Chapter 12 Spectrum Estimation

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30010, R.O.C.

Ergodicity based on Shift-invariant Event

Before we present the "ergodicity" perspective in the textbook, some preliminary and historical background on this term is given.

Definition (Shift-invariant event for one-sided processes) For a one-sided random process $\boldsymbol{X} = \{X_1, X_2, \ldots\}$ with alphabet $\mathcal{X} \subseteq \mathbb{R}$, let $\mathcal{X}^\infty$ be the set of all sequences $\boldsymbol{x} \triangleq (x_1, x_2, x_3, \ldots)$ of real numbers in $\mathcal{X}$. Denote by $\mathcal{F}_{\boldsymbol{X}}$ the smallest σ-field generated by all open sets of $\mathcal{X}^\infty$ (i.e., the Borel σ-field of $\mathcal{X}^\infty$). Then, an event E in $\mathcal{F}_{\boldsymbol{X}}$ is said to be $\mathbb{T}$-*invariant* with respect to the left-shift (or shift transformation) $\mathbb{T} : \mathcal{X}^\infty \to \mathcal{X}^\infty$ if

$$\mathbb{T}E \subseteq E,$$

where

$$\mathbb{T}E \triangleq \{\mathbb{T}\boldsymbol{x} : \boldsymbol{x} \in E\} \quad \text{and} \quad \mathbb{T}\boldsymbol{x} \triangleq \mathbb{T}(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots).$$

- In other words, $\mathbb{T}$ is equivalent to "chopping the first component."

Example. With

$$E_1 \triangleq \{(x_1=1,x_2=1,x_3=1,x_4=1,\ldots),(x_1=0,x_2=1,x_3=1,x_4=1,\ldots),\\ (x_1=0,x_2=0,x_3=1,x_4=1,\ldots)\}\,, \tag{12.1}$$

we have

$$\begin{aligned}\mathbb{T}E_1 &= \{(x_1=1,x_2=1,x_3=1,\ldots),(x_1=1,x_2=1,x_3=1\ldots),\\ &\qquad (x_1=0,x_2=1,x_3=1,\ldots)\}\\ &= \{(x_1=1,x_2=1,x_3=1,\ldots),(x_1=0,x_2=1,x_3=1,\ldots)\}\,.\end{aligned}$$

Thus, $\mathbb{T}E_1 \subseteq E_1$, which implies E_1 is $\mathbb{T}$-invariant.

Ergodicity based on Shift-invariant Event

Remarks.

- It can be proved that if $\mathbb{T}E \subseteq E$, then $\mathbb{T}^2E \subseteq \mathbb{T}E$. By induction, we can further obtain
$$\cdots \subseteq \mathbb{T}^3E \subseteq \mathbb{T}^2E \subseteq \mathbb{T}E \subseteq E.$$
- Thus, if an element say $(1,0,0,1,0,0,\ldots)$ is in a $\mathbb{T}$-invariant set E, then all its left-shift counterparts (i.e., $(0,0,1,0,0,1\ldots)$ and $(0,1,0,0,1,0,\ldots)$) should be contained in E.
- As a result, for a $\mathbb{T}$-invariant set E, an element and all its left-shift counterparts are either all in E or all outside E, but cannot be partially inside E.
- Hence, a "$\mathbb{T}$-invariant group" such as one containing
$$(1,0,0,1,0,0,\ldots), \quad (0,0,1,0,0,1\ldots) \quad \text{and} \quad (0,1,0,0,1,0,\ldots)$$
should be treated as an indecomposable group in $\mathbb{T}$-invariant sets.

Ergodicity based on Shift-invariant Event

- Although we are in particular interested in these "$\mathbb{T}$-invariant indecomposable groups," it is possible that some single "transient" element, such as $(0, 0, 1, 1, \ldots)$ in (12.1), is included in a $\mathbb{T}$-invariant set, and will be excluded after applying left-shift operation $\mathbb{T}$.

- This however can be resolved by introducing the "pseudo-inverse" operation $\mathbb{T}^{-1}$ (See page 3 of the below reference).

† P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*, American Mathematical Society, 1991.

- Note that $\mathbb{T}$ is a many-to-one mapping (See an example below), so its inverse operation in general does not exist!

Given $\mathcal{X} = \{0, 1\}$,

$$\mathbb{T}\{(0, 1, 0, 1, 0, 1 \ldots)\} = \mathbb{T}\{(1, 1, 0, 1, 0, 1 \ldots)\} = \{(1, 0, 1, 0, 1 \ldots)\}.$$

- The "pseudo-inverse" operation $\mathbb{T}^{-1}$ is defined as

$$\mathbb{T}^{-1}E \triangleq \{\boldsymbol{x} \in \mathcal{X}^\infty : \mathbb{T}\boldsymbol{x} \in E\}.$$

$$\mathbb{T}^{-1}\{(1, 0, 1, 0, 1 \ldots)\} = \{(0, 1, 0, 1, 0, 1 \ldots), (1, 1, 0, 1, 0, 1 \ldots)\}.$$

Ergodicity based on Shift-invariant Event

Definition (Ergodic set) A set is called the ergodic set if

$$\mathbb{T}^{-1}E = E.$$

- It can be shown that if

  $$\mathbb{T}^{-1}E = E,$$

  then

  $$\mathbb{T}E = \mathbb{T}(\mathbb{T}^{-1}E) = E,$$

  which in terms infer

  $$\cdots = \mathbb{T}^{-2}E = \mathbb{T}^{-1}E = E = \mathbb{T}E = \mathbb{T}^{2}E = \cdots.$$

  Thus, this definition excludes all "transient" elements from the ergodic set.

- It is named the *ergodic set* because as time goes by (the left-shift operator $\mathbb{T}$ can be regarded as a shift to a future time), the set always stays in the state that it has been before.

Ergodicity based on Shift-invariant Event

Ergodic Property (遍歷性)
名詞解釋: 遍歷性又稱各態遍歷性，是當力學體系從任一初態開始運動後，只要時間夠長，將要經過所有在能量曲面上的微觀運動狀態。因此無限長時間的時間平均等於範圍平均。(來源：國家教育研究院力學名詞辭典)

Oxford Dictionary - **Ergodic**. adj. Relating to or denoting systems or processes with the property that, given sufficient time, they include or impinge on all points in a given space and can be represented statistically by a reasonably large selection of points.

Example. Let $\mathtt{x}$ denote "don't-care," which can be either 1 or 0, and define

$$E_k = \{(\underbrace{\mathtt{x}, \mathtt{x}, \ldots, \mathtt{x}}_{k \text{ of them}}, 1, 0, 1, 0, 1, 0, \ldots), (\underbrace{\mathtt{x}, \mathtt{x}, \ldots, \mathtt{x}}_{k \text{ of them}}, 0, 1, 0, 1, 0, 1, \ldots)\}.$$

Then, $E = \cup_{k \geq 0} E_k$ is an ergodic set.

Ergodicity based on Shift-invariant Event

- For two-sided sequences, the two conditions below are equivalent:

$$\mathbb{T}^{-1}E = E \qquad \equiv \qquad \mathbb{T}E = E.$$

Definition (Ergodic process) A process is *ergodic* if any ergodic set has probability either 1 or 0.

Example. For an ergodic process, only one of the ergodic sets below can have probability one:

$$\left\{\boldsymbol{x} \in \{0,1\}^\infty : \lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = 0.0\right\}$$
$$\left\{\boldsymbol{x} \in \{0,1\}^\infty : \lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = 0.1\right\}$$
$$\left\{\boldsymbol{x} \in \{0,1\}^\infty : \lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = 0.2\right\}$$
$$\vdots$$
$$\left\{\boldsymbol{x} \in \{0,1\}^\infty : \lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = 0.9\right\}$$
$$\left\{\boldsymbol{x} \in \{0,1\}^\infty : \lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = 1.0\right\}$$

Ergodicity based on Shift-invariant Event

- The definition of ergodic processes has nothing to do with stationarity. It simply states that events that are unaffected by time-shifting (both left- and right-shifting) must have probability either zero or one.
- Ergodicity implies that all convergent sample averages converge to a constant (but not necessarily to the ensemble average), and stationarity assures that the time average converges to a random variable; hence, it is reasonably to expect that they jointly imply the ultimate time average equals the ensemble average (See the well-known *ergodic theorem* by Birkhoff and Khinchin.)

Theorem (Pointwise ergodic theorem) Consider a discrete-time stationary random process $\boldsymbol{X} = \{X_n\}_{n=1}^{\infty}$. For real-valued function $f(\cdot)$ on $\mathbb{R}$ with finite mean (i.e., $|E[f(X_n)]| < \infty$), there exists a random variable Y such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = Y \quad \text{with probability 1.}$$

If, in addition to stationarity, the process is also ergodic, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = E[f(X_1)] \quad \text{with probability 1.}$$

Ergodicity based on Shift-invariant Event

Remarks.

- In communications theory, one may assume that *the source is stationary* or *the source is stationary ergodic*. But it is rare to see the assumption of *the source being ergodic but non-stationary*.
- This is perhaps because an ergodic but non-stationary source not only does not facilitate the analytical study of communications problems, but may have limited application in practice.
- From this, we note that assumptions are made either to facilitate the analytical study of communications problems or to fit a specific need of applications. Without these two objectives, an assumption becomes of minor interest.
- This justifies that the *ergodicity* assumption usually comes after *stationarity* assumption. A specific example is the pointwise ergodic theorem, where the random processes considered is presumed to be stationary.

Closing Remarks on Ergodicity

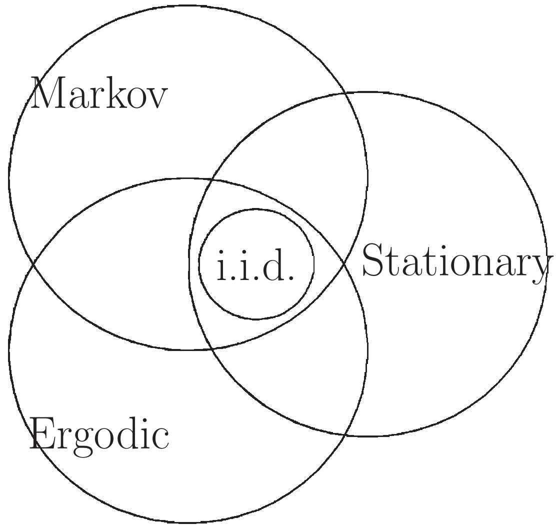
- The notion of ergodicity is often misinterpreted, since the definition is not very intuitive.
- It gets more confused as some engineering texts may provide a definition that a stationary process satisfying the ergodic theorem is also ergodic. Here is an example quoted from some text.

Definition. A stationary random process $\{X_n\}_{n=1}^{\infty}$ is called ergodic if for arbitrary integer k and function $f(\cdot)$ on $\mathcal{X}^k$ of finite mean,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{i+1}, \ldots, X_{i+k}) \xrightarrow{a.s.} E[f(X_1, \ldots, X_k)].$$

- This definition somehow implies that if a process is not stationary-ergodic, then the strong law of large numbers is violated (or the time average does not converge with probability 1 to its ensemble expectation). But this is not true (from its origin)!

Closing Remarks on Ergodicity



General relations of random processes.

- Indeed, the ergodic theorem is indeed a consequence of the original mathematical definition of ergodicity in terms of the shift-invariant property (See also pages 174-175 of the below reference).

† P. C. Shields, R. M. Gray and L. D. Davisson, *Random Processes: A Mathematical Approach for Engineers*, Prentice-Hall, 1986.

Closing Remarks on Ergodicity

Let us try to clarify the notion of ergodicity by the following remarks.

- The concept of ergodicity does not require stationarity. In other words, a non-stationary process can be ergodic.

- As mentioned earlier, stationarity and ergodicity imply the time average converges with probability 1 to the ensemble mean. Now if a process is stationary but not ergodic, then the time average still converges, but possibly not to the ensemble mean.

  Example. Let $\{A_n\}_{n=-\infty}^{\infty}$ and $\{B_n\}_{n=-\infty}^{\infty}$ be two i.i.d. binary 0-1 random variables with

$$\Pr\{A_n = 0\} = \Pr\{B_n = 1\} = \frac{1}{4}.$$

  Suppose that

$$X_n = \begin{cases} A_n, & \text{if } U = 1; \\ B_n, & \text{if } U = 0, \end{cases}$$

  where U is equiprobable binary random variable, and $\{A_n\}_{n=1}^{\infty}$, $\{B_n\}_{n=1}^{\infty}$ and U are independent. Then $\{X_n\}_{n=1}^{\infty}$ is stationary. Is the process ergodic? The answer is negative.

If the stationary process were ergodic, then from the pointwise ergodic theorem, its relative frequency would converge to

$$\begin{aligned}
\Pr(X_n = 1) &= \Pr(U = 1)\Pr(X_n = 1|U = 1) + \Pr(U = 0)\Pr(X_n = 1|U = 0) \\
&= \Pr(U = 1)\Pr(A_n = 1) + \Pr(U = 0)\Pr(B_n = 1) \\
&= \frac{1}{2}.
\end{aligned}$$

However, one should observe that the outputs of $(X_1, \ldots, X_n)$ form a Bernoulli process with relative frequency of 1's being either 3/4 or 1/4, depending on the value of U. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_n \xrightarrow{a.s.} Y,$$

where $\Pr(Y = 1/4) = \Pr(Y = 3/4) = 1/2$, which contradicts to the ergodic theorem. $\square$

Closing Remarks on Ergodicity

- From the above example, the pointwise ergodic theorem can actually be made useful in such a stationary but non-ergodic case, since the estimate with stationary ergodic process (either $\{A_n\}_{n=-\infty}^{\infty}$ or $\{B_n\}_{n=-\infty}^{\infty}$) is actually being observed by measuring the relative frequency (3/4 or 1/4).

- This renders a surprising fundamental result of random processes— *ergodic decomposition theorem:* Any stationary process is in fact a mixture of stationary ergodic processes, and hence one always observes a stationary ergodic outcome.

