KTH Electrical Engineering

Parameter Estimation and Waveform Fitting for Narrowband Signals

TOMAS ANDERSSON

TRITA–S3–SB-0540
ISSN 1103-8039
ISRN KTH/SB/R - - 05/40 - - SE
ISBN 91-7178-078-5

Doctoral Thesis in Signal Processing
Stockholm, Sweden 2005

Signal Processing
KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden
Tel. +46 8 790 6000, Fax+46 8 790 7260
http://www.s3.kth.se

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i signalbehandling tisdagen den 7 juni 2005 kl 13.00 i sal E3, Kungl Tekniska högskolan, Osquars backe 14, Stockholm.

Tryck: Universitetsservice US AB

Abstract

Frequency estimation has been studied for a large number of years. One reason for this is that the problem is easy to understand, but difficult to solve. Another reason, for sure, is the large number of applications that involve frequency estimation, e.g radar using frequency modulated continuous wave (FMCW) techniques where the distance to the target is embedded in the frequency, resonance sensor systems where the output signal is given as the frequency displacement from a nominal frequency, radio frequency identification systems (RFID) where frequency modulation is used in the communication link, etc. The requirement on the frequency estimator varies with the application and typical issues include: accuracy, precision or (bias) processing speed or complexity, and ability to handle multiple signals. A lot of solutions to different problems in this area has been proposed, but still several open questions remain.

The first part of this thesis addresses the problem of frequency estimation using low complexity algorithms. One way of achieving such an algorithm is to employ a coarse quantization on the input signal. In this thesis, a 1-bit quantizer is considered which enables the use of low complexity algorithms. Frequency estimation using look-up tables is studied and the properties of such an estimator are presented. By analyzing the look-up tables using the Hadamard transform a novel type of low-complexity frequency estimators is proposed. They use operations such as binary multiplication and addition of precalculated constants. This fact makes them suitable in applications where low complexity and high speed are major issues. A hardware demonstrator using the table look-up technique is designed and a prototype is analysed by real measurements.

Today, the interest of using digital signal processing instead of analog processing is almost absolute. For example, in testing analog-to-digital converters an important part is to fit a sinewave to the recorded data, as well as to calculate the parameters that in least-squares sense result

in the best fit. In this thesis, the sinewave fitting method included in the IEEE Standard 1057 is studied in some detail. Asymptotic Cramér-Rao bounds for three- and four model parameters are derived under the Gaussian assumption. Further, the sinewave fitting properties of the algorithm are analyzed by the parsimony principle. A novel model order selection criterion is proposed for waveform fitting methods in the case of a linear signal model. A generalization of this criterion is made to include the non-linear sinewave fitting application.

For multiple sinewave fitting applications two iterative algorithms are proposed. The first method is a combination of the standardized sinewave fit algorithm and the expectation maximization algorithm. The second algorithm is an extension of a single sinewave model to a multiple sinewave model employing the standardized sinewave fitting algorithm. Both algorithms are analysed by numerical means and are shown to accurately resolve multiple sinewaves and produce efficient estimates. Initialization issues of such algorithms are included to some extent.

Acknowledgements

My life as a Ph.D. student has certainly been a pleasant one. I want to take the opportunity to extend my greatest and heartfelt thanks to all of you who have helped and supported me in various ways.

Peter Händel, who has supervised my work, for his patience with my sometimes erratic progress, for guidance in the art of scientific writing and for being a living book and a never-ceasing source of references in the field of frequency estimation. It has sometimes been hard times but mostly exciting and very fun. You have among many things taught, or at least introduced, me to the art of time-planning.

Mikael Skoglund, who has acted as co-supervisor, for your support and also for your ideas that led to joint publications.

Karin Demin, for your guidance in the world of bureaucracy, for serving an uncountable number of cakes and for being the fundamental of order and good spirit within the Signal Processing group. I will never forget you.

Mats Bengtsson, for guidance in the world of LaTeX and its many peculiar tools that greatly improved the readability of this thesis. Also thanks for all car trips to Martin Olsson in our mission to make the group somewhat sweeter and bigger.

Nina Unkuri and Andreas Stenhall for putting a real meaning behind the words Computer Support. You never cease to impress with your ability to meet the demands of your users spread requisites. Without your support my computer would had been like a worn pen with broken nib.

George Jöngren, my old office neighbour, for teaching me how to talk through walls. Also, your passion for science and attitude when attacking new problems have been most inspiring.

Henrik Lundin, my current office neighbour, for collaboration leading to joint publication, your company on our conference journeys which

led to some unforgettable experiences. “Keep up that good old Swedish tradition!”

All the people in the Signal Processing and Communication Theory groups. I am proud to be a part of you. The positive and creative environment that you all support will be something that I not soon forget.

Also, many thanks to all my musician friends who very successfully managed to learn me how to make the most out of my spare time. Especially you members of SMusK, Terrassorkestern and Opus Big Band.

Lastly, I want to express my greatest gratitude to my family, mum, dad and Cecilia in particular, whose understanding, encourage and patience have made this work possible.

Tomas Andersson
Stockholm, May 2005

Contents

Chapter 1

Introduction

1.1 Background

In many applications it is a frequency contents of a signal that carry the information about some sought property. Examples include a frequency modulated communication system, a frequency modulated continuous wave (FMCW) radar system, resonance sensor systems, etc. An even longer list can be made with the frequency estimators that have been proposed during the last centuries. One of the pioneers in the area was Baron Gaspard Riche de Prony. He discovered that evenly spaced samples of a signal consisting of a sum of complex-valued exponentials obey the homogeneous difference equation [dP95]

$$x[n+p] + \alpha_{p-1}x[n+p-1] + \cdots + \alpha_0 x_0[n] = 0. \tag{1.1}$$

In todays leading edge signal processing algorithms one often uses a recursive formula when generating a sinewave. The recursive formula used is nothing but a special case of de Prony's general formula (1.1). Consider the problem of creating samples of a sinusoid at an angular frequency ω. Then given two initial values $y[0]$ and $y[1]$, the consecutive samples can be calculated using

$$y[n+2] = \alpha y[n+1] - y[n], \qquad n = 0, 1, 2, \ldots \tag{1.2}$$

where $\alpha = 2\cos(\omega)$ and the initial values are given by $y[0] = 0$ and $y[1] = \sin(\omega)$, respectively. Using (1.2), it is possible to generate N samples of the signal $y[n] = \sin(\omega n)$ using $N-1$ multiplications, $N-2$ additions and two sin-function calls. Performing the generation of a sinewave,

it is the sin function calls that demand the major computer resources. Even for a small number of samples, it is a substantial difference in numerical complexity using the recursive formula (1.2) compared with direct sin-function calls. Of course, there is a drawback using the recursive formula. Since the present sample depends on all the old samples, errors are accumulated. The error sources are in the addition/multiplication operations due to limited number precision. However, modern digital signal processors (DSP) often support floating point precision which keeps the error on an acceptable level.

The recursive sinewave formula (1.2) is a good example of a simplified suboptimal method. Here, suboptimal in the context of the trade-off between accuracy and numerical computations. Suboptimal algorithms are becoming more and more commonly used in todays applications. As stated previously, the reason is that the computational capacity is limited; either by price or by a limited power source in applications relying on a battery.

1.2 Model

As in (1.1) and (1.2), the thesis will address evenly spaced sampled signals. The signal part $s[n]$ is, in its general form, a sum of real valued sinusoids[1]

$$s[n] = \sum_{\ell=1}^{p} \alpha_\ell \sin(\omega_\ell n + \phi_\ell) + C_\ell. \tag{1.3}$$

Here, α_ℓ is a real valued constant describing the amplitude, $\omega_\ell = 2\pi f_\ell$ is the angular frequency and ϕ_ℓ describes the initial phase. The C_ℓ handles measurement data with non-zero mean value. A DC-level in data is often present in applications using analog to digital conversion since ADCs often have a unipolar signal input range. In some cases is it convenient to write the phase shifted model (1.3) as

$$s[n] = \sum_{\ell=1}^{p} A_\ell \cos(\omega_\ell n) + B_\ell \sin(\omega_\ell n) + C_\ell. \tag{1.4}$$

The mapping between the parameters A_ℓ, B_ℓ and α_ℓ, ϕ_ℓ is one to one according to

$$A_\ell = \alpha_\ell \sin\phi_\ell \qquad B_\ell = \alpha_\ell \cos\phi_\ell \tag{1.5}$$

[1] Note that C_ℓ is only introduced for notational simplicity. The actual DC-level is given by $C = \sum C_\ell$.

In this thesis only real valued signal models are considered. A motivation thereof is that the corresponding complex-valued signal model is already extensively investigated in the literature. Further, some of the topics in this thesis discuss signal processing on quantized signals collected from a measurement system, in such systems there is no advantage in having a complex-valued signal model.

It is further assumed that the measured signal $x[n]$ contains the signal and some additional measurement noise $w[n]$, that is

$$x[n] = s[n] + w[n], \qquad n = 1, 2, \ldots, N \tag{1.6}$$

It is difficult to make valid assumptions about the measurement noise. The noise term often acts as the parameter that describes everything not included in the signal model, that is thermal and quantization noise, model imperfections, etc. Though, when algorithms are to be evaluated a Gaussian assumption is often made to describe the noise. In this thesis, when needed the noise is modeled as white and Gaussian. The samples in a white Gaussian process are independent of each other, and each sample is fully described by its mean value and variance.

Often, it is convenient to use a vector representation to describe (1.6). A straightforward way of arriving at such a model is to stack the samples in column vectors, that is

$$\mathbf{s} = \sum_{\ell=1}^{p} \mathbf{H}_\ell \theta_\ell \tag{1.7}$$

with

$$\mathbf{H}_\ell = \begin{bmatrix} \cos\omega_\ell & \sin\omega_\ell & 1 \\ \cos 2\omega_\ell & \sin 2\omega_\ell & 1 \\ \vdots & \vdots & \vdots \\ \cos N\omega_\ell & \sin N\omega_\ell & 1 \end{bmatrix} \qquad \theta_\ell = \begin{bmatrix} A_\ell \\ B_\ell \\ C_\ell \end{bmatrix} \tag{1.8}$$

The measured samples $\mathbf{x} = [x[1] \ldots x[N]]^T$ can then be written as

$$\mathbf{x} = \mathbf{s} + \mathbf{w} \tag{1.9}$$

where $\mathbf{w} = [w[1] \ldots w[N]]^T$.

1.3 A Frequency Estimation Example

A classical paper on frequency estimation using discrete time data is the one by Rife and Borstyn [RB74]. In [RB74], the model (1.6) with $s[n]$ given by (1.3) in the special case $p = 1$ and $C = 0$. Here, the example of [RB74] is reviewed, for an arbitrary C.

Consider the single sinusoidal signal $s[n]$ with unknown parameters A, B, C and ω, that is (1.4) with $p = 1$. The measured signal is then given by $s[n]$ deteriorated by the additive noise component $w[n]$. The noise is assumed to be zero mean white Gaussian with variance σ^2. A typical realization of such a signal is displayed in Figure 1.1. The estimation problem in this case is to estimate the signal parameters using the measured data samples $\mathbf{x} = [x[1] \dots x[N]]^T$.

The probability density function (pdf)

$$p(\mathbf{x}; \theta, \omega) = \frac{1}{(2\pi\sigma^2)^N} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{H}\theta)^T(\mathbf{x} - \mathbf{H}\theta)\right] \tag{1.10}$$

describes the probability per infinitesimal volume of receiving the data samples $\mathbf{x}$ given a set of parameters $\{\theta, \omega\}$. Here the parameter vector θ equals $\theta = (A, B, C)^T$. In (1.10), $\mathbf{H}$ is implicitly dependent on the frequency ω. The maximum likelihood estimator (MLE) strives to maximize the pdf with respect to the unknown parameters for a given $\mathbf{x}$ and use those parameters as an estimate. That is,

$$[\hat{\theta}, \hat{\omega}] = \arg\max_{\theta,\omega} p(\mathbf{x}; \theta, \omega). \tag{1.11}$$

The expression (1.11) can be further simplified by taking the logarithm of (1.10), multiplying with -1 and removing the constant terms not dependent on θ or ω, that is

$$[\hat{\theta}, \hat{\omega}] = \arg\min_{\theta,\omega} \left[(\mathbf{x} - \mathbf{H}\theta)^T(\mathbf{x} - \mathbf{H}\theta)\right]. \tag{1.12}$$

If ω is known then $\mathbf{H}$ is a constant matrix and an estimate of θ is given by the least-squares solution as

$$\hat{\theta} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \tag{1.13}$$

In (1.13), $\mathbf{H}$ must have full column rank. This is generally true except when the angular frequency ω is equal to zero or a multiple of π. Using the least-squares solution (1.12), the MLE criterion function can be concentrated to one parameter,

$$g(\omega) = \mathbf{x}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}. \tag{1.14}$$

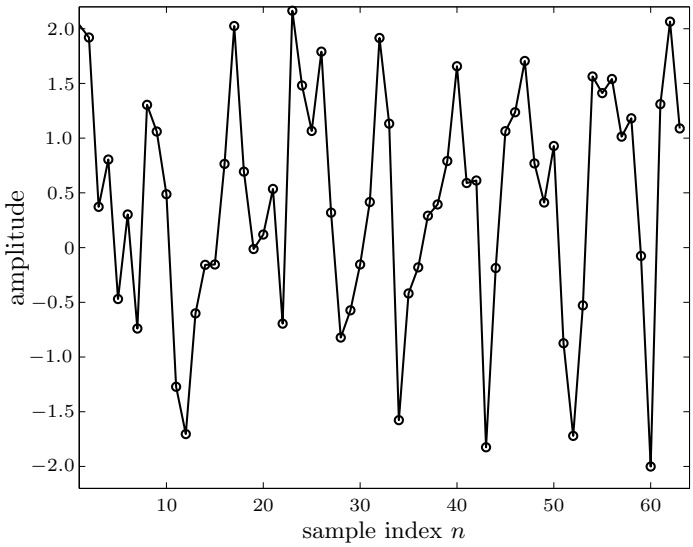


Figure 1.1: Example measurements on a single sinusoid disturbed by additive white Gaussian noise.

The frequency estimate is then given by maximizing $g(\omega)$, that is

$$\hat{\omega} = \arg\max_{\omega} g(\omega) \tag{1.15}$$

The equation (1.15) can be solved by a non-linear search or by using an iterative step method, i.e a Gauss-Newton iteration [SMFS89], [S1057]. However, when iterative methods are employed an initial estimate of the frequency is required. The perhaps most common way of solving an approximation of (1.15) is to calculate the discrete Fourier transform (DFT) of the signal $\mathbf{x}$ and then find the location of the dominating peak. This approximative solution corresponds to (1.15) if the inverse of $\mathbf{H}^T\mathbf{H}$ is replaced by a scaled identity matrix. This is a valid approximation for large N, that is if $N \gg 1/\omega$. Using the data samples in Figure 1.1, (1.14) is evaluated and displayed in Figure 1.2. From the peak location in Figure 1.2 an estimate of the frequency is obtained. In this example, the noise power is equal to the signal power. The signal to noise ratio (SNR) is defined by

$$\text{SNR} = \frac{\alpha^2}{2\sigma^2} = \frac{A^2 + B^2}{2\sigma^2} \tag{1.16}$$

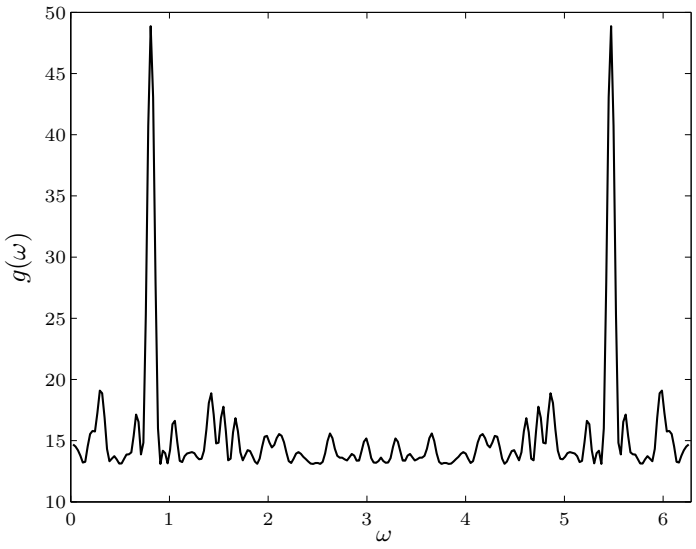


Figure 1.2: MLE criterion function $g(\omega)$. The true angular frequency is $\omega = 0.811$, $N = 64$ and SNR = 0dB.

The signal peaks in Figure 1.2 are easy to distinguish from the noise floor as long as the SNR is high enough. The threshold above which the signal peaks can be distinguished from the noise floor depends on the number of data samples N and the SNR. A rule of thumb is [SB85]

$$\mathrm{SNR}\frac{N}{\ln N} \gg 1 \tag{1.17}$$

In practical applications a factor 70 is suitable [SB85]. Using the estimate $\hat{\omega}$ obtained from (1.15) the least-square solution (1.13) can be used to obtain the parameters in θ. The estimated parameters versus the true parameters for the considered experimental data are listed in Table 1.1. As seen in Table 1.1 the estimated parameter values do not exactly coincide with the true ones. The accuracy of any estimation method is strongly dependent on the number of data samples N and the SNR. A lower bound on the variance of the frequency estimate from an unbiased estimator is given by the Cramér-Rao (CRB) [Kay93]. For the sought frequency, it is well known that a large N approximation of the CRB is

	true	estimated
α	1.00	1.052
ϕ	0.67	0.669
ω	0.811	0.812
A	0.620	0.653
B	0.784	0.825
C	0.410	0.423

Table 1.1: Comparison of the true and estimated parameter values.

given by [RB74]

$$\mathrm{var}\{\hat{\omega}\} \geq \frac{12}{\mathrm{SNR}N^3}. \tag{1.18}$$

1.4 Contributions and Outline

As indicated in the previous section, frequency estimation has been studied for a long period of time, and an enormous amount of references can be found in the literature. See, for example, the detailed list of references [Sto93]. However, there are still some white spots on the frequency estimation map. The aim of this thesis is to explore some of these white spots in order to get additional insight into this narrow-band research problem, as well as investigate the associated model-order and waveform fitting problems.

This thesis can be divided into three major parts. The first part (Part I, Chapter 2, 3 and 4) concerns parameter estimation utilizing coarse quantization in general, and frequency estimation in particular. The second part (Part II, Chapter 5 and 6) concerns properties and extensions of the IEEE standard 1057 (1241) sinewave fitting algorithm. The third, and final part, (Part III, Chapter 7 and 8) concerns parameter estimations of multiple sinewaves.

The chapters are written as individual parts and can be read independently from each other. Accordingly, some overlap in the contents of the chapters may exist as well as some differences in the used notation. An overview of the contents is given below.

Part I

Chapter 2

A method for fast frequency estimation by table look-up (FFETL) processing is presented. The estimation is based on data that has been quantized at one bit per sample, and all data processing is represented by a single table look-up operation, resulting in $\mathcal{O}(1)$-complexity. The performance of the proposed method is compared with the proper CRB for one bit quantized data, and the MLE for unquantized data by aid of Monte Carlo simulations. The FFETL method is shown to be (almost) statistically efficient over a wide range of SNRs, as encountered in practical applications. Practical aspects such as implementation issues, and the performance limitations due to quantization are discussed in some detail. The contents of this chapter have been published in

> Tomas Andersson, Mikael Skoglund, and Peter Händel. Frequency estimation by 1-bit quantization and table look-up processing. In *Proceedings European Signal Processing Conference*, pages 1807–1810, Tampere, Finland, September 2000. EURASIP.

Chapter 3

This chapter is devoted to implementation of the frequency estimator of Chapter 2 as well as an application on post-correction of analog-to-digital converters (ADCs). The table-lookup method presented in Chapter 2 is well suited for implementations requiring a low complexity. A prototype based on low-level logic circuits has been developed for demonstration purposes. In this chapter, the functionality of the prototype is described, as well as a few implementation issues are discussed. Some issues regarding low-complexity implementation of parameter estimators are included in

> P. Händel, M. Skoglund, T. Andersson, and A. Høst-Madsen. Method and apparatus for estimation physical parameters in a signal. *Swedish Patent 520067*, May 2003.

Recently, the experimental performance of the prototype was evaluated by aid of an available measurement test-bed, see [BAH05]. The evaluation is also available as

> Tomas Andersson and Peter Händel. Experiments on a hardware frequency estimator utilizing table look-up processing. Technical report IR-S3-SB-0541.

A tentative application of the fast frequency estimator is for error correction of ADCs. It is possible to take the frequency contents of the ADC output into account in order to make a more accurate ADC correction than methods only utilizing the present input value of the input. The joint work is presented in [LASH02]

> Henrik Lundin, Tomas Andersson, Mikael Skoglund, and Peter Händel. Analog-to-digital converter error correction using frequency selective tables. In *RadioVetenskap och Kommunikation (RVK)*, pages 487–490, June 2002.

Chapter 4

Fast analog-to-digital conversion with 1-bit per sample does not only make high sampling rates possible, but also reduces the required hardware complexity. For short data buffers or block lengths, it has been shown in Chapters 2-3 that tone frequency estimators can be implemented by a simple table look-up. In this chapter, an analysis is presented of such tables using the Hadamard transform. As an outcome of the analysis, a class of nonlinear estimators of low complexity is proposed. Their performance is evaluated using numerical simulations. Comparisons are made with the proper CRB and with the table look-up approach. Chapter 4 is available as

> Tomas Andersson, Mikael Skoglund, and Peter Händel. Frequency estimation utilizing the Hadamard transform. In *IEEE Workshop on Statistical Signal Processing*, Singapore, pages 409–412, August 2001.

Part II

In testing digital waveform recorders and analog-to-digital converters, an important part is to fit a sinusoidal model to recorded data, as well as to calculate the parameters that in least-squares sense result in the best fit. Algorithms performing a sinewave fit have been standardized in IEEE standard 1057 and IEEE standard 1241 [S1057,S1241]. Depending if the

sinewave frequency is known, or not, the algorithms are often denoted by the three-parameter fit and the four-parameter fit, respectively; where the three-parameter fit includes fitting of amplitude, initial phase and DC-offset. Normally, the four-parameter fit is employed, and software implementations of it are described in, for example, [MK01, Bla99]. A detailed performance comparison between the three- and four-parameter fit can be found in Chapter 5.

Notes on the performance of the four-parameter algorithm can be found in [BMS$^+$02, Hän00]. In [BMS$^+$02], the performance dependence on the initial estimates was addressed. The stop criterion of the iterative four-parameter sinewave fit was also given some attention. In [Hän00], the small error performance of the four-parameter algorithm is compared with the performance of an alternative nonlinear least-squares algorithm. The alternative algorithm in [Hän00] utilizes the fact that the least squares criterion can be concentrated with respect to three of the parameters, and thus the problem is reduced to a one-dimensional optimization problem. Some alternative methods to the four-parameter fit of [S1057] can be found in [GT97, HDM99, dSS01]. In this research area, the thesis contribution are listed below.

Chapter 5

Chapter 5 deals with some fundamental properties of the sinewave fit algorithm included in IEEE Standards 1057 and 1241 [S1057], [S1241]. Asymptotic Cramér-Rao bounds for three- and four model parameters are derived under the Gaussian assumption. Further, the sinewave fitting properties of the algorithm are analyzed by the parsimony principle [SS89]. A decision criterion whether to use the three- or four parameter model is derived. It is shown that a three parameter sinewave fit produces a better fit than the four parameter fit, if the frequency is known to be within an interval related to the number of samples and the signal-to-noise ratio. By a numerical analysis the theoretical results are shown to also be valid for the uniform noise model of quantization. Chapter 5 is available as

> Tomas Andersson and Peter Händel. IEEE standard 1057, Cramér-Rao bound and the parsimony principle. *IEEE Transactions on Instrumentation and Measurements*. In Press.

A short version is available as

> T. Andersson and P. Händel. IEEE-STD-1057, Cramér-Rao bound and the parsimony principle. *8th International Workshop on ADC Modeling and Testing*. Perugia, Italy September 8-10, 2003.

Chapter 6

Chapter 6 is an extension of Chapter 5, and presents a criterion for model order selection. By usage of the parsimony principle the mean sum-square-error is evaluated for models subject to imperfections in parameter values. In particular, model imperfections in different sinewave-fitting scenarios are analyzed. The analysis is carried out considering linear models. The obtained result is generalized to models incorporating non-linear parameters. Numerical illustrations are provided in order to gain insight of the behavior of model imperfections, as well as to numerically verify the theoretical results. The main contributions include a general result for linear signal models, as well as some novel results on sinewave-fitting. This work is presented in

> Tomas Andersson and Peter Händel. Robustness of wave-fitting with respect to uncertain parameter values. In *Proceedings IEEE Instrumentation and Measurement Technology Conference*, May 2005. Ottawa, Canada.

Part III

Multi-sinewave test methods require algorithms for multiple-tone parameter estimation. There exist a vast amount of publications on the topic [Sto93]. When several sinewaves are present in the signal one can generally not rely on algorithms designed for signals with only a single sinusoid. In this part of the thesis two algorithms are presented that jointly resolves several sinewaves. Emphasis has been made on practical convergence, i.e that the estimates of the parameters converge to the true parameters.

Chapter 7

In this chapter, the single sinewave fitting algorithm, described in Chapter 5, is extended to the multi-tone case. The main objective is to derive a multi-tone algorithm based on the standardized single-tone fit in IEEE

standards 1057 and 1241, respectively. By utilizing the expectation maximization (EM) algorithm in combination with a single-tone fit one can estimate the parameters for each sinewave independently. Further it is shown that the algorithm produces statistically efficient frequency estimates at high signal to noise ratios, that is the variance of the estimates reaches the CRB, independently of the actual number of tones present in the measured data. A full version of this chapter is presented as

> Tomas Andersson and Peter Händel. Multiple-tone estimation by IEEE standard 1057 and the expectation-maximization algorithm. *IEEE Transactions on Instrumentation and Measurements*. In press.

and a short version is available as

> Tomas Andersson and Peter Händel. Multiple-tone estimation by IEEE standard 1057 and the expectation-maximization algorithm. *IEEE Conf. on Instrumentation and Measurement*, Vail, CO, May 2003.

Chapter 8

This chapter presents a generalization of the IEEE four-parameter sinewave fit algorithm suitable to handle data comprising multiple sinewaves. The proposed method directly estimates the $3p + 1$ parameters of a p-tone model. The algorithm is analyzed numerically with emphasize on its convergence properties and statistical efficiency. The initialization of the algorithm is of major importance and an attempt to formulate a proper initialization procedure is presented. This work is presented as

> Tomas Andersson and Peter Händel. Toward a standardized multisinewave fit algorithm. In *9th European Workshop on ADC Modelling and Testing*, volume 1, pages 337–342, Athens, Greece, September 2004.

1.5 Conclusions and Future Research

With the present thesis as a starting point several directions of further research can be outlined. The conclusions and topics for further research are presented below, and not in a separate chapter at the end of the thesis.

1.5.1 Conclusions

High-speed Tone Frequency Estimation

In this thesis, among other things, a novel frequency estimator using table look-up processing has been proposed. The algorithm is derived using an input signal quantized with only 1-bit per sample. Methods for creating the table have been studied. Theoretical results and training approaches have been proposed. Further studies of the table used in the table look-up estimator using the Hadamard transform have resulted in a new class of low-complexity estimators. The studies have shown that such estimators are appropriate when the input signal is quantized with 1-bit.

The table look-up estimator is appropriate when a fast and low complexity processing is of importance. As an application example an ADC post correction application using the table look-up estimator has been presented. Further, the simplicity of the table look-up estimator has been visualised with the development of a demonstrator. Measurements on the demonstrator have shown that the performance is in accordance with performed numerical simulations.

Waveform Fitting

Standardized waveform fitting methods have been studied. Performance analyses employing the Cramér-Rao bound and the parsimony principle have been performed. The quality of the waveform fit is evaluated in terms of the mean sum-squared-error. A simple rule-of-thumb is derived suitable when selecting a proper estimation algorithm for the given problem. Also, the influence of quantization has been considered to some extent. In particular the presented analysis has been shown to be valid under the uniform noise model of quantization.

When using a linear model a simple criterion for model order selection has been derived. A generalization of this result to include non-linear models has been studies in the special case of sinewave fitting.

Multi-Tone Sinewave Fitting

An algorithm solving the multiple sinewave parameter estimation problem has been proposed. The algorithm utilizes the standardized wave-fit method along with the expectation maximization algorithm. Studies presented in this thesis have shown that the algorithm produces statistically

efficient estimates of the parameters. A second algorithm solving the multiple sinewave parameter fitting problem by extending the standardized single sinewave fitting algorithm has been proposed. Initialization issues and convergence has been studied by means of numerical simulations. Multi-tone waveform fitting is a non-trivial task and the proposed methods have potential to be useful tools for the practician in instrumentation and measurement.

1.5.2 Future Research

Fast methods for multi-bit data

The table look-up approach is highly memory consuming. The outcome from Chapter 4 can be viewed as way of trading memory against computations. In this thesis, these methods were studied when the input is quantized with 1-bit. Studies have shown [HMH00] that frequency estimation using 4-bit quantization is almost as good as using unquantized input data. An algorithm combining the 1-bit technique with the use of 4-bit data is an interesting topic which may be subject to future research.

Robustness of wave-fitting

The result derived in Chapter 5, about whether to choose a three- or a four-parameter method, is intuitive and easy to understand. The result has been generalised when a linear model is employed. The result also holds for a non-linear model in the special case of sinewave fitting. Further generalization of the result to a general non-linear model may be possible and is subject to further studies.

Easy-to-use algorithms

When constructing an algorithm to be subject for possible standardization, one must make the usage easy and *false proof*. In the two proposed multi-sinewave algorithms (Chapter 7 and Chapter 8, respectively) there are two issues that are *swepth under the carpet*, namely *i*) the detection of the number of sinewaves and finding initial estimates of the parameter to ensure convergence, and *ii*) since the both algorithms are iterative, when to stop the iterations.

i) If the tones are well separated in frequency the detection is generally not a problem. For instance the number of tones can be detected

by the most significant peaks in the periodogram. If the tones are closely separated in frequency and differs in amplitude other methods must be applied to solve the problem, i.e [Fuc88] [Fuc94].

ii) Often the user has some accuracy demands that could be used to decide when to stop iterating. That is, if the parameters converge. If, one the other hand, the parameters diverge this has to be detected to notify the user that the estemated parameters are not to be trusted.

These two issues are equally important in order to design an easy to use multiple-sinewave fitting algorithm. Important topics for further research include detection schemes and stop criterions for this type of algorithms.

Part I

High-speed Tone Frequency Estimation

Chapter 2

Frequency Estimation by Table Look-up

2.1 Introduction

Tone frequency estimation from an N-sequence of noise corrupted data

$$\{x[0], \ldots, x[N-1]\} \tag{2.1}$$

is a well-established research area, and several estimators have been proposed during the past decades. If the additive noise is white Gaussian, the maximum likelihood estimator (MLE) of the unknown frequency f is given by a non-linear least squares fit of a sinusoidal model to the samples (2.1) [RB74]. For a large sample-size N, the MLE is known as the location at which the periodogram $P(f)$, the magnitude squared Fourier transform of observations (2.1), attains its maximum. In practice an efficient approximation of the MLE can be implemented by aid of the fast Fourier transform (FFT) of the discrete time observation followed by a search for the maximum of the power spectral density [RB74]. An FFT-based implementation requires $\mathcal{O}(N \log N)$ floating point operations. Fast methods of order $\mathcal{O}(N)$ have been derived, e.g. by fitting a straight line to the unwrapped phase of data [Tre85, Kay89]; See [FL96] for efficient implementations of the algorithm in [Kay89]. The transformation from $x[n]$ to phase data is often implemented by a table look-up. Processing speed can be increased by replacing the floating point arithmetics with fixed point. It was shown in [HMH00] that 1-bit processing

is sufficient if a sampling rate beyond Nyqvist is employed.

In this chapter, faster methods based on 1-bit quantized observations will be studied. Using quantized data it is possible to design an algorithm of computational complexity $\mathcal{O}(1)$. Our scheme utilizes the fact that for a finite N there exist only a finite number of different possible realizations of the observed data, and hence the whole frequency estimator can be implemented using a table look-up approach. This approach is described in the following section.

