AUTO-REGRESSIVE SPECTRAL LINE ANALYSIS OF PIANO TONES

Thomas von Schroeter

Department of Computing
Imperial College of Science, Technology, and Medicine
180 Queen's Gate
London SW7 2BZ
ts9@ic.ac.uk

ABSTRACT

Three auto-regressive spectral estimation methods are experimentally tested with a view to musical applications: the Maximum Entropy method, Marple's MODCOVAR algorithm, and an efficient version of Prony spectral line estimation due to Cybenko. A performance analysis measuring the maximum relative error of their frequency estimates for a signal consisting of three sinusoids under variations of the model order (up to 20), signal length (60 to 200 samples) and noise level shows that unless the model order is close to 2/3 of the number of data points (i.e. when it is nearly ill-conditioned), Marple's algorithm gives by far the best results. In a separate experiment, Marple's algorithm was applied to recorded piano sounds; some preliminary results are shown which demonstrate its potential for fast multicomponent analysis.

1. INTRODUCTION

While auto-regressive spectral estimation methods have been popular in many areas of science for more than 20 years, it appears that investigations of the spectra of musical instruments have so far mainly relied on spectrogram methods. In a recent paper, the accuracy of frequency estimates obtained by the phase vocoder technique has been demonstrated for signals consisting of a small number of sinusoids whose frequencies are well separated from one another [1]. These conditions are certainly met for laboratory recordings when a single instrument is playing a single note [2, 3]. However, as soon as two notes are played together, a number of difficulties arise in practice: The *resolution* of spectrogram methods is limited to one sinusoid per filter band; spectral lines belonging to different notes which happen to lie in the same band cannot be resolved. *Spectral leakage* can compromise the accuracy of frequency estimates when components of different tones fall into neighbouring bands. Leakage can be reduced by increasing the frequency resolution, or reducing the transition bandwidth of the filters, but at the expense of time resolution. The *power distribution* among the fundamental and partials of a tone can be a source of further difficulties. In piano tones, the fundamental can contain more than 90% of the total power content; thus higher partials can easily be swamped by a noisy background. Detection of a sufficient number of partials would seem a necessary first step for polyphonic note identification.

In the light of these difficulties, auto-regressive (AR) methods hold some prospect for improvement for a number of reasons: First and foremost, the spectra of auto-regressive processes are better able to model sharp spectral features than the spectrogram, which gives a more or less uniformly sampled distribution of the average power across the whole frequency range. With AR methods, frequencies of isolated components can be determined much more accurately from short segments of the signal; the resolution of nearby components depends mainly on the model order and only to a lesser extent on the length of the signal segment. Besides, the order of an AR model, i.e. the number of poles of the underlying rational approximation to be fitted, can usually be chosen much smaller than the length of the signal segment to be analyzed; thus, in contrast to the spectrogram, considerable compression is achieved. Moreover, the trade-off between frequency and time resolution is far less dramatic with AR methods than with the spectrogram, and no explicit windowing is required. For a comprehensive survey of the most common fast spectrum analysis methods see [4, 5]; performance comparisons for some of the methods over short data segments were reported in [6, 7].

In an attempt to explore the potential of AR methods for music analysis, this paper is devoted to an experimental comparison of three of the auto-regressive methods reported in [4], namely the *Maximum Entropy method* (also known as all-pole or autocorrelation method), Marple's *MODCOVAR* algorithm [5], and *Prony spectral line estimation* in an efficient implementation due to Cybenko [8]. These methods were chosen to cover varying degrees of modelling freedom and computational complexity. Our performance criterion is the *relative* accuracy of frequency estimates under variations of the length of the signal segment, the noise level and also the order of the AR model, since as yet no reliable estimators of the correct model order seem to be known.

The paper is organized as follows: §2 reviews the algorithms, §3 contains the performance analysis, and §4 gives some preliminary results from an application of Marple's algorithm to recorded piano signals.

