International Journal of Wireless Communications and Mobile Computing
2021; 8(2): 22-38
http://www.sciencepublishinggroup.com/j/wcmc
doi: 10.11648/j.wcmc.20200802.12
ISSN: 2330-1007 (Print); ISSN: 2330-1015 (Online)

Asymptotically Optimal Low-Power Digital Filtering Using Adaptive Approximate Processing

Jeffrey Ludwig

Department of Mathematics, University of California, Irvine, United States of America

Email address:
jtludwig@uci.edu

To cite this article:
Jeffrey Ludwig. Asymptotically Optimal Low-Power Digital Filtering Using Adaptive Approximate Processing. *International Journal of Wireless Communications and Mobile Computing*. Vol. 8, No. 2, 2021, pp. 22-38. doi: 10.11648/j.wcmc.20200802.12

Received: May 28, 2021; **Accepted:** June 15, 2021; **Published:** July 13, 2021

Abstract: Techniques for reducing power consumption in digital circuits have become increasingly important because of the growing demand for portable multimedia devices. Digital filters, being ubiquitous in such devices, are a prime candidate for low-power design. We present a new algorithmic approach to low-power frequency-selective digital filtering which is based on the concepts of adaptive approximate processing. This approach is formalized by introducing the class of approximate filtering algorithms in which the order of a digital filter is dynamically varied to provide time-varying stopband attenuation in proportion to the time-varying signal-to-noise ratio (SNR) of the input signal, while maintaining a fixed SNR at the filter output. Since power consumption in digital filter implementations is proportional to the order of the filter, dynamically varying the filter order is a strategy which may be used to conserve power. From this practical technique we abstract a theoretical problem which involves the determination of an optimal filter order based on observations of the input data and a set of concrete assumptions on the statistics of the input signal. Two solutions to this theoretical problem are presented, and the key results are used to interpret the solution to the practical low-power filtering problem. We construct a framework to explore the statistical properties of approximate filtering algorithms and show that under certain assumptions, the performance of approximate filtering algorithms is asymptotically optimal.

Keywords: Low-power Signal Processing, Adaptive Filtering, Approximate Signal Processing

1. Introduction

Adaptive filtering and energy efficiency has been a topic of great interest in recent years. For example, the interest in reducing the power consumption of digital filters used in edge computing and sensor networks is growing rapidly [1]. A fundamental tradeoff between power consumption and accuracy has been utilized to determine energy-optimum design parameters for deep in-memory architectures for efficient hardware realizations of machine learning algorithms [2]. Remarkably, in a recent paper the authors note that "optimization for power is one of the most important design objectives in modern digital signal processing (DSP) applications," and then demonstrate the efficacy of a hybrid energy-efficient methodology for digital finite duration impulse response (FIR) filters [3].

Approximate signal processing may be used for applications in which it is desirable to dynamically adjust the quality of signal processing results to the availability of resources, such as time, bandwidth, memory, and power. Recently, excellent research has been accomplished in the area of incremental refinement structures for approximate signal processing in the context of power-efficient approximate multiplication for applications in which exact computation is not necessary [4-8]. The design of an energy efficient digital IIR filter using approximate multiplication, with a similar objective to the IIR filters presented in this paper, is presented in the research by Pilipović et al. [1]. Furthermore, approximate computing represents another recent innovation for minimizing power consumption via a variety of methods for reducing arithmetic activity [9-12].

The incessant demand for low-power signal processing algorithms has been persistent and consistent over the past three decades. The realm of energy efficiency has only grown in its intensity and breadth. In intensity, low-power speech recognition is now popular in mobile devices [13, 14]. The

evolution of energy efficient engineering has moved along stunningly creative paths to wearable devices [15], ingestible electronics [16, 17] and cryptographic engines [18].

In this paper we consider the practical problem of dynamically reducing the order of a frequency-selective digital filter, omnipresent in all of the previously mentioned state-of-the-art signal processing electronic systems, to conserve power. We will demonstrate that it is possible to dynamically vary the stopband attenuation provided by a digital filter to obtain the minimum amount of attenuation needed to continuously maintain a given output signal-to-noise ratio (SNR), and show that such adaptive approximate filtering algorithms significantly reduce the required average power consumption relative to that of conventional fixed-order filtering algorithms. From this practical problem we abstract a theoretical problem which involves the determination of an optimal filter order based on observations of the input data and a set of concrete assumptions on the statistics of the input signal. Two solutions to this theoretical problem will be presented, and the key results will be used to interpret the solution to the practical low-power filtering problem.

An underlying theory for approximate filtering is developed. We construct a framework to explore the statistical properties of approximate filtering algorithms, and show that under certain assumptions the performance of approximate filtering algorithms is asymptotically optimal. The focus of the algorithm development is on applications involving frequency-selective digital filtering in which the goal is to reject one or more frequency bands while keeping the remaining portions of the input spectrum largely unaltered. Examples of such applications include lowpass filtering for signal upsampling and downsampling, bandpass filtering for subband coding, and lowpass filtering for frequency-division multiplexing and demultiplexing. In addition, approximate filtering algorithms appear to be useful in other domains in which digital filters are used such as prediction, smoothing, echo cancellation, or equalization.

1.1. Overview

A brief summary of approximate filtering is now given. The basic idea is to begin filtering a given input signal with a frequency-selective digital filter of some nominal order, as shown in Figure 1. This filter has well-defined passband and stopband regions in frequency. After L output samples have been produced, we use the most recent block of L input and output samples to form an easily computable low-power estimate of the current input SNR, defined as the ratio of the input signal power in the passband of the filter to the input signal power in the stopband of the filter. In Figure 1 the decision module D uses the signal power estimates $\hat{P}_x$ and $\hat{P}_y$ to form an estimate of the temporally local input SNR. This estimate of the input SNR is then used to update the filter order to be the minimum value which guarantees that the output SNR, defined as the ratio of the output signal power in the passband of the filter to the output signal power in the stopband of the filter, will be greater than or equal to a pre-specified minimum tolerable output SNR. This filter order is then used to produce another block of L output samples, and the filter order update process is repeated.

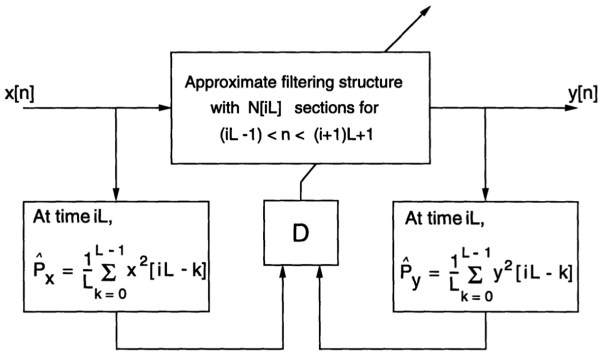

Figure 1. *An overview of approximate filtering. The adaptation strategy for updating the filter order after each new set of L output samples is defined by the decision module D.*

A key issue addressed in this paper is how well the low-power estimate of the filter order converges to the theoretical minimum order for situations satisfying certain statistical assumptions which are made in the derivation of the underlying theoretical framework for approximate filtering. Computer simulations are used to verify analytical results which we obtain in this paper that show that convergence to the correct filter order depends upon: 1) the number L of input and output samples used in estimating the input SNR, 2) the nominal order of the filter applied in generating the output samples that are used in estimating the input SNR, and 3) the proximity of the true input SNR to the boundaries in the input SNR space corresponding to changes in the optimal choice of filter order [19].

The adaptation mechanism used with an approximate filtering structure is designed to determine and use the filter with the smallest order while ensuring that the approximate filter output meets a pre-specified quality constraint. Minimization of the filter order used at any given time is desirable because of the resulting savings in power consumption by the underlying hardware [20]. The output quality criterion we use is designed to keep the output SNR (the ratio of the passband power to the stopband power in the filter output) above specified level. Other output quality constraints could easily be incorporated with minor modifications. One possible alternative output quality constraint is to keep the output signal power in the stopband of the filter below some pre-specified level [20].

1.2. Approximate Filter Structures

In approximate filtering algorithms the order of a frequency-selective digital filter is varied in a way defined by a control strategy and an approximate filter structure. A collection of frequency-selective digital filters, one for each filter order N in a given range $N_{\min} < N < N_{\max}$, constitutes an approximate filter structure $\mathcal{H}$. Each filter structure $\mathcal{H}$

must possess the property that its progressively higher order filters have progressively increased average attenuation in the stopband region(s) while maintaining close to unity gain in the passband region(s). Approximate filter structures represent an important element in the characterization of approximate filtering algorithms. Two classes of approximate filter structures, *truncation* and *replacement* filter structures, are important and used as a basis for classifying all approximate filter structures into one of four types [21]. A replacement filter structure is characterized by the relationship between the coefficients of filters of different orders being completely unconstrained; the coefficients of each individual filter may be selected or replaced independently. In a truncation filter structure this is not allowed. For a truncation filter structure with FIR constituent filter elements, the coefficients defining the lower order filters are constrained to be subsets of the coefficients defining the filter with maximum order $N_{\max}$. Similarly, in a truncation filter structure with IIR constituent elements, the pole/zero pairs defining the lower order filters are constrained to be subsets of the pole/zero pairs defining the filter with maximum order. Thus the lower order constituent elements in a truncation filter structure are truncated versions of the higher order constituent elements. It is clear that truncation filter structures may be described with fewer independent filter coefficients than replacement filter structures. Associated with this property is the fact that truncation filter structures require less memory, chip area, and bus accesses than replacement filter structures.

The frequency response magnitudes of the filters drawn from an exemplary approximate filter structure based on truncations of an IIR Butterworth filter are shown in Figure 2. The half-power frequency of the Butterworth filters is $\pi/2$. In this figure we show the magnitude-squared frequency responses for truncations of a 20th-order Butterworth filter with 3, 5, 7, 9, and 10 second-order sections. The key feature of an approximate filter structure is that the higher-order filters provide higher average stopband attenuation and thus have lower stopband power than the lower-order filters. This feature is clearly illustrated in Figure 2, and allows us incorporate a tradeoff between filter quality and filter cost into approximate filtering algorithms. The filter quality is measured by the average stopband attenuation, while the filter cost is measured by the required power consumption or equivalently the required filter order. The passband PB, stopband SB, and transition band TB regions for all filters in the approximate filter structure $\mathcal{H}$ are identical. The passband and stopband regions must be explicitly specified in the definition of an approximate filter structure, and by default the transition band is defined to span the remaining portions of the spectrum $\omega \in [-\pi, \pi]$ which are not included in the passband or stopband regions. Each of the individual filters which make up the constituent elements of the approximate filter structure $\mathcal{H}$ must be properly normalized. Possible normalizations include a unit energy normalization or a unity DC (zero frequency) gain normalization.

