

## Lecture 8

Topics:

- I) Numerical techniques for  $E(T|d)$
- II) sky cuts, Knox approximation.
- III) Sherman-Morrison-Woodbury
- IV) message passing
- V) Flat sky & Circulant embedding.

Computing the weiner filter  $E(s(\hat{n})|d)$  and generating conditional simulations  $s^*(\hat{n}) \sim P(s|d)$  are easy when one has all sky data of the form

$$d(\hat{n}) = s(\hat{n}) + n(\hat{n}), \quad \forall \hat{n} \in S^2$$

where both  $s$  &  $n$  are isotropic Gaussian random fields.

In particular

$$E(s_{\text{em}}|d) = \frac{C_e^{ss}}{C_e^{ss} + C_e^{nn}} d_{\text{em}}$$

$$s_{\text{em}}^* = \frac{C_e^{ss}}{C_e^{ss} + C_e^{nn}} d_{\text{em}} + \frac{C_e^{ss} C_e^{nn}}{C_e^{ss} + C_e^{nn}} w_{\text{em}}$$

↑  
unit var  
white noise

The main problem with this idealized model is that the Galaxy masks some directions so our data is more accurately modeled.

$$d(\hat{n}) = s(\hat{n}) + n(\hat{n}), \quad \forall \hat{n} \in \mathcal{D} \subsetneq S^2$$

(1) Since  $\mathcal{D} \subsetneq S^2$  we can no longer take the spherical harmonic transform.

(2)

$$\therefore E(s|d) = \text{generalized Wiener filter} \\ = \sum (\Sigma + N)^{-1} d$$

### Quick tip

In Julia instead of writing  $\text{inv}(\Sigma + N)^{-1} d$  use  $(\Sigma + N)^{-1} d$  which is more stable & avoids actually storing the matrix  $\text{inv}(\Sigma + N)$

### Knox approximation

For small sky surveys one can still generate a "signal + beam + white noise" model of the data that approximates the amount of information present in the data useful for studying statistical properties & projections of possible constraints.

Suppose  $d(\hat{n}) = s(\hat{n}) + n(\hat{n})$

where  $C_e^{nn} = \sigma^2 \exp\left(\frac{\theta^2}{8 \log^2 e} e(e+1)\right)$   
but only observed at  $\hat{n} \in \mathcal{D} \subset S^2$

$$\text{Let } f_{\text{sky}} := \frac{\text{area}(\mathcal{S})}{\text{area}(S^2)} \quad (3)$$

Then an idealized model for the info in  $d$  for  $s$  is

$$d'(\hat{n}) = s(\hat{n}) + r'(\hat{n})$$

where  $d'$  is observed on all  $\hat{n} \in S^2$

but  $C_s^{n'n'} = \frac{\delta^2}{f_{\text{sky}}} \exp\left(\frac{-b^2}{8\log 2} \ell(\ell_{\text{eff}})\right)$

e.g. if  $f_{\text{sky}} = \frac{1}{2}$  then

$$C_s^{n'n'} = 2 \cdot C_s^{nn}$$

which is what one would get with "half" the number of indep. obs.

Note: this is only an approximation since masking effectively correlates Sers.

$$d(\hat{n}) = s(\hat{n}) + n(\hat{n}), \hat{n} \in \mathcal{D}$$

is equiv to

$$d(\hat{n}) = I_{\mathcal{D}}(\hat{n}) \cdot s(\hat{n}) + n(\hat{n}), \hat{n} \in S^2$$

is "equiv" to convolution in the spectral domain

$$d_{\text{obs}} = \tilde{x} * \text{Sers.} + n_{\text{obs}}$$

Basis form of Generalized Wiener Filters (4)

Suppose

$$C^s(|\hat{n}_1 - \hat{n}_2|) = \text{cov}(s(\hat{n}_1), s(\hat{n}_2))$$

and  $d(\hat{n}) = s(\hat{n}) + n(\hat{n})$  is observed on  $\hat{n}_1, \hat{n}_2, \dots, \hat{n}_m \in S^2$