2. REVIEW OF THE ALGORITHMS

All three methods model the signal segment $s_n, n = 1..N$ as the sum of an *autoregressive process*,

$$s_{n+p} = -a_0 s_n - a_1 s_{n+1} - \ldots - a_{p-1} s_{n+p-1} \ , \ a_j \in \mathbb{C} \quad (1)$$

and noise. The number p of parameters, which is also the recursion depth, is called the *order* of the model. The spectral power density for a signal satisfying (1) in additive white noise of zero mean and variance ρ is

$$S(\omega) = \frac{\rho f_s^{-1}}{|a_0 + a_1 z + \ldots + a_{p-1} z^{p-1} + z^p|^2} \ , \quad z = e^{\mathrm{i}\omega} \quad (2)$$

where f_s denotes the sample rate, ρ the noise variance, and ω digital frequency (see e.g. [5], chapter 6). The ability of (2) to model sharp spectral features is due to its being a rational function in z rather than a polynomial, as is the discrete Fourier transform.

The *Maximum Entropy method* [4, 5] is based on a relation between the AR parameters and the expectation values $\phi_m = \langle s_n \overline{s}_{n-m} \rangle$ of the autocorrelation sequence which is known as the *Yule-Walker equation*:

$$\begin{bmatrix} \phi_0 & \phi_{-1} & \dots & \phi_{-p} \\ \phi_1 & \phi_0 & \dots & \phi_{1-p} \\ \vdots & \vdots & \dots & \vdots \\ \phi_p & \phi_{p-1} & \dots & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_{p-1} \\ \vdots \\ a_0 \end{bmatrix} = \begin{bmatrix} \rho \\ 0 \\ \vdots \\ 0 \end{bmatrix} . \quad (3)$$

In practice, estimates have to be substituted for the ϕ_m as the precise autocorrelation lags are not known. (3) has Toeplitz structure and can thus be solved with a fast algorithm using $O(p^2)$ multiplications. Given the parameters a_j, the dominant frequencies ω_k can be determined from the roots $z_k = |z_k| \exp(\mathrm{i}\omega_k)$ of the monic polynomial

$$P(z) = \sum_{j=0}^{p-1} a_j z^j + z^p = \prod_{k=1}^{p} (z - z_k) \quad (4)$$

which occurs in the numerator of 2. – (3) can be derived as the normal equation for the least squares problem of minimizing the squared 2-norm $\|f\|^2$ of the *forward prediction error* sequence

$$f_n = a_0 s_n + a_1 s_{n+1} + \ldots + a_{p-1} s_{n+p-1} + s_{n+p} \quad (5)$$

over all choices of the parameters $a_0, \ldots, a_{p-1}$. Alternatively, it can also be derived as the estimate which, given *exact knowledge* of the first p lags of the autocorrelation sequence, would maximize the process entropy over all possible choices of the unknown autocorrelation lags if the driving process is assumed to be Gaussian; hence the name of the method.

Marple's MODCOVAR algorithm [5] minimizes, instead of $\|f\|^2$, the sum $\|f\|^2 + \|b\|^2$ of the forward and backward prediction errors, where the *backward prediction error* sequence is

$$b_n = \overline{a}_0 s_n + \overline{a}_1 s_{n-1} + \ldots + \overline{a}_{p-1} s_{n-p+1} + s_{n-p} . \quad (6)$$

To avoid implicit windowing, the ranges of n in (5) and (6) are chosen such that only samples within the segment $1..N$ occur. The resulting normal equation can be written in the form

$$\begin{bmatrix} r_{00} & r_{01} \dots r_{0p} \\ r_{10} & r_{11} \dots r_{1p} \\ \vdots & \vdots \dots \vdots \\ r_{p0} & r_{p1} \dots r_{pp} \end{bmatrix} \begin{bmatrix} 1 \\ a_{p-1} \\ \vdots \\ a_0 \end{bmatrix} = \begin{bmatrix} \rho \\ 0 \\ \vdots \\ 0 \end{bmatrix} , \quad (7)$$

where

$$r_{ij} = \sum_{k=0}^{N-p-1} \left(s_{k+i} \overline{s}_{k+j} + s_{k+p-j} \overline{s}_{k+p-i} \right) .$$

(Bars denote complex conjugation.) Although (7) does not have as simple a structure as (3), it can be solved by a fast algorithm due to Marple [5] using $Np + 6p^2$ multiplications.

