sigma_x^2 = \frac{1}{\pi}\int_{\Omega_L}^{\pi} S_x^f(\Omega)d\Omega \tag{3.56}$$

$$= \sum_{m=M_L}^{M_H} f_m \left(1 - 2\arctan\left(\frac{(1+\beta_m)^2\tan(\Omega_L/2)}{1-\beta_m^2}\right)\right) \tag{3.57}$$

as the effective signal power over frequencies above Ω_L. This measure of signal power has the

useful property that it depends only slightly on the number of scales used in the description, as opposed to the variance of process. We use this notion of signal power when analyzing the performance of algorithms in Chapters 5 and 6.

3.4.3 Notational Conventions

We denote the number of states represented in the finite-order model as $M = M_H - M_L + 1$. For notational convenience in future chapters, we represent the finite spectrum

$$S_x^f(\Omega) = \sum_{m=M_L}^{M_H} S_{v_m}(\omega) = \sum_{m=1}^{M} S_{x_m}(\omega) \tag{3.58}$$

where $S_{x_m}(\Omega)$ is simply a reindexing of $S_{v_m}(\Omega)$, *i.e.* $x_m[n] = v_{m+M_L-1}$ with corresponding reindexed time constants $\tau_m = \beta_{m+M_L-1}$ and reindexed weights $g_m = f_{m+M_L-1}$.

Chapter 4

State Estimation with Discrete $1/f$ Processes

The linear, finite order models described in Section 3.4 are naturally represented by state-space system descriptions. This representation is desirable for estimation algorithms with $1/f$ processes due to the computational efficiency of recursive estimation using the Kalman filter and smoother. In addition, the state-space system description accounts for the presence of additive broadband measurement noise. In this chapter, we present two descriptions for $1/f$ processes which will be used for applications in parameter estimation, prediction, and smoothing in Chapters 5 and 6. We review the discrete-time Kalman filtering and Kalman smoothing equations that will be used in these applications.

A linear time-invariant single-output system may be described by the discrete-time state evolution and observation equations [1]

$$\boldsymbol{x}[n+1] = A\boldsymbol{x}[n] + B\boldsymbol{u}[n] \tag{4.1}$$

$$z[n] = C\boldsymbol{x}[n] + w[n] \tag{4.2}$$

for $n \geq 0$. The system state vector $\boldsymbol{x}[n]$ evolves in time according to the state evolution equation (4.1) which is driven by the input process $\boldsymbol{u}[n]$. The observation equation (4.2) describes the measurement process $z[n]$ as a linear transformation of the state vector obscured by a noise process $w[n]$. The input process and measurement process are assumed to be white, indepedendent from each other, and zero mean Gaussian processes with known

covariances. In addition, we specify a stochastic initial condition $\boldsymbol{x}[0]$ with known mean $\overline{\boldsymbol{x}}[0]$ and known covariance $R_x[0]$.

4.1 First-Order Autogressive Process

Each component process of the multiscale model is a wide-sense stationary first-order autoregressive (AR) process $x_m[n]$ parameterized by time-constant $1/\tau_m$ and weight g_m, with correlation function

$$R_{x_m}[k] = g_m \tau_m^{|k|}$$

which can be generated using initial condition $x_m[0] \sim N(0, g_m)$ and filtering equation

$$x_m[n] - \tau_m x_m[n-1] = \left(g_m(1-\tau_m^2)\right)^{1/2} u_m[n] \tag{4.3}$$

where the input process $u_m[n]$ is a white zero-mean Gaussian process with unit variance. For convenience, we define a new parameter $h_m = \left(g_m(1-\tau_m^2)\right)^{1/2}$. A state-space description for this first-order process with a one-dimensional state vector $x_m[n]$ is

$$\begin{aligned} x_m[n+1] &= \tau_m x_m[n] + h_m u_m[n] & (4.4) \\ z[n] &= x_m[n] & (4.5) \end{aligned}$$

with initial condition $x_m[0]$ having zero mean and covariance $R_{x_m}[0] = g_m$.

It will prove useful to equivalently describe the first order AR process using additional state variables. This concept of augmenting the signal model with additional states has been applied to several smoothing applications [1]. A state-space description for this first-order process with a two-dimensional state vector is

$$\begin{aligned} \boldsymbol{x_m}[n+1] &= \begin{bmatrix} \tau_m & 0 \\ 1 & 0 \end{bmatrix} \boldsymbol{x_m}[n] + \begin{bmatrix} h_m \\ 0 \end{bmatrix} u_m[n] & (4.6) \\ z[n] &= \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{x_m}[n] & (4.7) \end{aligned}$$

with initial condition $\boldsymbol{x_m}[0]$ having zero mean and covariance

$$R_{x_m}[0] = \begin{bmatrix} g_m & \tau_m g_m \\ \tau_m g_m & g_m \end{bmatrix}. \tag{4.8}$$

The state vector may be heuristically viewed as $\boldsymbol{x_m}[n] = \begin{bmatrix} x_m[n] & x_m[n-1] \end{bmatrix}^T$. In contrast with the single-state description, the information about the correlation between adjacent time samples is explicitly available.

4.2 Multiscale Model for $1/f$ Processes

A $1/f$ process described by the multiscale model presented in Chapter 3 is composed of the superposition of M uncorrelated component processes, each of which is described in the previous section. As each component process is a first-order AR process that may be described with either one or two states, the $1/f$ process can be described with either an M-state system or a $2M$-state system. In Chapter 5, the $2M$-state system is used for parameter estimation algorithm. In Chapter 6, the M-state system description is used for prediction and smoothing applications.

As described in Section 3.4, the M state model representing the $1/f$ process $x[n]$ has M component processes $\{x_1[n], \ldots, x_M[n]\}$ with governing time-constants $1/\tau_m$ and weights g_m. The state vector for the $1/f$ process consists of the total of all the component state vectors

$$\boldsymbol{x}[n] = \begin{bmatrix} \boldsymbol{x}_1[n] \\ \vdots \\ \boldsymbol{x}_N[n] \end{bmatrix} \tag{4.9}$$

where $\boldsymbol{x}_i[n]$ is a one-state or two-state vector corresponding to a first-order AR component process. The $1/f$ process is driven by a white M-dimensional Gaussian input vector

$$\boldsymbol{u}[n] = \begin{bmatrix} u_1[n] & u_2[n] & \ldots & u_M[n] \end{bmatrix}^T \tag{4.10}$$

with zero mean and covariance

$$R_{uu} = \mathrm{E}\left\{\boldsymbol{u}[n]\boldsymbol{u}[n]^T\right\} = I_{M\times M}$$

corresponding to M uncorrelated noise processes, each with unit variance.

The measurement process $z[n]$ is composed of the $1/f$ process corrupted by additive white measurement noise $w[n]$, described by the state-space equations

$$\boldsymbol{x}[n+1] = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_M \end{bmatrix} \boldsymbol{x}[n] + \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_M \end{bmatrix} \boldsymbol{u}[n] \tag{4.11}$$

$$z[n] = \begin{bmatrix} C_1 & \dots & C_M \end{bmatrix} \boldsymbol{x}[n] + w[n] \tag{4.12}$$

where $w[n]$ is a zero-mean white Gaussian noise sequence with variance σ_w^2. Each matrix is composed of component matrices corresponding to the component first-order AR process described in the previous section. For the M state system,

$$A_m = \tau_m, \quad B_m = h_m, \quad C_m = 1 \tag{4.13}$$

for $1 \leq m \leq M$. For the $2M$ state system,

$$A_m = \begin{bmatrix} \tau_m & 0 \\ 1 & 0 \end{bmatrix}, \quad B_m = \begin{bmatrix} h_m \\ 0 \end{bmatrix}, \quad C_m = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{4.14}$$

for $1 \leq m \leq M$. The initial condition $\boldsymbol{x}[0]$ is a zero-mean random vector with covariance

$$R_x[0] = \begin{bmatrix} R_{x_1}[0] & & 0 \\ & \ddots & \\ 0 & & R_{x_M}[0] \end{bmatrix} \tag{4.15}$$

where

$$R_{x_m}[0] = g_m \tag{4.16}$$

for the M state system, and

$$R_{x_m}[0] = \begin{bmatrix} g_m & \tau_m g_m \\ \tau_m g_m & g_m \end{bmatrix} \tag{4.17}$$

for the $2M$ state system.

4.3 Kalman Filter

In this section, we review the discrete-time Kalman filter, which will be used in applications of prediction described in Sections 6.1 and 6.2. The Kalman filter is a computationally efficient, recursive state estimator for the linear state-space model presented in the previous section. The state estimate is based on previous observations and is optimal in the minimum mean-squared error (MMSE) sense.

Given a general state-space system described by (4.1) and (4.2), the observation set

$$Z[n] = \{\boldsymbol{z}[0], \boldsymbol{z}[1], \ldots, \boldsymbol{z}[n]\} \tag{4.18}$$

consists of observations up to and including time n. The state estimation problem involves estimating the state $\boldsymbol{x}[n]$ given the observation set $Z[n]$. The Kalman filter recursively computes the state estimate $\hat{\boldsymbol{x}}[n]$ and the error covariance matrix

$$R_x[n] = \mathrm{E}\left\{(\boldsymbol{x}[n] - \hat{\boldsymbol{x}}[n])(\boldsymbol{x}[n] - \hat{\boldsymbol{x}}[n])^T \mid Z[n]\right\} \tag{4.19}$$

in two steps. In the propagation or prediction step, the state estimate and error covariance matrix given data up to time n are computed for time $n+1$, using the state estimate and error covariance matrix for time n. We denote the computed state estimate $\hat{\boldsymbol{x}}[n+1|n]$ and computed error covariance matrix $R_x[n+1|n]$, emphasizing that these quantities are based on observations up to time n. In the measurement update step, the new observation $y[n+1]$ is incorporated to generate the state estimate and error covariance matrix for time $n+1$. We denote the computed state estimate $\hat{\boldsymbol{x}}[n+1|n+1]$ and computed error covariance matrix $R_x[n+1|n+1]$, emphasizing that these quantities are based on observations up to time $n+1$.

The Kalman filter equations [1] are given in Table 4.1. The initialization step assumes that the state variables are zero mean with steady-state covariance $R_x[0]$. We then alternate between the update step and propagation step until we have generated the desired quantities. For state estimation algorithms, we are interested in $\hat{\boldsymbol{x}}[n|n]$ and $R_x[n|n]$. On the other hand, for the prediction algorithms of Chapter 6, we are interested in $\hat{\boldsymbol{x}}[n+1|n]$ and $R_x[n+1|n]$.

Initialization

$$\hat{\boldsymbol{x}}[0|-1] = 0 \tag{4.20}$$
$$R_x[0|-1] = R_x[0] \tag{4.21}$$

Propagation Equations

$$\hat{\boldsymbol{x}}[n+1|n] = A\hat{\boldsymbol{x}}[n|n] \tag{4.22}$$
$$R_x[n+1|n] = AR_x[n|n]A^T + BB^T \tag{4.23}$$

Update Equations

$$\hat{\boldsymbol{x}}[n+1|n+1] = \hat{\boldsymbol{x}}[n+1|n] + K_{n+1}\left(z[n+1] - C\hat{\boldsymbol{x}}[n+1|n]\right) \tag{4.24}$$
$$R_x[n+1|n+1] = \left(I - K_{n+1}C\right)R_x[n+1|n] \tag{4.25}$$

where

$$K_{n+1} = \frac{1}{CR_x[n+1|n]C^T + \sigma_w^2} R_x[n+1|n]C^T \tag{4.26}$$

is denoted the Kalman gain.

Table 4.1: Kalman Filter Equations

4.4 Kalman Smoother

In this section, we review the Kalman smoother, which will be applied to problems of parameter estimation in Chapter 5 and signal estimation in Section 6.3. We focus on the fixed-interval smoothing case [1], involving estimation of the state vector $\hat{\boldsymbol{x}}[n]$ for all times $0 \leq n \leq N$, based on the observation set

$$Z[N] = \{\boldsymbol{z}[0], \boldsymbol{z}[1], \ldots, \boldsymbol{z}[N]\}$$

consisting of observations up to and including time N. The Kalman smoother is optimal in the MMSE sense.

Again, we assume a general state-space system described by (4.1) and (4.2). The Kalman smoother computes the state estimates

$$\hat{\boldsymbol{x}}[n|N] = \mathrm{E}\left\{\boldsymbol{x}[n] \mid Z[N]\right\} \tag{4.27}$$

and the error covariance matrix

$$R_x[n|N] = \mathrm{E}\left\{(\boldsymbol{x}[n] - \hat{\boldsymbol{x}}[n|N])(\boldsymbol{x}[n] - \hat{\boldsymbol{x}}[n|N])^T \mid Z[N]\right\} \tag{4.28}$$

in three steps. The propagation and update steps are performed in the same manner as in the Kalman filter. A third smoothing step consists of a backward pass updating the state estimates with future observations.

The Kalman smoother equations [1] compute $\hat{\boldsymbol{x}}[n|N]$ and $R_x[n|N]$ for $0 \leq n \leq N$ according to the algorithm in Table 4.2. As before, the initialization step assumes that the state variables are zero mean with steady-state covariance $R_x[0]$.

Initialization

$$\hat{\boldsymbol{x}}[0|-1] = 0 \tag{4.29}$$
$$R_x[0|-1] = R_x[0] \tag{4.30}$$

For $n = 0, 1, 2, \ldots, N$, perform the following two steps:

Update Equations

$$\hat{\boldsymbol{x}}[n|n] = \hat{\boldsymbol{x}}[n|n-1] + K_n \left(z[n] - C\hat{\boldsymbol{x}}[n|n-1]\right) \tag{4.31}$$
$$R_x[n|n] = \left(I - K_n C\right) R_x[n|n-1] \tag{4.32}$$

where

$$K_n = \frac{1}{CR_x[n|n-1]C^T + \sigma_w^2} R_x[n|n-1]C^T \tag{4.33}$$

is denoted the Kalman gain.

Propagation Equations

$$\hat{\boldsymbol{x}}[n+1|n] = A\hat{\boldsymbol{x}}[n|n] \tag{4.34}$$
$$R_x[n+1|n] = AR_x[n|n]A^T + BB^T \tag{4.35}$$

For $n = N-1, N-2, \ldots, 0$ perform the following step:

Smoothing Equations

$$\hat{\boldsymbol{x}}[n|N] = \hat{\boldsymbol{x}}[n|n] + S_n \left(\hat{\boldsymbol{x}}[n+1|N] - A\hat{\boldsymbol{x}}[n|n]\right) \tag{4.36}$$
$$R_x[n|N] = R_x[n|n] + S_n \left(R_x[n+1|N] - R_x[n+1|n]\right) S_n^T \tag{4.37}$$

where

$$S_n = R_x[n|n]A^T R_x[n+1|n]^{-1} \tag{4.38}$$

Table 4.2: Kalman Smoothing Equations

Chapter 5

Parameter Estimation for $1/f$ Signals

Since the parameters of the $1/f$ signal are typically not known *a priori*, algorithms for parameter estimation are essential. In a typical scenario, only a finite-length observation sequence of a $1/f$ signal with unknown parameters is available. Additionally, these observations are invariably corrupted by broadband measurement noise. Some other approaches to parameter estimation for alternative models of $1/f$ behavior are given in [5] [13] and [20].