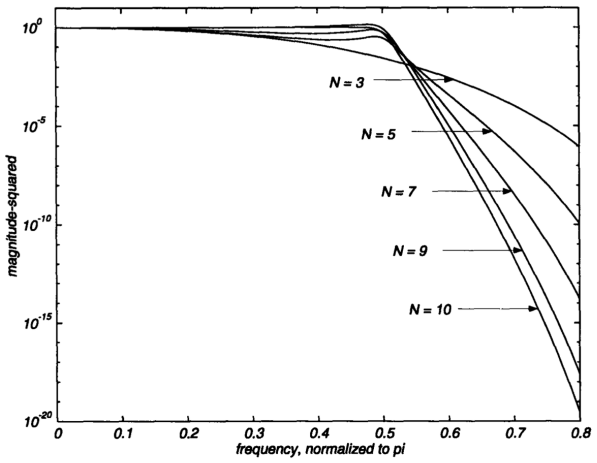

Figure 2. *Magnitude-squared frequency responses for truncations of a 20th-order Butterworth filter with 3, 5, 7, 9, and 10 second-order sections. The half-power frequency of the Butterworth filters is $\pi/2$.*

We have stated our intention to study the problem of dynamically reducing the order of a frequency-selective digital filter to conserve power while maintaining a desired level of output quality. We have stressed that the key is use vary the filter order over time to provide time-varying stopband attenuation in proportion to the time-varying SNR of the input signal, while maintaining a fixed SNR at the filter output. From the practical problem of dynamically reducing the order of a frequency-selective digital filter to conserve power, in the next section we abstract and explore an intimately related theoretical problem. The solutions to this theoretical problem will provide a basis for understanding and analyzing approximate filtering algorithms.

2. Problem Statement

In this section we introduce a theoretical problem, termed the *approximate filtering problem*, which involves the determination of an *optimal filter order* based on observations of input data and a set of concrete assumptions on the statistics of the input signal. Two solutions to this theoretical problem are presented. One solution is guided by a low-power approach and achieves suboptimal performance with an extremely low computational cost. A second solution is guided by a maximum likelihood objective and provides superior performance while requiring much more computation. While computationally impractical, the maximum likelihood approach provides valuable insight as well as a performance benchmark for comparison with the low-power solution. The key theoretical results are used to interpret the entire class of approximate filtering algorithms.

The fundamental theoretical problem we address in this paper is

Given

1. a set of L input samples $x[0] \cdots x[L-1]$ from a wide sense stationary (WSS) Gaussian random process $x[n]$

with power spectral density $S_x(\omega)$
2. a filter order set $\mathcal{N} = N_{\min} \cdots N_{\max}$ containing M elements
3. a filter structure $\mathcal{H} = \{h_{N\min}[n] \cdots h_{N\max}[n]\}$, containing M frequency-selective filter elements, all with passband $\omega \in PB$ and stopband $\omega \in SB$
4. a passband defined as $\omega \in PB$ for each filter in the filter structure $\mathcal{H}$
5. a stopband defined as $\omega \in SB$ for each filter in the filter structure $\mathcal{H}$
6. a minimum tolerable output signal-to-noise ratio $\mathrm{OSNR}_{\mathrm{tol}}$

Determine

The optimal filter order N^*, defined as the minimum order $N \in \mathcal{N}$ of the frequency-selective filter $h_N[n] \in \mathcal{H}$ which provides sufficient stopband attenuation to assure

$$\mathrm{OSNR}[N] \geq \mathrm{OSNR}_{\mathrm{tol}}, \tag{1}$$

where the output signal-to-noise ratio (SNR) is defined as

$$\mathrm{OSNR}[N] \triangleq \frac{P_y^{PB[N]}}{P_y^{SB[N]}}. \tag{2}$$

The output power spectral density is

$$S_y(\omega) = |H_N(\omega)|^2 S_x(\omega), \tag{3}$$

the output power in the passband is

$$\begin{aligned} P_y^{PB}[N] &= \frac{1}{2\pi}\int_{PB} S_y(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{PB} |H_N(\omega)|^2 S_x(\omega)d\omega, \end{aligned} \tag{4}$$

and the output power in the stopband is

$$\begin{aligned} P_y^{SB}[N] &= \frac{1}{2\pi}\int_{SB} S_y(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{SB} |H_N(\omega)|^2 S_x(\omega)d\omega. \end{aligned} \tag{5}$$

We refer to the problem in (1)-(5), as the *approximate filtering problem*. Any particular solution to the problem involves defining a method for reliably estimating the optimal filter order N^* based on observations of the input sequence $x[0] \cdots x[L-1]$. We measure the performance of a particular solution in terms of how accurately, on average, the solution determines the correct value for N^*. In this paper we will present two distinct solutions to the approximate filtering problem. The first is guided by a low-power consumption constraint and thus produces estimates of N^* which require very low computational overhead. We call this the *low-power* (LP) *solution* to the approximate filtering problem. The second solution we present is motivated by a maximum likelihood formulation and is therefore termed the *maximum likelihood* (ML) *solution* to the approximate filtering problem. We will show that the ML solution requires much more computation and offers slightly better performance than the LP solution in certain circumstances. While computationally prohibitive for practical use, the ML solution is conceptually insightful as its performance may be used as a meaningful benchmark for comparison with the performance of the low-power solution.

2.1. Summary of Low-power Approach

The low-power (LP) approach to finding N^* is computationally simple and will be shown to perform almost equally as well as the computationally prohibitive ML-based approach, even for signals statistically tailored to favor the ML-based approach.

A research group at Stanford University has designed and implemented an approximate filtering algorithm for the application of low-power interpolation and decimation. Excellent results have been obtained with a conventional interpolation/decimation system that was implemented and shown to consume an average of approximately 86.4 mW with a 5V power supply. Using an approximate filtering algorithm compiled onto a programmable processor, the power consumption was reduced by by 36% for the decimation system and 17% for the interpolation system [22].

The LP approach computes an estimate of the output SNR based on an estimate of the total *power difference* between the input and output of a digital frequency-selective filter. Conceptually, if the frequency-selective filter were an ideal piecewise-constant filter

$$H_{\mathrm{ideal}}(\omega) = \begin{cases} 1 & \text{if } \omega \in PB \\ 0 & \text{if } \omega \in SB \end{cases} \tag{6}$$

then the total power difference between the input and output signals would be equal to the exact input signal power in the stopband of the ideal filter. This is true since an ideal filter perfectly eliminates all the components of the input signal in its stopband, and passes with unity gain all the components of the input signal in its passband. Thus, ideally the output signal contains the exact passband components of the input signal, and nothing else. The input signal obviously contains both its passband and stopband signal components. Therefore the difference in input and output signal powers ideally gives the exact power in the stopband of the input signal. From this we may form a LP estimate $\widehat{\mathrm{OSNR}}_{\mathrm{LP}}[N, N_0]$ of the output SNR, and proceed to estimate N^* via (1) with $\mathrm{OSNR}[N]$ replaced by $\widehat{\mathrm{OSNR}}_{\mathrm{LP}}[N, N_0]$.

Ideal filters are not practically realizable. However, if we use a non-ideal filter which approximates an ideal filter, our LP estimate of the output SNR based on the difference in input and output signal powers will approximate the true output SNR. To the extent that this is a good approximation, the LP approach to computing N^* using (1) will be effective. The detailed derivation of the LP solution to the approximate filtering problem will be given in Section 3. A summary of the final result is presented now.

The LP estimate $\hat{N}^*_{\mathrm{LP}}$ of the optimal filter order N^* is determined by searching for the minimum value of $N \in \mathcal{N}$ satisfying

$$\widehat{\mathrm{OSNR}}[N, N_0] \geq \mathrm{OSNR}_{\mathrm{tol}}, \tag{7}$$

where the LP estimate of the output SNR is defined as

$$\widehat{\mathrm{OSNR}}[N, N_0] = \left[\frac{\mathbf{y}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}}\right] - \left(\frac{\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right] d\omega}{\int_{SB}\left|H_N(\omega)^2\right| d\omega}\right) - \left(\frac{\int_{SB}|H_{N_0}(\omega)|^2}{\int_{SB}|H_N(\omega)|^2}\right). \tag{8}$$

The expression in (8) may be rearranged with simple algebraic manipulations and substitutions to simplify the decision rule for selecting $\hat{N}^*_{\mathrm{LP}}$ to be the minimum filter order $N \in \mathcal{N}$ satisfying

$$R \geq R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}], \tag{9}$$

where R is the ratio

$$R = \left[\frac{\mathbf{y}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}}\right], \tag{10}$$

and $R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}]$ is a function of N parameterized by N_0 and $\mathrm{OSNR}_{\mathrm{tol}}$

$$R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}] = \left\{\mathrm{OSNR}_{\mathrm{tol}} + \left(\frac{\int_{SB}|H_{N_0}(\omega)|^2 d\omega}{\int_{SB}\left|H_N(\omega)^2\right| d\omega}\right)\right\} \times \frac{\int_{SB}|H_N(\omega)|^2 d\omega}{\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right] d\omega}$$

The signal vectors are $\mathbf{x} = [\,x[0]\,x[1]\,\cdots\,x[L-1]\,]^T$ and $\mathbf{y} = [\,y[0]\,y[1]\,\cdots\,y[L-1]\,]^T$, where the power window length L is the number of output samples which are produced by the filter $h_{N_0}[n] \in \mathcal{H}$ before $\hat{N}^*_{\mathrm{LP}}$ is computed. The nominal filter order N_0 may be chosen to be equal to any filter order $N \in \mathcal{N}$; however, as we shall see later, the choice $N_0 = N_{\max}$ produces the best results.

The LP solution to the approximate filtering problem takes advantage of the fact that during filtering L samples of the output $y[n]$ of the filter $h_{N_0}[n]$ are available, without additional computational cost. These samples are used to form the vector $\mathbf{y}$ and the ratio R. The function $R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}]$ is a function of N assuming N_0 and $\mathrm{OSNR}_{\mathrm{tol}}$ have been fixed, and its values may be easily computed and stored in advance. The rule for computing $\hat{N}^*_{\mathrm{LP}}$ is to search among the stored values of $R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}]$ to find amongst all those which satisfy $R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}] \leq R$ the unique one which corresponds to the minimum value of N. This value of N is defined to be $\hat{N}^*_{\mathrm{LP}}$.

We now consider a second solution to the approximate filtering problem, using an ML-based approach. This solution does not use the available output signal $y[n]$ to compute its estimate of N^*. In certain situations the ML solution will achieve better estimates of N^* than the LP solution, but this performance advantage comes at the expense of requiring more computation.

2.2. Summary of Maximum Likelihood Approach

An alternative strategy for determining the optimal filter order N^* defined in (1) involves computing an estimate $\hat{S}_x(\omega)$ of the input power spectral density (PSD) $S_x(\omega)$ from observations $x[0]\,x[1]\,\cdots\,x[L-1]$ of the WSS input random process $x[n]$, and using this PSD estimate to compute an estimate of the optimal filter order N^*. Because the PSD estimate $\hat{S}_x(\omega)$ is based on ML estimates of the all-pole parameters of the underlying random process which is assumed to be autoregressive (AR), we call this the ML approach. A summary of the final result of the ML approach to the approximate filtering problem is now presented, with the details of the derivation presented in Section 4.

