ECE 6123
Advanced Signal Processing
Spectral Estimation

Peter Willett

Fall 2017

1 Basics of Spectral Estimation

1.1 Introduction

We are all familiar with the discrete-time Fourier transform (DTFT) and the discrete Fourier transform (DFT) – the latter being implemented efficiently via the fast Fourier transform (FFT). The former useful for analyzing deterministic signals; the latter is more practical, and gives a way to understand the frequency behavior of a signal that is given to you as a time series, one that may not have an explicit expression that nicely sums to something compact or conversely whose DTF is amenable to integration. But what does it *mean* when we take the FFT of a random signal? Here we will explore this; we will when necessary assume the signal $\{x[n]\}_{n=0}^{N-1}$ is wss, zero mean and Guassian[1].

We begin this section by discussing the periodogram, which is the most obvious approach to spectral estimation: it has a big problem, which we will solve later. We continue with a discussion of the meaning of resolution. We then establish the relationship between spectral estimation and beamforming – it turns out that much of what we do can be used for array signal processing provided the source is monochromatic (or can be made to be so by filtering) and the array is a uniformly-spaced linear array (ULA).

The following sections deal with nonparametric and parametric spectral estimation. As the name implies, non-parametric spectral estimation makes no assumptions about the nature of the spectrum, and we look at the Bartlett, Welch and Capon approaches. Parametric methods do make such an assumption, and the ones we explore here are based on AR models and on modeling as sinusoids-plus-noise.

[1] This is only important when we are discussing the periodogram, so explore its consistency. We will assume in that section that $x[n] \in \Re$ for ease of explanation; the complex case is the same but notationally more difficult

1.2 The Periodogram

Recall that the power spectrum of a random process $\{x[n]\}$ is defined as

$$S(\omega) \equiv \sum_{k=-\infty}^{\infty} r[k] e^{-j\omega k} \tag{1}$$

where $\{r[k]\}$ is the (usual) autocorrelation $r[k] = \mathcal{E}\{x[n]x[n-k]^*\}$. How about we estimate it from our data $\{x[n]\}_{n=0}^{N-1}$ as

$$\hat{S}(\omega) \equiv \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j\omega n} \right|^2 \tag{2}$$

that is, as the DTFT magnitude square and suitably[2] scaled? Note that the periodogram is efficiently computed as

$$\hat{S}(\omega)\Big|_{\omega=\frac{2\pi k}{N}} = \frac{1}{N}|X(k)|^2 \tag{3}$$

where $X(k)$ is the k^{th} DFT (or FFT) output.

We need some statistical analysis of the periodogram. We begin with the mean:

$$\begin{aligned}
\mathcal{E}\{\hat{S}(\omega)\} &= \frac{1}{N}\mathcal{E}\left\{\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} x[n]x[m]^* e^{-j\omega(n-m)}\right\} && (4)\\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} r[n-m] e^{-j\omega(n-m)} && (5)\\
&= \frac{1}{N}\sum_{k=-(N-1)}^{N-1} (N-|k|) r[k] e^{-j\omega k} && (6)\\
&= S(\omega) \star \mathcal{F}\left[1 - \frac{|k|}{N}\right] && (7)\\
&= S(\omega) \star W_B(\omega) && (8)
\end{aligned}$$

where $W_B(\omega)$ is the DTFT of the (triangular) Bartlett window $w_B[k]$:

$$\begin{aligned}
W_B(\omega) &= \mathcal{F}\left[1 - \frac{|k|}{N}\right] && (9)\\
&= \mathcal{F}\left[w_B[k]\right] && (10)\\
&= \left(\frac{1}{N}\frac{\sin(\omega N/2)}{\sin(\omega/2)}\right)^2 && (11)
\end{aligned}$$

[2] We will soon see why the scaling.

That is, the expected value of the periodogram is a smoothed version of the true power spectrum: it gets convolved with the sinc-squared.

Turning now to the variance, we compute the second moment. We need here to – briefly – assume[3] that $\{x[n]\}$ is real and Gaussian. We use the fact that for jointly-Gaussian zero-mean random variables we have

$$\mathcal{E}\{ABCD\} \;=\; \mathcal{E}\{AB\}\mathcal{E}\{CD\} \;+\; \mathcal{E}\{AC\}\mathcal{E}\{BD\} \;+\; \mathcal{E}\{AD\}\mathcal{E}\{BC\} \tag{12}$$

We get

$$\begin{aligned}
&\mathcal{E}\{(\hat{S}(\omega))^2\} \\
&\quad= \frac{1}{N^2}\mathcal{E}\left\{\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} x[n]x[m]x[p]x[q]e^{-j\omega(n-m+p-q)}\right\} &(13)\\
&\quad= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} r[n-m]r[p-q]e^{-j\omega(n-m+p-q)} \\
&\qquad+ \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} r[n-q]r[m-p]e^{-j\omega(n-m+p-q)} \\
&\qquad+ \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} r[n-p]r[m-q]e^{-j\omega(n-m+p-q)} &(14)\\
&\quad= 2|S_1(\omega)|^2 \;+\; |S_2(\omega)|^2 &(15)
\end{aligned}$$

where

$$S_1(\omega) \;\equiv\; \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} r[n-m]e^{-j\omega(n-m)} \tag{16}$$

$$S_2(\omega) \;\equiv\; \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} r[n-m]e^{-j\omega(n+m)} \tag{17}$$

Comparing (16) to (6) we see from (8) that

$$S_1(\omega) \;=\; S(\omega) \star W_B(\omega) \tag{18}$$

On the other hand, we have

$$S_2(\omega) \;=\; \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} r[n-m]e^{-j\omega(n+m)} \tag{19}$$

[3]The Gaussian assumption is important. That of being real simplifies the notation.

$$= \frac{1}{N} \sum_{k=-(N-1)}^{N-1} \sum_{m=|k|}^{N-1-|k|} r[k] e^{-j\omega(k+2m)} \tag{20}$$

$$= \left(\frac{1}{N} \sum_{k=-(N-1)}^{-1} r[k] e^{-j\omega k} \left(\sum_{m=-k}^{N-1} e^{-j\omega 2m} \right) \right) + \left(\frac{1}{N} \sum_{k=0}^{N-1} r[k] e^{-j\omega k} \left(\sum_{m=0}^{N-1-k} e^{-j\omega 2m} \right) \right) \tag{21}$$

$$\tag{22}$$

As $N \to \infty$ the inner sums do not converge but are bounded – the bound does not grow with N – say, bounded in magnitude by C. We could therefore[4] write

$$|S_2(\omega)| \; < \; C \left| \frac{1}{N} \sum_{k=-(N-1)}^{N-1} r[k] e^{-j\omega k} \right| \tag{23}$$

and since the sum converges to the power spectrum, the term $S_2(\omega)$ is asymptotically zero. As such

$$\mathcal{E}\{(\hat{S}(\omega))^2\} = 2\left(S(\omega) \star W_B(\omega)\right)^2 \tag{24}$$

$$\mathrm{Var}\{\hat{S}(\omega)\} = \mathcal{E}\{(\hat{S}(\omega))^2\} - \left(\mathcal{E}\{\hat{S}(\omega)\}\right)^2 \tag{25}$$

$$= \left(S(\omega) \star W_B(\omega)\right)^2 \tag{26}$$

which leaves us the important message that *the periodogram is not consistent* – its variance does not decrease to zero as $N \to \infty$.

