Spectral estimation – Reading Assignments

Saif K. Mohammed and Erik G. Larsson

In the next three lectures, we aim to study techniques used to estimate the power spectrum of random processes. We consider a discrete-time model, where we are given a discrete-time sequence of zero mean random variables $\{y(t)\ ,\ t = 0, \pm 1, \pm 2, \dots\}$ which are assumed to be the output of a second-order stationary random process, sampled at integral multiples of the sampling period $T_s = 1/F_s$ where F_s is the sampling frequency. Second-order stationarity implies that the autocorrelation

$$\mathbb{E}\big[y(t)y^*(t-k)\big] \tag{1}$$

is only a function of k, and would be henceforth denoted by $r(k)$. The sequence $\{r(k)\}$ is sometimes referred to as the Auto Covariance Sequence (ACS). $r(k)$ satisfies some important properties like

$$\begin{aligned} r(k) &= r^*(-k) \\ r(0) &\geq |r(k)| \end{aligned} \tag{2}$$

The first equality follows from the definition in (1). The second inequality is left as a small exercise for the reader (see 1.3.4-1.3.6 in [1] for a proof).

The DTFT (Discrete-Time-Fourier-Transform) of the ACS is given by

$$\phi(w) = \sum_{k=-\infty}^{\infty} r(k)e^{-jwk} \tag{3}$$

where $w \in [-\pi, \pi)$ is the angular frequency, and is related to the frequency ($f\ ,\ |f| \leq F_s/2$) in Hz, as $w = 2\pi f/F_s$. The inverse relation to the DTFT in (3) is given by

$$r(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(w)e^{jwk}dw. \tag{4}$$

Substituting $k = 0$ in (4) we have

$$r(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(w)dw. \tag{5}$$

Further, from the definition of $r(k)$ in (1) it follows that $r(0)$ is the average power of the random sequence $\{y(t)\}$.

In fact, it can be shown that $\phi(w)$ has the physical interpretation of being the density function for the distribution of power across the frequency spectrum. That is, $\phi(w)\Delta w/\pi$ is the average power of $\{y(t)\}$ in the spectral range $(w - \Delta w\ ,\ w + \Delta w)$. Therefore, $\phi(w)$ is commonly referred to as the *power spectral density (PSD)* function of $y(t)$.

The *Spectral Estimation problem* can now be stated as follows

From a finite-length record $\{y(1), y(2), \dots, y(N)\}$ *of a second-order stationary random process, determine an estimate* $\widehat{\phi}(w)$ *of its power spectral density function* $\phi(w)$.

The spectral estimation problem is daunting since we are required to estimate an infinite number of values for $\phi(w)$ from a finite number of observations.

Nevertheless there are techniques to deal with this problem. Though we cannot cover all the possible techniques in this course, we at least aim to cover the most important ones. Roughly speaking, most of the known techniques so far can be broadly classified as “Parametric” or “Non-Parametric” methods. “Non-parametric” methods constitute the classical methods for spectral estimation, and are generally used when it is difficult to assume a signal model for $y(t)$. On the other hand when $y(t)$ is known to have an underlying parametric signal model (with high enough degree of confidence) then estimating $\phi(w)$ is tantamount to estimating the parameters of the underlying signal model. These methods are therefore referred to as “Parametric” methods for spectral estimation. Examples where “Parametric” methods can be used include situations in which the source signal is known to have very narrow peaks in its power spectrum. Another example could be in situations where we have sinusoids buried in noise. It is generally observed that the parametric methods provide better spectral estimates than the non-parametric methods , when the chosen signal model is a good fit to the observed random process.

We would roughly divide the spectral estimation techniques that we intend to cover in this course into three topics, with one lecture devoted to each topic.

The first lecture would aim to cover the non-parametric methods for spectral estimation, for which an introduction and specific reading assignments are detailed in Section I.

The second lecture would focus on parametric methods for spectral estimation. We would cover the AR, MA and the ARMA parametric signal models and techniques for parameter estimation (like for example the Yule-Walker method). We would also discuss computationally efficient algorithms for the Yule-Walker method (for e.g. the Levinson Durbin algorithm). Please refer to Section II for more details.