Prony spectrum estimation [4] is a method for fitting a signal with a complex linear combination of exponentials,

$$s(t) = \sum_k A_k z_k^t , \quad A_k,\ z_k \in \mathbb{C} , \quad (8)$$

where the A_k *and* the z_k are unknown. The model is flexible enough to apply to both damped and growing oscillations ($|z_k| \neq 1$). The special case of undamped real sinusoids ($|z_k| = 1$; z_k and A_k occurring in complex conjugate pairs) is referred to as *Prony spectral line estimation*. The connection with auto-regressive modelling is made by observing that any signal modelled by (8) satisfies the difference equation (1) with the parameters a_j given by (4). Conversely, given a noisy signal s, Prony spectral line estimation first fits the a_j by minimizing the prediction error

$$\|P(Z)s\|^2 = \sum_n |s_{n+p} + a_{p-1} s_{n+p-1} + \ldots + a_0 s_n|^2 \quad (9)$$

over all monic degree p polynomials P with roots on the unit circle, where $\|x\| = \sqrt{x^H x}$ denotes the 2-norm of a complex sequence x_n and Z the unit time shift operator defined by $(Zs)_j = s_{j+1}$. The roots z_k of the minimizing polynomial (4) then determine the digital frequencies $\omega_k = \arg(z_k)$, and the complex amplitudes A_k can be found by a least squares fit of (8) with the z_k obtained in the first step.

If the signal is real, the complex exponentials in the model, and hence the roots of P, occur in complex conjugate pairs. The redundancy can be eliminated by rewriting the model in real form,

$$s(t) = \sum_k A_k \cos(\omega_k t + \varphi_k) , \quad A_k,\ \omega_k,\ \varphi_k \in \mathbb{R} , \quad (10)$$

and observing that any sequence generated from it satisfies the difference equation $R(\Omega)s = 0$, where $\Omega = Z + Z^{-1}$ and

$$R(x) = \prod_k (x - 2\cos\omega_k) . \quad (11)$$

The polynomial R is real, monic, and has its zeros in the real interval $[-2, 2]$; compared to (8), the model order is halved. Conversely, for given signal s and model order p, Cybenko [8] shows 1) that the complex least squares problem (9) is equivalent to minimizing $\|R(\Omega)s\|$ over all real monic polynomials R of degree p *without* extra conditions about the location of its zeros (they are guaranteed to lie in $[-2, 2]$), and 2) that the minimizing polynomial R_p is the unique degree p monic orthogonal polynomial with respect to the inner product

$$\langle P, Q \rangle := (P(\Omega)s)^T Q(\Omega)s = s^T (P \cdot Q)(\Omega)s . \quad (12)$$

The task is therefore reduced to constructing this polynomial and finding its roots. This can be done efficiently by determining the coefficients in the double recursion

$$R_{j+1}(x) = (x - \alpha_{j+1}) R_j(x) - \beta_j^2 R_{j-1}(x) \quad (13)$$

starting with $R_{-1} \equiv 0$ and $R_0 \equiv 1$ up to order p, and finding the roots of R_p by computing the eigenvalues of the tridiagonal matrix

$$T_p = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \dots & 0 & 0 & 0 \\ \beta_1 & \alpha_2 & \beta_2 & \dots & 0 & 0 & 0 \\ 0 & \beta_2 & \alpha_3 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \beta_{p-2} & \alpha_{p-1} & \beta_{p-1} \\ 0 & 0 & 0 & \dots & 0 & \beta_{p-1} & \alpha_p \end{bmatrix} . \quad (14)$$

A more efficient version of the recursion is

$$\beta_{j+1} Q_{j+1}(x) = (x - \alpha_{j+1}) Q_j(x) - \beta_j Q_{j-1}(x) , \quad (15)$$

where the Q_j is the *orthonormal* polynomial of degree j with positive leading coefficient, starting with $Q_{-1} \equiv 0$ and $Q_0 \equiv 1/\|s\|$. Using the relation $\langle xP, Q\rangle = \langle P, xQ\rangle$, which is an immediate consequence of (12), the coefficients can be computed as

$$\alpha_{j+1} = \langle xQ_j, Q_j\rangle \ , \quad \beta_j = \langle xQ_j, Q_{j-1}\rangle \ , \quad \beta_0 = \|s\| \ . \ (16)$$