Notice that  $\forall \hat{n} \in S^2$

$$E(s(\hat{n}) | d) = \sum (\mathbb{I} + N)^{-1} d$$

single pixel value  
data vector

$$\text{where } \sum = (\dots, \underbrace{\text{cov}(s(\hat{n}), d(\hat{n}_i)), \dots}_{C^s(|\hat{n} - \hat{n}_i|)})$$

$$C^s(|\hat{n} - \hat{n}_i|)$$

$$\therefore E(s(\hat{n}) | d) = \sum_{i=1}^m \beta_i C^s(|\hat{n} - \hat{n}_i|)$$

where  $(\beta_1, \dots, \beta_m) = (\mathbb{I} + N)^{-1} d$  need to be computed just once for Wiener filtering at all  $\hat{n} \in S^2$

Sherman-Morrison-Woodbury formula (5)  
for low rank models with indep Noise

Suppose  $d = s + n$ ,  $s \sim N(0, \Sigma) \xrightarrow{\text{indep.}}$   
 $n \sim N(0, N)$

where  $N$  is diag &  $\Sigma$  is low rank in the  
sense that  $\Sigma = \begin{bmatrix} u & | & A & | & u^T \end{bmatrix}$ .

$\therefore s \sim \sum_{i=1}^m z_i \sqrt{\lambda_i} u_i$   
 $z_1, \dots, z_m$  iid  $N(0, 1)$ .  $u \in \mathbb{R}^{n \times m}$  diag matrix  
with  $\lambda \in \mathbb{R}^{m \times m}$

SMW formula:

$$(A + V D U^T)^{-1} = A^{-1} - A^{-1} V \left( D^{-1} + U^T A^{-1} V \right)^{-1} U^T A^{-1}$$

$A \in \mathbb{R}^{n \times n}$

$D \in \mathbb{R}^{m \times m}$

If  $m \ll n$  this  
is a low rank  
update of  $A$

This is an  
 $m \times m$  matrix.

$$\therefore (N + U \Lambda U^T)^{-1}$$

$$= N^{-1} - N^{-1} U \left( \Lambda^{-1} + U^T N^{-1} U \right)^{-1} U^T N^{-1}$$

diag matrix Rank  $m$

$\therefore$  easy to compute.

HODLR matrices (6)

This is a new technique for computing  
the inverse (& determinant) of a  
matrix  $A$  (which can be used  
for generalized wiener filtering  $\Sigma(\Sigma + N)^{-1} d$ )  
start by dividing  $A \in \mathbb{R}^n \times \mathbb{R}^n$  into  $2^m$  block

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_{2^m} \end{pmatrix}$$

where the diag blocks are  
small enough so computation of  
 $A_i^{-1}$  is easy.

Step 0:

$$A^{(1)} = \begin{pmatrix} A_1^{-1} & & & \\ & A_2^{-1} & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} A = \begin{pmatrix} I & I & * & \\ * & I & * & \\ & * & \ddots & \end{pmatrix}$$

Step 1: Pair up consecutive  $I$  blocks  
& form new blocks twice the size.

$$A^{(1)} = \begin{pmatrix} I & * & & & & \\ * & I & & & & \\ & & I & * & & \\ & & * & I & & \\ & & & & I & * \\ & & & & * & I \end{pmatrix}$$

Take each  $A_1^{(1)}, A_2^{(1)}, \dots$  & write ⑦

$$A_i^{(1)} = \begin{pmatrix} I & * \\ * & I \end{pmatrix} \approx \begin{pmatrix} I & VDU^T \\ UDV^T & I \end{pmatrix}$$

where  $D$  is a small matrix, i.e.  $VDU^T$  is low rank.

Now

$$\begin{aligned} A_i^{(1)} &= \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + \underbrace{\begin{pmatrix} V & 0 \\ 0 & u \end{pmatrix}}_{\text{rank 1}} \underbrace{\begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}}_{\text{rank 1}} \underbrace{\begin{pmatrix} 0 & u^T \\ v^T & 0 \end{pmatrix}}_{\text{rank 1}} \\ &= I + \tilde{V} \tilde{D} \tilde{u}^T \end{aligned}$$

$$\text{since } \begin{pmatrix} V & 0 \\ 0 & u \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} 0 & u^T \\ v^T & 0 \end{pmatrix}$$

$$= \begin{pmatrix} V & 0 \\ 0 & u \end{pmatrix} \begin{pmatrix} 0 & Du^T \\ Dv^T & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & VDU^T \\ UDV^T & 0 \end{pmatrix}$$