2.2 Frequency Estimation by Table Look-up

Consider the signal model

$$x[n] = s[n] + w[n], \qquad s[n] = A\sin(2\pi f n + \phi) \tag{2.2}$$

where $A > 0$ is the amplitude, ϕ the initial phase, and f is the normalized frequency, $0 < f < 1/2$, i.e. $f = F/f_s$ where F is the signal frequency and f_s is the sampling frequency. The noise is assumed white and Gaussian with variance σ^2. Our aim in this work is to devise an estimator, say $\hat{f}$, that strives to estimate the true value, say f_0, of the unknown frequency f, based on a block of observed data according to (2.1). We utilize the assumption that before the data is processed by the estimator, the observations $x[n]$ are quantized to form a *binary sequence* according to

$$y[n] = \mathrm{sign}(x[n]) \tag{2.3}$$

where $\mathrm{sign}(x) = 1$ for $x \geq 0$ and $\mathrm{sign}(x) = -1$ for $x < 0$. Our goal is then to find an estimator $\hat{f} : \{\pm 1\}^N \to \mathbf{R}$, operating on the observed and quantized data

$$\{y[0], \ldots, y[N-1]\} \tag{2.4}$$

that is optimal in the sense of minimum mean-square error (MMSE). That is, we strive to find the estimator that minimizes $\mathrm{E}[(\hat{f} - f)^2]$ over all possible $\hat{f}$.

One key observation of this work is that, because of the quantization, the number of possible sequences that can be observed by the estimator is finite. More precisely, we note that a particular observed sequence according to (2.4) of length N can be uniquely mapped to an integer $i \in \{0, 1, \ldots, M-1\}$, with $M = 2^N$. The mapping from an observed

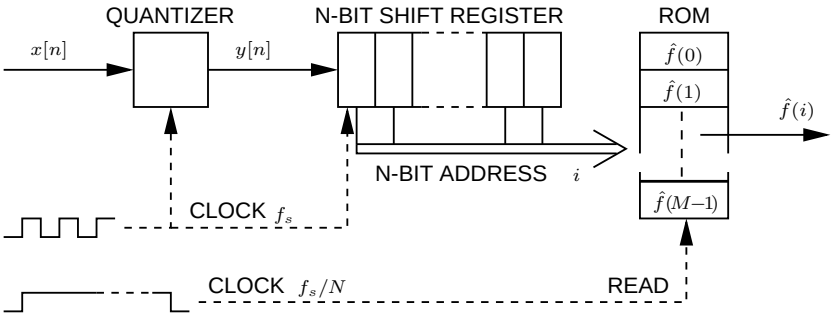


Figure 2.1: Exemplary description of frequency estimation by table look-up processing. An N-sequence of binary data $\{y[0], \ldots, y[N-1]\}$ defines the pointer i corresponding to a frequency estimate $\hat{f}(i)$. The stored table is designed according to the MMSE criterion.

sequence to the index i is here chosen as

$$i = \sum_{n=0}^{N-1} \frac{1 - y[n]}{2} \cdot 2^n. \tag{2.5}$$

Since the observed data is of finite resolution there can only be a finite number of possible estimator outputs. Thus, without loss of generality, all possible frequency estimators based on a sequence of quantized data, as in (2.4), can be implemented in two steps: (*a*) determine the index i that corresponds to the observed sequence according to (2.5), and (*b*) use this index as a pointer to an entry in a table, the *look-up table*

$$\{\hat{f}(0), \hat{f}(1), \ldots, \hat{f}(M-1)\} \tag{2.6}$$

containing all possible frequency estimates that can be produced by the estimator. Designing the best possible frequency estimator is then equivalent to constructing the table (2.6). Under the MMSE criterion [Kay93], we have that the table entries should be chosen as

$$\hat{f}(i) = \mathrm{E}[f \,|\, i] \tag{2.7}$$

where the conditional expectation can be computed under the assumption that the *a priori* distribution for the unknown frequency f is known.

When the table has been calculated and stored, the operation of the new frequency estimator can be illustrated as in Fig. 2.1.

2.3 Look-up Table Design

If the noise and the frequency distribution are known *a priori*, it is possible to calculate the look-up table analytically. However, if the noise has a different color than white and/or the frequency distribution is more complex than just uniform, the analytical expressions tend to be tedious to evaluate numerically. An alternative approach is then to train the table using a large set of training data. These two approaches are discussed in detail below.

2.3.1 Calculating the Look-up Table

In the case of white Gaussian noise and a uniform frequency distribution over $[0, 1/2)$, we can calculate the table entries (2.6) explicitly from (2.7). Let $p_\theta(\theta)$ denote the probability density function (pdf) of the stochastic quantity θ. We have from (2.7)

$$\hat{f}(i) = \mathrm{E}[f|i] = \int_0^{1/2} f \cdot p_{f|i}(f|i)\, df \tag{2.8}$$

where $p_{f|i}(f|i)$ is given by Bayes' rule as

$$p_{f|i}\,(f|i) = \frac{p_f(f)\; p_{i|f}(i|f)}{p_i(i)} \tag{2.9}$$

where the probability mass function (pmf) $p_i(i)$ is given by

$$p_i(i) = \int_0^{1/2} p_f(f)\; p_{i|f}(i|f) df. \tag{2.10}$$

We note that the index i is a function of the sequence (2.4), as given by (2.5). Hence, to find $p_i(i)$ and $p_{i|f}(i|f)$ we need to examine the individual samples $y[n]$. For white noise $w[n]$ in (2.2) the pmf for each $y[n]$ (conditioned on frequency and phase) is given as

$$p_{y|f,\phi}(y[n]|f,\phi) \quad = \frac{1}{2}\left[1 + y[n] \cdot \mathrm{erf}\Big(\sqrt{\mathrm{SNR}}\;\sin(2\pi f n + \phi)\Big)\right] \tag{2.11}$$

where erf(·)[1] is the error function and the signal to noise ratio is defined as

$$\text{SNR} = \frac{A^2}{2\sigma^2}. \tag{2.12}$$

Since $y[n]$ are independent for different n (conditioned on frequency and phase) we have

$$p_{i|f,\phi}(i|f,\phi) = \prod_{n=0}^{N-1} p_{y|f,\phi}(y[n]|f,\phi) \tag{2.13}$$

where $y[n],\ n = 0, \ldots, N-1$, gives i according to (2.5). It is reasonable to assume that the phase ϕ is uniformly distributed over $[0, 2\pi)$. Thus we can remove the phase dependency in (2.13) by integrating over ϕ according to

$$p_{i|f}(i|f) = \frac{1}{2\pi} \int_0^{2\pi} p_{i|f,\phi}(i|f,\phi)\ d\phi. \tag{2.14}$$

Using the above results we can then form a closed form expression for the table entries in (2.8) as

$$\hat{f}(i) = \frac{\displaystyle\int_0^{1/2} f \cdot g(i,f)\ df}{\displaystyle\int_0^{1/2} g(i,f)\ df} \tag{2.15}$$

where

$$g(i,f) = \int_0^{2\pi} \prod_{n=0}^{N-1} p_{y|f,\phi}(y[n]|f,\phi)\ d\phi. \tag{2.16}$$

In (2.16), $p_{y|f,\phi}(y[n]|f,\phi)$ is given by (2.11) and $\{y[0], \ldots, y[N-1]\}$ is related to i as described in (2.5).

From (2.15) we note that, for a given i, the value of $\hat{f}(i)$ is a function of the SNR only.

[1] $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$

2.3.2 Training the Look-up Table

A straightforward alternative approach to determine the table entries (2.6), for a given SNR, is to use a *training sequence* $\mathcal{T} = \{i_k\}_{k=1}^{K}$ (where each i_k corresponds to a particular length-N block of quantized data). Such a sequence can be obtained by simulating the assumed model for $y[n]$. That is, the k-th index i_k in the training sequence is determined as:

a) Draw a frequency f and an initial phase ϕ, according to known (or assumed) *a priori* distributions on these;

b) Draw noise samples $w[n]$ according to a known (or assumed) distribution and compute $x[n]$ according to (2.2), for $n = 0, \ldots, N-1$;

c) Quantize according to (2.3), and;

d) Determine the resulting index i_k according to (2.5).

Repeat steps *a*–*d* to get a new index, i_{k+1}. Now, given a training sequence $\mathcal{T}$, the i-th table entry, $\hat{f}(i)$, can be computed as the average over all frequencies f that gave rise to those sequences that correspond to index i. For completeness we let $\hat{f}(i) = 0$, for those i that are not in $\mathcal{T}$ (if any).

The main advantage of the training set approach is that it is relatively insensitive to the distribution of the noise and the *a priori* distributions for f and ϕ, and it may therefore be used, e.g., when the noise color is such that an analytical treatment according to Sec. 2.3.1 is infeasible.

2.3.3 Storage Complexity

It is readily realized that the size of the table (2.6) grows very fast (exponentially) with the block-size N. Consequently, the straightforward approach, as described above, of computing all table entries and then storing the whole table is not practical for N greater than, say, 20–25. However, we emphasize that there are several possible approaches that can be employed to compress the table and reduce its size.

It is straightforward to realize that a sequence $y[0], \ldots, y[N-1]$ and its complement, obtained by switching $-1 \leftrightarrow +1$, give rise to the same estimate $\hat{f}$. Hence, only $\hat{f}(i)$ for $i = 0, \ldots, M/2 - 1$ need to be stored (since $\hat{f}(M-1-i) = \hat{f}(i)$). Moreover, there are more sophisticated techniques to compress the table, for example similar to the one used in [SS98] (in a different application). Such techniques are subject to further study in Chapter 4.

2.4 Performance Evaluation

In this section we evaluate the performance of the considered frequency estimator.

2.4.1 Loss in Performance due to Quantization

A lower bound on the variance of any unbiased estimator is given by the Cramér-Rao bound (CRB) [Kay93]. For the model (2.2) the CRB of frequency is known to be inversely proportional to the SNR and the third power of N, see e.g. [RB74]. Due to the quantization in (2.3) inferior performance of any estimator employing $y[n]$ is expected over the MLE of frequency given $x[n]$. This loss in performance can be reduced by proper over-sampling prior to quantization [HMH00].

Further, for finite N all estimators employing $y[n]$ are biased (also asymptotically as $\sigma^2 \to 0$). For practical values of SNR and N, however, the squared bias may be negligible in comparison with the variance. Thus, a comparison between the CRB for (2.2) and for the augmented model (2.2)–(2.3) is indeed relevant. In [HMH00], it was shown that for 1-bit quantized data the CRB strongly depends on the frequency of the signal. But it was also shown that $\tau > 4$ times oversampling is as good as infinite quantization in terms of a frequency independent asymptotic CRB. Combining the CRBs for unquantized data and 1-bit data, we obtain an oversampling factor $\tau > 4$ for which we expect a similar lower bound on accuracy processing 1-bit observation $y[n]$ in place of $x[n]$. For high SNR (2.12) we obtain $\tau \approx 1.2\sqrt{\text{SNR}}$.

2.4.2 Training and Evaluation

A look-up table was determined using the training set approach of Sec. 2.3.2. The frequency f was taken as a set of realizations equally distributed over $[0, 1/2)$, and with ϕ according to a uniform distribution over $[0, 2\pi)$. Three look-up tables were trained at the different SNRs of 0 dB, 20 dB, and ∞ dB (no noise), respectively. The number of data points per block was $N = 16$, and the number of indices in the training set was $K = 10^{10}$, for the noisy cases, and $K = 10^7$ for the noise-free case.

The performance of the proposed method (denoted by FFETL-frequency estimation by table look-up) was tested on independent sets, of size $5{\cdot}10^5$, of evaluation data per SNR value.

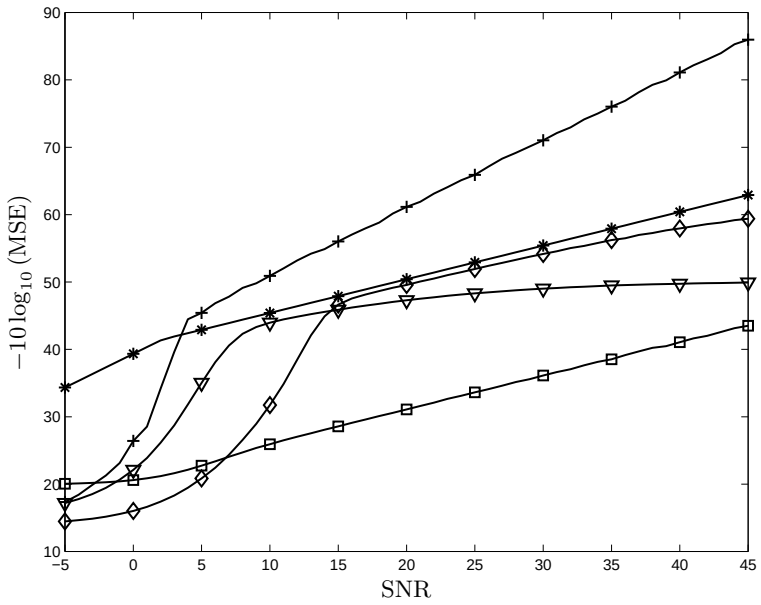


Figure 2.2: Empirical MSE as function of SNR for the proposed method. Three look-up tables trained at the different SNRs of 0 dB ($\triangledown$), 20 dB ($\Diamond$) and noise-free ($\square$) are compared. As reference, the performance of MLE for unquantized data ($+$) and the CRB ($*$) are displayed. The true signal frequency is $f_0 = 0.1$, and the data length is $N = 16$.

2.4.3 Performance versus SNR

In Fig. 2.2, the empirical mean-square error (MSE) is displayed for the three different tables. In this simulation we take the true frequency to be $f = f_0 = 0.1$. Clearly, the performance depends on the SNR level of the training data and we note that, by construction, FFETL is optimal (i.e. minimum MMSE) when working on data with the same SNR as the training data. The results displayed in Fig. 2.2 not only show this fact, but also the fact that choosing a lower SNR when training the look-up table results in a more robust estimator, i.e. with only a minor loss in performance at higher SNRs. The table trained by noise-free data results in the worst performance over the range of considered SNRs. For reference, the performance of the exact MLE for unquantized data [Kay93],

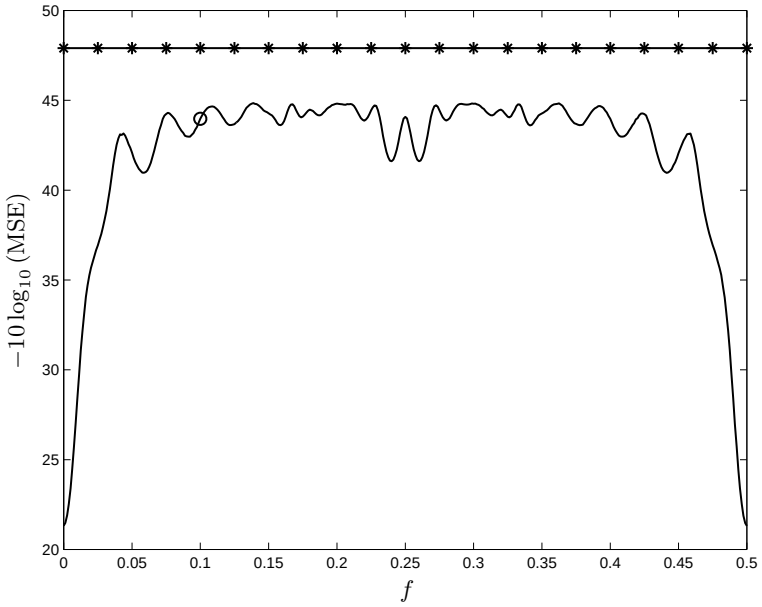


Figure 2.3: Empirical MSE (solid line) as function of frequency for the proposed method. SNR = 10 dB, data length $N = 16$ and the look-up table is trained at an SNR of 0 dB. As reference the asymptotic CRB is displayed ($*$). The frequency $f_0 = 0.1$ is marked with "$\circ$", corresponding to the result in Fig. 2.2 ($\triangledown$ at 10 dB).

and the asymptotic CRB for 1-bit quantized data [HMH00] are displayed in Fig. 2.2. Comparing with the performance of MLE indicates the loss of performance due to quantization (which can be compensated for, as discussed in Sec. 2.4.1), and also indicates an increased SNR threshold when processing 1-bit quantized data. However, it is obviously not "fair" to compare the performance of MLE with FFETL. A more reasonable benchmark is the asymptotic CRB for any unbiased estimator subject to 1-bit quantized data [HMH00]. One can note from Fig. 2.2 that in regions where the error variance dominates over the squared bias the performance of FFETL is close to the asymptotic CRB. In particular, FFETL with a look-up table trained at 20 dB has performance close to the asymptotic CRB over a wide range of SNRs (15–35 dB).

2.4.4 Performance versus Frequency

The performance of FFETL depends on the true signal frequency, f_0. The performance of FFETL as function of the true frequency is studied in Fig. 2.3. For frequencies f_0 near the boundaries at $f = 0$ and $f = 1/2$, FFETL is unable to provide a reasonable frequency estimate, resulting in a low MSE figure. This behavior is similar to the behavior of MLE of frequency for unquantized data. We observe, however, that the performance of FFETL at frequencies well within the region $(0, 1/2)$ is relatively constant.

2.5 Conclusions

A fast frequency estimator based on table look-up processing has been proposed. In an exemplary description an N-sequence of 1-bit quantized data is used as pointer to a 2^N-cell memory containing all of the possible different estimates. The look-up table, i.e. the set of memory entries, is constructed by minimizing an MMSE criterion. The performance of the described method has been evaluated by aid of Monte Carlo simulations and compared with the appropriate Cramér-Rao bound. It has been shown that the method is able to produce almost statistically efficient estimates of the signal frequency for a wide range of scenarios of practical interest.

Chapter 3

Implementation and Application

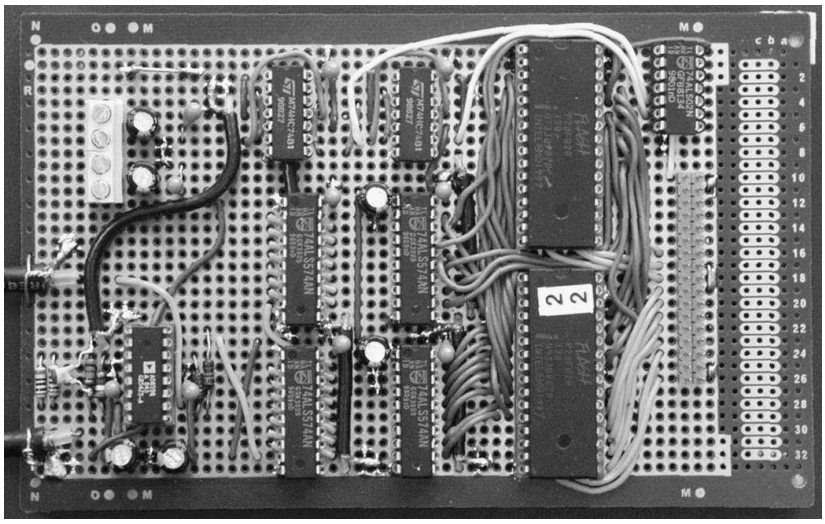
This chapter presents one implementation and an application of the table look-up estimator described in Chapter 2. The purpose with the prototype is to illustrate the simplicity of the estimator. The second part of this chapter present a joint work with ADC post correction. The frequency estimator is used in combination with static ADC post correction tables, which in combination make the ADC post correction scheme frequency-dependent.

3.1 A 20MHz Prototype

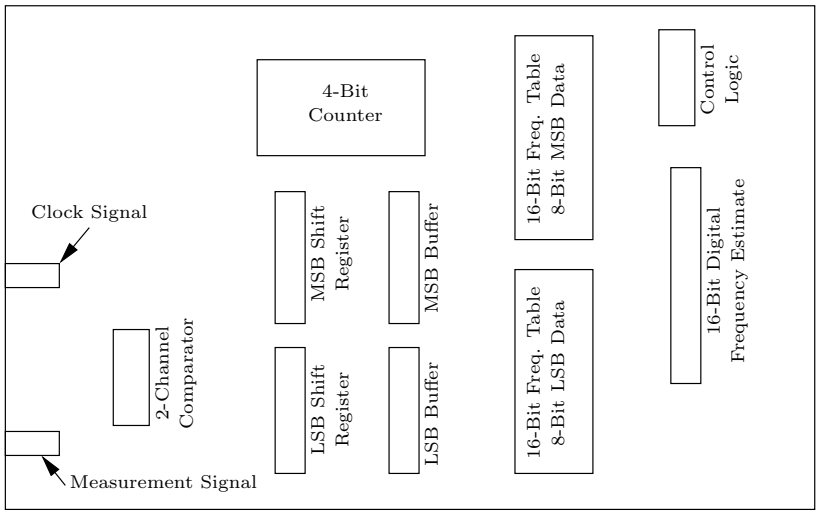
The main purpose of developing the prototype card was to illustrate the simplicity of the frequency estimator using table-look up described in Chapter 2. It is shown, with this prototype that it is possible to estimate the frequency of a signal up to 10MHz using on-the-shelf standard components. Figure 3.1(a) includes a picture of the prototype board. A description of the different circuits on the board is found in Figure 3.1(b).

3.1.1 Function

The names of the different parts on the prototype board are showed in Figure 3.1(b). In the following sections the function of each part is described shortly.



(a) Prototype Circuit Board



(b) Prototype Board Block Description

Figure 3.1: 20MHz hardware prototype

Input Signals

The prototype board takes two input signals, **a)** measurement signal with unknown frequency and **b)** a clock signal which is used as the sample frequency as well as a clock signal to the circuits on the board. The input signal amplitude should be stronger than $100m\mathrm{V}_{peak}$. The clock signal could be an arbitrarily shaped signal as long as the zero-crossings are well defined and the amplitude is larger than $500m$V. The prototype board support clock signals up to 20MHz.

1-bit Quantizer

The 1-bit quantizer is build around a 2-channel comparator AD8598 from Analog Devices. The channels work independently of each other. The first channel generates a square-wave clock signal by comparing the input signal with the ground (GND). The square-wave use standard TTL levels and has a rise and fall time below $10ns$. The second channel is used to quantize the measurement signal using the following specification.

$$y(t) = \begin{cases} 0\mathrm{V} & x(t) < 0\mathrm{V} \\ 5\mathrm{V} & x(t) \geq 0\mathrm{V} \end{cases} \tag{3.1}$$

where $x(t)$ is the analog measurement signal and $y(t)$ is the output signal from the quantizer. The output voltage of 5V corresponds to a well defined HIGH LEVEL TTL signal or for short a '1'.

Shift Register

The shift register is built using two 74574, each containing 8 synchronous clocked D type flip-flops. The D flip-flops are connected in a stack in such a way that the data from a previous D flip-flop are the input to the next D flip-flop. In this way a 16-bit shift register is obtained. The quantized input signal obtained from the 1-bit quantizer is used as input to the first D flip-flop.

4-bit Counter

The 4-bit counter is the board's control unit. It controls the operation of all the other units on the board by sending control signals at each operation cycle. This type of behavior is called *state space machine*(SSM). The SSM has 16 cycles, and thus requires a 4-bit control unit. The 4-bit counter is constructed using four D-flipflops in two 7474 forming an

asynchronous counter. By empirical experiment it was found out that this design was faster than the one circuit solution using a synchronous 4-bit counter (74169).

During the first 16 cycles, samples are collected in the shift register. After the 16-th cycle the input of the buffer is opened and the contents of the shift register is loaded. The output of the buffer is connected to the address input of the memory.

8 cycles after the samples have been loaded into the memory the memory output data is opened. The memory output is then driving the estimator output delivering a frequency estimate as a 16-bit word.

Buffer

The buffer consist of two 74574's each containing 8 synchronous clocked D-flip-flops. The output from the shift register is used as an input. Each 16-th cycle the content of the buffer is replaced with the contents of the shift register. The output is connected directly to the address input of the frequency table.

16-bit Frequency Table

The frequency table is constructed using two 64k $\times$ 8-bit flash PROM memories. A flash PROM is a programmable read only memory which can be reprogrammed using a special programming device. Using two 8-bit memories in parallel a 16-bit output word is obtained.

Control Logic

The control logic consists of four inverting or (NOR) gates in a 7402 circuit. The control logic is used by the 4-bit counter to generate the appropriate signals to control the buffer and the memory.

16-bit Digital Frequency Estimate

The output frequency estimate is given as a 16-bit word in the interval $[0, 1/2]$. The output interval is uniformly quantized in 2^{16} levels. The output signal is valid during the 10-th and the 16-th cycle. Each new frequency estimate is signaled by a logic control signal. The control signal is available on the output bus as well.

3.1.2 Measurement Setup

Two Marconi 2024 signal generators were used. One generator for the clock signal, from here called the *reference signal*, and the other generator to produce the measurement signal with *unknown* frequency, from here on denoted as the *input signal*. The internal clocks in the two generators were synchronized in order to avoid frequency drift between the input signal and the reference. The reference frequency was set to 20MHz. As noise generator, a Rohde & Schwarz SMU 200A Vector signal generator was used. The output from the noise generator was combined with the input signal using a passive combiner to form the noise-corrupted input signal to the frequency estimator. The effective SNR could be adjusted by varying the noise amplitude using passive dampers and also by fine tuning the amplitude of the sinewave.

The 16-bit estimate was recorded into a first-in-first-out buffer (FIFO). The contents of the FIFO were then recorded for further analysis by using a parallel digital data interface. The performance of the frequency estimator was tested on independent sets of estimates. For each frequency a set of $5 \cdot 10^4$ estimates were collected.

3.1.3 Performance

The empirical mean-square error (MSE) measured using the hardware prototype is displayed in Figure 3.2. Also plotted is the MSE obtained from numerical simulations, see also Figure 2.2. In this simulation the signal frequency was set to $f = f_0 = 0.1$ and the frequency look-up table was trained at an SNR equal to 20dB. Noted from Figure 3.2 is that the measured MSE using the hardware prototype follows the corresponding simulated curve. In the measurement setup there was some difficulties to adjust the SNR which resulted in an accuracy of about $\pm$1dB. Taking this into account the hardware prototype performs in accordance with the simulations.

3.1.4 Performance versus Frequency

The performance of FFETL depends on the signal frequency, f_0, of the input signal. The performance of FFETL as function of the input frequency is studied in Figure 3.3. For frequencies f_0 near the boundaries at $f = 0$ and $f = 1/2$, the estimation using a look-up table is unable to provide a reasonable frequency estimate, resulting in a low MSE figure.

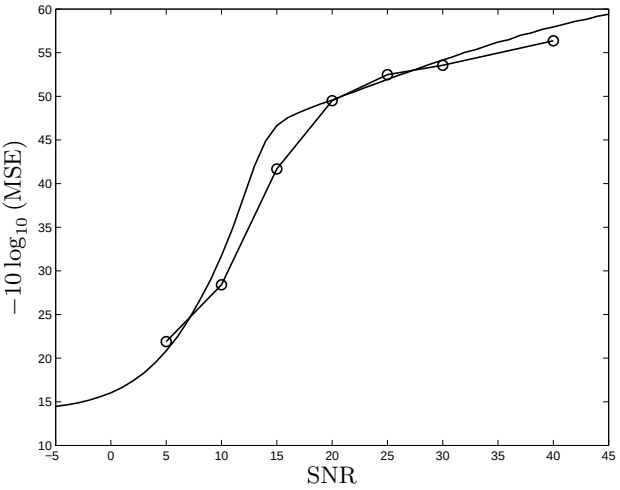


Figure 3.2: Empirical MSE as function of SNR using the hardware prototype (∘) as well as simulated performance (–). The true signal frequency is $f_0 = 0.1$, and the data length is $N = 16$.

This behavior is similar to the behavior of MLE of frequency for unquantized data. Also here, the performance of the hardware prototype is in accordance with the performance predicted by the numerical simulations.

3.1.5 Summary

A fast frequency estimator based on table look-up processing has been investigated. In an exemplary description an N-sequence of 1-bit quantized data is used as pointer to a 2^N-cell memory containing all of the possible different estimates. The look-up table, i.e. the set of memory entries, is constructed by minimizing an MMSE criterion. The performance of the described method has been evaluated by aid of experimental validation of a hardware prototype, Monte Carlo simulations and compared with the appropriate Cramér-Rao bound. Even though the circuits were specified to run at a maximum sampling rate of 20MHz the demonstrator worked stable using a sample frequency up to 40MHz. Using sub sampling the hardware prototype is expected to handle input signals up to ~ 100 MHz.

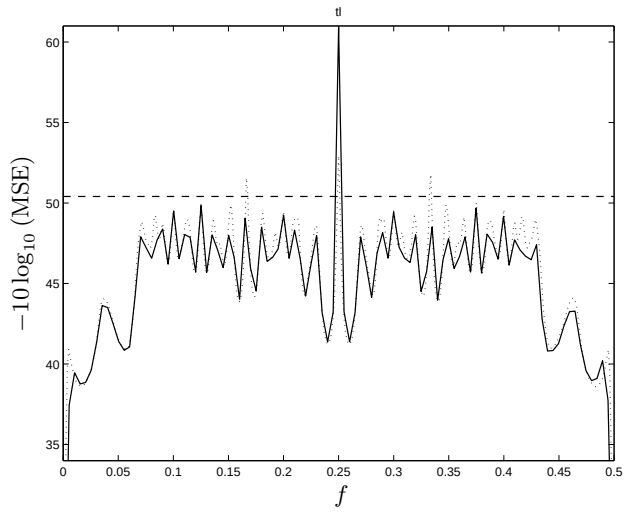


Figure 3.3: Empirical MSE (solid line) as function of frequency using the hardware prototype ($-$) and from numerical simulations ($\cdots$). SNR = 20 dB, data length $N = 16$ and the look-up table is trained at an SNR of 20 dB. As reference the asymptotic CRB is displayed ($--$).

3.2 ADC Correction

Error correction for analog-to-digital converters (ADCs) is considered. The frequency-dependent nature of ADC errors motivates the proposal of a novel scheme, incorporating look-up table correction and fast frequency estimation. The method is evaluated using experimental converter data, and the performance, measured in SFDR and SINAD, is found to be superior to that of non frequency-dependent correction methods.

The demand for highly linear ADCs is ever increasing. It is a well-known fact that practical ADCs suffer from various errors, e.g., gain, offset and linearity errors. These errors stem from numerous sources such as non-ideal spacing of transition levels and timing jitter, to mention a few, and they contribute to deterioration of the linearity of the converter. Several methods have been proposed to *externally* compensate for such errors, e.g., [HSP00, LSH01, IHK91, Mou89]. External in this case implies that digital signal processing methods which operate *outside* of the actual converter are used in the calibration and compensation schemes.

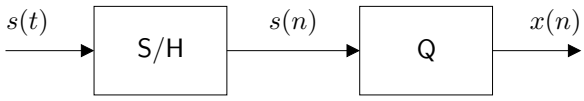


Figure 3.4: The ADC model. The first block is an ideal sample-and-hold circuit and the second block is an imperfect quantizer.

Here a b-bit ADC. The ADC will be modeled as an ideal sample-and-hold circuit followed by an imperfect quantizer, depicted in Figure 3.4. The sample-and-hold circuit samples the continuous-time input signal $s(t)$ at the sampling rate f_s, resulting in a discrete-time signal

$$s(n) = s(t)\big|_{t=n/f_s}. \tag{3.2}$$

The Q-block of Figure 3.4 then quantize $s(n)$ into one of the $M = 2^b$ output states $\{x_j\}$, $j = 0, \ldots, M-1$, and produces the corresponding output $x(n) = x_j$.

One frequently used method to correct ADCs is the look-up table correction. In classic look-up table correction, the correction table (containing the corrected output $\hat{s}_j$ associated with each ADC output state

x_j) is addressed using the present ADC output sample, $x(n)$. Obviously, this addressing yields the same correction for a given ADC output sample, regardless of the dynamic properties of the input signal. This is referred to as *static* correction. However, the errors of an ADC are in general frequency dependent. This often results in a severe performance loss for table look-up correction methods when applied at frequencies other than the calibration frequency. In this chapter a *frequency-selective correction* scheme is presented and evaluated. The method is based on two key components: a *fast frequency region estimator*, based on [ASH00], and a *correction table*. These are described in the following two sections. Results obtained using experimental ADC data are presented in Section 3.2.3.

3.2.1 Frequency Region Estimator

Tone frequency estimation from an N-sequence of noise corrupted data is a well known problem which can almost be considered solved. However, with "almost solved" we refer to the case where the number of data N is large, the signal to noise ratio (SNR) is high, and there exists an infinite amount of computer resources. In the ADC error correcting application this is not the case. The number of data N can not be made large, since we then will miss the information about the instantaneous frequency. In the considered application the SNR is usually high. In terms of computer resources this is an application where almost none are present.

A traditional way of constructing frequency estimators is by optimizing some criterion related to the frequency. The perhaps most commonly used method is the method of maximum likelihood, or approximate variants thereof [Kay93]. That is, choosing an estimate of the frequency in such a way that the model in use is the most likely given some data samples. In common for most frequency estimation methods is that the output frequency estimate is a continuous variable. Here, on the other hand, we consider the problem of finding the most probable region to hold the unknown frequency, out of a finite (small) set of regions.