The topic for the third lecture would be parametric methods for line-spectra, i.e., estimating the PSD of processes which have discrete-spectra, like for example sinusoids buried in noise. Specific methods covered would be the Non-linear Least Squares method and the ESPIRIT method. Another known method is that of MUSIC, which we have already covered as a method for estimating the angle of arrivals in array processing applications. We would also discuss the Filter Bank methods for spectral estimation with periodogram and the other windowing techniques as a special case. Of special mention would be deriving and understanding the data-dependent Capon filter method. Please refer to Section III for more details.

I. NON-PARAMETRIC METHODS

The most common methods are the *periodogram* and the correlogram method, and are basically equivalent to each other (Historically the periodogram method is more than 100 years old). The periodogram has good resolution (can resolve closely spaced spectral peaks, as long as the spacing is greater than or equal to $O(1/N)$) and the spectral estimates are asymptotically unbiased. However the variance of the spectral estimates does not decrease with increasing observation length. This issue is resolved to some extent by *windowing techniques*, like for example the Blackman-Tukey windowing technique. The periodogram is actually a special case of the Blackman-Tukey method where the window is rectangular. For most windowing techniques, there is an underlying trade-off between spectral resolution and variance of the spectral estimates. Typically, low variance implies loss of resolution and high resolution implies high variance. Therefore a pertinent question is whether or not the spectral estimates with these non-parametric methods are consistent ? The general answer is that, all non-parametric techniques are known to be *inconsistent* estimators of $\phi(w)$.

Reading assignments for this lecture, would aim to cover the above discussed aspects of the periodogram and windowing methods. Precise reading assignments are as follows:

1) Derive the periodogram and the correlogram methods, and show their equivalence. (Care should be taken to derive estimates of $r(k)$ (i.e., $\widehat{r}(k)$) from $\{y(t)\}$ in such a way that the estimated

$$\widehat{\phi}(w) = \sum_{k=-\infty}^{\infty} \widehat{r}(k) e^{-jwk} \tag{6}$$

is non-negative.

2) Bias analysis : Show that the periodogram is asymptotically (as $N \to \infty$) unbiased. Define spectral resolution and show that it is $O(1/N)$ for the periodogram method.

3) Variance analysis : Taking the example of a Gaussian white noise random process, show that the variance of the spectral estimates does not go to zero as $N \to \infty$. Show that this is also true if the observed random process is the output of a linear filter whose input is Gaussian white noise.

An intuition based reason of why this would happen for almost all second-order stationary processes is as follows. Basically $\widehat{r}(k)$ is an $O(1/\sqrt{N})$ estimate of $r(k)$. Therefore, with N observed samples, the summation in (6) has $2N-1$ summands each with an $O(1/\sqrt{N})$ error term. All these $O(1/\sqrt{N})$ error terms can add up to an overall error whose variance does not go to zero as $N \to \infty$.

4) In practice, the PSD is generally estimated at a finite number of points in the interval $[-\pi, \pi)$. Well know techniques for periodogram computation include FFT (Fast Fourier Transform).

5) Discuss the Blackman-Tukey windowing method as a method to improve the variance of the spectral estimates at the cost of reduced resolution.

Show that the Blackman-Tukey spectral estimator corresponds to a “locally” weighted average of the periodogram.

Show that the choice of the window length decides the trade-off between spectral resolution and statistical variance. To be precise, if the window length is $M < N$, then the resolution is $O(1/M)$ and the statistical variance is $O(M/N)$.

6) Solve exercise $C2.19$ in [1] (Specifically the problems (a), (b), (c) on both the bias and the variance properties of the periodogram estimate).

Matlab source code for computing the periodogram, Blackman–Tukey estimates is available in [2].

Most of the above topics are covered in [1]. A set of slides can also be found in [3].

II. Parametric methods

In many scenarios, the random process under observation can be closely approximated with a parametric signal model having much fewer parameters than the number of observations. The problem of spectral estimation is then simply the estimation of the parameters in the underlying signal model. If the number of parameters are few in number compared to the number of observed samples and the underlying model is an accurate enough approximation of the observed random process, then these methods usually give consistent spectral estimates.