1.3 Rayleigh Resolution Limit

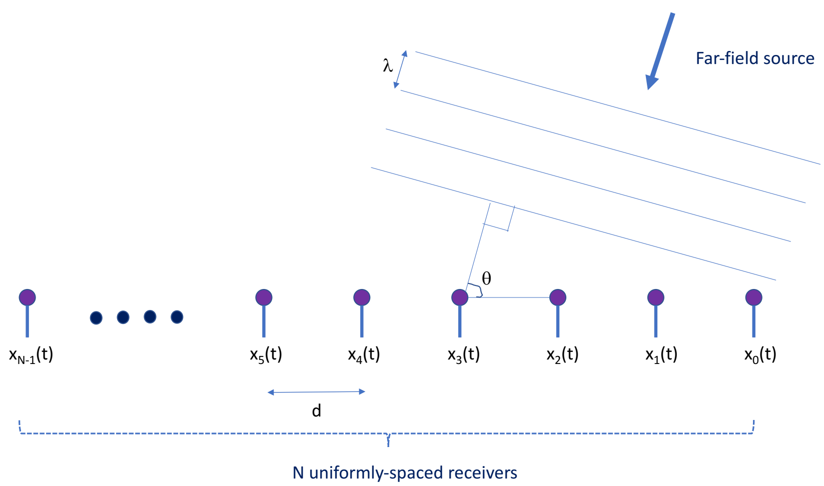
With reference to (8) and (11), two frequencies may appear, after convolution with $W_B(\omega)$, as a single spectral "bump". When this happens we say that the frequencies are not resolvable in the classical (periodogram-based) sense. Normally it is assumed such a merging happens when the two frequencies are closer together than the frequency spacing between the peak and first zero of $W_B(\omega)$, or $\frac{2\pi}{M}$. We call this the *Rayleigh resolution limit.*

1.4 Array Signal Processing

Consider a uniform linear array of sensors: microphones, hydrophones, radar receivers, etc. The uniform spacing is important for what follows here; but

[4] The argument could be made more precise, since for finite k both the inner sums converge to $\delta(\omega)$.

planar uniform arrays apply as well with greater complexity of notation. We require a far-field and monochromatic (single-frequency) source. Far-field means the wavefronts when they arrive at the sensor are planar (as opposed to curved). The monochromatic nature is important for the mathematics, but in fact one could assume that an FFT operation is occurring at the sensors, and the operations about to be described can be performed separately at each frequency and (possibly) combined. The notional setup is as pictured below.



The source is oriented at angle θ with respect to "horizontal" of the array – some people prefer to have θ with respect to broadside, the difference is will be that cos gets replaced by sin. Now suppose the source emits frequency f – the wavelength and speed of propagation are related to it as $f\lambda = c$. The signal received at the n^{th} sensor is

$$x_n(t) = A^{'} e^{j2\pi f(t - nd\cos(\theta)/c)} \tag{27}$$

where $A^{'}$ is a complex amplitude and d is the inter-sensor spacing. If all sensors sample at the same time, we could write

$$x[n] = Ae^{-j2\pi fnd\cos(\theta)/c} \tag{28}$$

$$= Ae^{-j2\pi n\left(\frac{d}{\lambda}\right)\cos(\theta)} \tag{29}$$

where we no longer need the time index and we've absorbed the phase caused by the sampling time into A. What is remarkable is that the signal now appears as a (spatial) sinusoid indexed by sensor number as opposed to time sample, and

$$\kappa = 2\pi\left(\frac{d}{\lambda}\right)\cos(\theta) \tag{30}$$

is the (spatial) frequency. An immediate consequence of this is that to avoid aliasing we need to have

$$2\pi\left(\frac{d}{\lambda}\right)\cos(\theta) \;<\; \pi \tag{31}$$

$$d\cos(\theta) \;<\; \frac{\lambda}{2} \tag{32}$$

and since $\cos(\theta) \leq 1$ this means that we must have

$$d \;<\; \frac{\lambda}{2} \tag{33}$$

in order to be sure there be no spatial aliasing at all.

Perhaps most interesting is that we see that we can apply our spectral estimation methods to the array processing problem: once we have the spatial frequency of the "sinusoid" we invert (30) to get the direction of arrival (DOA). The wrinkle is Rayleigh resolution, for which the limit is

$$2\pi\left(\frac{d}{\lambda}\right)\cos(\theta+\Delta) - 2\pi\left(\frac{d}{\lambda}\right)\cos(\theta) \;>\; \frac{2\pi}{M} \tag{34}$$

or with Δ small (and M sufficiently large),

$$\Delta \;>\; \frac{\lambda}{Md\sin(\theta)} \tag{35}$$

approximately. This gives an upper limit on Rayleigh resolution of two DOA's (we hope to do better!). One thing that is very noticeable is the deterioration of resolvability near "endfire" – when θ is close to 0 or π.

2 Nonparametric Spectral Estimation: The Bartlett and Welch Procedures

Inconsistency would seem to be a "deal-killer" for any estimator. But there is an easy fix. For data record $\{x[n]\}_{n=0}^{N-1}$, and assuming that $N = LM$, write

$$\hat{S}_i(\omega) \;\equiv\; \frac{1}{M}\left|\sum_{n=0}^{M-1} x[n+iM]e^{-j\omega n}\right|^2 \tag{36}$$

for $i = 0, \,, 1, \ldots, L-1$ and form the Bartlett spectral estimator as

$$\hat{S}_B(\omega) \;=\; \frac{1}{L}\sum_{i=0}^{L-1} S_i(\omega) \tag{37}$$

It is easy to see that

$$\mathrm{Var}\{\hat{S}_B(\omega)\} \;\approx\; \frac{1}{L}\left(S(\omega) \star W_B(\omega)\right)^2 \tag{38}$$

which indicates[5] that the Bartlett periodogram is indeed consistent. The price paid is that in this case

$$W_B(\omega) \;=\; \left(\frac{1}{M}\frac{\sin(\omega M/2)}{\sin(\omega/2)}\right)^2 \tag{39}$$

where (8) describes the mean of the Bartlett periodogram. Note that (39) does not change as N increases: the Bartlett periodogram converges, but converges to a smeared version of the power spectrum,

The Welch method somewhat improves on Bartlett in two ways: by allowing overlap (and hence better resolution due to a larger M in (39)) and by introducing windowing that can potentially reduce sidelobes (and hence eliminate interference of distant "loud" tones on quieter ones that the periodogram may be trying hard to discern). In the Welch approach we no longer require that $LM = N$, but continue with M as the length of the sections and L as the number of sections; call K the number of samples to jump between sections. We replace (36) by

$$\hat{S}_i(\omega) \;\equiv\; \frac{1}{M}\left|\sum_{n=0}^{M-1} w[n]x[n+iK]e^{-j\omega n}\right|^2 \tag{40}$$

where $w[n]$ is the window used, and the Welch periodogram $S_W(\omega)$ is formed from these exactly as $S_B(\omega)$ is in (37). Now we have

$$\mathcal{E}\{\hat{S}(\omega)\} \;=\; S(\omega) \star W(\omega) \tag{41}$$

where

$$W(\omega) \;\equiv\; \frac{1}{M}\left|\sum_{n=0}^{M-1} w[n]e^{-j\omega n}\right|^2 \tag{42}$$

Some careful analysis has shown that some degree of overlap is not too harmful: with 50% overlap (38) is increased by a factor $\frac{9}{8}$, approximately.

