

# Statistical Signal Processing

## Lecture 5

chapter 1: parameter estimation: deterministic parameters

- some optimality properties
- Maximum Likelihood estimation, examples
- Fischer Information Matrix
- Cramer-Rao lower bound on the MSE, example
- linear model
- asymptotic (large sample) properties
- recap: estimator properties and estimators
- simplified estimators: BLUE, (W)LS, method of moments

## Asymptotic (Large Sample) Properties

- asymptotic:  $n \rightarrow \infty$
- *asymptotically unbiased*:  $\lim_{n \rightarrow \infty} b_n(\theta) = 0$ ,  $\forall \theta \in \Theta$
- Example (mean and variance of Gaussian i.i.d. variables):

$$E[\widehat{\sigma}_{ML}^2 | \mu, \sigma^2] = \frac{n-1}{n} \sigma^2$$
$$b_n = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

$\widehat{\sigma}_{ML}^2$ : biased but asymptotically unbiased

- *consistency*: convergence of (a series of random vectors:)  $\widehat{\theta}_n \rightarrow \theta$ 
  - convergence in probability
  - mean square convergence
  - convergence with probability one
  - convergence in distribution

## Consistency

the sequence of estimates  $\widehat{\theta}(Y_n)$  is said to be

- *simply or weakly consistent* if

$$\lim_{n \rightarrow \infty} \Pr_{Y_n|\theta} \{ \|\widehat{\theta}(Y_n) - \theta\| < \epsilon \} = 1, \quad \forall \epsilon > 0, \quad \forall \theta \in \Theta$$

- *mean-square consistent* if

$$\lim_{n \rightarrow \infty} \text{MSE}_n = \lim_{n \rightarrow \infty} E_{Y_n|\theta} \|\widehat{\theta}(Y_n) - \theta\|^2 = 0, \quad \forall \theta \in \Theta$$

- *strongly consistent* if

$$\Pr_{Y_\infty|\theta} \{ \lim_{n \rightarrow \infty} \widehat{\theta}(Y_n) = \theta \} = 1, \quad \forall \theta \in \Theta$$

- Any of these 3 consistencies implies asymptotic unbiasedness. E.g. for mean-square:

$$\underbrace{E_{Y_n|\theta} \|\widehat{\theta}(Y_n) - \theta\|^2}_{\text{MSE}} = \underbrace{\|E_{Y_n|\theta} \widehat{\theta}(Y_n) - \theta\|^2}_{\text{bias}} + \underbrace{E_{Y_n|\theta} \|\widehat{\theta}(Y_n) - E_{Y_n|\theta} \widehat{\theta}\|^2}_{\text{variance}} \rightarrow 0$$

$$\Rightarrow \lim_{n \rightarrow \infty} E_{Y_n|\theta} \widehat{\theta}(Y_n) = \theta$$

## Consistency (2)

- Strong and mean-square consistency do not imply each other in general. Either implies weak consistency (e.g. use the Chebyshev inequality to show that mean-square consistency implies weak consistency), but not conversely. Except when  $\Theta$  is bounded: then weak consistency implies mean-square consistency.

- example: i.i.d.  $y_i \sim \mathcal{N}(\mu, \sigma^2)$ ,  $\theta = \mu$ ,  $\sigma^2$  known.  $\widehat{\mu}_{ML} = \bar{y}$

$$\text{Var}(\widehat{\mu}_{ML}) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0 \quad \text{mean-square consistent}$$

- example: i.i.d.  $y_i \sim U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ ,  $\widehat{\theta}_{ML} = \frac{y_{\min} + y_{\max}}{2}$

$$\begin{cases} y_{\min} \rightarrow \theta - \frac{1}{2} & \text{in probability} \\ y_{\max} \rightarrow \theta + \frac{1}{2} & \text{in probability} \end{cases} \quad \text{weak consistency}$$

