THE MVDR SPECTRUM AND SPEECH MODELING: A TUTORIAL

Manohar N. Murthi and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92039-0407 USA
manoharn@ece.ucsd.edu brao@ece.ucsd.edu

ABSTRACT

We present a brief tutorial on the Minimum Variance Distortionless Response Spectrum (MVDR). In particular, we highlight the filter-bank and all-pole model properties of the MVDR spectrum. The characterization of the MVDR spectrum as the output of a filter-bank leads to the observation that the MVDR approach results in a data dependent and frequency dependent filter-bank providing an interesting contrast to the uniform filter-bank structure implied by the FFT based methods. The all-pole model characterization provides an interesting contact with parametric methods and is contrasted to Linear Prediction based all-pole modeling. The efficacy of the MVDR spectrum is demonstrated with its application to speech modeling.

1. INTRODUCTION

Spectral analysis is a fundamental area of signal processing with a rich history and toolbox for the engineer [1, 2, 3, 4]. Depending upon the application, the engineer can choose among several different types of methods to analyze a signal's spectrum. For example, non-parametric spectral analysis methods such as periodogram based methods which rely upon the computational elegance of the Fast Fourier Transform are quite popular in obtaining estimates of spectral power at a set of frequencies. Parametric techniques such as Linear Prediction [5, 6], which is a well-known method among the general class of Auto-Regressive (AR) spectral estimates, are also useful for estimating a signal's spectra. Parametric techniques describe a signal's spectrum by using a small number of parameters that are easily estimated from data.

The choice of parametric or non-parametric spectral analysis methods often depends upon the type of signal being analyzed, and the context of the analysis. For example, the demands of a real-time application using low-power DSP chips in a data compression context often preclude the use of spectral analysis techniques that involve many iterative computations. Furthermore, the application often dictates the performance metrics used to evaluate competing analysis methods. In most spectral analysis textbooks, the fundamental issues of spectral resolution, variance of the estimate, and computational complexity are often used to compare and contrast competing spectral analysis methods. However, in certain applications, such as speech signal processing, other concerns often affect the type of analysis method used. For example, for the problem of estimating an envelope whose contour passes through the peaks of a voiced speech line spectrum, spectral resolution is not an important metric, and is often supplanted by measures of peak spectral distortion.

In this paper, we present a tutorial on the MVDR spectrum and discuss its application to speech. Our emphasis is upon providing insight into the MVDR approach. Consequently, we avoid mathematical derivations by suggesting appropriate references and concentrating on the interpretation of the results.

This work was supported by UC MICRO Grants 97-146, 98-125, Hughes Electronics, Nokia, and Qualcomm

2. THE MVDR SPECTRUM

The Minimum Variance Distortionless Response (MVDR) spectrum is a flexible spectral analysis method which is applicable in many signal processing contexts and possesses features of both parametric and non-parametric techniques. The MVDR spectrum was introduced by Capon [7], and is also known as the Capon spectrum, or the Maximum Likelihood Method (MLM) spectrum. The MVDR spectrum is a well-known method in array processing applications, and appears to be promising in other applications such as speech. The MVDR spectrum is given by [8, 2, 3, 4]

$$P_{MV}^{(M)}(\omega) = \frac{1}{\mathbf{v}^H(\omega)\mathbf{R}_{M+1}^{-1}\mathbf{v}(\omega)} \tag{1}$$

where $P_{MV}^{(M)}(\omega)$ is the Mth order MVDR spectrum for all frequencies ω, $\mathbf{v}(\omega) = [1, e^{j\omega}, e^{j2\omega}, \cdots, e^{jM\omega}]^T$ is a frequency tuning vector with $\mathbf{v}^H(\omega)$ its conjugate transpose, and $\mathbf{R}_{M+1}$ is the $(M+1)\times(M+1)$ Toeplitz autocorrelation matrix whose entries are obtained from the autocorrelation estimates of the signal. The spectrum's order M corresponds to the largest correlation lag in the Toeplitz matrix.