Now apply Sherman-Morrison-Woodbury

$$\begin{aligned} A_i^{(1)-1} &= (I + \tilde{V} \tilde{D} \tilde{u}^T)^{-1} \\ &= I - \tilde{V} \left( \tilde{D}^{-1} + \tilde{V}^T \tilde{u} \right)^{-1} \tilde{u}^T \end{aligned}$$

*These are low rank blocks.*

$$\therefore A^{(2)} := \begin{pmatrix} A_1^{(1)-1} & & & 0 \\ & A_2^{(1)-1} & & \\ 0 & & \ddots & \\ & & & \ddots \end{pmatrix} A^{(1)} \quad ⑧$$

$$= \begin{pmatrix} I & * & & * \\ * & I & * & \\ * & * & \ddots & \\ & & & \ddots \end{pmatrix}$$

Now set  $A^{(1)} \leftarrow A^{(2)}$  &  
Repeat step 1 & stop  
when you've got  $I$  as  $A^{(2)}$ .  
For the determinant Note Sylvester's theorem

$$\det(I_n + AB) = \det(I_m + \underbrace{BA}_{m \times m})$$

so the determinants of all the diag blocks  $A_1^{(1)}, A_2^{(1)}, \dots$  is easy.

### Message Passing

Developed recently by Wendelt et al... works in the situation where obs are of the form

$$d(\hat{n}) = s(\hat{n}) + n(\hat{n}), \quad \forall \hat{n} \in \mathcal{S} \subseteq S^2$$

s.t.  $s$  is an isotropic Gaussian random

field &

$$E(n(\hat{n}_1) n(\hat{n}_2)) = \underbrace{\sigma^2(\hat{n}_1)}_{\text{independent across pixels}} \delta(\hat{n}_1 - \hat{n}_2)$$

non-stationary variance

Extend  $d(\hat{n}) = s(\hat{n}) + n(\hat{n})$  to all  $\hat{n} \in S^2$

by setting  $\sigma^2(\hat{n}) = \infty$  on  $\hat{n} \in S^2 \setminus \mathcal{S}$ .

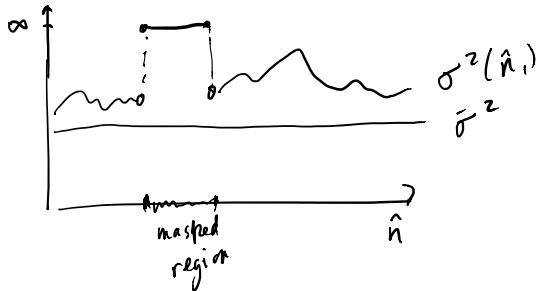
Now split the noise field  $n(\hat{n})$  into a homogeneous noise term & a non-stationary term  $n(\hat{n}) = \bar{n}(\hat{n}) + \tilde{n}(\hat{n})$  (a)

where  $\bar{n}$  &  $\tilde{n}$  are independent,

$$E(\bar{n}(\hat{n}_1)\bar{n}(\hat{n}_2)) = \bar{\sigma}^2 \delta(\hat{n}_1 - \hat{n}_2),$$

$$E(\tilde{n}(\hat{n}_1)\tilde{n}(\hat{n}_2)) = \tilde{\sigma}^2(\hat{n}_1) \delta(\hat{n}_1 - \hat{n}_2)$$

$$\therefore \bar{\sigma}^2 + \tilde{\sigma}^2(\hat{n}) = \sigma^2(\hat{n})$$



Define a "messenger field"  $\bar{s}(\hat{n})$  as follows

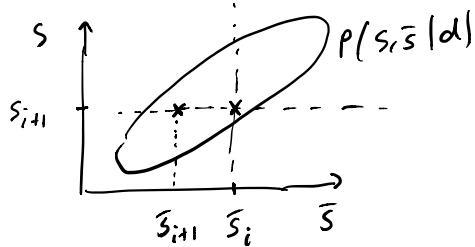
$$\begin{aligned} d(\hat{n}) &= s(\hat{n}) + n(\hat{n}) \\ &= s(\hat{n}) + \underbrace{\bar{n}(\hat{n}) + \tilde{n}(\hat{n})}_{=: \bar{s}(\hat{n})} \end{aligned}$$