$$\widehat{\theta}_{ML} \rightarrow \theta \quad \text{in probability}$$

mean-square consistency can also be shown

## Asymptotic Normality

- if  $\hat{\theta}_n$  consistent, then  $\tilde{\theta} \rightarrow 0$  in some sense
- introduce a magnifying glass:  $d_n(\hat{\theta}_n - \theta)$  where  $0 < d_{n-1} \leq d_n \rightarrow \infty$
- *convergence in distribution*: weaker than the 3 forms of convergence of sequences of random vectors mentioned before
- if  $d_n(\hat{\theta}_n - \theta) \xrightarrow{\text{in dist.}} \xi$ , some random vector, then the distribution of  $\xi$  useful as a measure for the limiting behavior of  $\hat{\theta}_n$
- usually  $d_n = \sqrt{n}$
- $\hat{\theta}_n$  *consistent asymptotically normal* (CAN) :  
if  $\hat{\theta}_n$  simply consistent and  $d_n(\hat{\theta}_n - \theta) \xrightarrow{\text{in dist.}} \mathcal{N}(0, \Xi(\theta))$   
CAN implies asympt. unbiased (which requires that bias  $\rightarrow 0$  faster than  $\frac{1}{d_n}$ ),  
 $\Xi$  = asymptotic normalized covariance of  $\hat{\theta}_n$ . We say that  $\hat{\theta}_n = \theta + \mathcal{O}_p(\frac{1}{d_n})$
- distinguish  $\Xi(\theta)$  from  $V(\theta) = \lim_{n \rightarrow \infty} d_n^2 C_{\hat{\theta}\hat{\theta}}(\theta)$  which may not even exist for a CAN estimate (if  $\hat{\theta}_n$  is simply but not mean-square consistent).  $V(\theta)$  exists for a mean-square consistent  $\hat{\theta}_n$ , but is not necessarily  $= \Xi(\theta)$ .
- Hence CAN can be used to formulate *interval estimators* on the basis of *point estimators*.

## Asymptotic Optimality of ML

- *asymptotic normalized information matrix* :  $J_0(\theta) = \lim_{n \rightarrow \infty} \frac{1}{d_n^2} J_n(\theta)$  if it exists  
( $J_0(\theta)$  = asymptotic average information per data sample  $y_n$  if  $d_n = \sqrt{n}$ )
- *best asymptotically normal* (BAN): CAN and  $\Xi(\theta) = J_0^{-1}(\theta)$   
also called *asymptotically efficient*
- under some regularity conditions (maximum of the likelihood function unique,  $y_i$  given  $\theta$  i.i.d.,...) the ML estimate is strongly consistent and BAN with  $d_n = \sqrt{n}$  ( $\Rightarrow$  another use of the CRB). In particular, the ML estimate is
  - asymptotically unbiased
  - asymptotically efficient (i.i.d.:  $J_n = nJ_1 \Rightarrow J_0 = J_1$ )
  - asymptotically normal
- example: i.i.d.  $y_i \sim \mathcal{N}(\mu, \sigma^2)$ ,  $\theta = \mu$ ,  $\sigma^2$  known.  $\hat{\mu}_{ML} = \bar{y}$

$$\hat{\mu}_{ML} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \longrightarrow \sqrt{n}(\hat{\mu}_{ML} - \mu) \sim \mathcal{N}(0, \sigma^2), \quad J_n = \frac{n}{\sigma^2} \Rightarrow J_0^{-1} = \sigma^2 = \Xi(\theta)$$

## Recap: Properties of Estimators $\hat{\theta}(Y)$

small sample (finite  $n$ ):

- *bias*:  $b_{\hat{\theta}}(\theta) = E_{Y|\theta} \hat{\theta}(Y) - \theta = -E_{Y|\theta} \tilde{\theta} = -m_{\tilde{\theta}} \quad (= 0, \forall \theta \in \Theta : \text{unbiased})$
- *error correlation*:  $R_{\tilde{\theta}\tilde{\theta}} = E_{Y|\theta} (\hat{\theta}(Y) - \theta) (\hat{\theta}(Y) - \theta)^T = C_{\tilde{\theta}\tilde{\theta}} + b_{\tilde{\theta}} b_{\tilde{\theta}}^T$

Cramer-Rao Bound :  $\hat{\theta}$  unbiased:  $R_{\tilde{\theta}\tilde{\theta}} = C_{\tilde{\theta}\tilde{\theta}} = C_{\hat{\theta}\hat{\theta}}$

$$C_{\tilde{\theta}\tilde{\theta}} \geq J^{-1}(\theta) \quad , \quad J(\theta) = -E_{Y|\theta} \frac{\partial}{\partial \theta} \left( \frac{\partial \ln f(Y|\theta)}{\partial \theta} \right)^T \quad \text{information matrix}$$

*efficient*:  $C_{\tilde{\theta}\tilde{\theta}} = J^{-1}(\theta) \quad , \quad \forall \theta \in \Theta \quad \Rightarrow \quad \hat{\theta}(Y) \text{ is UMVUE}$

large sample ( $n \rightarrow \infty$ ):