Example. One always observe either $A_1, A_2, A_3, \ldots$ or $B_1, B_2, B_3, \ldots$, depending on the value of U, for which both sequences are stationary ergodic. In other words, the time-stationary observation X_n satisfies

$$X_n = U \cdot A_n + (1 - U) \cdot B_n.$$

$\square$

Closing Remarks on Ergodicity

- In conclusion, ergodicity is not required for the strong law of large numbers to be effective. The next question is whether or not stationarity is required. Again the answer is negative.

- The main concern of the law of large numbers is the convergence of sample averages to its ensemble expectation.

- It should be reasonable to expect that random processes could exhibit transient behavior that violates the stationarity definition, yet the sample average still converges.

  Example. A finite-alphabet time-invariant (but not necessarily stationary) irreducible Markov chain satisfies the law of large numbers.

- Accordingly, one should not take the notions of stationarity and ergodicity too seriously (if the main concern is the law of large numbers) since they can be significantly weakened and still have laws of large numbers holding (i.e., time averages and relative frequencies have desired and well-defined limits).

The end of Preliminary Introduction of Ergodicity

12-1 Ergodicity

Concern:

- Can one estimate the mean $E[\boldsymbol{x}(t)]$ of a WSS process $\boldsymbol{x}(t)$ in terms of time samples $\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n), \cdots$?

Suppose that $\boldsymbol{x}(t)$ is defined over the probability system $(S, \mathcal{F}, P)$. Then,

$$E[\boldsymbol{x}(t)] = \int_S \boldsymbol{x}(t, \zeta) dP(\zeta) \quad \text{or} \quad \sum_{\zeta \in S} \boldsymbol{x}(t, \zeta) P(\zeta)$$

can be estimated by $\boldsymbol{x}(t, \zeta_1), \boldsymbol{x}(t, \zeta_2), \boldsymbol{x}(t, \zeta_3), \cdots$! This indicates that one needs to impractically take a large number of samples at the same time in order to estimate $E[\boldsymbol{x}(t)]$.

Definition (Mean-ergodic process) A process $\boldsymbol{x}(t)$ is *mean-ergodic* if for some constant η,

$$E\left[\,|\boldsymbol{\eta}_T - \eta|^2\,\right] \to 0 \text{ as } T \to \infty,$$

where

$$\boldsymbol{\eta}_T \triangleq \frac{1}{2T} \int_{-T}^{T} \boldsymbol{x}(t) dt.$$

$\boldsymbol{\eta}_T$ can then be further approximated by its Riemann integral, using $\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n), \cdots$.

Example 12-1 Is $\boldsymbol{x}(t) = \boldsymbol{c}$ mean-ergodic, where $\boldsymbol{c}$ is a nondegenerate random variable?

Answer:

$$\boldsymbol{\eta}_T = \frac{1}{2T}\int_{-T}^{T} \boldsymbol{x}(t)dt \; = \; \boldsymbol{c}.$$

Since

$$\lim_{T\to\infty} E[|\boldsymbol{\eta}_T - \eta|^2] = E[|\boldsymbol{c} - \eta|^2] \neq 0,$$

$\boldsymbol{x}(t)$ is not mean-ergodic. $\square$

Ergodicity

This process is stationary since

$$\boldsymbol{x}(t) = \boldsymbol{c} \quad \text{and} \quad \boldsymbol{x}(t-\tau) = \boldsymbol{c}$$

have the same statistics for any τ. However,

$$\Pr\left(\frac{\boldsymbol{x}(t_1) + \boldsymbol{x}(t_2) + \cdots + \boldsymbol{x}(t_n)}{n} \le c\right) = \Pr(\boldsymbol{c} \le c).$$

So

$$\frac{\boldsymbol{x}(t_1) + \boldsymbol{x}(t_2) + \cdots + \boldsymbol{x}(t_n)}{n}$$

converges in distribution (also in probability and with probability 1) to a random variable $\boldsymbol{c}$. This is not what mean-ergodicity wishes to see. The mean-ergodicity demands that

$$\frac{\boldsymbol{x}(t_1) + \boldsymbol{x}(t_2) + \cdots + \boldsymbol{x}(t_n)}{n}$$

converges in mean-sqaure to a constant c.

Variance of $\boldsymbol{\eta}_T$

- $\text{Var}[\boldsymbol{\eta}_T] = E[\boldsymbol{\eta}_T^2] - E^2[\boldsymbol{\eta}_T] = \dfrac{1}{4T^2}\displaystyle\int_{-T}^{T}\int_{-T}^{T}[R_{xx}(t,s) - \eta_x(t)\eta_x(s)]dtds$
- If $\boldsymbol{x}(t)$ is WSS, then

$$\begin{aligned}
\text{Var}[\boldsymbol{\eta}_T] &= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} R_{xx}(t-s)dtds - \eta_x^2 \\
&= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T-s}^{T-s} R_{xx}(u)duds - \eta_x^2 \\
&= \frac{1}{4T^2}\left(\int_{-2T}^{0} R_{xx}(u)\int_{-T-u}^{T} dsdu + \int_{0}^{2T} R_{xx}(u)\int_{-T}^{T-u} dsdu\right) - \eta_x^2 \\
&= \frac{1}{2T}\int_{-2T}^{2T} R_{xx}(u)\left(1 - \frac{|u|}{2T}\right)du - \eta_x^2 \\
&= \frac{1}{2T}\int_{-2T}^{2T} C_{xx}(u)\left(1 - \frac{|u|}{2T}\right)du.
\end{aligned}$$

Lemma A WSS process $\boldsymbol{x}(t)$ is mean-ergodic if, and only if, $\lim_{T\to\infty} \text{Var}[\boldsymbol{\eta}_T] = 0$.

Slutsky's Theorem

Theorem 12-1 (Slutsky's theorem) A WSS process $\boldsymbol{x}(t)$ is mean-ergodic if, and only if,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}C_{xx}(\tau)d\tau = 0.$$

Proof:

- By Cauchy-Schwartz inequality,

$$\text{Cov}[\boldsymbol{\eta}_T, \boldsymbol{x}(0)] \;\leq\; \sqrt{\text{Var}[\boldsymbol{\eta}_T]\text{Var}[\boldsymbol{x}(0)]}.$$

  Since $\text{Cov}[\boldsymbol{\eta}_T, \boldsymbol{x}(0)] = \frac{1}{2T}\int_{-T}^{T}C_{xx}(\tau)d\tau$, it is apparent that if $\text{Var}[\boldsymbol{\eta}_T] \to 0$,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}C_{xx}(\tau)d\tau = 0.$$

- On the contrary,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}C_{xx}(\tau)d\tau = 0$$

  implies the existence of T_0 for a given $\epsilon > 0$ such that

$$\left|\frac{1}{2T}\int_{-T}^{T}C_{xx}(\tau)d\tau\right| < \epsilon \text{ for every } T > T_0.$$

Slutsky's Theorem

This indicates that

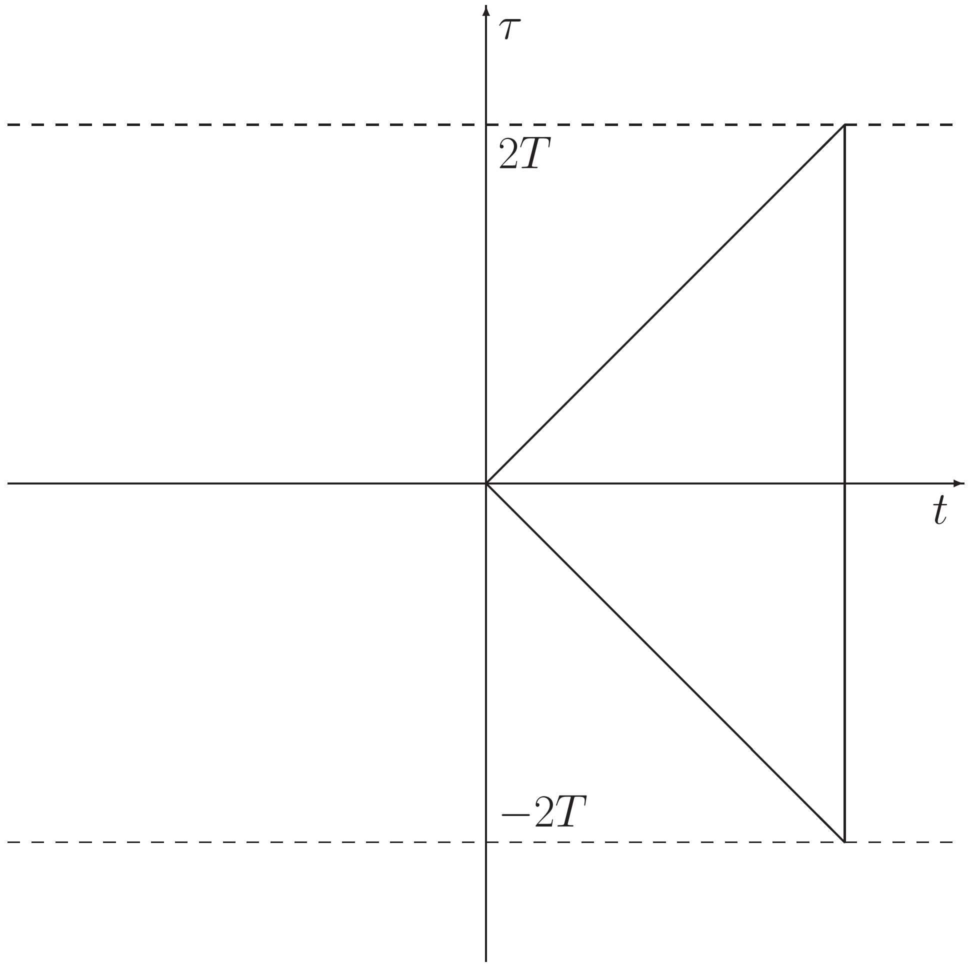
$$\left|\int_{-t}^{t} C_{xx}(\tau)d\tau\right| < 2t\epsilon \quad \text{for every } t > T_0.$$

Accordingly, for $2T > T_0$,

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{\eta}_T] &= \frac{1}{2T}\int_{-2T}^{2T} C_{xx}(\tau)\left(1-\frac{|\tau|}{2T}\right)d\tau \\
&= \frac{1}{4T^2}\int_{-2T}^{2T} C_{xx}(\tau)\,(2T-|\tau|)\,d\tau \\
&= \frac{1}{4T^2}\int_{-2T}^{2T} C_{xx}(\tau)\left(\int_{|\tau|}^{2T} dt\right)d\tau \quad \text{(See Slide 12-23.)} \\
&= \frac{1}{4T^2}\int_{0}^{2T}\left(\int_{-t}^{t} C_{xx}(\tau)d\tau\right)dt \\
&= \frac{1}{4T^2}\int_{0}^{T_0}\left(\int_{-t}^{t} C_{xx}(\tau)d\tau\right)dt + \frac{1}{4T^2}\int_{T_0}^{2T}\left(\int_{-t}^{t} C_{xx}(\tau)d\tau\right)dt
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{4T^2}\int_0^{T_0}\left(\int_{-t}^{t}|C_{xx}(\tau)|d\tau\right)dt + \frac{1}{4T^2}\int_{T_0}^{2T}\left|\int_{-t}^{t}C_{xx}(\tau)d\tau\right|dt\\
&\le \frac{1}{4T^2}\int_0^{T_0}\left(\int_{-t}^{t}C_{xx}(0)d\tau\right)dt + \frac{1}{4T^2}\int_{T_0}^{2T}(2t\epsilon)dt\\
&= \frac{T_0^2}{4T^2}C_{xx}(0) + \frac{4T^2 - T_0^2}{4T^2}\epsilon,
\end{aligned}
$$

which implies $\lim_{T\to\infty}\mathrm{Var}[\boldsymbol{\eta}_T] \le \epsilon$. The proof is completed by noting that ϵ can be made arbitrarily small. $\square$


τ
$2T$
t
$-2T$

Lemma (Sufficient conditions for mean-ergodicity) If

$$\left|\int_{-\infty}^{\infty} C_{xx}(\tau)d\tau\right| < \infty \quad \text{or} \quad \lim_{|\tau|\to\infty} |C_{xx}(\tau)| = 0,$$

then WSS $\boldsymbol{x}(t)$ is mean-ergodic.

Proof: Since $\left|\int_{-\infty}^{\infty} C_{xx}(\tau)d\tau\right| < \infty$ implies $\lim_{|\tau|\to\infty} |C_{xx}(\tau)| = 0$, it suffices to prove the sufficiency of the latter condition.

For any $\epsilon > 0$, there exists T_0 such that $|C_{xx}(\tau)| < \epsilon$ for $|\tau| > T_0$. Hence,

$$\begin{aligned}
\left|\frac{1}{2T}\int_{-T}^{T} C_{xx}(\tau)d\tau\right| &\le \frac{1}{2T}\int_{-T_0}^{T_0} |C_{xx}(\tau)|d\tau + \frac{1}{2T}\int_{T_0\le|\tau|<T} |C_{xx}(\tau)|d\tau \\
&\le \frac{1}{2T}\int_{-T_0}^{T_0} C_{xx}(0)d\tau + \frac{1}{2T}\int_{T_0\le|\tau|<T} \epsilon d\tau \\
&= C_{xx}(0)\frac{T_0}{T} + \epsilon\frac{T-T_0}{T} \to \epsilon \text{ as } T\to\infty.
\end{aligned}$$

The proof is completed by noting that ϵ can be made arbitrarily small. $\square$

Ergodicity for Discrete Processes

Concern:

- Can one estimate the mean $E\{\boldsymbol{x}[t]\}$ of a discrete WSS process $\boldsymbol{x}[t]$ in terms of time samples $\boldsymbol{x}[t_1], \boldsymbol{x}[t_2], \ldots, \boldsymbol{x}[t_n], \cdots$, where t_1, t_2, …, t_n, … are integers?