2.1. Computation of the MVDR spectrum

For parameterization purposes, the Mth order MVDR spectrum can be parametrically written as [9]

$$P_{MV}^{(M)}(\omega) = \frac{1}{\sum_{k=-M}^{M}\mu(k)e^{-j\omega k}}, \tag{2}$$

in which the $(M+1)$ coefficients $\mu(0), \cdots, \mu(M)$ completely specify the MVDR spectrum $P_{MV}^{(M)}(\omega)$.

By exploiting the the Toeplitz structure of $\mathbf{R}_{M+1}$, an efficient computation of the MVDR spectrum is also possible. First, an expression for the inverse of $\mathbf{R}_{M+1}$ (the Gohberg-Semencul formula [4]) can be obtained. Second, the MVDR parameters $\mu(k)$

can be related to the Linear Prediction spectrum. The Linear Prediction spectrum is parameterized by a set of Linear Prediction Coefficients $a_k, k = 1, \cdots, M$ with $a_0 = 1$, and a prediction error variance (gain) P_e. Consequently, the parameters $\mu(k)$, and the MVDR spectrum can be easily obtained using a method that involves a modest non-iterative computation based upon the parameters of a Linear Prediction spectrum [9, 8, 2]

$$\mu(k) = \begin{cases} \frac{1}{P_e}\sum_{i=0}^{M-k}(M+1-k-2i)a_i a_{i+k}^*, & k = 0,..,M \\ \mu^*(-k), & k = -M..-1. \end{cases} \tag{3}$$

Combined with the Levinson-Durbin algorithm [8, 2] for computing a_k and P_e, Equations 2 and 3 provide an efficient method for computing the MVDR spectrum.

2.2. Filter-Bank Interpretation

Although the MVDR spectrum is parametrically characterized by the $\mu(k)$ coefficients, it has an affinity with non-parametric filter-bank spectral analysis methods which we discuss now in order to provide insight into the methodology. For this purpose it is useful to recall the filter-bank interpretation of periodogram-based spectral analysis methods which rely upon the FFT [4]. In particular, the periodogram based estimate of the spectrum at any given frequency ω_l can be viewed as the power at the output of a bandpass filter with center frequency ω_l and filter characteristics defined by the length and nature of the analysis window. The periodogram computed via the FFT can be viewed as a bank of bandpass filters where the bandpass filters are both data and frequency *independent*.

Similar to periodogram-based methods, the MVDR spectral estimate for a set of analysis frequencies can also be *conceptually* viewed as the output of a bank of filters, with each filter centered at one of the analysis frequencies. In contrast to periodogram-based methods, the bandpass filter is both data and frequency *dependent*. The nature of the data and frequency dependent filter is what distinguishes MVDR from periodogram-based methods. This viewpoint can be readily understood by examining the formulation used in deriving the MVDR spectrum.

In particular, the MVDR spectral estimate of order M at frequency ω_l utilizes a specially designed FIR bandpass filter with impulse response $h_l[n]$ whose passband characteristics are simply defined as requiring the filter's response at the frequency of interest, ω_l, to be unity, i.e.

$$H_l(e^{j\omega_l}) = \sum_{k=0}^{M} h_l(k)e^{-j\omega_l k} = 1. \tag{4}$$

This constraint is known as the *distortionless constraint* and is central to the understanding of MVDR methods. In addition to satisfying this constraint, the FIR bandpass distortionless filter $h_l(n)$ minimizes the filter output power for a given input signal. This minimization of the output power defines the attenuation characteristics of the bandpass filter. In other words, the distortionless constraint ensures that the MVDR distortionless filter $h_l(n)$ will let the input signal components with frequency ω_l pass through undistorted (passband), and the minimization of the output power ensures that the remaining frequency components in the input signal are suppressed in an optimal manner (stopband).

Determination of the bandpass distortionless filter coefficients $h_l(n)$ results in a simple constrained optimization problem whose solution is given by [8, 2, 3, 4]

$$\mathbf{h_l} = \frac{\mathbf{R}_{M+1}^{-1}\mathbf{v}(\omega_l)}{\mathbf{v}^H(\omega_l)\mathbf{R}_{M+1}^{-1}\mathbf{v}(\omega_l)}. \tag{5}$$

Note that the filter coefficients are both dependent upon the center frequency ω_l and upon the data through the autocorrelation matrix $\mathbf{R}_{M+1}$. The power at the output of this filter is the MVDR spectral power estimate at frequency ω_l and is given by (1). Note that there is no need for explicitly computing the bandpass filter coefficients which are only needed for conceptual purposes.