- *asymptotically unbiased*:  $\lim_{n \rightarrow \infty} b_{\hat{\theta}}(\theta) = 0, \forall \theta \in \Theta$
- *consistency* (weak, in mean square, strong):  $\Rightarrow$  asymptotically unbiased
- *asymptotic normality*:

$$\text{BAN} \left\{ \begin{array}{l} \diamond \text{ weakly consistent} \\ \diamond \text{ asymptotically normal} \\ \diamond \text{ asymptotically efficient} \end{array} \right\} \text{ CAN}$$

## Recap: Estimation Techniques

- *Uniformly Minimum Variance Unbiased Estimator* (UMVUE):  
complicated (via "sufficient statistics")
- *Maximum likelihood* (ML):  $\hat{\theta}_{ML} = \arg \max_{\theta} f(Y|\theta)$

Qualities:

- ◇ if  $\exists$  efficient  $\hat{\theta} = \hat{\theta}_{eff}$  and  $\hat{\theta}_{ML}$  is obtained from  $\frac{\partial \ln f(Y|\theta)}{\partial \theta} = 0$   
 $\Rightarrow \hat{\theta}_{eff} = \hat{\theta}_{ML} = \hat{\theta}_{UMVUE}$
- ◇  $\hat{\theta}_{ML} = \text{BAN}$

Problems:

- ◇ what if  $f(Y|\theta)$  is unknown?
- ◇ if  $f(Y|\theta)$  is not concave (local maxima)
- *simplified estimators*:
  - ◇ *Best Linear Unbiased Estimator* (BLUE)  $\rightarrow$  linear model
  - ◇ *Method of Moments*
  - ◇ *Least-Squares* (LS)  $\rightarrow$  linear model

## Best Linear Unbiased Estimator (BLUE)

- deterministic analog of LMMSE in the Bayesian case
- *linear*:  $\hat{\theta}(Y) = F Y \quad (F : m \times n)$
- *unbiased*:  $E_{Y|\theta} \hat{\theta} = F E(Y|\theta) = \theta$
- *best = minimum variance*:  $\min C_{\hat{\theta}\hat{\theta}}$
- remarks:
  - BLUE inferior to UMVUE unless UMVUE is linear
  - generalizations:  $X = g(Y) : \hat{\theta}(Y) = F X = F g(Y)$  (linear in  $X$ )  
e.g.: linear in  $Y$  inappropriate if  $\theta \neq 0$  and  $E(Y|\theta) = 0$

### Example of $X = g(Y)$

- $y_i \sim \mathcal{N}(0, \sigma^2)$  i.i.d.,  $\theta = \sigma^2$ ,  $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$
- linear:  $\hat{\sigma}^2 = F Y \Rightarrow E_{Y|\sigma^2} \hat{\sigma}^2 = F E(Y|\sigma^2) = 0 \neq \sigma^2$   
no linear unbiased estimator  $\hat{\sigma}^2$  exists
- however, let  $x_i = y_i^2$ ,  $X = \begin{bmatrix} y_1^2 \\ \vdots \\ y_n^2 \end{bmatrix}$ ,  $E X = \begin{bmatrix} E y_1^2 \\ \vdots \\ E y_n^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ \vdots \\ \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{1}$
- $\hat{\sigma}^2 = F X \Rightarrow E_{Y|\sigma^2} \hat{\sigma}^2 = F E(X|\sigma^2) = \sigma^2 F \mathbf{1} = \sigma^2 \Rightarrow F \mathbf{1} = 1$
- for this problem:  $\hat{\sigma}^2_{UMVUE} = \frac{1}{n} \mathbf{1}^T X = \hat{\sigma}^2_{BLUE} \quad (F = \frac{1}{n} \mathbf{1}^T)$

## BLUE Assumptions

- unbiased:  $FE(Y|\theta) = \theta$ ,  $\forall \theta \in \Theta$   
unbiasedness and the requirement that a large class of linear unbiased estimators (many  $F$  satisfying  $FE(Y|\theta) = \theta$ ) should exist naturally lead to:
- *assumption 1*:  $E(Y|\theta) = H\theta$ ,  $(H : n \times m)$   
unbiasedness  $\rightarrow FH = I_m$  ( $\Rightarrow n \geq m$ )
- variance:  