Definition (Mean-ergodic process) A discrete process $\boldsymbol{x}[t]$ is *mean-ergodic* if for some constant η,

$$E\left[\,|\boldsymbol{\eta}_T - \eta|^2\,\right] \to 0 \text{ as } T \to \infty,$$

where

$$\boldsymbol{\eta}_T \triangleq \frac{1}{2T+1} \sum_{t=-T}^{T} \boldsymbol{x}[t].$$

- We can then similarly show that for a discrete WSS $\boldsymbol{x}[t]$,

$$\mathrm{Var}[\boldsymbol{\eta}_T] \;=\; \frac{1}{2T+1} \sum_{u=-2T}^{2T} C_{xx}[u] \left(1 - \frac{|u|}{2T+1}\right).$$

Lemma A WSS process $\boldsymbol{x}[t]$ is mean-ergodic if, and only if, $\lim_{T\to\infty} \mathrm{Var}[\boldsymbol{\eta}_T] = 0$.

Slutsky's Theorem for Discrete Processes

Theorem 12-2 (Slutsky's theorem for discrete processes) A discrete WSS process $\boldsymbol{x}[t]$ is mean-ergodic if, and only if,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{\tau=-T}^{T} C_{xx}[\tau] = 0.$$

Lemma (Sufficient conditions for mean-ergodicity) If

$$\lim_{|\tau|\to\infty} |C_{xx}[\tau]| = 0,$$

then WSS $\boldsymbol{x}[t]$ is mean ergodic.

Example 12-6(a) Suppose $\boldsymbol{x}[t]$ is zero-mean white with autocovariance function $C_{xx}[m] = P\delta[m]$. Then, as $\lim_{m\to\infty} C_{xx}[m] = 0$, $\boldsymbol{x}[t]$ is mean-ergodic, and the estimation variance equals:

$$\begin{aligned}
\text{Var}[\boldsymbol{\eta}_T] &= \frac{1}{2T+1} \sum_{u=-2T}^{2T} C_{xx}[u] \left(1 - \frac{|u|}{2T+1}\right) \\
&= \frac{1}{2T+1} \sum_{u=-2T}^{2T} P\delta[u] \left(1 - \frac{|u|}{2T+1}\right) = \frac{P}{2T+1}.
\end{aligned}$$

Example 12-6(b) In case $C_{xx}[m] = Pa^{|m|}$ for $|a| < 1$, $\boldsymbol{x}[t]$ is still mean-ergodic because $\lim_{m\to\infty} C_{xx}[m] = 0$. Also,

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{\eta}_T] &= \frac{1}{2T+1}\sum_{u=-2T}^{2T} C_{xx}[u]\left(1-\frac{|u|}{2T+1}\right)\\
&= \frac{1}{2T+1}\sum_{u=-2T}^{2T} Pa^{|u|}\left(1-\frac{|u|}{2T+1}\right)\\
&= \frac{P\left[(1-a^2)(2T+1)-2a+2a^{2T+2}\right]}{(1-a)^2(2T+1)^2}\\
&\approx \frac{P}{(2T+1)}\frac{(1+a)}{(1-a)} \text{ when } T \text{ large.}
\end{aligned}$$

Spectral Interpretation of Ergodicity

Examination of Mean-ergodicity in terms of Spectral

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{\eta}_T] &= \frac{1}{2T}\int_{-2T}^{2T} C_{xx}(u)\left(1-\frac{|u|}{2T}\right)du \\
&= \int_{-\infty}^{\infty} C_{xx}(u)\cdot\frac{1}{2T}\left(1-\frac{|u|}{2T}\right)\mathbf{1}\{|u|\le 2T\}du \\
&= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}^{c}(\omega)e^{j\omega u}d\omega\right)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin^2(T\nu)}{T^2\nu^2}e^{j\nu u}d\nu\right)du \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_{xx}^{c}(\omega)\frac{\sin^2(T\nu)}{T^2\nu^2}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j(\omega-\nu)u}du\right)d\nu d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_{xx}^{c}(\omega)\frac{\sin^2(T\nu)}{T^2\nu^2}\delta(\omega-\nu)d\nu d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}^{c}(\omega)\frac{\sin^2(T\omega)}{T^2\omega^2}d\omega,
\end{aligned}$$

where $S_{xx}^{c}(\omega)$ is the covariance spectrum of WSS $\boldsymbol{x}(t)$.

Spectral Interpretation of Ergodicity

Example. If $S^c_{xx}(\omega)$ consists of an impulse at the origin, then $\boldsymbol{x}(t)$ is not mean-ergodic because

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{\eta}_T] &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S^c_{xx}(\omega)\frac{\sin^2(T\omega)}{T^2\omega^2}d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \left[S_1(\omega) + 2\pi k_0\delta(\omega)\right]\frac{\sin^2(T\omega)}{T^2\omega^2}d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_1(\omega)\frac{\sin^2(T\omega)}{T^2\omega^2}d\omega + k_0 \not\to 0 \text{ as } T \to 0.
\end{aligned}$$

If process $\boldsymbol{x}(t)$ is regular, then $S^c_{xx}(\omega)$ contains no impulse at $\omega = 0$.

Further Generalization of $\boldsymbol{\eta}_T$

- $\boldsymbol{\eta}_T \triangleq \dfrac{1}{2T}\displaystyle\int_{-T}^{T} \boldsymbol{x}(t)dt$ is a special instance of $\boldsymbol{y}(t;T_0) \triangleq \dfrac{1}{T_0}\displaystyle\int_{t-T_0}^{t} \boldsymbol{x}(\alpha)d\alpha$ as

$$\boldsymbol{\eta}_T = \boldsymbol{y}(T;2T).$$

- If $\boldsymbol{x}(t)$ is WSS,

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{y}(t;T_0)] &= \frac{1}{T_0^2}\int_{t-T_0}^{t}\int_{t-T_0}^{t} C_{xx}(u-v)dudv \\
&= \frac{1}{T_0^2}\int_{t-T_0}^{t}\int_{t-v-T_0}^{t-v} C_{xx}(s)dsdv \\
&= \frac{1}{T_0^2}\int_{-T_0}^{0} C_{xx}(s)\int_{t-s-T_0}^{t} dvds + \frac{1}{T_0^2}\int_{0}^{T_0} C_{xx}(s)\int_{t-T_0}^{t-s} dvds \\
&= \frac{1}{T_0}\int_{-T_0}^{T_0} C_{xx}(s)\left(1-\frac{|s|}{T_0}\right)ds,
\end{aligned}$$

which is only a function of window size T_0.

Variance-Ergodicity

Concern:

- Can one estimate the variance of a WSS process $\boldsymbol{x}(t)$ in terms of time samples $\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n), \cdots$?

Concern with known mean and SSS:

- Can one estimate the variance $E[\boldsymbol{x}^2(t)] - \mu_x^2$ of a **SSS** process $\boldsymbol{x}(t)$ in terms of time samples $\boldsymbol{x}^2(t_1), \boldsymbol{x}^2(t_2), \ldots, \boldsymbol{x}^2(t_n), \cdots$?
- The answer has already been given in previous slides.

Example. For a zero-mean SSS Gaussian process $\boldsymbol{x}(t)$,

$$\begin{aligned}
C_{x^2x^2}(\tau) &= E[\boldsymbol{x}^2(t+\tau)\boldsymbol{x}^2(t)] - E[\boldsymbol{x}^2(t+\tau)]E[\boldsymbol{x}^2(t)] \\
&= \underbrace{E[\boldsymbol{x}^2(t+\tau)]E[\boldsymbol{x}^2(t)] + 2E^2[\boldsymbol{x}(t+\tau)\boldsymbol{x}(t)]}_{\text{See the next slide.}} - E[\boldsymbol{x}^2(t+\tau)]E[\boldsymbol{x}^2(t)] \\
&= 2C_{xx}^2(\tau).
\end{aligned}$$

Thus, zero-mean SSS Gaussian $\boldsymbol{x}(t)$ is variance-ergodic if, and only if,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} C_{xx}^2(\tau) d\tau = 0.$$

Variance-Ergodicity

For joint normal variables $\boldsymbol{x}$ and $\boldsymbol{y}$ with zero-mean, the pdf equals:

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}\begin{bmatrix} x & y\end{bmatrix}\Sigma^{-1}\begin{bmatrix} x \\ y\end{bmatrix}\right\} \\ &= \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{\sigma_y^2x^2-2\sigma_{xy}xy+\sigma_x^2y^2}{2(\sigma_x^2\sigma_y^2-\sigma_{xy}^2)}\right\}\end{aligned}$$

where $\Sigma = \begin{bmatrix}\sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2\end{bmatrix}$. This gives that $(x|y)$ is Gaussian distributed with mean $\sigma_{xy}y/\sigma_y^2$ and variance $(\sigma_x^2\sigma_y^2-\sigma_{xy}^2)/\sigma_y^2$, which implies that

$$\begin{aligned} E[\boldsymbol{x}^2\boldsymbol{y}^2] &= E[E[\boldsymbol{x}^2\boldsymbol{y}^2|\boldsymbol{y}]] = E[\boldsymbol{y}^2E[\boldsymbol{x}^2|\boldsymbol{y}]] \\ &= \frac{\sigma_x^2\sigma_y^2-\sigma_{xy}^2}{\sigma_y^2}E[\boldsymbol{y}^2]+\frac{\sigma_{xy}^2}{\sigma_y^4}E[\boldsymbol{y}^4] \\ &= \frac{\sigma_x^2\sigma_y^2-\sigma_{xy}^2}{\sigma_y^2}\sigma_y^2+\frac{\sigma_{xy}^2}{\sigma_y^4}(3\sigma_y^4) \\ &= \sigma_x^2\sigma_y^2+2\sigma_{xy}^2. \qquad \text{(See (6-199) in text.)}\end{aligned}$$

Variance-Ergodicity

Concern with unknown mean and SSS:

- Can one estimate the variance of a (real) **SSS** process $\boldsymbol{x}(t)$ in terms of time samples $\boldsymbol{x}^2(t_1), \boldsymbol{x}^2(t_2), \ldots, \boldsymbol{x}^2(t_n), \cdots$?
  - First, estimate $\boldsymbol{\eta}_T \triangleq \dfrac{1}{2T}\displaystyle\int_{-T}^{T} \boldsymbol{x}(t)dt$.
  - Then, estimate the variance of $\boldsymbol{x}(t)$ by

$$\hat{\boldsymbol{V}}_T = \frac{1}{2T}\int_{-T}^{T}[\boldsymbol{x}(t) - \boldsymbol{\eta}_T]^2 dt = \frac{1}{2T}\int_{-T}^{T}\boldsymbol{x}^2(t)dt - \boldsymbol{\eta}_T^2.$$

Remarks

- $\hat{\boldsymbol{V}}_T$ is a biased estimator, as contrary to

$$\boldsymbol{V}_T \triangleq \frac{1}{2T}\int_{-T}^{T}\boldsymbol{x}^2(t)dt - \eta_x^2 \quad \text{(estimator with known mean)}$$

  is an unbiased estimator.

- The estimation variance of $\hat{\boldsymbol{V}}_T$, however, is smaller than that of $\boldsymbol{V}_T$ in many cases, which is the reason that some use $\hat{\boldsymbol{V}}_T$ instead of $\boldsymbol{V}_T$ even when η_x is known.

Covariance-Ergodicity

Concern:

- Can one estimate the autocovariance function of a (real) zero-mean SSS process $\boldsymbol{x}(t)$ in terms of time samples $\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n), \cdots$?

Equivalent concern:

- Can one estimate the mean of a SSS process $\boldsymbol{x}(t+\tau)\boldsymbol{x}(t)$ in terms of time samples $\boldsymbol{x}(t_1+\tau)\boldsymbol{x}(t_1), \boldsymbol{x}(t_2+\tau)\boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n+\tau)\boldsymbol{x}(t_n), \cdots$?
- Answer: $\boldsymbol{C}_T(\tau) = \dfrac{1}{2T}\displaystyle\int_{-T}^{T} \boldsymbol{x}(t+\tau)\boldsymbol{x}(t)dt$

Example. If $\boldsymbol{x}(t)$ is a Gaussian zero-mean SSS process, then by letting $\boldsymbol{z}(t) = \boldsymbol{x}(t+\lambda)\boldsymbol{x}(t)$,

$$\begin{aligned} C_{zz}(\tau) &= E[\boldsymbol{x}(t+\tau+\lambda)\boldsymbol{x}(t+\tau)\boldsymbol{x}(t+\lambda)\boldsymbol{x}(t)] - E[\boldsymbol{x}(t+\tau+\lambda)\boldsymbol{x}(t+\tau)]E[\boldsymbol{x}(t+\lambda)\boldsymbol{x}(t)] \\ &= C_{xx}(\lambda+\tau)C_{xx}(\lambda-\tau) + C_{xx}^2(\tau). \end{aligned}$$

Thus, by Theorem 12-1 (cf. Slide 12-20), SSS Gaussian zero-mean $\boldsymbol{x}(t)$ is covariance-ergodic if, and only if,

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} [C_{xx}(\lambda+\tau)C_{xx}(\lambda-\tau) + C_{xx}^2(\tau)]d\tau = 0.$$

In addition, if $C_{xx}(\tau) \to 0$ as $|\tau| \to \infty$, then $\boldsymbol{x}(t)$ is mean-ergodic, variance-ergodic and covariance-ergodic.

Cross-Covariance-Ergodicity

Concern:

- Can one estimate the cross-covariance function of two zero-mean jointly SSS processes $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ in terms of their samples?
- Since the answer is very similar to other cases, the discussion about cross-covariance-ergodicity is omitted.

Bussgang's Theorem Revisited

Theorem (Bussgang's theorem) The cross-covariance $C_{xy}(\tau)$ of system input $\boldsymbol{x}(t)$ and system output $\boldsymbol{y}(t)$ for a stationary zero-mean Gaussian input and memoryless system $\boldsymbol{T}(\cdot)$ is proportional to $C_{xx}(\tau)$. I.e.,

$$C_{xy}(t_1 - t_2) \;=\; C_{xx}(t_1 - t_2)\frac{E[\boldsymbol{x}(t) \cdot g(\boldsymbol{x}(t))]}{R_{xx}(0)},$$

where $g(x_2) = E[\boldsymbol{T}(x_2)]$.