In terms of the vector quantities

$$q_j := Q_j(\Omega)s \ , \quad g_j := \Omega q_j - \beta_j q_{j-1} \ , \quad h_{j+1} := \beta_{j+1} q_{j+1} \ ,$$

the algorithm is as follows:

$$\begin{aligned} &h_0 \leftarrow s \ , \ \beta_0 \leftarrow \sqrt{s^T s} \ , \ q_0 \leftarrow \{0, \ldots, 0\} \text{ (length as } s) \ ; \\ &\text{for } j := 0 \text{ to } p-1 \text{ do } \{ \quad q_{j-1} \leftarrow q_j \ , \\ &\qquad\qquad\qquad\quad q_j \leftarrow h_j/\beta_j \ , \\ &\qquad\qquad\qquad\quad g_j \leftarrow \Omega q_j - \beta_j q_{j-1} \ , \\ &\qquad\qquad\qquad \alpha_{j+1} \leftarrow g_j^T q_j \ , \\ &\qquad\qquad\qquad h_{j+1} \leftarrow g_j - \alpha_{j+1} q_j \ , \\ &\qquad\qquad\qquad \beta_{j+1} \leftarrow \sqrt{h_{j+1}^T h_{j+1}} \ \} \end{aligned}$$

It requires about $4Np$ multiplications, plus p^3 for the eigenvalues of T_p.

3. ACCURACY OF FREQUENCY MEASUREMENTS

All numerical experiments were carried out in *Mathematica*. The methods were applied to the following synthesized test signal (a longer segment of which was used in [8]):

$$s_n = 5.5\cos(1.3n) + 5.5\cos(0.2n) + 1.7\cos(2.5n) \ , \quad (17)$$

where $n = 1..200$. For the measurements in noise, 100 different realizations of a Gaussian random sequence w_n with zero mean and variance 1 (average noise to signal ratio $\|w\|/\|s\| \approx 0.18$) were synthesized and separately added to the signal. Further parameters were the length N of the signal segment ($N = 50$, 100 and 200), and the number of real sinusoids modelled; for Prony line estimation, this number coincides with the model order p, while for the other two methods it is equal to $p/2$.

To avoid implicit windowing, and hence bias, in its operation on finite sequences, $\Omega = Z + Z^{-1}$ was realized by adding the two opposite time shifts of its argument excluding the extreme components which would fall on zeros if the sequences were zero-padded. To be able to carry out the ensuing vector operations, this has to be matched by a shortening of the other vector quantities. Thus their dimension drops by 2 in each iteration. Likewise, for the Maximum Entropy method, the *unbiased* autocorrelation estimate

$$\phi_m = \phi_{-m} = \frac{1}{N-m} \sum_{n=0}^{N-m-1} s_n s_{n+m} \ , \quad 0 \le m \le p$$

was used; the Toeplitz solver was adapted from [9], section 2.8.

From the parameters obtained with each of the methods, the zeros z_k (resp. x_k) of the resulting polynomials were computed using *Mathematica*'s standard routines `NSolve` (for the Maximum Entropy method and Marple's algorithm) and `Eigenvalues` (for Cybenko's method) and converted to digital frequencies according to $\omega_k = \arg(z_k)$ resp. $\omega_k = \arccos(x_k/2)$. Among the resulting set of positive frequencies, the ones closest to the three true frequencies ω_k^* were selected, and the maximal relative error

$$\max\{|\omega_k - \omega_k^*|/\omega_k^* \ , \ k = 1, 2, 3\} \quad (18)$$

was computed. (For $p \le 3$ all measured frequencies were retained, and the relative error computed with the closest true frequency.)

The results of our simulations are shown in Fig. 1 and suggest the following *conclusions*: 1) Marple's algorithm is consistently superior, typically by a factor of 2 to 4 in the relative accuracy, except for model orders close to $2N/3$. (For $p > 2N/3$ the matrix in (7) is singular.) The next best method is the Maximum Entropy method, with Prony line estimation finishing a distant third. This is somewhat surprising given that the Prony line model matches the test data more closely than the other two methods. It has previously been suggested that constraining the poles to lie on the unit circle leads to a statistical bias in the frequency estimates proportional to the noise variance [10]; however this would not explain the comparatively poor performance in the absence of noise. 2) Not suprisingly, for clean data the relative accuracy of all estimators improves fairly monotonically with the model order (in fact, Marple's algorithm stops at the correct order $p = 6$); for noisy data the accuracy of the Prony and Maximum Entropy methods show quite drastic fluctuations. 3) Specifically for musical applications, an upper bound for the relative frequency error (18) of a quarter tone, or $2^{1/24} - 1 \approx 2.9\%$, would certainly be a minimum requirement. For segment lengths of at least 100 samples and clean data, this seems to be clearly satisfied by the Maximum Entropy method and Marple's algorithm, while Prony spectral line estimation is ruled out.