$$\begin{aligned} C_{\tilde{\theta}\tilde{\theta}} &= C_{\hat{\theta}\hat{\theta}} = E_{Y|\theta} (\hat{\theta} - E_{Y|\theta} \hat{\theta}) (\hat{\theta} - E_{Y|\theta} \hat{\theta})^T \\ &= E_{Y|\theta} (FY - FE(Y|\theta)) (FY - FE(Y|\theta))^T \\ &= FE_{Y|\theta} (Y - E(Y|\theta)) (Y - E(Y|\theta))^T F^T = FC_{YY}(\theta) F^T \end{aligned}$$
- *assumption 2*:  $C_{YY}(\theta) = c(\theta)C$   
 $c(\theta) (> 0, \forall \theta)$  is a scalar function of  $\theta$ ,  $C > 0$  is constant w.r.t.  $\theta$

## BLUE Optimization Problem

- $\min_{\hat{\theta}: E_{Y|\theta} \hat{\theta}(Y) = \theta} C_{\tilde{\theta}\tilde{\theta}} \rightarrow \min_{F: FH=I} FCFT$
- introduce matrix square root  $B (n \times n)$  of  $C = C^T > 0 (n \times n)$ :  $C = BB^T$   
notation:  $B = C^{1/2}$ ,  $C^{T/2} = (C^{1/2})^T$ ,  $C = C^{1/2}C^{T/2}$ ,  $C^{-1} = C^{-T/2}C^{-1/2}$
- Consider a vector space of matrices with  $n$  columns with matrix inner product  $\langle X_1, X_2 \rangle = X_1 X_2^T$ . Take  $X_1 = H^T C^{-T/2}$ ,  $X_2 = F C^{1/2}$ . With  $FH = I$ :  

$$\left\langle \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right\rangle = \begin{bmatrix} H^T C^{-T/2} \\ F C^{1/2} \end{bmatrix} \begin{bmatrix} H^T C^{-T/2} \\ F C^{1/2} \end{bmatrix}^T = \begin{bmatrix} H^T C^{-1} H & I \\ I & F C F^T \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \geq 0$$
- From the Schur Complements Lemma,  $R_{22} \geq R_{21} R_{11}^{-1} R_{12}$  with equality iff  $X_2 = R_{21} R_{11}^{-1} X_1$ .
- Hence  $\min_{F: FH=I} FCFT = (H^T C^{-1} H)^{-1}$   
for  $F = (H^T C^{-1} H)^{-1} H^T C^{-1} = (H^T C_{YY}^{-1} H)^{-1} H^T C_{YY}^{-1}$ .
- Or  $\hat{\theta}_{BLUE} = (H^T C^{-1} H)^{-1} H^T C^{-1} Y = (H^T C_{YY}^{-1} H)^{-1} H^T C_{YY}^{-1} Y$   
with  $C_{\tilde{\theta}\tilde{\theta}} = F C_{YY} F^T = c(\theta) F C F^T = c(\theta) (H^T C^{-1} H)^{-1} = (H^T C_{YY}^{-1} H)^{-1}$

## BLUE: Example Cont'd and Recap

Example Cont'd:

- $y_i \sim \mathcal{N}(0, \sigma^2)$  i.i.d.,  $\theta = \sigma^2$ ,  $x_i = y_i^2$ ,  $\widehat{\sigma^2} = F X$
  - BLUE assumptions OK:  $E(X|\sigma^2) = \mathbf{1} \sigma^2 = H \theta$ ,  $C_{XX} = 2\sigma^4 I = c(\theta) C$
- $$R_{x_i x_j} = E y_i^2 y_j^2 = \begin{cases} \sigma^4 & , i \neq j \\ 3\sigma^4 & , i = j \end{cases} \Rightarrow R_{XX} = 2\sigma^4 I + \sigma^4 \mathbf{1}\mathbf{1}^T, C_{XX} = R_{XX} - m_X m_X^T = 2\sigma^4 I$$
- $\widehat{\sigma^2}_{BLUE} = (H^T C^{-1} H)^{-1} H^T C^{-1} X = \frac{1}{n} \mathbf{1}^T X = \overline{y^2}$
  - $C_{\widehat{\sigma^2} \widehat{\sigma^2}}(\sigma^2) = (H^T C_{XX}^{-1} H)^{-1} = \frac{2\sigma^4}{n}$   $H = \mathbf{1}, C = I, c(\theta) = 2\sigma^4$
  - note: this example is not a linear model!