Now consider both  $s, \bar{s}$  as unknowns. The posterior  $P(s, \bar{s} | d)$  is Gaussian and is the stationary distribution of a Gibbs Markov Chain:

for  $i = 1, 2, \dots$

Step  $i+1$ :  $s_{i+1} \sim P(s | \bar{s}_i, d)$

$\bar{s}_{i+1} \sim P(\bar{s} | s_{i+1}, d)$



Instead of sampling the conditional the algorithm simply alternates Wiener filters: (10)

$$\text{step } i+1: s_{i+1} = E(s | \bar{s}_i, d)$$

$$\bar{s}_{i+1} = E(\bar{s} | s_{i+1}, d)$$

The key is that each Wiener filter can be "diagonalized". Indeed

$$\begin{aligned} (s_{i+1})_{em} &= E(s_{em} | \bar{s}_i, d) \\ &= E(s_{em} | \bar{s}_i) \\ &= \frac{C_e^{ss}}{C_e^{ss} + C_e^{\bar{n}\bar{n}}} (\bar{s}_i)_{em} \end{aligned}$$

Also  $\bar{s}_{i+1} = E(\bar{s} | s_{i+1}, d)$  can be derived as a solution to

$$\begin{aligned} \nabla_{\bar{s}} \log P(s_{i+1}, \bar{s}, d) &= 0 \\ &= \nabla_{\bar{s}} \left[ \log P(d | \bar{s}, s_{i+1}) + \log P(\bar{s} | s_{i+1}) + \log P(s_{i+1}) \right] \\ &= \nabla_{\bar{s}} \left[ - \frac{(d - \bar{s})^T \tilde{N}^{-1} (d - \bar{s})}{2} - \frac{(\bar{s} - s_{i+1})^T \bar{N}^{-1} (\bar{s} - s_{i+1})}{2} \right] \end{aligned}$$

where  $\tilde{N} = E(\tilde{n} \tilde{n}^T) = \text{diag}$

$\bar{N} = E(\bar{n} \bar{n}^T)$

$$= + \tilde{N}^{-1} (d - \bar{s}) - \bar{N}^{-1} (\bar{s} - s_{i+1})$$

$$\therefore \bar{s}_{i+1} = (\tilde{N}^{-1} + \bar{N}^{-1})^{-1} [\tilde{N}^{-1} d + \bar{N}^{-1} s_{i+1}]$$

$\nwarrow \swarrow$  diagonal matrices

## Messenger Algorithm 1:

For  $i=1, 2, \dots$

step  $i+1$ :

$$(s_{i+1})_{em} = \frac{C_e^{ss}}{C_e^{ss} + C_e^{\bar{N}\bar{N}}} (\bar{s}_i)_{em}$$

$$\bar{s}_{i+1} = (\tilde{N}^{-1} + \bar{N}^{-1})^{-1} [\tilde{N}^{-1}d + \bar{N}^{-1}s_{i+1}]$$

To analyze the algorithm write the iterations

$$s_{i+1} = (\Sigma + \lambda\bar{N})^{-1} \bar{s}_i$$

$$\bar{s}_{i+1} = (\tilde{N}^{-1} + (\lambda\bar{N})^{-1})^{-1} [\tilde{N}^{-1}d + (\lambda\bar{N})^{-1}s_{i+1}]$$

where  $\lambda=1$ . A fixed point satisfies

$$s = (\Sigma^{-1} + \bar{N}^{-1})^{-1} \bar{N}^{-1} \bar{s}$$

$$\bar{s} = (\tilde{N}^{-1} + \bar{N}^{-1})^{-1} [\tilde{N}^{-1}d + \bar{N}^{-1}s]$$

$$\therefore (\Sigma^{-1} + \bar{N}^{-1})s = \bar{N}^{-1}\bar{s}$$

$$(\tilde{N}^{-1} + \bar{N}^{-1})\bar{s} = \tilde{N}^{-1}d + \bar{N}^{-1}s$$

$$\therefore (\tilde{N}^{-1} + \bar{N}^{-1})\bar{N}(\Sigma^{-1} + \bar{N}^{-1})s = \tilde{N}^{-1}d + \bar{N}^{-1}s$$