Consider the input $s(n)$ to the quantizer in Figure 3.4 to be modelled as a sine wave and additive Gaussian noise. The input is then given by,

$$s(n) = A\sin(2\pi f_0 n + \phi) + w(n) \tag{3.3}$$

where $A > 0$ is the real valued amplitude, ϕ is the initial phase, f_0 is the unknown normalized frequency, $0 < f_0 < 1/2$, and $w(k)$ is the noise with variance σ^2. The output from the quantizer $x(n)$ is a b-bit quantized

version of $s(n)$. It has been shown [ASH00] that there exists a high-performance frequency estimator of low complexity employing only 1-bit of the input signal. The use of 1-bit data also has the advantage that the estimator does not depend on the power of the input signal, that is no gain control is needed. Here, we are not limited to use 1-bit data but the resulting structure with a table look-up procedure is tractable since it supports the demand of a fast estimator of low complexity.

Operation

The frequency estimator input $y(n)$ is given by the most significant bit (msb) of $x(n)$,

$$y(n) = \text{sign}(s(n)), \tag{3.4}$$

where $\text{sign}(x) = 1$ for $x \geq 0$ and $\text{sign}(x) = -1$ for $x < 0$. By collecting N successive binary samples at each time instant n, that is

$$\{y(n), \ldots, y(n-N+1)\}. \tag{3.5}$$

The number of possible input sequence are finite and can be uniquely mapped onto an integer $i \in \{0, \ldots, 2^N - 1\}$. The index i is then used as a pointer to an entry in a *frequency region estimation table*, see Figures 3.5–3.6. Finally, the i-th table entry contains a region estimate $\hat{F}(n) \in \{F_1, \ldots, F_K\}$, indicating that the instantaneous signal frequency is within the k-th frequency region. The frequency regions F_k are defined as,

$$F_k = \{f \in [0, 1/2) : |f - f_k| \leq |f - f_l|,\ l = 1, \ldots, K\} \tag{3.6}$$

where $k = 1, \ldots, K$. In this chapter, the frequencies f_l have been chosen equally spaced over the region $[0, 1/2)$, but could be chosen arbitrary over the space of possible input frequencies.

Design

As a frequency region estimate we choose the region that maximizes the probability of including the unknown frequency f_0, that is

$$\hat{F}(n) = \arg\max_{\forall F_k} \Pr\{f_0 \in F_k | y(n), \ldots, y(n-N+1)\} \tag{3.7}$$

A straightforward way to obtain the table is to use a training approach [ASH00]. Given a set of data samples $x(n)$ based on different frequencies,

within the regions $F_1, \ldots, F_K$, it is possible to build a training set $\mathcal{T} = \{i_l\}_{l=1}^{L}$, where each i_l corresponds to a block of N samples of the msb in $x(n)$. The samples $x(n)$ are generated using an input of a single sinusoid, at a known frequency, disturbed by noise. Hence, to each block i_l there is a corresponding true frequency belonging to one of the regions $f_1, \ldots, f_K$. Now, given a training set $\mathcal{T}$ the i -th table entry can be computed as the index of the most probable frequency region over those i_l corresponding to the index i. For completeness, we let $\hat{f}_k(i) = \lceil K/2 \rceil$ for those i that are not in the training set $\mathcal{T}$.

3.2.2 Correction Table

Static ADC correction yields the same corrected value $\hat{s}_j$ given the ADC output x_j, regardless of the signal frequency, while the errors sought to mitigate for in general are frequency dependent. The correction scheme presented here utilizes a frequency selective correction table. This is accomplished by extending the usual one-dimensional correction table of classical look-up table compensation to a two-dimensional table, using *both* the present ADC output $x(n) = x_j$ *and* the present frequency region estimate $\hat{F}(n) = F_k$ for addressing. This method can also be interpreted as selecting a specific one-dimensional correction table for each frequency estimate $F_k \in \{F_1, \ldots, F_K\}$. Thus, the corrected output $\hat{s}(n)$ is the table entry $\hat{s}_{j,\,k}$ associated with x_j and F_k. The correction system has two operation modes, *compensation* and *correction*, which are briefly described below, see [LASH02], for a detailed description.

Compensation and Calibration

In compensation mode, i.e. normal ADC operation with correction engaged, the ADC output sample, $x(n)$, is mapped through the correction table to a compensated output value $\hat{s}(n)$. The correction is determined by the present ADC output together with the current frequency region estimate, as depicted in Figure 3.5. Thus, the compensation becomes

$$\begin{gathered} s(t) \rightarrow (x_j,\, F_k) \rightarrow \hat{s}_{j,\,k} = \hat{s}(n) \\ \hat{s}_{j,\,k} \in \{\hat{s}_{i,\,\ell}\}_{(i,\,\ell)=(0,\,1)}^{(M-1,\,K)} . \end{gathered} \tag{3.8}$$

With this structure, the compensation is made dynamic, with table addressing depending on the frequency contents of the signal.

Prior to using the correction table for compensation, it must be calibrated. Generally, calibration is performed with a calibration signal $s(t)$

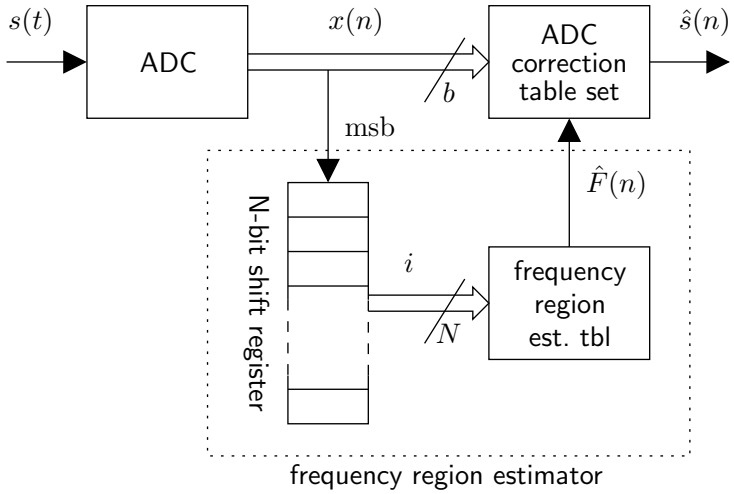


Figure 3.5: Compensation system outline. The frequency region estimator selects the appropriate ADC correction table.

applied to the ADC input. The calibration system outline is shown in Figure 3.6. The table entries $\hat{s}_{j,\,k}$ should be selected such that the resulting conversion $s(t) \rightarrow \big(x_j,\, F_k\big) \rightarrow \hat{s}_{j,\,k} = \hat{s}(n)$ is "better" than without correction. The employed design criterion is to minimize the mean squared error, $\mathrm{E}[(\hat{s}(n) - s(n))^2]$, where $\mathrm{E}[\cdot]$ denotes the expected value. Since the selection of $\hat{s}(n) = \hat{s}_{j,\,k}$ depends on the ADC output $x(n)$ and the frequency region estimate $\hat{F}(n)$, the criterion becomes

$$\hat{s}_{j,\,k} = \arg\min_{\hat{s}} \mathrm{E}\big[(\hat{s} - s(n))^2 \,\big|\, x(n) = x_j,\, \hat{F}(n) = F_k\big] \tag{3.9}$$

It can be shown [HSP00, Llo82] that in order to minimize the criterion (3.9), $\hat{s}_{j,\,k}$ should be set to the mean value of all input samples, $s(n)$, that were quantized into the value x_j while the frequency region estimate was equal to F_k. Under the interpretation that the frequency region estimate $\hat{F}(n)$ selects which table, out of a set of one-dimensional correction tables, to use, the result above is equivalent to saying that the correction value s_j in the k-th table should be set to the mean of all samples, $s(n)$, that produced the ADC output $x(n) = x_j$ while the k-th table was selected.

We see that in order to calibrate the correction table, the discrete time versions $s(n)$ of the analog calibration signal must be known. However,

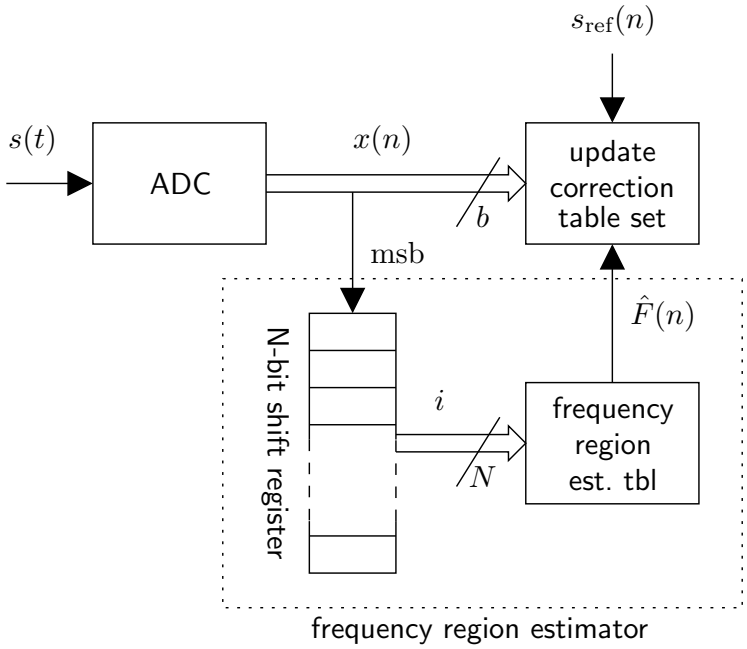


Figure 3.6: ADC correction table calibration system. The reference signal $s_{\mathrm{ref}}(n)$ is an estimate of the input signal samples $s(n)$.

these are in general not available and must therefore be estimated with some estimate $s_{\mathrm{ref}}(n)$. This estimate can be obtained in several ways; a "better" ADC in parallel with the ADC under test, a digitally generated calibration signal fed to the ADC through a digital-to-analog converter [TL97], or signal reconstruction using optimal filtering [HSP00] are all feasible methods for producing $s_{\mathrm{ref}}(n)$.

3.2.3 Performance

The proposed method has been evaluated with experimental ADC data from an Analog Devices AD876 10-bit converter, running at 20 MHz sampling frequency. The ADC correction table was calibrated using sinusoid calibration signals at several different frequencies. The calibration signal estimate $\hat{s}(n)$ was obtained using the optimal filtering method proposed in [HSP00].

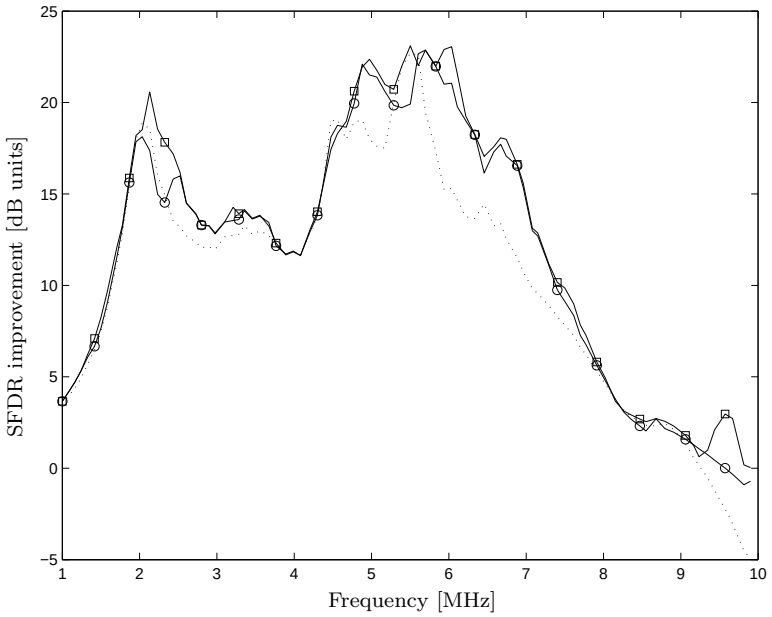


Figure 3.7: SFDR improvements for frequency-selective correction (solid lines) compared with static correction (dotted line). Circles represent a frequency estimator with 8 regions ($K = 8$) and squares represent 16 regions ($K = 16$).

Spurious-free dynamic range (SFDR) and signal-to-noise and distortion ratio (SINAD) [S1241] are used to evaluate the method. The performance is presented as SFDR and SINAD improvements compared to the uncompensated case, and is shown in Figures 3.7 and 3.8, respectively. The performance for a static correction scheme is also plotted in Figure 3.7-3.8 for comparison.

The frequency-selective correction was evaluated for two test cases: the first case having 8 frequency regions ($K = 8$) and the second case having 16 regions ($K = 16$). Both cases comprise a 16-bit shift-register ($N = 16$). The results indicate that the frequency-selective correction method is superior to the frequency-static method in general, but also that increasing the number of frequency ranges K from 8 to 16 does not give any significant improvement.

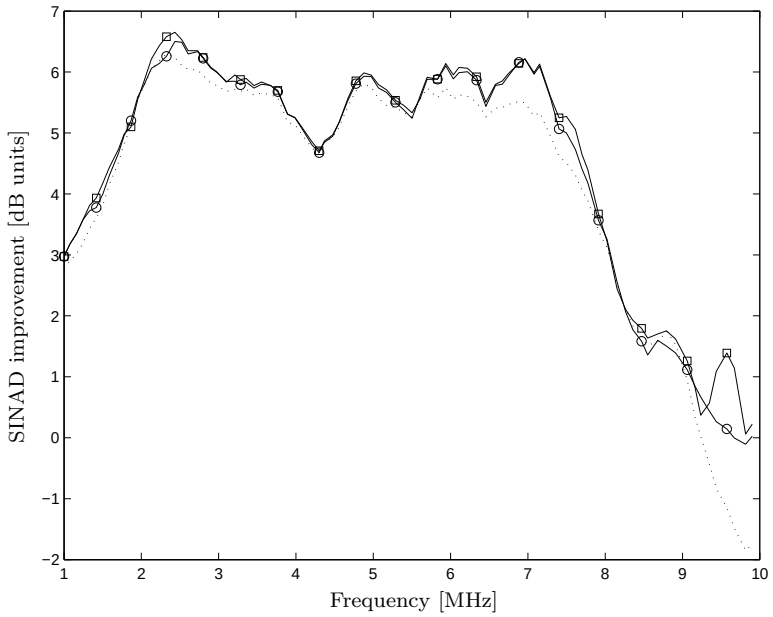


Figure 3.8: SINAD improvements for frequency-selective correction (solid lines) compared with static correction (dotted line). Circles represent a frequency estimator with 8 regions ($K = 8$) and squares represent 16 regions ($K = 16$).

We see from the results in Figure 3.7-3.8 that the SFDR is improved with between zero and 7 dB, while the improvement in SINAD is below 1 dB except at frequencies near the Nyquist frequency (10 MHz), where the SINAD improvement is approximately 2 dB. As opposed to the case of static correction, the improvement for frequency selective correction never fall below zero, i.e., the performance of the corrected ADC will never be inferior to that of the uncompensated ADC.

3.2.4 Conclusions on ADC-calibration

We have in this chapter derived an extension of classic look-up table ADC correction. The extension comprises a frequency selective look-up table method, using a fast frequency region estimator together with a

two-dimensional correction table. The motivation for using a frequency selective correction method was that the errors of AD converters in general vary with frequency. The proposed method was evaluated using experimental ADC data, and results showed that the ADC performance, measured as SFDR and SINAD, improved compared with the performance of a static correction scheme.

Chapter 4

Frequency Estimation Utilizing the Hadamard Transform

4.1 Introduction

Frequency estimation by using table-look-up methods has been discussed in Chapter 2. One of the conclusions were that the table grows exponentially with the number of data samples N. In the case of 32 data sample this would mean a table of $\sim 4 \cdot 10^9$ entries. Here the problem with table-look-up methods due to limited memory resources is addressed. By utilizing the Hadamard transform a class of suboptimal frequency estimators is derived. The key-point is that memory resources are traded against computations. However, the utilization of the one-bit quantization is taken into account when deriving the algorithm to make them fast and easy to implement.

Tone frequency estimation from an N-sequence

$$\{x[0], \ldots, x[N-1]\} \tag{4.1}$$

of noise corrupted data is a well-established research area and several estimators have been proposed during the past decades. In this chapter, the considered signal model is

$$x[n] = s[n] + e[n], \qquad s[n] = A\sin(2\pi f n + \phi) \tag{4.2}$$

where $A > 0$ is the amplitude, ϕ the initial phase, and f is the normalized frequency, $0 < f < 1/2$, i.e. $f = F/f_s$ where F is the signal frequency and f_s is the sampling frequency. The frequency f is an unknown parameter and the phase ϕ is assumed to be uniformly distributed over the interval $[0, 2\pi]$ (and independent of other signal parameters). The noise is assumed white Gaussian with variance σ^2.

The observed data $y[n]$ is assumed to be a a quantized version of $x[n]$ forming a *binary sequence*

$$\{y[0], \ldots, y[N-1]\} \tag{4.3}$$

according to

$$y[n] = \text{sign}(x[n]) \tag{4.4}$$

where $\text{sign}(x) = 1$ for $x \geq 0$ and $\text{sign}(x) = -1$ for $x < 0$. In an electronic circuit we would represent such binary data by ones and zeros.

Here we consider estimators that strive to estimate the true value, say f_0 (a deterministic constant), of the unknown frequency f, based on a *binary sequence* of the observed data according to (4.3). The goal is to find an estimator $\hat{f} : \{\pm 1\}^N \to \mathbb{R}$, operating on the observed and quantized data and optimal in the sense of minimum mean square error (MMSE). That is, to find the estimator that minimizes $\text{E}[(\hat{f} - f)^2]$ subject to an assumed *a priori* distribution for the unknown frequency f. The *a priori* distribution for the frequency is a design parameter of the estimator.

Because of the quantization, the number of possible different sequences (4.3) is finite. Hence, a particular observed sequence, of length N, can always be mapped to an index $i \in \{0, \ldots, M-1\}$, with $M = 2^N$, where the mapping from an observed sequence to the index i is chosen as

$$i = \sum_{n=0}^{N-1} \frac{1 - y[n]}{2} \, 2^n. \tag{4.5}$$

Since there is only a finite number of possible observed sequences, there is also a finite number of possible estimator outputs. Thus any estimator can be implemented in two steps: (*a*) determine the index i that corresponds to the observed sequence according to (4.5), and (*b*) use this index as a pointer to an entry in a *table*

$$\{\hat{f}(0), \hat{f}(1), \ldots, \hat{f}(M-1)\} \tag{4.6}$$

containing all possible frequency estimates. Under the MMSE criterion we have that the table entries should be chosen as

$$\hat{f}(i) = \mathrm{E}[f \,|\, i]. \tag{4.7}$$

where the expectation is with respect to the assumed *a priori* distribution for f, the phase and the noise, conditioned on the observed sequence (as represented by the index i). In Chapter 2 methods for computing estimator tables (4.6) based on (4.7) were studied. The performance of the resulting MMSE estimator was also investigated. As demonstrated in Chapter 2, table based frequency estimation performs well compared, e.g., with the Cramér–Rao bound for one-bit quantized data [HMH00]. However, the size of the table grows exponentially with the block-length N, and the method is hence not feasible for block-lengths larger than, say, 24–26 samples. The aim here is to investigate methods to *compress* the table, that is, characterizing the set of possible estimates $\hat{f}$ using (much) less than 2^N table entries. The main tool in achieving such compression is the Hadamard transform, as explained next.

4.2 The Hadamard Transform

Any function $\gamma : \{0, \ldots, M-1\} \to \mathbb{R}$, where $M = 2^N$ and with a finite domain represented by the integers $\{0, \ldots, M-1\}$, can be expanded as

$$\gamma(i) = \mathbf{t}^{\mathrm{T}}\, \mathbf{h}(i), \quad \text{with}$$

$$\mathbf{h}(i) \triangleq \begin{bmatrix} 1 \\ y[N-1] \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ y[0] \end{bmatrix} = \begin{bmatrix} 1 \\ y[0] \\ y[1] \\ y[0]y[1] \\ y[2] \\ y[0]y[2] \\ y[1]y[2] \\ \vdots \\ \prod_{n=0}^{N-1} y[n] \end{bmatrix} \tag{4.8}$$

where $\otimes$ denotes the Kronecker matrix product and with the relation between the index i and the binary variables $\{y[n]\}$ defined as in (4.5).

The vector $\mathbf{t}$, with elements $\{t_m\}$, is then the *Hadamard transform* of $\mathbf{g} = [\,\gamma(0)\cdots\gamma(M-1)\,]^{\mathrm{T}}$, computed as

$$\mathbf{t} = 2^{-N}\,\mathbf{H}\,\mathbf{g} \tag{4.9}$$

where $\mathbf{H}$ is the size $M \times M$ *Hadamard matrix*, with rows $\mathbf{h}^T(0)$, ..., $\mathbf{h}^T(M-1)$. Computing $\mathbf{t}$, as in (4.9), requires $\mathcal{O}(NM)$ operations [MS77]. Note that the representation $\gamma(i) = \mathbf{t}^T\mathbf{h}(i)$ gives the value $\gamma(i)$ in terms of the "bits" $\{y[n]\}$ of the index i. This property has proven to be of great use in synthesis and analysis of quantizers [HKS95]. In the application studied here, the finite-domain function of interest is the estimator $\hat{f}(i)$, and the binary variables $\{y[n]\}$ are the one-bit quantized data samples (4.4). By using (4.8) the Hadamard transform can be employed to represent this estimator as

$$\begin{aligned}\hat{f}(i) = \mathbf{t}^{\mathrm{T}}\mathbf{h}(i) = \sum_{m=0}^{M-1} t_m h_m(i) = t_0 + t_1\,y[0] + t_2\,y[1] \\ + t_3\,y[0]y[1] + \cdots + t_{M-1}\prod_{n=0}^{N-1} y[n].\end{aligned} \tag{4.10}$$

That is, $\hat{f}$ can be represented in terms of the transform coefficients $\{t_m\}$ and *all possible different products that can be formed using the variables* $\{y[n]\}$. For a given $\hat{f}(i)$ the coefficients $\{t_m\}$ (the t-coefficients, for short) are calculated via the Hadamard transform. It is important to note that the representation (4.10) is exact.

Noting that the estimator $\hat{f}$ is completely defined by the t-coefficients it is possible to use (4.10) as a basis for reducing the number of parameters needed in implementing the estimator. However, since there are M different t_m nothing is gained by using (4.10) to implement the estimator (on the contrary there is a loss in computational complexity since the sum in (4.10) needs to be calculated, while a table look-up implementation based on (4.6) basically requires no computation at all). It is reasonable, however, to assume that not all of the t-coefficients are significant (in the sense that some of them are zero or close-to zero). Hence, if the t-coefficients that are most significant can be identified, only these have to be stored (setting "insignificant" coefficients to zero) to compute an approximate estimate using (4.10). Compared with using a table look-up implementation such an approach can be used to trade storage complexity for computations.

4.3 Table Analysis

Consider a known table used in a table look-up frequency estimator (4.6), say $\mathbf{g}$. That is

$$\mathbf{g} = \left[\hat{f}(0), \hat{f}(1), \ldots, \hat{f}(M-1)\right]^{\mathrm{T}}. \tag{4.11}$$

The table entries in $\mathbf{g}$ can be expressed as a function of the corresponding t-coefficients and a binary representation of the entry index, as in (4.10). To illustrate the structure of the t-coefficients a table (4.11) trained at SNR $= A^2/(2\sigma^2) = 20$ dB and for a block-length $N = 16$ is used, according to [ASH00]. The t-coefficients for this table are computed, as in (4.9), and their normalized magnitudes $|t_i|/t_0$ are displayed in Figure 4.1. Note that there exist coefficients that are significantly larger in magnitude than the rest (marked in Figure 4.1 above the dashed line). In further analyzing the t-coefficients one can note that all the dominant t-coefficients correspond to a weight two product in the sum (4.10), i.e t_3 is multiplied with the product $y[0]y[1]$ and t_6 is multiplied with $y[1]y[2]$ and so forth. From this, the dominant t-coefficients can be divided into two sets:

- $\mathcal{A}$) t-coefficients that correspond to a weight two product of neighboring samples. For example correspond the coefficient t_{12} to the product $y[2]y[3]$ and the coefficient t_{24} correspond to the product $y[3]y[4]$.
- $\mathcal{B}$) t-coefficients that correspond to a weight two product of samples separated by a distance of an even number of samples. The set $\mathcal{B}$ is exemplified by t_9 corresponding to the binary product $y[0]y[3]$, or t_{33} corresponding to $y[0]y[5]$.

The coefficient t_0 is included in both sets. Neighboring samples are separated by a zero distance, hence set $\mathcal{A}$ is a subset of $\mathcal{B}$. Using one of the sets $\mathcal{A}$ or $\mathcal{B}$ an approximation of each entry in the true $\mathbf{g}$ can be formed. These entries form an estimate of the table $\hat{\mathbf{g}}$. By calculating an entry estimate only when needed, it is possible to reduce the memory complexity since fewer coefficients need to be stored. Accordingly, the memory complexity is reduced from storing the entire table with 2^N coefficients to N or $N^2/4 + 1$ using set $\mathcal{A}$ or $\mathcal{B}$, respectively. That is, a reduction from an exponential to a polynomial relation between the block length and the number of coefficients. A block diagram of a type-$\mathcal{A}$ estimator is given in Figure 4.2.

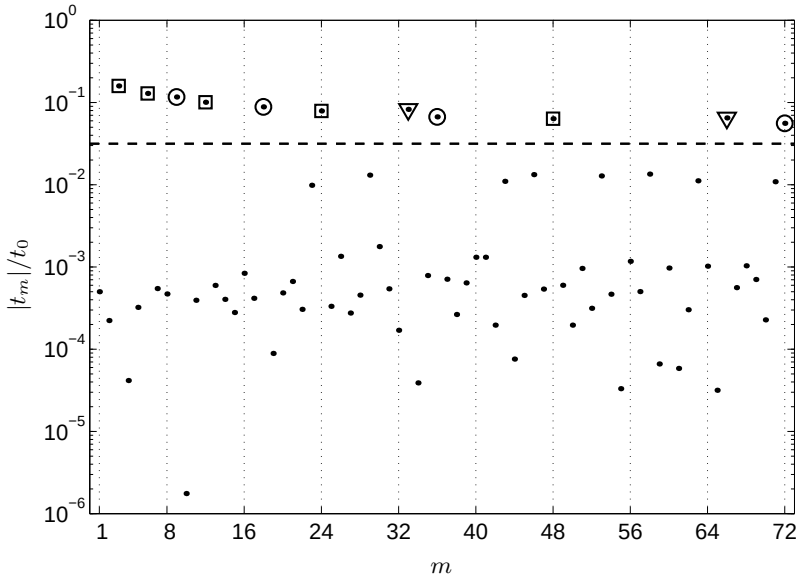


Figure 4.1: The normalized magnitudes of the first 72 t-coefficients in (4.10) for a fixed table of size $M = 2^{16}$. The coefficients above the dashed line correspond to weight two binary products of neighboring samples ($\square$), samples at distance 3 ($\circ$) and 5 (∇), respectively.

4.4 Estimator Design

In was shown above how to form an approximation of each table entry using a reduced set of t-coefficients. Calculating the entire set of t-coefficients requires storage of the full table $\mathbf{g}$. This is not feasible for, say, $N > 26$. The structure of the approximate estimator is, however, independent of N. Here, we use the structure of such an estimator and calculate the corresponding reduced set of coefficients under the MMSE criterion.

Let $\tilde{\mathbf{h}}_{\mathcal{A}}(i)$ and $\tilde{\mathbf{h}}_{\mathcal{B}}(i)$ denote vectors containing the signal products in (4.10) corresponding to the t-coefficients in the sets $\mathcal{A}$ and $\mathcal{B}$, respectively.

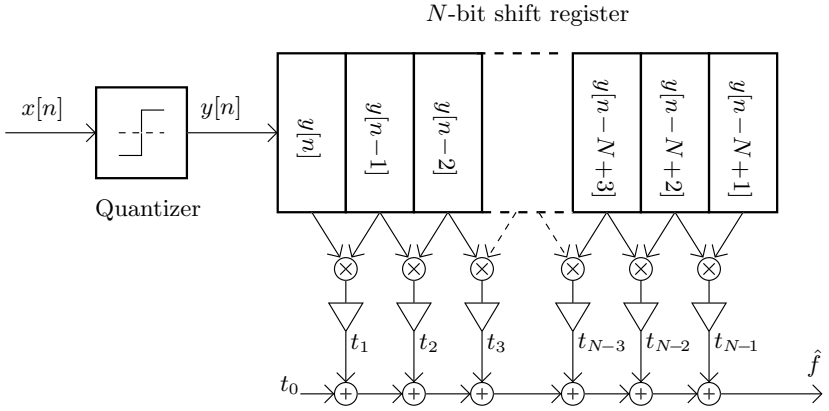


Figure 4.2: A proposed estimator where neighboring binary products of weight two are used (type-$\mathcal{A}$).

That is,

$$\tilde{\mathbf{h}}_{\mathcal{A}}(i) = \begin{bmatrix} 1 \\ y[0]y[1] \\ y[1]y[2] \\ \vdots \end{bmatrix}, \qquad \tilde{\mathbf{h}}_{\mathcal{B}}(i) = \begin{bmatrix} 1 \\ y[0]y[1] \\ y[1]y[2] \\ y[0]y[3] \\ y[2]y[3] \\ y[1]y[4] \\ y[3]y[4] \\ \vdots \end{bmatrix} \tag{4.12}$$

where the relation between the index i and the sequence $y[n]$ is given by (4.5). We denote the corresponding vectors with t-coefficients by $\tilde{\mathbf{t}}_{\mathcal{A}}$ and $\tilde{\mathbf{t}}_{\mathcal{B}}$, respectively. We can now formulate two corresponding frequency estimators as

$$\hat{f}_{\mathcal{A}}(i) = \tilde{\mathbf{t}}_{\mathcal{A}}^{\mathrm{T}} \tilde{\mathbf{h}}_{\mathcal{A}}(i), \tag{4.13}$$

$$\hat{f}_{\mathcal{B}}(i) = \tilde{\mathbf{t}}_{\mathcal{B}}^{\mathrm{T}} \tilde{\mathbf{h}}_{\mathcal{B}}(i). \tag{4.14}$$

In order to optimize the performance of the estimators in (4.13),(4.14) let $\tilde{\mathbf{t}}_k$ be a design parameter to be chosen optimally. Using the MMSE

criterion $\tilde{\mathbf{t}}_k$ is given by

$$\begin{aligned}\tilde{\mathbf{t}}_k &= \arg\min_{\mathbf{a}} \mathrm{E}(f - \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{h}}_k(i))^2 \\ &= (\mathrm{E}[\tilde{\mathbf{h}}_k(i)\tilde{\mathbf{h}}_k(i)^{\mathrm{T}}])^{-1}\ \mathrm{E}[\tilde{\mathbf{h}}_k(i)f] \qquad k = \mathcal{A}, \mathcal{B}\end{aligned} \tag{4.15}$$

where the expectation is with respect to frequency f, phase ϕ and noise $e[n]$.

A feasible approach to calculate the expectations needed in (4.15) is by aid of Monte Carlo integration. Such a training procedure for the problem at hand is discussed in Chapter 2.

4.5 Relations to the MLE

The relation between the estimators obtained by the present approach and the method of maximum likelihood (MLE) for unquantized data is considered in this section. A natural approach is to employ the MLE, or approximations thereof, directly to one-bit quantized data and hope that it will provide accurate estimates also in the one-bit quantized case.

The MLE is given by a non-linear least-squares fit of the sinusoidal model (4.2) to the observed data $\{x[n]\}$ in (4.1). In order to continue, we consider the Hilbert transformed counterpart to (4.2), that is

$$z[n] = A\exp(j\phi)\exp(j2\pi f) + e[n] \tag{4.16}$$

where A is the real valued amplitude, ϕ is the initial phase and $e[n]$ is complex-valued circular Gaussian noise with variance $2\sigma^2$. Tretter showed in [Tre85] that for high SNR the additive noise can be written as a phase noise, and that the approximation of MLE is obtained solving a (linear) least squares problem. The difficulty with this estimator is that the phase sequence needs to be unwrapped. A remedy to this problem was given in [Kay89] where the estimator was rewritten in differenced phase data. The latter estimator is known as the weighted phase averager (WPA). Normalizing the differenced phase data $\angle\{z[n]z^*[n-1]\}$ by 2π (where * denotes conjugate) implies that the MLE is well approximated by

$$\hat{f} = \sum_{n=1}^{N-1} w_n f_n \qquad w_n = \frac{6n(N-n)}{N(N^2-1)} \tag{4.17}$$

where $f_n = \angle\{z[n]z^*[n-1]\}/2\pi$. The scalar f_n is a two-points estimate of the sought frequency. Thus, a high-SNR approximation of the MLE of

frequency is given by a parabolicaly weighted sum of two-points estimates of the sought frequency.