Recap: BLUE assumptions:

- $\begin{cases} (1) E(Y|\theta) = H \theta \\ (2) C_{YY}(\theta) = c(\theta) C \end{cases}$

Only need to know the first two moments of  $f(Y|\theta)$  which need to satisfy these assumptions. The higher-order moments of  $f(Y|\theta)$ : don't need to know, can be arbitrary functions of  $\theta$ . So the problem should more or less look like a linear model problem, up to the second-order moments.

## BLUE: Linear Model

- $Y = H \theta + V$ ,  $EV = 0$ ,  $EVV^T = C_{VV}$   
( $EV$  and  $C_{VV}$  independent of  $\theta$ , only first two moments of  $V$  specified)
- BLUE assumptions satisfied:
 
$$\begin{cases} E(Y|\theta) = H \theta \\ C_{YY}(\theta) = E_{Y|\theta} (Y - E(Y|\theta)) (Y - E(Y|\theta))^T = E_V V V^T = C_{VV} = C \quad (c(\theta) = 1) \end{cases}$$
- $\widehat{\theta}_{BLUE} = (H^T C_{VV}^{-1} H)^{-1} H^T C_{VV}^{-1} Y$  with  $C_{\widehat{\theta}\widehat{\theta}} = (H^T C_{VV}^{-1} H)^{-1}$
- If  $V \sim \mathcal{N}(0, C_{VV})$  then  $\widehat{\theta}_{BLUE} = \widehat{\theta}_{ML} = \text{efficient} \Rightarrow \widehat{\theta}_{UMVUE}$

## Method of Moments

Principle:

- $m$  unknown parameters  $\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}$
- $f(Y|\theta)$  depends on  $\theta \Rightarrow$  its moments also
- take  $m$  moments  $\mu = g(\theta) = \begin{bmatrix} g_1(\theta) \\ \vdots \\ g_m(\theta) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}$

such that  $g(\cdot)$  is invertible, i.e.  $\theta = g^{-1}(\mu)$  : can determine  $\theta$  from  $\mu$ .

- estimate the moments:  $\hat{\mu}$  (e.g. sample moments)
- method of moments:  $\hat{\theta}_{MM} = g^{-1}(\hat{\mu})$

## Method of Moments: Example 1

- $y_i, i = 1, \dots, n$  i.i.d.,  $f(y|\theta)$  mixture distribution,  $\theta$  mixture parameter

$$f(y|\theta) = (1-\theta)\phi_1(y) + \theta\phi_2(y), \quad \phi_k(y) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{y^2}{2\sigma_k^2}}, k = 1, 2$$

- $\mu = E(y^2|\theta) = (1-\theta)\sigma_1^2 + \theta\sigma_2^2 = g(\theta) \Rightarrow \theta = g^{-1}(\mu) = \frac{\mu - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}$
- $\hat{\theta}_{MM} = g^{-1}(\hat{\mu}) = \frac{\hat{\mu} - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}, \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i^2$  sample mean squared value
- bias:  $E\hat{\theta} = \frac{1}{\sigma_2^2 - \sigma_1^2} E\hat{\mu} - \frac{\sigma_1^2}{\sigma_2^2 - \sigma_1^2} = \frac{1}{\sigma_2^2 - \sigma_1^2} \mu - \frac{\sigma_1^2}{\sigma_2^2 - \sigma_1^2} = \theta$  : unbiased



## Method of Moments: Example 1 (cont'd)

- $$\begin{aligned} \text{Var}(\hat{\theta}) &= \text{Var}\left(\frac{1}{\sigma_2^2 - \sigma_1^2} \hat{\mu} - \frac{\sigma_1^2}{\sigma_2^2 - \sigma_1^2}\right) = \frac{1}{(\sigma_2^2 - \sigma_1^2)^2} \text{Var}(\hat{\mu}) = \frac{1}{(\sigma_2^2 - \sigma_1^2)^2} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n y_i^2\right) \\ &= \frac{1}{(\sigma_2^2 - \sigma_1^2)^2} \sum_{i=1}^n \text{Var}\left(\frac{1}{n} y_i^2\right) = \frac{1}{(\sigma_2^2 - \sigma_1^2)^2} \sum_{i=1}^n \frac{1}{n^2} \text{Var}(y_i^2) = \frac{1}{(\sigma_2^2 - \sigma_1^2)^2} \frac{1}{n} \text{Var}(y^2) \end{aligned}$$
- $$\begin{aligned} f(y|\theta) &= (1-\theta) \phi_1(y) + \theta \phi_2(y) \\ \text{Var}(y^2) &= Ey^4 - (Ey^2)^2, \quad Ey^2 = (1-\theta) \sigma_1^2 + \theta \sigma_2^2 \\ Ey^4 &= (1-\theta) 3\sigma_1^4 + \theta 3\sigma_2^4 \end{aligned}$$
- $$\Rightarrow \text{Var}(\hat{\theta}_{MM}) = \frac{3(1-\theta)\sigma_1^4 + 3\theta\sigma_2^4 - [(1-\theta)\sigma_1^2 + \theta\sigma_2^2]^2}{n(\sigma_1^2 - \sigma_2^2)^2} \xrightarrow{n \rightarrow \infty} 0$$