A closer examination of the Mth order ($(M+1)$ taps) distortionless filter $h_l(n)$ indicates that it uses one degree of freedom to maintain the distortionless constraint which allows the input signal power at ω_l to pass through without distortion. The remaining M degrees of freedom are used to cancel the rest of the input signal at the other frequencies in an optimal (minimum output power) manner. In particular, the distortionless filter $h_l(n)$ has a passband at ω_l only, and the stopband ranges over the rest of the frequencies. The data and frequency dependent nature of the distortionless filter allows it to vary the nulls of its stopband depending upon the input signal while preserving a unity gain passband at its analysis frequency ω_l.

Example 1:*Comparison of spectral estimates for two sinusoids.* In this example, a signal consisting of the sum of two closely spaced sinusoids, $s(n) = 10\cos(\omega_0 n) + 8\cos(\omega_1 n)$, is constructed in which $\omega_0 = 0.4\pi$ and $\omega_1 = 0.4156\pi$. The true autocorrelations $r(m)$ of the signal $s(n)$ are calculated and used to determine Blackman-Tukey and MVDR spectral estimates. The Blackman-Tukey estimate uses $r(0), \cdots, r(7)$, and consequently a two-sided window of length 15. The MVDR estimate uses $r(0), \cdots, r(3)$, corresponding to an order of $M = 3$. The results are shown in Figure 1. In Figure 1(a), the Blackman-Tukey estimate is displayed for the powers of the 2 sinusoids, which are denoted by the two circles. The Blackman-Tukey Estimate is unable to determine the powers of both sinusoids exactly because of its fixed bandpass filter structure. In Figure 1(b), the MVDR spectral estimate for the two powers is denoted by the solid line. Note that the MVDR spectral estimates of the 2 sinusoidal powers are exact. The frequency response of the distortionless filter $h_{\omega_0}(n)$ used to measure the power at ω_0 is denoted by the dotted line. $|H_{\omega_0}(\omega)|$ features a distortionless response (unity gain) at the frequency of interest (passband) ω_0, and adjusts its stopband to place a null at ω_1, thereby minimizing its output power. Therefore the data and frequency dependent filters used to obtain the spectral power estimates in the MVDR method utilize their degrees of freedom in a more judicious manner than the periodogram estimate's bandpass filters.

2.3. All-Pole Model Interpretation

In addition to its similarities with non-parametric spectral analysis techniques, the MVDR spectrum also has a connection with parametric models such as all-pole filters, and has a deeper connection with Linear Prediction which we now examine.

All-pole models, $1/B(z)$, which are stable and causal IIR filters with roots inside the unit circle, are popular in spectral modeling. Linear Prediction is the most popular method for determining all-pole models. The Linear Prediction spectrum parameters P_e and a_k define a spectrum

$$P_{LP}^{(M)}(\omega) = \frac{P_e}{|\mathbf{v}^H(\omega)\mathbf{a}|^2} = \frac{P_e}{|A(e^{j\omega})|^2} \tag{6}$$

where $\mathbf{a} = [1, a_1, \cdots, a_M]^T$, and $A(z) = 1 + a_1 z^{-1} + \cdots + a_M z^{-M}$. The associated all-pole filter $1/A(z)$ is useful for different signal analysis tasks.

The Mth order MVDR spectrum $P_{MV}^{(M)}(\omega)$ also defines an Mth order all-pole filter. In particular Equation 1 can be rewritten as

$$P_{MV}^{(M)}(\omega) = \frac{1}{|B(e^{j\omega})|^2}, \tag{7}$$

where $1/|B(e^{j\omega})|^2$ is the magnitude response of the Mth order MVDR all-pole filter $1/B(z)$. The MVDR filter $B(z)$ can be found by a spectral factorization of $B(z)B(z^{-1}) = \sum_{k=-M}^{M} \mu(k)z^{-k}$ used in the MVDR parametric spectrum of Eq. 2, although an approximation of $B(z)$ can be found by simpler methods.