$$\therefore (\tilde{N}^{-1}\bar{N} + I)(\Sigma^{-1} + \bar{N}^{-1})s = \tilde{N}^{-1}d + \bar{N}^{-1}s$$

$$\therefore (\tilde{N}^{-1}\bar{N}\Sigma^{-1} + \Sigma^{-1} + \bar{N}^{-1} + \bar{N}^{-1})s = \tilde{N}^{-1}d + \bar{N}^{-1}s$$

$$\therefore (\bar{N}\Sigma^{-1} + \tilde{N}\Sigma^{-1} + I)s = d$$

$$\therefore (\bar{N} + \tilde{N} + \Sigma)\Sigma^{-1}s = d$$

$$\therefore s = \Sigma(\bar{N} + \tilde{N} + \Sigma)^{-1}d$$

↑  
from  $\text{Nilson}$   
 $\text{filter}$ .

(11)

In a similar way the residuals satisfy

$$\varepsilon_{i+1} = \left[ \sum (\Sigma + \lambda\bar{N})^{-1} \right] [(\lambda\bar{N})(\tilde{N} + \lambda\bar{N})^{-1}] \varepsilon_i$$

Since the spectral radius ( $\rho(A) = \max_{\text{eigenvalue}} |\text{eigenvalue}|$ ) of a product of matrices satisfies  $\rho(AB) \leq \rho(A)\rho(B)$

We have

$$\rho \left( \left[ \sum (\Sigma + \lambda\bar{N})^{-1} \right] [(\lambda\bar{N})(\tilde{N} + \lambda\bar{N})^{-1}] \right) < 1$$

$\therefore$  The iterations are geometrically convergent.

Note: Slower convergence rate when  $\Sigma$  is large &  $\lambda\bar{N}$  is small

$\therefore$  to speed up the convergence choose a "cooling schedule"  
 $\lambda_1 > \lambda_2 > \dots \rightarrow 1$ .

## Flat sky Approximation

Ref Wayne Hu's paper  
 "Weak lensing of the CMB: A harmonic Approach" (13)

Goal: For  $\hat{n} \in S^2$  near the pole find a GRF  $\{S^{FS}(x) : x \in \mathbb{R}^2\}$

s.t.  $S^{FS}(x) \approx S(\underline{\theta}, \underline{\varphi})$ , where  $x = (\theta \cos \varphi, \theta \sin \varphi)$ ,  $\theta \approx 0$ .  
 Polar coords

This will allow us to study statistical problems on  $\mathbb{R}^2$  & relate them (by analogy) to CMB problems on  $S^2$ .

For  $\ell = 0, 1, 2, \dots$  &  $\varphi \in [0, 2\pi)$  define

$$\xi_\ell(\varphi) := \sum_{n=-\ell}^{\ell} \frac{\sin \sqrt{\ell}}{i^n \sqrt{2\pi}} e^{in\varphi}$$

$$S^{FS}(x) := \sum_{\ell=0}^{\infty} \int_0^{2\pi} \xi_\ell(\varphi) \exp(i\varphi \cdot (\ell \cos \varphi, \ell \sin \varphi)) \frac{d\varphi}{2\pi}$$

Notice that  $\forall m \in \mathbb{Z}$

$$\begin{aligned} \int_0^{2\pi} e^{-im\varphi} \xi_\ell(\varphi) \frac{d\varphi}{2\pi} &= \sum_{n=-\ell}^{\ell} \frac{\sin \sqrt{\ell}}{i^n \sqrt{2\pi}} \underbrace{\int_0^{2\pi} e^{-im\varphi} e^{in\varphi} \frac{d\varphi}{2\pi}}_{= \delta(m-n)} \\ &= \frac{\sin \sqrt{\ell}}{i^m \sqrt{2\pi}} I_{\{|m| \leq \ell\}} \end{aligned}$$

Now suppose  $x = (\theta_x \cos \varphi_x, \theta_x \sin \varphi_x)$ .