[5]The approximation is that the limited dependency between M-blocks of data has been ignored. Statisticians would invoke a "mixing" condition.

3 Nonparametric Spectral Estimation: MVDR

This approach, sometimes known as the MVDR (for "minimum-variance distortionless response" which is nicely descriptive) and sometimes as the Capon method (which is less so) is an excellent way to "listen" to weak frequencies (or directions) without fear of interference from other stronger ones. In fact, the stronger these interferers are, the less problems they cause. The idea is to form a "filter"

$$y_\omega[n] \;=\; \mathbf{w}(\omega)^H \mathbf{x}_n \tag{43}$$

whose output[6] represents what is present at frequency ω: the expected output $y[n]$ should contain what is at frequency ω and as little else as possible, and the expected output power is the power at that frequency. The MVDR idea is to select $\mathbf{w}(\omega)$ such that

$$\mathbf{w}(\omega) \;=\; \arg\min_{\mathbf{w}} \left\{ \mathcal{E}\{|y_\omega[n]|^2\} \right\} \text{ subject to } \mathbf{w}(\omega)^H \mathbf{q}(\omega) = 1 \tag{44}$$

where

$$\mathbf{q}(\omega) \;\equiv\; \begin{pmatrix} 1 \\ e^{-j\omega} \\ e^{-j2\omega} \\ \vdots \\ e^{-j(M-1)\omega} \end{pmatrix} \tag{45}$$

is a sinusoid (vector) at frequency ω. Notionally, then we want to listen faithfully to frequency ω (the constraint); but we want to minimize all interference (the minimization). If there is a strong frequency component at frequency ω' it is reasonable to expect the minimization to place a zero accordingly:

$$W_\omega(z)|_{z=e^{j\omega'}} \;\equiv\; \sum_{k=0}^{M-1} w_\omega[k]^* z^{-k}|_{z=e^{j\omega'}} \;=\; \mathbf{w}(\omega)^H \mathbf{q}(\omega') \;\approx\; 0 \tag{46}$$

At any rate, we have

$$\mathcal{E}\{|y_\omega[n]|^2\} \;=\; \mathbf{w}(\omega)^H \mathbf{R} \mathbf{w}(\omega) \tag{47}$$

[6] It is interesting to consider this in light of the interpretation of spectral estimation applied to array processing: one can actually *listen* in a particular direction (spatial frequency) by forming these $y[n]$'s at all (temporal) frequencies and then constructing the time-series coming from that by in the inverse DFT. The filter to be used to do this will appear shortly as (50).

and we solve the minimization via Lagrange multipliers as

$$\mathbf{R}\mathbf{w}(\omega) \;=\; \lambda \mathbf{q}(\omega) \tag{48}$$

Substituting back we have

$$\lambda \;=\; \frac{1}{\mathbf{q}(\omega)^H \mathbf{R}^{-1} \mathbf{q}(\omega)} \tag{49}$$

which gives us

$$\mathbf{w}(\omega) \;=\; \frac{\mathbf{R}^{-1}\mathbf{q}(\omega)}{\mathbf{q}(\omega)^H \mathbf{R}^{-1} \mathbf{q}(\omega)} \tag{50}$$

and hence

$$\begin{aligned} \hat{S}_{mvdr}(\omega) &= \mathcal{E}\{|y[n]|^2\} && (51)\\ &= \frac{\mathbf{q}(\omega)^H \mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1} \mathbf{q}(\omega)}{(\mathbf{q}(\omega)^H \mathbf{R}^{-1} \mathbf{q}(\omega))^2} && (52)\\ &= \frac{1}{\mathbf{q}(\omega)^H \mathbf{R}^{-1} \mathbf{q}(\omega)} && (53) \end{aligned}$$

4 Parametric Spectral Estimation: AR Modeling

4.1 The Yule-Walker Approach

This is rather obvious, given what we have seen before. Assume that we have estimated autocorrelations $\{r[k]\}_{k=0}^M$. We solve the augmented Yule-Walker equations and have thence

$$\hat{S}_{yw}(\omega) \;=\; \frac{P_M}{|1 + \sum_{k=1}^M a_k^* e^{-jk\omega}|^2} \tag{54}$$

where P_M is the same as σ_ν^2 as seen before. Levinson-Durbin will simplify the solution to the YW equations.

4.2 Maximum Entropy Spectral Estimation

It is an interesting fact that the YW spectral estimate is the *maximum-entropy* spectral estimator of the spectrum given knowledge of the M autocorrelations $\{r[k]\}_{k=0}^M$. That is, amongst all the (wss) random processes that have $\{r[k]\}_{k=0}^M$ as their first $M+1$ values, the M^{th}-order AR model is the "most random" in the sense of Shannon's entropy – it is "better" than any other AR order, or ARMA or MA or sinusoid-plus-noise (etc.) model in this sense. This course does not pre-suppose any familiarity with information theory, so we won't prove this.

The intuition is that the entropy (disorder) of a *wss* random process is related to the variance of the prediction error. Suppose we knew $\{r[k]\}_{k=0}^{M}$, and our prediction error power was σ_M^2. Now instead suppose we know more: we know $\{r[k]\}_{k=0}^{N}$, where $N > M$. It is tautologically true that we have $\sigma_N^2 \leq \sigma_M^2$, meaning knowing more autocorrelations must help in reducing entropy. The only situation in which it does not help (i.e., $\sigma_N^2 = \sigma_M^2$ for $N > M$) is when the process is AR of order M, since in that case the coefficients used to predict $u[n]$ and multiply $\{u[n-M-1], u[n-M-2], \ldots\}$ are all zero. Hence the AR process is maximally unpredictable amongst all *wss* random processes for which $\{r[k]\}_{k=0}^{M}$ are known.