Assuming that $S_x(\omega)$ corresponds to a pth-order AR random process, the ML-based estimate $\hat{S}x(\omega)$ is defined as

$$\hat{S}x(\omega) = \hat{\sigma}^2_u \left|1 + \sum_{m=1}^{p} \hat{a}_m e^{-j\omega m}\right|^{-2}, \tag{11}$$

where the ML-based estimates $\hat{\mathbf{a}} = [\,\hat{a}_1\,\hat{a}_2\,\cdots\,\hat{a}_p\,]^T$ and $\hat{\sigma}^2_u$ for the parameters $\mathbf{a} = [\,a_1\,a_2\cdots a_p]^T$ and σ^2_u are computed via the well-known Yule-Walker equations. This method for AR parameter estimation is well known as the autocorrelation method. The final result is that an ML estimate $\hat{N}^*_{\mathrm{ML}}$ of N^* may be computed via selecting the minimum value of $N \in \mathcal{N}$ satisfying

$$\widehat{\mathrm{OSNR}}_{\mathrm{ML}}[N] \geq \mathrm{OSNR}_{\mathrm{tol}}, \tag{12}$$

where the ML estimate of the output SNR is defined as

$$\widehat{\mathrm{OSNR}}_{\mathrm{ML}}[N] \triangleq \frac{[\hat{P}^{PB}_y[N]]_{\mathrm{ML}}}{[\hat{P}^{SB}_y[N]]_{\mathrm{ML}}}, \tag{13}$$

the ML estimate of the output power in the passband is defined

as

$$[\hat{P}_y^{PB}[N]]_{\mathrm{ML}} = \frac{1}{2\pi}\int_{PB}[\hat{S}_y(\omega)]_{\mathrm{ML}}d\omega$$
$$= \frac{1}{2\pi}\int_{PB}[\hat{S}_x(\omega)]_{\mathrm{ML}}|H_N(\omega)|^2 d\omega, \quad (14)$$

and the ML estimate of the output power in the stopband is defined as

$$[\hat{P}_y^{SB}[N]]_{\mathrm{ML}} = \frac{1}{2\pi}\int_{SB}[\hat{S}_y(\omega)]_{\mathrm{ML}}d\omega$$
$$= \frac{1}{2\pi}\int_{SB}[\hat{S}_x(\omega)]_{\mathrm{ML}}|H_N(\omega)|^2 d\omega. \quad (15)$$

Once the ML estimate $\hat{S}_x(\omega)$ of $S_x(\omega)$ has been computed, the quantities $[\hat{P}_y^{PB}[N]]_{\mathrm{ML}}$, $[\hat{P}_y^{SB}[N]]_{\mathrm{ML}}$, and $\widehat{\mathrm{OSNR}}_{\mathrm{ML}}[N]$ may be evaluated for each value of $N \in \mathcal{N}$ and frequency-selective filter $h_N[n] \in \mathcal{H}$, and $\hat{N}^*_{\mathrm{ML}}$ may be determined via (12). This approach will be shown to experimentally produce excellent estimates of N^*, especially when the input signal is synthetically generated to be a true AR WSS random process. This is an intuitively natural result. Unfortunately the ML approach is not practically viable due to its excessive computational requirements. However it will serve as a meaningful benchmark for performance comparison with the LP approach which is the focal point of this paper.

Having now formally presented the statement of the approximate filtering problem as well as overviews of the two solutions which are developed in this paper, we move on to the detailed derivations of each solution. First we formulate the LP estimate $\hat{N}^*_{\mathrm{LP}}$ in Section 3, and then we formulate the ML estimate $\hat{N}^*_{\mathrm{ML}}$ in Section 4. The capabilities of this algorithm for reducing power consumption in digital filtering applications are truly remarkable.

3. Derivation of Low-power Solution

In this section we develop the low-power (LP) solution to the approximate filtering problem summarized in (1). The LP solution provides a method for computing the estimate $\hat{N}^*_{\mathrm{LP}}$ of the optimal filter order N^* based on low-power operations. The LP method is necessarily computationally simple, and thus it has the advantage of requiring significantly less average power than the ML-based solution which will be given in Section 4. After presenting some underlying assumptions, our approach to deriving an expression for $\hat{N}^*_{\mathrm{LP}}$ begins with determining a LP estimate $\widehat{ISNR}_{\mathrm{LP}}[N_0]$ of the input SNR based on the difference between the input power and the output power of a frequency-selective filter $h_{N_0}[n] \in \mathcal{H}$ with nominal order $N_0 \in \mathcal{N}$. We use our estimate $\widehat{ISNR}_{\mathrm{LP}}[N_0]$ to produce an expression for a LP estimate $\mathrm{OSNR}_{\mathrm{LP}}[N, N_0]$ of the output SNR, which can be substituted into (1) for OSNR[N] to determine the LP solution $\hat{N}^*_{\mathrm{LP}}$ to the approximate filtering problem.

[1] we assume that all random processes discussed in this paper are ergodic

To begin we suppose that a discrete-time WSS random process[1] $x[n]$ with power spectrum $S_x(\omega)$ is filtered using a digital frequency-selective filter with impulse response $h_{N_0}[n] \in \mathcal{H}$ and order $N_0 \in \mathcal{N}$ to obtain an output signal $y[n]$. We assume that the filter $h_{N_0}[n]$ is taken from an approximate filter structure $\mathcal{H}$ and thus has a well-defined spectral passband PB, stopband SB, and transition band TB. We make the following key assumptions:

Assumption 1

$S_x(\omega)$ is equal to an unknown constant σ^2_{SB} in the stopband region of $\mathcal{H}$

$$S_x(\omega) = \sigma^2_{SB} \qquad \omega \in SB \quad (16)$$

Assumption 2

For frequencies in the passband of $\mathcal{H}, |H_N(\omega)|^2$ is approximately equal to unity

$$|H_N(\omega)|^2 \approx 1 \qquad \omega \in PB \quad (17)$$

This is assumed to be true for all $N \in \mathcal{N}$, and thus for all $H_N(\omega) \in \mathcal{H}$.

Assumption 3

$S_x(\omega)$ is negligible in the transition band of $\mathcal{H}$

$$S_x(\omega) \approx 0 \qquad \omega \in TB \quad (18)$$

Furthermore, this implies that

$$\int_{TB} S_x(\omega) f(\omega) d\omega \approx 0 \quad (19)$$

for any finite, continuous, differentiable function $f(\omega)$.

We note that no assumption is made about the shape of the function $S_x(\omega)$ within the passband. $S_x(\omega)$ in the passband is finite but otherwise arbitrary. Assumption 2 states that $S_x(\omega)$ is negligible in the transition band. This is reasonable for situations in which the input stopband and passband components are separated by a guard band, as is the case in a whole host of communications applications [23].

As mentioned earlier, our first step is to determine a LP estimate of the input SNR under the stated assumptions. We define the signal-to-noise ratio (SNR) as the ratio of the signal power in the passband of $\mathcal{H}$ to the signal power in the stopband of $\mathcal{H}$. The input SNR may be expressed as

$$\mathrm{ISNR} \triangleq \frac{P_x^{PB}}{P_x^{SB}}, \quad (20)$$

where

$$P_x^{PB} = \frac{1}{2\pi}\int_{PB} S_x(\omega) d\omega, \quad (21)$$

and

$$P_x^{SB} = \frac{1}{2\pi}\int_{SB} S_x(\omega) d\omega. \quad (22)$$

By invoking the Assumption 1 which states that $S_x(\omega)$ is equal to an unknown constant σ^2_{SB} in the stopband, it follows

that

$$\text{ISNR} = \frac{\frac{1}{2\pi}\int_{PB} S_x(\omega)d\omega}{\frac{1}{2\pi}\int_{SB}\sigma_{SB}^2 d\omega} = \frac{\frac{1}{2\pi}\int_{PB} S_x(\omega)d\omega}{\frac{1}{2\pi}\sigma_{SB}^2\Delta_{SB}}, \quad (23)$$

where Δ_{SB} is the spectral width of the stopband. For example, if the stopband is defined as $\pi/2 \leq |\omega| \leq \pi$, then $\Delta_{SB} = \pi$. Assumption 2 states that $|H_N(\omega)|^2 \approx 1$ for $\omega \in PB$. Since $S_y(\omega) = S_x(\omega)|H_N(\omega)|^2$, Assumption 2 implies that $S_y(\omega) \approx S_x(\omega)$ for $\omega \in PB$, and (23) becomes

$$\text{ISNR} \approx \text{ISNR}[N_0] = \frac{\frac{1}{2\pi}\int_{PB} S_y(\omega)d\omega}{\frac{1}{2\pi}\sigma_{SB}^2\Delta_{SB}} = \frac{P_y^{PB}[N_0]}{\frac{1}{2\pi}\sigma_{SB}^2\Delta_{SB}}, \quad (24)$$

where $P_y^{PB}[N_0]$ is the output power in the passband which was previously defined in (4). We note that the approximate expression in (24) for the input SNR is a function of N_0 due to its dependence on $P_y^{PB}[N_0]$. We now proceed to find an expression for $P_y^{PB}[N_0]$. First note that the total output power

$$\begin{aligned} P_y[N_0] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_y(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(\omega)|H_{N_0}(\omega)|^2 d\omega. \end{aligned} \quad (25)$$

may be written as the sum of three spectrally-disjoint components

$$P_y[N_0] = P_y^{PB}[N_0] + P_y^{SB}[N_0] + P_y^{TB}[N_0], \quad (26)$$

where $P_y^{PB}[N_0]$ is given in (4), $P_y^{SB}[N_0]$ is given in (5), and

$$\begin{aligned} P_y^{TB}[N_0] &= \frac{1}{2\pi}\int_{TB} S_y(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{TB} S_x(\omega)|H_{N_0}(\omega)|^2 d\omega. \end{aligned} \quad (27)$$

If we now invoke the Assumption 3 which states that $S_x(\omega)$ is negligible in the transition band, then $P_y^{TB}[N_0] \approx 0$, and rearranging (26) produces

$$P_y^{PB}[N_0] \approx P_y[N_0] - P_y^{SB}[N_0]. \quad (28)$$

We now examine the term $P_y^{SB}[N_0]$. Combining the definition of $P_y^{SB}[N_0]$ in (5) with Assumption 1 which states that $S_x(\omega) = \sigma_{SB}^2$ for $\omega \in SB$, we obtain

$$\begin{aligned} P_y^{SB}[N_0] &= \frac{1}{2\pi}\int_{SB} S_x(\omega)|H_{N_0}(\omega)|^2 d\omega \\ &= \frac{1}{2\pi}\sigma_{SB}^2\int_{SB} |H_{N_0}(\omega)|^2 d\omega \end{aligned} \quad (29)$$

In order to obtain a more elementary expression for $P_y^{SB}[N_0]$, it is apparent from (29) that an expression for the unknown parameter σ_{SB}^2 is needed. For this purpose we consider the difference in input and output signal power

$$P_x - P_y[N_0] = \frac{1}{2\pi}\int_{-\pi}^{\pi}[S_x(\omega) - S_y(\omega)]d\omega \quad (30)$$

where the total input signal power P_x, is defined as

$$P_x = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(\omega)d\omega. \quad (31)$$

As was shown for the total output power $P_y[N_0]$ in (26), the difference in input and output signal power may similarly be broken up into its spectrally disjoint passband, stopband, and transition band components

$$\begin{aligned} P_x - P_y[N_0] &= \frac{1}{2\pi}\int_{PB}[S_x(\omega) - S_y(\omega)]d\omega \\ &+ \frac{1}{2\pi}\int_{SB}[S_x(\omega) - S_y(\omega)]d\omega \\ &+ \frac{1}{2\pi}\int_{TB}[S_x(\omega) - S_y(\omega)]d\omega \end{aligned} \quad (32)$$

