Lecture 8

(1)

Topics:

I) Numerical techniques for $E(T|d)$
II) sky cuts, knox approximation.
III) Sherman-Moorison-Woodbury
IV) message passing
V) Conjugate Gradient
VI) 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

$$E(s_{\ell m}|d) = \frac{C_\ell^{ss}}{C_\ell^{ss} + C_\ell^{nn}} d_{\ell m}$$

$$s^*_{\ell m} = \frac{C_\ell^{ss}}{C_\ell^{ss}+C_\ell^{nn}} d_{\ell m} + \frac{\sqrt{C_\ell^{ss} C_\ell^{nn}}}{\sqrt{C_\ell^{ss}+C_\ell^{nn}}} w_{\ell m}$$

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.

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

(2)

Since $\Omega \subsetneq S^2$ we can no longer take the spherical Harmonic transform.

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

Quick tip

In Julia instead of writing `inv(Σ+N)*d` use `(Σ+N)\d` which is more stable & avoids actually storing the matrix `inv(Σ+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 $C_\ell^{nn} = \delta^2 \exp\left(\frac{b^2}{8 \log 2} \ell(\ell+1)\right)$

but only observed at $\hat{n} \in \Omega \subset S^2$

Let $f_{sky} := \dfrac{area(\Omega)}{area(S^2)}$.

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

$$d'(\hat n) = s(\hat n) + n'(\hat n)$$

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

but $C_\ell^{n'n'} = \dfrac{\sigma^2}{f_{sky}} \exp\left(\dfrac{b^2}{8\log 2}\ell(\ell+1)\right)$

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

$$C_\ell^{n'n'} = 2 \cdot C_\ell^{nn}$$

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

Note: this is only an approximation since masking effectively correlates $s_{\ell m}$'s.

$$d(\hat n) = s(\hat n) + n(\hat n), \quad \hat n \in \Omega$$

is equiv to

$$d(\hat n) = I_\Omega(\hat n) \cdot s(\hat n) + n(\hat n), \quad \hat n \in S^2$$

is "equiv" to convolution in the spectral domain

$$d_{\ell m} = \text{"}\tilde{I} * s_{\ell m}\text{'s"} + n_{\ell m}$$

Basis form of Generalized Wiener Filters

Suppose

$$C^s(|\hat n_1 - \hat n_2|) = cov(s(\hat n_1), s(\hat n_2))$$

and $d(\hat n) = s(\hat n) + n(\hat n)$ is observed on $\hat n_1, \hat n_2, \ldots, \hat n_m \in S^2$

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

$$E(s(\hat n) \mid d) = \Sigma (\Sigma + N)^{-1} d$$

single pixel value

data vector

where $\Sigma = (\cdots, \underbrace{cov(s(\hat n), d(\hat n_i))}_{C^s(|\hat n - \hat n_i|)}, \ldots)$

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

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

Sherman-Morrison-Woodbury formula for low Rank Models with indep Noise

Suppose $d = s + n$, $\left.\begin{array}{l} s \sim N(0,\Sigma) \\ n \sim N(0,N) \end{array}\right\}$ indep.

where N is diag & Σ is low rank in the sense that $\Sigma = \underbrace{\boxed{U}}_{U \in \mathbb{R}^{n\times m}} \; \underbrace{\boxed{\Lambda}}_{\Lambda \in \mathbb{R}^{m\times m}} \; \boxed{U^T}$.

$\Lambda \in \mathbb{R}^{m\times m}$ diag matrix with diag > 0

$U \in \mathbb{R}^{n\times m}$

$\therefore \; s \sim \sum_{i=1}^{m} z_i \sqrt{\lambda_i}\, u_i$

$z_1, \dots, z_m$ iid $N(0,1)$.

SMW formula:

$$\underbrace{(A + VDU^T)^{-1}} = A^{-1} - A^{-1}V\underbrace{\left(D^{-1} + U^T A^{-1} V\right)^{-1}}_{\text{This is an } m\times m \text{ matrix.}} U^T A^{-1}$$

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

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

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

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

$$= N^{-1}d - N^{-1}U\big(\Lambda^{-1} + \underbrace{U^T N^{-1} U}_{\text{Rank } m}\big)^{-1} U^T N^{-1} d$$

diag matrix

$\therefore$ easy to compute.