In terms of spectral modeling, the MVDR spectrum is smoother in appearance than the Linear Prediction spectrum. In fact, the smoothness of the MVDR spectrum can be explained by the fact that the MVDR spectrum of order M is the harmonic mean of the order 0 to M Linear Prediction spectra. This property, discovered by Burg [10], is given by

$$\frac{1}{P_{MV}^{(M)}(\omega)} = \sum_{k=0}^{M} \frac{1}{P_{LP}^{(k)}(\omega)}, \tag{8}$$

where $P_{MV}^{(M)}(\omega)$ is the Mth order MVDR spectrum and $P_{LP}^{(k)}(\omega)$ is the kth order Linear Prediction spectrum. As a result of this averaging of the low order LP spectra, the Mth order MVDR spectrum has a smoother appearance than the Mth order Linear Prediction spectrum. This averaging also results in the Mth order MVDR spectrum having lower resolution than the corresponding Mth order Linear Prediction spectrum. On the other hand, the harmonic averaging results in the MVDR spectrum having a lower variance (more statistical reliability) than the Linear Prediction spectrum.

3. APPLICATION TO SPEECH MODELING

Speech is one of the most important signals considered in signal processing. We now present some background on speech signals in order to facilitate the understanding of MVDR all-pole modeling of speech spectra. Speech signals can be broadly classified into two classes: voiced speech, and unvoiced speech. Voiced speech is roughly periodic, and can be analytically approximated in the frequency domain as a discrete line spectrum with the spectral lines occurring at harmonics of a fundamental frequency which corresponds to the pitch of the speaker. Unvoiced speech tends to look more like noise, and is often approximated as the output of an IIR filter excited by white noise.

All-pole models are a fundamental part of many speech processing systems, most particularly speech coders, and are used to model the short-term spectrum of speech for 20-30 millisecond segments [11, 12, 13, 14, 15]. The all-pole filter $1/A(z)$ attempts to provide a spectral envelope that provides a gross spectral shape. For the case of voiced speech, which is the most prevalent and perceptually important type of speech, the all-pole spectral envelope's contour should smoothly pass through the peaks of the voiced speech harmonic spectral powers. It is not necessary for the all-pole spectrum to resolve the different harmonic peaks. Therefore, spectral resolution, which is often a very important metric for comparing differing spectral analysis methods, is not an important metric in all-pole modeling of speech. Rather, in determining the all-pole model parameters, envelope estimation concerns such as fitting a set of spectral peaks are more important. In fact, logarithmic measures of distortion between the voiced speech peak powers and the all-pole envelope contour are commonly used.

Linear Prediction based all-pole models tend to overestimate and overemphasize the harmonic powers of voiced speech spectra, particularly for high pitch speakers. The Linear Prediction spectrum's Mean Square Error criterion forces the LP filter to attempt to cancel the harmonics of a voiced speech signal, causing the all-pole spectrum's poles to move closer to the unit circle as the filter order is raised. As a result, the LP all-pole envelope's contour for high pitch and some medium pitch voiced speech tends to be unnaturally sharp, leading to metallic sounding speech, and several ad-hoc schemes are utilized to reduce these effects. Consequently, all-pole models which produce more accurate modeling of voiced speech harmonic powers are of interest [16, 17]. The MVDR all-pole spectrum of speech, first suggested in [18], and more thoroughly described in [19], appears to offer a reasonable alternative as we show next with the help of an example.

Example 2: *MVDR Variable Order Modeling of Voiced Speech.* In this experiment, a voiced speech spectrum of pitch 160 Hz is modeled by MVDR all-pole spectra of increasing filter orders. The results are shown in Figure 2. In the figure, the voiced speech spectrum consists of a set of roughly $L = 25$ spectral peaks that are at multiples of the pitch frequency. The harmonic nature of the voiced speech signal suggests that voiced speech can be analytically approximated as a discrete line spectrum. The valleys of the voiced speech spectrum are not as important from a perceptual modeling point of view. The MVDR all-pole spectra are plotted for filter orders of $M = 20, 30, 40$, and 50. As the filter order is increased, the MVDR all-pole envelope's fit of the voiced speech harmonic powers improves, a feature that the Linear Prediction spectrum does not possess.