- By Bussgangs' theorem, to estimate $C_{xx}(\tau)$, it suffices to estimate $C_{xy}(\tau)$ for some properly chosen system $\boldsymbol{T}(\cdot)$.

  Example. Choose $\boldsymbol{y}(t) = \text{sgn}[\boldsymbol{x}(t)]$ (cf. Hard Limiter) for zero-mean Gaussian SSS $\boldsymbol{x}(t)$. Then,

$$C_{xx}(\tau) = \sqrt{\frac{\pi C_{xx}(0)}{2}}C_{xy}(\tau) = \frac{\pi}{2}C_{xy}(0)C_{xy}(\tau)$$

  and $C_{xy}(\tau)$ can be estimated by

$$\frac{1}{2T}\int_{-T}^{T} \boldsymbol{x}(t+\tau)\text{sgn}[\boldsymbol{x}(t)]dt.$$

Distribution-Ergodic Processes

Concern:

- Can one estimate the cdf $F_{\boldsymbol{x}}(x) \triangleq \Pr[\boldsymbol{x}(t) \leq x]$ of a SSS process $\boldsymbol{x}(t)$ in terms of time samples $\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n), \cdots$?

Equivalent concern:

- Can one estimate the mean of a SSS process $\boldsymbol{y}(t) \triangleq \mathbf{1}\{\boldsymbol{x}(t) \leq x\}$ in terms of time samples?
- This is exactly the mean-ergodicity.
- Since

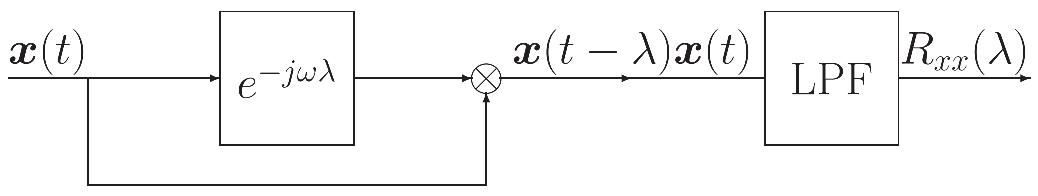
$$\begin{aligned} C_{yy}(\tau) &= E[\boldsymbol{y}(t+\tau)\boldsymbol{y}^*(t)] - E[\boldsymbol{y}(t+\tau)]E[\boldsymbol{y}^*(t)] \\ &= \Pr\{\boldsymbol{x}(t+\tau) \leq x \text{ and } \boldsymbol{x}(t) \leq x\} - \Pr\{\boldsymbol{x}(t+\tau) \leq x\} \Pr\{\boldsymbol{x}(t) \leq x\} \\ &= F_{\boldsymbol{x}}(x, x; 0, \tau) - F_{\boldsymbol{x}}^2(x), \end{aligned}$$

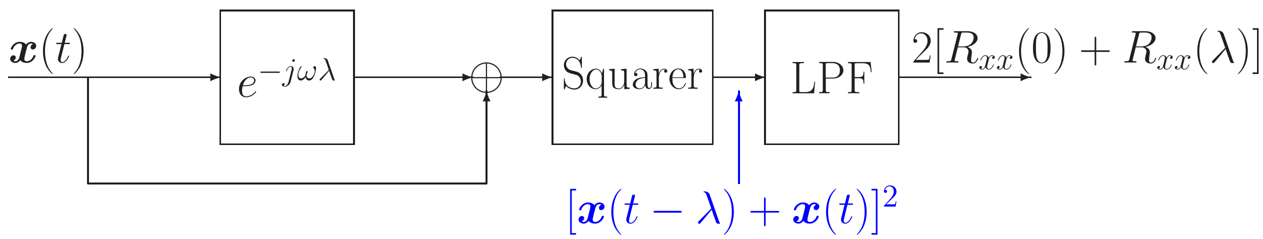
a SSS process $\boldsymbol{x}(t)$ is distribution-ergodic if

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} F_{\boldsymbol{x}}(x, x; 0, \tau) d\tau = F_{\boldsymbol{x}}^2(x).$$

Implementation of Ergodicity Estimation

- A **corelometer** is a physical device measuring the autocorrelation $R_{xx}(\lambda)$ of a (WSS) process $\boldsymbol{x}(t)$.
- Below are two possible structures for correlometer, where one uses a multiplier and the other uses an adder with squarer. The LPF can be treated as an integrator.

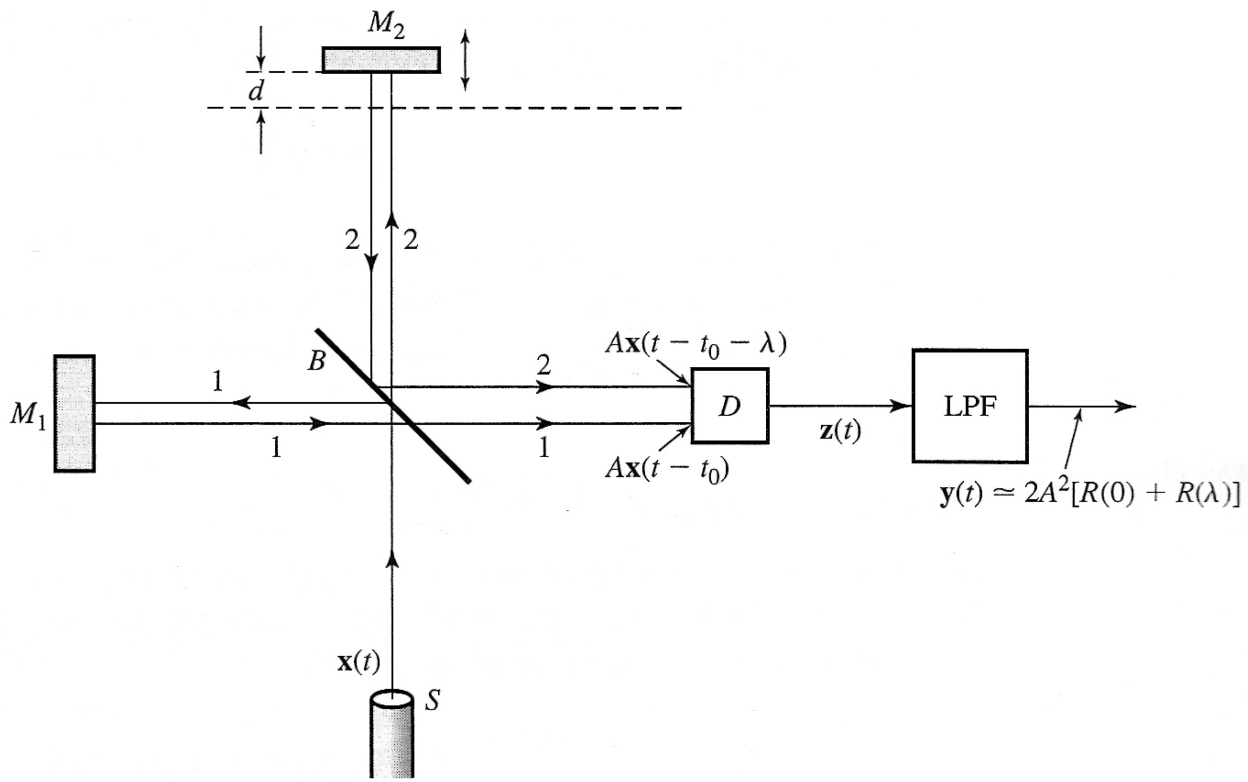




Implementation of Ergodicity Estimation

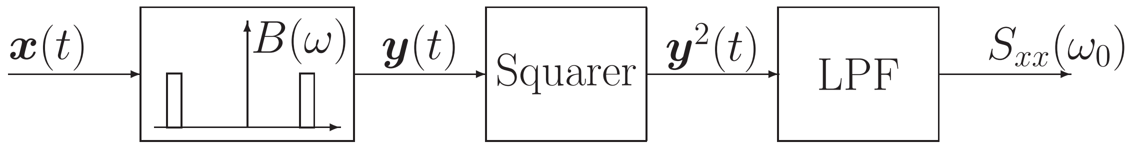
Michelson interferometer (Corelometer)

- $\lambda = 2d/c$ and $t_0 = l/c$, where c is the light speed.
- D is a squarer; hence, $\boldsymbol{z}(t) = A^2[\boldsymbol{x}(t - t_0 - \lambda) + \boldsymbol{x}(t - t_0)]^2$.



Implementation of Ergodicity Estimation

- A **spectrometer** is a physical device measuring the Fourier transform $S_{xx}(\omega)$ of $R_{xx}(\lambda)$.
- $E[\boldsymbol{y}^2(t)] = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{\infty} S_{yy}(\omega)d\omega = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{\infty} S_{xx}(\omega)|B(\omega)|^2 d\omega \approx S_{xx}(\omega_0)$



Implementation of Ergodicity Estimation

Fabry-Pérot interferometer (Spectrometer)

- $B(\omega) = \dfrac{1}{1 - r^2 e^{-j2\omega d/c}} = \sum_{n=0}^{\infty} r^{2n} e^{-j2n\omega d/c}$, where $r \approx 1$ is the reflection coefficient of the two places P_1 and P_2, and c is the light speed in the medium between the plate (See (a) in the next slide).

- Notably,

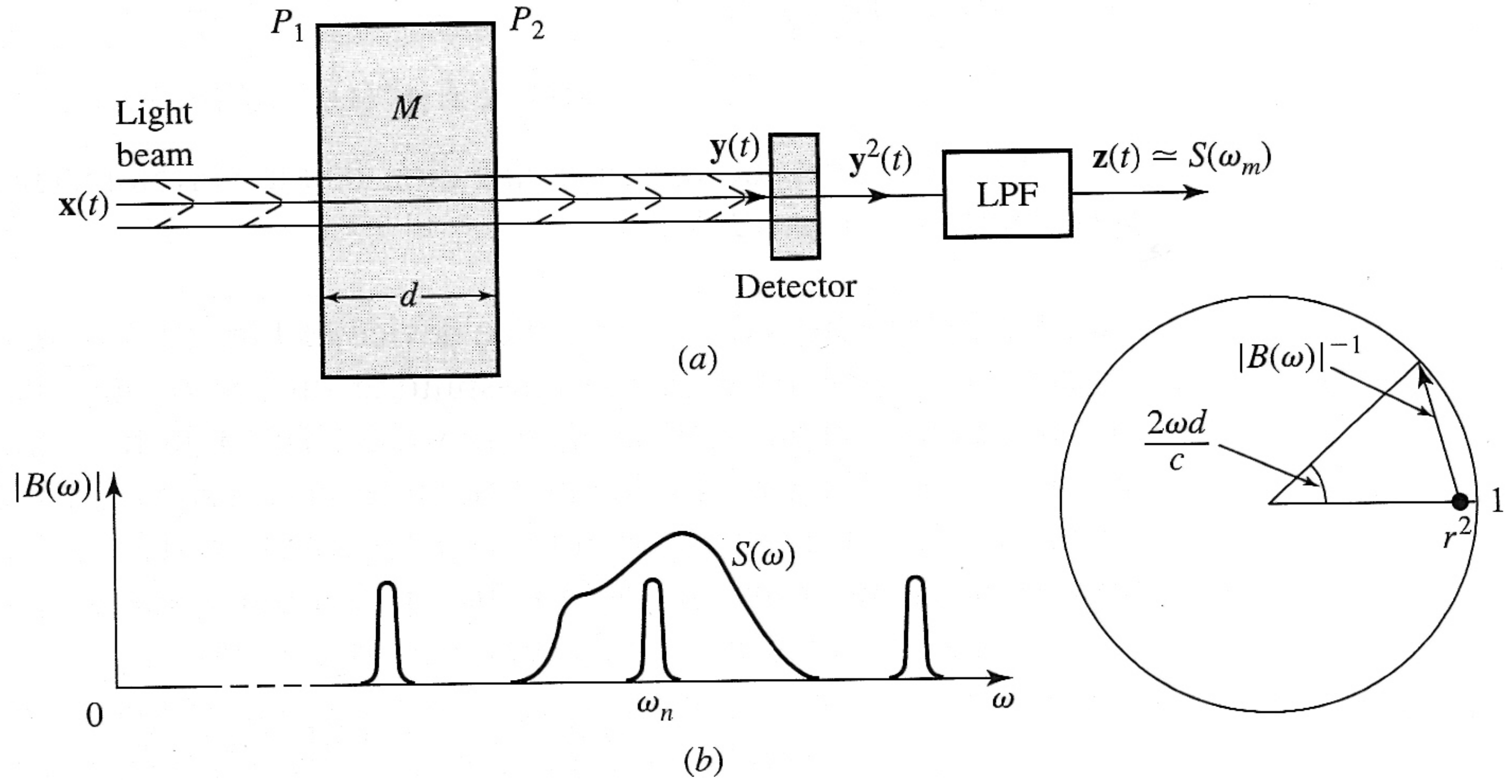
$$|B(\omega)|^2 = \frac{1}{1 + r^4 - 2r^2 \cos(2\omega d/c)}$$