HODLR matrices

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

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

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

Step 0:

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

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

$$A^{(1)} = \begin{pmatrix} \underbrace{\begin{matrix} I & * \\ * & I \end{matrix}}_{A_1^{(1)}} & * & * \\ * & \underbrace{\begin{matrix} I & * \\ * & I \end{matrix}}_{A_2^{(1)}} & * \\ * & * & \ddots \end{pmatrix}$$

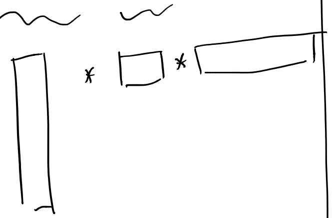
Take each $A_1^{(1)}, A_2^{(1)}, \dots$ & write

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

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

Now

$$A_i^{(1)} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + \begin{pmatrix} V & 0 \\ 0 & U \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}\begin{pmatrix} 0 & U^T \\ V^T & 0 \end{pmatrix}$$



$$= I + \tilde{V}\tilde{D}\tilde{U}^T$$

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

$$A_i^{(1)^{-1}} = (I + \tilde{V}\tilde{D}\tilde{U}^T)^{-1}$$

$$= I - \underbrace{\tilde{V}(\tilde{D}^{-1} + \tilde{V}^T\tilde{U})^{-1}\tilde{U}^T}_{\text{These are low rank themselves.}}$$

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

$$= \left(\begin{array}{c|c|c} I & * & * \\ \hline * & I & * \\ \hline * & * & \ddots \end{array}\right)$$

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 thm

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

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

Message Passing

Developed recently by Wandelt et al...

works in the situation where obs are of the form

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

s.t. s is an isotropic Gaussian random field &

$$E(n(\hat{n}_1)\, n(\hat{n}_2)) = \underbrace{\sigma^2(\hat{n}_1)\,\delta(\hat{n}_1 - \hat{n}_2)}_{\text{independent accross pixel but with non-stationary variane}}$$

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 $\hat{n} \in S^2 \setminus \Omega$.

Now split the noise field $n(\hat{n})$ into a homogenious noise term & a non-stationary term

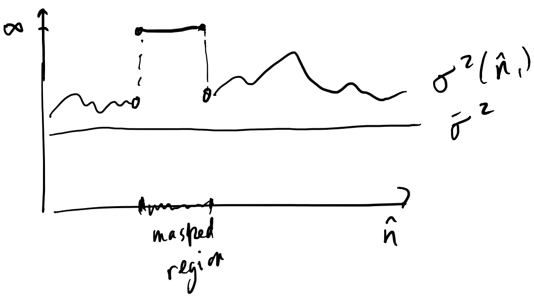
$$n(\hat{n}) = \bar{n}(\hat{n}) + \tilde{n}(\hat{n}) \qquad (9)$$

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

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

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

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



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

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

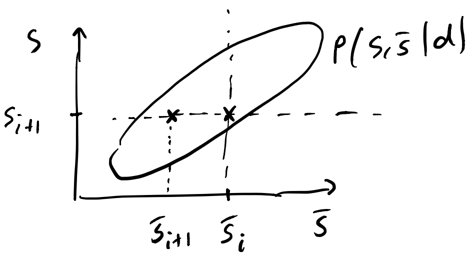
$$= \underbrace{s(\hat{n}) + \bar{n}(\hat{n})}_{=:\ \bar{s}(\hat{n})} + \tilde{n}(\hat{n})$$

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

for $i = 1, 2, \ldots$

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

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



(10)

Instead of sampling the conditional the algorithm simply alternates wiener filters:

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

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

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

$$(s_{i+1})_{\ell m} = E(s_{\ell m} \mid \bar{s}_i, d)$$

$$= E(s_{\ell m} \mid \bar{s}_i)$$

$$= \frac{C_\ell^{ss}}{C_\ell^{ss} + C_\ell^{\bar{n}\bar{n}}} (\bar{s}_i)_{\ell m}$$