4.3 Relationship to MVDR

Note that we have

$$\hat{S}_{mvdr}(\omega) = \frac{1}{\mathbf{q}(\omega)^H \mathbf{R}^{-1} \mathbf{q}(\omega)} \tag{55}$$

Now from our earlier work we know that

$$\mathbf{R}^{-1} = \mathbf{L}^H \mathbf{D}^{-1} \mathbf{L} \tag{56}$$

where

$$\mathbf{D} = \begin{pmatrix} P_0 & 0 & 0 & \ldots & 0 \\ 0 & P_1 & 0 & \ldots & 0 \\ 0 & 0 & P_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & P_M \end{pmatrix} \tag{57}$$

in which P_i is the i^{th}-order prediction error and

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ a_{1,1} & 1 & 0 & \ldots & 0 \\ a_{2,2} & a_{2,1} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{M,M} & a_{M,M-1} & a_{M,M-2} & \ldots & 1 \end{pmatrix} \tag{58}$$

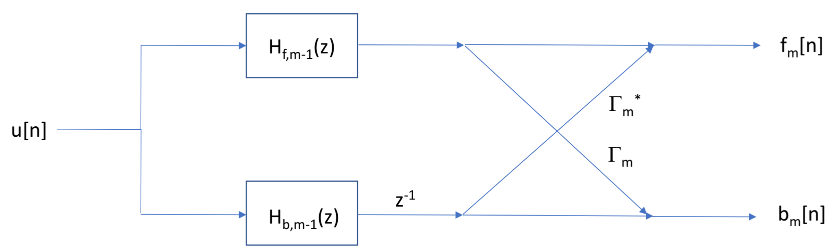
is a matrix of AR predictors. So we are able to write

$$\hat{S}_{mvdr}(\omega) = \left(\sum_{m=0}^{M} \frac{1}{P_m \hat{S}_{yw,m}(\omega)} \right)^{-1} \tag{59}$$

where $\hat{S}_{yw,m}(\omega)$ is the m^{th}-order YW spectral estimate. The MVDR spectral estimate is consequently the "parallel resistors"-weighted sum of YW spectra.

4.4 The Burg Algorithm

The YW spectral estimation approach has two steps: first estimate the correlations, then insert these to YW, presumably efficiently solved via Levinson-Durbin. The Burg approach begins from an earlier place: it assumes only that a record of data is available. There is no need to estimate correlations, Burg estimates the spectrum directly. Now, below is repeated the *lattice* interpretation of the m^{th}-order forward- and backward-error prediction filters (PEFs) from the section on linear prediction that we enjoyed earlier.



Let us suppose, as in the figure, that we have $\{f_{m-1}[n]\}$ & $\{b_{m-1}[n]\}$; that is, we are trying to find the m^{th}-order model and have worked from model order 1, then 2, all the way up to $m-1$. The notion is that we choose Γ_m to minimize the prediction error.

Let us recall from an earlier section of the course that if we posed

$$J(\mathbf{w}) \;=\; \sigma_d^2 - 2\Re\{\mathbf{w}^H\mathbf{p}\} + \mathbf{w}^H\mathbf{R}\mathbf{w} \tag{60}$$

then we could write

$$\nabla\mathbf{w}J(\mathbf{w}) \;=\; -2\mathbf{p} + 2\mathbf{R}\mathbf{w} \tag{61}$$

which is a nice reference – complex derivatives / gradients are sometimes hard to remember. We also have

$$J(w) \;=\; \sigma_d^2 - 2\Re\{w^*p\} + |w|^2R \tag{62}$$

$$\frac{dJ(w)}{dw} \;=\; -2p + 2Rw \tag{63}$$

when these are particularized to scalars – apologies that this is belabored.

Suppose we want to minimize the m^{th}-order forward prediction error

$$\mathcal{E}\{|f_m[n]|^2\} \;=\; \mathcal{E}\{|f_{m-1}[n] + \Gamma^* b_{m-1}[n-1]|^2\} \tag{64}$$

We take the gradient and get

$$0 = \nabla_\Gamma \left(\mathcal{E}\{|f_{m-1}[n] + \Gamma_m^* b_{m-1}[n-1]|^2\} \right) \quad (65)$$

$$= 2\mathcal{E}\{b_{m-1}[n-1]f_{m-1}[n]^*\} \; + \; 2\Gamma_m \mathcal{E}\{|b_{m-1}[n-1]|^2\} \quad (66)$$

or we get the minimizing reflection coefficient

$$\Gamma_m = \frac{-\mathcal{E}\{b_{m-1}[n-1]f_{m-1}[n]^*\}}{\mathcal{E}\{|b_{m-1}[n-1]|^2\}} \quad (67)$$

Now it is also interesting to minimize

$$\mathcal{E}\{|b_m[n]|^2\} = \mathcal{E}\{|b_{m-1}[n-1] + \Gamma f_{m-1}[n]|^2\} \quad (68)$$

We take the gradient and get

$$0 = \nabla_\Gamma \left(\mathcal{E}\{|b_{m-1}[n-1] + \Gamma_m f_{m-1}[n]|^2\} \right) \quad (69)$$

$$= \nabla_\Gamma \left(\mathcal{E}\{|b_{m-1}^*[n-1] + \Gamma_m^* f_{m-1}[n]^*|^2\} \right) \quad (70)$$

$$= 2\mathcal{E}\{b_{m-1}[n-1]f_{m-1}[n]^*\} \; + \; 2\Gamma_m^* \mathcal{E}\{|f_{m-1}[n]|^2\} \quad (71)$$

and we now get the minimizing reflection coefficient

$$\Gamma_m = \frac{-\mathcal{E}\{b_{m-1}[n-1]f_{m-1}[n]^*\}}{\mathcal{E}\{|f_{m-1}[n]|^2\}} \quad (72)$$

The symmetry is pleasing between the two; but it is perhaps strange to have b and f treated differently. So the Burg approach is actually to minimize

$$\mathcal{E}\{|f_m[n]|^2\} \; + \; \mathcal{E}\{|b_m[n]|^2\} \quad (73)$$

and the solution is easily seen to be

$$\Gamma_m = \frac{-2\mathcal{E}\{b_{m-1}[n-1]f_{m-1}[n]^*\}}{\mathcal{E}\{|b_{m-1}[n-1]|^2\} + \mathcal{E}\{|f_{m-1}[n]|^2\}} \quad (74)$$

The Burg spectral estimate $\hat{S}_{burg}(\omega)$ is the AR spectrum (like (54)) that uses the Γ_m's as its reflection coefficients. Levinson-Durbin offers an easy way to transform these into AR parameters (the a's), and P_M is directly estimable from $\mathcal{E}\{|f_m[n]|^2$, and the expectations necessary to calculate Γ_m are estimated from $f_{m-1}[n]$ and $b_{m-1}[n]$. Burg offers a slick way to build up the AR model step by step directly from the data. There is some evidence that the Burg spectrum is more "peaky" than the YW spectrum (i.e., sinusoids stand out more clearly). This may have to do with the fact that its zeros have to be inside the unit circle (since mathematically $|\Gamma_m| \leq 1$) whereas with estimated $r[k]$'s this may not be true for the YW estimator[7].

[7] The idea is that zeros that "want" to get arbitrarily close to the unit circle but can't "escape" it can do so with Burg; whereas with YW they can escape and become less close to the unit circle.