In this first part of this chapter, we consider the problem of jointly estimating signal and noise parameters for the case of Gaussian $1/f$ processes corrupted by stationary white Gaussian noise. We develop algorithms for Maximum Likelihood (ML) parameter estimation exploiting the state-space model and Kalman smoothing equations described in Chapter 4. The performance of the estimator is analyzed with Monte Carlo simulations on synthetic data.

In the second part of this chapter, we consider problems in which the $1/f$ process obscures a deterministic signal of interest. The deterministic signal is parameterized as a linear combination of a finite set of basis signals. We develop algorithms for ML parameter estimation of the deterministic signal jointly with the corrupting $1/f$ process and white Gaussian noise. Of particular interest are cases involving affine signals, which have application in the analysis of economic time-series and ocean temperature data. We examine in more detail the special case in which the parameters of the $1/f$ noise are known.

5.1 $1/f$ Signal in White Gaussian Noise

In this section, we consider the problem in which we have observations $z[n]$ of a discrete zero-mean Gaussian $1/f$ process $x[n]$ corrupted by zero-mean independent identically distributed (i.i.d.) Gaussian noise $w[n]$ that is statistically independent of $x[n]$, so

$$z[n] = x[n] + w[n], \quad 0 \leq n \leq N-1 \tag{5.1}$$

where N is the length of the observed data. Each finite length sequence may be viewed as an N-length column vector

$$\boldsymbol{z} = \begin{bmatrix} z[0] \\ \vdots \\ z[N-1] \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x[0] \\ \vdots \\ x[N-1] \end{bmatrix}, \quad \boldsymbol{w} = \begin{bmatrix} w[0] \\ \vdots \\ w[N-1] \end{bmatrix}. \tag{5.2}$$

The goal is to estimate the vector of unknown parameters

$$\boldsymbol{\theta} = \{\gamma, \sigma^2, \sigma_w^2\} \tag{5.3}$$

where $0 < \gamma < 2$ and $\sigma^2 > 0$ are parameters of the $1/f$ process $x[n]$ and $\sigma_w^2 > 0$ is the variance of the i.i.d. additive Gaussian noise process $w[n]$.

The observed data is a Gaussian random vector of length N with probability density function (p.d.f.)

$$f_Z(\boldsymbol{z}; \boldsymbol{\theta}) = [\det(2\pi\Lambda_Z(\boldsymbol{\theta}))]^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{z}^T\Lambda_Z^{-1}(\boldsymbol{\theta})\boldsymbol{z}\right] \tag{5.4}$$

indexed by the vector of unknown parameters $\boldsymbol{\theta}$. The $N \times N$ covariance matrix of the observation vector is given by

$$\Lambda_Z(\boldsymbol{\theta}) = \Lambda_X(\gamma, \sigma^2) + \Lambda_W(\sigma_w^2) \tag{5.5}$$

where $\Lambda_X(\gamma, \sigma^2)$ is the covariance matrix of the $1/f$ signal vector and $\Lambda_W(\sigma_w^2)$ is the covariance matrix of the additive Gaussian noise vector. The covariance of the $1/f$ process is

given by

$$\Lambda_X(\gamma, \sigma^2) = \begin{bmatrix} R_X[0] & R_X[1] & \dots & R_X[N-1] \\ R_X[1] & R_X[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_X[1] \\ R_X[N-1] & \dots & R_X[1] & R_X[0] \end{bmatrix} \tag{5.6}$$

where $R_X[n]$ is the covariance function of the $1/f$ process with parameters γ and σ^2, as given in Chapter 3. The covariance of the additive Gaussian noise process is simply

$$\Lambda_W(\sigma_w^2) = \sigma_w^2 I \tag{5.7}$$

where I is the $N \times N$ identity matrix.

The ML estimate $\hat{\boldsymbol{\theta}}_{ML}$ of $\boldsymbol{\theta}$ is defined by

$$\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta} \in \Theta} \log f_Z(\boldsymbol{z}; \boldsymbol{\theta}) \tag{5.8}$$

where Θ represents the region over which the parameters are allowed to range. Solving directly for the ML estimate is generally a difficult multiparameter optimization problem; in addition, the calculation of the inverse in (5.4) is computationally intensive.

The structure of the discrete $1/f$ model suggests using an iterative estimate-maximize (EM) algorithm [4] to simplify computations. As described in Section 3.3, the discrete finite multiscale model represents the $1/f$ signal as the superposition of M uncorrelated single time-constant processes

$$x[n] = \sum_{m=1}^{M} x_m[n] \tag{5.9}$$

which have correlation functions

$$R_m[k] = g_m(\gamma, \sigma^2) \tau_m^{-|k|}. \tag{5.10}$$

We assume that the observations are well-represented by the finite-order model and that the order parameters $\{\Delta, M_L, M_H\}$ are known. As described in Section 3.4, this assumption implies that the spectral behavior of the observations is $1/f$ only over a known, finite frequency range. In practical scenarios, this may not always be a realistic assumption; we discuss this further in Section 5.1.2.

The EM algorithm for this problem efficiently estimates the unknown parameters by estimating the statistics of these component processes. A detailed derivation of the EM algorithm for this problem is presented in Appendix B.1. The algorithm is described below and summarized in Table 5.1 in terms of $\gamma^{[l]}$, $\sigma^{2[l]}$, and $\sigma_w^{2[l]}$ which denote the estimates of parameters γ, σ^2, and σ_w^2, respectively, that are generated on the lth iteration of the algorithm. The EM algorithm is guaranteed to converge to a stationary point of the likelihood function.

E Step

The E step calculates the statistics of the complete data, which includes the component states as well as information about the white noise variance. These expectations can be efficiently computed using the Kalman smoothing equations in Table 4.2 with a $2M$-state system described by equations (4.11), (4.12), (4.14) and (4.17), representing the $1/f$ signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$ and white noise with parameter $\sigma_w^{2[l]}$.

M Step

The M step updates the parameter estimate by performing the maximization

$$\max_{\boldsymbol{\theta}} U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]}) \rightarrow \boldsymbol{\theta}^{[l+1]}. \tag{5.19}$$

The updated white noise variance estimate is straightforward, while the updated $1/f$ signal parameters require more complicated algebraic manipulation. Due to the tridiagonal structure of H_m^{-1}, only certain elements of

$$\widehat{\boldsymbol{x}_m \boldsymbol{x}_m^T} = \mathrm{E}\left\{ \boldsymbol{x}_m \boldsymbol{x}_m^T \mid \boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\theta}^{[l]} \right\} \tag{5.20}$$

corresponding to those calculated in the E step are necessary. A root-finding algorithm such as bracketing and bisection can be used to solve (5.15).

5.1.1 Special Case of Known Noise Statistics

If the variance of the white noise σ_w^2 is known, the algorithm simplifies. During the E step, the Kalman smoothing is performed using the known noise parameter σ_w^2 rather than the

E Step

$$\hat{x}_m^{[l]}[n] = \mathrm{E}\left\{x_m[n] \mid Z[N]; \boldsymbol{\theta}^{[l]}\right\} \tag{5.11}$$

$$\widehat{(x_m[n]x_k[n])}^{[l]} = \mathrm{E}\left\{(x_m[n] - \hat{x}_m[n])(x_k[n] - \hat{x}_k[n]) \mid Z[N]; \boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_k[n], \quad k = 1, \ldots, M \tag{5.12}$$

$$\widehat{(x_m[n]x_m[n-1])}^{[l]} = \mathrm{E}\left\{(x_m[n] - \hat{x}_m[n])\,(x_m[n-1] - \hat{x}_m[n-1]) \mid Z[N]; \boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_m[n-1], \quad n = 1, \ldots, N-1 \tag{5.13}$$

for $m = 1, \ldots, M$.

M Step

$$\sigma_w^{2[l+1]} = \frac{1}{N}\sum_{n=0}^{N-1}\left[z^2[n] - 2z[n]\sum_{m=1}^{M}\hat{x}_m^{[l]}[n] + \sum_{m=1}^{M}\sum_{k=1}^{M}\widehat{(x_m[n]x_k[n])}^{[l]}\right] \tag{5.14}$$

$$\gamma^{[l+1]} \longleftarrow \sum_{m=1}^{M}\left(m - \frac{M+1}{2}\right)\frac{\Delta^{\gamma^{[l+1]}(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) = 0 \tag{5.15}$$

$$\sigma^{2[l+1]} = \frac{1}{NM}\sum_{m=1}^{M}\frac{\Delta^{\gamma^{[l+1]}(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) \tag{5.16}$$

where

$$q_m(\Delta) = \frac{\Delta^{2m}}{\tau_m - \tau_m^{-1}} \tag{5.17}$$

$$H_m^{-1} = \frac{1}{1-\tau_m^2}\cdot\begin{bmatrix} 1 & -\tau_m & & & 0 \\ -\tau_m & 1+\tau_m^2 & -\tau_m & & \\ & \ddots & \ddots & \ddots & \\ & & -\tau_m & 1+\tau_m^2 & -\tau_m \\ 0 & & & -\tau_m & 1 \end{bmatrix}. \tag{5.18}$$

Table 5.1: EM algorithm for parameter estimation of $1/f$ signal in white noise of unknown statistics

E Step

$$\hat{x}_m^{[l]}[n] = \mathrm{E}\left\{x_m[n] \mid Z[N]; \boldsymbol{\theta}^{[l]}\right\} \tag{5.21}$$

$$\begin{aligned}(x_m[n]x_m[n-1])^{\wedge[l]} &= \mathrm{E}\left\{(x_m[n] - \hat{x_m}[n])\,(x_m[n-1] - \hat{x_m}[n-1]) \mid Z[N]; \boldsymbol{\theta}^{[l]}\right\} \\ &\quad + \hat{x}_m[n]\hat{x}_m[n-1], \quad n = 1, \ldots, N-1\end{aligned} \tag{5.22}$$

for $m = 1, \ldots, M$.

M Step

$$\gamma^{[l+1]} \longleftarrow \sum_{m=1}^{M} \left(m - \frac{M+1}{2}\right) \frac{\Delta^{\gamma^{[l+1]}(m+M_L-1)}}{q_m(\Delta)} \mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) = 0 \tag{5.23}$$

$$\sigma^{2[l+1]} = \frac{1}{NM} \sum_{m=1}^{M} \frac{\Delta^{\gamma^{[l+1]}(m+M_L-1)}}{q_m(\Delta)} \mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) \tag{5.24}$$

where

$$q_m(\Delta) = \frac{\Delta^{2m}}{\tau_m - \tau_m^{-1}} \tag{5.25}$$

$$H_m^{-1} = \frac{1}{1-\tau_m^2} \cdot \begin{bmatrix} 1 & -\tau_m & & & 0 \\ -\tau_m & 1+\tau_m^2 & -\tau_m & & \\ & \ddots & \ddots & \ddots & \\ & & -\tau_m & 1+\tau_m^2 & -\tau_m \\ 0 & & & -\tau_m & 1 \end{bmatrix}. \tag{5.26}$$

Table 5.2: EM algorithm for parameter estimation of $1/f$ signal in white noise of known statistics

estimated parameter $\sigma_w^{2[l]}$. In the M step, we do not update the noise parameter σ_w^2. This case of the algorithm is summarized in Table 5.2. As before, due to the tridiagonal structure of H_m^{-1}, only those elements of $\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}$ corresponding to the quantities calculated in the E step are necessary. The EM algorithm converges substantially faster in the case of known noise statistics.

5.1.2 Model Order Effects

The model order parameters $\{\Delta, M_L, M_H\}$ for the parameter estimation algorithm are assumed to be appropriate for the observation data sequence. In Sections 3.2 and 3.4, it was shown that parameter Δ governed the amplitude and ripple of the nearly-$1/f$ behavior

of the model, while parameters M_L and M_H governed the low and high frequency behavior, respectively, of the discrete $1/f$ model. This section discusses the effects of mismatches between true and assumed model order parameters.

The Maximum-Likelihood parameter estimates as described in (5.4) and (5.8) depend on the covariance Λ_Z of the $1/f$ process model. In most scenarios involving real data, the observed sequence exhibits $1/f$ behavior over all frequencies. In this case, the most appropriate choice of order for the observed data sequence would be the infinite-order model given in Section 3.3. Therefore, the finite-order model should closely approximate the covariance of the infinite-order model. First, we consider the sensitivity of the parameter estimation to different choices of M_H. Recall from (3.38) that the variance of high frequency processes decreases exponentially, *i.e.* $f_m \approx \sigma^2 \Delta^{-\gamma m}$. In this case, we can approach the covariance of the infinite-order model as closely as desired. Therefore the parameter estimates will converge to a fixed value as M_H is increased. Figure 5-1 plots the RMS error in estimating γ as the order parameter M_H of the estimator is varied, with M_L fixed. The results of 64 trials were averaged to obtain these results. In each trial, the simulated data of length 50 is generated using the parameter M_L and M_{Hgen} shown in the figure. Parameter estimation is then performed on the data using various values of M_H in the estimation algorithm.

Next, we consider the sensitivity of the parameter estimation algorithm to different choices of M_L. Recall from Section 3.3.2 that for $\gamma < 1$, the low-frequency component processes have decreasing power f_m as m decreases. Again, the parameter estimates will converge to a fixed value as M_L is decreased. However, for the case $\gamma \geq 1$, the low-frequency component processes have non-decreasing power as m decreases; consequently the covariance of a finite-order $1/f$ process model does not converge. The parameter estimates have a strong dependence on M_L in this case. In general, decreasing M_L adds low-frequency component processes to the estimator and decreases the estimate of γ. Increasing M_L discards low-frequency component processes and increases the estimate of γ. This effect becomes more pronounced as γ is increased. Figure 5-2 plots the RMS error in estimating γ as the order parameter M_L of the estimator is varied, with M_H fixed. The results of 64 trials were averaged to obtain these results. In each trial, the simulated data of length 50 is generated using the parameter M_{Lgen} and M_H shown in the figure. Parameter estimation is then performed on the data using various values of M_L in the estimation algorithm.