Among the class of random processes which are suited for parametric methods, a very common class is the class of processes whose PSD is a rational function of e^{-jw} (i.e., a ratio of two polynomials in e^{-jw}).

$$\phi(w) = \frac{\sum_{k=-m}^{m} \gamma_k e^{-jwk}}{\sum_{k=-n}^{n} \rho_k e^{-jwk}} \tag{7}$$

where $\gamma_{-k} = \gamma_k^*$ and $\rho_{-k} = \rho_k^*$. An important observation here is that, the Weierstrass Theorem in Calculus asserts that any continuous PSD can be approximated arbitrarily closely by a rational PSD of the form (7) with sufficiently large degrees m and n. This observation is another reason why the rational PSD model in (7) is of significant interest among researchers in the spectral estimation community.

Further since $\phi(w) \geq 0$, from the Spectral Factorization Theorem it follows that (7) can be expressed as

$$\phi(w) = \left|\frac{B(w)}{A(w)}\right|^2 \sigma^2 \tag{8}$$

where $\sigma > 0$ and

$$\begin{aligned} A(w) &= \sum_{i=0}^{n} a_i e^{-jwi} \quad a_0 = 1 \\ B(w) &= \sum_{i=0}^{n} b_i e^{-jwi} \quad b_0 = 1 \end{aligned} \tag{9}$$

With $z = e^{jw}$, the PSD in the Z-domain can be expressed as

$$\phi(z) = \sigma^2 \frac{B(z)B^*(1/z^*)}{A(z)A^*(1/z^*)} \tag{10}$$

where

$$\begin{aligned} A(z) &= \sum_{i=0}^{n} a_i z^{-i} \\ A^*(1/z^*) = [A(1/z^*)]^* &= \sum_{i=0}^{n} a_i^* z^i \end{aligned} \tag{11}$$

and similarly for $B(z)$ and $B^*(1/z^*)$. Since we have assume that $\phi(w)$ is finite and does not vanish for any w, it follows that $A(z)$ and $B(z)$ do not have any zero on the unit circle. Further, in (10) we can assume that the zeros of both $A(z)$ and $B(z)$ are inside the unit circle. This is always possible since, if z is a zero of $B(z)B^*(1/z^*)$ then so is $1/z^*$, and the same holds true for $A(z)A^*(1/z^*)$. Hence the zeros of both the numerator and the denominator in (10) can be rearranged to get the desired form.

From (8), we also observe that $\phi(w)$ is the PSD of a signal obtained by filtering white noise of power σ^2 through a filter with the Z-transfer function $B(z)/A(z)$. Also this filter is stable, since all its poles are within the unit circle. We can therefore write the output time-domain sequence as

$$\begin{aligned} y(t) &= \frac{B(q)}{A(q)} e(t) \\ A(q)y(t) &= B(q)e(t) \end{aligned} \tag{12}$$

where

$$\begin{aligned} q^{-1} &= \text{unit time delay operator} \\ e(t) &= \text{white noise of variance } \sigma^2 \end{aligned} \tag{13}$$

Therefore, using the spectral factorization theorem, we have been able to transform the rational PSD in (7) into the *signal model* in (12). The spectral estimation problem is therefore to simply estimate the parameters $\{a_i\}$, $\{b_i\}$ and σ^2.

The class of processes with rational PSD is further divided into three sub-classes of processes.

i) ARMA (Auto-regressive Moving average)

$$A(q)y(t) = B(q)e(t) \tag{14}$$

ii) AR (Auto-regressive)

$$A(q)y(t) = e(t) \tag{15}$$

iii) MA (Moving average)

$$y(t) = B(q)e(t) \tag{16}$$

Assuming the underlying signal model in (12), we can estimate the signal model parameters $\{a_i\}$, $\{b_i\}$ and σ^2 from the observed sample sequence $(y(1), \ldots, y(N))$. Once we have the parameter estimates $\{\hat{a}_i\}_{i=1}^n$, $\{\hat{b}_i\}_{i=1}^m$ and $\hat{\sigma}^2$, the estimated PSD is given by

$$\widehat{\phi}(w) = \left|\frac{\widehat{B}(w)}{\widehat{A}(w)}\right|^2 \hat{\sigma}^2 \tag{17}$$

where

$$\begin{aligned} \widehat{A}(w) &= \sum_{i=0}^{n} \hat{a}_i e^{-jwi} \quad \hat{a}_0 = 1 \\ \widehat{B}(w) &= \sum_{i=0}^{n} \hat{b}_i e^{-jwi} \quad \hat{b}_0 = 1 \end{aligned} \tag{18}$$

Another important question is regarding the choice of the polynomial degrees m and n respectively. Generally, either this information is known *a priori*, or else the PSD estimates are evaluated for a range of values of m, n and the best is chosen according to some criterion.