$\Rightarrow \hat{\theta}_{MM} = \text{mean-square consistent}$

## MM Example 2: Sinusoid in White Noise

- $y_k = s_k + v_k = A \cos(\omega k + \phi) + v_k, \quad k = 1, \dots, n$
- $$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad S = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}, \quad V = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \quad \theta = \begin{bmatrix} A \\ \omega \\ \sigma_v^2 \end{bmatrix}, \quad \Theta : A > 0, \omega \in [0, \pi], \sigma_v^2 > 0$$
- distributions:  $\phi \sim \mathcal{U}[0, 2\pi]$  independent of  $\theta, V$ ;  $EV = 0, EVV^T = \sigma_v^2 I_n$

randomness:  $f(Y, \phi|\theta) = f(\phi|\theta) f(Y|\theta, \phi) = f(\phi) f_{V|\sigma_v^2}(Y - S(A, \omega, \phi) | \sigma_v^2)$

in what follows: only first and second moments of  $V$  needed
- mean:  $E_{Y, \phi|\theta} y_k = AE \cos(\omega k + \phi) + Ev_k = 0$

covariance sequence:

$$\begin{aligned} r_{yy}(i) &= Ey_k y_{k+i} = A^2 E \cos(\omega k + \phi) \cos(\omega k + \phi + \omega i) \\ &\quad + AE \cos(\omega k + \phi) Ev_{k+i} + AE \cos(\omega k + \phi + \omega i) Ev_k + Ev_k v_{k+i} \\ &= \frac{A^2}{2} E \cos(2\omega k + 2\phi + \omega i) + \frac{A^2}{2} E \cos(\omega i) + \sigma_v^2 \delta_{i0} \\ &= \frac{A^2}{2} \cos(\omega i) + \sigma_v^2 \delta_{i0} \end{aligned}$$

## MM Example 2: Sinusoid in White Noise (2)

- moments:  $\mu = \begin{bmatrix} r_{yy}(0) \\ r_{yy}(1) \\ r_{yy}(2) \end{bmatrix} = \begin{bmatrix} \frac{A^2}{2} + \sigma_v^2 \\ \frac{A^2}{2} \cos(\omega) \\ \frac{A^2}{2} \cos(2\omega) \end{bmatrix} = g(\theta)$
- $\theta = g^{-1}(\mu)$ :
$$\omega = \begin{cases} \arccos\left(\frac{r_{yy}(2) + \sqrt{r_{yy}^2(2) + 8r_{yy}^2(1)}}{4r_{yy}(1)}\right) & , r_{yy}(1) \neq 0 \\ \frac{\pi}{2} & , r_{yy}(1) = 0 \end{cases}$$
$$A = \begin{cases} \sqrt{\frac{2r_{yy}(1)}{\cos(\omega)}} & , r_{yy}(1) \neq 0 \\ \sqrt{-2r_{yy}(2)} & , r_{yy}(1) = 0 \end{cases}, \quad \sigma_v^2 = r_{yy}(0) - \frac{A^2}{2}$$
- sample moments  $\hat{\mu}$ :  $\hat{r}_{yy}(i) = \frac{1}{n} \sum_{k=1}^{n-i} y_k y_{k+i}$ ,  $i = 0, 1, 2$

## Method of Moments: Properties

- $\hat{\mu}$  easy to compute,  $\hat{\theta}_{MM} = g^{-1}(\hat{\mu})$  straightforward if  $\mu$  chosen well, hence  $\hat{\theta}_{MM}$  easy to determine and easy to implement
- no optimality properties but usually consistent (since  $\hat{\mu}$  consistent)
- if performance of  $\hat{\theta}_{MM}$  not satisfactory, can use  $\hat{\theta}_{MM}$  as initialization in an iterative optimization procedure that finds  $\hat{\theta}_{ML}$