The ability of the MVDR model to provide a better envelope with increasing model order can be understood by using the filter bank interpretation of the MVDR spectrum (sec. 2.2). Consider the MVDR spectrum at the lth harmonic of a voiced speech segment with pitch frequency ω_0. The MVDR spectrum is the power at the output of a data and frequency dependent bandpass filter. The distortionless response property of this filter ensures that the lth harmonic's power passes through undistorted. The minimization of the output power ensures that the contributions from the other interfering harmonics are minimized. However, if the model order is smaller than the number of interfering harmonic components, their contributions cannot be completely nullified, leading to a positive bias. As the model order is increased, the MVDR spectrum's ability to suppress the interfering harmonics is enhanced, thereby reducing the bias, and increasing the accuracy of the envelope.

As the previous example suggested, the high order MVDR spectrum can serve as a reference spectrum for speech coding applications. In particular, reduced order ($M = 10$ to 14) filters suitable for coding applications can be determined using techniques inspired by the MVDR approach [18, 20].

For modeling unvoiced speech, the linear prediction approach leads to reasonable all-pole models. The MVDR spectrum provides a smoother envelope than Linear Prediction all-pole models and also offers adequate modeling of unvoiced spectra. More detailed simulation results can be found in [19].

4. CONCLUSION

The MVDR approach has a unique perch between parametric and non-parametric spectral analysis techniques. With its data-adaptive filter-bank interpretation, the MVDR spectrum offers attractive properties for spectral analysis. Furthermore, the MVDR spectrum has a simple parameterization and computation that makes it attractive for real-time signal processing.

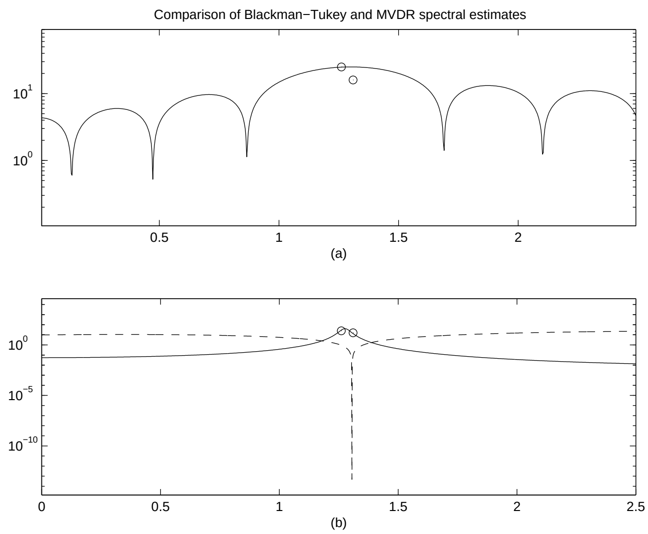

Figure 1: Comparison of Blackman-Tukey and MVDR spectral estimates of two closely spaced sinusoids. (a) Blackman-Tukey Estimate $M = 7$ (b) MVDR spectral estimate (solid line) of $M = 3$. The frequency response of the distortionless filter for measuring ω_0 is denoted by the dotted line. Note that this distortionless filter places a deep null at the competing sinusoid at ω_1. From *Example 1*.