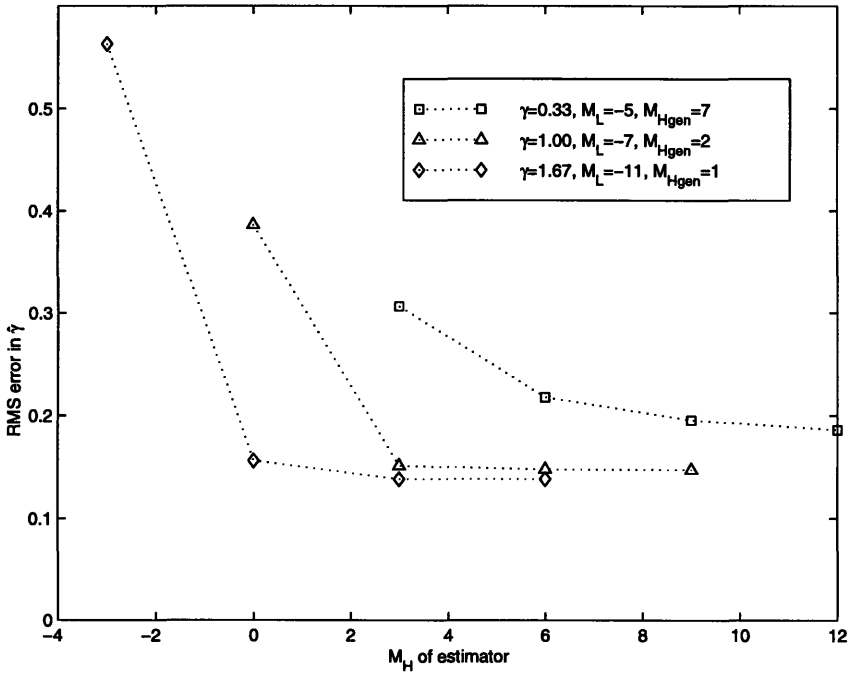


Figure 5-1: RMS error in estimates of γ as a function of order parameter M_H

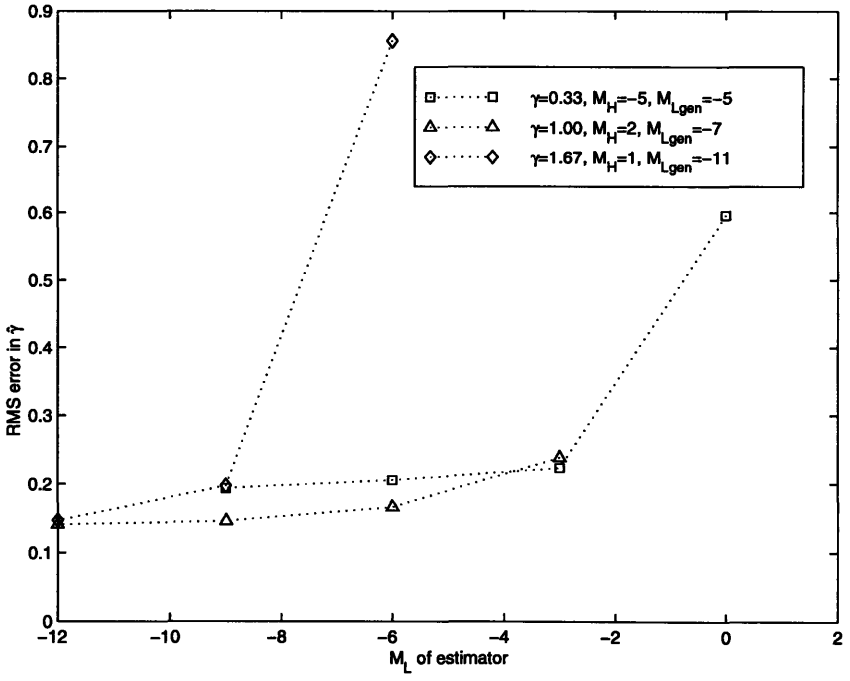


Figure 5-2: RMS error in estimates of γ as a function of order parameter M_L

Although a $1/f$ process with parameter $\gamma \geq 1$ has infinite power, a finite-length observation of the process has finite power. We assume that the data spectrum exhibits a rolloff between $1/f$ and white behavior at some frequency. A proper choice of M_L should generate a model with a spectrum exhibiting the same rolloff frequency as the data spectrum; however, this rolloff frequency cannot be easily determined. We therefore expect that the choice of M_L will not be optimal, and consequently the parameter estimation algorithm will be biased.

The choice of Δ determines the amplitude and frequency spacing of the ripple superimposed upon the $1/f$ spectrum, as discussed in Section 3.2. As $\Delta \to 1$, the finite-order model converges to ideal $1/f$ behavior; consequently the parameter estimates converge to a fixed value. However, as Δ is decreased, additional component processes are necessary to generate $1/f$ spectral behavior over the same frequency range.

When choosing the model order parameters for the estimator, it is important to note that the rate of convergence of the EM algorithm slows as the number of component processes increases. Furthermore, the number of computations in the estimation step of each iteration is proportional to the square of the number of component processes.

5.1.3 Simulations

The performance of the parameter estimation algorithm is analyzed using Monte Carlo simulations on synthetic discrete samples of $1/f$ processes, both for the noise-free case and the case where the data is corrupted by stationary white Gaussian noise. The $1/f$ processes are generated using the state-space description of Chapter 4. The model order parameters is this section are chosen to be $\Delta = 4, M_L = -10$, and $M_H = 10$ for both the generated process and parameter estimation algorithm.

We evaluate the performance of the special form of the EM algorithm corresponding to the noise-free case discussed in Section 5.1.1. Figure 5-3 shows the RMS error in the estimates of γ and σ^2 as a function of observation length, for various values of γ. The results from 64 trials were averaged to obtain the error estimates. In some cases, the error in the estimates increased with additional data samples; this is most likely a result of the relatively small number of trials used in this simulation, rather than a property of the estimator.

We also evaluate the performance of the general EM algorithm in several SNR scenarios.

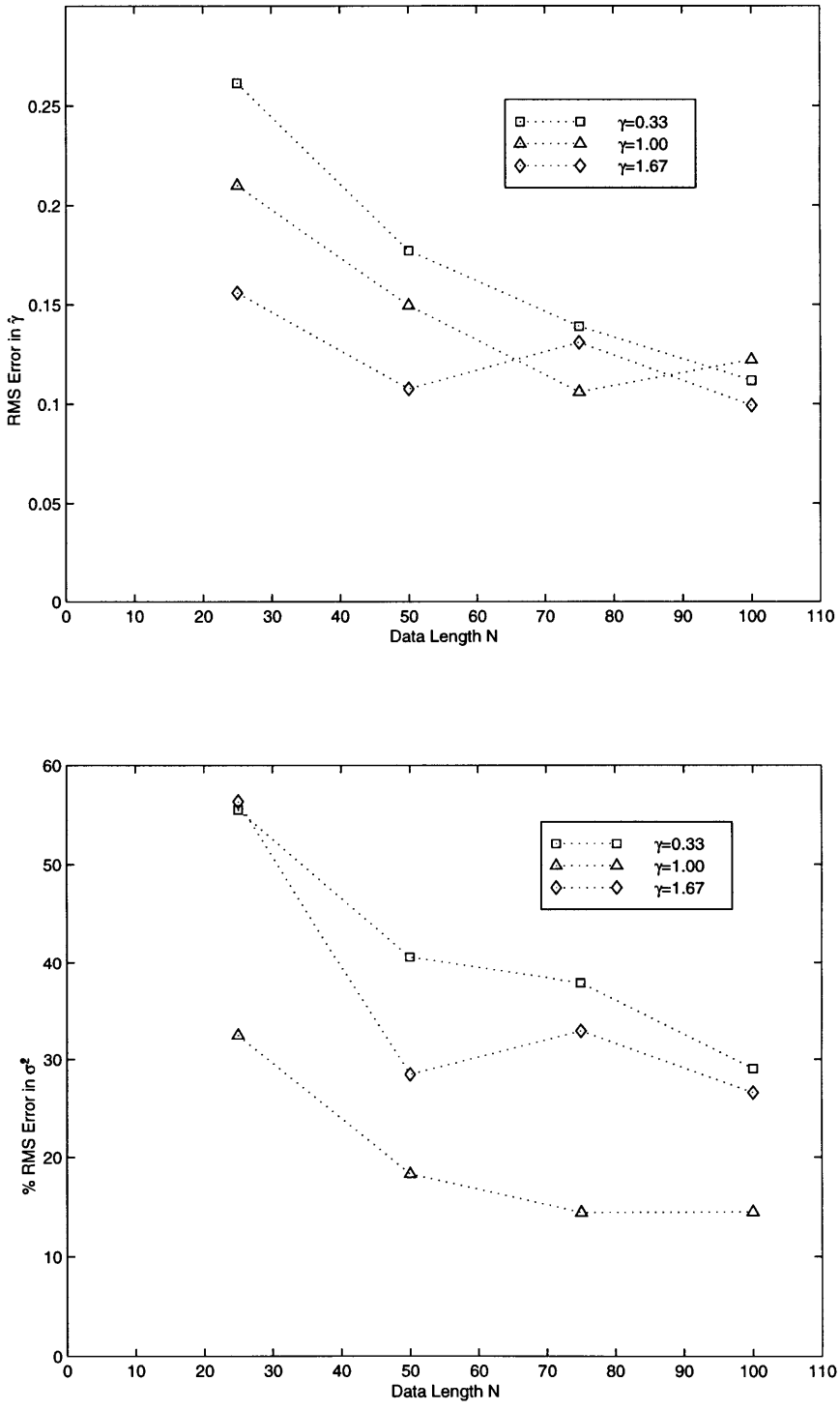


Figure 5-3: RMS error in estimates of γ and σ^2 as a function of data length N in the special case of no noise

We define the signal power in this case to be

$$E_s = \frac{1}{\pi} \int_{\pi/N}^{\pi} S_x(\Omega) d\Omega \tag{5.27}$$

where $S_x(\Omega)$ is the power spectrum of the $1/f$ process. The signal-to-noise ratio is therefore

$$\mathrm{SNR} = \log_{10} \frac{\mathrm{E_s}}{\sigma_{\mathrm{w}}^2}. \tag{5.28}$$

Figure 5-4 shows the RMS error in the estimates of γ and σ^2 as a function of the SNR, for various values of γ. The results from 64 trials were averaged to obtain the error estimates. The data length is these simulations is 50 samples. In some cases, the error in the estimates increases as SNR is increased; again, this is probably a result of the relatively small number of trials used in this simulation, as opposed to a property of the estimator. We see that as γ increases, the effects of white noise on the parameter estimates decreases. This is to be expected since for large values of γ, low-frequency component processes dominate and are easily distinguishable from white noise. For small values of γ, high-frequency component processes dominate; the values of these component processes are hard to estimate due to their similarity with the obscuring white noise.

5.2 $1/f$ Signal and Deterministic Signal

Section 5.1 considers applications in which the $1/f$ process is the signal is of primary interest. In other applications, the $1/f$ process is a noise process obscuring another signal of interest. This section considers the problem of estimating the parameters of a deterministic signal obscured by an unknown $1/f$ noise process as well as white Gaussian measurement noise.

We formulate the problem as follows. Suppose we have observations $z[n]$ of a deterministic signal $s[n]$ obscured by a discrete zero-mean Gaussian $1/f$ noise process $x[n]$ in addition to zero-mean independent identically distributed (i.i.d.) Gaussian noise $w[n]$ that is statistically independent of $x[n]$, so

$$z[n] = s[n] + x[n] + w[n], \quad 0 \leq n \leq N-1 \tag{5.29}$$

where N is the length of the observed data. In addition, the signal is parameterized as a

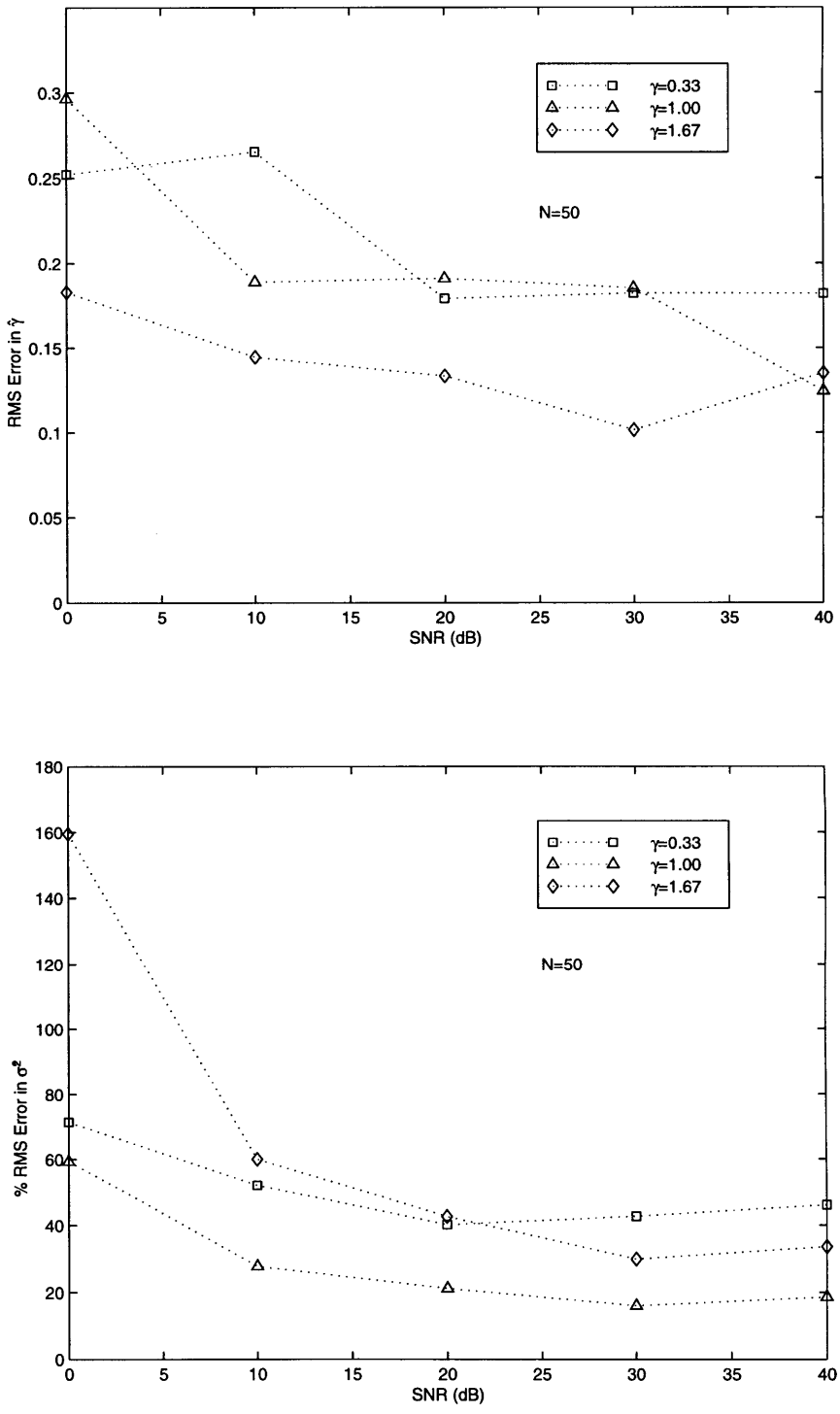


Figure 5-4: RMS error in estimates of γ and σ^2 as a function of SNR with data length $N = 50$

linear combination of a finite set of known basis signals, *i.e.*,

$$s[n] = \sum_{p=1}^{P} \lambda_p b_p[n] \tag{5.30}$$

for real parameters $\lambda_1, \ldots, \lambda_P$. We may view each finite length sequence as an N-length column vector

$$\boldsymbol{z} = \begin{bmatrix} z[0] \\ \vdots \\ z[N-1] \end{bmatrix}, \quad \boldsymbol{s} = \begin{bmatrix} s[0] \\ \vdots \\ s[N-1] \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x[0] \\ \vdots \\ x[N-1] \end{bmatrix}, \quad \boldsymbol{w} = \begin{bmatrix} w[0] \\ \vdots \\ w[N-1] \end{bmatrix}. \tag{5.31}$$

We wish to estimate the vector of unknown parameters

$$\boldsymbol{\theta} = \{\lambda_1, \ldots, \lambda_P, \gamma, \sigma^2, \sigma_w^2\} \tag{5.32}$$

where $\lambda_1, \ldots, \lambda_P$ parameterize the deterministic signal $s[n]$, $0 < \gamma < 2$ and $\sigma^2 > 0$ are parameters for the $1/f$ process $x[n]$ and $\sigma_w^2 > 0$ is the variance of the i.i.d. additive Gaussian noise process $w[n]$. As in the previous section, maximizing the likelihood function of the observed data directly is difficult.