5 Parametric Spectral Estimation: Sinusoids in White Noise

5.1 Justification of the Sinusoid Model

Let us begin with an arbitrary (Toeplitz) correlation matrix $\mathbf{R}$, and define

$$\tilde{\mathbf{R}} \equiv \mathbf{R} - \lambda_{min}\mathbf{I} \tag{75}$$

It is clear that $\tilde{\mathbf{R}}$ shares the same eigenvectors as $\mathbf{R}$, while each of its eigenvalues is reduced by λ_{min}. There is at least one zero eigenvalue, and let us call the associated eigenvector $\mathbf{g}$. We have

$$\begin{aligned}
0 &= \mathbf{g}^H\tilde{\mathbf{R}}\mathbf{g} && (76)\\
&= \sum_{m=0}^{M-1}\sum_{n=0}^{M-1} g[m]^* g[n] r[m-n] && (77)\\
&= \sum_{m=0}^{M-1}\sum_{n=0}^{M-1} g[m]^* g[n] \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{S}(\omega) e^{j\omega(m-n)} d\omega && (78)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{S}(\omega) \sum_{m=0}^{M-1}\sum_{n=0}^{M-1} g[m]^* g[n] e^{j\omega(m-n)} d\omega && (79)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{S}(\omega)|G(\omega)|^2 d\omega && (80)
\end{aligned}$$

We have blithely and obviously defined

$$\begin{aligned}
\mathbf{g} &\equiv \begin{pmatrix} g[0] \\ g[1] \\ g[2] \\ \vdots \\ g[M-1] \end{pmatrix} && (81)\\
G(z) &\equiv \sum_{m=0}^{M-1} g[m] z^{-k} && (82)\\
G(\omega) &\equiv \sum_{m=0}^{M-1} g[m] e^{-j\omega m} && (83)\\
S(\omega) &\equiv \sum_{k=-\infty}^{\infty} r[k] e^{-j\omega k} && (84)\\
\tilde{S}(\omega) &\equiv \sum_{k=-\infty}^{\infty} [r[k] - \lambda_{min}\delta[k]] e^{-j\omega k} && (85)
\end{aligned}$$

where $\delta[k]$ is the unit impulse in the DSP sense. It is important to note that no claim is made that $S(\omega)$ be the actual power spectral density; in fact, it is only *one of* the power spectra whose first M autocorrelations match those of the true random process.

Now from (82) it is seen that $G(z)$ is a polynomial of order $M-1$, and hence it has $M-1$ roots (zeros). Some (or all) of these may be on the unit circle, so (83) can hence be zero for at most $M-1$ values. And since $|G(\omega)|^2 \geq 0$ (80) makes it clear that we have

$$\tilde{S}(\omega)|G(\omega)|^2 \;=\; 0 \quad \forall\omega \tag{86}$$

which tells us that $\tilde{S}(\omega)$ can be non-zero at only those ω's for which $G(\omega) = 0$. There are only at most $M-1$ such ω's and hence we know that we can write

$$\tilde{S}(\omega) \;=\; \sum_{n=1}^{M-1} p_n \delta(\omega - \omega_n) \tag{87}$$

$$S(\omega) \;=\; \sigma^2 + \sum_{n=1}^{M-1} p_n \delta(\omega - \omega_n) \tag{88}$$

where the p_k's are nonnegative real numbers (some can be zero), and hence

$$r[k] \;=\; \sigma^2\delta[k] + \sum_{n=1}^{M-1} p_n e^{j\omega_n k} \tag{89}$$

This (89) tells us a remarkable thing: the first M correlations of any *wss* random process can be written as the sum of a δ-function and $M-1$ complex sinusoids. Put another way – and a bit more notionally – any random process can be thought of as arising from sinusoids plus white noise. This is a backdoor proof of the Caratheodory Theorem. Note that none of this is meant to imply that *all* power spectra have the form (88); what is shown is that for any *wss* random process for which we know the fist M autocorrelations $\{r[k]\}_{k=0}^{M-1}$ there *exists* a random process consistent with those autocorrelations that has form (88). This is perhaps a statement that is parallel to that relating to AR processes: there are many *wss* random processes that have $\{r[k]\}_{k=0}^{M-1}$, but amongst them the one with maximum entropy is the AR process of order $M-1$.

We end by proffering

$$\mathbf{R} \;=\; \sigma^2\mathbf{I} + \sum_{n=1}^{M-1} p_n \mathbf{q}(\omega_n)\mathbf{q}(\omega_n)^H \tag{90}$$

in which $\mathbf{q}(\omega)$ is as in (45), as a general model the correlation matrix of a *wss* random process. Note that there is no reason to expect that the ω_n's are related either to each other or to the "DFT frequencies" – actually, their values are what need to be sought; and to be general we should allow some (or all) p_n's to be zero.

5.2 Pisarenko Harmonic Decomposition

The discussion in the previous section tells us that the eigendecomposition of $\mathbf{R}$ is key, and suggests the following prescription.

1. Estimate $\mathbf{R}$.
2. Find the minimum eigenvalue of $\mathbf{R}$: λ_{min}. We know that σ^2 in (90) is λ_{min}.
3. Find the eigenvector $\mathbf{g}$ that corresponds to λ_{min}.
4. Find the roots of $G(z)$ (see (90)).
5. Keep those roots that on the unit circle[8] and label them $z_m = e^{j\omega_m}$.
6. Solve the Vandermonde system

$$\begin{pmatrix} r[1] \\ r[2] \\ \vdots \\ r[M-1] \end{pmatrix} = \begin{pmatrix} e^{j\omega_1} & e^{j\omega_2} & \dots & e^{j\omega_{M-1}} \\ e^{j2\omega_1} & e^{j2\omega_2} & \dots & e^{j2\omega_{M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(M-1)\omega_1} & e^{j(M-1)\omega_2} & \dots & e^{j(M-1)\omega_{M-1}} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{M-1} \end{pmatrix} \tag{91}$$

This looks great. And unfortunately it doesn't work very well. The problem is in steps (1) & (3): when a correlation matrix is *estimated* rather than analytically given, the eigenvector polynomial's roots are not especially inclined to be on the unit circle. Notionally, the concern is that essentially all the estimation hard work is performed by the eigenvector corresponding to the *minimum* eigenvalue; and exactly this eigenvalue is by its nature the least well estimated.

[8] In any sort of "practice" roots that are close to the unit circle will do.

5.3 MUSIC

First, this has nothing to do with horns and violins. It stands for **mu**ltiple **si**gnal **c**lassification. Let us work with the ideas from the Pisarenko analysis. First, let us assume that we have

$$\mathbf{R} \;=\; \sigma^2\mathbf{I} + \sum_{n=1}^{L} p_n \mathbf{q}(\omega_n)\mathbf{q}(\omega_n)^H \tag{92}$$

where the only difference from (90) is that in (92) the signal is assumed to contain $L < M - 1$ sinusoids. That implies that the multiplicity of the minimum eigenvalue (i.e., σ^2) is $M - L > 1$. This is useful, since with a larger "noise-subspace" suggests more accurate estimation of it: Pisarenko works perfectly well in theory, it's the practice with estimated $\mathbf{R}$ where it can fail.