  and

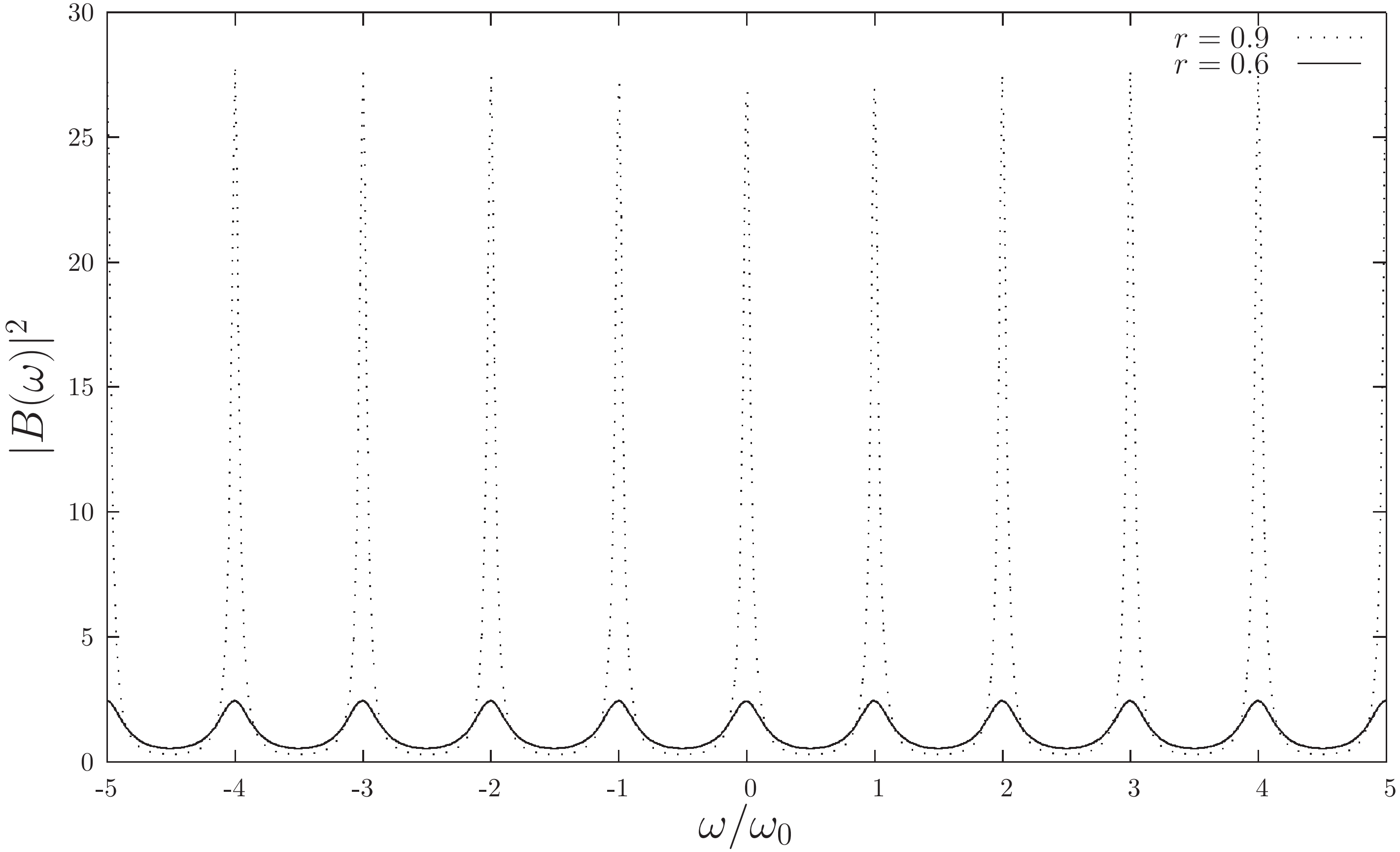
$$B(\omega) \approx \frac{1}{2} + \frac{\omega_0}{2} \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) - j\frac{1}{2} \cot\left(\pi \frac{\omega}{\omega_0}\right) \text{ when } r \to 1, \text{ where } \omega_0 = \pi\frac{c}{d}.$$

  So, if $S_{xx}(\omega)$ only overlaps with one impulse in $B(\omega)$, then $B(\omega)$ functions like the ideal single-impulse filter (See (b) in the next slide).

Implementation of Ergodicity Estimation



Implementation of Ergodicity Estimation



The end of Section 12-1 Ergodicity

12-2 Spectrum Estimation

Concern:

- Can one estimate the power spectrum $S_{xx}(\omega)$ of a real WSS process $\boldsymbol{x}(t)$ in terms of $\boldsymbol{x}_T(t)$, where

$$\boldsymbol{x}_T(t) \triangleq \boldsymbol{x}(t)p_T(t) \qquad p_T(t) \triangleq \begin{cases} 1, & |t| < T \\ 0, & |t| > T \end{cases}$$

Definition (Periodogram) The periodogram of a process is defined as

$$\boldsymbol{S}_T(\omega) = \frac{1}{2T}\left|\int_{-T}^{T} \boldsymbol{x}(t)e^{-j\omega t}dt\right|^2.$$

- The periodogram is the normalized absolute square of the Fourier transform of the known segment $\boldsymbol{x}_T(t)$, i.e.,

$$\boldsymbol{S}_T(\omega) = \frac{1}{2T}|\boldsymbol{X}_T(\omega)|^2 \qquad \boldsymbol{X}_T(\omega) = \int_{-\infty}^{\infty} \boldsymbol{x}_T(t)e^{-j\omega t}dt.$$

Spectral Bias

Theorem 12-4 Let $\boldsymbol{y}(t) = \boldsymbol{x}_T(t) = \boldsymbol{x}(t)p_T(t)$. For WSS $\boldsymbol{x}(t)$, $\boldsymbol{S}_T(\omega) = (1/(2T))|\boldsymbol{Y}(\omega)|^2$ is an asymptotically unbiased estimator of $S_{xx}(\omega)$.
If, in addition,

$$\mathrm{Cov}\{|\boldsymbol{X}_T(u)|^2, |\boldsymbol{X}_T(v)|^2\} = \mathrm{Cov}\{|\boldsymbol{Y}(u)|^2, |\boldsymbol{Y}(v)|^2\} = S_{yy}^2(u,v) + S_{yy}^2(u,-v),$$

$$\mathrm{Var}[\boldsymbol{S}_T(\omega)] \approx \begin{cases} S_{xx}^2(\omega), & |\omega| \gg \dfrac{1}{T} \\ 2S_{xx}^2(0), & \omega = 0 \end{cases}$$

Proof: It can be shown that the autocorrelation function of $\boldsymbol{y}(t) = p_T(t)\boldsymbol{x}(t)$ is:

$$R_{yy}(t_1, t_2) = R_{xx}(t_1 - t_2)p_T(t_1)p_T^*(t_2).$$

Spectral Bias

Hence,

$$
\begin{aligned}
S_{yy}(u_1,u_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xx}(t_1-t_2)p_T(t_1)p_T^*(t_2)e^{-j(u_1t_1+u_2t_2)}dt_1dt_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(f)e^{j(t_1-t_2)f}df\right) p_T(t_1)p_T^*(t_2)e^{-j(u_1t_1+u_2t_2)}dt_1dt_2 \\
&= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} p_T(t_1)e^{-j(u_1-f)t_1}dt_1 \int_{-\infty}^{\infty} p_T^*(t_2)e^{j(-u_2-f)t_2}dt_2\right) S_{xx}(f)df \\
&= \int_{-\infty}^{\infty}\frac{1}{2\pi}\left(\frac{2\sin[T(u_1-f)]}{(u_1-f)}\right)\left(\frac{2\sin[T(u_2+f)]}{(u_2+f)}\right) S_{xx}(f)df
\end{aligned}
$$

(Here, we also show that $S_{yy}(u_1,u_2)$ is real! This will be used in Slides 12-51 and 12-73.)

$$
\begin{aligned}
&\approx \int_{-\infty}^{\infty}\frac{1}{2\pi}\left(\frac{2\sin[T(u_1-f)]}{(u_1-f)}\mathbf{1}\left\{-k\le\frac{T(u_1-f)}{\pi}\le k\right\}\right) \\
&\qquad \left(\frac{2\sin[T(u_2+f)]}{(u_2+f)}\mathbf{1}\left\{-k\le\frac{T(u_2+f)}{\pi}\le k\right\}\right) S_{xx}(f)df \text{ for } k \text{ large} \\
&= \int_{-\infty}^{\infty}\frac{1}{2\pi}\left(\frac{2\sin[T(u_1-f)]}{(u_1-f)}\mathbf{1}\left\{u_1-\frac{\pi k}{T}\le f\le u_1+\frac{\pi k}{T}\right\}\right) \\
&\qquad \left(\frac{2\sin[T(u_2+f)]}{(u_2+f)}\mathbf{1}\left\{-u_2-\frac{\pi k}{T}\le f\le -u_2+\frac{\pi k}{T}\right\}\right) S_{xx}(f)df.
\end{aligned}
$$

Spectral Bias

Accordingly, it is fair to say that $S_{yy}(u_1, u_2) \approx 0$ if

$$u_1 + \frac{\pi k}{T} \le -u_2 - \frac{\pi k}{T} \text{ or } -u_2 + \frac{\pi k}{T} \le u_1 - \frac{\pi k}{T},$$

which is equivalent to saying that $|u_1+u_2| \gg 1/T$. (I.e., $S_{yy}(\omega, \omega) \approx 0$ if $|\omega| \gg 1/T$.)

On the other hand, for T large,

$$\begin{aligned} S_{yy}(\omega, -\omega) &= 2T\int_{-\infty}^{\infty} \frac{\sin^2[T(\omega - f)]}{\pi T(\omega - f)^2} S_{xx}(f)df \\ &\approx 2T\int_{-\infty}^{\infty} \delta(\omega - f) S_{xx}(f)df \quad \text{(See the next slide.)} \\ &= 2TS_{xx}(\omega). \end{aligned}$$

Consequently, by the lemma in Slide 11-78,

$$E[\boldsymbol{S}_T(\omega)] = \frac{1}{2T}R_{YY}(\omega, \omega) = \frac{1}{2T}S_{yy}(\omega, -\omega) \to S_{xx}(\omega) \text{ as } T \to \infty.$$

$$\begin{aligned} \mathrm{Var}[\boldsymbol{S}_T(\omega)] &= \mathrm{Cov}\{\boldsymbol{S}_T(\omega), \boldsymbol{S}_T(\omega)\} \\ &= \frac{1}{4T^2}\mathrm{Cov}\{|\boldsymbol{X}_T(\omega)|^2, |\boldsymbol{X}_T(\omega)|^2\} \\ &= \frac{1}{4T^2}\Big[\ \underbrace{S_{yy}^2(\omega, \omega)}_{\approx 0 \text{ if } \omega \gg 1/T} + \underbrace{S_{yy}^2(\omega, -\omega)}_{\approx 4T^2 S_{xx}^2(\omega) \text{ if } T \text{ large}}\ \Big] \approx \begin{cases} S_{xx}^2(\omega), & \omega \gg 1/T \\ 2S_{xx}^2(0), & \omega = 0 \end{cases} \end{aligned}$$

□

Spectral Bias

That $\displaystyle\int_{-\infty}^{\infty}\frac{\sin^2(Ts)}{\pi T s^2}ds = 1$, $\displaystyle\lim_{s\to 0}\frac{\sin^2(Ts)}{\pi T s^2} = \frac{T}{\pi}$, and $\displaystyle\left|\frac{\sin^2(Ts)}{\pi T s^2}\right| \le \frac{1}{\pi T s^2}$ for $s \neq 0$ implies that $\displaystyle\lim_{T\to\infty}\frac{\sin^2(Ts)}{\pi T s^2} = \delta(s)$.

- Although it is asymptotic unbias, $\boldsymbol{S}_T(\omega)$ is still a bias estimate of $S_{xx}(\omega)$ for finite T.

Modified Periodogram

- One can reduce the bias by introducing a data window $c(t)$ to obtain:

$$\boldsymbol{S}_T(\omega; c) = \frac{1}{2T}\left|\int_{-T}^{T} c(t)\boldsymbol{x}(t)e^{-j\omega t}dt\right|^2,$$

which is named the $\boxed{\textit{modified periodogram}}$.

- The modified periodogram is the normalized absolute square of the Fourier transform of the known segment $\boldsymbol{x}_T(t; c) = c_T(t)\boldsymbol{x}(t)$, where $c_T(t) = c(t)p_T(t)$, i.e.,

$$\boldsymbol{S}_T(\omega; c) = \frac{1}{2T}|\boldsymbol{X}_T(\omega; c)|^2 \qquad \boldsymbol{X}_T(\omega; c) = \int_{-\infty}^{\infty} \boldsymbol{x}_T(t; c)e^{-j\omega t}dt.$$

- The choice that minimizes the bias of $\boldsymbol{S}_T(\omega; c)$ will be introduced later.

Modified Periodogram

Let $\boldsymbol{y}(t) = c_T(t)\boldsymbol{x}(t)$. Then

$$R_{yy}(t_1, t_2) = R_{xx}(t_1 - t_2)c_T(t_1)c_T^*(t_2).$$

Hence,

$$\begin{aligned}
S_{yy}(u_1, u_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xx}(t_1 - t_2)c_T(t_1)c_T^*(t_2)e^{-j(u_1t_1+u_2t_2)}dt_1dt_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(f)e^{j(t_1-t_2)f}df\right)c_T(t_1)c_T^*(t_2)e^{-j(u_1t_1+u_2t_2)}dt_1dt_2 \\
&= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} c_T(t_1)e^{-j(u_1-f)t_1}dt_1\int_{-\infty}^{\infty} c_T^*(t_2)e^{j(-u_2-f)t_2}dt_2\right)S_{xx}(f)df \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} C_T(u_1 - f)C_T^*(-u_2 - f)S_{xx}(f)df
\end{aligned}$$

Consequently, by the lemma in Slide 11-78,

$$E[\boldsymbol{S}_T(\omega; c)] = \frac{1}{2T}R_{YY}(\omega, \omega) = \frac{1}{2T}S_{yy}(\omega, -\omega) = \frac{1}{4\pi T}\int_{-\infty}^{\infty} S_{xx}(f)|C_T(\omega - f)|^2df.$$

This will be used later in Slide 12-56.

Remarks on the Condition of Theorem 12-4

Theorem 11-3. An **example** of $\boldsymbol{y}(t)$ for the condition that

$$\text{Cov}\{|\boldsymbol{Y}(u)|^2, |\boldsymbol{Y}(v)|^2\} = S_{yy}^2(u,v) + S_{yy}^2(u,-v)$$

is real Gaussian with zero mean.