(14)

$$\begin{aligned} S^{FS}(x) &= \sum_{l=0}^{\infty} \int_0^{2\pi} \xi_l(\varphi) \underbrace{\exp(ix \cdot (l \cos \varphi, l \sin \varphi))}_{\downarrow} \frac{d\varphi}{2\pi} \\ &= \sum_{l=0}^{\infty} \int_0^{2\pi} \xi_l(\varphi) \left[ \sum_{m \in \mathbb{Z}} i^m J_m(l \theta_x) e^{im(\varphi_x - \varphi)} \right] \frac{d\varphi}{2\pi} \\ &\stackrel{\text{Jacobi-Arnold expansion}}{=} \sum_{l=0}^{\infty} \sum_{m \in \mathbb{Z}} i^m J_m(l \theta_x) e^{im\varphi_x} \underbrace{\int_0^{2\pi} \xi_l(\varphi) e^{-im\varphi} \frac{d\varphi}{2\pi}}_{= \frac{S_{lm} \sqrt{l}}{i^m \sqrt{2\pi}} I_{\{\mid m \mid \leq l\}}} \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{S_{lm} \sqrt{l}}{\sqrt{2\pi}} J_m(l \theta_x) e^{im\varphi_x} \\ &\approx \frac{\sqrt{2\pi}}{\sqrt{l}} Y_{lm}(\theta_x, \varphi_x) \end{aligned}$$

for  $\theta_x \approx 0$  by. Gradshteyn &

$$\approx \sum_{lm} S_{lm} Y_{lm}(\theta_x, \varphi_x) \quad \text{Ryzhik eq. 8.722}$$

as was to be shown.

Notice also that

$$\begin{aligned} S_k^{FS} &= \int_{\mathbb{R}^2} \frac{dx}{2\pi} e^{-ik \cdot x} S^{FS}(x) \\ &= \sum_{l=0}^{\infty} \int_0^{2\pi} \frac{d\varphi}{2\pi} \int_{\mathbb{R}^2} \frac{dx}{2\pi} e^{-ik \cdot x} \xi_l(\varphi) \exp(ix \cdot (l \cos \varphi, l \sin \varphi)) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{l=0}^{\infty} \int_0^{2\pi} \frac{d\varphi}{2\pi} \sum_l e_l(\varphi) \int_{k - \left( \begin{array}{c} l \cos \varphi \\ l \sin \varphi \end{array} \right)}^{(2\pi)} \\
 &= \begin{cases} \frac{d\varphi}{dk} \sum_l e_l(\varphi) & \text{when } k = (l \cos \varphi, l \sin \varphi) \text{ & } l \in \mathbb{Z}^+ \\ 0 & \text{o.w.} \end{cases}
 \end{aligned}$$

$$\text{Note } \frac{d\psi}{dk} = \frac{d\phi}{\ell d\theta d\phi} = \frac{1}{\ell d\theta} \approx \frac{1}{\ell}$$

$$\therefore E\left(S_k^{FS} \overline{S_{k'}^{FS}}\right) = E\left[\left(\frac{dy}{dk}\right)^2 \sum_{n=-l}^l \sum_{n'=-l'}^{l'} \frac{\sin \sqrt{\ell}}{i^n \sqrt{2\pi}} e^{inY} \frac{\sin \sqrt{\ell'}}{i^{n'} \sqrt{2\pi}} e^{in'Y'}\right]$$

$$= \left( \frac{d\psi}{dk} \right)^2 \sum_{n=-e}^e \sum_{n'=-e'}^{e'} \frac{\sqrt{e} \sqrt{e'}}{2\pi} E(s_n s_{e'n'}) e^{in\varphi - in'\varphi'} \underbrace{\left[ s_n, s_{n'}, C_\ell^{ss} \right]}_0$$

$$= \left( \frac{d\psi}{dk} \right)^2 \sum_{n=-\ell}^{\ell} \frac{\ell!}{2\pi} \delta_{\ell n'} \left[ {}_e^{ss} e^{in(\varphi - \varphi')} \right]$$

$$= \left( \frac{dy}{dk} \right)^2 e^{\ell} \text{See, } C_e^{ss} \sum_{n=-\ell}^{\ell} e^{\frac{in(\varphi - \varphi')}{2\pi}}$$

$$= I_{k=k'} \left( \frac{1}{e} \right)^2 \frac{d}{dq} C_e^{ss} \approx \delta_{q-q'}$$

$$\approx \frac{I_{k=p'}}{\ell d \ell dk} C_e^{ss} \approx S_{k=p'} C_e^{ss}$$

This "shows" we can use  $C^{ss}$  to approximate the 2-d spectral density of  $S^{FS}$ .

Circulant embedding.