In the considered case in this chapter, the signal model is real-valued. A key motivation for using one-bit processing is the low complexity on the digital side, and accordingly we also strive to minimize the analog preprocessing of data. Thus, we consider a one-bit implementation employing observations (4.4) only.

The observed data (4.4) spans the measurement interval $(N-1)/f_s$. For simplicity, consider a time interval that covers an integer number of periods of $s[n]$ in (4.2), Then, the measurement interval can be expressed in the number of zero-crossings of $s[n]$ as $p/2f_s$, where p is the number of zero crossings. Accordingly, the sought frequency can (as $\sigma^2 \to 0$) be expressed as

$$f_0 = \frac{p}{2}(N-1) \tag{4.18}$$

With the unipodal representation of $y[n]$, the number of zero crossings can be expressed as

$$p = \sum_{n=1}^{N-1} \frac{1 - y[n]y[n-1]}{2} \triangleq \sum_{n=1}^{N-1} p_n \tag{4.19}$$

where p_n is a two-points estimate of the fractional number of zero crossing between two consecutive samples. Accordingly, a two-points estimate of the sought frequency is given by

$$f_n = \frac{p_n}{2} \tag{4.20}$$

Thus, a one-bit implementation of the estimator (4.17) is

$$\hat{f} = \sum_{n=1}^{N-1} w_n \frac{p_n}{2} = \frac{1}{4} \sum_{n=1}^{N-1} w_n \left(1 - y[n]y[n-1]\right). \tag{4.21}$$

By observing the similarity between (4.13) and (4.21) it is possible to form an estimator with same structure as (4.13) but with an alternative derivation. The estimator $\hat{f}_{\mathrm{MLE}}(i)$ follows from (4.21) by a straightforward calculation, that is

$$\hat{f}_{\mathrm{MLE}}(i) \approx \hat{f} = \frac{1}{4} + \sum_{n=0}^{N-2} b_{n+1}\, y[n]y[n+1] \tag{4.22}$$

where $b_n = -w_n/4$. The estimator (4.22) is a high SNR ($\sigma^2 \to 0$) approximation of the MLE for unquantized data applied to binary measurements.

At high SNR, it is possible to interchange the phase angle calculation and the summation in the estimator (4.17). Then we end up with an estimator known as the weighted linear predictor [Kay89]. For finite, but high SNR the weights $\{b_n\}$ can be optimized depending calculated as shown below. In [CKQ94], an SNR dependent optimal weighting scheme was derived. After appropriate scaling, we have that for finite (but high) SNR the coefficients b_n in (4.22) can be approximated by [CKQ94]

$$b_n = -\frac{\sinh(\theta/2)}{4} \times \frac{\sinh(N\theta) - \sinh(n\theta) - \sinh(\{N-n\}\theta)}{(N-1)\sinh(\theta/2)\sinh(N\theta) - 2\sinh(N\theta/2)\sinh(\{N-1\}\theta/2)} \tag{4.23}$$

where $\theta = \log \alpha$ with $\alpha = 1 + \beta/2 + (\beta^2/4 + \beta)^{1/2}$ and the scalar β given by $\beta = 1/\text{SNR}$.

In conclusion, note the structural similarity between (4.22) and the estimator in (4.13). Figure 4.3 displays a comparison between the coefficients obtained from (4.15) versus the coefficients calculated using the MLE approach (4.23). The similarity between the $\tilde{\mathbf{t}}_{\mathcal{A}}$ and the b_n coefficients are convincing. From formula (4.23) the b_n coefficients are formed like a second order parabola while the $\tilde{\mathbf{t}}_{\mathcal{A}}$ has a higher order parabolic shape. This difference is possible due to approximations in the derivation of (4.23). The $\tilde{\mathbf{t}}_{\mathcal{A}}$ coefficients are, in fact optimized for this specific scenario without any approximations. A drawback is though, that the $\tilde{\mathbf{t}}_{\mathcal{A}}$ coefficients need to be re-optimized if the preferences are to be changed. However, the derivation of the $\tilde{\mathbf{t}}_{\mathcal{A}}$ coefficients are not only restricted to a frequency estimation scenario, but could be used in any scenario where an estimator is needed. Though not verified in simulations, the two frequency estimators using the sets $\tilde{\mathbf{t}}_{\mathcal{A}}$ and b_n of coefficients are expected to perform equally well.

4.6 Numerical Evaluation

In Figure 4.4, the empirical mean square error (MSE) is shown as function of SNR for a data record of length $N = 16$. The performance of the estimator using the full table $\mathbf{g}$ in (4.11) is compared with using subsets of parameters, that is type-$\mathcal{A}$ and type-$\mathcal{B}$ in (4.13) and (4.14), respectively. As reference, the asymptotic ($N \to \infty$) CRB for the given signal model

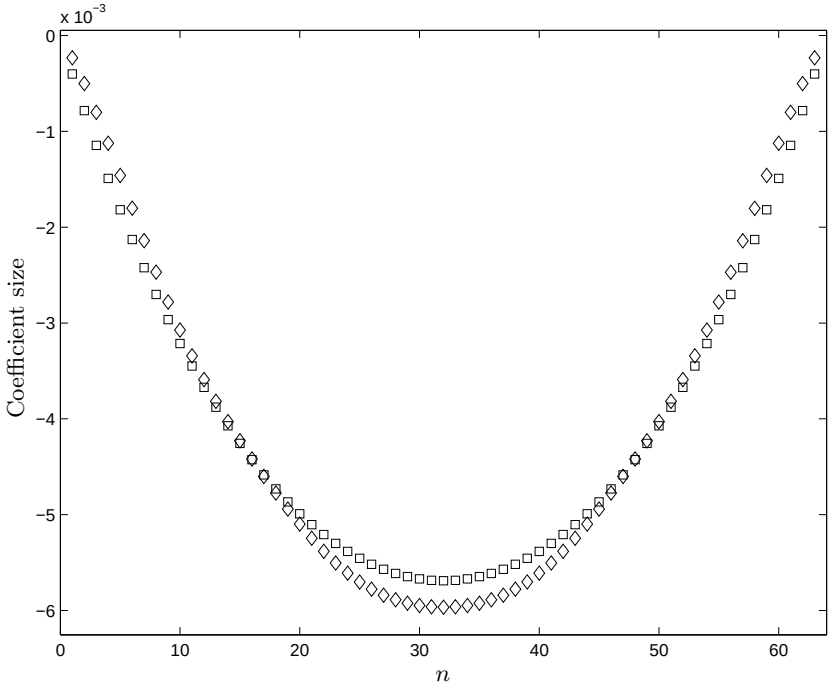


Figure 4.3: Comparison of t-coefficients $\tilde{\mathbf{t}}_{\mathcal{A}}$ ($\Diamond$) (4.15), and b_n ($\square$) (4.23). In both cases, evaluation has been performed using $N = 64$ and SNR = 20dB.

is included [HMH00]. The table $\mathbf{g}$ in (4.11) is obtained using a training approach discussed in [ASH00]. The t-coefficients $\tilde{\mathbf{t}}_{\mathcal{A}}$ and $\tilde{\mathbf{t}}_{\mathcal{B}}$ for $\hat{f}_{\mathcal{A}}(i)$ and $\hat{f}_{\mathcal{B}}(i)$ are calculated according to (4.15) using Monte Carlo integration at SNR = 20 dB. The *a priori* distribution of f is chosen as a uniform distribution on the interval $[\varepsilon, 0.5 - \varepsilon]$ where ε is a design parameter and has been set to $\varepsilon = 0.04$. Our experience indicates that a smaller value of ε typically results in a significant performance reduction while a larger value does not appear to influence the performance negatively. In Figure 4.4 (as well as in Figure 4.5), the performance is evaluated for the signal in (4.2) with the true frequency $f_0 = 0.1$. Further the MSE figures are averaged over 100.000 independent trials. From Figure 4.4, we note a decreased performance when the complexity of the estimator is reduced.

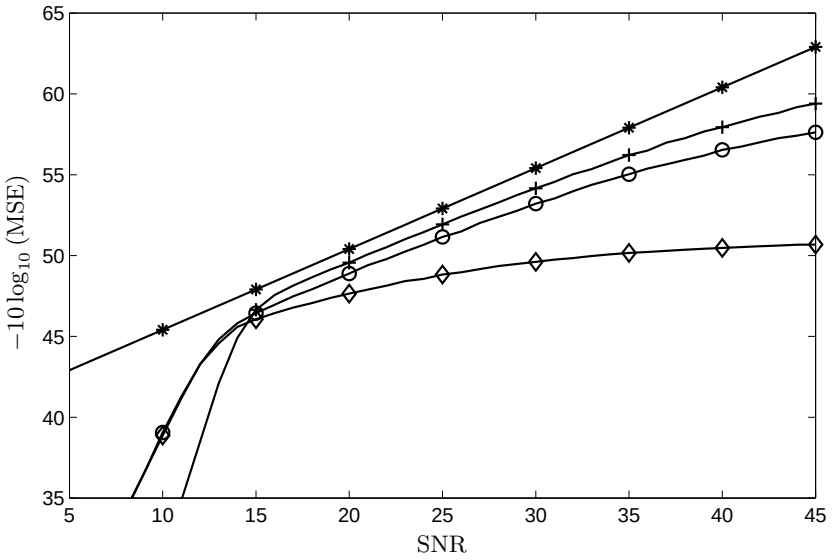


Figure 4.4: Performance of the proposed estimators for $N=16$. Displayed are: CRB ($*$), table look-up estimator (4.6) ($+$), estimator $\hat{f}_{\mathcal{A}}(i)$ ($\Diamond$) and $\hat{f}_{\mathcal{B}}(i)$ ($\circ$).

We observe further that for high SNRs the performance of (4.11) starts to deviate from the CRB due to a non-negligible bias term in the MSE. For $\tilde{f}_{\mathcal{A}}(i)$ and $\tilde{f}_{\mathcal{B}}(i)$ the bias is even more significant.

The experiment is repeated in Figure 4.5, but now for $N = 64$. In this case, it is not feasible to implement (4.11) and it is therefore excluded from the comparison. From the figure, we note that the performance of $\hat{f}_{\mathcal{B}}(i)$ almost coincides with the asymptotic CRB for all SNRs above a threshold at about 15 dB. We also note that the difference in performance between $\hat{f}_{\mathcal{A}}(i)$ and $\hat{f}_{\mathcal{B}}(i)$ is negligible for low SNRs. At high SNRs the difference is more significant. In Figure 4.6, the empirical MSE is shown as a function of the unknown signal frequency f_0 at a fixed SNR = 20 dB and block length $N = 64$. As a reference the asymptotic CRB is displayed. We observe that both the estimators, $\hat{f}_{\mathcal{A}}(i)$ and $\hat{f}_{\mathcal{B}}(i)$ performs well, except at frequencies near 0 or 0.5, and that the difference in performance between them is negligible. However, the performance

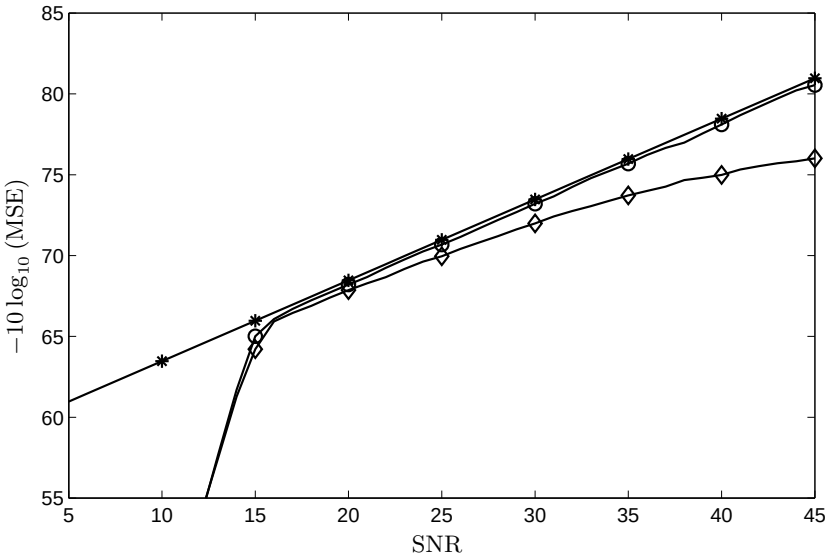


Figure 4.5: Performance of the proposed estimators for $N=64$. Displayed are: CRB ($*$), estimator $\hat{f}_{\mathcal{A}}(i)$ ($\Diamond$) and $\hat{f}_{\mathcal{B}}(i)$ ($\circ$).

is dependent on the unknown signal frequency f_0 and for some isolated frequencies the performance is significantly deteriorated.

4.7 Summary and Conclusions

We have shown that the table based approach used as a frequency estimator in Chapter 2 can be transformed using the Hadamard transform to an equivalent representation based on a sum over binary products and a set of coefficients. We have also investigated how the set of coefficients can be reduced, and how such reduction makes it possible to handle large blocks of data. We furthermore showed how the remaining coefficients can be optimized to increase the performance of the estimator with reduced complexity. An alternative derivation of a coefficient set using the MLE frequency estimator were also performed. The performance of the new estimators was then evaluated by aid of simulations and their performance was compared with the appropriate Cramér–Rao bound. The

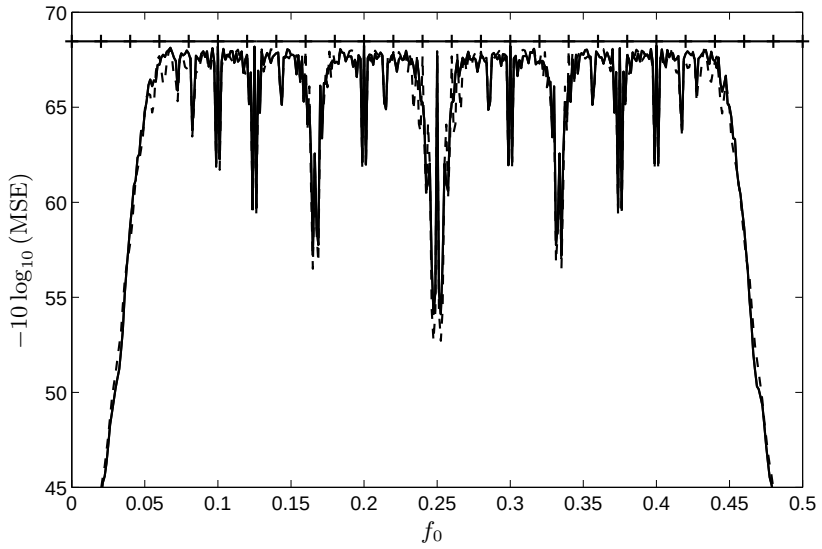


Figure 4.6: Empirical MSE as a function of the frequency f for $N=64$ and SNR = 20 dB. Displayed are: CRB (+), $\hat{f}_{\mathcal{A}}(i)$ (- -) and $\hat{f}_{\mathcal{B}}(i)$ (–).

simulations indicated that the considered methods are able to produce nearly statistically efficient estimates.

Part II

Waveform Fitting

Chapter 5

Model Order Selection in Standardized Sinewave Fitting

An important problem in many applications is to fit a sinusoidal model to recorded data, as well as to calculate the parameters that in least-squares result in the best fit. Algorithms have been standardized in IEEE Standard 1057 and IEEE Standard 1241 [S1057, S1241]. For easy reference, the three- and four-parameter sinewave fit algorithms of [S1241, Sect. 4.1.4] are hereafter denoted as the three- and four-parameter algorithm, respectively. Software implementations of the four-parameter algorithm can be found in [MK01, Bla99], and investigations of its performance is the main topic in [ASDM01, BMS+03, Hän00]. In this chapter the performance of the standardized three- and four-parameter methods is studied.

For example when testing waveform recorder or ADCs it is known that the three- and four parameter algorithm produce biased estimates, due to harmonic distortion originating from either non-coherent sampling or nonlinearities [PS96, HP03]. Parametric modeling of the harmonic distortion is the basis for the improved waveform fits in [PS96]. Modifications of the three- and four-parameter algorithms are beyond the scope of the chapter, and the analysis is based on an assumption of unbiased estimation.

The aim of this chapter is to investigate some fundamental properties of the three- and four parameter algorithms and answer the question

whether to use the former or latter for a given scenario. Under the Gaussian assumption a thorough answer is provided, both theoretically and by aid of extensive computer simulations. It is further shown that the derived results are applicable under the ideal noise model of quantization, described by a uniform probability density function. Finally, the applicability of the result of fitting sine waves in ADC testing is briefly discussed, that is in scenarios where the quantization noise is clearly not uniform neither independent of the sinewave [KB04].

The remaining part of this chapter is organized as follows. In Section 5.1, the signal model is introduced and the non-linear least squares for parameter estimation is reviewed. In Section 5.2, the asymptotic Cramér-Rao bound (CRB) using the model (5.1) is derived for three and four unknown parameters, respectively. In section 5.3, the parsimony principle is used in order to derive an expression on the mean-squared-error. When a three parameter model is employed the mean-squared-error is frequency dependent, and this dependency is studied in some detail. A numerical example illustrating the parsimony principle, as well as a discussion of the impact of the presented theory in a practical estimation scenario and finally, the uniform model of quantization, are gathered in Section 5.4. Conclusions that are drawn are presented in Section 5.5.

5.1 Signal Model and Non-linear Least Squares

Assume that the data vector (a column vector) contains the sequence of measurement samples $\mathbf{x} = [x_1 \dots x_N]^T$ taken at time instants $\{t_1, \dots, t_N\}$. It is further assumed that data can be modeled by

$$s_n(\boldsymbol{\vartheta}) = A\cos(\omega t_n) + B\sin(\omega t_n) + C \tag{5.1}$$

where A, B, C and ω are (known or unknown) constants. The angular frequency ω is related to the signal frequency by $\omega = 2\pi f$, where f is the signal frequency in Hertz. For short, ω is from here on called frequency. Stressing the dependence of $s_n(\boldsymbol{\vartheta})$ on the *generic parameter vector* $\boldsymbol{\vartheta}$ turns out to be convenient for the following discussion, where the unknown parameters are gathered in $\boldsymbol{\vartheta}$. Throughout this chapter, the vector $\boldsymbol{\vartheta}$ represents either the set of three parameters (A, B, C), or a set of four parameters, depending on if the frequency ω is known or not. The sinewave fit problem is solved by minimizing the sum-squared-

error [S1057], [S1241]

$$V(\boldsymbol{\vartheta}) = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - s_n(\boldsymbol{\vartheta})\right)^2 \tag{5.2}$$

with respect to the unknown parameters $\boldsymbol{\vartheta}$. Consider a signal model where the measurements are described by

$$x_n = s_n(\boldsymbol{\vartheta}) + w_n \tag{5.3}$$

where x_n is the observation and $s_n(\boldsymbol{\vartheta})$ the underlying sinewave (5.1) described by the parameter vector $\boldsymbol{\vartheta}$. The process w_n describes the modeling error, noise, etc, and is assumed to be a zero-mean white Gaussian stochastic process with variance σ^2. The Gaussian assumption is crucial to the validity of the theoretical results derived in this chapter and may seem restrictive. However, in many practical cases where the measurement noise is independent of the signal s_n, the Gaussian assumption is a reasonable approximation, and the derived results are (at least approximately) valid. One such scenario is estimation of the frequencies of sinewaves from quantized measurements where the measurement noise is significantly larger than the quantization noise [AH05a]. A numerical example for this scenario is included in Section 5.4.2. By aid of numerical simulations, it is shown that the derived results are applicable under the uniform noise model of quantization. On the other hand, in ADC-testing the Gaussian assumption is not valid, and the quantization noise is neither Gaussian nor independent of the sinewave [KB04]. In fact, the ideal quantization noise is deterministic. Accordingly, the derived results are not directly applicable in testing ADCs.

Consider the particular parameter vector $\boldsymbol{\vartheta} = \boldsymbol{\theta}$, where

$$\boldsymbol{\theta} = [\theta^T \ \omega]^T \tag{5.4}$$

and

$$\theta = \begin{bmatrix} A & B & C \end{bmatrix}^T. \tag{5.5}$$

Let $\mathbf{D}(\omega)$ be the $N \times 3$ matrix

$$\mathbf{D}(\omega) = \begin{bmatrix} \cos\omega t_1 & \sin\omega t_1 & 1 \\ \vdots & \vdots & \vdots \\ \cos\omega t_N & \sin\omega t_N & 1 \end{bmatrix}. \tag{5.6}$$

Then, the sum-squared-error (5.2) can be written as

$$V(\boldsymbol{\theta}) = V(\omega, \theta) = \frac{1}{N} \left(\mathbf{x} - \mathbf{D}(\omega)\, \theta\right)^T \left(\mathbf{x} - \mathbf{D}(\omega)\, \theta\right). \tag{5.7}$$

When the frequency ω is known, (5.7) is minimized in the least-squares sense by solving the set of linear equations $\mathbf{D}(\omega)\theta = \mathbf{x}$ [Kay93]. If $\mathbf{D}(\omega)$ has full column rank the solution is given by

$$\hat{\theta} = \left(\mathbf{D}(\omega)^T \mathbf{D}(\omega)\right)^{-1} \mathbf{D}(\omega)^T \mathbf{x}. \tag{5.8}$$

One should not that a direct calculation of (5.8) may be numerically imprecise and from an implementation point of view it is recommended to use some matrix factorization algorithm [GvL96]. When the frequency is unknown, the criterion (5.7) can be concentrated with respect to θ by plugging in the least-squares solution (5.8) into (5.7) [Kay93]. Thus,

$$V(\omega) = \frac{1}{N} \left(\mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{\Pi}(\omega) \mathbf{x}\right) \tag{5.9}$$

where $\Pi(\omega)$ is the projection matrix

$$\mathbf{\Pi}(\omega) = \mathbf{D}\left(\omega\right)(\mathbf{D}(\omega)^T\, \mathbf{D}(\omega))^{-1}\, \mathbf{D}(\omega)^T. \tag{5.10}$$

It is straightforward to show that ω can be found by a one-dimensional search for the maximum of [Kay93]

$$g(\omega) = \mathbf{x}^T\, \mathbf{D}(\omega)\, \left(\mathbf{D}(\omega)^T\, \mathbf{D}(\omega)\right)^{-1} \mathbf{D}(\omega)^T\, \mathbf{x} \tag{5.11}$$

The dependency of (5.11) on ω is non-trivial. Although, efficient algorithms exist for this class of non-linear least-squares problems. One may note that for large N (that is, as $N \to \infty$) the columns in $\mathbf{D}(\omega)$ become orthogonal. Thus, $\mathbf{D}(\omega)^T\, \mathbf{D}(\omega)$ becomes a diagonal matrix with elements $[N/2 \quad N/2 \quad N]$ in the main diagonal. Accordingly, $g(\omega)$ in (5.11) is well approximated by $g_\infty(\omega)$, where

$$\begin{aligned} g_\infty(\omega) &= \frac{2}{N} \left(\sum_{n=1}^{N} \cos(\omega t_n) x_n\right)^2 + \frac{2}{N} \left(\sum_{n=1}^{N} \sin(\omega t_n) x_n\right)^2 + \frac{1}{N} \left(\sum_{n=1}^{N} x_n\right)^2 \\ &= 2P(\omega) + \frac{1}{N} \left(\sum_{n=1}^{N} x_n\right)^2. \end{aligned} \tag{5.12}$$

The definition of the periodogram $P(\omega)$ in (5.12) follows the standard literature [Kay93]

$$P(\omega) = \frac{1}{N}\left|\sum_{n=1}^{N} x_n \exp(-j\omega t_n)\right|^2 . \tag{5.13}$$

Once (5.11) has been maximized and the corresponding argument (say $\hat{\omega}$) has been determined, the unknowns in θ are obtained by a least-squares fit (5.8), replacing ω in (5.8) with $\hat{\omega}$.

5.2 Cramér-Rao Bound

A lower bound on the accuracy (covariance) of any unbiased estimator is given by the CRB, that is $\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) \geq \mathrm{CRB}(\boldsymbol{\vartheta})$ where $\geq$ is to be interpreted as that the difference $\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) - \mathrm{CRB}(\boldsymbol{\vartheta})$ is positive semidefinite. The CRB is given by the inverse of the Fisher information matrix $\mathbf{J}(\boldsymbol{\vartheta})$, that is $\mathrm{CRB}(\boldsymbol{\vartheta}) = \mathbf{J}(\boldsymbol{\vartheta})^{-1}$. The Fisher information matrix is given by [Kay93]

$$\mathbf{J}(\boldsymbol{\vartheta}) = \mathrm{E}\left\{\left(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)\left(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)^T\right\} \tag{5.14}$$

where $p(\mathbf{x};\boldsymbol{\vartheta})$ denotes the probability density function (pdf), and where the derivative is evaluated at the true parameters. By the Gaussian assumption on w_n, the pdf for $\mathbf{x}$ is given by

$$p(\mathbf{x};\boldsymbol{\vartheta}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})\right)^T\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})\right)\right]. \tag{5.15}$$

The derivative of a scalar function $y(\boldsymbol{\vartheta})$ with respect to the vector $\boldsymbol{\vartheta}$ is defined as the vector

$$\frac{\partial y(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \begin{bmatrix} \dfrac{\partial y(\boldsymbol{\vartheta})}{\partial \vartheta_1} \\ \vdots \\ \dfrac{\partial y(\boldsymbol{\vartheta})}{\partial \vartheta_p} \end{bmatrix} \tag{5.16}$$

where $p = \dim(\boldsymbol{\vartheta})$ and $\{\vartheta_1 \ldots \vartheta_p\}$ are the elements in $\boldsymbol{\vartheta}$. Now, calculating (5.14) using (5.15)-(5.16) yields

$$\mathbf{J}(\boldsymbol{\vartheta}) = \frac{1}{\sigma^2}\sum_{n=1}^{N} \boldsymbol{\psi}_n \boldsymbol{\psi}_n^T \tag{5.17}$$

where,

$$\boldsymbol{\psi}_n = \frac{\partial s_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \begin{bmatrix} \dfrac{\partial s_n(\boldsymbol{\vartheta})}{\partial \vartheta_1} \\ \vdots \\ \dfrac{\partial s_n(\boldsymbol{\vartheta})}{\partial \vartheta_p} \end{bmatrix}. \tag{5.18}$$

A derivation of (5.17) can be found in Appendix 5.A. In the forthcoming discussion the following formula describing the elements of the Fisher information matrix $J(\boldsymbol{\vartheta})$ is useful.

$$[\mathbf{J}(\boldsymbol{\vartheta})]_{\ell,r} = \frac{1}{\sigma^2} \sum_{n=1}^{N} [\boldsymbol{\psi}_n]_\ell [\boldsymbol{\psi}_n]_r. \tag{5.19}$$

In (5.19), $[\cdot]_{\ell,r}$ denotes the ℓ, r-th element of the matrix within the parentheses and $[\cdot]_\ell$ denotes the ℓ-th element in the vector. With $\boldsymbol{\vartheta} = \boldsymbol{\theta}$ according to (5.4), we have to calculate

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} J_{AA} & J_{AB} & J_{AC} & J_{A\omega} \\ J_{AB} & J_{BB} & J_{BC} & J_{B\omega} \\ J_{AC} & J_{BC} & J_{CC} & J_{C\omega} \\ J_{A\omega} & J_{B\omega} & J_{C\omega} & J_{\omega\omega} \end{bmatrix} \tag{5.20}$$

where $J_{AA} = [\mathbf{J}(\boldsymbol{\theta})]_{1,1}$ according to (5.19), etc. The derivatives in (5.19) are straightforward to calculate and can be found in [Hän00]. The expressions derived in [Hän00] are suitable for numerical evaluation of the information matrix, and the CRB follows by numerical inversion of the result; See [Hän00] for examples on the performance of the frequency estimate using IEEE standard 1057 [S1057] compared with the exact CRB. The results in [Hän00] are a straightforward generalization of the well-known results in [RB74]; The signal model is extended to include the DC-level (the C parameter).

5.2.1 Asymptotic CRB

In order to get analytical insight, we derive an asymptotic expression valid for large N (as $N \to \infty$) and uniform sampling at f_s Hertz, that is

$$t_n = \frac{n}{f_s}, \quad n = 1, \ldots, N. \tag{5.21}$$

Without loss of generality f_s has from here on in the chapter been set to $f_s = 1\text{Hz}$. Starting from the results in [Hän00], straightforward calculations of (5.20) yield

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{2\sigma^2}\left(\mathbf{J}_1 + \mathbf{J}_2\right) \tag{5.22}$$

where

$$\mathbf{J}_1 = \begin{bmatrix} N & 0 & 0 & -\frac{BN^2}{2} \\ 0 & N & 0 & \frac{AN^2}{2} \\ 0 & 0 & 2N & 0 \\ -\frac{BN^2}{2} & \frac{AN^2}{2} & 0 & \frac{(A^2+B^2)N^3}{3} \end{bmatrix} \tag{5.23}$$

and

$$\mathbf{J}_2 = \begin{bmatrix} \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(N) \\ \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(N) \\ \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(N) \\ \mathcal{O}(N) & \mathcal{O}(N) & \mathcal{O}(N) & \mathcal{O}(N^2) \end{bmatrix}. \tag{5.24}$$

In (5.24), $\mathcal{O}(x)$ denotes a quantity that is asymptotically linear in x, that is $\lim_{x\to\infty} |\mathcal{O}(x)/x| = c$ where $0 < c < \infty$, and $\mathcal{O}(1)$ denotes a bounded quantity. Now, the CRB yields

$$\text{CRB}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})^{-1} = 2\sigma^2\left(\mathbf{J}_1 + \mathbf{J}_2\right)^{-1} \simeq 2\sigma^2\mathbf{J}_1^{-1} \tag{5.25}$$

where the second equality follows from (5.22), and $\simeq$ denotes an equality where only the dominant terms have been retained. A proof of the last equality for a related signal model can be found in Appendix B of [NP86]. Modifications needed for the model under study are straightforward.

5.2.2 Three Parameter Model

For known frequency, that is $\boldsymbol{\vartheta} = \theta = [A \quad B \quad C]^T$ according to (5.5), the corresponding information matrix is given by the the upper left 3×3 sub-matrix of (5.20). For large N the sub-matrix is diagonal, c.f., (5.23), and inversion of it is straightforward. The diagonal elements of the inverse yield the lower bound on the variance of the estimates, that is

$$\text{var}(\hat{A}) \quad \geq \quad \text{CRB}(A) \simeq \frac{2\sigma^2}{N} \tag{5.26}$$

$$\text{var}(\hat{B}) \quad \geq \quad \text{CRB}(B) \simeq \frac{2\sigma^2}{N} \tag{5.27}$$

$$\text{var}(\hat{C}) \quad \geq \quad \text{CRB}(C) \simeq \frac{\sigma^2}{N} \tag{5.28}$$

Since the upper left 3×3 sub-matrix of $\mathbf{J}_1$ in (5.23) is diagonal, the lower bound on estimation accuracy of A and B does not depend if the offset is assumed known or is estimated. Accordingly, a two parameter model with only A and B as unknown parameters results in the bounds (5.26)-(5.27).

5.2.3 Four Parameter Model

A four parameter model corresponds to $\boldsymbol{\theta} = [\theta^T \quad \omega]^T$ as in (5.4). Straightforward calculations of the inverse of (5.23) yield

$$\mathbf{J}_1^{-1} = \begin{bmatrix} \frac{1}{N}\left(1+\frac{3B^2}{\alpha^2}\right) & -\frac{3AB}{\alpha^2 N} & 0 & \frac{6B}{\alpha^2 N^2} \\ -\frac{3AB}{\alpha^2 N} & \frac{1}{N}\left(1+\frac{3A^2}{\alpha^2}\right) & 0 & \frac{6A}{\alpha^2 N^2} \\ 0 & 0 & \frac{1}{2N} & 0 \\ \frac{6B}{\alpha^2 N^2} & -\frac{6A}{\alpha^2 N^2} & 0 & \frac{12}{\alpha^2 N^3} \end{bmatrix} \tag{5.29}$$

where $\alpha^2 = A^2 + B^2$, that is the signal amplitude squared. Accordingly, the CRB on the parameters yields

$$\text{var}(\hat{A}) \quad \geq \quad \text{CRB}(A) \simeq \frac{2\sigma^2}{N}\left(1+\frac{3B^2}{\alpha^2}\right) \tag{5.30}$$

$$\text{var}(\hat{B}) \quad \geq \quad \text{CRB}(B) \simeq \frac{2\sigma^2}{N}\left(1+\frac{3A^2}{\alpha^2}\right) \tag{5.31}$$

$$\text{var}(\hat{C}) \quad \geq \quad \text{CRB}(C) \simeq \frac{\sigma^2}{N} \tag{5.32}$$

$$\text{var}(\hat{\omega}) \quad \geq \quad \text{CRB}(\omega) \simeq \frac{24\sigma^2}{\alpha^2 N^3} \tag{5.33}$$

As above, the result of A, B is independent of C. Also, CRB(ω) is asymptotically independent of the fact if C is estimated or assumed known. One may note from (5.30)-(5.31) that the CRB's for the A and B parameters are 1 to 4 times the corresponding CRB's in (5.26) and (5.27), respectively. In order to further analyze the above CRB's, a re-parameterized model is considered below.