In this lecture our main focus would be on methods to estimate the signal parameters of the underlying signal model. Here is a precise list of suggested reading assignment.

1) Derive the Yule-Walker method for estimating the signal parameters of the AR(m) signal model used to model random processes whose PSD have narrow peaks.

2) When there is no *a priori* information about the degrees of the polynomial, signal model parameter estimation is usually done for a range of values of the degree of the polynomials. Estimating the parameters for so many signal models (a different signal model for each value of n) can be prohibitive in terms of complexity. However, due to the Hermitian-Toeplitz structure of the auto covariance matrix of an AR process, it is possible to calculate the signal parameters incrementally. That is when going from a signal model with polynomial degree n to a signal model with polynomial degree $n+1$, we need not re-calculate the parameters of the new signal model using the YW method. Instead, the parameters of the signal model with polynomial degree $n+1$ can be calculated as an update on the parameters of the signal model with polynomial degree n. This is basically what the Levinson Durbin algorithm does.

Study the Levinson Durbin algorithm and state the gain achieved in terms of savings in computational complexity.

3) For processes whose underlying signal model is MA, the PSD depends only on a *finite* number of autocorrelation lags, and therefore the Blackman-Tukey windowing method is directly applicable in this case. Since we have already covered the Blackman-Tukey method in the first lecture, we can refer to the first lecture for details.

4) For ARMA signals, a modified Yule-Walker method can be used to first compute the AR parameters, followed by computing the MA parameters. Interested readers are referred to Section 3.7 in [1].

5) Solve exercise C 3.16 (a,e,f,g) in [1]. The matlab programs are readily available at [2].

In general, the non-parametric methods result in consistent and better spectral estimates than the parametric methods as long as the underlying signal model is a good approximation to the observed random process.

III. Parametric methods for Line Spectra and the Filter Bank method

In the first half of the third lecture, continuing with the parametric approach taken in lecture 2, we would firstly focus on parametric methods for estimating PSD of random processes having discrete spectra. In the second half of this lecture, we would study Filter Bank methods for spectral estimation.

For random processes with discrete spectra, the time-domain signal model used is

$$\begin{aligned} y(t) &= x(t) + e(t) \\ x(t) &= \sum_{k=1}^{n} \alpha_k e^{j(w_k t + \varphi_k)} \end{aligned} \quad (19)$$

where $x(t)$ denotes the noise-free complex valued sinusoidal signal; $\{\alpha_k\}$, $\{w_k\}$, $\{\varphi_k\}$ being its amplitude, angular frequencies and initial phases. $e(t)$ is an additive observation noise. The initial phases are usually assumed to be independent random variables distributed uniformly in $[-\pi, \pi)$.

With the underlying signal model in (19), the PSD of $y(t)$ can be shown to be

$$\phi(w) = 2\pi \sum_{p=1}^{n} \alpha_p^2 \delta(w - w_p) + \sigma^2 \quad (20)$$

where $\delta(w - w_p)$ is the dirac-delta function having the property

$$\int_{-\pi}^{\pi} F(w)\delta(w - w_p)dw = F(w_p). \quad (21)$$

From (20) it is clear that the PSD has a noise floor of σ^2 and has n vertical lines (or impulses) located at the sinusoidal frequencies $\{w_k\}$. Owing to this appearance, $\phi(w)$ is commonly referred to as a line or discrete spectrum.

From the PSD in (20) it is clear that, estimating $\phi(w)$ is equivalent to estimating the parameters $\{w_k\}$, $\{\alpha_k\}$ and σ^2. In many applications, the parameters of interest are usually $\{w_k\}$. Even otherwise, we could first estimate $\{w_k\}$ by any of the methods we discuss in this lecture. With the estimated frequencies, we can then formulate a linear regression problem to estimate the amplitudes $\{\alpha_k\}$ (for more details please see 4.1.10 and 4.1.11 in [1]).