4. AUTO-REGRESSIVE SPECTRA OF PIANO SOUNDS

The remaining figures show some preliminary results of an application of Marple's MODCOVAR algorithm to piano sounds. These were taken from a recording of Bach's Fugue in C major from part I of "The Well-Tempered Clavier", sampled at 5 kHz using *SoundEdit* on a Macintosh, and then processed as above using 20 poles. Due to a fault in the tape recorder used as input device, the entire spectrum is shifted to the lower end by about a semitone.

Fig. 2 shows the spectrum of a time slice of 1000 samples (0.2s) from the first tone of the piece. The note being played is C_4; however, comparison with the fundamentals and partials of B_3 (shown as dots) reveals the frequency shift just mentioned. Nevertheless, the fundamental and 5 partials are clearly visible. Fig. 3 shows a frequency peak spectrum of the first 10 seconds of the piece with frequencies given in units of semitones above $A_3 = 220$ Hz. Fundamentals and up to 5 partials of the longer notes are visible as stripes; to resolve shorter ones (like the demisemiquavers G_4 and F_4 in the first bar at about frame 100) would probably require higher sample rates. Only poles lying within a distance of 0.01 from the unit circle in the complex plane are retained in order to filter out spurious poles. The fact that information about the relative size of spectral peaks is available at no additional computational cost represents a practical advantage of unconstrained pole search over the Prony line method.

CPU timings in *Mathematica* (on a PC with 400 MHz Pentium II processor) for the pole search over the entire signal segment shown in the figure were about 22 seconds for an uncompiled version of Marple's algorithm, and 7 seconds for a partly compiled implementation of the Maximum Entropy method.

5. ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council of Great Britain.