Consider

$$s_n(\alpha, \phi, C, \omega) = \alpha \sin(\omega t_n + \phi) + C \tag{5.34}$$

where $A = \alpha \sin\phi$ and $B = \alpha \cos\phi$. Let

$$\boldsymbol{\Psi} = \begin{bmatrix} \alpha & \phi & C & \omega \end{bmatrix}^T \tag{5.35}$$

then the CRB for this parameter vector follows from the general relationship [Kay93]

$$\mathrm{CRB}(\boldsymbol{\Psi}) = \left[\frac{\partial \boldsymbol{\Psi}}{\partial \boldsymbol{\theta}}\right] \mathrm{CRB}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\Psi}}{\partial \boldsymbol{\theta}}\right]^T \tag{5.36}$$

where the k, r-th element of $[\partial \boldsymbol{\Psi}/\partial \boldsymbol{\theta}]$ is $\partial \boldsymbol{\Psi}_k/\partial \boldsymbol{\theta}_r$. Here,

$$\left[\frac{\partial \boldsymbol{\Psi}}{\partial \boldsymbol{\theta}}\right] = \begin{pmatrix} \sin\phi & \cos\phi & 0 & 0 \\ \frac{\cos\phi}{\alpha} & -\frac{\sin\phi}{\alpha} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.37}$$

The CRB for the re-parameterized model follows from

$$\mathrm{var}(\hat{\alpha}) \geq \mathrm{CRB}(\alpha) \simeq \frac{2\sigma^2}{N} \tag{5.38}$$

$$\mathrm{var}(\hat{\phi}) \geq \mathrm{CRB}(\phi) \simeq \begin{cases} \dfrac{2\sigma^2}{N\alpha^2} & \text{three parameter model} \\ \dfrac{8\sigma^2}{N\alpha^2} & \text{four parameter model} \end{cases} \tag{5.39}$$

$$\mathrm{var}(\hat{C}) \geq \mathrm{CRB}(C) \simeq \frac{\sigma^2}{N} \tag{5.40}$$

$$\mathrm{var}(\hat{\omega}) \geq \mathrm{CRB}(\omega) \simeq \frac{24\sigma^2}{\alpha^2 N^3} \quad \text{four parameter model} \tag{5.41}$$

In conclusion from (5.38)–(5.41), the uncertainty in the phase ϕ is dependent on whether ω is known or not, while the uncertainty in the amplitude α and DC-level C is independent of this fact.

In wave-form fitting, the quality of the parameter estimates is not as important as the quality of the actual wave-form fit, that is, the value of (5.2) for the given estimate. Clearly, for a given set of data the minimum value of (5.2) is reduced if the number of free parameters is increased. In fact with a suitable parameter vector with N entries, it is possible to obtain a perfect fit for a sequence of N input data, that is, the sum-squared-error (5.2) can be forced to zero. On the other hand, increasing the number of parameters implies that the estimator fits the parameters to the noise, and not to the signal. Thus, one should strive to use as few parameters as possible, but still have a model flexible enough to describe the behavior of the signal. The question to be asked is if the three- or four-parameter model should be used. Clearly, if the frequency is known one should use the three-parameter model, since more accurate estimates

of the parameters are expected, according to the CRB. On the other hand, if the frequency is unknown one should use the four-parameter model. In some applications, often the frequency is partially known, or known to be within some frequency range. In the forthcoming Section, the problem when to use the three- and four-parameter models is studied by aid of the parsimony principle.

5.3 The Parsimony Principle

Consider the criterion (5.2) for the signal (5.3) described by a vector $\boldsymbol{\vartheta}_0$, that is, the set of true but unknown parameters. Then the expected value of (5.2) is given by

$$\mathrm{E}[V(\boldsymbol{\vartheta}_0)] = \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}[(x_n - s_n(\boldsymbol{\vartheta}_0))^2] = \sigma^2 \tag{5.42}$$

Here the expectation is with respect to the measurement noise. Thus, when an estimate (say, $\hat{\boldsymbol{\vartheta}}$) is exact $\hat{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}_0$, the residual is white noise and has minimum variance. Now, assuming an estimate $\hat{\boldsymbol{\vartheta}}$ based on some *past* data, then a Taylor series expansion of $\varepsilon_n(\boldsymbol{\vartheta}) = x_n - s_n(\boldsymbol{\vartheta})$ around $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ gives

$$\mathrm{E}[V(\hat{\boldsymbol{\vartheta}})] = \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}\left[\left(\varepsilon_n(\boldsymbol{\vartheta}_0) + \left.\frac{\partial\varepsilon_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}^{T}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)\right)^2\right]. \tag{5.43}$$

Then, using the notation introduced in (5.18)

$$\boldsymbol{\psi}_n(\boldsymbol{\vartheta}_0) = \left.\frac{\partial s_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} = -\left.\frac{\partial\varepsilon_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \tag{5.44}$$

where the definition of $\varepsilon_n(\boldsymbol{\vartheta})$ was used in the second equality. Now, it follows that

$$\begin{aligned}
\mathrm{E}[V(\hat{\boldsymbol{\vartheta}})] &= \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}\left[\left(\varepsilon_n(\boldsymbol{\vartheta}_0) - \boldsymbol{\psi}_n^T(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)\right)^2\right] \\
&= \sigma^2 + \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}\left[\boldsymbol{\psi}_n^T(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)\boldsymbol{\psi}_n^T(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)\right] \\
&= \sigma^2 + \mathrm{Tr}\left\{\frac{1}{N}\mathrm{E}[(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)^T]\sum_{n=1}^{N}\boldsymbol{\psi}_n(\boldsymbol{\vartheta}_0)\boldsymbol{\psi}_n^T(\boldsymbol{\vartheta}_0)\right\} \\
&= \sigma^2 + \frac{\sigma^2}{N}\mathrm{Tr}\left\{\mathrm{cov}(\hat{\boldsymbol{\vartheta}})\,\mathrm{CRB}(\boldsymbol{\vartheta})^{-1}\right\} \qquad (5.45)
\end{aligned}$$

where in the second equality it is assumed that the estimate is unbiased. In the third equality the trace (Tr) operator is used to change the order of which the vectors are multiplied. Also in the third equality the order of summation and the trace operator has changed order. From (5.45), one note that replacing $\boldsymbol{\vartheta}_0$ with an estimate $\hat{\boldsymbol{\vartheta}}$ results in an increased mean-squared-error, where the effects of the estimate is given by the second term in (5.45). In the fourth equality the definition of the Fisher information matrix (5.17) has been used. If an efficient estimator is used to estimate the parameters in $\boldsymbol{\vartheta}$, the covariance $\mathrm{cov}(\hat{\boldsymbol{\vartheta}})$ equals the $\mathrm{CRB}(\boldsymbol{\vartheta})$. Hence, the term within the trace operator in (5.45) equals an identity matrix of size p, where $p = \dim(\boldsymbol{\vartheta})$. If a less accurate estimator is used $\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) - \mathrm{CRB}(\boldsymbol{\vartheta}) > 0$. Accordingly, the trace term in (5.45) is bounded below by p, so that

$$\mathrm{E}[V(\hat{\boldsymbol{\vartheta}})] \quad \geq \quad \sigma^2\left(1 + \frac{p}{N}\right). \qquad (5.46)$$

According to (5.46), the expected sum-squared-error increases with an increasing model order p. This is a known result and a general derivation of the parsimony principle can be found in [SS89]. As already stated, (5.46) holds with equality when an efficient estimator is employed. For uniform sampling according to (5.21) there exist in the considered scenario several unbiased estimators such that $\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) \simeq \mathrm{CRB}(\boldsymbol{\vartheta})$ [Kay93].

From (5.46) it is clear that the three parameter fit always (in expectation) results in a smaller sum-squared-error (5.2) than the four-parameter method. However, a three parameter method requires the frequency ω to be known. In many practical cases the frequency is known up to some

uncertainty quantity, a quantity that here is denoted by ω_δ. Accordingly, the frequency variable of the three-parameter fit (say $\tilde{\omega}$) deviates from the actual frequency ω_0, that is

$$|\tilde{\omega} - \omega_0| \leq \omega_\delta. \tag{5.47}$$

The question addressed is when a four parameter method performs better than a three parameter method, in terms of a smaller expected sum-squared-error. Clearly, a four parameter method results in a constant mean-squared-error independent of ω_δ, as seen from (5.46). In the three parameter case, however, the expectation of (5.9) depends on ω, as analyzed below.

5.3.1 Mean-squared-error analysis

Let $\mathcal{V}(\omega)$ denote the expectation (w.r.t the noise w_n) of (5.9), that is

$$\mathcal{V}(\omega) = \mathrm{E}\left[V(\omega)\right]. \tag{5.48}$$

Hence, $\mathcal{V}(\omega)$ is the expected sum-squared-error, and is a measure of the expected quality of the wave-form fit. A second order Taylor series expansion of $\mathcal{V}(\omega)$ around the true frequency $\omega = \omega_0$ gives

$$\mathcal{V}(\omega) \simeq \mathcal{V}(\omega_0) + \left.\frac{\partial \mathcal{V}(\omega)}{\partial \omega}\right|_{\omega=\omega_0} (\omega - \omega_0) + \frac{1}{2} \left.\frac{\partial^2 \mathcal{V}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_0} (\omega - \omega_0)^2 \tag{5.49}$$

The first two derivatives of $\mathcal{V}(\omega)$ with respect to ω are given below, that is

$$\frac{\partial \mathcal{V}(\omega)}{\partial \omega} = \mathrm{E}\left\{ -\frac{2}{N} [\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})]^T \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \omega} \right\} \tag{5.50}$$

$$\frac{\partial^2 \mathcal{V}(\omega)}{\partial \omega^2} = \mathrm{E}\left\{ \frac{2}{N} \frac{\partial \mathbf{s}^T(\boldsymbol{\theta})}{\partial \omega} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \omega} \right\} - \mathrm{E}\left\{ \frac{2}{N} \left[\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})\right]^T \frac{\partial^2 \mathbf{s}(\boldsymbol{\theta})}{\partial \omega^2} \right\}. \tag{5.51}$$

The used estimator is assumed unbiased which implies that the expectation of the difference $[\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})]$ in (5.50) evaluated for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ by definition equals the zero vector. Therefore, the first derivative of $\mathcal{V}(\omega)$ with respect to ω in (5.49) equals zero. The same argument holds in (5.51), that is putting the second term to zero as well. In the reconstruction of $s_n(\boldsymbol{\theta})$, the estimate of the three parameters in $\hat{\theta}$ depends on the assumed frequency $\tilde{\omega}$, see (5.8). For notational simplicity this dependency is not

explicitly shown in the calculations below. Replacing θ with its least-squares solution (5.8) and taking the derivative with respect to ω using the chain rule yields

$$\begin{aligned}\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \omega} =& \frac{\partial}{\partial \omega}\mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{x} \\ =& \mathbf{D}_\omega[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{x} + \mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}_\omega^T\mathbf{x} \\ & - 2\mathbf{D}\mathbf{D}^T\mathbf{D}\left[\mathbf{D}_\omega^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}\mathbf{D}^T\mathbf{x} \end{aligned} \tag{5.52}$$

where the subscript $[\cdot]_\omega$ denotes an element-wise derivative with respect to ω. The differentiating rules applied on the inverse of $\mathbf{D}^T\mathbf{D}$ can be found in Appendix 5.B (see (5.59)). The derivation of the second derivative (5.51) is straightforward, but tedious. It can also be found in Appendix 5.B. Evaluated at $\omega = \omega_0$, the second derivative can be expressed as

$$\left.\frac{\partial^2 \mathcal{V}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_0} \simeq \frac{(A^2+B^2)N^2}{12} \tag{5.53}$$

Inserting (5.53) in the Taylor series expansion (5.49), the expected value of $V(\omega)$ using a three parameter model is described by a second order parabola

$$\mathcal{V}(\omega) \simeq \sigma^2(1+\frac{3}{N}) + \frac{(A^2+B^2)N^2}{24}(\omega-\omega_0)^2. \tag{5.54}$$

The mean-squared-error (5.54) is a main result in this chapter. To summarize, applying a four-parameter fit results in a mean-squared-error given by (5.46) for $p = 4$. For a three-parameter method the corresponding result is given by (5.54), which is a function of the difference between the actual signal frequency ω_0 and the frequency variable of the algorithm $\omega = \tilde{\omega}$.

The *break-even* frequency for which the considered methods perform equally well is obtained by setting (5.46) for $p = 4$ equal to (5.54) and solving for $|\omega - \omega_0|$. The result yields

$$|\omega - \omega_0| = \sqrt{\frac{24\sigma^2}{\alpha^2 N^3}}. \tag{5.55}$$

It is interesting to note that the break-even distance from ω_0 equals the square root of the CRB(ω) in (5.33). Hence, if the algorithm variable $\tilde{\omega}$ is guaranteed to be close enough (as given by (5.55)) to ω_0, the three-parameter model should be used in favor of a four-parameter model.

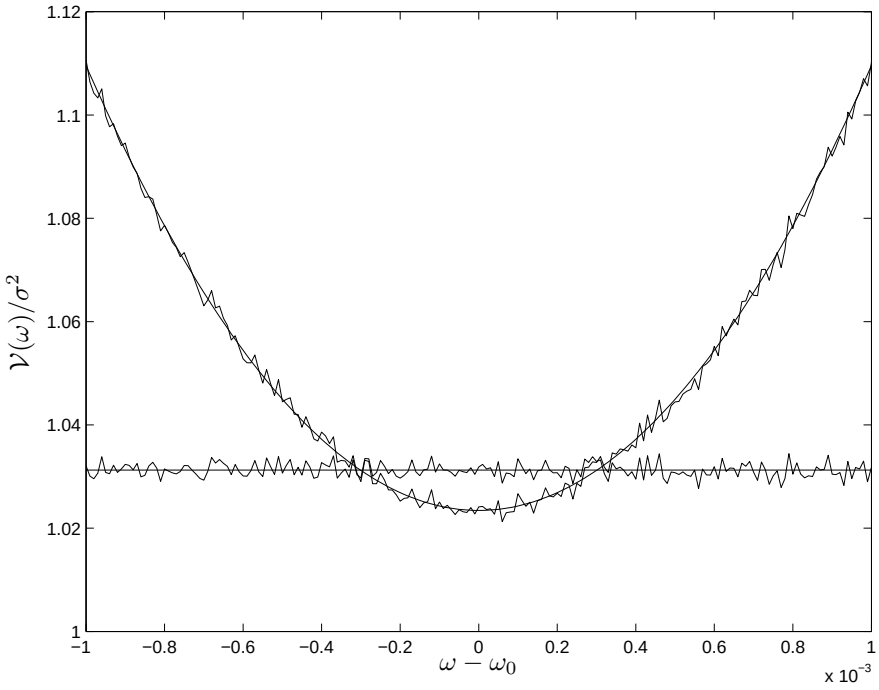


Figure 5.1: The mean-squared-error $\mathcal{V}(\omega)$ normalized with the noise variance σ^2 for a three- (parabola) and four-parameter (straight line) model. The dashed lines are numerical evaluations of the residual using the different models. The curves have been obtained for $N = 128$ and SNR = 18dB.

5.4 Discussion and Numerical results

The theoretical results in section 5.2 and 5.3 are based on the Gaussian assumption. The purpose of this section is twofold. First, the theoretical results are illustrated by numerical examples under the Gaussian assumption. Secondly, the applicability of the theoretical results in practical scenarios is investigated. A typical estimation scenario with quantized measurements is investigated followed by an investigation of applicability of the theoretical results on ADC-testing.

5.4.1 Illustration of the Parsimony Principle

In Figure 5.1, the mean-squared-error $\mathcal{V}(\omega)$ (5.48) is shown, both the theoretical results (5.46) for $p = 4$ and (5.54) for $p = 3$, respectively; and experimental evaluations thereof based on Monte-Carlo simulations. The results have been obtained for the Gaussian scenario for $N = 128$ and a signal-to-noise ratio SNR $= \alpha^2/(2\sigma^2)$ with $\alpha^2 = A^2 + B^2$ at 18dB. In the evaluation, the sampling rate $f_s = 1$Hz and the actual frequency ω_0 has been randomly and uniformly drawn in the region $[(2\pi)/N, \pi - (2\pi)/N]$. The DC-level C has been randomly and uniformly drawn in the region $[-0.1, 0.1]$. The mean-squared-error has been evaluated using a frequency with a variable displacement from the true frequency for the three-parameter method, and a maximum-likelihood estimate of the signal frequency for the four-parameter fit, respectively. In the simulations, the expectation has been approximated with the sample mean, using $2.5 \cdot 10^5$ independent Monte-Carlo simulations at each frequency point.

In Figure 5.1, it is seen that for small deviations of ω from ω_0 a sinewave fit using a three-parameter model performs better than a four-parameter model in terms of a smaller mean-squared-error, as predicted by the theory. Further, there is an excellent agreement between theoretical results and the results obtained from Monte-Carlo simulations.

5.4.2 Quantization in an Estimation Scenario

Finding the maximum likelihood estimates of the signal parameters $\boldsymbol{\vartheta}$ is under the Gaussian assumption equivalent with minimizing the sum-squared-error (5.9). The previous discussion indicates that the waveform fit performs better (that is, a lower mean-squared-error) if the frequency is known. In this example the waveform fit performance is investigated when the measurements are quantized. The use of quantized data will be in conflict with the considered signal model. However, the quantization can be modeled by the noise term w_n, which is not Gaussian anymore. In Figure 5.2, the mean-squared-error (5.48) using quantized measurements is shown. The measurement signal has been quantized with a uniform 5-bit quantizer. The range of the quantizer influences the numerical value of the mean-squared-error. To fully investigate the influence of the gain control prior to quantization is out of the scope of this chapter. The quantizer input range is tuned to fully cover the signal, and the full scale range (FSR) equal to 2.4 has been used. The signal- and noise levels, the number of data points and the number of Monte-Carlo runs are the same

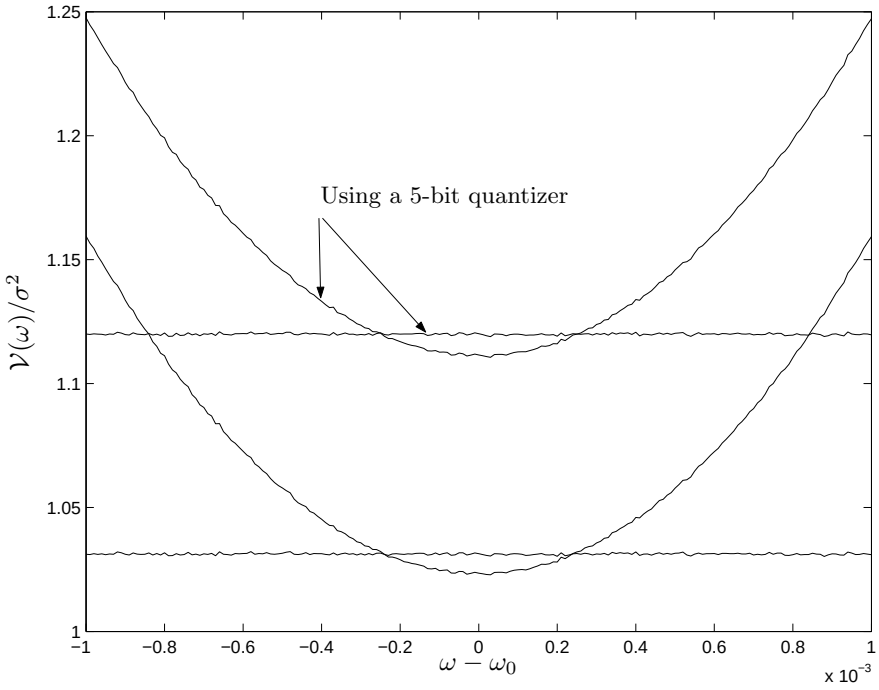


Figure 5.2: The mean-squared-error $\mathcal{V}(\omega)$ normalized with the noise variance σ^2 for a three- (parabola) and four-parameter (straight line) model. The lines show numerical evaluations of the residual using the different models. The two topmost curves are obtained when the measurement signal has been quantized with a 5-bit quantizer. The two lower curves show the residual power using un-quantized measurements. The curves have been obtained for $N = 128$ and SNR = 20dB.

as in the setup resulting in Figure 5.1. The signal-to-noise-and-distortion ratio is in this case neglectable compared with the Gaussian noise variance. In an estimation scenario this is often the case. It is seen from Figure 5.2 that the mean-squared-error increases when the measurements are quantized compared with the corresponding results for un-quantized data. However, the parabolic shape of the mean-squared-error using the three-parameter model is unchanged, and the intersection points with the mean-squared-error using the four-parameter model coincides with

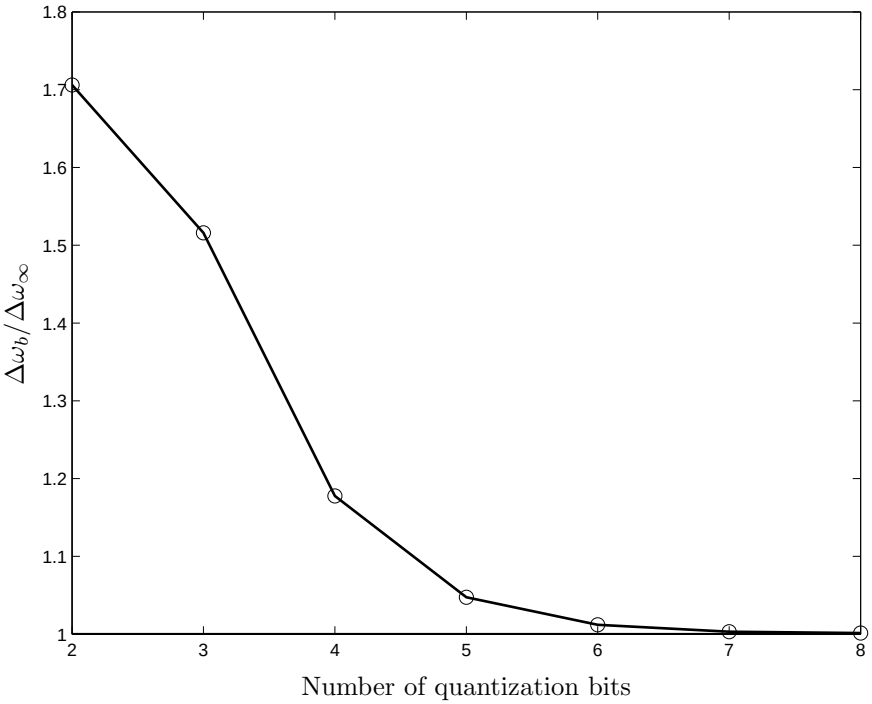


Figure 5.3: The *break-even* frequency $\Delta\omega_b$ for quantized data. The break-even frequency is normalized by $\Delta\omega_\infty$ given by (5.54). The curve has been obtained for $N = 128$ and SNR = 20dB.

the corresponding point for un-quantized measurements. A formula for this intersection is found in (5.55).

A numerical evaluation of the break-even frequency as a function of the number of bits in the quantization has been performed. The intersection occurs when the three-parameter algorithm is used with a frequency variable $\tilde{\omega}$ that differs from ω_0. Let $\Delta\omega_b$ be the difference $|\omega - \omega_0|$ where the intersection occurs, that is the break-even frequency. Here, the subscript b denotes the number of bits used in the quantizer. In Figure 5.3, $\Delta\omega_b$ is plotted for the quantization levels $b \in \{2, 3, 4, 5, 6, 7, 8\}$. From the figure it can be noted that $\Delta\omega_b$ increases with a decreasing number of bits. In the case of $b = 5$, $\Delta\omega_5$ is about 6% larger than for the un-quantized case $\Delta\omega_\infty$. It can also be noted that if the measurement

data is quantized with 8-bits then the effect of quantization is negligible, for the considered SNR. The deviation of $\Delta\omega_b$ from $\Delta\omega_\infty$ is due to the quantization, where each bit of reduction decreases the signal-to-noise-and-distortion ratio by 6 dB. From Figure 5.3, it is evident that for the considered scenario the quantization noise has an impact on the break-even frequency for quantizers with less than 6 bit.

5.4.3 Uniform Noise Model of Quantization and ADC Testing

In this example the validity of the derived results is investigated under the uniform noise model of quantization. Consider a 3-bit quantizer with a FSR of 2.46 and a signal with unit amplitude, that results in a signal- and distortion ratio (SNDR) equal to 18dB. The mean-squared-error is displayed in Figure 5.4. The simulation setup is identical to the one in Section 5.4.1. From the diagram in Figure 5.4 we can conclude that the derived results are also valid under the uniform noise model of quantization.

It is worth mention that the uniform noise model of quantization is a poor approximation in testing ADCs [KB04]. In particular the pdf is not uniform with strong peaks whose location depends on the signal amplitude and DC-offset, as well as a dependency between the quantization error and the signal. The implication is that neither the three- or four-parameter model describe the measured data, and more complicated models are needed for unbiased estimation such as the one in [PS96].

5.5 Conclusions

The IEEE standard 1057 for tone frequency estimation and signal recovery has been considered, and a performance analysis employing the Cramér-Rao bound (CRB) and the parsimony principle has been performed. The asymptotic CRB analysis shows, among other things, that the accuracy of the amplitude estimate is independent of the DC-level. The quality of the wave form fit is evaluated in terms of the expected sum-squared-error, and a simple rule (5.55) is derived, that is suitable as a rule-of-thumb when selecting a proper estimation algorithm for the given problem. The rule (5.55) is shown to be valid also under the uniform noise model of quantization.

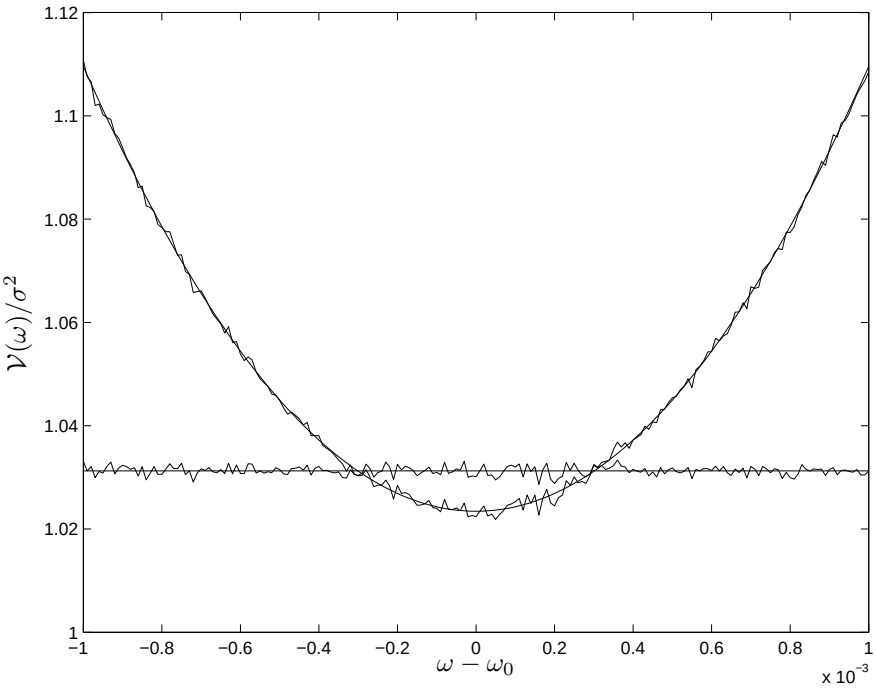


Figure 5.4: The mean-squared-error $\mathcal{V}(\omega)$ normalized with the noise variance σ^2 for a three- (parabola) and four-parameter (straight line) model. The dashed lines are numerical evaluations of the residual using the different models. The curves have been obtained for $N = 128$ and a uniform noise at SNR = 18dB.

Appendix 5.A Derivation of the Fisher information matrix $J(\boldsymbol{\vartheta})$ (5.17)

The quantity $\partial \ln p(\mathbf{x};\boldsymbol{\vartheta}/\partial\boldsymbol{\vartheta})$ in (5.14) can be expanded as

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} =& \frac{\partial}{\partial\boldsymbol{\vartheta}}\left\{-\frac{1}{N}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})\right)^T\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})\right)\right\} \\ =& -\frac{1}{\sigma^2}\frac{\partial \mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})\right) \end{aligned} \tag{5.56}$$

The quantity $\partial\mathbf{s}(\boldsymbol{\vartheta})^T/\partial\boldsymbol{\vartheta}$ is a $(p\times N)$ matrix, following from the definition of (5.16). Let $\mathbf{v}$ be a column vector with the elements $\{v_1 \dots v_N\}$ which are independent of $\boldsymbol{\vartheta}$. Then

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\vartheta}}(\mathbf{s}(\boldsymbol{\vartheta})^T\mathbf{v}) =& \begin{bmatrix}\frac{\partial}{\partial\boldsymbol{\vartheta}_1}\mathbf{s}(\boldsymbol{\vartheta})^T\mathbf{v} \\ \vdots \\ \frac{\partial}{\partial\boldsymbol{\vartheta}_p}\mathbf{s}(\boldsymbol{\vartheta})^T\mathbf{v}\end{bmatrix} = \begin{bmatrix}\frac{\partial}{\partial\boldsymbol{\vartheta}_1}s_1(\boldsymbol{\vartheta}) & \dots & \frac{\partial}{\partial\boldsymbol{\vartheta}_1}s_N(\boldsymbol{\vartheta}) \\ \vdots & & \\ \frac{\partial}{\partial\boldsymbol{\vartheta}_p}s_1(\boldsymbol{\vartheta}) & \dots & \frac{\partial}{\partial\boldsymbol{\vartheta}_p}s_N(\boldsymbol{\vartheta})\end{bmatrix}\mathbf{v} \\ =& \frac{\partial\mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\mathbf{v} \end{aligned} \tag{5.57}$$

Now inserting (5.56) and (5.57) with $\mathbf{v} = (\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})) = [w_1 \quad \dots \quad w_N]^T$ into (5.14), one has

$$\begin{aligned}\mathbf{J}(\boldsymbol{\vartheta}) =& \frac{1}{\sigma^4}\,\mathrm{E}\left\{\frac{\partial\mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})\right)\left(\mathbf{x}-\mathbf{s}(\boldsymbol{\vartheta})\right)^T\left(\frac{\partial\mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\right)^T\right\} \\ =& \frac{1}{\sigma^4}\frac{\partial\mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\,\mathrm{E}\{\begin{bmatrix}w_1^2 & w_1w_2 & \dots & w_1w_N \\ w_2w_1 & w_2^2 & \dots & w_2w_N \\ \vdots & \vdots & \ddots & \vdots \\ w_Nw_1 & w_Nw_2 & \dots & w_N^2\end{bmatrix}\}\left(\frac{\partial\mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\right)^T \\ =& \frac{1}{\sigma^2}\frac{\partial\mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\left(\frac{\partial\mathbf{s}(\boldsymbol{\vartheta})^T}{\partial\boldsymbol{\vartheta}}\right)^T. \end{aligned} \tag{5.58}$$

The third equality follows from the assumption that the noise terms w_n are uncorrelated, and thus the noise covariance matrix equals $\sigma^2\mathbf{I}$. Using the notation introduced in (5.18), the result (5.17) follows. The derivations above follows the one in [Kay93] closely.