Proof:

- Let $\boldsymbol{Y}(u) = \boldsymbol{A}(u) + j\boldsymbol{B}(u)$ and $\boldsymbol{Y}(v) = \boldsymbol{A}(v) + j\boldsymbol{B}(v)$, where $\boldsymbol{A}(u)$, $\boldsymbol{B}(u)$, $\boldsymbol{A}(v)$ and $\boldsymbol{B}(v)$ are jointly Gaussian (because $\boldsymbol{y}(t)$ is Gaussian).

- Derive

$$\begin{aligned}
&\text{Cov}\{|\boldsymbol{Y}(u)|^2, |\boldsymbol{Y}(v)|^2\} \\
&= E[(\boldsymbol{A}^2(u) + \boldsymbol{B}^2(u))(\boldsymbol{A}^2(v) + \boldsymbol{B}^2(v))] - E[\boldsymbol{A}^2(u) + \boldsymbol{B}^2(u)]E[\boldsymbol{A}^2(v) + \boldsymbol{B}^2(v)] \\
&= E[\boldsymbol{A}^2(u)\boldsymbol{A}^2(v)] + E[\boldsymbol{A}^2(u)\boldsymbol{B}^2(v)] + E[\boldsymbol{B}^2(u)\boldsymbol{A}^2(v)] + E[\boldsymbol{B}^2(u)\boldsymbol{B}^2(v)] \\
&\quad - E[\boldsymbol{A}^2(u)]E[\boldsymbol{A}^2(v)] - E[\boldsymbol{A}^2(u)]E[\boldsymbol{B}^2(v)] - E[\boldsymbol{B}^2(u)]E[\boldsymbol{A}^2(v)] - E[\boldsymbol{B}^2(u)]E[\boldsymbol{B}^2(v)] \\
&= 2E^2[\boldsymbol{A}(u)\boldsymbol{A}(v)] + 2E^2[\boldsymbol{A}(u)\boldsymbol{B}(v)] + 2E^2[\boldsymbol{B}(u)\boldsymbol{A}(v)] + 2E^2[\boldsymbol{B}(u)\boldsymbol{B}(v)]
\end{aligned}$$

where the last step follow from the fact that

$$E[\boldsymbol{x}^2\boldsymbol{y}^2] - E[\boldsymbol{x}^2]E[\boldsymbol{y}^2] = 2E^2[\boldsymbol{x}\boldsymbol{y}]$$

for any jointly Gaussian $(\boldsymbol{x}, \boldsymbol{y})$.

Remarks on the Condition of Theorem 12-4

- We further derive

$$
\begin{aligned}
S_{yy}(u,-v) &= R_{YY}(u,v) = E[\boldsymbol{Y}(u)\boldsymbol{Y}^*(v)] \\
&= E[(\boldsymbol{A}(u) + j\boldsymbol{B}(u))(\boldsymbol{A}(v) - j\boldsymbol{B}(v))] \\
&= \underbrace{E[\boldsymbol{A}(u)\boldsymbol{A}(v)]}_{C} + \underbrace{E[\boldsymbol{B}(u)\boldsymbol{B}(v)]}_{D} - j\,\underbrace{E[\boldsymbol{A}(u)\boldsymbol{B}(v)]}_{E} + j\,\underbrace{E[\boldsymbol{B}(u)\boldsymbol{A}(v)]}_{F}
\end{aligned}
$$

and

$$
\begin{aligned}
S_{yy}(u,v) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{yy}(t_1,t_2)e^{-j(ut_1+vt_2)}dt_1dt_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[\boldsymbol{y}(t_1)\boldsymbol{y}(t_2)]e^{-j(ut_1+vt_2)}dt_1dt_2 \\
&= E[\boldsymbol{Y}(u)\boldsymbol{Y}(v)] \\
&= E[(\boldsymbol{A}(u) + j\boldsymbol{B}(u))(\boldsymbol{A}(v) + j\boldsymbol{B}(v))] \\
&= \underbrace{E[\boldsymbol{A}(u)\boldsymbol{A}(v)]}_{C} - \underbrace{E[\boldsymbol{B}(u)\boldsymbol{B}(v)]}_{D} + j\,\underbrace{E[\boldsymbol{A}(u)\boldsymbol{B}(v)]}_{E} + j\,\underbrace{E[\boldsymbol{B}(u)\boldsymbol{A}(v)]}_{F}.
\end{aligned}
$$

Remarks on the Condition of Theorem 12-4

- Since $S_{yy}(u,v)$ is real due to WSS $\boldsymbol{x}(t)$ (See Slide 12-46),

$$\begin{aligned} 2C &= S_{yy}(u,-v) + S_{yy}(u,v) \\ 2D &= S_{yy}(u,-v) - S_{yy}(u,v) \\ 2E &= 0 \\ 2F &= 0 \end{aligned}$$

- Consequently,

$$4C^2 + 4D^2 + 4E^2 + 4F^2 = 2S_{yy}^2(u,-v) + 2S_{yy}^2(u,v).$$

$\square$

$S_{yy}(u,v)$ may be complex under $\boldsymbol{y}(t) = \boldsymbol{x}(t)c(t)$ for general real-valued Gaussian $\boldsymbol{x}(t)$ (not necessarily WSS) and real $c(t)$ (not necessarily symmetric). As such, we have

$$\mathrm{Cov}\{|\boldsymbol{Y}(u)|^2, |\boldsymbol{Y}(v)|^2\} = |S_{yy}(u,-v)|^2 + |S_{yy}(u,v)|^2.$$

Remarks

- The **estimation variance** (namely, $E[|\boldsymbol{S}_T(\omega; c) - S_{xx}(\omega)|^2]$) is larger than the **variance of the estimate** (namely, $\mathrm{Var}[\boldsymbol{S}_T(\omega; c)]$) for every T, but their difference will decrease to zero as $T \to \infty$, if the estimator is asymptotically unbiased.

$$\begin{aligned}
&E[|\boldsymbol{S}_T(\omega; c) - S_{xx}(\omega)|^2] - \mathrm{Var}[\boldsymbol{S}_T(\omega; c)] \\
&\quad = \cancel{E[|\boldsymbol{S}_T(\omega; c)|^2]} - S_{xx}(\omega)E[\boldsymbol{S}_T^*(\omega; c)] - S_{xx}^*(\omega)E[\boldsymbol{S}_T(\omega; c)] + |S_{xx}(\omega)|^2 \\
&\qquad -\cancel{E[|\boldsymbol{S}_T(\omega; c)|^2]} + |E[\boldsymbol{S}_T(\omega; c)]|^2 \\
&\quad = |E[\boldsymbol{S}_T(\omega; c)] - S_{xx}(\omega)|^2 .
\end{aligned}$$

- *Fact without formal proof*: The estimation variance can not be made zero even if T is large, and is asymptotically lower-bounded by $S_{xx}^2(\omega)$.
- *Fact without formal proof*: Hence, the use of data window does not help much in decreasing either the estimation variance or the variance of the estimate (in asymptotic sense).

Smoothed Spectrum

- One way to improve the estimation variance is to smooth the spectrum:

$$\boldsymbol{S}_{T,w}(\omega;c) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W(y)\boldsymbol{S}_T(\omega - y;c)dy$$

  at the price of a slightly larger bias. Notably, $\boldsymbol{S}_T(\omega;c)$ is a biased estimator for finite T.

- It can be anticipated that:

$$\begin{aligned}
\boldsymbol{s}_{T,w}(\tau;c) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \boldsymbol{S}_{T,w}(\omega;c)e^{j\omega\tau}d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} W(y)\boldsymbol{S}_T(\omega - y;c)dy\right)e^{j\omega\tau}d\omega \\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty} W(y)\int_{-\infty}^{\infty} \boldsymbol{S}_T(\omega - y;c)e^{j\omega\tau}d\omega dy, u = \omega - y \\
&= \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} W(y)e^{jy\tau}dy\right)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \boldsymbol{S}_T(u;c)e^{ju\tau}du\right) \\
&= w(\tau)\boldsymbol{s}_T(\tau;c),
\end{aligned}$$

  where $w(\tau)$ is called the *lag window* with the property $w(0) = 1$, and its Fourier transform $W(\omega)$ is called the *spectral window*.

Bias and variance of $\boldsymbol{S}_{T,w}(\omega;c)$

- Bias:

$$\begin{aligned}
E[\boldsymbol{S}_{T,w}(\omega;c)] &= \frac{1}{2\pi}\int_{-\infty}^{\infty} W(y)E[\boldsymbol{S}_T(\omega-y;c)]dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} W(y)\left(\int_{-\infty}^{\infty} S_{xx}(v)\cdot\frac{1}{4\pi T}|C_T(\omega-y-v)|^2 dv\right)dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(v)\left(\frac{1}{4\pi T}\int_{-\infty}^{\infty} W(y)|C_T(\omega-y-v)|^2 dy\right)dv.
\end{aligned}$$

Hence, under certain condition that makes valid the bounded convergence theorem, and using $\lim_{T\to\infty}\dfrac{1}{4\pi T}|C_T(\omega)|^2=\delta(\omega)$, we obtain

$$\begin{aligned}
\lim_{T\to\infty} E[\boldsymbol{S}_{T,w}(\omega;c)] &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(v)\left(\int_{-\infty}^{\infty} W(y)\lim_{T\to\infty}\frac{1}{4\pi T}|C_T(\omega-y-v)|^2 dy\right)dv \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(v)\left(\int_{-\infty}^{\infty} W(y)\delta(\omega-y-v)dy\right)dv \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(v)W(\omega-v)dv.
\end{aligned}$$

If $W(\omega)$ remains constant in a very short duration around origin, and zero outside, then one can retain $\lim_{T\to\infty} E[\boldsymbol{S}_{T,w}(\omega;c)] \approx S_{xx}(\omega)$.

Smoothed Spectrum

- Variance:

$$\mathrm{Var}[\boldsymbol{S}_{T,w}(\omega;c)] = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(y)W^*(z)\mathrm{Cov}\{\boldsymbol{S}_T(\omega-y;c),\boldsymbol{S}_T(\omega-z;c)\}dydz$$

Let us examine the impact of the spectral window on the behavior of the estimation variance of $\boldsymbol{S}_{T,w}(\omega;c)$ through an "artificial" example.

Example. Assume that

$$\begin{cases} c(t)=1 \quad \text{(No data window)} \\ R_{xx}(\tau)=(N_0/2)\delta(\tau) \\ W_0(\omega)=4\sin^2(\omega/2)/\omega^2 \quad \text{and} \quad w_0(\tau)=(1-|\tau|)\mathbf{1}\{|\tau|\le 1\} \text{ Bartlett spectral window} \\ W(\omega)=MW_0(M\omega)=4\sin^2(M\omega/2)/(M\omega^2) \quad \text{and} \quad w(\tau)=w_0(\tau/M) \\ E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(s_1)\boldsymbol{x}^*(t_2)\boldsymbol{x}(s_2)] = R_{xx}(t_1-s_1)R_{xx}(s_2-t_2)+R_{xx}(t_1-t_2)R_{xx}(s_2-s_1) \end{cases}$$

An example for $E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(s_1)\boldsymbol{x}^*(t_2)\boldsymbol{x}(s_2)] = R_{xx}(t_1-s_1)R_{xx}(s_2-t_2)+R_{xx}(t_1-t_2)R_{xx}(s_2-s_1)$ is that $\boldsymbol{x}(t)=\sum_{n=-\infty}^{\infty}\boldsymbol{c}_n e^{jn\theta t}$ for independent complex zero-mean Gaussian $\{\boldsymbol{c}_n\}_{n=-\infty}^{\infty}$ with variance $E[|\boldsymbol{c}_n|^2]=\sigma_n^2$ and $E[\boldsymbol{c}_n^2]=0$ (Thus, $E[|\boldsymbol{c}_n|^4]=2E^2[|\boldsymbol{c}_n|^2]$). See the next slide.

Smoothed Spectrum

$$
\begin{aligned}
&E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(s_1)\boldsymbol{x}^*(t_2)\boldsymbol{x}(s_2)] \\
&\quad = E\left[\left(\sum_{n=-\infty}^{\infty} \boldsymbol{c}_n e^{jn\theta t_1}\right)\left(\sum_{m=-\infty}^{\infty} \boldsymbol{c}_m^* e^{-jm\theta s_1}\right)\left(\sum_{k=-\infty}^{\infty} \boldsymbol{c}_k^* e^{-jk\theta t_2}\right)\left(\sum_{\ell=-\infty}^{\infty} \boldsymbol{c}_\ell e^{j\ell\theta s_2}\right)\right] \\
&\quad = \sum_{n=-\infty}^{\infty} E[|\boldsymbol{c}_n|^4] e^{jn\theta(t_1-s_1-t_2+s_2)} + \underbrace{\sum_{n=-\infty}^{\infty}\sum_{k=-\infty,k\neq n}^{\infty} E[|\boldsymbol{c}_n|^2]E[|\boldsymbol{c}_k|^2] e^{jn\theta(t_1-s_1)} e^{jk\theta(s_2-t_2)}}_{n=m\neq k=\ell} \\
&\qquad + \underbrace{\sum_{n=-\infty}^{\infty}\sum_{m=-\infty,m\neq n}^{\infty} E[|\boldsymbol{c}_n|^2]E[|\boldsymbol{c}_m|^2] e^{jn\theta(t_1-t_2)} e^{jm\theta(s_2-s_1)}}_{n=k\neq m=\ell} + \underbrace{\cancel{\sum_{n=-\infty}^{\infty}\sum_{m=-\infty,m\neq n}^{\infty} E[\boldsymbol{c}_n^2]E[(\boldsymbol{c}_m^*)^2] e^{jn\theta(t_1+s_2)} e^{-jm\theta(s_1+t_2)}}}_{n=\ell\neq m=k} \\
&\quad = \sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} E[|\boldsymbol{c}_n|^2]E[|\boldsymbol{c}_k|^2] e^{jn\theta(t_1-s_1)} e^{jk\theta(s_2-t_2)} + \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} E[|\boldsymbol{c}_n|^2]E[|\boldsymbol{c}_m|^2] e^{jn\theta(t_1-t_2)} e^{jm\theta(s_2-s_1)} \\
&\quad = R_{xx}(t_1-s_1)R_{xx}(s_2-t_2) + R_{xx}(t_1-t_2)R_{xx}(s_2-s_1),
\end{aligned}
$$

where

$$
R_{xx}(\tau) = E[\boldsymbol{x}(t+\tau)\boldsymbol{x}^*(t)] = E\left[\left(\sum_{n=-\infty}^{\infty} \boldsymbol{c}_n e^{jn\theta(t+\tau)}\right)\left(\sum_{m=-\infty}^{\infty} \boldsymbol{c}_m^* e^{-jm\theta t}\right)\right] = \sum_{n=-\infty}^{\infty} E[|\boldsymbol{c}_n|^2] e^{jn\theta\tau}.
$$