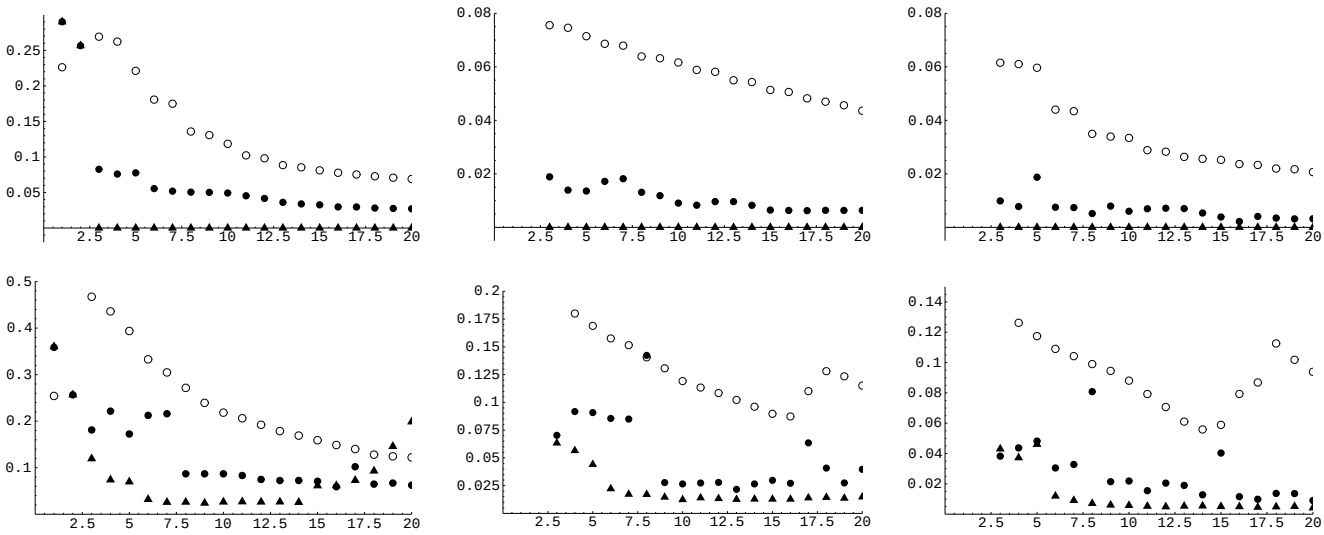

Figure 1: *Relative error* (18) *of the frequency measurement (vertical axis) as a function of the number of sinusoids modelled. Dots: maximum entropy method; circles: Prony method; triangles: Marple's MODCOVAR algorithm. Top row: clean data s_n, bottom row: noisy data $s_n + w_n$ with $\|w\|/\|s\| \approx 0.18$. From left to right: segment length $N = 60$, 100, and 200 samples.*

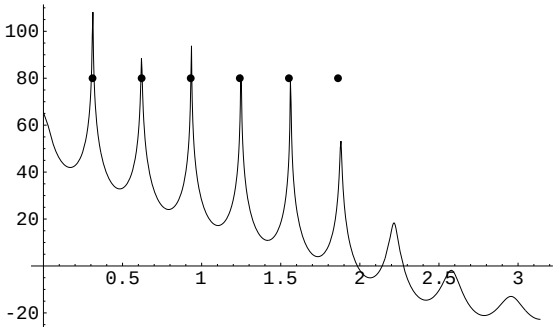

Figure 2: *Auto-regressive power spectrum $S(\omega)$ in dB of a single piano tone, obtained from 1000 samples of a recorded signal sampled at 5 kHz using Marple's algorithm with 20 poles. The dots indicate the location of the exact multiples of B_3 (246.94 Hz).*

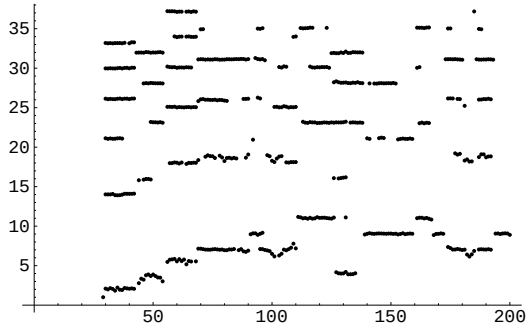

Figure 3: *Auto-regressive peak spectrum of the first 10 seconds of a recording of Bach's Fugue in C Major BWV 846, sampled at 5 kHz and processed using Marple's algorithm with 20 poles. Horizontal axis: time in frames of 250 samples (50 ms); vertical axis: frequency in semitones above A_3 (220 Hz).*

6. REFERENCES

[1] M. S. Puckette and J. C. Brown, "Accuracy of Frequency Estimates Using the Phase Vocoder," *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 166–176, 1998.

[2] J. C. Brown and M. S. Puckette, "A high resolution fundamental frequency determination based on phase changes of the Fourier transform," *J. Acoust. Soc. Am.*, vol. 94, pp. 662–667, 1993.

[3] J. C. Brown, "Frequency ratios of spectral components of musical sounds," *J. Acoust. Soc. Am.*, vol. 99, 1996.

[4] S. M. Kay and S. L. Marple, "Spectrum Analysis – A Modern Perspective," *Proc. IEEE*, vol. 69, pp. 1380–1419, 1981.

[5] S. L. Marple, *Digital spectral analysis with applications*, Prentice-Hall (Englewood Cliffs, New Jersey), 1987.

[6] P. Barone, "Some Practical Remarks on the Extended Prony Method of Spectrum Analysis," *Proc. IEEE*, vol. 76, pp. 284–285, 1988.

[7] K. M. M. Prabhu and K. Bhoopathy Bagan, "Resolution capability of nonlinear spectral estimation methods for short data lengths," *IEE Proceedings, Part F*, vol. 136, pp. 135–142, 1989.

[8] G. Cybenko, "Fast approximation of dominant harmonics," *SIAM J. Sci. Stat. Comput.*, vol. 5, pp. 317–331, 1984.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in PASCAL. The Art of Scientific Computing*, Cambridge University Press, 1989.

[10] P. Stoica and A. Nehorai, "On Linear Prediction Models Constrained to Have Unit-Modulus Poles and Their Use for Sinusoidal Frequency Estimation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, pp. 940–942, 1988.