Appendix 5.B The Derivative of the Mean-squared-error

Let $\mathbf{M}$ be a matrix where all elements are differentiable with respect to ω and with an inverse such that $\mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$. Then

$$\begin{aligned} 0 = \frac{\partial \mathbf{I}}{\partial \omega} = \frac{\partial}{\partial \omega}(\mathbf{M}\mathbf{M}^{-1}) = \mathbf{M}_\omega \mathbf{M}^{-1} + \mathbf{M}[\mathbf{M}^{-1}]_\omega \\ \Rightarrow \frac{\partial}{\partial \omega}\mathbf{M}^{-1} = -\mathbf{M}^{-1}\mathbf{M}_\omega \mathbf{M}^{-1}. \end{aligned} \tag{5.59}$$

In (5.59) the notation $[\cdot]_\omega$ denotes derivation of $[\cdot]$ with respect to ω. Using (5.59) the derivative (5.52) can be expressed as

$$\begin{aligned} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \omega} =& \frac{\partial}{\partial \omega}\mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{x} = \mathbf{D}_\omega[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{x} + \mathbf{D}[\mathbf{D}^T\mathbf{D}]_\omega^{-1}\mathbf{D}^T\mathbf{x} \\ &+ \mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}_\omega^T\mathbf{x} - \mathbf{D}(\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1} \\ &+ \left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1})\mathbf{D}^T\mathbf{x} \end{aligned} \tag{5.60}$$

Using the fact that for large N (that is $N \to \infty$), $\mathbf{D}_\omega^T\mathbf{D}$ is symmetric and (5.60) can be expressed

$$\begin{aligned} \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \omega} =& \mathbf{D}_\omega[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{x} + \mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}_\omega^T\mathbf{x} \\ &- 2\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{x} \end{aligned} \tag{5.61}$$

The product of (5.61) with its transpose, as appeared in (5.51) equals

$$\begin{aligned}
\mathrm{E}\left\{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta})}{\partial\omega}\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial\omega}\right\} =& \mathbf{x}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{x} && (T_I)\\
&+ 2\mathbf{x}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{x} && (T_{II})\\
&- 4\mathbf{x}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\\
&\quad \mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{x} && (T_{III})\\
&+ \mathbf{x}^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{x} && (T_{IV})\\
&- 4\mathbf{x}^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\\
&\quad \mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{x} && (T_V)\\
&+ 4\mathbf{x}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\\
&\quad \mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{x} && (T_{VI})
\end{aligned} \tag{5.62}$$

As $N \gg 1/\omega$, the quantity $\mathbf{D}^T\mathbf{D}$ above can be approximated by the constant matrix

$$\mathbf{D}^T\mathbf{D} \simeq \begin{bmatrix} \frac{N}{2} & 0 & 0 \\ 0 & \frac{N}{2} & 0 \\ 0 & 0 & N \end{bmatrix}. \tag{5.63}$$

The same approximation is applied to $\mathbf{D}_\omega^T\mathbf{D}$ and $\mathbf{D}_\omega^T\mathbf{D}_\omega$, that is

$$\mathbf{D}_\omega^T\mathbf{D}_\omega \simeq \begin{bmatrix} \frac{N^3}{6} & 0 & 0 \\ 0 & \frac{N^3}{6} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{5.64}$$

and

$$\mathbf{D}_\omega^T\mathbf{D} \simeq \begin{bmatrix} 0 & \frac{N^2}{4} & 0 \\ \frac{N^2}{4} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{5.65}$$

Further exploiting the fact that $\mathbf{D}(\omega_0)\theta_0 = \mathbf{s}(\boldsymbol{\theta}_0)$, implies that

$$\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}_0) + \mathbf{w} = \mathbf{D}(\omega_0)\theta_0 + \mathbf{w}. \tag{5.66}$$

Now, each term in (5.62) can be expressed in terms of the true parameters $\boldsymbol{\theta}_0$. Towards that end let $\mathrm{E}\left\{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta})}{\partial \omega}\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \omega}\right\} = T_I + \cdots + T_{VI}$, then

$$\begin{aligned}
T_I =&\theta_0^T\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}\theta_0\\
&+2\theta_0^T\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\underbrace{\mathrm{E}\{\mathbf{w}\}}_{=0}\\
&+\mathrm{E}\ \mathrm{Tr}\left\{\mathbf{w}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{w}\right\}\\
=&\theta_0^T\mathbf{D}_\omega^T\mathbf{D}_\omega\theta_0+\sigma^2\mathrm{Tr}\left\{\mathbf{D}_\omega^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\right\}\\
\simeq&\frac{(A^2+B^2)N^3}{6}+\frac{\sigma^2N^2}{3}
\end{aligned}\tag{5.67}$$

$$\begin{aligned}
T_{II} =&2\theta_0^T\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\theta_0+4\theta_0^T\ldots\mathrm{E}\{\mathbf{w}\}\\
&+2\,\mathrm{E}\ \mathrm{Tr}\left\{\mathbf{w}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{w}\right\}\\
=&2\theta_0^T\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\theta_0\\
&+2\sigma^2\mathrm{Tr}\left\{\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\right\}\\
\simeq&\frac{(A^2+B^2)N^3}{4}+\sigma^2N^2
\end{aligned}\tag{5.68}$$

where $(\ldots)$ indicates a product of factors that is multiplied by the zero vector $\mathrm{E}\{\mathbf{w}\}$.

$$\begin{aligned}
T_{III} =&-4\theta_0^T\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}\theta_0\\
&-8\theta_0^T\ldots\mathrm{E}\{\mathbf{w}\}\\
&-4\,\mathrm{E}\,\mathrm{Tr}\left\{\mathbf{w}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{w}\right\}=\ldots\\
\simeq&-\frac{(A^2+B^2)N^3}{2}-2\sigma^2N^2
\end{aligned}\tag{5.69}$$

$$\begin{aligned}
T_{IV} =&\theta_0^T\mathbf{D}^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\theta_0+2\theta_0^T\ldots\mathrm{E}\{\mathbf{w}\}\\
&+\mathrm{E}\ \mathrm{Tr}\left\{\mathbf{w}^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}^T\mathbf{D}\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{w}\right\}\\
=&\theta_0^T\mathbf{D}^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{D}_\omega^T\mathbf{D}\theta_0+\sigma^2\mathrm{Tr}\left\{\mathbf{D}_\omega^T\mathbf{D}_\omega\left[\mathbf{D}^T\mathbf{D}\right]^{-1}\right\}\\
\simeq&\frac{(A^2+B^2)N^3}{8}+\frac{2\sigma^2N^2}{3}
\end{aligned}\tag{5.70}$$

$$\begin{aligned} T_V =& -4\theta_0^T \mathbf{D}^T \mathbf{D}_\omega \left[\mathbf{D}^T\mathbf{D}\right]^{-1} \mathbf{D}_\omega^T \mathbf{D}\theta_0 - 8\theta_0^T \ldots \mathrm{E}\{\mathbf{w}\} \\ & - 4\,\mathrm{E}\ \mathrm{Tr}\left\{\mathbf{w}^T \mathbf{D}_\omega \left[\mathbf{D}^T\mathbf{D}\right]^{-1} \mathbf{D}_\omega^T \mathbf{D} \left[\mathbf{D}^T\mathbf{D}\right]^{-1} \mathbf{D}^T\mathbf{w}\right\} \\ \simeq& -\frac{(A^2+B^2)N^3}{2} - 2\sigma^2 N^2 \end{aligned} \tag{5.71}$$

$$\begin{aligned} T_{VI} =& 4\theta_0^T \mathbf{D}_\omega^T \mathbf{D} \left[\mathbf{D}^T\mathbf{D}\right]^{-1} \mathbf{D}_\omega^T \mathbf{D}\theta_0 + 8 \ldots \mathrm{E}\{\mathbf{w}\} \\ & + 4\,\mathrm{E}\ \mathrm{Tr}\left\{\mathbf{w}^T \mathbf{D} \left[\mathbf{D}^T\mathbf{D}\right]^{-1} \mathbf{D}_\omega^T \mathbf{D} \left[\mathbf{D}^T\mathbf{D}\right]^{-1} \mathbf{D}_\omega^T \mathbf{D} \left[\mathbf{D}^T\mathbf{D}\right]^{-1} \mathbf{D}^T\mathbf{w}\right\} \\ \simeq& \frac{(A^2+B^2)N^3}{2} + 2\sigma^2 N^2 \end{aligned} \tag{5.72}$$

The sum of all terms $(T_I + \ldots T_{VI})$ yields

$$\begin{aligned} \mathrm{E}\left\{\frac{\partial \mathbf{s}^T(\theta)}{\partial\omega}\frac{\partial \mathbf{s}(\theta)}{\partial\omega}\right\} \simeq& (A^2+B^2)N^3\left(\frac{1}{6}+\frac{1}{4}-\frac{1}{2}+\frac{1}{8}-\frac{1}{2}+\frac{1}{2}\right) \\ & + \sigma^2 N^2\left(\frac{1}{3}+1-2+\frac{2}{3}-2+2\right) \\ =& \frac{(A^2+B^2)N^3}{24} \end{aligned} \tag{5.73}$$

By (5.73), equation (5.53) directly follows by multiplication with $2/N$.

Chapter 6

Model Order Selection in Waveform fitting

6.1 Introduction

Waveform fitting based on digital data is a common problem in instrumentation and measurements. In this chapter, model order selection for accurate fitting of waveforms is considered for a linear signal model in general, and a sinewave model in particular. The problem at hand is to derive guidelines for the user how to select among different structures within a hierarchical set of models. For models that are linear in the sought parameters, a general relation between parameter uncertainty and quality of wave-fit is derived. Studied as well is the sinewave fit, where the sought frequency enters the model in a non-linear fashion. Presented results are extensions of the work presented in Chapter 5.

A crucial question in system identification is the proper selection of model structure. Consider a parametric model $\mathcal{M}_1$ described by the parameter vector θ_1, where θ_1 has n_1 entries. Further consider the model $\mathcal{M}_2$ described by θ_2 with $n_2 > n_1$ entries. The models $\mathcal{M}_1$ and $\mathcal{M}_2$ are hierarchically related, meaning that the more restricted model $\mathcal{M}_1$ is a subset of $\mathcal{M}_2$. With $\theta_2 = (\theta_1^T, \eta^T)^T$ (where T denotes the transpose operation), the hierarchical relationship can be expressed by

$$\mathcal{M}_1(\theta_1) = \mathcal{M}_2(\theta_2)|_{\theta_2=(\theta_1^T, \bar{\eta}^T)^T} \tag{6.1}$$

Here, the vector η of dimension $\dim(\eta) = n_2 - n_1$ denotes the parameters

of $\mathcal{M}_2$ that are not included in θ_1. Due to the hierarchical relationship between the models, (6.1) is fulfilled when η in $\mathcal{M}_2$ is replaced by the *known* set of variables $\bar{\eta}$, where $\bar{\eta}$ is implicitly included in $\mathcal{M}_1$. An important example of hierarchical models in waveform fitting is the sinewave model, as outlined below.

6.1.1 Sinewave-fitting

In sinewave-fitting two models are employed [S1057], namely the more restrictive three parameter model (here, corresponding to $\mathcal{M}_1$) with sought parameters

$$\theta_1 = (A,\, B,\, C)^T \tag{6.2}$$

and the four parameter model (that is, $\mathcal{M}_2$) with

$$\theta_2 = (\underbrace{A,\, B,\, C}_{\theta_1},\, \underbrace{\omega}_{\eta})^T \tag{6.3}$$

The model structure $\mathcal{M}$ that contains $\mathcal{M}_1$ as well as $\mathcal{M}_2$ is given by

$$x[n;\, \theta] = A\sin(\omega\, n) + B\cos(\omega\, n) + C \tag{6.4}$$

In (6.4), θ denotes a generic parameter vector. Further, the variable ω is assumed to be a *known* constant $\omega = \bar{\omega}$ for $\mathcal{M}_1$, whereas for $\mathcal{M}_2$ it is gathered in the vector θ_2 of sought parameters.

6.2 The Parsimony Principle

In scenarios where all four parameters A, B, C and ω are unknown, clearly $\mathcal{M}_2$ should be used as the basis for a parametric estimation procedure, and thus all $n_2 = 4$ parameter values are sought for. In scenarios where the more restrictive model $\mathcal{M}_1$ is sufficient to describe the behavior of the measurement, it should be used in favor of $\mathcal{M}_2$ due to the parsimony principle [SS89], briefly reviewed below. Indeed, for a given set of measurement data

$$\{y[0], \ldots, y[N-1]\} \tag{6.5}$$

the more flexible model (that is $\mathcal{M}_2$) will result in a smaller minimum sum-squared-error, than $\mathcal{M}_1$. In other terms, for a given set of data (6.5) it holds that

$$\min_{\theta_2} V(\theta_2) \leq \min_{\theta_1} V(\theta_1) \tag{6.6}$$

where $V(\theta)$ (for a generic parameter vector θ) equals the sum-squared-error

$$V(\theta) = \frac{1}{N} \sum_{\ell=0}^{N-1} (y[\ell] - x[\ell;\, \theta])^2 \tag{6.7}$$

In (6.7), $x[n;\, \theta]$ denotes the signal model parameterized by the generic parameter vector θ (for example, sinewave-fitting according to (6.4)). A small sum-squared-error is not necessarily a measure of a good fit of the model to the measurements. The reason behind this it that the minimizing argument of $V(\theta)$ not only fits the data (6.5) to the parametric model $x[n;\, \theta]$, but also to the measurement noise and model imperfections. In order to analyze this behavior, we consider an additive model of the measurements as

$$y[n] = x[n;\, \bar{\theta}] + e[n] \tag{6.8}$$

In (6.8), $y[n]$ denotes the measurements, $x[n;\, \bar{\theta}]$ the signal model described by the set of true parameters $\theta = \bar{\theta}$, and $e[n]$ models noise, model imperfections and so alike. It is further assumed that $e[n]$ is zero-mean white Gaussian noise with variance σ^2. Inserting (6.8) into (6.7), and taking expectation with respect to the additive noise yields, the mean sum-squared-error (MSSE)

$$\mathcal{V}(\bar{\theta}) \triangleq \frac{1}{N} \mathrm{E}[V(\bar{\theta})] = \frac{1}{N} \sum_{\ell=0}^{N-1} \mathrm{E}[e(n)^2] = \sigma^2 \tag{6.9}$$

Thus, a perfect waveform fit yields a MSSE given by the variance of the additive noise. The result (6.9) is however of little practical use since θ is indeed unknown and has to be estimated. In this paper, an estimate is denoted by $\hat{\theta}$. Due to the additive noise it no longer holds that $\hat{\theta} = \bar{\theta}$. In fact, the estimate $\hat{\theta}$ is a stochastic variable that may be characterized by its first and second order moments, that is mean value and covariance matrix. If not otherwise stated we assume that $\hat{\theta}$ is a statistically efficient estimate, that is unbiased and with a covariance matrix corresponding to the Cramér-Rao lower bound (CRB) [Kay93]. Now, the MSSE is given by, [SS89]

$$\mathcal{V}(\hat{\theta}) \simeq \sigma^2 \left(1 + \frac{p}{N}\right) \tag{6.10}$$

where $p = \dim(\theta)$. In (6.10), $\simeq$ denotes an approximate equality where only the dominant terms have been retained. In order to illustrate the importance of (6.10), we consider the sinewave-fitting application.

6.2.1 Sinewave-fitting (cont'd)

Sinewave-fitting with respect to $\mathcal{M}_1$ and $\mathcal{M}_2$ is considered in [S1057]. The implication of (6.10) to sinewave-fitting is that *i)* if the sinewave frequency $\bar{\omega}$ is known the three parameter fit should be used in favor of the four parameter fit. The MSSE is given by (6.10) with $p = \dim(\theta_1) = 3$. *ii)* If, on the other hand, the four parameter fit is employed the resulting error variance is slightly increased, that is given by (6.10) with $p = \dim(\theta_2) = 4$. From a robustness point of view, it is often argued that one should always employ the four parameter fit, regardless of the extra parameter resulting in an inferior MSSE. This robustness aspect was studied in some detail in [AH05a] where the influence of frequency errors on the MSSE was studied. Clearly, if the frequency is not perfectly known, there is a boarder line beyond which the four parameter fit outperforms the three parameter fit. In particular, it was shown that for the three parameter fit [AH05a]

$$\mathcal{V}(\hat{\theta}_1, \Delta\omega) \simeq \sigma^2 \left(1 + \frac{3}{N}\right) + \frac{(A^2 + B^2)N^2}{24}\Delta\omega^2 \tag{6.11}$$

where $\Delta\omega$ is the frequency error $\Delta\omega = \omega - \bar{\omega}$. For small errors $|\Delta\omega| \approx 0$ the MSSE (6.11) is smaller than (6.10) for $p = 4$, and thus the three parameter fit should still be employed. On the other hand, for large frequency errors $|\Delta\omega| \gg 0$ the four parameter fit should be employed, resulting in a better fit through its flexibility to adjust the frequency parameter. For $\Delta\omega = 0$, the expression (6.11) coincides with (6.10) for $p = 3$. An illustration of the theoretical results above can be found in Figure 6.1. Also included in the diagram of Figure 6.1 is results obtained from extensive Monte-Carlo simulations; See Chapter 5 for details.

6.3 Linear Models

Consider a hierarchical structure linear in θ, that is the waveform $x[n; \theta]$ can be written in matrix form as

$$\mathbf{x}[\theta] = \mathbf{H}\theta_1 + \mathbf{G}\eta \tag{6.12}$$

where the n-th row in $\mathbf{x}[\theta]$ corresponds to $x[n, \theta]$. $\mathbf{H}$ and $\mathbf{G}$ are the known matrices of size $(N \times n_1)$ and $(N \times n_2 - n_1)$, respectively. Further, we assume a low order model $\mathcal{M}_1$ described by θ_1 of dimension n_1, and a high order model $\mathcal{M}_2$ of dimension n_2. The models are assumed hierarchically

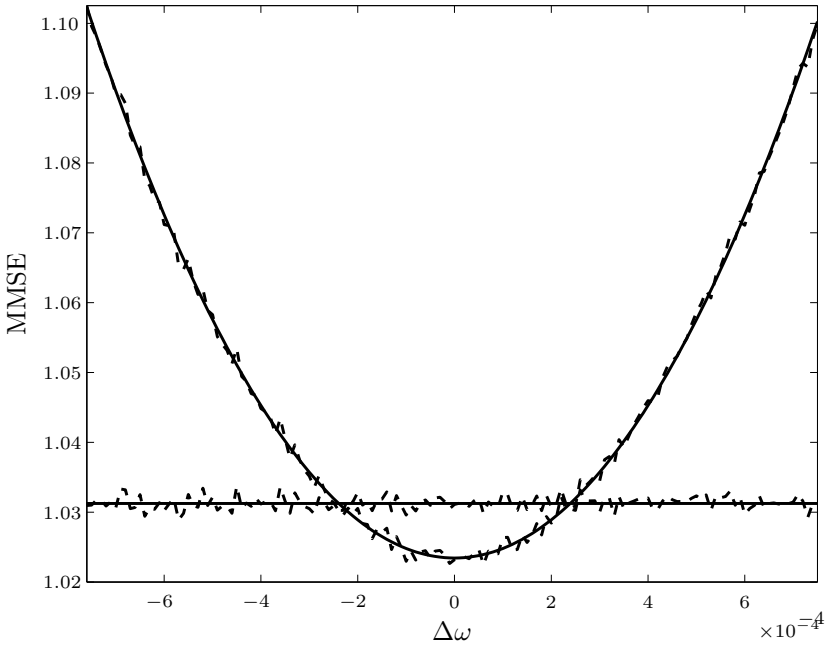


Figure 6.1: The MSSE normalized by the noise variance σ^2 for a three- (parabola) and four-parameter model (straight line). The dashed lines are numerical evaluations of the MSSE using the different models. The curves have been obtained for $N = 128$ and SNR = 20dB.

related with $\theta_2 = (\theta_1^T, \eta^T)^T$, where $\dim(\eta) = n_2 - n_1$. The parameter difference $\Delta\eta = \eta - \bar{\eta}$ describes the error in the variable η when applying $\mathcal{M}_1$ for solving the estimation problem. The goodness of the fit when using $\mathcal{M}_2$ in terms of MSSE is directly related to the number of unknown parameters n_2 and given by (6.10) with $p = n_2$.

The use of structure $\mathcal{M}_2$ requires an estimate of θ_2. If all columns in the model matrices $\mathbf{H}$ and $\mathbf{G}$ are linearly independent, an unbiased and minimum variance estimate of θ_2 can be found by [Kay93]

$$\hat{\theta}_2 = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\eta} \end{bmatrix} = \begin{bmatrix} \mathbf{H}^T\mathbf{H} & \mathbf{H}^T\mathbf{G} \\ \mathbf{G}^T\mathbf{H} & \mathbf{G}^T\mathbf{G} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}^T \\ \mathbf{G}^T \end{bmatrix} \mathbf{y}. \tag{6.13}$$

In (6.13), $\mathbf{y}$ denotes the vector of measurements. Further, a lower limit on the covariance of the estimate (6.13) is given by the CRB. The CRB

is determined by the model and the noise variance as [Kay93]

$$\mathrm{CRB}\left(\begin{bmatrix}\theta_1\\ \eta\end{bmatrix}\right) = \sigma^2 \begin{bmatrix}\mathbf{H}^T\mathbf{H} & \mathbf{H}^T\mathbf{G}\\ \mathbf{G}^T\mathbf{H} & \mathbf{G}^T\mathbf{G}\end{bmatrix}^{-1}. \tag{6.14}$$

The CRB for $\hat{\theta}_1$ and $\hat{\eta}$ is given by the upper left $n_1 \times n_1$, and the lower right $(n_2 - n_1) \times (n_2 - n_1)$ block matrices of (6.14), respectively. By performing a UDL factorization [SS89], the inverse in (6.14) results in

$$\mathrm{CRB}(\theta_1) = \sigma^2 \left[\mathbf{H}^T \Pi_G^{\perp} \mathbf{H}\right]^{-1} \tag{6.15}$$

$$\mathrm{CRB}(\eta) = \sigma^2 \left[\mathbf{G}^T \Pi_H^{\perp} \mathbf{G}\right]^{-1} \tag{6.16}$$

where the projection matrix (for a matrix $\mathbf{A}$) $\Pi_A^{\perp}$, is defined by $\Pi_A^{\perp} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}$.

When the uncertainty in η is small (that is, the norm of $\Delta\eta$ is small), $\mathcal{M}_1$ is expected to result in a lower MSSE than $\mathcal{M}_2$. The essential question is how the fit is dependent on $\Delta\eta$ and for which $\Delta\eta$ $\mathcal{M}_1$ is a better model than $\mathcal{M}_2$. When using the model $\mathcal{M}_1$ the parameter vector $\hat{\theta}_1$ is estimated by [Kay93]

$$\hat{\theta}_1 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{G}\eta). \tag{6.17}$$

Now, the MSSE using model $\mathcal{M}_1$ and $\hat{\theta}_1$ according to (6.17) can be expressed as

$$\begin{aligned}\mathcal{V}(\hat{\theta}_1, \eta) &= \frac{1}{N}\,\mathrm{E}\,||\mathbf{y} - \mathbf{H}\hat{\theta}_1 - \mathbf{G}\eta||^2\\ &= \frac{1}{N}\,\mathrm{E}\,||\Pi_H^{\perp}\mathbf{y} - \Pi_H^{\perp}\mathbf{G}\eta||^2\end{aligned} \tag{6.18}$$

where for any vector $\mathbf{z}$ the vector norm is defined as $||\mathbf{z}||^2 = \mathbf{z}^T\mathbf{z}$. For small uncertainties in η, the MSSE (6.18) can be expanded in a Taylor series as

$$\begin{aligned}\mathcal{V}(\hat{\theta}_1, \Delta\eta) =& \mathcal{V}(\hat{\theta}_1, \bar{\eta}) + \left[\frac{\partial}{\partial\eta}\mathcal{V}(\hat{\theta}_1, \eta)\right]^T\Bigg|_{\eta=\bar{\eta}} \Delta\eta\\ &+ \frac{1}{2}\Delta\eta^T\left[\nabla_\eta^2\mathcal{V}(\hat{\theta}_1, \eta)\right]\Big|_{\eta=\bar{\eta}}\Delta\eta + \ldots\end{aligned} \tag{6.19}$$

where the $\nabla_\eta^2[\cdot]$ denotes the Hessian with respect to η. The first term in (6.19) is given by equation (6.10) using $\hat{\theta} = \hat{\theta}_1$. Differentiation of

(6.18) with respect to η and insertion of $\eta = \bar{\eta}$ results in (decomposing $\mathbf{y} = \mathbf{x} + \mathbf{e}$)

$$\begin{aligned}\frac{\partial}{\partial \eta}\, \mathcal{V}(\hat{\theta}_1, \eta)\Big|_{\eta=\bar{\eta}} &= -\frac{2}{N}\, \mathrm{E}\left\{\mathbf{G}^T \Pi_H^{\perp}\, (\mathbf{y} - \mathbf{G}\eta)\right\} \\ &= -\frac{2}{N}\, \mathrm{E}\left\{\mathbf{G}^T \Pi_H^{\perp} \left(\underbrace{\mathbf{H}\bar{\theta}_1 + \mathbf{G}\bar{\eta} + \mathbf{e}}_{=\mathbf{y}} - \mathbf{G}\bar{\eta} \right) \right\} \\ &= -\frac{2}{N} \mathbf{G}^T \Pi_H^{\perp}\, \mathrm{E}\left\{\mathbf{e}\right\} = 0 \end{aligned} \tag{6.20}$$

The Hessian of (6.18) with respect to η is given by

$$\nabla_{\eta}^2 \left[\mathcal{V}(\hat{\theta}_1, \eta) \right] = \frac{2}{N} \mathbf{G}^T \Pi_H^{\perp} \Pi_H^{\perp} \mathbf{G} = \frac{2}{N} \mathbf{G}^T \Pi_H^{\perp} \mathbf{G}. \tag{6.21}$$

In the last equality the fact that $\Pi_H^{\perp}$ is an idempotent matrix is used. Further, higher than second order derivatives of (6.18) are zero. Observing (6.16) it is evident that the Hessian of (6.18) is coupled to the CRB(η). Using (6.10), (6.20) and (6.21) the Taylor series expansion (6.19) can be written as

$$\mathcal{V}(\hat{\theta}_1, \Delta\eta) \simeq \sigma^2 \left(1 + \frac{n_1}{N}\right) + \frac{\sigma^2}{N} \Delta\eta^T\, \mathrm{CRB}(\eta)^{-1} \Delta\eta. \tag{6.22}$$

In terms of MSSE the goodness of the fit (6.22) using model $\mathcal{M}_1$ is given as a function of the uncertainty $\Delta\eta$ and the CRB(η). Comparing (6.22) with (6.10) for $p = n_2$ one can conclude that $\mathcal{M}_1$ should be used in favor of $\mathcal{M}_2$ when

$$\Delta\eta^T\, \mathrm{CRB}(\eta)^{-1} \Delta\eta \leq \dim(\eta). \tag{6.23}$$

In the special case of $\dim(\eta) = 1$, the results above reduces to

$$|\Delta\eta| \leq \sqrt{\mathrm{CRB}(\eta)}. \tag{6.24}$$

Intuitively the above result make sense since a large variance of $\hat{\eta}$ makes information of η more valuable, and *vice versa*, if η can be accurately estimated model $\mathcal{M}_2$ can be used instead of $\mathcal{M}_1$ without loss of performance.

6.4 Application to Sinewave-fitting

6.4.1 Sinewave-fitting with Known Frequency

Consider the sinewave model (6.4) with known frequency. Let the unknowns A and B be contained in the parameter vector θ_1 of dimension $\dim(\theta_1) = 2$. Let the DC-level C be regarded as the parameter η in $\mathcal{M}_2$, whereas it is regarded as a known variable in $\mathcal{M}_1$. With appropriate selection of the matrices $\mathbf{H}$ and $\mathbf{G}$ the model set-up coincides with the linear model in (6.12). Applying the result (6.24) one can conclude that the model $\mathcal{M}_1$ should be used in favor of $\mathcal{M}_2$ if

$$|C - \bar{C}| \leq \sqrt{\frac{\sigma^2}{N}}. \tag{6.25}$$

That is, if the uncertainty in the DC-level is small according to (6.25) it can be treated as a known constant, or if it is known to have a magnitude smaller than $\sigma/\sqrt{N}$ it can be disregarded from the model. A numerical illustration of this property is displayed in Figure 6.2.

6.4.2 Sinewave-fitting with Unknown Frequency

An extension of the previous example to the case when the frequency is unknown is made by consider the following signal model,

$$\mathbf{x}[\theta] = \mathbf{H}(\omega) \begin{bmatrix} A \\ B \end{bmatrix} + \mathbf{1}C \tag{6.26}$$

where each row in $\mathbf{x}[\theta]$ is equal to $x[n;\theta]$ in (6.4) using $\theta = (A, B, C, \omega)^T$. Here the model $\mathcal{M}_1$ is described by the parameter vector $\theta_1 = (A, B, \omega)^T$ of dimension $n_1 = 3$. The second model $\mathcal{M}_2$ is described by the parameter vector $\theta_2 = (\theta_1^T, C)^T$. Because ω enters the problem in a non-linear fashion the result in (6.24) is not directly applicable. As in the previous example the $\mathrm{CRB}(C)$ is given by σ^2/N.

The minimum variance estimator of ω given a set of measurements $\mathbf{y}$ is given by [Kay93]

$$\hat{\omega} = \ \arg\max_\omega \ \ \mathbf{y}^T \Pi_\omega \mathbf{y} - \mathbf{1}^T \Pi_\omega \mathbf{1} C^2. \tag{6.27}$$

Here, the projection matrix Π_ω is given by $\Pi_\omega = \mathbf{H}(\omega)(\mathbf{H}(\omega)^T\mathbf{H}(\omega))^{-1}\mathbf{H}(\omega)^T$. If $N \gg 2\pi/\omega$ then the second term in (6.27) $\mathbf{1}^T\Pi_\omega\mathbf{1} \approx 0$, making $\hat{\omega}$ independent of the DC-level C. Given $\hat{\omega}$ the estimate of A and B is now

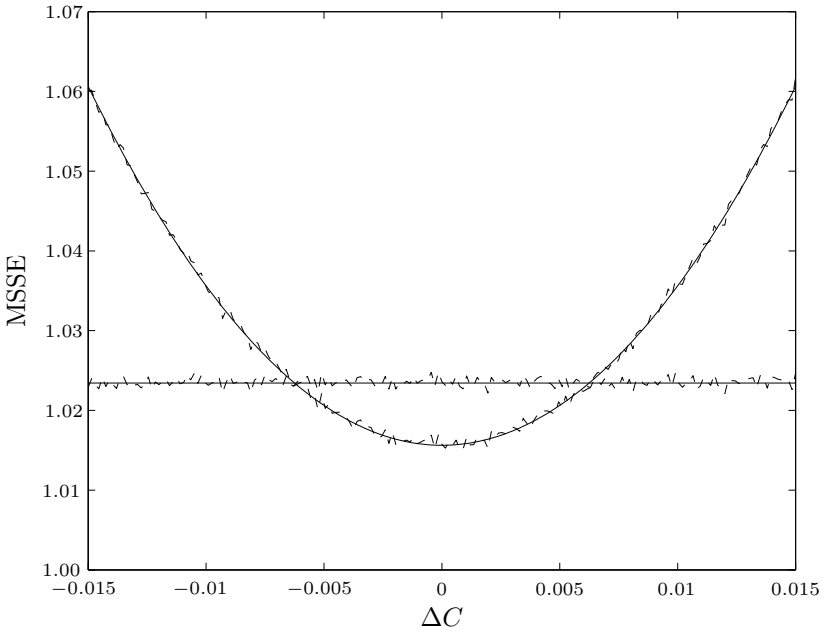


Figure 6.2: The MSSE normalized with the noise variance σ^2 for the two-parameter (A, B) case (parabola) and three-parameter (A, B, C) (straight line) case. The dashed lines are numerical evaluations of the MSSE using the different models. The curves have been obtained for $N = 128$ and SNR $= 20dB$.

given by

$$\begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} = \mathbf{H}(\hat{\omega})^T \mathbf{H}(\hat{\omega}))^{-1} \mathbf{H}(\hat{\omega})^T (\mathbf{y} - \mathbf{1}C) \tag{6.28}$$

Using (6.26) and (6.27) the MSSE $\mathcal{V}(\hat{\theta}_1, C)$ can be written similar to (6.18) as

$$\mathcal{V}(\hat{\theta}_1, C) = \frac{1}{N} \mathrm{E} \left|\left| \Pi_{\hat{\omega}}^{\perp} \mathbf{y} - \Pi_{\hat{\omega}}^{\perp} \mathbf{1}C \right|\right|^2 \tag{6.29}$$

where $\Pi_{\hat{\omega}}^{\perp} = \mathbf{I} - \Pi_{\hat{\omega}}$. The Hessian of (6.29) with respect to C is given by

$$\nabla^2 \left[\mathcal{V}(\hat{\theta}_1, C) \right] = \frac{2}{N} \mathbf{1}^T \Pi_{\hat{\omega}}^{\perp} \mathbf{1} \simeq 2. \tag{6.30}$$

This is a result that is in accordance with (6.21) and results in: Model

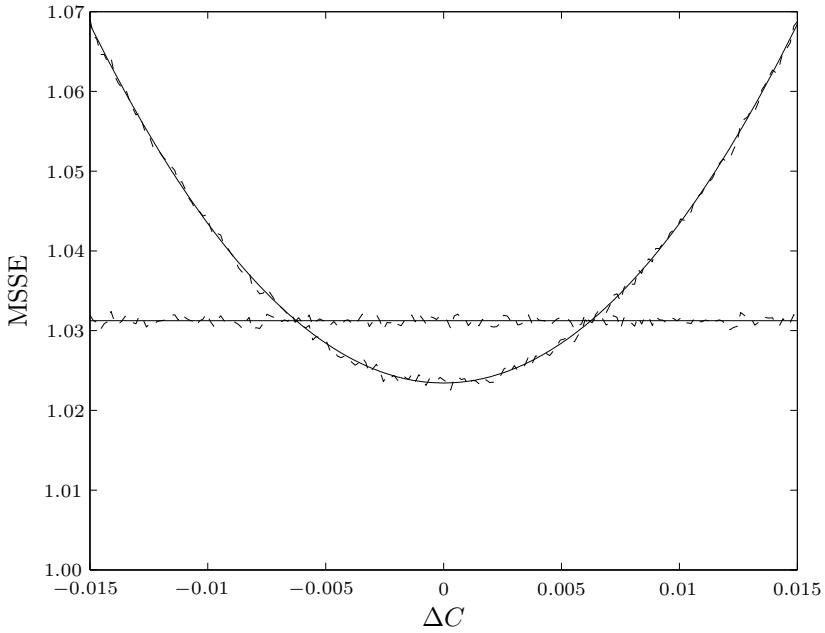


Figure 6.3: The MSSE normalized with the noise variance σ^2 for the three-parameter (A, B, ω) case (parabola) and four-parameter(A, B, C, ω) (straight line) case. The dashed lines are numerical evaluations of the MSSE using the different models. The curves have been obtained for $N = 128$ and SNR $= 20dB$.