We expand this and incorporate Assumption 1 to produce

$$\begin{aligned} P_x - P_y[N_0] &= \frac{1}{2\pi}\int_{PB} S_x(\omega)\left[1 - |H_N(\omega)|^2\right]d\omega \\ &+ \frac{1}{2\pi}\sigma_{SB}^2\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right]d\omega \\ &+ \frac{1}{2\pi}\int_{TB} S_x(\omega)\left[1 - |H_{N_0}(\omega)|^2\right]d\omega \end{aligned} \quad (33)$$

We now recall Assumption 2 which states that $|H_{N_0}(\omega)|^2 \approx 1$ for $\omega \in PB$. Under this assumption the first addend in the right-hand side of (33) is approximately zero. Furthermore, Assumption 3 which states that $S_x(\omega)$ is negligible in the transition band, we may argue that the third addend in the right-hand side of (33) is approximately zero. We therefore obtain

$$P_x - P_y[N_0] \approx \sigma_{SB}^2\frac{1}{2\pi}\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right]d\omega, \quad (34)$$

which may be rearranged to yield

$$\sigma_{SB}^2 \approx (P_x - P_y[N_0])\left(\frac{1}{2\pi}\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right]d\omega\right)^{-1} \quad (35)$$

Substituting this approximate expression for σ_{SB}^2 into (29) we obtain

$$\begin{aligned} P_y^{SB}[N_0] &= \frac{1}{2\pi}\sigma_{SB}^2\int_{SB}|H_{N_0}(\omega)|^2 d\omega \\ &\approx (P_x - P_y[N_0])P_h^{SB}[N_0], \end{aligned} \quad (36)$$

where

$$\begin{aligned} P_h^{SB}[N_0] &= \left(\frac{1}{2\pi}\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right]d\omega\right)^{-1} \\ &\times \left(\frac{1}{2\pi}\int_{SB}|H_{N_0}(\omega)|^2 d\omega\right) \end{aligned} \quad (37)$$

is a particularly relevant measure of the spectral quality of the filter $H_{N_0}(\omega)$. Note that $P_h^{SB}[N_0] = 0$ in the case of an ideal filter which was defined in (6). We now incorporate (28) into (24) and obtain

$$\mathrm{ISNR}[N_0] = \frac{P_y^{PB}[N_0]}{\frac{1}{2\pi}\sigma_{SB}^2\Delta_{SB}} \approx \frac{P_y[N_0] - P_y^{SB}[N_0]}{\frac{1}{2\pi}\sigma_{SB}^2\Delta_{SB}}. \quad (38)$$

Substituting in the approximate expression in (36) for $P_y^{SB}[N_0]$ produces

$$\mathrm{ISNR}[N_0] \approx \frac{P_y[N_0] - (P_x - P_y[N_0])P_h^{SB}[N_0]}{\frac{1}{2\pi}\sigma_{SB}^2\Delta_{SB}} \quad (39)$$

Plugging in our approximate expression in (35) for σ_{SB}^2 produces

$$\mathrm{ISNR}[N_0] \approx \left[\frac{P_y[N_0]}{P_x - P_y[N_0]}\right]\left(\frac{1}{\Delta_{SB}}\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right]d\omega\right) - \left(\frac{1}{\Delta_{SB}}\int_{SB}|H_{N_0}(\omega)|^2 d\omega\right). \quad (40)$$

Armed with this expression for $\mathrm{ISNR}[N_0]$ which is valid under the stated assumptions, we now turn to the problem of computing low-power estimates of the signal-related quantities in (40): the total input power P_x and the total output power $P_y[N_0]$.

3.1. Low-power Estimation

To obtain the LP estimate $\mathrm{ISNR}_{\mathrm{LP}}[N_0]$ of the input SNR based on (40), suppose that we have applied a filter of order N_0 to the input $x[n]$ and have obtained L output samples prior to and including time n. We may then obtain the following estimates

$$\hat{P}_x = \frac{1}{L}\sum_{k=0}^{L-1} x^2[n-k] = \mathbf{x}^T\mathbf{x}, \quad (41)$$

and

$$\hat{P}_y[N_0] = \frac{1}{L}\sum_{k=0}^{L-1} y^2[n-k] = \mathbf{y}^T\mathbf{y}, \quad (42)$$

where the $L \times 1$ signal vectors are defined as $\mathbf{x} = [\,x[n-L+1]\,\cdots\,x[n-1]\,x[n]\,]^T, \mathbf{y} = [\,y[n-L+1]\,\cdots\,y[n-1]\,y[n]\,]^T$. We note that the explicit dependence of $\hat{P}_y[N_0]$ and $\hat{P}_x$ on n and L is omitted for notational simplicity. By incorporating the estimates $\hat{P}_y[N_0]$ and $\hat{P}_x$ into (40) in place of $P_y[N_0]$ and P_x, respectively, we obtain the LP estimate $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}[N_0]$ for the input SNR

$$\widehat{\mathrm{ISNR}}_{\mathrm{LP}}[N_0] = \left[\frac{\mathbf{y}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}}\right]\left(\frac{1}{\Delta_{SB}}\int_{SB}\left[1 - |H_{N_0}(\omega)|^2\right]d\omega\right) - \left(\frac{1}{\Delta_{SB}}\int_{SB}|H_{N_0}(\omega)|^2 d\omega\right). \quad (43)$$

This is our final expression for $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}[N_0]$. We observe that $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}[N_0]$ is easily computable from the signal-dependent quantities $\mathbf{y}^T\mathbf{y}$ and $\mathbf{x}^T\mathbf{x}$ and the integrals of the filter $h_{N_0}[n]$ that appear in (43).

In order to compute $\hat{N}_{\mathrm{LP}}^*$ an estimate $\widehat{\mathrm{OSNR}}_{\mathrm{LP}}[N, N_0]$ of the output SNR is required. To proceed, we define $\mathrm{SNRI}[N]$, the signal-to-noise ratio improvement factor, as the multiplicative factor by which the input SNR is multiplied by to obtain the output SNR. This will enable us to easily obtain our estimate $\widehat{\mathrm{OSNR}}_{\mathrm{LP}}[N, N_0]$ of the output SNR from our estimate $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}[N_0]$ of the input SNR via a simple multiplication by $\mathrm{SNRI}[N]$. The SNR improvement factor is clearly a function of the filter order $N \in \mathcal{N}$. The signal-to-noise ratio improvement factor satisfies the relationship

$$\mathrm{ISNR} \cdot \mathrm{SNRI[N]} = \mathrm{OSNR[N]}, \quad (44)$$

which we may rearrange with substitution of the definitions for ISNR and $\mathrm{OSNR}[N]$ from (20) and (2), respectively, to produce

$$\mathrm{SNRI}[N] = \frac{\mathrm{OSNR}[N]}{\mathrm{ISNR}} = \frac{P_y^{PB}[N]}{P_y^{SB}[N]} \cdot \frac{P_x^{SB}}{P_x^{PB}}. \quad (45)$$

If we substitute in the definitions of $P_y^{PB}[N], P_y^{SB}[N], P_x^{SB}$, and P_x^{PB}, we arrive at

$$\mathrm{SNRI}[N] \approx \left(\frac{\int_{PB} S_x(\omega)|H_N(\omega)|^2 d\omega}{\int_{SB} S_x(\omega)|H_N(\omega)|^2 d\omega}\right)\left(\frac{\int_{SB} S_x(\omega) d\omega}{\int_{PB} S_x(\omega) d\omega}\right). \quad (46)$$

If we now invoke the Assumption 1 and Assumption 2 which were stated at the beginning of Section 3, our expression for $\mathrm{SNRI}[N]$ reduces to

$$\mathrm{SNRI}[N] \approx \frac{\Delta_{SB}}{\int_{SB}|H_N(\omega)|^2 d\omega}. \quad (47)$$

Thus, the SNR improvement factor is inversely proportional to the power in the stopband of the filter $H_N(\omega)$. We noted earlier in Section 2 that the frequency selective filters in $\mathcal{H}$ possess the property that as the filter order N increases, the average stopband attenuation also increases, and consequently the total stopband power decreases. Thus it is clear from (47) that the SNR improvement factor increases as the filter order N increases. The function $\mathrm{SNRI}[N]$ is plotted vs. the filter order N for the Parks-McLellan FIR replacement filter structure in Figure 3 and for the Butterworth IIR truncation filter structure in Figure 4. In each case the stopband is defined as $\omega \in [5\pi/8, \pi]$. These two filter structures are an essential component to low-power adaptive filter design [21]. As is clear from Figure 3 and Figure 4, the function $\mathrm{SNRI}[N]$ monotonically increases with N. We refer to plots of the function $\mathrm{SNRI}[N]$ as the performance profile for a given approximate filter structure $\mathcal{H}$.

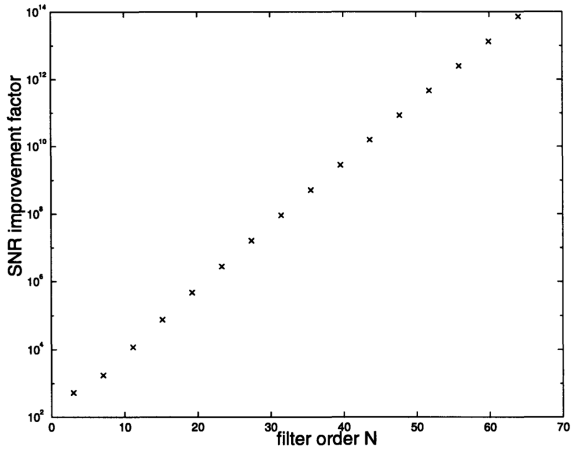

Figure 3. Performance profile for the Parks-McLellan FIR replacement filter structure. The stopband is defined as $\omega \in [5\pi/8, \pi]$.

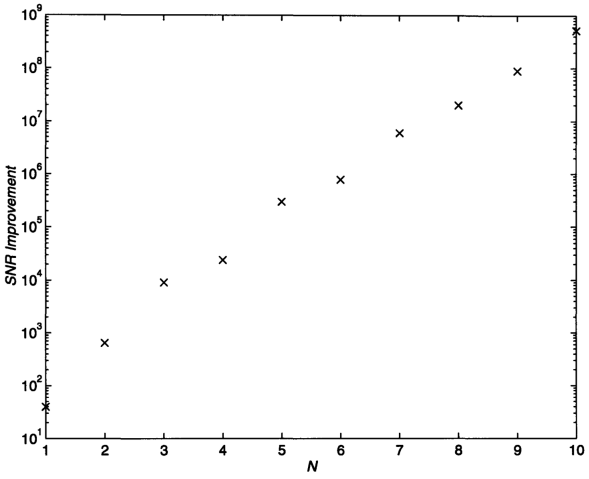

Figure 4. Performance profile for the Butterworth IIR truncation filter structure. The stopband is defined as $\omega \in [5\pi/8, \pi]$.

If the input SNR is relatively low we must select a relatively high filter order to obtain a sufficiently large SNR improvement factor to assure that the output SNR is maintained above the minimum tolerable level $\mathrm{OSNR}_{\mathrm{tol}}$. Conversely, when the input SNR is relatively high we will be able to select a relatively low filter order which will provide an SNR improvement factor that will assure $\mathrm{OSNR}[N] \geq \mathrm{OSNR}_{\mathrm{tol}}$.