The EM algorithm for this problem is similar to that presented in Section 5.1. A detailed derivation of the EM algorithm for this problem is presented in Appendix B.2. The algorithm is summarized below, where $\lambda_1^{[l]}, \ldots, \lambda_P^{[l]}$, $\gamma^{[l]}$, $\sigma^{2[l]}$, and $\sigma_w^{2[l]}$ denote the estimates generated on the lth iteration of parameters $\lambda_1, \ldots, \lambda_P$, γ, σ^2, and σ_w^2, respectively. The algorithm is summarized in Table 5.3. We omit the definitions of $q_m(\Delta)$ and H_m^{-1} which can be found in Table 5.1.

E Step

As in the previous algorithm, the E step calculates the statistics of the complete data, which includes the component states as well as information about the white noise variance, using the most recent parameter estimates. Again, these expectations can be efficiently computed using the Kalman smoothing equations (4.31)–(4.38) with a $2M$-state system described by equations (4.11), (4.12), (4.14) and (4.17), representing the $1/f$ signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$ and white noise with parameter $\sigma_w^{2[l]}$.

E Step

$$
\begin{aligned}
z'[n] &= z[n] - \sum_{p=1}^{P} \lambda_p^{[l]} b_p[n] && (5.33)\\
Z'[N] &= \{z'[0], \ldots, z'[N]\} && (5.34)\\
\hat{x}_m^{[l]}[n] &= \mathrm{E}\left\{x_m[n] \mid Z'[N]; \boldsymbol{\theta}^{[l]}\right\} && (5.35)\\
(x_m[n]x_k[n])\hat{}^{[l]} &= \mathrm{E}\left\{(x_m[n] - \hat{x}_m[n])(x_k[n] - \hat{x}_k[n]) \mid Z'[N]; \boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_k[n], \\
&\quad k = 1, \ldots, M && (5.36)\\
(x_m[n]x_m[n-1])\hat{}^{[l]} &= \mathrm{E}\left\{(x_m[n] - \hat{x_m}[n])\,(x_m[n-1] - \hat{x_m}[n-1]) \mid Z'[N]; \boldsymbol{\theta}^{[l]}\right\} \\
&\quad + \hat{x}_m[n]\hat{x}_m[n-1], \quad n = 1, \ldots, N-1 && (5.37)
\end{aligned}
$$

for $m = 1, \ldots, M$.

M Step

$$
\begin{aligned}
\sigma_w^{2[l+1]} &= \frac{1}{N}\sum_{n=0}^{N-1}\Bigg[z^2[n] - 2z[n]\sum_{p=1}^{P}\lambda_p^{[l]}b_p[n] \\
&\quad + \sum_{p=1}^{P}\sum_{q=1}^{Q}\lambda_p^{[l]}\lambda_q^{[l]}b_p[n]b_q[n] - 2z[n]\sum_{m=1}^{M}\hat{x}_m^{[l]}[n] \\
&\quad + 2\sum_{m=1}^{M}\sum_{p=1}^{P}\lambda_p^{[l]}\hat{x}_m^{[l]}[n]b_p[n] + \sum_{m=1}^{M}\sum_{k=1}^{M}(x_m[n]x_k[n])\hat{}^{[l]}\Bigg] && (5.38)\\
\gamma^{[l+1]} &\longleftarrow \sum_{m=1}^{M}\left(m - \frac{M+1}{2}\right)\frac{\Delta^{\gamma^{[l+1]}(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) = 0 && (5.39)\\
\sigma^{2[l+1]} &= \frac{1}{NM}\sum_{m=1}^{M}\frac{\Delta^{\gamma^{[l+1]}(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) && (5.40)\\
\lambda_p^{[l+1]} &= \left(\sum_{n=1}^{N}b_p^2[n]\right)^{-1}\sum_{n=1}^{N}\left[b_p[n]\left(z[n] - \sum_{k\neq p}\lambda_k^{[l]}b_k[n] - \sum_{m=1}^{M}\hat{x}_m^{[l]}[n]\right)\right] && (5.41)
\end{aligned}
$$

Table 5.3: EM algorithm for parameter estimation of deterministic signal obscured by $1/f$ and white noise

Since we have available observations of the $1/f$ signal, white noise, and an additional deterministic signal, the sequence that is input into the Kalman smoothing equation is the observation sequence with the current estimate of the deterministic sequence subtracted out,

$$z'[n] = z[n] - \sum_{p=1}^{P} \lambda_p^{[l]} b_p[n]. \tag{5.42}$$

M Step

The M step updates the parameter estimate by performing the maximization

$$\max_{\boldsymbol{\theta}} U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]}) \rightarrow \boldsymbol{\theta}^{[l+1]}. \tag{5.43}$$

The maximization of the $1/f$ signal parameters are unchanged. The results of the M step are derived in detail in Appendix B.2.

5.2.1 Special Case of Known $1/f$ Parameters

In this section, we consider a scenario in which the parameters of the $1/f$ process and white measurement noise are known. In this case, the problem reduces to estimating the parameters of a deterministic signal in colored noise of known statistics.

The problem formulation follows that of Section 5.2 through equations (5.29), (5.30), (5.31). In this case, we wish the estimate the vector of unknown parameters

$$\boldsymbol{\theta} = \begin{bmatrix} \lambda_1 & \dots & \lambda_P \end{bmatrix}^T \tag{5.44}$$

which parameterize the deterministic signal $s[n]$. In this case, maximization of the parameters can be performed directly.

In vector notation, the problem formulation can be expressed as

$$\boldsymbol{z} = G\boldsymbol{\theta} + \boldsymbol{x} + \boldsymbol{w} \tag{5.45}$$

where

$$G = \begin{bmatrix} b_1[0] & b_2[0] & \dots & b_P[0] \\ \vdots & \vdots & & \vdots \\ b_1[N-1] & b_2[N-1] & \dots & b_P[N-1] \end{bmatrix} \tag{5.46}$$

and $\boldsymbol{x}$ and $\boldsymbol{w}$ have zero mean and covariances $\Lambda_X(\gamma, \sigma^2)$ from (5.6) and $\Lambda_W(\sigma_w^2)$ from (5.7), respectively. The maximum likelihood parameters $\hat{\boldsymbol{\theta}}_{ML}(\boldsymbol{z})$ with covariances Λ_{ML} are given by

$$\hat{\boldsymbol{\theta}}_{ML}(\boldsymbol{z}) = \left(G^T(\Lambda_X + \Lambda_W)^{-1}G\right)^{-1} G^T(\Lambda_X + \Lambda_W)^{-1}\boldsymbol{z} \tag{5.47}$$

$$\Lambda_{ML} = \left(G^T(\Lambda_X + \Lambda_W)^{-1}G\right)^{-1}. \tag{5.48}$$

We consider the special case of estimating a deterministic affine signal in $1/f$ noise of known parameters, with $\sigma_w^2 = 0$. In this case, the unknown parameters $\boldsymbol{\theta} = [\ \lambda_1 \quad \lambda_2\ ]^T$ represent the offset and slope of the deterministic signal, and

$$G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}. \tag{5.49}$$

Estimation of the offset of the affine signal has a strong dependence on the apparent steady value of the process and is therefore sensitive to the model order parameter M_L for $\gamma \geq 1$. Therefore, we are principally interested in the estimation of the slope of the signal. Figure 5-5 plots the error covariance in the estimate of the slope as a function of the data length N. The error covariances are normalized so that the error for each value of γ is the same at $N = 2$. The error covariances exhibit polynomial-type decay as a function of data length. For Brownian motion, the error covariance is proportional to $1/N$, whereas for white noise, the error covariance is asymptotically proportional to $1/N^3$. As γ increases, it becomes increasingly difficult to estimate the slope of an affine signal in $1/f$ noise.

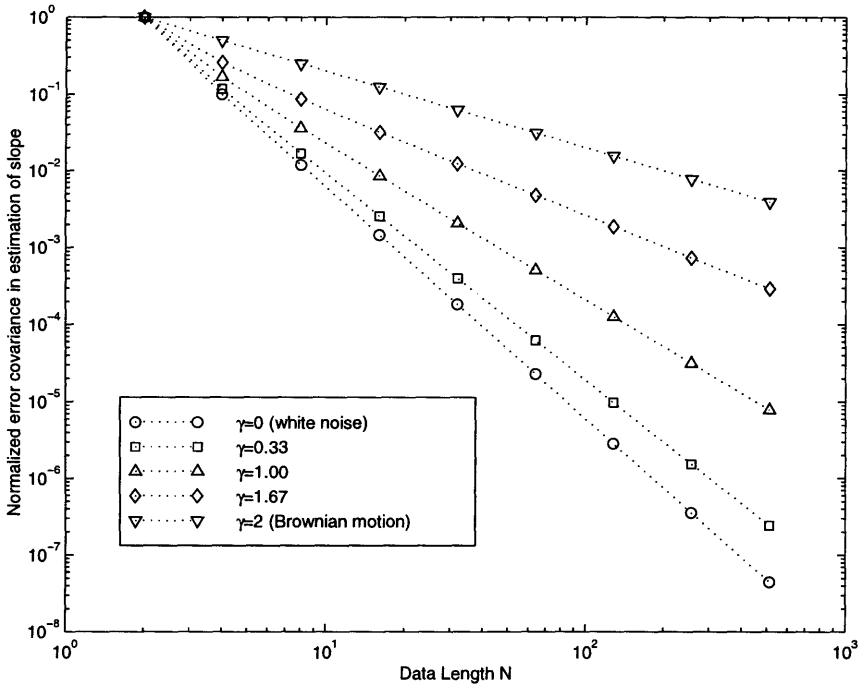


Figure 5-5: Error covariance in the estimate of the slope of a deterministic affine signal in $1/f$ noise of known parameters as a function of data length N

Chapter 6

Prediction and Smoothing of $1/f$ Signals

In this chapter, we consider the $1/f$ signal estimation problems of prediction and smoothing. Our general problem formulation is to consider the estimation of the past, present, or future values of a $1/f$ signal $s[n]$ from noisy observations $z[n]$ of the form

$$z[n] = s[n] + w[n], \quad 0 \leq n \leq N \tag{6.1}$$

where $w[n]$ is an i.i.d. Gaussian noise sequence with zero mean and variance σ_w^2. The observations can be represented with the multiscale state-space description from Section 4.2. We assume that the signal and noise parameters $\gamma, \sigma^2, \sigma_w^2$ are known, although in practice, these parameters must be estimated from the observed data using the algorithms in Chapter 5. We exploit the state-space model in Chapters 3 and 4 to design computationally efficient recursive algorithms based on the Kalman filter and smoother.

In the first part of the chapter, we consider the problem of predicting the future values of the $1/f$ signal given the observations. Algorithms for single-step and multi-step prediction are presented and analyzed in the special case of no noise. In the second part of the chapter, we consider of smoothing of the noisy observations of the $1/f$ signal over a fixed time interval. The smoothing algorithm is analyzed for several SNR scenarios.

6.1 Single-Step Prediction

Prediction pertains to the problem of estimating the future values of a random process given observations of past values. In this section, we consider estimating a single time sample into the future. In Section 6.2, we extend the algorithm to address the problem of estimating multiple time samples into the future. The properties of the predictor for $1/f$ processes are compared to those for autoregressive and moving-average processes.

Given the observation set $Z[n] = \{z[0], z[1], \ldots, z[n]\}$ of a $1/f$ process, single-step prediction optimally estimates the next observation sample $s[n+1]$. The state estimate $\hat{\boldsymbol{x}}[N+1|N]$ and state error covariance $R_x[N+1|N]$ for the system can be determined by applying the Kalman filtering equations in Table 4.1. Using this information, the single-step prediction estimate

$$\begin{aligned} \hat{s}[N+1|N] &= \mathrm{E}\{x[N+1] \mid Z[N]\} \\ &= \mathrm{E}\{C\boldsymbol{x}[N+1] \mid Z[N]\} \\ &= C\hat{\boldsymbol{x}}[N+1|N] \end{aligned} \tag{6.2}$$

and the prediction error covariance

$$\begin{aligned} R_s[N+1|N] &= \mathrm{E}\left\{(z[N+1] - \hat{z}[N+1|N])^2 \mid Z[N]\right\} \\ &= \mathrm{E}\left\{(C\boldsymbol{x}[N+1] - C\hat{\boldsymbol{x}}[N+1|N])^2 \mid Z[N]\right\} \\ &= CR_x[N+1|N]C^T \end{aligned} \tag{6.3}$$

are easily calculated. The single-step prediction algorithm is summarized in Table 6.1. Note that the prediction error covariance is independent of the observed data and dependent only on the number of observed time samples N.