$\mathcal{M}_1$ shall be used in favor of $\mathcal{M}_2$ if relation (6.25) holds, independent of the fact that ω is known or not. In the latter case the result holds asymptotically as $N \gg 2\pi/\omega$. However, this is a weak assumption since there is no point in estimating ω if $\omega < 2\pi N$. This limit is referred to as the Fourier resolution [Kay93]. A numerical illustration of the MSSE in the current scenario is displayed in Figure 6.3.

6.5 Conclusions

A simple criterion for model order selection has been derived (6.23). This criterion holds when the model is linear in the unknown parameters. The correction term in the MSSE (c.f the second term in (6.19)) holds exactly

for linear models. A generalization of the result to include non-linear parameters have been studied in the special case with one non-linear parameter. Numerical illustrations of the theoretical results have shown good agreement with the theoretical analysis.

From the two examples above we can conclude that estimation of the DC-level is superfluous if $|C| < \sqrt{\sigma^2/N}$ in sinewave-fitting, as well as in linear models with a CRB associated to the DC-level given by σ^2/N. The assumption under which the results have been derived includes a Gaussian requirement on the white noise additive model imperfections, and that the model $\mathcal{M}_2$ describes the signal correctly. Thus, the result may not be applicable for example in ADC-testing where the considered noise model is questionable [KB04].

Part III

Multi-Tone Sinewave Fitting

Chapter 7

Multi-Tone Parameter Estimation using IEEE 1057 Sinewave Fit and the EM-algorithm

The aim of this work is to present an efficient algorithm for multiple-tone parameter estimation. The algorithm is inspired by the expectation-maximization algorithm, and it utilizes the IEEE standard 1057 for single tone parameter estimation. In the derivation of the algorithm it is assumed that the number of tones are known and that the frequencies are well separated. The algorithm is evaluated using noisy data consisting of multiple real-valued tones. The performance of the frequency estimator is studied and compared with the asymptotic Cramér-Rao bound (CRB). It is shown that the algorithm produces statistically efficient frequency estimates at high signal to noise ratios, that is the variance of the estimates reaches the CRB. Finally, it is illustrated that the algorithm can produce efficient estimates independent of the number of tones in the input signal.

In this chapter, we consider the problem of estimating the parameters of *multiple* real-valued tones by aid of the IEEE standard 1057, in combination with the expectation-maximization (EM) algorithm. The EM algorithm was introduced and formulated in [DLR77] and has received

attention in different areas. Feder and Weinstein [FW88] have address the problem of finding multiple direction of arrival angles using the EM algorithm, that is the spatial analogy to the temporal version presented in this paper. The main advantage of the EM algorithm over alternative algorithms is that it provides an iterative solution where a multiple parameter estimation problem is decoupled into several single parameter estimation problems [FW88]. In particular, we consider a signal consisting of p superimposed tones, and the problem of estimating its parameters. This is a classic problem and it has received some attention in the past [RB76, SMFS89]. In [RB76], the parameter estimation problem was formulated and the maximum likelihood (ML) estimator was studied. The ML estimator, which is both non-linear and involves a multi dimensional search, was found hard to resolve. Methods based on the discrete Fourier transform (DFT) were proposed as an approximation of the ML estimator. However, in the case when more than one tone is present a DFT-based method will produce biased estimates. Accordingly, some windowing methods were proposed to reduce the bias [RB76]. The bias problem was solved in [SMFS89] were the true ML was derived and evaluated. An iterative Gauss-Newton algorithm was proposed as a way to minimize the ML criterion function. In the iterative algorithm proper initial estimates are essential, and the importance of them was also given some attention in [SMFS89]. The method proposed in [SMFS89] was shown to produce statistically efficient frequency estimates. However, the method is computationally intensive and requires a $(3p \times 3p)$ matrix inversion to be solved in each iteration. Here, we present a solution to the multiple-tone estimation problem that utilizes both the EM algorithm and a four-parameter fit implementation of the IEEE standard 1057. The EM algorithm is employed to decouple the problem into p separate parameter estimation problems, that is p single-tone problems. Once the signal has been decomposed a four-parameter fit can be used to efficiently find the sought parameters. The number of single-tones p is assumed to be known. This is a reasonable assumption in many applications where the number of sought waveforms is known. For unknown p, a method for estimating p is presented in [Fuc88].

7.1 Signal Model

Consider N samples of the signal $x[n]$ which consists of p number of real-valued tones contaminated by an additive measurement noise. Each tone

can be modeled as

$$s_i[n] = A_i \cos(\omega_i n) + B_i \sin(\omega_i n) + C_i, \quad n = 1, \ldots, N \tag{7.1}$$

where C_i is a DC-term, and $i = 1, \ldots, p$. The constants A_i, B_i and C_i are all assumed to be unknown. The constant angular frequencies ω_i are also considered as unknown parameters. Here, $\omega_i = 2\pi f_i / f_s$ were f_i is the signal frequency in Hertz and f_s is the sampling frequency. The model (7.1) is equivalent with modeling each tone as an amplitude- and phase shifted sinusoid, that is

$$s_i[n] = \alpha_i \sin(\omega_i n + \phi_i) + C_i \tag{7.2}$$

where $A_i = \alpha_i \sin(\phi_i)$ and $B_i = \alpha_i \cos(\phi_i)$.

The measured signal $x[n]$ is a sum of the p tones and an additional noise term $w[n]$, that is

$$x[n] = \sum_{i=1}^{p} s_i[n] + w[n], \quad n = 1, \ldots, N. \tag{7.3}$$

The noise $w[n]$ is assumed to be zero-mean white Gaussian with variance σ^2. The assumption that $w[n]$ is Gaussian may appear somewhat restrictive. However, if the Gaussian hypothesis fails to be true, the method may still be applicable but will no longer provide ML estimates. However, the estimator still yields minimum variance estimates if $w[n]$ is an i.i.d sequence [SMFS89]. If $w[n]$ is a colored sequence it is shown in [SMFS89] that an ML estimator under white Gaussian noise condition gives accurate estimates.

Further, it is assumed that the frequencies are unique and well separated, meaning that

$$|\omega_i - \omega_j| \gg \frac{1}{N} \quad \text{for all} \quad i \neq j. \tag{7.4}$$

The implication of (7.4) is that the estimation problem may be solved by Fourier-based methods in order to generate proper initial values. The proposed method does not rely on (7.4), and thus it may work for signals with closely spaced frequencies as well. In [Fuc88] it is shown that sinewaves with frequencies even closer than (7.4) can be detected. The main focus in this paper, however, is to present a practical algorithm for multiple-tone estimation, and apply it to the common problem when

(7.4) is fulfilled. In practice, if a DFT-based method is used to initialize the method, equation (7.4) reads

$$|\omega_i - \omega_j| > \frac{6\pi}{N} \quad \text{for all} \quad i \neq j. \tag{7.5}$$

This bound ensures that neighboring signal peaks in the DFT are well separated.

For convenience in the further discussion a vector notation is introduced. The signal samples $s_i[n]$ are stacked in a column vector as

$$\mathbf{s}_i = \begin{bmatrix} s_i[1] & s_i[2] & \dots & s_i[N] \end{bmatrix}^T \tag{7.6}$$

where T denotes transpose. Each tone can then be written as

$$\mathbf{s}_i = \mathbf{H}_i \theta_i \tag{7.7}$$

where $\mathbf{H}_i$ contains the signal basis functions as

$$\mathbf{H}_i = \begin{bmatrix} \cos\omega_i & \sin\omega_i & 1 \\ \cos 2\omega_i & \sin 2\omega_i & 1 \\ \vdots & \vdots & \vdots \\ \cos N\omega_i & \sin N\omega_i & 1 \end{bmatrix} \tag{7.8}$$

and the linear parameters are gathered in

$$\theta_i = \begin{bmatrix} A_i & B_i & C_i \end{bmatrix}^T. \tag{7.9}$$

Accordingly, the data model (7.3) can be written as

$$\mathbf{x} = \mathcal{H}\boldsymbol{\theta} + \mathbf{w} \tag{7.10}$$

where $\mathcal{H}$ is the $(N \times 2p+1)$-matrix

$$\mathcal{H} = \begin{bmatrix} 1 & \cos\omega_1 & \sin\omega_1 & \dots & \cos\omega_p & \sin\omega_p \\ 1 & \cos 2\omega_1 & \sin 2\omega_1 & \dots & \cos 2\omega_p & \sin 2\omega_p \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & \cos N\omega_1 & \sin N\omega_1 & \dots & \cos N\omega_p & \sin N\omega_p \end{bmatrix} \tag{7.11}$$

and $\boldsymbol{\theta}$ is the parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} C & A_1 & B_1 & A_2 & B_2 & \dots & A_p & B_p \end{bmatrix}^T. \tag{7.12}$$

In (7.12), the DC-level C has been introduced as the sum of individual DC-components, that is

$$C = \sum_{i=1}^{p} C_i. \tag{7.13}$$

For the case $p = 1$, the model (7.10)-(7.13) is used in the IEEE standards [S1057,S1241]. In the case of several tones, that is $p > 1$, it is not possible to resolve the individual C_i's. As a consequence, the DC-level (7.13) is included in the parameter vector (7.12).

The data model (7.10) is convenient to work with as all the unknown linear parameters are collected in $\boldsymbol{\theta}$. The unknown frequencies ω_i are implicitly collected in $\mathcal{H}$. For convenience they are stacked in the parameter vector $\boldsymbol{\omega}$ as

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_1 & \omega_2 & \dots & \omega_p \end{bmatrix}^T. \tag{7.14}$$

7.2 Algorithm

The aim of the proposed algorithm is to estimate the parameters in (7.12) and (7.14). In estimation methods derived from the method of ML, the criterion function to be optimized depends on the probability density function (pdf) of the data $\mathbf{x}$ given the parameters [Kay93], where

$$\mathbf{x} = \begin{bmatrix} x[1] & x[2] & \dots & x[N] \end{bmatrix}^T. \tag{7.15}$$

In the considered scenario the pdf is denoted by $p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\omega})$, that is a function of the sought $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$. Here, the pdf (and its logarithm) is both multi-dimensional and nonlinear in the sought parameters and maximization of the pdf with respect to the unknown parameters is in general difficult. Our proposal is therefor to decompose the observed data $\mathbf{x}$ into p new data sets. Let the observed data $\mathbf{x}$ be a sum of p individual data sets $\mathbf{y}_i$,

$$\mathbf{x} = \sum_{i=1}^{p} \mathbf{y}_i \tag{7.16}$$

where each set $\mathbf{y}_i$ includes only one single-tone component disturbed by noise as

$$\mathbf{y}_i = \mathbf{s}_i + \mathbf{w}_i, \quad i = 1 \dots p. \tag{7.17}$$

In (7.17), $\mathbf{w}_i$ is a decomposition of the noise term $\mathbf{w}$ in (7.10), chosen as

$$\mathbf{w}_i = \beta_i \mathbf{w}. \tag{7.18}$$

A brief discussion on how to choose the coefficients β_i is made in [FW88]. It follows directly from (7.3) and (7.16) that the coefficients β_i can be chosen arbitrary as long as they sum up to one, that is

$$\sum_{i=1}^{p} \beta_i = 1. \tag{7.19}$$

From each decomposed data set $\mathbf{y}_i$ one can find the corresponding unknown parameters θ_i and ω_i. The problem of estimating the parameters of a single tone disturbed by noise is well known, and it may be efficiently solved employing an implementation of the IEEE standard 1057 sinewave fit [S1057]. In short, the proposed algorithm consists of the following steps:

A) Find initial estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$,

B) find estimates $\hat{\mathbf{y}}_i$ of the separable signals $\mathbf{y}_i$, and

C) use $\hat{\mathbf{y}}_i$ to maximize $p_{\mathbf{y}_i}(\mathbf{y}_i; \theta_i, \omega_i)$.

Then, by iterating through the steps **B**-**C**, a local minima of $p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\omega})$ can be found. This type of algorithm is often refereed to as the EM algorithm. Step **B** corresponds to the *expectation* step in the EM algorithm, and step **C** to the *maximization* step [FW88, DLR77]. Details about the steps **A**-**C** are given below.

The initialization procedure is crucial for the convergence of the proposed method (as well as for other methods for the problem at hand). Clearly the method converges even if the separation is narrower than (7.5), *if* a proper initialization can be performed. If the method converges to the global minima its performance is expected to be close to the CRB. For well separated spectral components the CRB is approximately given by the single tone CRB [Hän00], whereas for closely spaced components approximate expressions for the CRB may be found in [Swi95]. These approximate CRB expressions may be used as a guideline on algorithm error variance as function of frequency separation.

7.2.1 Initial Estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$

The initial estimates of the sought parameters can be found by performing an N-points DFT. If N is chosen as a power of two, fast Fourier transforms (FFT) methods can be employed, hence reducing the computational complexity. The p largest peaks in the magnitude of the DFT are

identified, with the DC-level excluded. The peak locations are chosen as initial estimates of the frequencies. Amplitude, phase and C_i estimates are all obtained from the DFT as

$$\hat{\theta}_i = \begin{bmatrix} \hat{A}_i \\ \hat{B}_i \\ \hat{C}_i \end{bmatrix} = \frac{1}{N} \begin{bmatrix} 2 \cdot \mathrm{Re}\{X(k_i)\} \\ -2 \cdot \mathrm{Im}\{X(k_i)\} \\ \frac{1}{p} X(0) \end{bmatrix}, \quad i = 1, \ldots, p \tag{7.20}$$

where $X(k)$ is the N-points DFT of the input signal $\mathbf{x}$ and k_i is the frequency bin corresponding to the actual peak of $|X(k)|$. The quantities $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part of the quantity between the brackets, respectively. The initial value of the C_i parameter is rather arbitrary obtained by dividing the DC-level of $\mathbf{x}$ into p equal parts.

At high SNR, the performance of the initial frequency estimates is given by the grid size of the DFT. With N samples, the error $\hat{\omega}_i - \omega_i$ is typically uniformly distributed over an interval of length $2\pi/N$, implying that the performance is approaching a mean-squared-error (MSE) of

$$\mathrm{MSE}(\hat{\omega}_i) = \mathrm{E}\{(\hat{\omega}_i - \omega_i)^2\} = \frac{\pi^2}{3N^2}. \tag{7.21}$$

In a practical case with $N = 128$ this corresponds to a MSE of -37dB. Standard methods such as zero-padding or interpolation may be used in order to increase the resolution of the DFT calculations.

The approach taken is to identify the dominating peaks as spectral components. The number of significant (in some sense) peaks has to coincide with p, in order to avoid under- or over-modeling. In addition, one should note that the proposed method does not incorporate prior knowledge on harmonic or folded components, and thus all spectral peaks are assumed independent. Designing a method utilizing a harmonic or folded signal model is straightforward [NP86] [PS96] and such an algorithm is expected to have superior performance in terms of error variance, but worse robustness properties.

In general, over-fitting the number of parameters is preferred over under-fitting. If p is larger than the actual number of tones including harmonic components, the strongest noise peaks present will be modelled as sinewave components. However, this over-fitting does not degrade the performance in large samples, other than increasing the computational load. On the other hand, if p is less than the actual number of tones the accuracy of the frequency estimates is degraded [RB76]. Accordingly, a

proper strategy is to select a large p in combination with a procedure for detection of over-fitting. An applicable procedure can be found in [SHS94].

7.2.2 Estimates of $\hat{\mathbf{y}}_i$

When an initial estimate of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ has been obtained, the ML estimate of $\mathbf{y}_i$ is given by [FW88]

$$\hat{\mathbf{y}}_i = \hat{\mathbf{H}}_i\hat{\theta}_i + \beta_i \left[\mathbf{x} - \sum_{\ell=1}^{p} \hat{\mathbf{H}}_\ell\hat{\theta}_\ell\right] \tag{7.22}$$

where $\hat{\mathbf{H}}_i$ and $\hat{\theta}_i$ are estimates of $\mathbf{H}_i$ in (7.8) and θ_i in (7.9), respectively. The second term in (7.22) can be interpreted as an estimate of $\mathbf{w}_i$. The coefficient β_i is related to the decomposition of $\mathbf{w}$, as given by (7.18). One can choose β_i arbitrary as long as (7.19) is fulfilled. One way of choosing the β_i's is to make the signal-to-noise ratio (γ_i) equal for all signals $\hat{\mathbf{y}}_i$, where γ_i is defined as

$$\gamma_i = \frac{\alpha_i^2}{2\sigma_i^2} = \frac{A_i^2 + B_i^2}{2\sigma_i^2} = \frac{\theta_i^T\theta_i}{2\sigma_i^2}. \tag{7.23}$$

Here σ_i^2 is the variance of the noise term $\mathbf{w}_i$ in the separated signal $\mathbf{y}_i$ (7.17). The relation between the variance of $\mathbf{w}_i$ and $\mathbf{w}$ is obtained by evaluating the variance of both sides of (7.18), that is

$$\sigma_i^2 = \beta_i^2\sigma^2. \tag{7.24}$$

In order to determine the constants β_i we first manipulate (7.23) and (7.24), giving us

$$\frac{\beta_i}{|\alpha_i|} = \sqrt{\frac{1}{2\gamma_i\sigma^2}}. \tag{7.25}$$

Since SNR_i is chosen to be constant and independent of i, it follows that (7.25) also is constant. From (7.19) it follows that

$$1 = \sum_i \beta_i = \sum_i \underbrace{\left(\frac{\beta_i}{|\alpha_i|}\right)}_{\text{independent of } i} |\alpha_i| \Rightarrow \beta_i = \frac{|\alpha_i|}{\sum_i |\alpha_i|}. \tag{7.26}$$

When forming the signal $\mathbf{y}_i$ from (7.22) the β_i is replaced with an estimate $\hat{\beta}_i$. This estimate is formed by replacing the quantity $|\alpha_i|$ in (7.26) with its estimated equivalence $\sqrt{\hat{\theta}_i^T\hat{\theta}_i}$.

7.2.3 Maximization of $p_{\hat{\mathbf{y}}_i}(\hat{\mathbf{y}}_i; \theta_i, \omega_i)$

After performing the decomposition of $\mathbf{x}$ as described above, the signal $\hat{\mathbf{y}}_i$ consists of one *strong* tone with parameters near the true θ_i and ω_i. After applying the IEEE standard 1057 sinewave fit on the decoupled signals, estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ are obtained. The obtained estimates are used to decouple the data once again. Thus, improving the estimates of the single-tones ω_i and θ_i. This iteration may continue until the estimates have converged. The full algorithm is summarized in Table 7.1. In Table 7.1, the level of convergence is measured by the Euclidean norm of the frequency update vector $\tilde{\boldsymbol{\omega}}$, that is

$$\left|\tilde{\boldsymbol{\omega}}^{(r)}\right| = \left|\hat{\boldsymbol{\omega}}^{(r)} - \hat{\boldsymbol{\omega}}^{(r-1)}\right|. \tag{7.27}$$

The iterations continue as long as $|\tilde{\boldsymbol{\omega}}^{(r)}|$ is larger than a constant ε. Here, r corresponds to the current iteration step, see Table 7.1. In the case where the algorithm fails to resolve the frequencies, an upper limit on the number of iterations should be set. This happens, for example, when the SNR is below the SNR-threshold that occurs in non-linear estimation problems.

---

a) Find initial estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$. Denote them $\hat{\boldsymbol{\theta}}^{(0)}$ and $\hat{\boldsymbol{\omega}}^{(0)}$.
b) Let $r = 0$ and $\tilde{\boldsymbol{\omega}}^{(r)} = 1$.
c) While $|\tilde{\boldsymbol{\omega}}^{(r)}| > \varepsilon$
d) Build $H_1 \dots H_p$ using $\hat{\boldsymbol{\theta}}^{(r)}$ and $\hat{\boldsymbol{\omega}}^{(r)}$ according to (7.8).
e) Build p data sets $\hat{\mathbf{y}}_1 \dots \hat{\mathbf{y}}_p$ according to (7.22).
for $i = 1 \dots p$
f) Estimate θ_i and ω_i from $\hat{\mathbf{y}}_i$ using IEEE standard 1057 [S1057].
end
g) Form $\hat{\boldsymbol{\theta}}^{(r+1)}$ and $\hat{\boldsymbol{\omega}}^{(r+1)}$ using outcomes from step **f)**.
h) Form $\tilde{\boldsymbol{\omega}}^{(r+1)}$ according to (7.27).
i) Let $r = r + 1$.
j) end

Table 7.1: Multi-Tone Parameter Estimation by aid of the IEEE Standard 1057 and the EM Algorithm.

7.3 Simulation Examples

The algorithm in Table 7.1 has been evaluated in three different scenarios (case *i*)- *iii*), below). The number of tones in each scenario is chosen as: In case *i*) $p=2$, in case *ii*) $p=10$ and in case *iii*) $p=3$, respectively. In all cases, the signal has been disturbed by zero-mean white Gaussian noise and the number of samples is chosen to $N=128$. The amplitudes $\{\alpha_i\}$ in the first two cases equal to 1. In the third case $\alpha_1 = 1$, $\alpha_2 = 10^{-1/4}$ and $\alpha_3 = 10^{-1/2}$. With the latter choice of amplitudes, the individual SNRs are related as $\mathrm{SNR}_1 = \mathrm{SNR}_2 + 5\mathrm{dB}$ and $\mathrm{SNR}_2 = \mathrm{SNR}_3 + 5\mathrm{dB}$, where the individual SNR_i is defined by

$$\mathrm{SNR}_i = \frac{\alpha_i^2}{2\sigma^2}. \tag{7.28}$$

The initial phases $\{\phi_i\}$ have been drawn from a uniform distribution within the interval $[0, 2\pi)$. Although the algorithm provides estimates of all unknown parameters, the main purpose in this paper is to investigate the performance of the frequency estimates. The unknown frequencies $\{\omega_i\}$ have been drawn from a uniform distribution within the interval $[0, \pi)$, conditionally that (7.5) is fulfilled. The noise variance σ^2 of $w[n]$ in (7.3) is varied in such a way that the local SNR_i (7.28) is varied in the range -20 to 70dB. In case *iii*) where the signal amplitude varies from tone to tone, the variance σ^2 has been chosen in such a way that the SNR_1 (that is, the local SNR of the strongest signal) varies in the range -20 to 70dB.

The algorithm is implemented as described in Table 7.1. The number of iterations in the IEEE standard 1057 sinewave fit was fixed to four. In step **c)**, (according to Table 7.1) ε was set to $1 \cdot 10^{-7}$. The number of iterations carried out in the loop **c)**–**j)** varied, depending on the actual SNR and the number of tones. In the case *i*), $4-8$ iterations were enough, and in the case *ii*) and *iii*)$10 - 15$. If the algorithm failed to converge the iterations were stopped after 40 in both cases. For SNRs lower than -5dB, the maximum number of iterations was commonly reached.

In Figures 7.1-7.3, the evaluated performance in terms of empirical MSE is shown, based on 5×10^4 independent runs. Here the parameter of interest is the frequency. As a comparison, the asymptotic Cramér Rao bound (CRB) is included in the figures [Kay93]. The CRB is a lower bound on the variance of an unbiased frequency estimate. The exact CRB is highly frequency dependent, whereas the asymptotic CRB is frequency

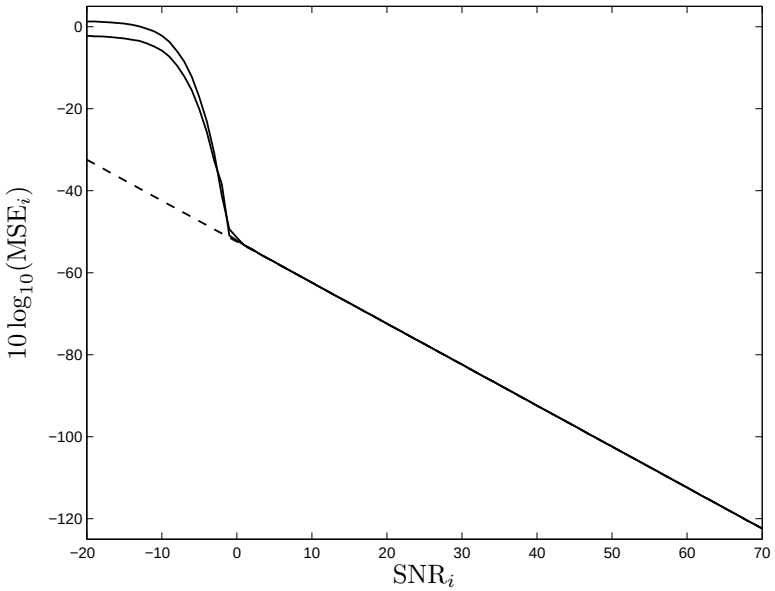


Figure 7.1: Performance of the algorithm when estimating the frequencies of two well separated tones-case *i*) The number of data samples is $N = 128$. Empirical MSE_i (solid lines) as a function of the SNR_i, for $i = 1, 2$. The asymptotic CRB (dashed line) is given as a reference.

independent and given by [Hän00],

$$\mathrm{var}\{\hat{\omega}_i\} \geq \frac{12}{N^3 \mathrm{SNR}_i}. \tag{7.29}$$

From the numerical evaluations, presented in Figure 7.1 and 7.2, one can see that the variance of the frequency estimates reaches the CRB for SNRs above a certain threshold value. Also in case *iii*) where the amplitudes are varied, the frequency estimates reach the CRB. Here, it can be seen that for the second strongest tone ($i = 2$) the MSE is increased by 5dB compared with the MSE of the strongest tone, which is in fully accordance with (7.29). Hence, the proposed estimator produces statistically efficient estimates in the considered examples, above the threshold SNR. From Figures 7.1 and 7.2 it is also shown that the SNR-threshold below which the frequency estimates are deteriorated and the algorithm

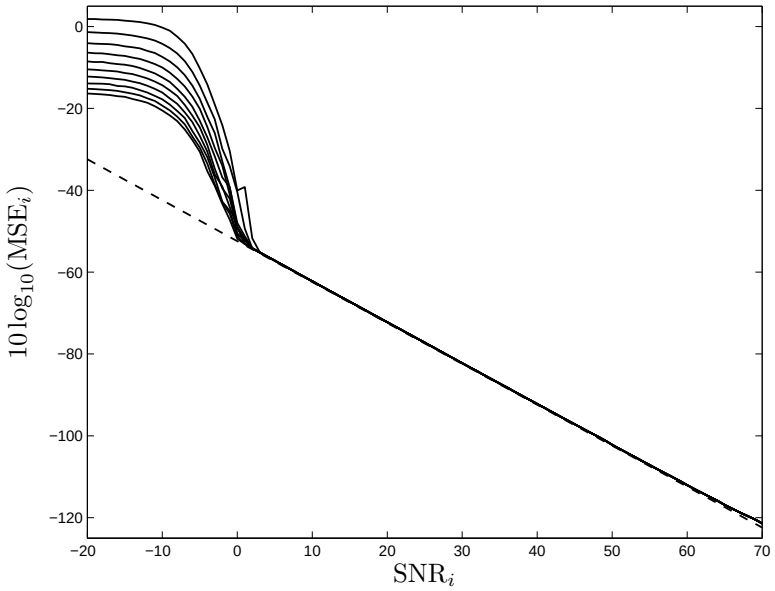


Figure 7.2: Performance of the algorithm when estimating the frequencies of ten well separated tones–case *ii*). The number of data samples is $N = 128$. The individual empirical MSE_i (solid lines) as a function of the local SNR_i, for $i = 1, \ldots, 10$. The asymptotic CRB (dashed line) is given as a reference.

fails to resolve the different frequencies is independent on the number of sinewaves. In Figure 7.3, the SNR-threshold is located at the local SNR_i equal to -5dB. In a single-tone frequency estimation scenario it is known that the SNR-threshold is only dependent on the number of data samples N [SB85]. In the case of $N = 128$ the theoretical threshold is -5dB according to [SB85], a result that is fully in accordance with the numerical results achieved by the algorithm presented in this paper.

7.4 Conclusions

In this paper, a novel algorithm for multiple-tone parameter estimation has been proposed. By numerical evaluation it is shown that the method

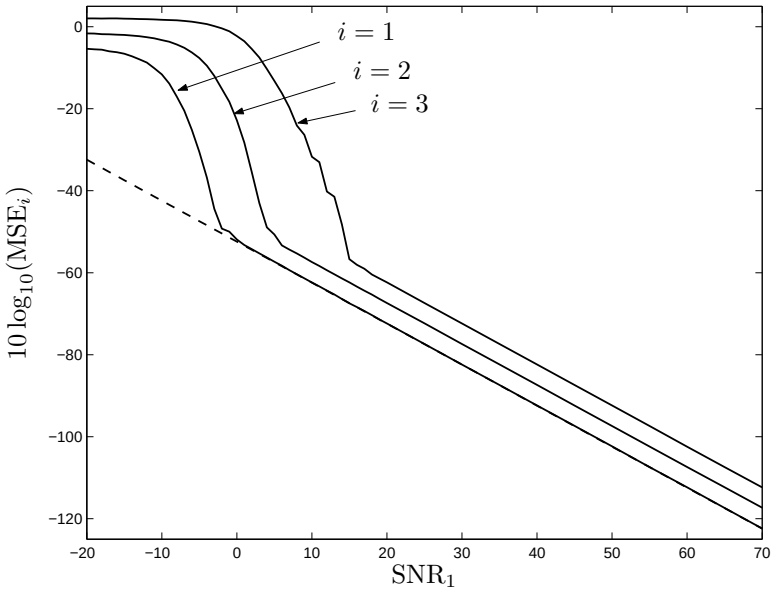


Figure 7.3: Performance of the algorithm when estimating the frequencies of three well separated tones–case *iii*). Here, the amplitudes are $1(i = 1)$, $10^{-1/4}(i = 2)$ and $10^{-1/2}(i = 3)$. The number of data samples is $N = 128$. Empirical MSE_i (solid line) as a function of the local SNR of the strongest tone (SNR_1). The asymptotic CRB (dashed line) corresponding to SNR_1 is given as a reference.

is statistically efficient for high SNR, i.e. the frequency estimation MSE coincides with the asymptotic CRB. The algorithm handles several tones, as long as they are separated in frequency, without degrading the performance on the estimated parameters. The SNR-threshold below which the algorithm fails to resolve the frequencies is independent on the number of sinewaves present in the input signal and depends only on the number of data samples N. The numerical complexity is approximately linear in the number of tones p. In the case of $p=1$ the complexity is given by the complexity of the employed implementation of the IEEE standard 1057 sinewave fit.