To determine the LP solution to estimating N^*, we replace the exact ISNR in (44) with the LP estimate $\mathrm{ISNR}_{\mathrm{LP}}[N_0]$ given in (43), to obtain

$$\begin{aligned}\widehat{\mathrm{OSNR}}_{\mathrm{LP}}[N, N_0] &= \widehat{\mathrm{ISNR}}_{\mathrm{LP}}[N_0] \cdot \mathrm{SNRI}[N] \\ &= \left[\frac{\mathbf{y}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}}\right] \left(\frac{\displaystyle\int_{SB} \left[1 - |H_{N_0}(\omega)|^2\right] d\omega}{\displaystyle\int_{SB} |H_N(\omega)|^2 d\omega}\right) - \left(\frac{\displaystyle\int_{SB} |H_{N_0}(\omega)|^2 d\omega}{\displaystyle\int_{SB} |H_N(\omega)|^2 d\omega}\right),\end{aligned} \tag{48}$$

which we may compare to $\mathrm{OSNR}_{\mathrm{tol}}$ to determine the low-power estimate $\hat{N}^*_{\mathrm{LP}}$ for the optimal filter order N^* as the minimum filter order $N \in \mathcal{N}$ satisfying

$$\widehat{\mathrm{OSNR}}_{\mathrm{LP}}[N, N_0] \geq \mathrm{OSNR}_{\mathrm{tol}}. \tag{49}$$

The expression in (48) may be rearranged with simple algebraic manipulations and substitutions to produce

$$\left[\frac{\mathbf{y}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}}\right] \geq \left\{\mathrm{OSNR}_{\mathrm{tol}} + \left(\frac{\displaystyle\int_{SB} |H_{N_0}(\omega)|^2 d\omega}{\displaystyle\int_{SB} |H_N(\omega)|^2 d\omega}\right)\right\} \times \frac{\displaystyle\int_{SB} |H_N(\omega)|^2 d\omega}{\displaystyle\int_{SB} \left[1 - |H_{N_0}(\omega)|^2\right] d\omega}. \tag{50}$$

By defining the ratio of quadratic forms in the above expression as

$$R = \left[\frac{\mathbf{y}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}}\right], \tag{51}$$

and the function $R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}]$ as

$$R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}] = \left\{\mathrm{OSNR}_{\mathrm{tol}} + \left(\frac{\displaystyle\int_{SB} |H_{N_0}(\omega)|^2 d\omega}{\displaystyle\int_{SB} |H_N(\omega)|^2 d\omega}\right)\right\} \times \frac{\displaystyle\int_{SB} |H_N(\omega)|^2 d\omega}{\displaystyle\int_{SB} \left[1 - |H_{N_0}(\omega)|^2\right] d\omega}. \tag{52}$$

the decision rule for selecting $\hat{N}^*_{\mathrm{LP}}$ simplifies to choosing the *minimum* filter order $N \in \mathcal{N}$ satisfying

$$R \geq R_{\mathrm{tol}}[N; N_0, \mathrm{OSNR}_{\mathrm{tol}}]. \tag{53}$$

The notation $R_{\rm tol}[N; N_0, {\rm OSNR}_{\rm tol}]$ has been used to emphasize that $R_{\rm tol}[N; N_0, {\rm OSNR}_{\rm tol}]$ is a function of the filter order N and is parameterized by the nominal filter order N_0 and the minimum tolerable output SNR ${\rm OSNR}_{\rm tol}$. This enforces the fact that ${\rm OSNR}_{\rm tol}$ and N_0 are application-specific parameters which are to be fixed in advance, leaving $R_{\rm tol}[N; N_0, {\rm OSNR}_{\rm tol}]$ a monotonic function of the single variable N. Note that the only signal-dependent quantity in (53) is the ratio of quadratic forms R, which was defined in (51). As a final note, if we desire to avoid the power hungry division involved in computing R, we may use an alternative form for the decision rule for selecting $\hat{N}^*_{\rm LP}$. The resulting simplified decision rule for selecting $\hat{N}^*_{\rm LP}$ is to to choose the *minimum* filter order $N \in \mathcal{H}$ satisfying

$$\mathbf{y}^T\mathbf{y} \geq (\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}) R_{\rm tol}[N; N_0, {\rm OSNR}_{\rm tol}] \qquad (54)$$

The low-power decision rule is now summarized.

Summary of Method for Determining $\hat{N}^*_{\rm LP}$

1. Fix the values of the application-specific parameters N_0, L, and ${\rm OSNR}_{\rm tol}$
2. Compute R using (51) and the signal vectors $\mathbf{x}$ and $\mathbf{y}$ defined in (41) and (42), respectively
3. Determine $\hat{N}^*_{\rm LP}$ as the minimum value of N for which

$$R \geq R_{\rm tol}[N; N_0, {\rm OSNR}_{\rm tol}],$$

as described in (53) or (54)

In summary, the LP solution to the approximate filtering problem invokes three explicit assumptions and relies on the signal-dependent estimates $\hat{P}_x$ and $\hat{P}_y[N_0]$ given in (41) and (42), respectively. For situations in which the three assumptions are valid and in which $\hat{P}_x$ and $\hat{P}_y[N_0]$ are good estimates of P_x and $P_y[N_0]$, respectively, we expect excellent estimator performance using the LP estimate $\hat{N}^*_{\rm LP}$ for N^*. This issue is explored in the next section. The function ${\rm SNRI}[N]$ and the nature of its dependence on N and $\mathcal{H}$ have been studied extensively [21].

3.2. Convergence Analysis

It is of interest to determine the degree to which the low-power filter order estimate $\hat{N}^*_{\rm LP}$ converges to the theoretically optimal filter order N^* for input signals that satisfy the assumptions underlying the derivation of $\hat{N}^*_{\rm LP}$. In this section, we illustrate empirically that better convergence is obtained as the duration L over which $\hat{P}_x$ and $\hat{P}_y[N_0]$ are computed is made longer, and also as the nominal filter order N_0 is made larger. We also observe and discuss the fact that since optimal filter order selections partition the range of possible input SNR values, the relation of the actual input SNR to the boundaries in this partitioning is an important factor in determining whether or not the truly optimal filter order N^* is exactly determined by the LP estimation method.

For our convergence analysis, we assume that the input signal satisfies the same conditions that were stipulated in the derivation of our expression for $\hat{N}^*_{\rm LP}$ in Section 3. This means that we assume the input signal $x[n]$ is a WSS random process. When L consecutive samples of the output $y[n]$ are produced using a filter of order N_0, it follows that these output samples also belong to a WSS random process. We conclude that $\hat{P}_x$ and $\hat{P}_y[N_0]$ as defined in (41) and (42), respectively, represent estimates of the zero-lag auto-correlation values of $x[n]$ and $y[n]$, respectively. These well-known estimators converge to the true values of the zero-lag autocorrelations as L is made larger. Since $\hat{P}_x$ and $\hat{P}_y[N_0]$ are the only signal-related quantities used in obtaining the input SNR estimate in accordance with (43), we expect the input SNR estimate ${\rm ISNR}_{\rm LP}[N_0]$ to converge to the true input SNR as L and N_0 are made larger.

To verify the influence of the estimation interval L on the input SNR estimate, we applied the LP estimation method of (53) to a synthetically generated random signal $x[n]$. This signal was designed to have a flat spectrum in the passband $|\omega| \in [0, 3\pi/8]$, a flat spectrum in the stopband $|\omega| \in [5\pi/8, \pi]$, negligible energy in the transition band $|\omega| \in [3\pi/8, 5\pi/8]$, and a fixed SNR throughout its 10,000 point duration. The signal was filtered using an order-N_0 digital Butterworth filter. The L consecutive input and output samples (starting from the 1000th sample to avoid filter startup transient effects) were used to obtain the LP estimate ${\rm ISNR}_{\rm LP}[N_0]$ of the input SNR. For a case where the true SNR of the input signal $x[n]$ was set to 0.07, in Figure 5 we show the LP estimates of the input SNR obtained for different values of L in the range $1 \leq L \leq 4000$ and N_0 in the range $4 \leq N_0 \leq 10$. It is clear from Figure 5 that as L and N_0 increase, the LP estimate of the input SNR visually converges to the true input SNR of 0.07. It should be noted that lower values of N_0 correspond to frequency response shapes which violate the underlying assumptions to a greater degree. We must keep in mind that unless the filter $h_{N_0}[n]$ is ideal as in (6), the estimate ${\rm ISNR}_{\rm LP}[N_0]$ will never truly converge to the true input SNR, no matter how large L is made.

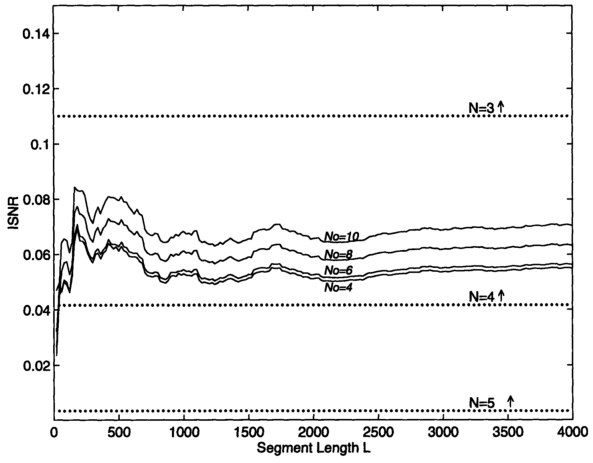

Figure 5. *Solid curves represent input SNR estimates as a function of L and N_0. The actual SNR for the input signal is 0.07. The straight dotted lines indicate the partitioning of the input SNR space by optimal values for the number of filter sections to use in order to obtain an output SNR of at least 1000.*

In Figure 5, we have also indicated the partitioning of the input SNR space in accordance with the corresponding optimal

filter order N* which should be used to ensure a minimum tolerable output SNR of 1000. Except for very small values of L, it is seen that the LP estimate of the input SNR leads to $\hat{N}^*_{\rm LP} = N^*$. This result is dependent on the fact that an input SNR of .07 happens to fall near the middle of the input SNR space corresponding to $N^* = 4$. For example, if the actual input SNR had been made equal to 0.04, the LP estimates of the input SNR may have crossed into the incorrect $N^* = 5$ region for a larger set of values of L.