Due to the long-term dependence characteristic of $1/f$ signals, we expect that the qualitative behavior of the prediction error covariance for the case of $1/f$ signals is different from the case of autoregressive moving-average (ARMA) signals. We analyze the single-step prediction algorithm for the special case $\sigma_w^2 = 0$. To facilitate this analysis, we note a few important properties of the estimator. A prediction made using no data has an error covariance equal to the variance of the process, *i.e.*, $R_s[0|-1] = \sigma_s^2$. We view $R_s[N+1|N]$

Kalman Filter

Apply Kalman filter to observations $Z[N]$ to generate $\hat{\boldsymbol{x}}[N+1|N]$ and $R_x[N+1|N]$.

$1/f$ Process Prediction

$$\hat{s}[N+1|N] = C\hat{\boldsymbol{x}}[N+1|N] \tag{6.4}$$
$$R_s[N+1|N] = CR_x[N+1|N]C^T \tag{6.5}$$

Table 6.1: Single-step prediction algorithm

as the error covariance of a prediction using N previous time samples. Using additional previous time samples never increases the prediction error covariance:

$$R_s[N+2|N+1] \leq R_s[N+1|N]. \tag{6.6}$$

While additional data will generally improve the quality of the prediction, there is a lower bound on the prediction error covariance. This lower bound

$$R_s[\infty|\infty] = \lim_{N\to\infty} R_s[N+1|N] \tag{6.7}$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log S_s(\Omega)d\Omega \tag{6.8}$$

is the minimum prediction error covariance, where $S_s(\Omega)$ is the power spectrum of $s[n]$. We are interested in analyzing how quickly the prediction error covariance approaches this lower bound as the number of previous time samples are increased. Therefore, we consider the effect of increasing N on the quantity $R_s[N+1|N] - R_s[\infty|\infty]$ for autoregressive, moving-average, and $1/f$ processes.

Autoregressive Processes

An autoregressive process $s[n]$ of order K is defined by

$$s[n] + a_1 s[n-1] + \ldots + a_K s[n-K] = w[n] \tag{6.9}$$

where the a_i are constants and $\{w[n]\}$ is a sequence of independent identically distributed Gaussian random variables with variance σ_w^2. The minimum error covariance is $R_s[\infty|\infty] = \sigma_w^2$ and is achieved for $N = K$. For an autoregressive process of order K, the memory of the process is completely contained in the past K time samples.

Moving Average Processes

A moving average process $s[n]$ of order K is defined by

$$s[n] = w[n] + c_1 w[n-1] + \ldots + c_K w[n-K] \tag{6.10}$$

where the c_i are constants and $\{w[n]\}$ is a sequence of independent identically distributed Gaussian random variables with variance σ_w^2. The variance of the process is

$$\sigma_s^2 = \sigma_w^2(1 + c_1^2 + \ldots + c_K^2) \tag{6.11}$$

and the minimum error covariance is $R_s[\infty|\infty] = \sigma_w^2$. In contrast to the autoregressive case, the minimum error covariance is not achieved exactly for finite N. Figure 6-1 shows $R_s[N+1|N] - R_s[\infty|\infty]$ vs. N for a moving average process described by $c_1 = .8, c_2 = .6$. The prediction error covariance approaches the lower bound exponentially as N increases.

$1/f$ Processes

We examine three $1/f$ processes with $\gamma = 0.33, 1.00, 1.67$. We see in Figure 6-2 that the prediction error covariance approaches the minimum error covariance in a polynomial fashion. We observe that

$$R_s[N+1|N] - R_s[\infty|\infty] \propto \frac{1}{N} \tag{6.12}$$

approximately for each of these processes, for $N \geq 2$. This behavior, in contrast with the more rapid approach to the lower bound in the moving average and autoregressive cases, suggests that samples from the distant past of $1/f$ processes have a greater impact than samples from the distant past of ARMA processes on the performance of the algorithm in predicting future data samples. This is to be expected due to long-term correlation characteristic of $1/f$ processes.

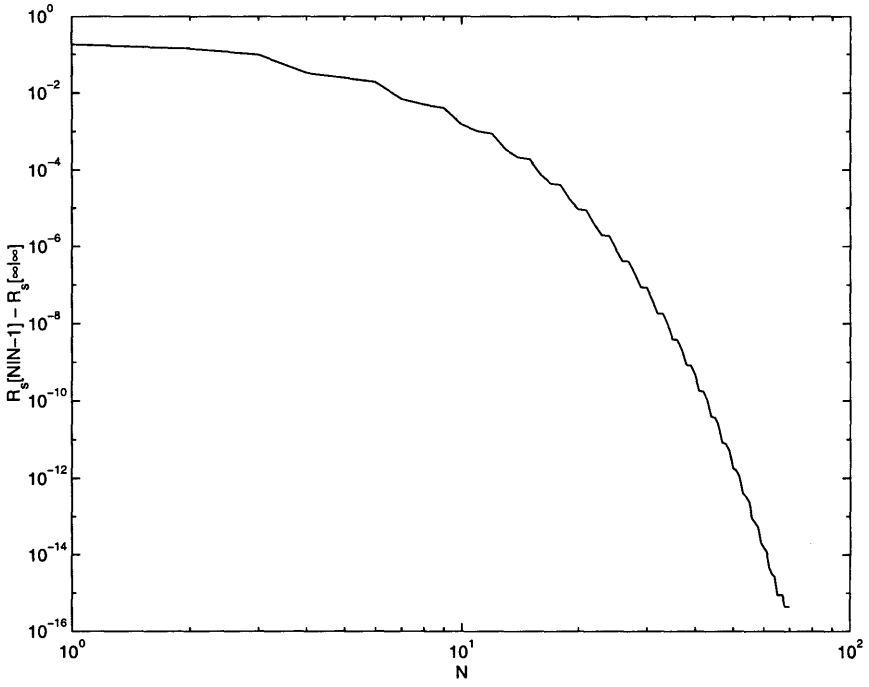


Figure 6-1: $R_s[N+1|N] - R_s[\infty|\infty]$ vs. N for a moving average process

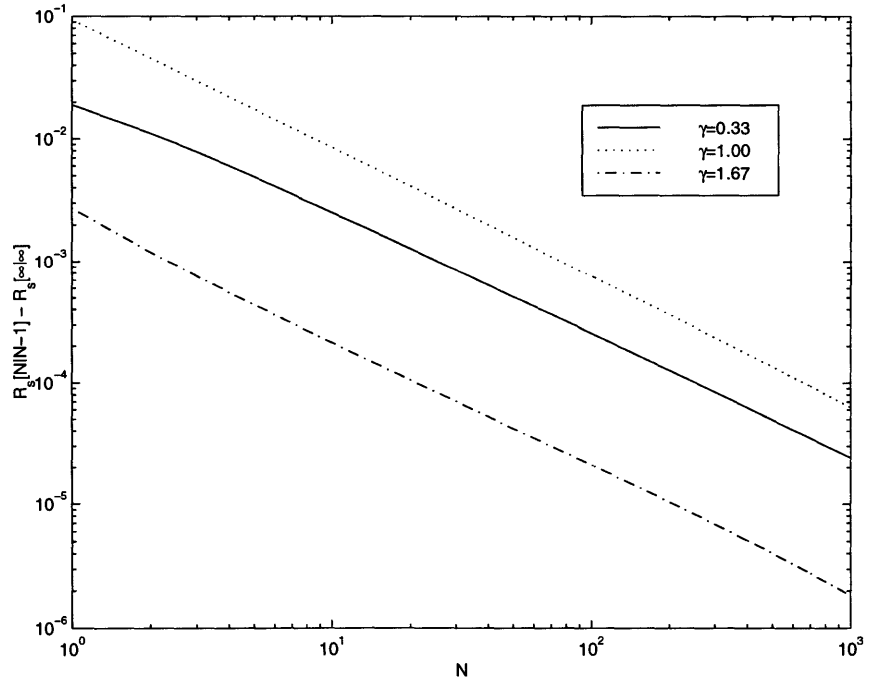


Figure 6-2: $R_s[N+1|N] - R_s[\infty|\infty]$ vs. N for $1/f$ processes

Kalman Filter

Apply Kalman filter to observations $Z[N]$ to generate $\hat{\boldsymbol{x}}[N+1|N]$ and $R_x[N+1|N]$.

For $m = 1, \ldots, M-1$, perform the following step

Propagation Equations

$$\hat{\boldsymbol{x}}[N+m+1|N] = A\hat{\boldsymbol{x}}[N+m|N] \tag{6.15}$$
$$R_x[N+m+1|n] = AR_x[N+m|N]A^T + BB^T \tag{6.16}$$

$1/f$ Process Prediction

$$\hat{s}[N+M|N] = C\hat{\boldsymbol{x}}[N+M|N] \tag{6.17}$$
$$R_s[N+M|N] = CR_x[N+M|N]C^T \tag{6.18}$$

Table 6.2: Multi-step prediction algorithm

6.2 Multi-Step Prediction

Given the observation set $Z[N]$ of a $1/f$ process, multi-step prediction optimally estimates the Mth succeeding time sample $s[N+M]$, for $M > 1$. The state estimate $\hat{\boldsymbol{x}}[N+1|N]$ and state error covariance $R_x[N+1|N]$ can be determined through application of the Kalman filtering equations in Table 4.1. This multi-step prediction is performed iteratively by using the propagation equations of the Kalman filter. The omission of the update equation reflects the lack of additional observation information $Z[N+1], \ldots, Z[N+M-1]$. This algorithm produces the state estimate $\hat{\boldsymbol{x}}[N+M|N]$ and state error covariance $R_x[N+M|N]$, which are used to calculate the multi-step prediction estimate

$$\hat{s}[N+M|N] = \mathrm{E}\left\{s[N+M]|Z[N]\right\} \tag{6.13}$$

and prediction error covariance

$$R_s[N+M|N] = \mathrm{E}\left\{(s[N+M] - \hat{s}[N+M|N])^2 \mid Z[N]\right\}. \tag{6.14}$$

The multi-step prediction algorithm is summarized in Table 6.2.

As in Section 6.1, we analyze the behavior of the prediction error covariance and com-

pare it to the prediction error covariance for autoregressive and moving-average processes. For this analysis, we assume that $\sigma_w^2 = 0$ and let $N = 10000$ so that the prediction error covariance has approximately reached a steady state at the minimum prediction error covariance. As M increases, the prediction error covariance increases monotonically:

$$R_s[N+M+1|N] \geq R_s[N+M|N]. \tag{6.19}$$

The maximum value of the prediction error covariance is the same as the error covariance of predicting the signal with no data, which is equal to the variance of the signal:

$$\lim_{M\to\infty} R_s[N+M|N] = \sigma_s^2. \tag{6.20}$$

We are interested in how quickly the prediction error covariance approaches this upper bound as M increases. Therefore, we consider $\sigma_s^2 - R_s[N+M|N]$ as we increase M for autoregressive, moving-average, and $1/f$ processes.

Autoregressive Processes

For this example, we use an third-order autoregressive process with coefficients $a_1 = 2.4, a_2 = 1.92, a_3 = 0.512$. Figure 6-3 shows $\sigma_s^2 - R_s[N+M|N]$ vs. M for this process. There is a rapid (faster than polynomial) approach to the maximum error covariance as M increases. This suggests that as the distance from the predicted time sample to the observed data increases, the value of the observed data for predicting the point quickly decreases.

Moving-Average Processes

By the definition of moving-average processes in (6.10), a sample $s[n]$ of a Kth order moving-average process is uncorrelated to samples $s[n-K-1], s[n-K-2], \ldots$ more than K points in the past. Therefore,

$$R_s[N+M|N] = \sigma_s^2 \tag{6.21}$$

for $M > K$. Observations of a Kth moving-average process are useful only in predicting future data within K samples of the last observation sample.

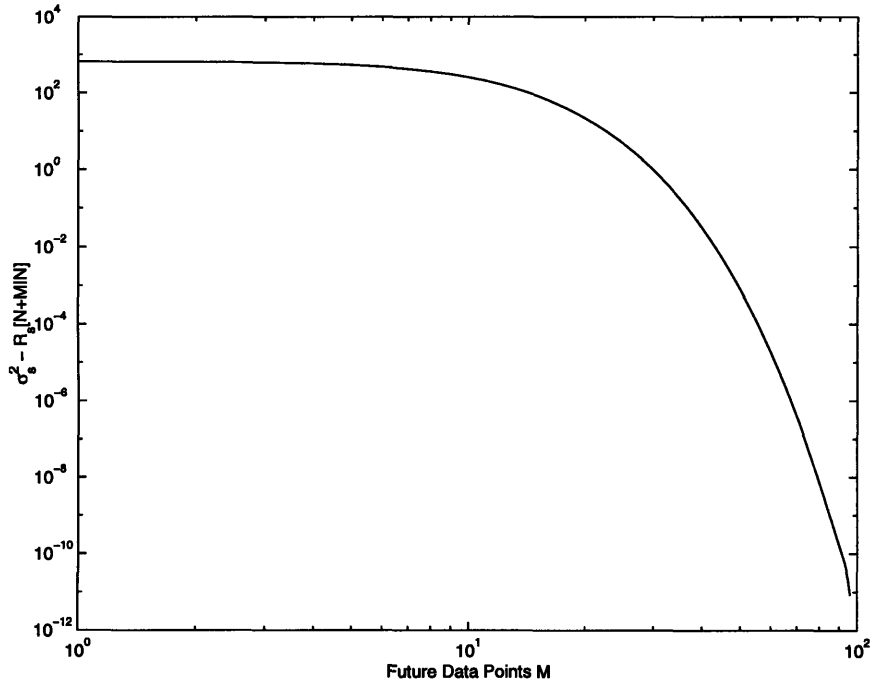


Figure 6-3: $\sigma_s^2 - R_s[N+M|N]$ vs. M for autoregressive process

$1/f$ Processes

We examine three $1/f$ processes with $\gamma = 0.33, 1.00, 1.67$. We see in Figure 6-4 that the prediction error covariance approaches the maximum error covariance in a polynomial fashion. For $\gamma = 1$ and $\gamma = 1.67$, this decay is extremely slow. This behavior, in contrast with the more rapid approach to the lower bound in the moving average case, is also due to the long-term correlation characteristic of $1/f$ processes. As the distance from the predicted time sample to the observed data increases, the value of the observed data for predicting the point decreases slowly. As γ increases, corresponding to a stronger dependence, this decrease becomes even slower.

6.3 Smoothing

An observed $1/f$ signal is frequently obscured by a background of additive stationary white noise. Smoothing of the $1/f$ signal attempts to extract the signal from the noise. In this section, we exploit the linear state-space system to design algorithms for smoothing.