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

$$\nabla_{\bar{s}} \log P(s_{i+1}, \bar{s}, d) = 0$$

$$= \nabla_{\bar{s}} \left[\log P(d \mid \bar{s}, s_{i+1}) + \log P(\bar{s} \mid s_{i+1}) + \cancel{\log P(s_{i+1})} \right]$$

$$= \nabla_{\bar{s}} \left[-\frac{(d - \bar{s})^T \tilde{N}^{-1} (d - \bar{s})}{2} - \frac{(\bar{s} - s_{i+1})^T \bar{N}^{-1} (\bar{s} - s_{i+1})}{2} \right]$$

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

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

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

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

diagonal matrices

Messanger Algorithm 1:

For $i=1,2,\dots$

step $i+1$:

$$(s_{i+1})_{\ell m} = \frac{C_\ell^{ss}}{C_\ell^{ss}+C_\ell^{\bar n\bar n}}(\bar s_i)_{\ell m}$$

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

To analyze the algorithm write the iterations

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

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

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

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

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

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

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

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

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

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

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

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

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

$\uparrow$

true wiener filter.

In a similary way the residuals satisfy

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

Since the spectral radius ($\rho(A)=$ max abs eigenvalue) of a product of matrices satisfies

$$\rho(AB)\le\rho(A)\rho(B)$$

we have

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

$\therefore$ the iterations are geometrically convergent.

Note: Slower convergence rate when Σ is large & $\lambda\bar N$ is small

$\therefore$ to speed up the convergence choose a "cooling schedual"

$$\lambda_1>\lambda_2>\cdots\longrightarrow 1.$$

Small Tweak to Message Passing

Since the spectral radius of $\Sigma(\Sigma+N)^{-1}$ is a factor in the convergence rate, the algorithm can be slow for modes which have high signal-to-noise ratio.

$\therefore$ It appears helpful to split out low rank & high $\frac{S}{N}$ modes from s:

$$d = s + n$$
$$= L + \underbrace{H + \bar{n}}_{=\bar{H}} + \tilde{n}$$

where $L \sim N(0, \Sigma_L)$, $H \sim N(0, \Sigma_H)$ } indep

Σ_L is low rank

e.g. $L(\hat{n}) = \sum_{\ell m}^{\ell_{max}} a_{\ell m} Y_{\ell m}(\hat{n})$. ← not too large.

Now the conditional WFs are

$$E(H \mid \bar{H}, \cancel{L}, \cancel{d})$$
$$E(\bar{H} \mid H, L, d)$$
$$E(L \mid \underbrace{H, \bar{H}, d}_{d - \bar{H} = L + \tilde{n}})$$

← These steps are similar to regular message passing

This one equals

$$\Sigma_L \left[\Sigma_L + \tilde{N}\right]^{-1}(d - \bar{H})$$

$\tilde{N}^{-1}$ is easy & Σ_L is a low rank update to $\tilde{N}$ so Sherman-Morrison-Woodbury applies.

Preconditioned Conjugate gradient

Notice that the computation of

$$E(s|d) = \Sigma(\Sigma+N)^{-1}d$$
$$= (\Sigma^{-1} + N^{-1})^{-1} N^{-1} d$$

is equivalent to solving

$$Ax = b$$

where $A = (\Sigma^{-1} + N^{-1})$ ← symmetric positive definite & full Rank.

$b = N^{-1}d$.

Claim: $f(x) = \frac{1}{2}x^T A x - x^T b$ has a unique minimizer

$$x^* = \underset{x}{\operatorname{argmin}}\, f(x)$$

which satisfies $Ax^* = b$.

Proof: $f(x)$ is quadratic & $f(x) \to \infty$ as $x \to \infty$ by positive definitness & full Rank of A

$\therefore$ x^* satisfies $\underbrace{\nabla f(x^*)}_{= Ax^* - b} = 0$.

unique by full Rank.