Now, note that due to the orthogonality property of the eigenvectors of a Hermitian matrix we for have all of these "minimal" eigenvectors $\{\mathbf{g}_m\}_{m=L+1}^{M}$ that

$$\mathbf{g}_m^H\mathbf{q}(\omega_n) \qquad n = 1, 2, \ldots, L \tag{93}$$

This means that the MUSIC spectral estimator

$$\hat{S}music(\omega) \;=\; \frac{1}{\sum_{m=L+1}^{M} |\mathbf{g}_m^H\mathbf{q}(\omega)|^2} \tag{94}$$

should have strong peaks at $\omega = \omega_n$, $n = 1, 2, \ldots, L$. Note that MUSIC is not really a spectral estimator, in the sense that it does not provide complete information about the true spectrum $S(\omega)$. All it tries to do – and it succeeds quite nicely – is to show the sinusoidal frequencies as peaks. In the array processing application these peaks would be DOA's.

Now, as a practical matter we can form the $\mathbf{g}$'s directly from the estimated autocorrelation matrix $\hat{\mathbf{R}}$. But we could also use the techniques that we have learned about the SVD, and form

$$\mathbf{A}^H \;=\; \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_N \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \tag{95}$$

and write

$$\mathbf{A} \;=\; \mathbf{U\Sigma V}^H \tag{96}$$

and recall that since

$$\hat{\mathbf{R}} \;=\; \frac{1}{N}\mathbf{A}^H\mathbf{A} \tag{97}$$

we consequently find the eigenvectors of $\hat{\mathbf{R}}$ in the unitary matrix $\mathbf{V}$. It is often useful to write

$$\mathbf{V} \;=\; (\mathbf{V}_s \;\; \mathbf{V}_n) \tag{98}$$

where these contain eigenvectors respectively from the "signal" and "noise" subspaces. So we could write

$$\hat{S}music(\omega) \;=\; \frac{1}{\mathbf{q}(\omega)^H \mathbf{V}_n \mathbf{V}_n^H \mathbf{q}(\omega)} \tag{99}$$

which is the noise-subspace version of MUSIC. The signal-subspace version is

$$\hat{S}music(\omega) \;=\; \frac{1}{M - \mathbf{q}(\omega)^H \mathbf{V}_s \mathbf{V}_s^H \mathbf{q}(\omega)} \tag{100}$$

Another variant of MUSIC is to write (99) as

$$\hat{S}music(z) \;=\; \frac{1}{\mathbf{z}^H \mathbf{V}_n \mathbf{V}_n^H \mathbf{z}} \tag{101}$$

where

$$\mathbf{z} \;\equiv\; \begin{pmatrix} 1 \\ z \\ z^2 \\ \vdots \\ z^{M-1} \end{pmatrix} \tag{102}$$

If we form

$$D(z) \;\equiv\; |\mathbf{V}_n^H \mathbf{z}|^2 \tag{103}$$

$$\;=\; H(z)H(1/z^*)^* \tag{104}$$

then the angles of the roots of $H(z)$ should provide the peaks of $\hat{S}_{music}(\omega)$. This is, not surprisingly, referred to as root-MUSIC.

Finally let us recall

$$\hat{S}_{mvdr}(\omega) \;=\; \frac{1}{\mathbf{q}(\omega)^H \mathbf{R}^{-1} \mathbf{q}(\omega)} \tag{105}$$

Now we can write

$$\mathbf{R} \;=\; \sum_{m=1}^{L} \lambda_m \mathbf{g}_m \mathbf{g}_m^H \;+\; \lambda_{min} \sum_{m=L+1}^{M} \mathbf{g}_m \mathbf{g}_m^H \tag{106}$$

Suppose we "enhanced" the signal subspace by a factor κ:

$$\mathbf{R}_\kappa \;=\; \sum_{m=1}^{L} \kappa \lambda_m \mathbf{g}_m \mathbf{g}_m^H \;+\; \lambda_{min} \sum_{m=L+1}^{M} \mathbf{g}_m \mathbf{g}_m^H \tag{107}$$

and thence

$$\mathbf{R}_\kappa^{-1} \;=\; \sum_{m=1}^{L} \frac{1}{\kappa\lambda_m}\mathbf{g}_m\mathbf{g}_m^H \;+\; \frac{1}{\lambda_{min}}\sum_{m=L+1}^{M}\mathbf{g}_m\mathbf{g}_m^H \tag{108}$$

It is easy to see that

$$\hat{S}_{music}(\omega) \;=\; \lim_{\kappa\to\infty}\left\{\frac{\lambda_{min}^{-1}}{\mathbf{q}(\omega)^H\mathbf{R}_\kappa^{-1}\mathbf{q}(\omega)}\right) \tag{109}$$

meaning that MUSIC is the essentially same as MVDR with asymptotic enhancement of the signal subspace.

There is one more note about MUSIC – and it's an important one. Let us go right back to (92) and re-write as

$$\mathbf{R} \;=\; \sigma^2\mathbf{I} \;+\; \sum_{n=1}^{L} p_n\mathbf{q}(\theta_n)\mathbf{q}(\theta_n)^H \tag{110}$$

where the difference is that these $\mathbf{q}$-vectors are parameterized not by frequency (ω) but in some other way (θ). An example would be that the observations $\mathbf{x}_n$ are from a general array of sensors and θ_n is a representation of the position (in three dimensions) of the n^{th} source. If we can write, via physics, the signal that we would *expect* (in a noise-free situation) to observe[9] at the array elements $\mathbf{x}_n$, then $\mathbf{R}$ according to (110) is a valid representation of the correlation matrix. The MUSIC idea works acceptably here too: when θ is "swept" along all its possible values[10], the MUSIC peaks should be observed at the θ_n's. This is why the "SI" in MUSIC is for "signal" not "sinusoid" – it's more general than just sinusoids.

5.4 The Minimum-Norm Method

In the signal-subspace version of MUSIC we recognized that $|\mathbf{V}_s\mathbf{q}(\omega_n)|^2 = M$ for any signal-space frequency ω_n; and we get a spectral peak by taking the reciprocal of $M - |\mathbf{V}_s\mathbf{q}(\omega_n)|^2$. Minimum-norm attempts to form that directly by seeking a "filter" $\mathbf{a}$ such that

$$\mathbf{V}_s^H\mathbf{a} \;=\; 0 \tag{111}$$

[9] This might, for example, be via electromagnetic modeling that accounts for all propagation paths and reflections that would be encountered by a source at θ.

[10] This may take some doing if θ is multi-dimensional. For example, if θ is two-dimensional, such as azimuth / range, then the MUSIC "spectrum" is a surface.