To close this section, we offer two remarks regarding the convergence properties of the LP optimal filter order estimate $\hat{N}^*_{\rm LP}$. First, we note that

$$\lim_{N_0\to\infty}\lim_{L\to\infty}\widehat{\rm ISNR}_{\rm LP}[N_0] = {\rm ISNR}. \tag{55}$$

This statement simply elucidates the fact that as $N_0 \to \infty$, the filter $h_{N_0)[n]}$ becomes ideal with no transition band. In the limiting case of (55) the power difference $P_x - P_y[N_0]$ converges to the true value of P_x^{PB}, and consequently the ratio $R = \mathbf{y}^T\mathbf{y}/(\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y})$ converges to the true input SNR. While in practice we will never be able to realize the conditions of this limiting case, the result is nevertheless insightful. Secondly, we observe that under Assumption 1

$$\lim_{N\to\infty}{\rm SNRI}[N] = \frac{\Delta_{SB}}{\displaystyle\int_{SB}|H_N(\omega)|^2 d\omega}, \tag{56}$$

which is the same expression we get for SNRI[N] when we invoke the assumptions presented in the derivation of the LP estimation method. Remarkably, then, we may conclude that

$$\lim_{N\to\infty}\lim_{N_0\to\infty}\lim_{L\to\infty} = \hat{N}^*_{\rm LP}, \tag{57}$$

which implies that if we use sufficiently large values of L and N_0 to compute ${\rm ISNR}_{\rm LP}[N_0]$, then we can expect asymptotically optimal performance as the filter orders $N \in \mathcal{N}$ that we search over to compute $\hat{N}^*_{\rm LP}$ increase without bound. A second note we make in closing is that by introducing the $L \times L$ convolution matrix

$$\mathbf{H}[N_0] = \begin{bmatrix} h_{N_0}[0] & 0 & \cdots & 0 \\ h_{N_0}[1] & h_{N_0}[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_0}[L-1] & h_{N_0}[L-2] & \cdots & h_{N_0}[0] \end{bmatrix} \tag{58}$$

we may express the vector $\mathbf{y}$ as

$$\mathbf{y} = \mathbf{H}[N_0]\mathbf{x}, \tag{59}$$

and thus the expression for R in (51) simplifies to

$$R = \frac{\mathbf{x}^T\mathbf{A}[N_0]\mathbf{x}}{\mathbf{x}^T\mathbf{B}[N_0]\mathbf{x}} \tag{60}$$

where the $L \times L$ matrices $\mathbf{A}$ and $\mathbf{B}$ satisfy

$$\mathbf{A} = \mathbf{H}^T[N_0]\mathbf{H}[N_0] \tag{61}$$

$$\mathbf{B} = (\mathbf{I} - \mathbf{H}[N_0])^T(\mathbf{I} - \mathbf{H}[N_0]). \tag{62}$$

In this formulation R has the form of a ratio of two quadratic forms in the random vector $\mathbf{x}$. It is interesting to note that if the vector $\mathbf{x}$ is a multivariate Gaussian random vector, an extremely complicated nevertheless computable expression may be obtained for the variance of R as a function of the filter coefficients in $\mathbf{H}$ and the power window length L. Various forms of the variance of the random variable R may be found in the literature [24-27]. A future direction of this research is to analytically evaluate this variance and compare it to the sample variance computed in computer simulations. Furthermore, the problem of *designing* the filter $h_N[n]$ which appears in the matrix $\mathbf{H}[N]$ to produce filter structures $\mathcal{H}$ which *minimize the variance* of R provides an exciting and challenging future avenue to pursue in the area of approximate filtering algorithms.

In addition, expressions for the mean and variance of $\mathbf{x}^T\mathbf{A}[N_0]\mathbf{x}$ and $\mathbf{x}^T\mathbf{B}[N_0]\mathbf{x}$ for a multivariate Gaussian random vector $\mathbf{x}$ with zero mean and covariance matrix $\mathbf{\Sigma}$ may be obtained in closed form [28]. Specifically, they are given by

$$E(\mathbf{x}^T\mathbf{A}[N_0]\mathbf{x}) = {\rm trace}(\mathbf{\Sigma}\mathbf{A}[N_0]) \tag{63}$$

$${\rm VAR}(\mathbf{x}^T\mathbf{A}[N_0]\mathbf{x}) = 2\,{\rm trace}(\mathbf{\Sigma}\mathbf{A}[N_0])^2, \tag{64}$$

and

$$E(\mathbf{x}^T\mathbf{B}[N_0]\mathbf{x}) = {\rm trace}(\mathbf{\Sigma}\mathbf{B}[N_0]) \tag{65}$$

$${\rm VAR}(\mathbf{x}^T\mathbf{B}[N_0]\mathbf{x}) = 2\,{\rm trace}(\mathbf{\Sigma}\mathbf{B}[N_0])^2. \tag{66}$$

While these expressions do not give the true mean and variance of the random variable R or $\hat{N}^*_{\rm LP}$, they do offer insight into the statistical properties of two signal-dependent quantities $\mathbf{y}^T\mathbf{y} = \mathbf{x}^T\mathbf{A}[N_0]\mathbf{x}$ and $(\mathbf{y}^T\mathbf{y} - \mathbf{x}^T\mathbf{x}) = \mathbf{x}^T\mathbf{B}[N_0]\mathbf{x}$ involved in the simplified division-free decision rule for determining $\hat{N}^*_{\rm LP}$ given in (54), and thus are worth mentioning here.

3.3. Numerical Example

In this section we present a simple numerical example to demonstrate the efficacy of the LP solution to the approximate filtering problem. We first synthetically generated a random driving noise signal which consisted of independent and identically distributed samples distributed according to a unit variance, zero mean Gaussian probability density function (PDF). This driving noise sequence was then filtered with a 30th-order all-pole filter to create a WSS Gaussian random process. A plot of the PSD of this random process is shown in Figure 6.

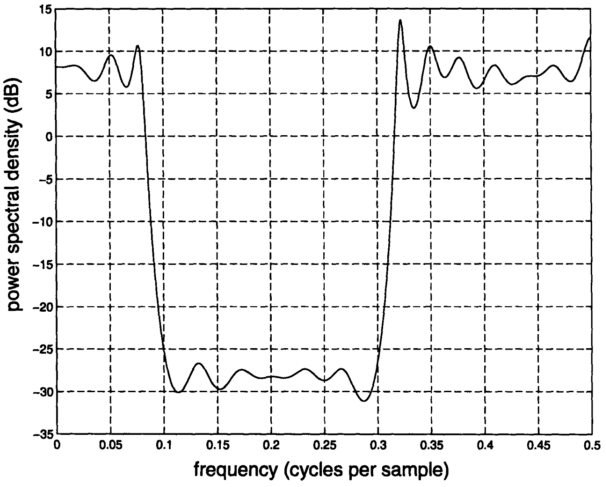

Figure 6. *Power spectral density of the 30th-order AR process which is used as the input signal in the numerical example.*

The all-pole filter parameters were selected to assure the power spectral density $S_x(\omega)$ was negligible in the transition band, defined in this example to be $3\pi/8 \leq |\omega| \leq 5\pi/8$.

The exact all-pole parameters that were used in this example are given in a more elaborate research study [21]. The passband was defined to be $0 \leq |\omega| \leq 3\pi/8$ and the stopband was defined as $5\pi/8 \leq |\omega| \leq \pi$. As can be seen from the spectral shape of $S_x(\omega)$, the input power spectral density is negligible in the transition band and relatively flat in the stopband in accordance with the assumptions underlying the development of the LP solution to the approximate filtering problem.

Table 1. *Summary of the results of the numerical example using the LP estimator in which the true value of $N^* = 10$ and the true input SNR is 0.4831. The results are tabulated for power window lengths of $L = 50, 100, 500, 1000,$ and 5000.*

L	$\hat{\mathrm{N}}^*_{\mathrm{LP}}$		$\widehat{\mathrm{ISNR}}_{\mathrm{LP}}$	
	Sample Mean	Sample STD	Sample Mean	Sample STD
50	7.5700	2.0013	0.6243	0.6540
100	7.7100	1.5718	0.5009	0.2320
500	7.2300	1.4829	0.5046	0.0987
1000	7.2000	1.4771	0.4998	0.0678
5000	7.6800	1.4967	0.4809	0.0260

In Table 1 we summarize the LP estimation results after 100 Monte Carlo trials were performed for each of the values of $L = 50, 100, 500, 1000,$ and 5000. The sample mean and sample standard deviation (STD) are listed in Table 1. Clearly as L increases, the quality of our LP estimates N^*_{LP} and $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}$ improve. This is evident from the fact that the sample standard deviations decrease as L increases. In addition, the low-power estimate $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}$ of the input SNR converges towards the true value of 0.4831 as L increases. In this example the Parks-McLellan FIR replacement approximate filter structure was used with $N_0 = N_{\max} = 64$ and $N_{\min} = 3$.

In Figure 7 we have plotted four histograms of the actual LP estimates of the input SNR for the same 100 Monte Carlo trials. Each histogram corresponds the estimates of the input SNR for different values of L. As L increases the estimates *tighten up* around their means. From the entries in Table 1 it is clear that while $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}$ converges towards the true input SNR as L increases, the estimate $\hat{N}^*_{\mathrm{LP}}$ does not coverage to N^* in this example. This is a consequence of the fact that theoretically $\widehat{\mathrm{ISNR}}_{\mathrm{LP}}$ converges to ISNR as L and N_0 increase without bound, while the convergence of $\hat{N}^*_{\mathrm{LP}}$ to N^* requires L, N_0, and $N_{\max}$ to increase without bound.

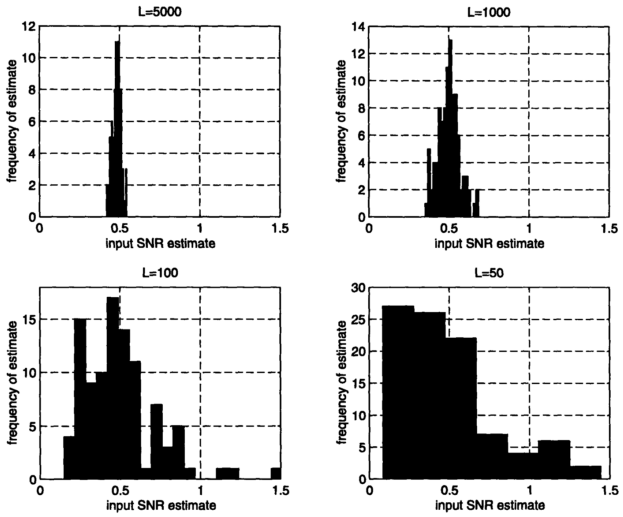

Figure 7. *Histograms of the LP input SNR estimates for L = 5000, 1000, 100, and 50. Each histogram represents the results of 100 Monte Carlo simulations.*

We shall revisit this same numerical example in Section 4.2 and evaluate the performance of the ML solution for comparison to the results given here.

4. Derivation of Maximum Likelihood Solution

In this section we assume $S_x(\omega)$ is the PSD of an autoregressive (AR) Gaussian random process, and we use a maximum likelihood (ML) objective to find estimates for the parameters defining this process. As we shall see the PSD $S_x(\omega)$ is a function of these parameters, and this function is one-to-one (invertible). Therefore we may easily obtain the ML estimate $[\hat{S}_x(\omega)]_{\mathrm{ML}}$ by invoking the well-known invariance property of the ML estimator [29].

By inspecting (1) we observe that the optimal filter order N^* is not in one-to-one correspondence with $S_x(\omega)$. This is true since many different functions $\hat{S}_x(\omega)$ could result in the same ratio of integrals which define $\mathrm{OSNR}[N]$, and thus many different functions $\hat{S}_x(\omega)$ could result in the same N^*. Nevertheless, we shall still be able to find a maximum *modified* likelihood estimate, which we shall denote $\hat{N}^*_{\mathrm{ML}}$, for

the optimal filter order N^* which is based on the maximum likelihood estimate $[\hat{S}_x(\omega)]_{\rm ML}$. We will show that the estimate $\hat{N}^*_{\rm ML}$, although not the estimate which truly maximizes the associated likelihood function, instead maximizes a modified likelihood function. Thus the performance of $\hat{N}^*_{\rm ML}$ will serve as a meaningful benchmark for comparison with the LP estimate $\hat{N}^*_{\rm LP}$ of the optimal filter order which was presented in Section 3.