Chapter 8

Toward a Standardized Multiple-Sinewave Fit Algorithm

Sinewave test methods have for a long period of time been dominating in testing digital devices. The extraction of the parameters of a single tone is well known, and there exists a standardized method [S1057,S1241]. There are several physical parameters that can not be measured by single-sinewave tests. One example is the inter-modulation test for analog to digital converters (ADCs) [S1057, S1241]. In the inter-modulation test, one must not only estimate the parameters of the two tones that excite the ADC, but also the number of harmonics introduced by the nonlinearity of the ADC.

Unfortunately, it is more difficult from a measurement to resolve multiple sinewaves than a single one. If one ignores the fact that several sinewaves are present and use an estimator designed for a single tone the estimates get biased due to spectral leakage [RB76]. Several researchers in the area have tried to resolve this problem in several different ways. One attractive method is to use the maximum likelihood (ML) estimator [Kay93]. However, there exists no closed form solution for the multi-tone model due to the highly nonlinear criterion function. There exist some iterative approaches to solve the ML problem, among others the ones in [SMFS89] and the method presented in Chapter 7, respectively. The accuracy of these methods reaches the corresponding Cramér-Rao

bound (CRB). The one presented in Chapter 7, however, has superior SNR threshold than the method in [SMFS89].

The aim of this chapter is to review some results on frequency estimation and present their implications on the instrumentation and measurement set-up. Further, a generalization of the four-parameter fit of [S1057, S1241] is presented, that is able to estimate the $3p + 1$ parameters of a multi-tone model.

8.1 Requirements on a Sinewave-fit Algorithm

The multi-tone estimation problem may be separated into subproblems like signal detection, algorithm initialization and parameter extraction. In the instrumentation and measurement set-up, we are typically interested in detection of line spectral components above the noise floor (that is, the spurious frequencies), as well as the level of the noise floor itself.

8.1.1 Cramér-Rao Bound and Signal Model

In the literature, a complex-valued signal model is often employed. At first glance, this complex-valued model seems to make the analysis more complicated than using a real-valued model. In fact, the opposite holds true that can be seen, for example, from the exact CRB for a single complex-valued exponential signal (or, cisoid) disturbed by additive white Gaussian noise. A well known result on the achievable accuracy of any unbiased estimator of the normalized angular frequency ω for evenly sampled data is given by [RB74]

$$\mathrm{Var}(\hat{\omega}) \geq \frac{6}{\mathrm{SNR}\, N^2\,(N-1)} \tag{8.1}$$

where $\hat{\omega}$ denotes an estimate of ω. Here and from now on, N denotes the number of available samples. Further, SNR denotes the signal-to-noise ratio. In the real-valued case, the corresponding CRB is more complicated and results in an expression that is dependent on the signal frequency and the initial phase [Hän00]. However, an asymptotic (as $N \to \infty$) expression is known to be

$$\mathrm{Var}(\hat{\omega}) \geq \frac{12}{\mathrm{SNR}\, N^3}. \tag{8.2}$$

Moreover, in a test environment without I/Q-modulation there exclusively exist real-valued signals which make a complex-valued signal model improper. In this paper, the following multi-sinewave signal model is employed

$$s[n] = C + \sum_{\ell=1}^{p} A_\ell \cos \omega_\ell t_n + B_\ell \sin \omega_\ell t_n, \quad n = 1, \ldots, N \tag{8.3}$$

where t_n denotes the (normalized) sampling instants. The parameters A_ℓ, B_ℓ and C are all assumed to be unknown constants. The constant angular frequencies ω_ℓ are also considered as unknown parameters. Here, $\omega_\ell = 2\pi f_\ell / f_s$ where f_ℓ is the signal frequency in Hertz and f_s is the sampling frequency. In (8.3), regular sampling at f_s Hertz corresponds to an integer $t_n = n$. The number of sinewaves p is assumed to be known. This is a reasonable assumption in many applications where the number of sought waveforms is known. Estimation of p is further discussed in Section 8.2.1. The model (8.3) is equivalent with modeling each tone as an amplitude- and phase shifted sinusoid, that is

$$s[n] = C + \sum_{\ell=1}^{p} \alpha_\ell \sin(\omega_\ell t_n + \phi_\ell) \tag{8.4}$$

where $A_\ell = \alpha_\ell \sin \phi_\ell$ and $B_\ell = \alpha_\ell \cos \phi_\ell$.

The measured signal $x[n]$ is a sum of the signal (8.3) and an additional noise term $w[n]$, that is

$$x[n] = s[n] + w[n], \quad n = 1, \ldots, N. \tag{8.5}$$

The noise is assumed to be zero-mean white Gaussian with variance σ^2. The assumption that $w[n]$ is Gaussian is restrictive, but is, on the other hand, only used in order to assess the performance of the algorithm by a comparison with the CRB. If the Gaussian noise hypothesis fails it is shown in [SJL97] that a least-squares fit asymptotically results in an efficient estimator. Further, if the covariance matrix of the estimates only depends on the second order statistics of the data, a least-squares fit will result in the minimum-variance estimate [SMFS89].

8.1.2 Frequency Resolution

Frequency resolution is an important topic that has to be discussed in some detail. In general terms, resolution of two line spectral components

is a function of the separation between them, the amplitudes, as well as the level of the noise floor. We define frequency resolution as the minimum angular frequency separation between two neighboring sinewaves $\Delta\omega = |\omega_i - \omega_j|$ for which both spectral components can be detected and then estimated accurately. We emphasize that frequency resolution in model based estimation differs from the Fourier or Rayleigh resolution, roughly determined by [Kay88]

$$\Delta\omega = \frac{2\pi}{N}. \tag{8.6}$$

Accordingly, proper high-resolution methods resolve line spectral components within the Fourier resolution. Accurate estimation of closely-spaced tones has been presented in [Fuc94]. At SNR = 10dB and for $N = 100$, the method in [Fuc94] is shown to resolve two equipowered sinewaves with frequencies as close as $\Delta\omega = \pi/(2N)$, that is the fourth of the Fourier resolution (8.6). Further, the method is shown to resolve two sinewaves whose amplitudes differ by 10dB at half the Fourier resolution.

8.1.3 Performance of Multi-tone Methods

The performance one can expect from a proper method is a relevant topic. The performance depends on several causes, but the two most important items are to detect the correct number of tones and the initialization procedure for the fine-tuning of the parameter estimator. If detection and initialization are performed in a correct and successful way the error variance of the overall method is expected to be close to the CRB. If the spectral components are well separated in frequency each parameter estimate is expected to reach its corresponding CRB, that for the frequency approximately coincides with the single tone CRB in (8.2). However, if two sinewaves are closely located in frequency the single tone assumption is not valid. In [Swi95], it is shown that if the frequency separation $\Delta\omega$ is larger than about 1.5 times the Fourier resolution (8.6) then the CRB in the dual tone case basically coincides with (8.2).

8.2 A generalized IEEE 1057 algorithm

In Chapter 7, an algorithm for multi-tone parameter estimation based on the IEEE Standard 1057 four-parameter fit in combination with the expectation-maximization (EM) algorithm is presented. The EM-algorithm

was employed to separate the measurements into p single tone components that were the input to p parallel four-parameter fits. Here, an alternative to that method is derived, given by an extension of the four-parameter fit to a $3p$+1-parameter fit, where p denotes the number of spectral components.

The proposed algorithm can be divided into two steps. First, the individual spectral components are detected and initial estimates are formed. The second step increases the accuracy of the estimates by successive iterations. Each step is crucial in order to obtain a fully automated multi-sinewave estimator. In this Section the attention is concentrated on the second iterative fine-tuning step.

8.2.1 A procedure for algorithm initialization

A general issue in non-linear parameter estimation is the threshold-effect that occurs at a certain SNR, below witch the estimates are deteriorated. In the considered case, this threshold depends on the number of data samples [SB85]. In [Kno97], an indicator quantity $\gamma = N\,\mathrm{SNR}/\log_e N$ was introduced. It was shown (by aid of the Barankin bound) that $\gamma \geq 70$ always pull the single tone ML estimator out of the threshold region. The initialization can be performed by searching for the p largest peaks in the periodogram. There exist fast and efficient methods to compute the periodogram and therefore this is an attractive approach. A drawback with a periodogram-based method is the poor frequency resolution, determined by the Rayleigh resolution. Another drawback with the periodogram is that a strong sinewave shadows weaker ones. This masking effect results in an inferior frequency resolution than (8.6). When the periodogram is successful in resolving the individual sinewaves one may expect an estimation accuracy resulting in a mean-squared-error (MSE) of (for $\ell = 1, \ldots, p$)

$$\mathrm{MSE}(\hat{\omega}_\ell) = \mathrm{E}\{(\hat{\omega}_\ell - \omega_\ell)^2\} = \frac{\pi^2}{3N^2}. \tag{8.7}$$

The MSE in (8.7) results in a root-MSE (RMSE) of order $2/N$. The RMSE indicates the size of the minimum convergence radius of any iterative algorithm used for fine-tuning of the estimates.

Given initial values $\{\hat{\omega}_\ell\}$, estimates of the unknown $\{A_\ell\}$, $\{B_\ell\}$ and C can be found by solving a linear system of equations. Using a vector notation, the signal $s[n]$ in (8.3) can be written as

$$s = \mathbf{H}\theta, \tag{8.8}$$

where s is the signal vector

$$s = \begin{bmatrix} s[1] & \dots & s[N] \end{bmatrix}^T, \tag{8.9}$$

(T denotes transpose) and θ the parameter vector

$$\theta = \begin{bmatrix} C & A_1 & B_1 & A_2 & B_2 & \dots & A_p & B_p \end{bmatrix}^T. \tag{8.10}$$

Further, $\mathbf{H}$ is a $N \times 2p+1$ matrix given by

$$\mathbf{H} = \begin{bmatrix} 1 & \cos\omega_1 t_1 & \sin\omega_1 t_1 & \dots & \cos\omega_p t_1 & \sin\omega_p t_1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & \cos\omega_1 t_N & \sin\omega_1 t_N & \dots & \cos\omega_p t_N & \sin\omega_p t_N \end{bmatrix}. \tag{8.11}$$

Given initial estimates $\{\hat{\omega}_\ell\}$ for $\ell = 1, \dots, p$ the in least-squares sense optimal estimate of θ (8.10) is given by (if the matrix $\mathbf{H}^T\mathbf{H}$ is invertible)

$$\hat{\theta} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T x. \tag{8.12}$$

In (8.12), $\mathbf{H}$ is formed by plugging in the $\{\hat{\omega}_\ell\}$ into (8.11), and the vector x contains the measurements, that is

$$x = \begin{bmatrix} x[1] & \dots & x[N] \end{bmatrix}^T. \tag{8.13}$$

8.2.2 A $3p+1$ parameter fit algorithm

The algorithm is a generalization of the four-parameter sinewave fit algorithm [S1057, S1241] to handle p-tone data. In the four-parameter sinewave fit the nonlinear signal is linearized around the previous frequency estimate resulting in a linear signal model. Given the frequency estimates $\hat{\omega}_\ell^{(r)}$ at the iteration step r, a Taylor series expansion around $\hat{\omega}_\ell^{(r)}$ can be performed as

$$\cos\omega_\ell t_n \approx \cos\hat{\omega}_\ell^{(r)} t_n - t_n \sin\hat{\omega}_\ell^{(r)} t_n \cdot \Delta\omega_\ell^{(r)} \tag{8.14}$$

and

$$\sin\omega_\ell t_n \approx \sin\hat{\omega}_\ell^{(r)} t_n + t_n \cos\hat{\omega}_\ell^{(r)} t_n \cdot \Delta\omega_\ell^{(r)}, \tag{8.15}$$

where $\Delta\omega_\ell^{(r)} = \omega_\ell - \hat{\omega}_\ell^{(r)}$. Inserting (8.14) and (8.15) in (8.3) results in the approximate signal model

$$\begin{aligned} s[n;\vartheta_r] \approx C^{(r)} + \sum_{\ell=1}^{p} & A^{(r)} \cos\hat{\omega}_\ell^{(r)} t_n + B^{(r)} \sin\hat{\omega}_\ell^{(r)} t_n + \\ & - \hat{A}^{(r-1)} \Delta\omega_\ell^{(r)} t_n \sin\hat{\omega}_\ell^{(r)} + \hat{B}^{(r-1)} \Delta\omega_\ell^{(r)} t_n \cos\hat{\omega}_\ell^{(r)}, \end{aligned} \tag{8.16}$$

where the two estimates $\hat{A}^{(r-1)}$ and $\hat{B}^{(r-1)}$ from iteration $(r-1)$ have been inserted in the last two terms. In (8.16), the multi-sinewave parameter vector ϑ_r is given by

$$\vartheta_r = \begin{bmatrix} C^{(r)} & A_1^{(r)} & B_1^{(r)} & \Delta\omega_1^{(r)} & \dots & A_p^{(r)} & B_p^{(r)} & \Delta\omega_p^{(r)} \end{bmatrix}^T. \quad (8.17)$$

Using the vector ϑ_r, (8.16) can be written using matrix notation as

$$s[\vartheta_r] \approx \mathbf{D}_r \vartheta_r. \quad (8.18)$$

The matrix $\mathbf{D}_r$ forming the set of basis functions is given by

$$\mathbf{D}_r = \begin{bmatrix} \mathbf{1} & D_1^{(r)} & \dots & D_p^{(r)} \end{bmatrix} \quad (8.19)$$

where the vector $\mathbf{1} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$ is of length N, and the submatrices $D_\ell^{(r)}$ are given by

$$D_\ell^{(r)} = \begin{bmatrix} \cos\omega_\ell^{(r)} t_1 & \sin\omega_\ell^{(r)} t_1 & -A_\ell^{(r-1)} t_1 \cos\omega_\ell^{(r)} t_1 + B_\ell^{(r-1)} t_1 \sin\omega_\ell^{(r)} t_1 \\ \vdots & \vdots & \vdots \\ \cos\omega_\ell^{(r)} t_N & \sin\omega_\ell^{(r)} t_N & -A_\ell^{(r-1)} t_N \cos\omega_\ell^{(r)} t_N + B_\ell^{(r-1)} t_N \sin\omega_\ell^{(r)} t_N \end{bmatrix}. \quad (8.20)$$

The least-squares solution of ϑ_r can be computed according to

$$\vartheta_r = \left(\mathbf{D}_r^T \mathbf{D}_r\right)^{-1} \mathbf{D}_r^T x. \quad (8.21)$$

For each iteration of (8.21) the frequency estimates $\{\hat{\omega}_\ell^{(r)}\}$ are updated according to

$$\omega_\ell^{(r+1)} = \omega_\ell^{(r)} + \Delta\omega_\ell^{(r)}, \quad \ell = 1, \dots, p. \quad (8.22)$$

The iterations are stopped when sufficient precision is reached for some $\{\varepsilon_\ell\}$ according to $|\Delta\omega_\ell^{(r)}| \leq \varepsilon_\ell$ for all $\ell = 1, \dots, p$. One should note that a direct calculation of (8.21) may be numerically imprecise and from an implementation point of view it is recommended to use some matrix factorization algorithm [BMS+02].

8.3 Numerical evaluations

The proposed algorithm has been evaluated in four different scenarios, all based on equidistant sampling $t_n = 1, 2, \ldots$. The first two scenarios (a and b) have been evaluated using the initialization routine in Section 8.2.1. In the latter two scenarios (c and d) the iterative algorithm is investigated (for simplicity) when initialization is performed around the true frequency, according to $\hat{\omega}_\ell = \omega_\ell + u$, where u is a random variable with uniform distribution in the interval $[-1/(2N), 1/(2N)]$. Here, the initialization is guaranteed to be within the expected resolution of a periodogram based initialization method. The number of data samples N is set to $N = 128$ in all scenarios. The normalized angular frequencies $\{\omega_\ell\}$ have been drawn from a uniform distribution within the interval $[0, \pi)$, which corresponds to a frequency between 0 and $f_s/2$ Hertz. The initial phases $\{\phi_\ell\}$ are chosen from a uniform distribution within the interval $[0, 2\pi)$. The amplitude settings in each scenario are given in Table 8.1.

	p	α_ℓ
a)	8	$1 \ \forall \ \ell$
b)	3	$\{1, 10^{-1/4}, 10^{-1/2}\}$
c)	8	$1 \ \forall \ \ell$
d)	3	$\{1, 10^{-1/4}, 10^{-1/2}\}$

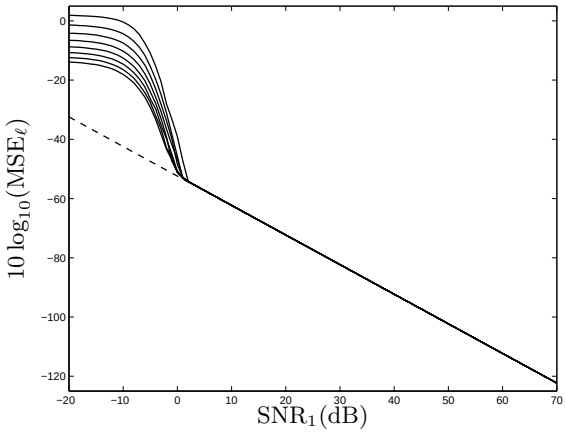
Table 8.1: Parameter values used in the numerical evaluations.

The results (based on 10^5 independent runs) are presented in Figure 8.1 and Figure 8.2. For high SNRs, each frequency estimate has a corresponding variance that coincides with the CRB (8.2). For low SNRs, in cases a)–b) there is a distinct threshold where the algorithm fails to resolve the frequencies, at about SNR = 0dB. However, this threshold is inherent from the initialization where the periodogram method is used. In cases c)–d), the threshold is not that distinct as in cases a)–b), and accurate frequency estimates is obtained for SNRs below 0dB.

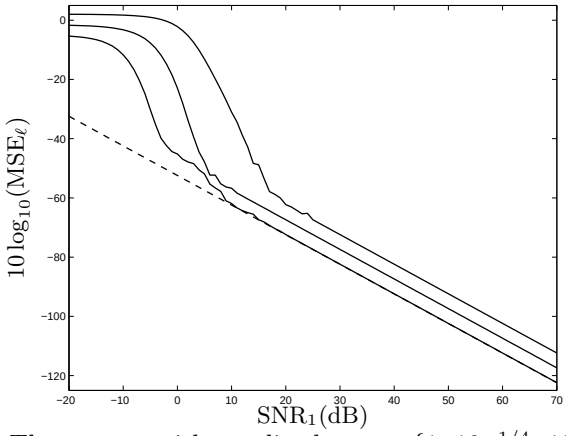
In Figures 8.1(b) and 8.2(b), it is observed that the proposed algorithm enables to find proper frequency estimates for signals where the amplitudes differ as much as 10dB, without affecting the performance. The threshold effect in Figure 8.1(b) is not as distinct as in Figure 8.2(b), making it most probable a result of the particular initialization procedure used.

8.4 Summary

A real-valued multi-sinewave model has been proposed and the performance of a novel method has been discussed. The four-parameter sinewave fit of [S1057,S1241] has been generalized to a multi-sinewave fit using $3p + 1$ parameters. Its performance has been studied by numerical evaluations. The initialization of the algorithm has been briefly discussed and its influence on the performance of the proposed algorithm has been somewhat illustrated.

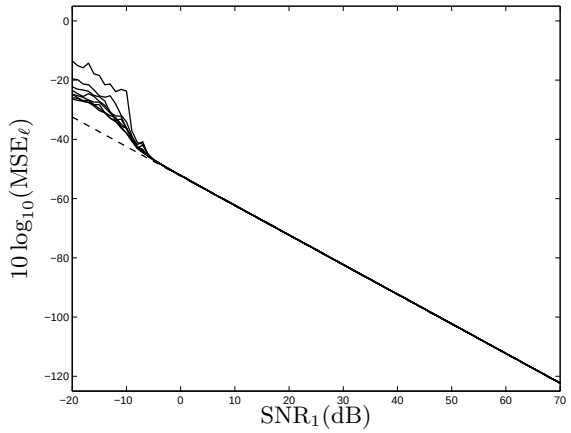


(a) – Eight tones with amplitudes $\alpha_\ell = 1$ for all ℓ.

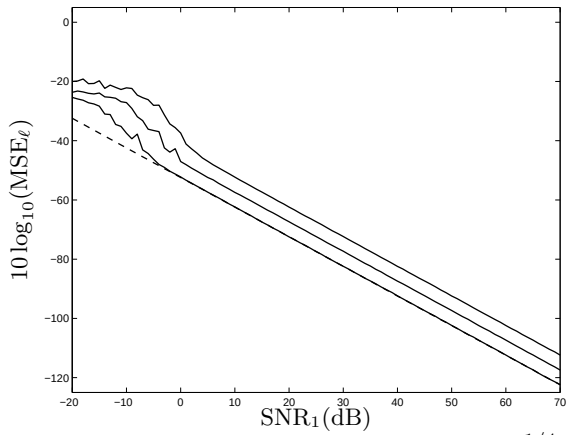


(b) – Three tones with amplitudes $\alpha_\ell \in \{1,\, 10^{-1/4},\, 10^{-1/2}\}$.

Figure 8.1: MSE of the algorithm when estimating the frequencies. The number of data samples is $N = 128$. Initialization is performed using the periodogram method. The individual empiric MSE_ℓ (solid line) as a function of the $\mathrm{SNR}_1 = \alpha_1/(2\sigma^2)$. The asymptotic CRB corresponding to SNR_1 (dashed line) is given as a reference.



(a) – Eight tones with amplitudes $\alpha_\ell = 1$ for all ℓ.



(b) – Three tones with the amplitudes $\alpha_\ell \in \{1,\, 10^{-1/4},\, 10^{-1/2}\}$.

Figure 8.2: MSE of the algorithm when estimating the frequencies. The number of data samples is $N = 128$. Random initialization is performed according to $\hat{\omega}_\ell = \omega_\ell + u$, where u is a random variable with uniform distribution in the interval $[-1/(2N), 1/(2N)]$. The individual empiric MSE_ℓ (solid line) as a function of the $\mathrm{SNR}_1 = \alpha_1/(2\sigma^2)$. The asymptotic CRB corresponding to SNR_1 (dashed line) is given as a reference.

Bibliography

[AH03a] Tomas Andersson and Peter Händel. IEEE standard 1057, Cramér-Rao bound and the parsimony principle. In *International Workshop on ADC Modelling and Testing*, pages 231–234, Perugia, Italy, September 2003.

[AH03b] Tomas Andersson and Peter Händel. Multiple-tone estimation by IEEE standard 1057 and the expectation-maximization algorithm. In *IEEE Conference on Instrumentation and Measurement*, Vail, CO, May 2003.

[AH04] Tomas Andersson and Peter Händel. Toward a standardized multi-sinewave fit algorithm. In *9th European Workshop on ADC Modelling and Testing*, volume 1, pages 337–342, Athens, Greece, September 2004.

[AH05a] Tomas Andersson and Peter Händel. IEEE standard 1057, Cramér-Rao bound and the parsimony principle. *IEEE Transactions on Instrumentation and Measurements*, January 2005. In press.

[AH05b] Tomas Andersson and Peter Händel. Multiple-tone estimation by IEEE standard 1057 and the expectation-maximization algorithm. *IEEE Transactions on Instrumentation and Measurements*, January 2005. In press.

[AH05c] Tomas Andersson and Peter Händel. Robustness of wave-fitting with respect to uncertain parameter values. In *Proceedings IEEE Instrumentation and Measurement Technology Conference*, Ottawa, Canada, May 2005.

[ASDM01] P. Arpaia, A. Cruz Serra, P. Daponte, and C. L Monteiro. A critical note to IEEE 1057-94 standard on hysteretic ADC dynamic testing. *IEEE Transactions on Instrumentation and Measurement*, 50(4):941–948, August 2001.

[ASH00] Tomas Andersson, Mikael Skoglund, and Peter Händel. Frequency estimation by 1-bit quantization and table look-up processing. In *Proceedings European Signal Processing Conference*, pages 1807–1810, Tampere, Finland, September 2000.

[ASH01] Tomas Andersson, Mikael Skoglund, and Peter Händel. Frequency estimation utilizing the hadamard transform. In *IEEE Workshop on Statistical Signal Processing*, pages 409–412, August 2001.

[ASH05] Tomas Andersson, Mikael Skoglund, and Peter Händel. Experiments on a hardware frequency estimator utilizing table look-up processing. Technical report, Royal Institute of Technology (KTH), May 2005. IR-S3-SB-0541.

[BAH05] Niclas Björsell, Olav Anderson, and Peter Händel. High dynamic range test-bed for characterization of analog-to-digital converters up to 500 msps. In *14th IMEKO Symposium on New Technologies in Measurement and Instrumentation*, Gdynia, Polen, September 12–15 2005.

[Bla99] J.J. Blair. Sine-fitting software for IEEE standards 1057 and 1241. In *the 16th IEEE Instr. and Meas. Technology Conference*, volume 3, pages 1504–06, Venice, Italy, May 1999. IMTC/99.

[BMS+02] TamásŽoltán Bilau, Tamás Megyeri, Attila Sárhegyi, János Márkus, and István Kollár. Four parameter fitting of sine wave testing results: iteration and convergence. In *4-th International Conference on Advanced A/D and D/A Conversion Techniques and their Applications & 7-th European Workshop on ADC Modelling and Testing*, pages 185–189, Prague, Czech Republic, June 26-28 2002.

[BMS+03] TamásŽoltán Bilau, Tamás Megyeri, Attila Sárhegyi, János Márkus, and István Kollár. Four parameter fitting of sine

wave testing results: iteration and convergence. *Computer Standards and Interfaces*, 26(1):51–56, 2003.

[CKQ94] V. Clarkson, P. J. Kootsookos, and B. G. Quinn. Analysis of the variance threshold of kay's weighted predictor frequency estimator. *IEEE Transactions on Signal Processing*, 42(9):2370–2379, 1994.

[DLR77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39(1):1–38, 1977.

[dP95] Baron Gaspard Riche de Prony. Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'École Polytechnique*, 1(22):24–76, 1795.

[dSS01] M. Fonseca da Silva and A. Cruz Serra. Improving convergence of sine fitting algorithms. In *6th European Workshop on ADC Modeling and Testing*, pages 121–124, Lisbon, Portugal, September 13-14 2001.

[FL96] P.D. Fiore and S.W. Lang. Efficient phase-only frequency estimation. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2809–2812, Atlanta, GA, May 7-10 1996.

[Fuc88] J.J Fuchs. Estimating the number of sinusoids in additive white noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(12):1846–1853, December 1988.

[Fuc94] Jean-Jacques Fuchs. Multiscale identification of real sinusoids in noise. *Automatica*, 30(1):147–155, 1994.

[FW88] M. Feder and E. Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(4):477–489, April 1988.

[GT97] N. Giaquinto and A. Trotta. Fast and accurate ADC testing via an enhanced sine wave fitting algorithm. *IEEE Trans. on Instrumentation and Measurement*, 46(4):1020–1024, August 1997.

[GvL96] Gene H. Golub and Charles F. van Loan. *Matrix Computations*. John Hopkins University Press, third edition, 1996.

[Hän00] Peter Händel. Properties of the IEEE-STD-1057 four parameter sine wave fit algorithm. *IEEE Transactions on Instrumentation and Measurement*, 49:1189–1193, December 2000.

[HDM99] Djamel Haddadi, Dominique Dallet, and Philippe Marchegay. A globally convergent Newton method for characterizing A/D converters. In *4th European Workshop on ADC Modeling and Testing*, pages 72–77, Bordeaux, 1999.

[HKS95] P. Hedelin, P. Knagenhjelm, and M. Skoglund. Theory for transmission of vector quantization dat a. In W. B. Kleijn and K. K. Paliwal, editors, *Speech coding and synthesis*, chapter 10, pages 347–396. Elsevier Science, 1995.

[HMH00] Anders Host-Madsen and Peter Händel. Effects of sampling and quantization on single-tone frequency estimation. *IEEE Transactions on Signal Processing*, 48:650–662, November 2000.

[HP03] Konrad Hejn and Andrej Pacut. Effective resolution of analog to digital converters. *IEEE Instrumentation and Measurement Magazine*, 6(3):48–55, September 2003.

[HSAHM03] Peter Händel, Mikael Skoglund, Tomas Andersson, and Anders Höst-Madsen. Method and apparatus for estimation physical parameters in a signal. Swedish Patent 520067, May 2003.

[HSP00] Peter Händel, Mikael Skoglund, and Mikael Pettersson. A calibration scheme for imperfect quantizers. *IEEE Transactions on Instrumentation and Measurement*, 49:1063–1068, October 2000.

[IHK91] F. H. Irons, D. M. Hummels, and S. P. Kennedy. Improved compensation for analog-to-digital converters. *IEEE Transactions on Circuits and Systems*, 38(8):958–961, August 1991.

[Kay88] Steven M. Kay. *Modern Spectral Estimation*. Prentice Hall, 1988.

[Kay89] S. Kay. A fast and accurate single frequency estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):1987–1990, 1989.

[Kay93] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*, volume 1. Prentice Hall, 1993.

[KB04] István Kollár and Jerome J. Blair. Improved determination of the best fitting sine wave in ADC testing. In *IEEE Instrumentation and Measurement Technology Conference*, volume 2, pages 829–834, May 2004.

[Kno97] Lue Knockaert. The Barankin bound and threshold behavior in frequency estimation. *IEEE Transactions on Signal Processing*, 45(9):2398–2401, september 1997.

[LASH02] Henrik Lundin, Tomas Andersson, Mikael Skoglund, and Peter Händel. Analog-to-digital converter error correction using frequency selective tables. In *RadioVetenskap och Kommunikation (RVK)*, pages 487–490, Stockholm, Sweden, June 2002.

[Llo82] Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, IT-28(2):129–137, March 1982.

[LSH01] Henrik Lundin, Mikael Skoglund, and Peter Händel. On external calibration of analog-to-digital converters. In *IEEE Workshop on Statistical Signal Processing*, pages 377–380, Singapore, August 2001.

[MK01] J. Márkus and I. Kollár. Standard framework for IEEE-STD-1241 in Matlab. In *IEEE Instrumentation and Measurement Technology Conference*, pages 1847–52, Budapest, Hungary, May 21-23 2001. IMTC/2001.

[Mou89] D. Moulin. Real-time equalization of A/D converter nonlinearities. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume 1, pages 262–267. IEEE, 1989.

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, Amsterdam, 1977.

[NP86] A. Nehorai and B. Porat. Adaptive comb filtering for harmonic signal enhancement. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(5):1124–1138, October 1986.

[PS96] Rik Pintelon and Johan Schoukens. An improved sine-wave fitting procedure for characterizing data acquisition channels. *IEEE Transactions on Instrumentation and Measurement*, 45(2):588 – 593, April 1996.

[RB74] D.C. Rife and R.R. Boorstyn. Single tone parameter estimation from discrete-time observations. *IEEE Transactions on Information Theory*, 20(5):591–598, 1974.

[RB76] D.C. Rife and R.R. Boorstyn. Multiple tone parameter estimation from discrete-time observations. *The Bell System Technical Journal*, 55(9):1389–1410, November 1976.

[S1057] IEEE Standard for digitizing waveform recorders. 1994.

[S1241] IEEE Standard for terminology and test methods for analog-to-digital converters. 2000.

[SB85] A.O Steinhardt and C. Bretherton. Threshold in frequency estimation. In *Proc. 1985 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1273–1276, Tampa, FL, March 26-29 1985.

[SHS94] Petre Stoica, Peter Händel, and Torsten Söderström. Approximate maximum likelihood frequency estimation. *Automatica*, 30(1):131–145, 1994.

[SJL97] Petre Stoica, Andeas Jakobsson, and Jian Li. Cisoid parameter estimation in the colored noise case: Asymtotic Cramér-Rao bound, maximum likelihood and nonlinear least-squares. *IEEE Transactions on Signal Processing*, 45(8):2048–2059, August 1997.

[SMFS89] P. Stoica, R.L. Moses, B. Friedlander, and T. Söderström. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Transactions*

on Acoustics, Speech and Signal Processing, ASSP-37:378–392, March 1989.

[SS89] Torsten Söderström and Petre Stoica. *System Identification*. Prentice Hall International, 1989.

[SS98] Mikael Skoglund and Jan Skoglund. On nonlinear utilization of intervector dependency in vector quantization. In *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 361–364, Seattle, USA, May 1998.

[Sto93] Petre Stoica. List of references on spectral line analysis. *Signal Processing*, pages 329–340, 1993.

[Swi95] D. N. Swingler. Further simple approximations to the Cramér–Rao lower bound on frequency estimates for closely–spaced sinusoids. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 43(1):367–369, January 1995.

[TL97] J. Tsimbinos and K. V. Lever. Improved error-table compensation of A/D converters. *IEE Proceedings - Circuits, Devices and Systems*, 144(6):343–349, December 1997.

[Tre85] S.A Tretter. Estimating the frequency of a noisy sinusoid by linear regression. *IEEE Transactions on Information Theory*, 31(6):832–835, 1985.