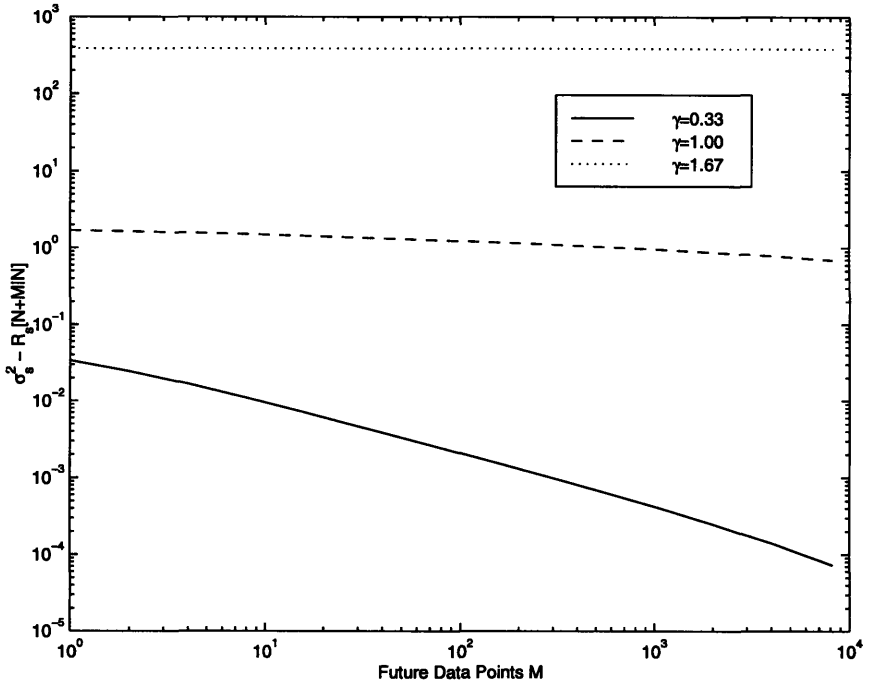


Figure 6-4: $\sigma_s^2 - R_s[N+M|N]$ vs. M for $1/f$ processes

Our problem formulation is to consider the estimation of values of a $1/f$ signal $s[n]$ from $0 \leq n \leq N$ from noisy observations $Z[N] = \{z[0], z[1], \ldots, z[N]\}$ of the form

$$z[n] = s[n] + w[n] \tag{6.22}$$

where $w[n]$ is an i.i.d. Gaussian noise sequence with zero mean and variance σ_w^2. We note that the observations can be described with the multiscale state-space description from Section 4.2. For this section, we assume that the signal and noise parameters $\gamma, \sigma^2, \sigma_w^2$ are known. In practice, these parameters must be estimated from the observed data.

We first apply the discrete-time Kalman smoothing equations in Table 4.2 to the observed data. The state estimates and error covariance are then used to smooth the $1/f$ process. The smoothing algorithm is summarized in Table 6.3.

We measure the performance of these estimators in terms of SNR gain of the signal estimate, as a function of the SNR over relevant frequencies of the observations. We define

Kalman Smoothing

Apply Kalman smoothing equations observations $Z[N]$ to generate $\hat{\boldsymbol{x}}[n|N]$ and $R_x[n|N]$ for $0 \leq n \leq N$.

$1/f$ Process Estimation

$$\hat{s}[n] = C\hat{\boldsymbol{x}}[n|N] \tag{6.23}$$
$$R_s[n] = CR_x[n|N]C^T \tag{6.24}$$

Table 6.3: Smoothing algorithm using finite-interval data

the signal energy in this case to be

$$E_s = \frac{1}{\pi}\int_{\pi/N}^{\pi} S_x(\Omega)d\Omega \tag{6.25}$$

where $S_x(\Omega)$ is the power spectrum of the $1/f$ process. The signal-to-noise ratio is therefore

$$SNR = \log_{10}\frac{E_s}{\sigma_w^2}. \tag{6.26}$$

The mean-squared estimation error over the finite observation length is

$$\epsilon = \frac{1}{N}\sum_{n=0}^{N} R_s[n]. \tag{6.27}$$

The SNR gain is defined as

$$\text{SNR gain} = \frac{\sigma^2}{\epsilon}. \tag{6.28}$$

Figure 6-5 shows the SNR gain of the signal estimates for various values of γ as a function of the SNR of the observations. The sequence length is $N = 1000$ and the values of M_L and M_H for the finite model are chosen according to the last column of Table 3.2. The estimator achieves a larger gain as γ is increased. For larger values of γ, more energy is concentrated in low-frequency scales with strong long-term correlation; these scales are easily distinguished from the noise. At smaller values of γ, more energy is concentrated at high-frequency scales with weak long-term correlation; these scales are similar to the noise and are consequently difficult to estimate. Similar results for smoothing were achieved using

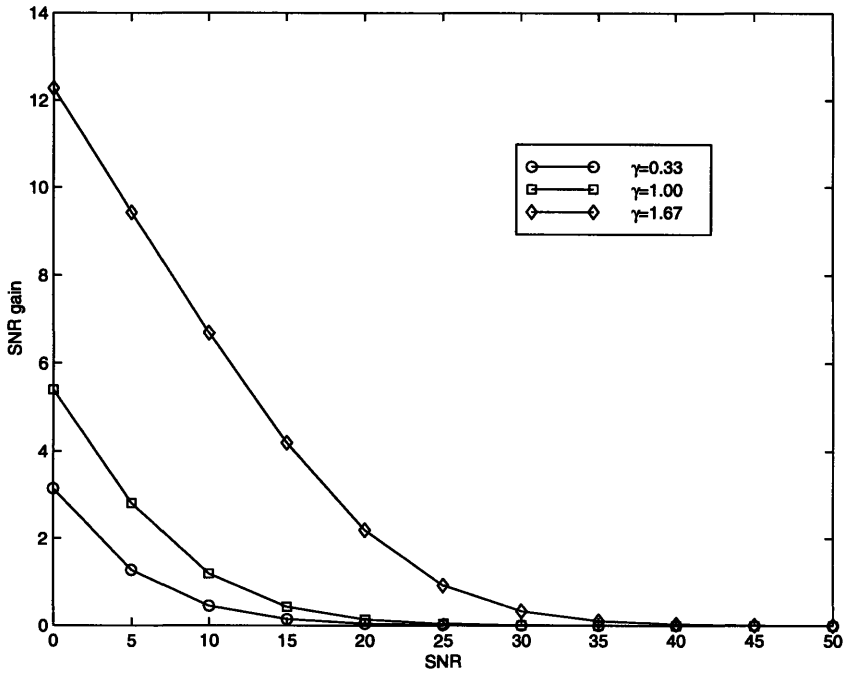
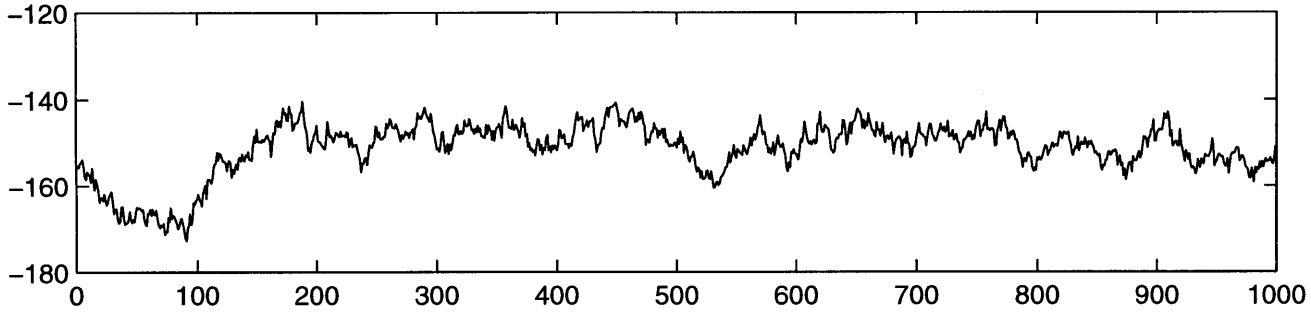


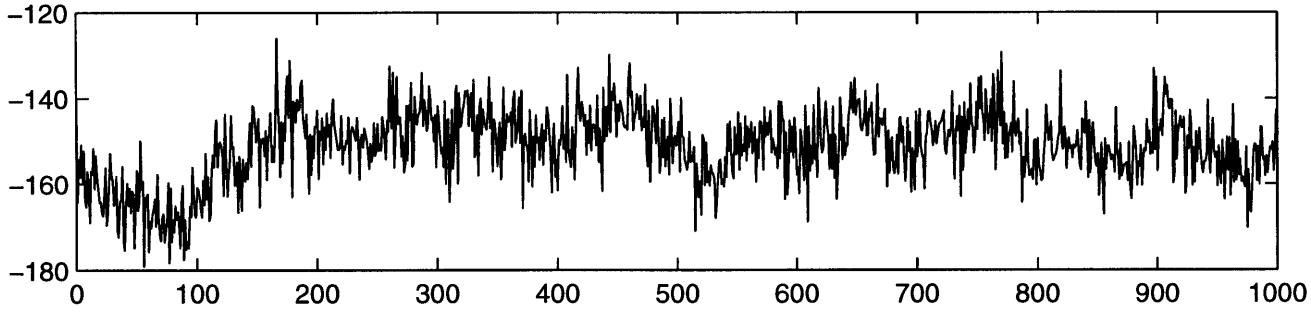
Figure 6-5: SNR gain (dB) of the signal estimate as a function of SNR (dB) of the observations over relevant frequencies for various γ

wavelet-based algorithms in [20].

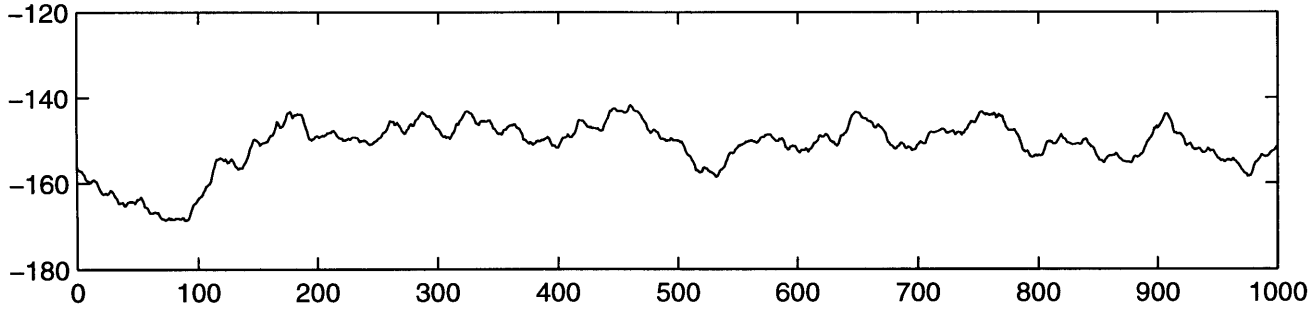
Finally, Figure 6-6 presents an example of smoothing of a $1/f$ signal of length $N = 200$ obscured by additive white Gaussian noise. The SNR is 0 dB. The parameters of the data set are estimated using the parameter estimation algorithm presented in Section 5.1. Using these parameters, smoothing is performed on the noisy data set. The estimated $1/f$ parameter is $\hat{\gamma} = 1.70$ and the SNR gain of the smoothing algorithm is 9.5 dB.



(a) $1/f$ signal with $\gamma = 1.67$



(b) Signal corrupted by additive white Gaussian noise. SNR = 0dB



(c) Smoothed estimate of $1/f$ signal given noisy observations.
$\hat{\gamma} = 1.70$, SNR gain = 9.5dB

Figure 6-6: Example of signal smoothing

Chapter 7

Conclusions

The $1/f$ family of fractal random processes models a wide range of physical and man-made phenomena. In this thesis, we have introduced a multiscale state-space representation for $1/f$ processes suitable for addressing several signal processing problems that have previously had no practical solution, such as parameter estimation of a deterministic signal in $1/f$ noise and prediction of a $1/f$ signal in white noise.

This thesis has focused on discrete, finite-length Gaussian $1/f$ signals. A significant contribution of this thesis is the frequency-based characterization for discrete $1/f$ processes in Chapter 3 based on the concept of self-similarity at low frequencies. We have developed a multiscale state-space representation for $1/f$ processes satisfying this characterization. This representation is composed of a finite number of first-order autoregressive component processes and is naturally represented with a state-space system description given in Chapter 4. We have analyzed the effects of various choices of component processes on the spectrum of the process.

Using this representation, in Chapter 5 we have presented iterative algorithms for estimating the parameters of $1/f$ signals in white noise that exploit the computational efficiency of the Kalman smoother. A performance analysis based on Monte Carlo simulations has demonstrated the robustness of the estimator in the presence of noise even for relatively small data lengths. In the same chapter, we have presented algorithms for estimating the parameters of a deterministic signal in $1/f$ and white noise. A preliminary analysis has been performed for the special case of affine signals in $1/f$ and white noise of known statistics. A performance analysis for the general case of unknown $1/f$ and white noise statistics

provides a possible future direction for research. Generally, theoretical characterizations for the performance of these estimators based on tools such as the Cramer-Rao bounds could prove valuable.

Several potential extensions to these parameter estimation algorithms provide interesting opportunities for additional research. Since these estimator are sensitive to mismatches between true and assumed model order parameters, algorithms for jointly estimating the order of the model as well as the parameters of the signal are preferable. Another improvement would be to obtain sequential or adaptive parameter estimation algorithms, possibly by replacing the Kalman smoother with a Kalman filter or fixed-lag smoother [1]. Such sequential EM algorithms have been considered in, *e.g.* [6] [18].

The thesis focuses on $1/f$ signal estimation in white Gaussian noise in Chapter 6. A principal contribution of this thesis is the development of practical algorithms for predicting future values of the $1/f$ signal given noisy observations. An empirical analysis of the behavior of the prediction error covariance for both the single-step and multi-step prediction case compares the characteristic long-term memory of $1/f$ processes to the relatively short memory of the well-known autoregressive and moving-average processes. A theoretical justification explaining the behavior of the prediction error covariance could yield additional insight into the memory characteristics of $1/f$ processes. Furthermore, several additional signal processing problems are reasonably straightforward to address using the methods in this thesis. The problem of interpolation between data points of $1/f$ signals is potentially very interesting due to the self-similarity of $1/f$ processes.