By this, we also obtain

$$
S_{xx}(\omega) = \sum_{n=-\infty}^{\infty} E[|\boldsymbol{c}_n|^2]\delta(\omega - n\theta) \approx \frac{N_0}{2} \quad \text{if } \theta \text{ is very small and } E[|\boldsymbol{c}_n|^2] = \frac{N_0}{2}.
$$

Smoothed Spectrum

Then,

$$
\begin{aligned}
&\mathrm{Cov}[\boldsymbol{S}_T(\omega_1;c),\boldsymbol{S}_T(\omega_2;c)] \\
&= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}(E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(s_1)\boldsymbol{x}^*(t_2)\boldsymbol{x}(s_2)]-E[\boldsymbol{x}(t_1)\boldsymbol{x}^*(s_1)]E[\boldsymbol{x}^*(t_2)\boldsymbol{x}(s_2)]) \\
&\quad e^{-j\omega_1(t_1-s_1)}e^{j\omega_2(t_2-s_2)}dt_1dt_2ds_1ds_2 \\
&= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}R_{xx}(t_1-t_2)e^{-j\omega_1 t_1}e^{j\omega_2 t_2}dt_1dt_2\times\int_{-T}^{T}\int_{-T}^{T}R_{xx}(s_2-s_1)e^{j\omega_1 s_1}e^{-j\omega_2 s_2}ds_2ds_1 \\
&= \frac{N_0^2\sin^2(T(\omega_1-\omega_2))}{4T^2(\omega_1-\omega_2)^2},
\end{aligned}
$$

where

$$
\begin{aligned}
\int_{-T}^{T}\int_{-T}^{T}R_{xx}(t_1-t_2)e^{-j\omega_1 t_1}e^{j\omega_2 t_2}dt_1dt_2 &= \int_{-T}^{T}\int_{-T}^{T}\frac{N_0}{2}\delta(t_1-t_2)e^{-j\omega_1 t_1}e^{j\omega_2 t_2}dt_1dt_2 \\
&= \frac{N_0}{2}\int_{-T}^{T}e^{-j(\omega_1-\omega_2)t_2}dt_2 \\
&= \frac{N_0\sin(T(\omega_1-\omega_2))}{(\omega_1-\omega_2)}.
\end{aligned}
$$

Smoothed Spectrum

$$
\begin{aligned}
&\mathrm{Var}[\boldsymbol{S}_{T,w}(\omega;c)]\\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(y)W^*(z)\mathrm{Cov}\{\boldsymbol{S}_T(\omega-y;c),\boldsymbol{S}_T(\omega-z;c)\}dydz\\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{16\sin^2(My/2)\sin^2(Mz/2)}{M^2y^2z^2}\left(\frac{N_0^2\sin^2(T(y-z))}{4T^2(y-z)^2}\right)dydz\\
&= \frac{N_0^2}{\pi^2T^2M^2}\int_{-\infty}^{\infty}\frac{\sin^2(My/2)}{y^2}\int_{-\infty}^{\infty}\frac{\sin^2(Mz/2)\sin^2(T(y-z))}{z^2(y-z)^2}dzdy\\
&= \frac{N_0^2}{\pi TM^2}\int_{-\infty}^{\infty}\frac{\sin^2(My/2)}{y^2}\int_{-\infty}^{\infty}\frac{\sin^2(Mz/2)}{z^2}\left(\frac{\sin^2(T(y-z))}{\pi T(y-z)^2}\right)dzdy\\
&\approx \frac{N_0^2}{16\pi T}\int_{-\infty}^{\infty}\underbrace{\frac{4\sin^2(My/2)}{My^2}}_{W(y)}\int_{-\infty}^{\infty}\underbrace{\frac{4\sin^2(Mz/2)}{Mz^2}}_{W(z)}\delta(y-z)dzdy \text{ at } T \text{ large}\\
&= \frac{N_0^2}{16\pi T}\left(\frac{4\pi}{3}M\right) = \frac{MN_0^2}{12T}\quad\left(=\frac{E_w}{2T}S_{xx}^2(\omega),\ \text{where}\ \begin{cases} S_{xx}(\omega)=\frac{N_0}{2}\\ E_w=\frac{1}{2\pi}\int_{-\infty}^{\infty}W^2(\omega)d\omega=\frac{2}{3}M\end{cases}\right),
\end{aligned}
$$

which decreases to zero as T large, where we use the $\delta(\cdot)$ approximation from Slide 12-48.

Selection of Spectral Window

Tradeoff in the selection of spectral window

- **Bias**: $W(\omega)$ remains constant in a very short duration around origin, and zero outside, in order to have small bias.
- **Variance**: $\text{Var}[\boldsymbol{S}_{T,w}(\omega;c)] \approx \frac{E_w}{2T} S_{xx}^2(\omega)$ at large T, where $E_w = \frac{1}{2\pi}\int_{-\infty}^{\infty} W^2(\omega)d\omega$. This implies that E_w must be small, compared to T.

Question: For a fixed $w_0(\tau)$ with $w_0(\tau) = 0$ for $|\tau| > 1$, define $w(\tau) = w_0(\tau/M)$; then $W(\omega) = MW_0(M\omega)$. Can we find an M that results in small bias and small variance at the same time?

Answer: It is apparent that M shall be **large** for small bias. However,

$$E_w = \frac{1}{2\pi}\int_{-\infty}^{\infty} W^2(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} M^2 W_0^2(M\omega)d\omega = ME_{w_0}$$

implies that M shall be **small** for small variance. □

- Hence, a tradeoff between bias and estimate variance in the selection of M must be made.

Suggestions for the selection of M

1. The spectral window $W(\omega)$ must be positive and its area must equal 2π (as $w(0) = 1$). This ensures the positivity and consistency of the estimate. Some choices of spectral windows are listed below.

	$w_0(\tau)$	$W_0(\omega)$	m_2
Bartlett	$(1 - \lvert\tau\rvert)\mathbf{1}\{\lvert\tau\rvert < \mathbf{1}\}$	$4\sin^2(\omega/2)/\omega^2$	∞
Tukey	$\frac{1}{2}(1 + \cos(\pi\tau))\mathbf{1}\{\lvert\tau\rvert < \mathbf{1}\}$	$\pi^2 \sin(\omega)/[\omega(\pi^2 - \omega^2)]$	$\pi^2/2$
Papoulis	$[\frac{1}{\pi}\lvert\sin(\tau)\rvert + (1 - \lvert\tau\rvert)\cos(\pi\tau)]\mathbf{1}\{\lvert\tau\rvert < \mathbf{1}\}$	$8\pi^2\cos^2(\omega/2)/(\pi^2 - \omega^2)$	π^2

2. The spectral window $W(\omega)$ must go to zero rapidly as ω increases. This reduces the influence of the distant spectrum to the measured location.

3. A usual measure for duration of $W(\omega)$ is the second moment

$$m_2 \triangleq \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 W(\omega) d\omega = \frac{1}{2\pi M^2}\int_{-\infty}^{\infty} \omega^2 W_0(\omega) d\omega.$$

   This shall be small for small bias. (Recall that $E_w = ME_{w_0}$ must be small for small variance of the estimate!)

Minimum Bias Data Window

- It is hard to find an optimal data window to, say, minimize the bias (cf. Slide 12-50):

$$E[\boldsymbol{S}_T(\omega;c)] \;=\; \int_{-\infty}^{\infty} S_{xx}(v)\cdot\frac{1}{4\pi T}|C_T(\omega-v)|^2 dv,$$

where

$$C_T(\omega)\triangleq\int_{-T}^{T} c(t)e^{-j\omega t}dt.$$

- However, under the condition that $S_{xx}(v)$ can be well-approximated by parabolic function, we can show that the truncated cosine data window minimizes the bias.

Sketch of the proof:

$$\begin{aligned} S_{xx}(w-v) &\approx S_{xx}(\omega-v)\big|_{v=0} + v\,\frac{\partial S_{xx}(\omega-v)}{\partial v}\bigg|_{v=0} + \frac{v^2}{2}\,\frac{\partial^2 S_{xx}(\omega-v)}{\partial v^2}\bigg|_{v=0} \\ &= S_{xx}(\omega) - vS'_{xx}(\omega) + \frac{v^2}{2}S''_{xx}(\omega) \end{aligned}$$

Minimum Bias Data Window

implies

$$
\begin{aligned}
E[\boldsymbol{S}_T(\omega;c)] &= \int_{-\infty}^{\infty} S_{xx}(\omega - v)\cdot\frac{1}{4\pi T}|C_T(v)|^2 dv \\
&\approx \int_{-\infty}^{\infty} S_{xx}(\omega)\cdot\frac{1}{4\pi T}|C_T(v)|^2 dv - \int_{-\infty}^{\infty} vS'_{xx}(\omega)\cdot\frac{1}{4\pi T}|C_T(v)|^2 dv \\
&\quad + \int_{-\infty}^{\infty} \frac{v^2}{2}S''_{xx}(\omega)\cdot\frac{1}{4\pi T}|C_T(v)|^2 dv \\
&= S_{xx}(\omega) + \frac{S''_{xx}(\omega)}{8\pi T}\int_{-\infty}^{\infty} v^2|C_T(v)|^2 dv,
\end{aligned}
$$

if $\displaystyle\int_{-\infty}^{\infty}\frac{1}{4\pi T}|C_T(v)|^2 dv = \frac{1}{2T}\int_{-T}^{T}|c(t)|^2 dt = 1$, and $|C_T(v)| = |C_T(-v)|$.

Hence, a data window, which minimizes

$$
\frac{1}{2\pi}\int_{-\infty}^{\infty}|(j\omega)C_T(\omega)|^2 d\omega = \int_{-T}^{T}|c'(t)|^2 dt
$$

subject to the above constraints, minimizes the bias.

Minimum Bias Data Window

It can then be shown that under

$$\int_{-\infty}^{\infty} \frac{1}{4\pi T} |C_T(v)|^2 dv = \frac{1}{2T} \int_{-T}^{T} |c(t)|^2 dt = 1 \quad \text{and} \quad |C_T(v)|^2 = |C_T(-v)|^2,$$

the data window that minimizes

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} v^2 |C_T(v)|^2 dv = \int_{-T}^{T} |c'(t)|^2 dt$$

is:

$$c^\diamond(t) = \sqrt{2} \cos\left(\frac{\pi}{2T} t\right) \mathbf{1}\{|t| \leq T\} \Leftrightarrow C_T^\diamond(\omega) = 4\sqrt{2}\pi T \frac{\cos(T\omega)}{(\pi^2 - 4T^2\omega^2)}.$$

This data window is named the *truncated cosine data window.* $\square$

Minimum Bias Data Window

Suppose $c(t) \geq 0$ for $|t| < T$, and boundary condition $c(T) = c(-T) = 0$. We wish to minimize

$$\int_{-T}^{T} [c'(t)]^2 dt$$

subject to

$$\int_{-T}^{T} c^2(t) dt = 2T.$$

By using Lagrange multipliers technique, we turn to minimize

$$\int_{-T}^{T} [c'(t)]^2 dt - \lambda \left(\int_{-\infty}^{\infty} c^2(t) dt - 2T \right) = \int_{-T}^{T} \underbrace{\left([c'(t)]^2 - \lambda c^2(t) \right)}_{\mathcal{L}(t,c,c')} dt + 2\lambda T$$

Euler-Lagrange Equation for Single function of single variable with higher derivatives

The stationary values of the functional

$$I[f] = \int_{x_0}^{x_1} \mathcal{L}(x, f, f', \ldots, f^{(k)}) dx$$

can be obtained from the Euler-Lagrange equation:

$$\frac{\partial \mathcal{L}}{\partial f} - \frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial f'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial \mathcal{L}}{\partial f''}\right) - \cdots + (-1)^k \frac{d^k}{dx^k}\left(\frac{\partial \mathcal{L}}{\partial f^{(k)}}\right) = 0$$

under fixed boundary conditions for the function itself as well as for the first $k-1$ derivatives.

We have $\mathcal{L}(t,c,c') = [c'(t)]^2 - \lambda c^2(t)$, which implies

$$\frac{\partial \mathcal{L}}{\partial c} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial c'}\right) = (-2\lambda c(t)) - \frac{d}{dt}(2c'(t)) = 0 \Leftrightarrow c''(t) + \lambda c(t) = 0$$

Theorem 8.6 [Tom M. Apostoal, *Calculus*, pp. 326, Volume 1, 2nd Edition, 1967] The solution of the equation $y''(x) + by(x) = 0$ is

$$y(x) = c_1 u_1(x) + c_2 u_2(x),$$

where c_1 and c_2 are constants determined by the initial condition, and

1. $u_1(x) = 1$ and $u_2(x) = x$ if $b = 0$;
2. $u_1(x) = e^{kx}$ and $u_2(x) = e^{-kx}$ if $b = -k^2 < 0$;
3. $u_1(x) = \cos(kx)$ and $u_2(x) = \sin(kx)$ if $b = k^2 > 0$.

$i)$ $c(t) = c_1\cos(kt) + c_2\sin(kt)$ with $\lambda = k^2$.

$ii)$ $c(T) = c(-T) = 0 \Rightarrow c_1\cos(kT) + c_2\sin(kT) = c_1\cos(kT) - c_2\sin(kT) = 0$

$\Rightarrow c_1\cos(kT) = c_2\sin(kT) = 0$ (Note that we cannot have $c_1 = c_2 = 0$!)