Most of the methods known to estimate the frequencies assuming the sinusoidal signal model are called *high-resolution* methods since they can achieve a spectral resolution better than $1/N$ (Note that the periodogram achieved a spectral resolution of $1/N$). In addition, the frequency estimates of these high-resolution methods are consistent. Most of the methods that we intend to cover in this lecture assume knowledge of the number of sinusoidal components, i.e. n. In general, n can be estimated using model order selection techniques covered in a previous lecture in this course.

The methods for frequency estimation, that we would discuss are

i) Non-linear least squares method, where the estimation problem is posed as a non-linear least squares problem.

$$(\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\varphi}}) = \arg \min_{\boldsymbol{\omega}, \boldsymbol{\alpha}, \boldsymbol{\varphi}} \sum_{t=1}^{N} \left| y(t) - \sum_{k=1}^{n} \alpha_k e^{j(w_k t + \varphi_k)} \right|^2 \quad (22)$$

where $\boldsymbol{\omega}$ is the vector of the angular frequencies, and similarly for $\boldsymbol{\alpha}$ and $\boldsymbol{\varphi}$. Note that this method corresponds to the maximum likelihood estimate if the noise process is Gaussian.

This method has implicitly been discussed (basically the method of parameter concentration) in a previous lecture where we discussed the paper on Cramer-Rao bound, MUSIC and ML for the array processing applications.

ii) Methods based on the covariance matrix model where the information about the angular sinusoidal frequencies is present in the noise subspace corresponding to the autocorrelation matrix of

$$\tilde{y}(t) = (y(t), y(t-1), \cdots, y(t-m+1))^T. \quad (23)$$

These methods are usually called subspace methods, one of which (MUSIC) has already been discussed in a previous lecture. Another well known subspace method is ESPIRIT (Estimation of signal parameters via Rotational Invariance techniques).

Precise reading assignments would involve,

1) Show that if the spacing between the sinusoidal angular frequencies is larger than $1/N$, then the periodogram method can also resolve all the n frequencies.

In the second half of this lecture, we will focus on Filter bank methods for spectral estimation. The basic philosophy of these methods, is to filter the signal $y(t)$ through a very narrow band filter centered at w, followed by a measurement of the average power at the filter output. If $\phi(w)$ is fairly constant in the passband of this filter, then a simple scaling of the measured average power at the filter output would result in an accurate enough estimate of the PSD $\phi(w)$. To estimate $\phi(w)$ at many different values of w would therefore need a "bank" of filters, and hence the name Filter Bank method. It can be shown that the periodogram, and the windowing methods are special cases of the Filter bank method, where only one filter output sample is used to measure the average power at the filter output.

Precise reading assignments would include

1) Discuss the Filter Bank interpretation of the periodogram and explain why the variance of the periodogram does not decrease with increasing N.

Using the filter bank approach motivate as to why the modified periodogram methods like Bartlett and Welch have a reduced variance at the cost of loss in resolution.

2) Slepian baseband FIR filters. The Slepian baseband filters are designed to maximize the power of the in-band filtered output signal relative to the total power of the filtered output signal, when the filter is excited with white noise.

Present a derivation of the Slepian baseband filter's impulse response. Note that this would involve a result on the eigen vectors of real symmetric Toeplitz matrices as $N \to \infty$.

3) The spectral estimates derived using the Slepian baseband filter are designed to be selective for a white noise input. It might be valuable to design a filter which is as selective as possible, not for the white noise input, but for the observed random process. This is the main idea behind the Capon filter, which is essentially a data-dependent bandpass filter. Compared to the Slepian filters, the Capon filter usually has a higher spectral resolution, i.e., a higher ability to resolve finer details in the PSD $\phi(w)$.

Derive the Capon filter. Comment on using the Capon filter to estimate sinusoids in noise.

REFERENCES

[1] P. Stoica and R. Moses, "Introduction to Spectral Analysis, *Prentice Hall*, 1997.
[2] www.prenhall.com/~stoica
[3] http://www2.ece.ohio-state.edu/ randy/SAtext/