Appendix A

Proof of Proposition 3.1

We first show that $S_y(\Omega) < S_x(\Omega)$ for any Ω such that $0 < \Omega < \pi$.

$$
\begin{aligned}
S_y(\Omega) &= \sum_{m=-\infty}^{\infty} \frac{\sigma^2 \Delta^{(2-\gamma)m}}{\Omega^2 + \Delta^{2m}} && \text{(A.1)} \\
&\overset{(a)}{=} \sum_{m=-\infty}^{\infty} \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}\right)} && \text{(A.2)} \\
&\overset{(b)}{<} \sum_{m=-\infty}^{\infty} \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m\cos\Omega} && \text{(A.3)} \\
&= S_x(\Omega) && \text{(A.4)}
\end{aligned}
$$

Step (a) is derived by substituting equations (3.25) and (3.26). Step (b) follows since β_m is positive for all m, and $\cos\Omega > 1 - \frac{\Omega^2}{2}$ for $0 < \Omega < \pi$. □

Next, we show that $S_x(\Omega) < \frac{12}{12-\Omega_0^2} S_y(\Omega)$ for any Ω such that $0 < \Omega < \Omega_0$.

$$
\begin{aligned}
S_x(\Omega) &= \sum_{m=-\infty}^{\infty} \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m\cos\Omega} && \text{(A.5)} \\
&\overset{(c)}{<} \sum_{m=-\infty}^{\infty} \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}+\frac{\Omega^4}{24}\right)} && \text{(A.6)} \\
&= \sum_{m=-\infty}^{\infty} \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}\right)} \cdot \frac{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}\right)}{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}+\frac{\Omega^4}{24}\right)} && \text{(A.7)} \\
&\overset{(d)}{<} \sum_{m=-\infty}^{\infty} \frac{12}{12-\Omega^2} \cdot \frac{f_m(1-\beta_m^2)}{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}\right)} && \text{(A.8)} \\
&\overset{(e)}{<} \frac{12}{12-\Omega_0^2} S_y(\Omega) && \text{(A.9)}
\end{aligned}
$$

Step (c) holds since β_m is positive for all m, and $\cos\Omega < 1 - \frac{\Omega^2}{2} + \frac{\Omega^4}{24}$ for $0 < \Omega < \pi$. Step (d) follows since

$$\frac{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}\right)}{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}+\frac{\Omega^4}{24}\right)} = 1 + \frac{2\beta_m(\Omega^4/24)}{1+\beta_m^2 - 2\beta_m\left(1-\frac{\Omega^2}{2}+\frac{\Omega^4}{24}\right)} \tag{A.10}$$

$$= 1 + \frac{2\beta_m(\Omega^4/24)}{(1-\beta_m)^2 + 2\beta_m\left(\frac{\Omega^2}{2}-\frac{\Omega^4}{24}\right)} \tag{A.11}$$

$$< 1 + \frac{2\beta_m(\Omega^4/24)}{2\beta_m\left(\frac{\Omega^2}{2}-\frac{\Omega^4}{24}\right)} \tag{A.12}$$

$$< 1 + \frac{\Omega^2}{12-\Omega^2} \tag{A.13}$$

$$= \frac{12}{12-\Omega^2} \tag{A.14}$$

Step (e) follows since $\Omega < \Omega_0$. □

Appendix B

The EM Parameter Estimation Algorithm

This appendix contains the derivations for the EM algorithm for the estimation of a signal and noise parameters for the case of a $1/f$ signal in white noise, presented in Section 5.1, and the related EM algorithm for the estimation of signal and noise parameters for the case of a $1/f$ signal and deterministic signal in white noise, presented in Section 5.2.

Recall from Section 3.3 that a discrete $1/f$ signal with parameters γ and σ^2 is represented as the superposition of M uncorrelated single time-constant processes

$$x[n] = \sum_{m=1}^{M} x_m[n] \tag{B.1}$$

which have correlation functions $R_m[k] = g_m(\gamma, \sigma^2)\tau_m^{-|k|}$. The time constants $\tau_m$are independent of parameters γ and σ^2. The weights are dependent on parameters γ and σ^2 of the $1/f$ signal:

$$g_m(\gamma, \sigma^2) = \left(\sigma^2 \Delta^{-\gamma(m+M_L-1)}\right) q_m(\Delta) \tag{B.2}$$

where $M_L - 1$ is the indexing offset described in Section 3.4.3 and

$$q_m(\Delta) = \frac{\Delta^{2m}}{\tau_m - \tau_m^{-1}} \tag{B.3}$$

is a function independent of the $1/f$ process parameters γ and σ^2.

B.1 $1/f$ Signal in White Gaussian Noise

In this section, we derive the EM algorithm for the estimation of a signal and noise parameters $\boldsymbol{\theta} = \{\gamma, \sigma^2, \sigma_w^2\}$ for the case of a $1/f$ signal in white noise, presented in Section 5.1. The signal is observed over a finite time length so each component process can be viewed as an N-length column vector $\boldsymbol{x_m}$. Denote the MN-length column vector

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_M} \end{bmatrix}. \tag{B.4}$$

We define the incomplete data as the observed signal $\boldsymbol{z}$. The complete data is defined as the samples of the observed signal $\boldsymbol{z}$ together with the samples of each individual component process of the $1/f$ signal:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{x} \end{bmatrix}. \tag{B.5}$$

The EM algorithm for the problem is defined [4] as

E step: Compute

$$U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]}). \tag{B.6}$$

M step:

$$\boldsymbol{\theta}^{[l+1]} = \arg\max_{\boldsymbol{\theta}} U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]}). \tag{B.7}$$

where

$$U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]}) \equiv \mathrm{E}\left\{\log f_Y(\boldsymbol{y}; \boldsymbol{\theta}) \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} \tag{B.8}$$

E step

From Bayes's rule,

$$f_Y(\boldsymbol{y}; \boldsymbol{\theta}) = f_X(\boldsymbol{x}; \boldsymbol{\theta}) \cdot f_{Z|X}(\boldsymbol{z} \mid \boldsymbol{x}; \boldsymbol{\theta}) \tag{B.9}$$

and equivalently,

$$\log f_Y(\boldsymbol{y}; \boldsymbol{\theta}) = \log f_X(\boldsymbol{x}; \boldsymbol{\theta}) + \log f_{Z|X}(\boldsymbol{z} \mid \boldsymbol{x}; \boldsymbol{\theta}) \tag{B.10}$$

The E step of the algorithm computes the conditional expectation

$$U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]}) = \mathrm{E}\left\{\log f_Y(\boldsymbol{y}; \boldsymbol{\theta}) \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} \tag{B.11}$$

$$= \mathrm{E}\left\{\log f_X(\boldsymbol{x}; \boldsymbol{\theta}) \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} + \mathrm{E}\left\{\log f_{Z|X}(\boldsymbol{z} \mid \boldsymbol{x}; \boldsymbol{\theta}) \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} \tag{B.12}$$

in an efficient manner through the Kalman smoother described in Table 4.2.

From (5.1) and (B.1),

$$\log f_{Z|X}(\boldsymbol{z} \mid \boldsymbol{x}; \boldsymbol{\theta}) = \frac{N}{2} \log 2\pi\sigma_w^2 - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} \left[z[n] - \sum_{m=1}^{M} x_m[n] \right]^2 \tag{B.13}$$

Taking the conditional expectation given $\boldsymbol{Z} = \boldsymbol{z}$ at parameter value $\boldsymbol{\theta}^{[l]}$

$$\mathrm{E}\left\{\log f_{Z|X}(\boldsymbol{z} \mid \boldsymbol{x}; \boldsymbol{\theta}) \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} \tag{B.14}$$

$$= -\frac{N}{2} \log 2\pi\sigma_w^2 - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} \left[z^2[n] - 2z[n] \sum_{m=1}^{M} \hat{x}_m^{[l]}[n] + \sum_{m=1}^{M} \sum_{k=1}^{M} (x_m[n] x_k[n])\hat{}^{[l]} \right]$$

where we define

$$(\hat{\cdot})^{[l]} \equiv \mathrm{E}\left\{\cdot \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} \tag{B.15}$$

These conditional expectations can be computed using the Kalman smoother with a $2M$-state system given by equations (4.11), (4.12), (4.14) and (4.17), describing the $1/f$ signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$ with white noise with parameter $\sigma_w^{2[l]}$. Specifically,

$$\hat{x}_m^{[l]}[n] = \mathrm{E}\left\{x_m[n] \mid Z[N]; \boldsymbol{\theta}^{[l]}\right\} \tag{B.16}$$

is a single element of the state estimate $\hat{\boldsymbol{x}}[n \mid N]$ given by the smoothing equation (4.36), and

$$(x_m[n] x_k[n])\hat{}^{[l]} = \mathrm{E}\left\{(x_m[n] - \hat{x}_m[n])(x_k[n] - \hat{x}_k[n]) \mid Z[N]; \boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_k[n] \tag{B.17}$$

where the first term in the sum is a single element of the error covariance matrix $R_{\boldsymbol{x}}[n \mid N]$ given by the smoothing equation (4.37).

The component processes have a multivariate Gaussian probability density

$$f_X(\boldsymbol{x};\boldsymbol{\theta}) = \left[\det\left(2\pi\Lambda_X(\boldsymbol{\theta})\right)\right]^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{x}^T\Lambda_X^{-1}(\boldsymbol{\theta})\boldsymbol{x}\right] \tag{B.18}$$

where $\Lambda_X = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^T\right\}$. Since the component processes are uncorrelated,

$$\Lambda_X = \begin{bmatrix} \Lambda_{X_1} & & 0 \\ & \ddots & \\ 0 & & \Lambda_{X_M} \end{bmatrix} \tag{B.19}$$

where the component matrices are

$$\Lambda_{X_m} = g_m(\sigma^{2[l]}, \gamma^{[l]}) \cdot \underbrace{\begin{bmatrix} 1 & \tau_m & \dots & \tau_m^{N-1} \\ \tau_m & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tau_m \\ \tau_m^{N-1} & \dots & \tau_m & 1 \end{bmatrix}}_{H_m} \tag{B.20}$$

denoting the covariance of the m^{th} component process and the matrix H_m is independent of $\boldsymbol{\theta}$. The inverse of this matrix can be explicitly determined as

$$\Lambda_X^{-1} = \begin{bmatrix} \Lambda_{X_1}^{-1} & & 0 \\ & \ddots & \\ 0 & & \Lambda_{X_M}^{-1} \end{bmatrix} \tag{B.21}$$

where the component inverses are known $N \times N$ tridiagonal matrices

$$\Lambda_{X_m}^{-1} = \frac{1}{g_m(\sigma^{2[l]}, \gamma^{[l]})} \cdot \frac{1}{(1-\tau_m^2)} \cdot \underbrace{\begin{bmatrix} 1 & -\tau_m & & & 0 \\ -\tau_m & 1+\tau_m^2 & -\tau_m & & \\ & \ddots & \ddots & \ddots & \\ & & -\tau_m & 1+\tau_m^2 & -\tau_m \\ 0 & & & -\tau_m & 1 \end{bmatrix}}_{H_m^{-1}}. \tag{B.22}$$

Taking the logarithm of (B.18),

$$\begin{aligned}\log f_X(\boldsymbol{x};\boldsymbol{\theta}) &= c - \frac{1}{2}\left[\log\det\Lambda_X(\boldsymbol{\theta}) + \boldsymbol{x}^T\Lambda_X^{-1}(\boldsymbol{\theta})\boldsymbol{x}\right] \\ &= c - \frac{1}{2}\left[\log\det\Lambda_X(\boldsymbol{\theta}) + \mathrm{tr}\left(\Lambda_X^{-1}(\boldsymbol{\theta})\boldsymbol{x}\boldsymbol{x}^T\right)\right] \end{aligned} \tag{B.23}$$

where c is a constant independent of $\boldsymbol{\theta}$. The conditional expectation given $\boldsymbol{Z} = \boldsymbol{z}$ at parameter value $\boldsymbol{\theta}^{[l]}$ is

$$\mathrm{E}\left\{\log f_X(\boldsymbol{x};\boldsymbol{\theta}) \mid \boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\} = c - \frac{1}{2}\left[\log\det\Lambda_X(\boldsymbol{\theta}) + \mathrm{tr}\left(\Lambda_X^{-1}(\boldsymbol{\theta})\widehat{\boldsymbol{x}\boldsymbol{x}^T}\right)\right] \tag{B.24}$$

where $\widehat{\boldsymbol{x}\boldsymbol{x}^T} = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^T \mid \boldsymbol{Z} = \boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\}$. This can be simplified by exploiting the block diagonal structure of Λ_X and substituting for $\Lambda_{X_m}(\boldsymbol{\theta})$ using (B.2), (B.20) and (B.22):

$$\begin{aligned}&\mathrm{E}\left\{\log f_X(\boldsymbol{x};\boldsymbol{\theta}) \mid \boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\} \\ &= c - \frac{1}{2}\left[\sum_{m=1}^{M}\log\det\Lambda_{X_m}(\boldsymbol{\theta}) + \sum_{m=1}^{M}\mathrm{tr}\left(\Lambda_{X_m}^{-1}(\boldsymbol{\theta})\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right)\right] &\text{(B.25)} \\ &= c_2 - \frac{1}{2}\left[\sum_{m=1}^{M}N\log g_m(\sigma^2,\gamma) + \sum_{m=1}^{M}\frac{1}{g_m(\sigma^2,\gamma)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right)\right] &\text{(B.26)} \\ &= c_3 - \frac{1}{2}\left[NM\log\sigma^2 - \frac{NM}{2}(M+2M_L-1)\gamma\log\Delta\right. \\ &\quad \left. + \frac{1}{\sigma^2}\sum_{m=1}^{M}\frac{\Delta^{\gamma(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right)\right] &\text{(B.27)}\end{aligned}$$

where

$$\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T} = \mathrm{E}\left\{\boldsymbol{x}_m\boldsymbol{x}_m^T \mid \boldsymbol{Z} = \boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\} \tag{B.28}$$

and c_2 and c_3 are constants independent of $\boldsymbol{\theta}$. It is necessary to compute only two subsets of elements of the matrix $\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}$ due to the tridiagonal structure of $\Lambda_{X_m}^{-1}(\boldsymbol{\theta})$. The first subset corresponds to the diagonal elements of $\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}$:

$$\mathrm{E}\left\{x_m[n]^2 \mid \boldsymbol{Z} = \boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\} = \mathrm{E}\left\{(x_m[n] - \hat{x}_m[n])^2 \mid \boldsymbol{Z} = \boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]^2 \tag{B.29}$$

for $n = 0,\ldots,N-1$, where we define $\hat{x}_m[n] = \mathrm{E}\left\{x_m[n] \mid \boldsymbol{Z} = \boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\}$. The second set

corresponds to the off-diagonal elements of $\widehat{\boldsymbol{x}_m \boldsymbol{x}_m^T}$:

$$\begin{aligned} &\mathrm{E}\left\{x_m[n]x_m[n-1] \mid \boldsymbol{Z}=\boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\} \\ &= \mathrm{E}\left\{\left(x_m[n]-\hat{x_m}[n]\right)\left(x_m[n-1]-\hat{x_m}[n-1]\right) \mid \boldsymbol{Z}=\boldsymbol{z};\boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_m[n-1]. \end{aligned} \tag{B.30}$$

for $n = 0, \ldots, N-1$. The quantities in (B.29) and (B.30) can be efficiently computed using the $2M$ state Kalman smoothing equations from Table 4.2 with the system described by equations (4.11), (4.12), (4.14) and (4.17), again describing the $1/f$ signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$ with white noise with parameters $\sigma_w^{2[l]}$.