4.1. Maximum Likelihood Estimation

In this section we first present an expression for the asymptotic ML estimates of the AR parameters of a Gaussian random process. Directly following the insightful presentation in the textbook by Kay [29], we consider the random process generated as the output $x[n]$ of a stable, causal all-pole filter

$$H(z) = \frac{1}{A(z)} \tag{67}$$

excited at the input by a zero-mean white Gaussian noise sequence $u[n]$. The pth-order polynomial function $A(z)$ is defined by the AR filter parameters $[\, a_1\, a_2\, \cdots\, a_p\,]$ as

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}. \tag{68}$$

If the all-pole filter $H(z)$ is stable, the excitation noise sequence $u[n]$ assures that the output $x[n]$ is a WSS random process. The effect of the filter is to color the white noise sequence $u[n]$. The AR model is capable of producing a wide variety of PSD functions, depending on the choice of the AR filter parameters $[\, a_1\, a_2\, \cdots\, a_p\,]$ and excitation noise variance σ_u^2. The problem is to estimate the parameters $[\, a_1\, a_2\, \cdots\, a_p\,]$ and σ_u^2 based on the observed data sequence $[\, x[0]\, \cdots\, x[L-1]\,]$. Once the parameter estimates $[\, \hat{a}_1\, \hat{a}_2\, \cdots\, \hat{a}_p\,]$ and $\hat{\sigma}_u^2$ are computed, the PSD is estimated as

$$\hat{S}_x(\omega) = \hat{\sigma}_u^2 \left| 1 + \sum_{m=1}^{p} \hat{a}_m e^{-j\omega m} \right|^{-2} \tag{69}$$

We now given expressions for the asymptotic ML estimates for the parameters $[\, a_1\, a_2\, \cdots\, a_p\,]$ and σ_u^2. First, the estimated autocorrelation function is

$$\hat{r}_{xx}[k] = \begin{cases} \frac{1}{L} \displaystyle\sum_{n=0}^{L-1-|k|} x[n]x[n+|k| & |k| \le L-1 \\ 0 & |k| \ge L \end{cases}. \tag{70}$$

The set of equations to be solved for the asymptotic ML estimate of the AR filter parameters $\mathbf{a}$ is

$$\sum_{i=1} p\hat{r}_{xx}[k-l] = -\hat{r}_{xx}[k] \qquad k = 1, 2, \cdots, p \tag{71}$$

which can be rewritten in matrix form as

$$\begin{bmatrix} \hat{r}_{xx}[0] & \hat{r}_{xx}[1] & \cdots & \hat{r}_{xx}[p-1] \\ \hat{r}_{xx}[1] & \hat{r}_{xx}[0] & \cdots & \hat{r}_{xx}[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{xx}[p-1] & \hat{r}_{xx}[p-2] & \cdots & \hat{r}_{xx}[0] \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix}$$

$$= - \begin{bmatrix} \hat{r}_{xx}[1] \\ \hat{r}_{xx}[2] \\ \vdots \\ \hat{r}_{xx}[p] \end{bmatrix}. \tag{72}$$

These are the well-known *Yule-Walker equations*, which may be recursively solved using the Levinson recursion algorithm [30]. What is left is to solve for the asymptotic ML estimate σ_u^2. The asymptotic ML estimate is given by

$$\hat{\sigma}_u^2 = \hat{r}_{xx}[0] + \sum_{k=1}^{p} \hat{a}_k \hat{r}_{xx}[k]. \tag{73}$$

Thus, the asymptotic ML estimates $\hat{\mathbf{a}}$ and $\hat{\sigma}_u^2$ for the parameters $\mathbf{a}$ and σ_u^2 are given in (72) and (73), respectively. These estimates converge to the true ML estimates as $L \to \infty$, and yield reasonable estimates for sufficiently large finite values L. We recall that our ML estimate of the PSD is

$$[\hat{S}_x(\omega)]_{\rm ML} = \hat{\sigma}_u^2 \left| 1 + \sum_{m=1}^{p} \hat{a}_m e^{j\omega m} \right|^{-2}. \tag{74}$$

This estimate may be used to determine an ML estimate $\hat{N}^*_{\rm ML}$ of the optimal filter order N^*. This ML-based estimate is produced by choosing $\hat{N}^*_{\rm ML}$ to be the *minimum order* $N \in \mathcal{N}$ of the frequency-selective filter $h_N[n] \in \mathcal{H}$ which provides sufficient stopband attenuation to assure

$$\widehat{\rm OSNR}_{\rm ML} \ge {\rm OSNR}_{\rm tol}, \tag{75}$$

where the ML-based estimate of the output SNR is defined as

$$\widehat{\rm OSNR}_{\rm ML} \triangleq \frac{[\hat{P}_y^{PB}]_{\rm ML}}{[\hat{P}_y^{SB}]_{\rm ML}}, \tag{76}$$

the ML-based estimate of the output power in the passband is defined as

$$[\hat{P}_y^{PB}[N]]_{\rm ML} = \frac{1}{2\pi} \int_{PB} [\hat{S}_y(\omega)]_{\rm ML} d\omega$$

$$= \frac{1}{2\pi} \int_{PB} [\hat{S}_x(\omega)]_{\rm ML} |H_N(\omega)|^2 d\omega, \tag{77}$$

and the ML-based estimate of the output power in the stopband is defined as

$$[\hat{P}_y^{SB}[N]]_{\mathrm{ML}} = \frac{1}{2\pi}\int_{SB}[\hat{S}_y(\omega)]_{\mathrm{ML}}d\omega$$
$$= \frac{1}{2\pi}\int_{SB}[\hat{S}_x(\omega)]_{\mathrm{ML}}|H_N(\omega)|^2 d\omega. \quad (78)$$

Consequently, the estimate $\hat{N}^*_{\mathrm{ML}}$ maximizes the modified likelihood function which is related to the true likelihood function as explained in the textbook by Kay [29].

Summary of Method for Determining $\hat{N}^*_{\mathrm{ML}}$

1. Given observations $[\,x[0]\;\cdots\;x[L-1]\,]$, compute the ML estimates of $\hat{\mathbf{a}}$ and $\hat{\sigma}_u^2$ using (73) and (73), respectively
2. Compute $[\hat{S}_x(\omega)]_{\mathrm{ML}}$ via (74)
3. Determine $\hat{N}^*_{\mathrm{ML}}$ according to (75)

4.2. Numerical Example

In this section we present the results of using the ML estimation method for determining N^* on the same numerical example that was presented in Section 3.3. Numerical results are given to demonstrate the performance of the ML method. Recall from Section 3.3 that this example involves a random driving noise signal consisting of independent and identically distributed samples distributed according to a unit variance, zero mean Gaussian PDF. This driving noise signal was filtered with an 30th-order all-pole filter to create a WSS Gaussian random process with a PSD which was shown previously in Figure 6. In Table 2 we summarize the ML estimation results after 100 Monte Carlo trials were performed for each of the values of $L = 50, 100, 500, 1000$, and 5000. The sample mean and sample standard deviation (STD) are listed Table 2. Clearly as L increases, the quality of the ML optimal filter order estimate $\hat{N}^*_{\mathrm{ML}}$ improves since its standard deviation decreases as L increases. In addition, the ML estimate of the optimal filter order converges towards the true value of $N^* = 10$ as L becomes larger.

In the simulations we used the Yule-Walker equations and the Levinson recursion to solve for the ML AR coefficients which give the ML power spectrum estimate and thus the ML estimate of the optimal filter order $\hat{N}^*_{\mathrm{ML}}$. While the Yule-Walker equations give the asymptotic ML estimates of the AR parameters which converge to the true ML estimates as $L \to \infty$, it is well known that another method produces better estimates for finite data records [31]. This method of AR parameter estimation is known as the forward-backward least-squares method. Using the forward-backward least-squares method to compute the AR parameter estimates would probably improve the performance of the estimator $\hat{N}^*_{\mathrm{ML}}$. In our simulations here we used the Yule-Walker AR parameter estimates since we were guided by the mathematical optimality of the ML approach. Small errors in the AR parameter estimates should not have a significant effect on the estimator $\hat{N}^*_{\mathrm{ML}}$. This is true since $\hat{N}^*_{\mathrm{ML}}$ is based on a ratio of integrals of the power spectrum. Small errors in the AR parameter estimates will be integrated out when computing $\hat{N}^*_{\mathrm{ML}}$.

As a final note, we observe that in determining the ML estimates for the numerical example we assumed the order of the AR process was known to be equal to 30. This introduces an element of unfairness when comparing the performance of the ML and LP approaches. In practice the AR order would not be known exactly, and would thus have to be estimated. Furthermore, this example was specifically tailored for the ML approach, since the input signal was synthetically generated to represent a true AR random process. This provides a second reason why we would expect the ML estimates to outperform the LP estimates.

Table 2. *Summary of the results of the numerical example using the ML estimator in which the true value of $N^* = 10$ and the true input SNR is 0.4831. The results are tabulated for power window lengths of $L = 50, 100, 500, 1000$, and 5000.*

L	$\hat{\mathbf{N}}^*_{\mathbf{ML}}$	
	Sample Mean	**Sample STD**
50	11.0300	1.5983
100	11.2000	1.5176
500	10.6000	1.2060
1000	10.7200	1.2877
5000	10.2400	0.8180

5. Experimental Design and Results of the Low-power Adaptive IIR Digital Filter

In this section we present computer simulations which show that significant power savings may be achieved when the order of an IIR digital filter is dynamically varied to provide time-varying stopband attenuation in proportion to the time-varying SNR of the input signal, while maintaining a fixed level of output quality. We highlight experiments involving speech signals to demonstrate the practical viability of the low-power adaptive IIR digital filter presented in this paper. As we shall see, significant reduction in power consumption over a fixed-order IIR digital filter is achieved in simulations involving the demultiplexing frequency-division multiplexed (FDM) speech signals.

We illustrate the potential of the low-power adaptive IIR digital filter to reduce power consumption in speech processing. We use a Butterworth truncation filter structure with 10 second-order sections. This approximate filter structure and the adaptation control strategy described in Section 3 was applied to two speech signals which had been frequency-division multiplexed. The power window length was chosen to be $L = 100$ and the minimum tolerable output SNR was set to 1,000. The IIR filters in the Butterworth truncation filter structure each had a half-power frequency of $\pi/2$. The stopband was defined to be between $5\pi/8$ and π, while the passband was defined to be between 0 and $3\pi/8$. One speech signal was spectrally centered in

the passband region of the lowpass filter and the other was modulated into the stopband region of the lowpass filter. The sampling rate for each of the speech signal was 16,000 Hz. Figure 8 shows the speech signal in the passband, the speech signal in the stopband, and the evolution of the number of filter sections used by the approximate filtering technique. Since we are using cascades of second order subsections, the power consumption is directly proportional to the number of active subsections, which can be enabled or disabled with low overhead. Since we are calculating the model order over a window length $L = 100$, less than 1% overhead is needed for these calculations. Examination of the figure shows that as would be expected, the number of filter sections is large when the input SNR is small. Furthermore, the number of filter sections is small when the input SNR is high. If we compare the power consumption for in the low-power adaptive filter to the power consumption of a fixed IIR filter order need to handle the worst-case signal statistics, we see that a power reduction of approximately 63% is achieved in this simulation. This can be best understand with a visual inspection of Figure 8, where we can see the time-varying number of filter sections as a function of sample number ranges between 2 and 10, while the number of sections in a fixed-order IIR filter that would ensure we are able to maintain the minimum tolerable output SNR of 1,000 at all times is 8.