$iii)$ Since $\cos(x) = \sin(x) = 0$ has no solution, one of c_1 and c_2 is zero.

Then, $c(t) \geq 0$ for $|t| \leq T$ implies $c_2 = 0$, and hence, $k = \frac{\pi}{2T}$.

$iv)$ $\int_{-T}^{T} c^2(t)dt = 2T$ implies $c_1 = \sqrt{2}$.

Minimum Bias Spectral Window

- It is hard to find an optimal spectral window to, say, minimize the asymptotic bias (cf. Slide 12-56):

$$\lim_{T\to\infty} E[\boldsymbol{S}_{T,w}(\omega)] \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(v)W(\omega - v)dv.$$

- However, under the condition that $S_{xx}(v)$ can be well-approximated by parabolic function, we can show that the Papoulis spectral window minimizes the asymptotic bias.

Sketch of the proof:

$$\begin{aligned} S_{xx}(w-v) &\approx S_{xx}(\omega - v)\big|_{v=0} + v\left.\frac{\partial S_{xx}(\omega - v)}{\partial v}\right|_{v=0} + \frac{v^2}{2}\left.\frac{\partial^2 S_{xx}(\omega - v)}{\partial v^2}\right|_{v=0} \\ &= S_{xx}(\omega) - vS'_{xx}(\omega) + \frac{v^2}{2}S''_{xx}(\omega) \end{aligned}$$

Minimum Bias Spectral Window

implies

$$\begin{aligned}
\lim_{T\to\infty} E[\boldsymbol{S}_{T,w}(\omega)] &= \frac{1}{2\pi}\int_{-\infty}^{\infty} W(v)S_{xx}(\omega - v)dv \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} W(v)S_{xx}(\omega)dv - \frac{1}{2\pi}\int_{-\infty}^{\infty} W(v)vS'_{xx}(\omega)dv \\
&\quad + \frac{1}{2\pi}\int_{-\infty}^{\infty} W(v)\frac{v^2}{2}S''_{xx}(\omega)dv \\
&= S_{xx}(\omega) + \frac{1}{4\pi}S''_{xx}(\omega)\int_{-\infty}^{\infty} v^2 W(v)dv,
\end{aligned}$$

if $\frac{1}{2\pi}\int_{-\infty}^{\infty} W(v)dv = 1$ and $W(v) = W(-v)$.

Hence, a spectral window, which minimizes

$$m_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} v^2 W(v)dv$$

subject to the above constraints, minimizes the asymptotic bias.

Minimum Bias Spectral Window

It can then be shown that under

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} W(v)dv = 1 \quad \text{and} \quad W(v) = W(-v) \quad \text{and} \quad \underbrace{W(\omega) \geq 0}_{\text{additional constraint}},$$

the spectral window that minimizes

$$m_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} v^2 W(v)dv$$

is:

$$W(\omega) = \frac{1}{2T}|C_T^{\diamond}(\omega)|^2 = 16\pi^2 T \frac{\cos^2(T\omega)}{(\pi^2 - 4T^2\omega^2)^2}$$

This is exactly what is obtained in text by assigning $M = 2T$. You may compare this solution to the one in Slide 12-65 and you shall find that $W(\omega)$ is proportional to $|C_T^{\diamond}(\omega)|^2$.

This spectral window is named the *Papoulis spectral window*. □

LMS Spectral Windows

- Another criterion for the optimality of spectral windows is the minimization of the MS estimation error defined as:

$$\text{bias}^2 + \text{variance} \;=\; \left|E[\boldsymbol{S}_{T,w}(\omega;c)] - S_{xx}(\omega)\right|^2 + \text{Var}[\boldsymbol{S}_{T,w}(\omega;c)]$$

- It is called the ML estimation error because

$$\begin{aligned}\text{bias}^2 + \text{variance} \;&=\; \cancel{|E[\boldsymbol{S}_{T,w}(\omega;c)]|^2} - E[\boldsymbol{S}_{T,w}(\omega;c)]S^*_{xx}(\omega) - E[\boldsymbol{S}^*_{T,w}(\omega;c)]S_{xx}(\omega) + |S_{xx}(\omega)|^2 \\ &\quad + E[|\boldsymbol{S}_{T,w}(\omega;c)|^2] - \cancel{|E[\boldsymbol{S}_{T,w}(\omega;c)]|^2} \\ &=\; E[|\boldsymbol{S}_{T,w}(\omega;c) - S_{xx}(\omega)|^2]\end{aligned}$$

In such case, we may say that $\boldsymbol{S}_{T,w}(\omega;c)$ can be well-approximated or well-modeled by $S_{xx}(\omega) + \boldsymbol{v}_w(\omega;c)$, where $\boldsymbol{v}_w(\omega;c)$ is an approximate noise.

LMS Spectral Windows

- (For simplicity, we disregard the data window c and spectral window w in this derivation.)

  From Theorem 12-4, we know that for T very large, $1/T$ is very small. Hence, we can say that $S_{yy}(u,v)$ and $S_{yy}(u,-v)$ are close to zero for most u and v considered (i.e., $|u+v| \gg 1/T$ and $|u-v| \gg 1/T$ are true for most u and v).

- Thus, by $\boldsymbol{S}_T(\omega) = S_{xx}(\omega) + \boldsymbol{v}(\omega)$,

$$\begin{aligned}
E[\boldsymbol{v}(u)\boldsymbol{v}^*(v)] &= \mathrm{Cov}\{\boldsymbol{S}_T(u), \boldsymbol{S}_T(v)\} \\
&= \frac{1}{4T^2}\mathrm{Cov}\{|\boldsymbol{X}_T(u)|^2, |\boldsymbol{X}_T(v)|^2\} \\
&= \frac{1}{4T^2}(S_{yy}^2(u,v) + S_{yy}^2(u,-v)) \approx 0.
\end{aligned}$$

LMS Spectral Windows

- When $|u+v| \gg \frac{1}{T}$, $S_{xx}(\omega)$ is almost a constant within the "effective integration range" of $S_{yy}(u,-v)$. (See below).

$$
\begin{aligned}
E[\boldsymbol{v}(u)\boldsymbol{v}^*(v)] &= \mathrm{Cov}\{\boldsymbol{S}_T(u), \boldsymbol{S}_T(v)\} \\
&= \frac{1}{4T^2}\mathrm{Cov}\{|\boldsymbol{X}_T(u)|^2, |\boldsymbol{X}_T(v)|^2\} \\
&\approx \frac{1}{4T^2}S_{yy}^2(u,-v) \quad (S_{yy}(u,v) \approx 0 \text{ when } |u+v| \gg 1/T \text{ from Theorem 12-4}) \\
&= \frac{1}{4T^2}\left(\int_{-\infty}^{\infty}\frac{2}{\pi}\frac{\sin[T(u-f)]\sin[T(v-f)]}{(u-f)(v-f)}S_{xx}(f)df\right)^2 \quad \text{(From Slide 12-46)} \\
&\approx \frac{1}{4T^2}\left(S_{xx}(u)\int_{-\infty}^{\infty}\frac{2}{\pi}\frac{\sin[T(u-f)]\sin[T(v-f)]}{(u-f)(v-f)}df\right)^2 \quad (S_{xx}(u) \approx \text{constant}) \\
&= \frac{1}{4T^2}\left(S_{xx}(u)\frac{2\sin[T(u-v)]}{(u-v)}\right)^2 \\
&= \frac{\pi}{T}S_{xx}^2(u)\frac{\sin^2[T(u-v)]}{\pi T(u-v)^2} \\
&\approx \frac{\pi}{T}S_{xx}^2(u)\delta(u-v), \quad \text{(See Slide 12-48.)}
\end{aligned}
$$

where

$$
\begin{aligned}
\int_{-\infty}^{\infty} \frac{2}{\pi} \frac{\sin[T(u-f)]\sin[T(v-f)]}{(u-f)(v-f)} df &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{2\sin[T(u-f)]}{(u-f)} \right) \left(\frac{2\sin[T(v-f)]}{(v-f)} \right) df \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-T}^{T} e^{-j(u-f)t_1} dt_1 \right) \left(\int_{-T}^{T} e^{-j(v-f)t_2} dt_2 \right) df \\
&= \int_{-T}^{T} \int_{-T}^{T} e^{-j(ut_1+vt_2)} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{jf(t_1+t_2)} df \right) dt_1 dt_2 \\
&= \int_{-T}^{T} \int_{-T}^{T} e^{-j(ut_1+vt_2)} \delta(t_1+t_2) dt_1 dt_2 \\
&= \int_{-T}^{T} e^{j(u-v)t_2} dt_2 \\
&= \frac{2\sin[T(u-v)]}{(u-v)}.
\end{aligned}
$$

LMS Spectral Windows

The above "rough" derivation formulates, e.g., the following research problem.

Question: Determine the best Δ that minimizes the MS estimation error for moving average spectral window of area 1, provided that $S_{xx}(\omega-\alpha) \approx S_{xx}(\omega) - \alpha S'_{xx}(\omega) + (\alpha^2/2)S''_{xx}(\omega)$, and $\boldsymbol{S}_T(\omega) \approx S_{xx}(\omega) + \boldsymbol{v}(\omega)$, where $\boldsymbol{v}(\omega)$ is zero mean with covariance function $E[\boldsymbol{v}(u)\boldsymbol{v}^*(v)] = (\pi/T)S^2_{xx}(u)\delta(u-v)$.

Answer:

- Since

$$\boldsymbol{S}_{T,w}(\omega) = \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} \boldsymbol{S}_T(\omega-\alpha)d\alpha \approx \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} S_{xx}(\omega-\alpha)d\alpha + \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} \boldsymbol{v}(\omega-\alpha)d\alpha,$$

we derive

$$\begin{aligned} E[\boldsymbol{S}_{T,w}(\omega)] - S_{xx}(\omega) &\approx \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} S_{xx}(\omega-\alpha)d\alpha - S_{xx}(\omega) \\ &\approx \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} \frac{\alpha^2}{2} S''_{xx}(\omega)d\alpha \\ &= S''_{xx}(\omega)\frac{\Delta^2}{6}. \end{aligned}$$

LMS Spectral Windows

By observing that

$$\boldsymbol{S}_{T,w}(\omega) = \frac{1}{2\Delta}\int_{\omega-\Delta}^{\omega+\Delta} \boldsymbol{S}_T(\alpha)d\alpha \approx S_{xx}(\omega) + \frac{1}{2\Delta}\int_{\omega-\Delta}^{\omega+\Delta} \boldsymbol{v}(\alpha)d\alpha,$$

we further derive

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{S}_{T,w}(\omega)] &= E[|\boldsymbol{S}_{T,w}(\omega) - E[\boldsymbol{S}_{T,w}(\omega)]|^2] \\
&= E\left[\left|\frac{1}{2\Delta}\int_{\omega-\Delta}^{\omega+\Delta} \boldsymbol{v}(\alpha)d\alpha\right|^2\right] \\
&= \frac{1}{4\Delta^2}\int_{\omega-\Delta}^{\omega+\Delta}\int_{\omega-\Delta}^{\omega+\Delta} \frac{\pi}{T} S_{xx}^2(\alpha)\delta(\alpha-\beta)d\alpha d\beta \\
&= \frac{\pi}{4\Delta^2 T}\int_{\omega-\Delta}^{\omega+\Delta} S_{xx}^2(\alpha)d\alpha \\
&\approx \frac{\pi}{4\Delta^2 T} S_{xx}^2(\omega)[2\Delta] \quad \text{(By mean-value theorem)} \\
&= S_{xx}^2(\omega)\frac{\pi}{2\Delta T}.
\end{aligned}$$

Hence, we need to find the Δ that minimizes

$$S_{xx}^2(\omega)\frac{\pi}{2\Delta T} + (S''_{xx}(\omega))^2\frac{\Delta^4}{36},$$

which implies that $\Delta^* = \left(\frac{9\pi}{2T}\right)^{1/5}\left(\frac{S_{xx}(\omega)}{S''_{xx}(\omega)}\right)^{2/5}$. $\square$

Remark

- For the optimal Δ^*,

$$\text{bias} = \left(\frac{\pi^2}{384T^2}\right)^{1/5} S_{xx}^{4/5}(\omega)[S''_{xx}(\omega)]^{1/5}$$

and

$$\sqrt{\text{variance}} = \left(\frac{\pi^2}{12T^2}\right)^{1/5} S_{xx}^{4/5}(\omega)[S''_{xx}(\omega)]^{1/5}.$$

Hence,

$$\frac{\text{bias}}{\sqrt{\text{variance}}} = \frac{1}{2}.$$

In other words, standard deviation is equal to twice of the bias, which is named *two-to-one rule.*

LMS Spectral Windows

Final remark

- The best Δ^* is actually a function of ω, which means that the moving average window size is varying with ω!

Question: Determine the best $W(\omega)$ (bandlimited to Δ and of area 1) that minimizes the MS estimation error, provided that $S_{xx}(\omega - \alpha) \approx S_{xx}(\omega) - \alpha S'_{xx}(\omega) + (\alpha^2/2)S''_{xx}(\omega)$, and $\boldsymbol{S}_T(\omega) \approx S_{xx}(\omega) + \boldsymbol{v}(\omega)$, where $\boldsymbol{v}(\omega)$ is zero mean with covariance function $E[\boldsymbol{v}(u)\boldsymbol{v}^*(v)] = (\pi/T)S^2_{xx}(u)\delta(u-v)$.

Answer by Priestley:

$$\boldsymbol{S}_{T,w}(\omega) = \frac{3}{4\Delta}\int_{-\Delta}^{\Delta} \boldsymbol{S}_T(\omega - \alpha)\left(1 - \frac{\alpha^2}{\Delta^2}\right) d\alpha,$$

where $\Delta = \left(\frac{15\pi}{T}\right)^{1/5}\left(\frac{S_{xx}(\omega)}{S''_{xx}(\omega)}\right)^{2/5}$.

The end of Section 12-2 Spectrum Estimation