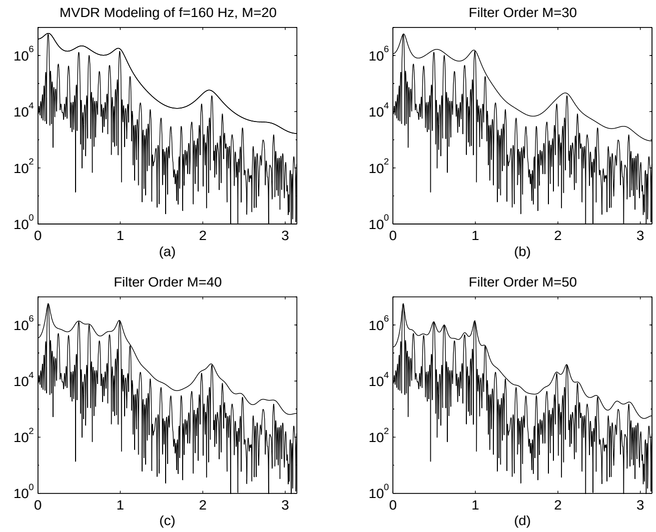

Figure 2: MVDR spectral envelope modeling of a f=160 Hz voiced speech spectrum. (a) filter order M=20. (b) M=30. (c) M=40. (d) M=50. From *Example 2.*

5. REFERENCES

[1] S.M. Kay and S.L. Marple Jr. “Spectrum analysis – A modern perspective”. *Proc. IEEE*, 69:1380–1419, Nov. 1981.

[2] S.L. Marple Jr. *Digital Spectral Analysis with Applications.* Prentice-Hall, Englewood Cliffs, NJ, 1987.

[3] S.M. Kay *Modern Spectral Estimation: Theory and Application.* Prentice-Hall, Englewood Cliffs, NJ, 1988.

[4] P. Stoica and R. Moses *Spectral Analysis* Prentice-Hall, Upper Saddle River, New Jersey, 1997.

[5] J. Makhoul. “Linear prediction - A tutorial review”. *Proceedings IEEE*, 63(4):501–580, April 1975.

[6] J.D. Markel and A.H. Gray Jr. *Linear Prediction of Speech*. Springer-Verlag, New York, 1976.

[7] J. Capon “High-resolution frequency-wavenumber spectrum analysis” *Proceedings of the IEEE*, vol. 57, (no. 8):1408-18, August 1969.

[8] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Englewood Cliffs,NJ, 1991.

[9] B.R. Musicus “Fast MLM power spectrum estimation from uniformly spaced correlations”. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, (no. 5):133-5, Oct. 1985.

[10] J.P. Burg “The relationship between maximum entropy spectra and maximum likelihood spectra”. *Geophysica*, vol. 37, pp.375-376.

[11] L.R. Rabiner and R.W. Schafer. *Digital Processing of Speech Signals.* Prentice-Hall, Englewood Cliffs, NJ, 1978.

[12] D. O’Shaughnessy. *Speech Communication: Human and Machine.* Addison-Wesley, 1987.

[13] W. B. Kleijn and K. K. Paliwal (editors). *Speech Coding and Synthesis.* Elsevier, 1995.

[14] L.R. Rabiner and B.H. Juang *Fundamentals of Speech Recognition.* Prentice-Hall, 1993.

[15] J .R. Deller, J. G. Proakis, and J. H. L. Hansen *Discrete-Time Processing of Speech Signals*, Prentice-Hall, Upper Saddle River, NJ, 1987.

[16] A. El-Jaroudi and J. Makhoul. “Discrete All-Pole Modeling”. *IEEE Trans. Signal Processing*, Feb. 1991.

[17] R. McAulay and T. Quatieri “Sinusoidal Coding ”. In *Speech Coding and Synthesis,* W.B. Kleijn and K.K. Paliwal, Eds., Elsevier 1995.

[18] M. N. Murthi and B. D. Rao. “Minimum Variance Distortionless Response (MVDR) Modeling of Voiced Speech”. In *Proc. ICASSP 97,* Munich, Germany, April 1997.

[19] M. N. Murthi and B. D. Rao. “All-Pole Modeling of Speech based on the Minimum Variance Distortionless Response Spectrum”. *Report No. 9804, Center For Wireless Communications,* University of California, San Diego.

[20] M. N. Murthi and B. D. Rao. “All-Pole Model Parameter Estimation for Voiced Speech”. In *1997 IEEE Workshop on Speech Coding For Telecommunications Proceedings*, Pocono Manor, PA, 1997.