However, $\mathbf{V}_s^H$ is a "short / fat" matrix, so the solution is underdetermined. Naturally, then we seek the **a** with minimum norm – that is the SVD idea. Let us write (111) in linear-predictor format with

$$\mathbf{a} = \begin{pmatrix} 1 \\ -\mathbf{w} \end{pmatrix} \tag{112}$$

and likewise partition

$$\mathbf{V}_s = \begin{pmatrix} \mathbf{g}_s^T \\ \mathbf{G}_s \end{pmatrix} \tag{113}$$

$$\mathbf{V}_n = \begin{pmatrix} \mathbf{g}_n^T \\ \mathbf{G}_n \end{pmatrix} \tag{114}$$

which isolates the top rows of the two matrices. We have from (111)

$$\begin{aligned} 0 &= \mathbf{V}_s^H \mathbf{a} && (115) \\ &= \begin{pmatrix} \mathbf{g}_s^* & \mathbf{G}_s^H \end{pmatrix} \begin{pmatrix} 1 \\ -\mathbf{w} \end{pmatrix} && (116) \\ \mathbf{G}_s^H \mathbf{w} &= \mathbf{g}_s^* && (117) \\ \mathbf{G}_s^T \mathbf{w}^* &= \mathbf{g}_s && (118) \end{aligned}$$

We seek to minimize $\mathbf{w}^H\mathbf{w}$ subject to (118). We have

$$\begin{aligned} \nabla \left(\mathbf{w}^H \mathbf{w} - 2\lambda^T (\mathbf{G}_s^T \mathbf{w}^* - \mathbf{g}_s) \right) &= 0 && (119) \\ \mathbf{w} &= \mathbf{G}_s \lambda && (120) \end{aligned}$$

so reinstatement of the constraint gives us

$$\begin{aligned} (\mathbf{G}_s^T \mathbf{G}_s^*) \lambda^* &= \mathbf{g}_s && (121) \\ (\mathbf{G}_s^H \mathbf{G}_s) \lambda &= \mathbf{g}_s^* && (122) \\ \lambda &= \left(\mathbf{G}_s^H \mathbf{G}_s \right)^{-1} \mathbf{g}_s^* && (123) \\ \mathbf{w} &= \mathbf{G}_s \left(\mathbf{G}_s^H \mathbf{G}_s \right)^{-1} \mathbf{g}_s^* && (124) \end{aligned}$$

Let us simplify. We have

$$\begin{aligned} \mathbf{I} &= \mathbf{V}_s^H \mathbf{V}_s && (125) \\ &= \begin{pmatrix} \mathbf{g}_s^* & \mathbf{G}_s^H \end{pmatrix} \begin{pmatrix} \mathbf{g}_s^T \\ \mathbf{G}_s \end{pmatrix} && (126) \\ \mathbf{G}_s^H \mathbf{G}_s &= \mathbf{I} - \mathbf{g}_s^* \mathbf{g}_s^T && (127) \\ \left(\mathbf{G}_s^H \mathbf{G}_s \right)^{-1} &= \mathbf{I} + \frac{\mathbf{g}_s^* \mathbf{g}_s^T}{1 - \mathbf{g}_s^T \mathbf{g}_s^*} && (128) \end{aligned}$$

where (128) follows via the matrix-inversion lemma. We thus substitute back to (124) to get

$$\mathbf{w} = \mathbf{G}_s \left(\mathbf{I} + \frac{\mathbf{g}_s^* \mathbf{g}_s^T}{1 - \mathbf{g}_s^T \mathbf{g}_s^*} \right) \mathbf{g}_s^* \tag{129}$$

$$= \mathbf{G}_s \frac{\mathbf{g}_s^* - (\mathbf{g}_s^T \mathbf{g}_s^*)\mathbf{g}_s^* + \mathbf{g}_s^*(\mathbf{g}_s^T \mathbf{g}_s^*)}{1 - \mathbf{g}_s^T \mathbf{g}_s^*} \tag{130}$$

$$= (1 - \mathbf{g}_s^T \mathbf{g}_s^*)^{-1} \mathbf{G}_s \mathbf{g}_s^* \tag{131}$$

hence

$$\mathbf{a} = \begin{pmatrix} 1 \\ -(1 - \mathbf{g}_s^T \mathbf{g}_s^*)^{-1} \mathbf{G}_s \mathbf{g}_s^* \end{pmatrix} \tag{132}$$

An expression equivalent to (132) is also available in terms of the noise subspace. Write

$$\mathbf{I} = \mathbf{V}\mathbf{V}^H \tag{133}$$

$$= (\mathbf{V}_s \ \ \mathbf{V}_n) \begin{pmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{pmatrix} \tag{134}$$

$$= \begin{pmatrix} \mathbf{g}_s^T & \mathbf{g}_n^T \\ \mathbf{G}_s & \mathbf{G}_n \end{pmatrix} \begin{pmatrix} \mathbf{g}_s^* & \mathbf{G}_s^H \\ \mathbf{g}_n^* & \mathbf{G}_n^H \end{pmatrix} \tag{135}$$

hence we have (136)-(138)

$$\mathbf{g}_s^T \mathbf{g}_s^* + \mathbf{g}_n^T \mathbf{g}_n^* = 1 \tag{136}$$

$$\mathbf{g}_s^T \mathbf{G}_s^H + \mathbf{g}_n^T \mathbf{G}_n^H = \mathbf{0} \tag{137}$$

$$\mathbf{G}_s \mathbf{G}_s^H + \mathbf{G}_n \mathbf{G}_n^H = \mathbf{I} \tag{138}$$

We can therefore write

$$\mathbf{a} = \begin{pmatrix} 1 \\ (\mathbf{g}_n^T \mathbf{g}_n^*)^{-1} \mathbf{G}_n \mathbf{g}_n^* \end{pmatrix} \tag{139}$$

which is an alternative expression for (132). We can write

$$\hat{S}_{mn}(\omega) = \frac{1}{|\mathbf{a}^H \mathbf{q}(\omega)|^2} \tag{140}$$

and either (132) or (139) can be used.

An interpretation is as follows. We “enhance” the correlation matrix to

$$\mathbf{R}' \equiv \lim_{\kappa \to \infty} \left\{ \frac{1}{\kappa} \mathbf{R}_\kappa \right\} \tag{141}$$

$$= \mathbf{V}_s \mathbf{V}_s^H \tag{142}$$

where $\mathbf{R}_\kappa$ is as in (107) – that is, $\mathbf{R}'$ contains only the signal subspace. We want to find a filter $\mathbf{a}$ of the form (112) such that the output power is zero – then the frequency response is zero at the frequencies contained in the signal subspace and (140) has (∞) peaks at those frequencies. But the output power of the minimum-norm filter $\mathbf{a}$ is