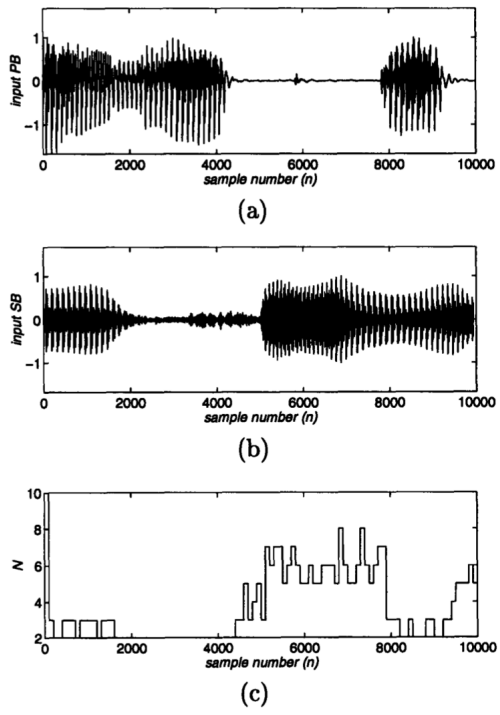

Figure 8. *Demultiplexing of FDM speech using low power frequency selective filtering. (a) passband speech, (b) stopband speech, and (c) number of filter sections as a function of sample number.*

For comparison of the of the power savings achieved by the adaptive digital filter presented in this paper to other results recently reported in the literature, we present the findings of three other experiments. First, a recent state-of-the-art digital FIR filter using data-driven clock gating and multibit flip-flops combined achieved 22% to 25% power reduction compared to that using a conventional design [3]. Secondly, in another very recent study, a novel design for an energy-efficient IIR digital filter achieved nearly 63% reduction in energy with a negligible deviation of the frequency response from the standard implementation [1]. Finally, a low power reconfigurable FIR digital filter based on dual mode operation achieved power savings up to 37.97% in simulations using speech signal processing, similar to the simulations in this section [32]. These results demonstrate how low-power digital signal processing continues to be an area of focused interest and innovation.

The low-power (LP) approach to finding N^* is computationally simple and viable for direct implementation in hardware for practical applications. A research group at Stanford University has designed and implemented an approximate filtering algorithm for the application of low-power interpolation and decimation. Excellent results have been obtained with a conventional interpolation/decimation system that was implemented and shown to consume an average of approximately 86.4 mW with a 5V power supply. Using an approximate filtering algorithm compiled onto a programmable processor, the power consumption was reduced by 36% for the decimation system and 17% for the interpolation system [22].

6. Conclusion

We have considered the practical problem of dynamically reducing the order of a frequency-selective digital filter to reduce average power consumption, and presented the class of approximate filtering algorithms for which this is accomplished. Approximate filtering algorithms were developed by abstracting a theory from the practical low-power filtering problem. The theory centered on the problem of determining an optimal filter order based on observations of the input data and a set of concrete assumptions on the statistics of the input signal. We explored the statistical properties of this theory, and showed that under certain assumptions the class of approximate filtering algorithms is asymptotically optimal. The theory served the purpose of aiding us in understanding interpreting the performance of approximate filtering algorithms.

The low-power approach to adaptively finding the optimal filter order is computationally simple and viable for direct implementation in hardware for practical applications in which power savings in the range of 17% to 35% have been achieved. The theoretical and numerical simulation results demonstrate the efficacy and versatility of asymptotically optimal low-power digital filtering using adaptive approximate processing, and contribute to the evolution of low-power digital signal processing which continues to be an area of focused interest and innovation.

References

[1] Pilipović, R.; Risojević, V.; Bulić, P. On the Design of an Energy Efficient Digital IIR A-Weighting Filter using Approximate Multiplication. *Sensors* 2021, *21*, 732. https://doi.org/10.3390/s21030732

[2] M. Kang, S. K. Gonugondla and N. R. Shanbhag, Deep In-Memory Architectures in SRAM: An Analog Approach to Approximate Computing, in Proceedings of the IEEE, vol. 108, no. 12, pp. 2251-2275, Dec. 2020, https://doi.org/10.1109/JPROC.2020.3034117

[3] Agathoklis, P.; Touil, L.; Hamdi, A.; Gassoumi, I.; Mtibaa, A. Design of Low-Power Structural FIR Filter Using Data-Driven Clock Gating and Multibit Flip-Flops. *Journal of Electrical and Computer Engineering* 2020 https://doi.org/10.1155/2020/8108591

[4] Kim, M. S.; Garcia, A. A. D. B.; Oliveira, L. T.; Hermida, R.; Bagherzadeh, N. Efficient Mitchell's Approximate Log Multipliers for Convolutional Neural Networks. *IEEE Trans. Comput.* 2018, *68*, 660-675.

[5] Liu, W.; Xu, J.; Wang, D.; Wang, C.; Montuschi, P.; Lombardi, F. Design and Evaluation of Approximate Logarithmic Multipliers for Low Power Error-Tolerant Applications. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2018, *65*, 2856-2868.

[6] Pilipović, R.; Bulić, P. On the Design of Logarithmic Multiplier Using Radix-4 Booth Encoding. *IEEE Access* 2020, *8*, 64578-64590.

[7] Leon, V.; Zervakis, G.; Soudris, D.; Pekmestzi, K. Approximate Hybrid High Radix Encoding for Energy-Efficient Inexact Multipliers. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 2018, *26*, 421-430.

[8] Liu, W.; Cao, T.; Yin, P.; Zhu, Y.; Wang, C.; Swartzlander, E.E.; Lombardi, F. Design and Analysis of Approximate Redundant Binary Multipliers. *IEEE Trans. Comput.* 2019, *68*, 804-819.

[9] Agrawal, A.; Choi, J.; Gopalakrishnan, K.; Gupta, S.; Nair, R.; Oh, J.; Prener, D.A.; Shukla, S.; Srinivasan, V.; Sura, Z. Approximate computing: Challenges and opportunities. In *Proceedings of the 2016 IEEE International Conference on Rebooting Computing (ICRC)*, San Diego, CA, USA, 17-19 October 2016; pp. 1-8.

[10] Mittal, S. A survey of techniques for approximate computing. *ACM Comput. Surv. (CSUR)* 2016, *48, 62*.

[11] Jerger, N. E.; Miguel, J. S. Approximate Computing. *IEEE Micro* 2018, *38*, 8-10.

[12] Eeckhout, L. Approximate Computing, Intelligent Computing. *IEEE Micro* 2018, *38*, 6-7.

[13] Price M., J. Glass, A. P. Chandrakasan, A Low-Power Speech Recognizer and Voice Activity Detector Using Deep Neural Networks, IEEE Journal of Solid-State Circuits, vol. 53, no. 1, pp. 66-75, Jan. 2018.

[14] Calhoun, B. H., D. C. Daly, N. Verma, D. Finchelstein, D. D. Wentzloff, A. Wang, S.-H. Cho, and A. P. Chandrakasan, Design Considerations for Ultra-low Energy Wireless Microsensor Nodes, IEEE Transactions on Computers. pp. 727-749, June 2005.

[15] Tikekar M., V. Sze, A. P. Chandrakasan, A Fully Integrated Energy-Efficient H.265/HEVC Decoder With eDRAM for Wearable Devices, IEEE Journal of Solid-State Circuits, vol. 53, no. 8, pp.2368-2377, Aug. 2018.

[16] Steiger C., A. Abramson, P. Nadeau, A. P. Chandrakasan, R. Langer, G. Traverso, Ingestible electronics for diagnostics and therapy, Nature Reviews Materials, vol. 4, no. 2, pp. 83-98, Dec. 2018.

[17] Mimee M., P. Nadeau, A. Hayward, S. Carim, S. Flanagan, L. Jerger, J. Collins, S. McDonnell, R. Swartwout, R. J. Citorik, V. Bulovic, R. Langer, G. Traverso, A. P. Chandrakasan, T. K. Lu, An ingestible bacterial-electronic system to monitor gastrointestinal health, Science, vol. 360, no. 6391, pp. 915-918, May 2018.

[18] Banerjee U., A. Wright, C. Juvekar, M. Waller, Arvind, A. P. Chandrakasan, An Energy-Efficient Reconfigurable DTLS Cryptographic Engine for Securing Internet-of-Things Applications, IEEE Journal of Solid-State Circuits, vol. 54, no. 8, pp. 2339-2352, Aug. 2019.

[19] J. T. Ludwig, S. H. Nawab, and A. P. Chandrakasan. Convergence results on adaptive approximate filtering. In Advanced Signal Processing Algorithms (F. T. Luk, ed.), Proceedings of SPIE, Denver, CO, August 1996.

[20] J. T. Ludwig, S. H. Nawab, and A. P. Chandrakasan. Low-power digital filtering using approximate processing. IEEE Journal on Solid State Circuits, 31 (3): 395-400, March 1996.

[21] J. T. Ludwig. Low-power Digital Filtering Using Adaptive Approximate Processing. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT RLE. September 2, 1997.

[22] C. J. Pan. A Low-power digital filter for decimation and interpolation using approximate processing. International Solid State Circuits Conference, pages 102-103, February, 1997.

[23] E. Lee and D. G. Messerschmitt. Digital Communication. Kluwer Academic Publishers, Boston, MA, 1994.

[24] G. A. Ghazal. Moments of the ratio two dependent quadratic forms. Statistics and Probability Letters, 20 (4): 313-315, 1994.

[25] Lieberman. Saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables. Journal of the American Statistical Association, 89 (427): 924-928, 1994.

[26] M. D. Smith. Expectations of ratios of quadratic forms. Journal of Multivariate Analysis, 31: 244-257, 1989.

[27] M. D. Smith. Comparing approximations to the expectation of a ratio of quadratic forms in normal variables. Econometric Reviews, 15 (1): 81-95, 1996.

[28] L. L. Scharf. Statistical Signal Processing: Detection, Estimation, and Time Series Analysis. Addison-Wesley Publishing Company, Reading, MA, 1991.

[29] S. M. Kay. Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall, Upper Saddle River, NJ, 1993.

[30] S. M. Kay. Modern Spectrum Estimation: Theory and Application. Prentice Hall, Englewood Cliffs, NJ, 1988.

[31] Boaz Porat. Digital Processing of Random Signals: Theory and Methods. Prentice Hall, Englewood Cliffs, NJ, 1994.

[32] Padmapriya, S.; Lakshmi Prabha, V. Design of an efficient dual mode reconfigurable FIR filter architecture in speech signal processing, *Microprocessors and Microsystems*, 2015 *39, 7*, 521-528.