In summary, the E step computes, for $m = 1, \ldots, M$, the conditional expectations

$$\hat{x}_m^{[l]}[n] = \mathrm{E}\left\{x_m[n] \mid Z[N];\boldsymbol{\theta}^{[l]}\right\} \tag{B.31}$$

$$\widehat{x_m[n]x_k[n]}^{[l]} = \mathrm{E}\left\{(x_m[n]-\hat{x}_m[n])(x_k[n]-\hat{x}_k[n]) \mid Z[N];\boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_k[n], \quad k = 1, \ldots, M \tag{B.32}$$

$$\begin{aligned} \widehat{x_m[n]x_m[n-1]}^{[l]} &= \mathrm{E}\left\{\left(x_m[n]-\hat{x_m}[n]\right)\left(x_m[n-1]-\hat{x_m}[n-1]\right) \mid Z[N];\boldsymbol{\theta}^{[l]}\right\} \\ &\quad + \hat{x}_m[n]\hat{x}_m[n-1], \quad n = 1, \ldots, N-1 \end{aligned} \tag{B.33}$$

which are used to form

$$\begin{aligned} &U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]}) = \\ &\quad c - \frac{N}{2}\log 2\pi\sigma_w^2 - \frac{1}{2\sigma_w^2}\sum_{n=0}^{N-1}\left[z^2[n] - 2z[n]\sum_{m=1}^{M}\hat{x}_m^{[l]}[n] + \sum_{m=1}^{M}\sum_{k=1}^{M}\widehat{(x_m[n]x_k[n])}^{[l]}\right] \\ &\quad - \frac{1}{2}\left[NM\log\sigma^2 - \frac{NM}{2}(M + 2M_L - 1)\gamma\log\Delta\right. \\ &\quad \left. + \frac{1}{\sigma^2}\sum_{m=1}^{M}\frac{\Delta^{\gamma(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right)\right] \end{aligned} \tag{B.34}$$

M step

The M step maximizes $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]})$ as given by (B.34). We differentiate $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]})$ with respect to each of the parameters of $\boldsymbol{\theta}$ to obtain the local extrema. This maximization can be separated into two independent sections since $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]})$ can be expressed as the sum of two terms by (B.12).

The first term is described by (B.13) and is dependent only on σ_w^2. The maximization

of this step gives us

$$\hat{\sigma}_w^2 = \frac{1}{N}\sum_{n=0}^{N-1}\left[z^2[n] - 2z[n]\sum_{m=1}^{M}\hat{x}_m^{[l]}[n] + \sum_{m=1}^{M}\sum_{k=1}^{M}(x_m[n]x_k[n])^{\wedge[l]}\right]. \tag{B.35}$$

The second term is described by (B.27) and is dependent on γ and σ^2. Differentiation with respect to each parameter gives us

$$\frac{NM\hat{\sigma}^2}{2}(M + 2M_L - 1) = \sum_{m=1}^{M}(m + M_L - 1)\frac{\Delta^{\hat{\gamma}(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) \tag{B.36}$$

$$\hat{\sigma}^2 = \frac{1}{NM}\sum_{m=1}^{M}\frac{\Delta^{\hat{\gamma}(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) \tag{B.37}$$

Eliminating $\hat{\sigma}$ we obtain that $\hat{\gamma}$ is the solution of the polynomial equation

$$\sum_{m=1}^{M}\left(m - \frac{M+1}{2}\right)\frac{\Delta^{\hat{\gamma}(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right) = 0. \tag{B.38}$$

The solution to this polynomial equation is unique.

B.2 $1/f$ Signal and Deterministic Signal

In this section, we derive the EM algorithm for the estimation of a signal and noise parameters $\boldsymbol{\theta} = \{\lambda_1, \ldots, \lambda_P, \gamma, \sigma^2, \sigma_w^2\}$ for the case of a $1/f$ signal and deterministic signal in white noise, presented in Section 5.2. As before, we denote the MN-length column vector

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_M} \end{bmatrix} \tag{B.39}$$

which represents the component processes of the $1/f$ signal. We define the incomplete data as the observed signal $\boldsymbol{z}$. The complete data is defined as the samples of the observed signal $\boldsymbol{z}$ together with the samples of each individual component process of the $1/f$ signal:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{x} \end{bmatrix}. \tag{B.40}$$

E step

The E step for this situation can be broken into two steps. First, we form a modified observation sequence

$$Z'[N] = \{z'[0], \ldots, z'[N]\} \tag{B.41}$$

which consists of the observations of the $1/f$ signal and deterministic signal in white noise with the current estimate of the deterministic signal removed

$$z'[n] = z[n] - \sum_{p=1}^{P} \lambda_p^{[l]} b_p[n]. \tag{B.42}$$

Second, we compute the conditional expectations of the $1/f$ signal components with the Kalman smoothing equations, using the modified observation sequence. This computation is identical to the computation performed in Appendix B.1

From (5.29), (5.30), and (B.1),

$$\log f_{Z|X}(\boldsymbol{z} \mid \boldsymbol{x}; \boldsymbol{\theta}) = \frac{N}{2} \log 2\pi\sigma_w^2 - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} \left[z[n] - \sum_{m=1}^{M} x_m[n] - \sum_{p=1}^{P} \lambda_p b_p[n] \right]^2 \tag{B.43}$$

Taking the conditional expectation given $\boldsymbol{Z} = \boldsymbol{z}$ at parameter value $\boldsymbol{\theta}^{[l]}$

$$\begin{aligned} &\mathrm{E}\left\{ \log f_{Z|X}(\boldsymbol{z} \mid \boldsymbol{x}; \boldsymbol{\theta}) \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]} \right\} \\ &= -\frac{N}{2} \log 2\pi\sigma_w^2 - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} \left[z^2[n] - 2z[n] \sum_{p=1}^{P} \lambda_p^{[l]} b_p[n] + \sum_{p=1}^{P} \sum_{q=1}^{Q} \lambda_p^{[l]} \lambda_q^{[l]} b_p[n] b_q[n] \right. \\ &\quad \left. - 2z[n] \sum_{m=1}^{M} \hat{x}_m^{[l]}[n] + 2 \sum_{m=1}^{M} \sum_{p=1}^{P} \lambda_p^{[l]} \hat{x}_m^{[l]}[n] b_p[n] + \sum_{m=1}^{M} \sum_{k=1}^{M} (x_m[n] x_k[n])\hat{}^{[l]} \right] \end{aligned} \tag{B.44}$$

where we define

$$(\dot{})^{[l]} \equiv \mathrm{E}\left\{ \cdot \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]} \right\} \tag{B.45}$$

These conditional expectations can be computed using the Kalman smoother with a $2M$-state system given by equations (4.11), (4.12), (4.14) and (4.17), describing the $1/f$ signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$ with white noise with parameter $\sigma_w^{2[l]}$. The input for the

Kalman smoothing equations is the modified observation sequence $Z'[N]$. Specifically,

$$\hat{x}_m^{[l]}[n] = \mathrm{E}\left\{x_m[n] \mid Z'[N]; \boldsymbol{\theta}^{[l]}\right\} \tag{B.46}$$

is a single element of the state estimate $\hat{\boldsymbol{x}}[n \mid N]$ given by the smoothing equation (4.36), and

$$\widehat{(x_m[n]x_k[n])}^{[l]} = \mathrm{E}\left\{(x_m[n] - \hat{x}_m[n])(x_k[n] - \hat{x}_k[n]) \mid Z'[N]; \boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_k[n] \tag{B.47}$$

where the first term in the sum is a single element of the error covariance matrix $R_x[n \mid N]$ given by the smoothing equation (4.37).

The probability density function of the component processes has the same form as given in (B.18). As in (B.27),

$$\begin{aligned}\mathrm{E}\left\{\log f_X(\boldsymbol{x};\boldsymbol{\theta}) \mid \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} &= c - \frac{1}{2}\left[NM\log\sigma^2 - \frac{NM}{2}(M + 2M_L - 1)\gamma\log\Delta\right. \\ &\quad \left. + \frac{1}{\sigma^2}\sum_{m=1}^{M}\frac{\Delta^{\gamma(m+M_L-1)}}{q_m(\Delta)}\mathrm{tr}\left(H_m^{-1}\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T}\right)\right]\end{aligned} \tag{B.48}$$

where $\widehat{\boldsymbol{x}_m\boldsymbol{x}_m^T} = \mathrm{E}\left\{\boldsymbol{x}_m\boldsymbol{x}_m^T \mid \boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\}$. As before, we calculate

$$\mathrm{E}\left\{(x_m[n] - \hat{x}_m[n])^2 \mid \boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]^2 \tag{B.49}$$

and

$$\mathrm{E}\left\{(x_m[n] - \hat{x_m}[n])\,(x_m[n-1] - \hat{x_m}[n-1]) \mid \boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\} + \hat{x}_m[n]\hat{x}_m[n-1]. \tag{B.50}$$

where we define $\hat{x}_m[n] = \mathrm{E}\left\{x_m[n] \mid \boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\theta}^{[l]}\right\}$. The quantities in (B.49) and (B.50) can be efficiently computed using the $2M$ state Kalman smoothing equations (4.31)–(4.38) with the system described by equations (4.11), (4.12), (4.14) and (4.17), again describing the $1/f$ signal with parameters $\gamma^{[l]}$ and $\sigma^{2[l]}$ with white noise with parameters $\sigma_w^{2[l]}$. The input into the Kalman filter is the modified observation sequence $Z'[N]$.

M step

The M step maximizes $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]})$. We differentiate $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]})$ with respect to each of the parameters of $\boldsymbol{\theta}$ to obtain the local extrema. This maximization can be separated into two independent sections since $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{[l]})$ can be expressed as the sum of two terms by (B.12).

The first term is described by (B.44) and is dependent only on the white noise parameter σ_w^2 and the deterministic signal parameters $\{\lambda_1, \ldots, \lambda_P\}$. The maximization of this step gives us

$$\begin{aligned}
\sigma_w^{2[l+1]} &= \frac{1}{N}\sum_{n=0}^{N-1}\Bigg[z^2[n] - 2z[n]\sum_{p=1}^{P}\lambda_p^{[l]}b_p[n] \\
&\quad + \sum_{p=1}^{P}\sum_{q=1}^{Q}\lambda_p^{[l]}\lambda_q^{[l]}b_p[n]b_q[n] - 2z[n]\sum_{m=1}^{M}\hat{x}_m^{[l]}[n] \\
&\quad + 2\sum_{m=1}^{M}\sum_{p=1}^{P}\lambda_p^{[l]}\hat{x}_m^{[l]}[n]b_p[n] + \sum_{m=1}^{M}\sum_{k=1}^{M}(x_m[n]x_k[n])\hat{}^{[l]}\Bigg] \qquad \text{(B.51)} \\
\lambda_p^{[l+1]} &= \left(\sum_{n=1}^{N}b_p^2[n]\right)^{-1}\sum_{n=1}^{N}\left[b_p[n]\left(z[n] - \sum_{k\neq p}\lambda_k^{[l]}b_k[n] - \sum_{m=1}^{M}\hat{x}_m^{[l]}[n]\right)\right] \qquad \text{(B.52)}
\end{aligned}$$

for $p = 1, \ldots, P$.

The second term is described by (B.48) and is dependent on γ and σ^2. This maximization is identical to the corresponding maximization in the M step of Appendix B.1.

Bibliography

[1] Brian D. O. Anderson and John B. Moore. *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, NJ, 1979.

[2] R. J. Barton and V. H. Poor. Signal detection in fractional Gaussian noise. *IEEE Transactions on Information Theory*, IT-34:943–959, September 1988.

[3] D. C. Champeney. *A Handbook of Fourier Theorems.* Cambridge University Press, Cambridge, England, 1987.

[4] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society*, pages 1–38, 1977.

[5] Mohamed Deriche and Ahmed H. Tewfik. Maximum likelihood estimation of the parameters of discrete fractionally differenced gaussian noise process. *IEEE Transactions on Signal Processing*, 41(10):2977–2989, October 1993.

[6] Meir Feder. *Statistical Signal Processing using a Class of Iterative Estimation Algorithms.* PhD thesis, MIT, September 1987.

[7] Meir Feder and Ehud Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Transactions ASSP*, 36(4):477–489, April 1988.

[8] C. W. Granger and R. Joyeux. An introduction to long memory time series models and fractional differencing. *J. Time Series Anal.*, 1(1), 1980.

[9] Carl W. Helstrom. *Probability and Stochastic Processes for Engineers.* Prentice Hall, Englewood Cliffs, NJ, 1991.

[10] J. R. M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

[11] Marvin S. Keshner. $1/f$ noise. *Proceedings of the IEEE*, 70(3):212–218, March 1982.

[12] Warren M. Lam. *Multiscale Methods for the Analysis and Application of Fractal Point Processes and Queues.* PhD thesis, MIT, January 1997.

[13] 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 Transactions on Medical Imaging*, MI-5:152–161, September 1986.

[14] B. B. Mandelbrot. Some noises with $1/f$ spectrum, a bridge between direct current and white noise. *IEEE Trans. Inform. Theory*, IT-13:289–298, 1967.

[15] B. B. Mandelbrot and H. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–436, October 1968.

[16] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1983.

[17] A. van der Ziel. On the noise spectra of semi-conductor noise and of flicker effect. *Physica*, 16(4):359–372, 1950.

[18] E. Weinstein, A.V. Oppenheim, M. Feder, and J. R. Buck. Iterative and sequential algorithms for multi-sensor signal enhancement. *IEEE Transactions on Signal Processing*, 42(3):846–859, April 1994.

[19] G. W. Wornell. A Karhunen-Loeve-like expansion for $1/f$ processes via wavelets. *IEEE Trans. Inform. Theory*, 36:859–861, July 1990.

[20] G. W. Wornell. *Signal Processing with Fractals: A Wavelet-Based Approach.* Prentice-Hall, Upper Saddle River, NJ, 1995.