$$|\mathbf{R}'\mathbf{a}|^2 \;=\; 0 \tag{143}$$

or

$$|\mathbf{V}_s^H\mathbf{a}|^2 \;=\; 0 \tag{144}$$

or

$$\mathbf{G}_s^H\mathbf{w} \;=\; \mathbf{g}_s^* \tag{145}$$

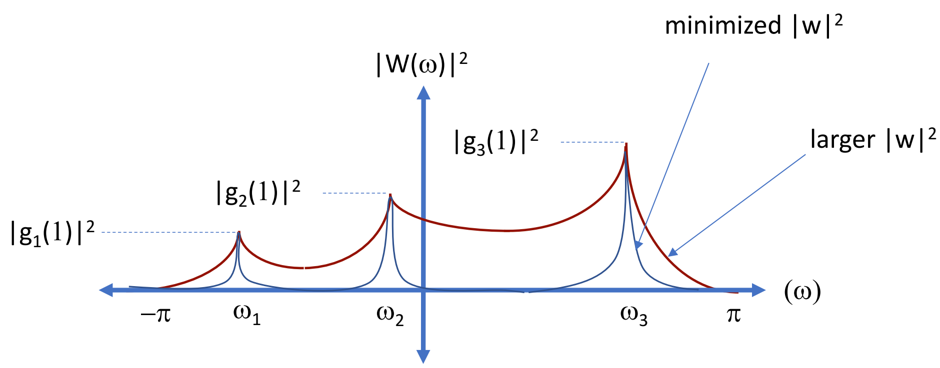
Writing the n^{th} row of (145) and conjugating, we have

$$\sum_{m=1}^{M=1} w[m]^* e^{j\omega_n m} \;=\; g_s[n] \tag{146}$$

which specifies that $|W(\omega_n)|^2 = |g_s[n]|^2$ Now, (145) is underdetermined for $\mathbf{w}$; hence the "minimum-norm" idea is to minimize $\mathbf{w}^H\mathbf{w}$. The reason this is interesting is that

$$|\mathbf{w}|^2 \;=\; \sum_{m=1}^{M-1} |w[m]|^2 \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} |W(\omega)|^2 d\omega \tag{147}$$

by Parseval. So if we minimize $|\mathbf{a}|^2 = 1 + |\mathbf{w}|^2$ we are actually minimizing the area under the integral of the magnitude-squared prediction filter, which – notionally at least – forces the filter to sharpen its focus on sinusoids. This is as pictured below.



5.5 ESPRIT

Actually the same person invented both MUSIC and ESPRIT (Professor Thomas Kailath), hence they have cool names. ESPRIT stands for **e**stimation of **s**inusoid **p**arameters by **r**otational-**i**nvariant **t**echniques – whose relevance is perhaps a little murky, but which does sound quite uplifting. Suppose we write as usual when looking for sinusoids

$$x[n] = \sum_{l=1}^{L} b_l e^{j\omega_l n} + w[n] \tag{148}$$

where $w[n]$ is the usual AWGN. In matrix form we have

$$\mathbf{x}_n = \begin{pmatrix} x[n] \\ x[n-1] \\ x[n-2] \\ \vdots \\ x[n-M+1] \end{pmatrix} \tag{149}$$

$$= \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{-j\omega_1} & e^{-j\omega_2} & \dots & e^{-j\omega_L} \\ e^{-j2\omega_1} & e^{-j2\omega_2} & \dots & e^{-j2\omega_L} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j(M-1)\omega_1} & e^{-j(M-1)\omega_2} & \dots & e^{-j(M-1)\omega_L} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_L \end{pmatrix} + \begin{pmatrix} w[n] \\ w[n-1] \\ w[n-2] \\ \vdots \\ w[n-M+1] \end{pmatrix} \tag{150}$$

$$= \mathbf{Sb} + \mathbf{w}_n \tag{151}$$

where $\mathbf{S}$ is $M \times L$. Now suppose we write $y[n] = x[n+1]$. Then we have

$$\mathbf{y}_n = \mathbf{S\Omega}^* \mathbf{b} + \mathbf{w}_n \tag{152}$$

where

$$\mathbf{\Omega} = \begin{pmatrix} e^{-j\omega_1} & 0 & 0 & \dots & 0 \\ 0 & e^{-j\omega_2} & 0 & \dots & 0 \\ 0 & 0 & e^{-j\omega_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & e^{-j\omega_L} \end{pmatrix} \tag{153}$$

Now we have

$$\mathbf{R}_{xx} \;=\; \mathbf{SPS}^H \;+\; \sigma^2\mathbf{I} \tag{154}$$

where[11] $\mathbf{R}_{xx} \equiv \mathcal{E}\{\mathbf{x}_n\mathbf{x}_n^H\}$ and

$$\mathbf{P} \;=\; \mathcal{E}\{\mathbf{bb}^H\} \;=\; \begin{pmatrix} P_1 & 0 & 0 & \dots & 0 \\ 0 & P_2 & 0 & \dots & 0 \\ 0 & 0 & P_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & P_L \end{pmatrix} \tag{155}$$

and $P_i \equiv \mathcal{E}\{|b_i|^2\}$. We also have

$$\mathbf{R}_{xy} \;=\; \mathbf{SP\Omega S}^H \;+\; \sigma^2\mathbf{\Gamma} \tag{156}$$

where $\mathbf{R}_{xy} \equiv \mathcal{E}\{\mathbf{x}_n\mathbf{y}_n^H\}$ and

$$\mathbf{\Gamma} \;\equiv\; \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \tag{157}$$

Define

$$\mathbf{C}_{xx} \;\equiv\; \mathbf{R}_{xx} \;-\; \sigma^2\mathbf{I} \tag{158}$$

$$\mathbf{C}_{xy} \;\equiv\; \mathbf{R}_{xy} \;-\; \sigma^2\mathbf{\Gamma} \tag{159}$$

Then solving the generalized eigenvalue equation

$$(\mathbf{C}_{xx} \;-\; \lambda\mathbf{C}_{xy})\mathbf{g} \;=\; 0 \tag{160}$$

is tantamount to looking for solutions λ to

$$\mathbf{SPS}^H \;-\; \lambda\mathbf{SP\Omega S}^H \;=\; 0 \tag{161}$$

$$\mathbf{SP}(\mathbf{I} \;-\; \lambda\mathbf{\Omega})\mathbf{S}^H \;=\; 0 \tag{162}$$

yields the sinusoids $\{e^{j\omega_l}\}$ directly as the solutions λ. The solution to (160) is sometimes known as a *matrix pencil.* We can write (160) as

$$(\mathbf{C}_{xx}\mathbf{C}_{xy}^{-1} \;-\; \lambda\mathbf{I})(\mathbf{C}_{xy}\mathbf{g}) \;=\; 0 \tag{163}$$

[11]The ponderous subscript notation is necessary here.

or

$$(\mathbf{C}_{xx}\mathbf{C}_{xy}^{-1})(\mathbf{C}_{xy}\mathbf{g}) \;=\; \lambda(\mathbf{C}_{xy}\mathbf{g}) \tag{164}$$

which is a standard equation for eigenstuff. So λ and $\mathbf{C}_{xy}\mathbf{g}$ from (164) solve (160). There are more-efficient solutions, however. And beyond our scope here is that ESPRIT is actually a total least-squares (TLS) solution.