$\blacksquare$

Gradient desent produces a sequence $x_1, x_2, \ldots$ s.t.

$$x_{k+1} = x_k - \varepsilon \nabla f(x_k)$$
$$= x_k - \varepsilon(A x_k - b)$$

Since A isn't diagonal it couples the coordinates & minization can't be performed one coordinate at a time.

However, if we could change basis of x

$$x = \sum_{i=1}^{n} \underbrace{\beta_i}_{\text{coeffs i.e. new basis coords}} \overbrace{v_i}^{\text{basis vectors}}$$

s.t. $v_i^T A v_j = 0$ (i.e. orthogonal w.r.t $\langle v, w \rangle_A = v^T A w$)

Then the objective as a function of $\beta_1, \ldots, \beta_n$ has the form:

$$f(\beta_1, \ldots, \beta_n) = \frac{1}{2} \sum_{i,j=1}^{n} \beta_i \beta_j \underbrace{v_i^T A v_j}_{= \delta_{ij}} - \sum_{i=1}^{n} \beta_i v_i^T b$$

$$= \frac{1}{2} \sum_{i=1}^{n} \beta_i^2 - \sum_{i=1}^{n} \beta_i v_i^T b$$

which can be minimized w.r.t. each coordinate individually.

minimizing β_i is found by

$$\beta_i = v_i^T b$$

The conjugate Gradient algorithm produces a minimizing sequence

$$x_1, x_2, \ldots$$

s.t. $\underbrace{x_{k+1} - x_k}_{\beta_{k+1} v_{k+1}}, \underbrace{x_k - x_{k-1}}_{\beta_k v_k}, \ldots$ are orthogonal w.r.t. $\langle \cdot, \cdot \rangle_A$ &

$$x_{k+1} = (x_{k+1} - x_k) + (x_k - x_{k-1}) + \cdots + (x_1 - \overbrace{x_0}^{\text{inital guess}})$$

$$\overset{(*)}{=} \sum_{i=1}^{k+1} \beta_i v_i \quad \text{s.t. } \beta_i = v_i^T b$$

$$= (x_i - x_{i-1})^T b$$

i.e. minimizes $f(\beta_1, \ldots, \beta_n)$ in coordinate i.

Basic Conjugate Gradient algorithm

$x_0 = 0$

$r_0 = b - A x_0$

$p_0 = r_0$

for $k = 0, 1, \ldots$

$$x_{k+1} = x_k + \|r_k\|^2 \frac{p_k}{p_k^T A p_k}$$

$$r_{k+1} = r_k - \|r_k\|^2 \frac{A p_k}{p_k^T A p_k}$$

Stop if $\|r_{k+1}\| \le \varepsilon \|b\|$

$$p_{k+1} = r_{k+1} + \frac{\|r_{k+1}\|^2}{\|r_k\|^2} p_k$$

end

Note that $(*)$ holds at step $k = 0$ since

$$x_1 - x_0 = \frac{\|b\|^2}{\sqrt{b^T A b}} \underbrace{\frac{b}{\sqrt{b^T A b}}}_{v_1} = \beta_1 v_1 \quad \text{where } \beta = v_1^T b$$

$$\left(\text{so } v_1^T b = \frac{\|b\|^2}{\sqrt{b^T A b}}\right)$$

The other steps are shown similarly.

Note: One never needs to actually construct A... just have an algorithm to compute $x \mapsto Ax$.

e.g. when $A = (\Sigma^{-1} + N^{-1})$ & obs are on the full sky

$$= c\mathcal{F}^T (\underbrace{\Lambda_\Sigma^{-1} + \Lambda_N^{-1}}_{\text{diagonal}}) \underbrace{\mathcal{F}}_{\text{FFT or spherical harmonic transform}}$$

So $Ax = \underbrace{c\mathcal{F}^T}_{\text{IFFT}} (\underbrace{\Lambda_\Sigma^{-1} + \Lambda_N^{-1}}_{\text{coord mult}}) \underbrace{\mathcal{F} x}_{\text{FFT}}$

Note:

when $\text{length}(x) = n$ any set of n conjugate vectors (i.e. with $\langle v, w \rangle_A = 0$ if $v \neq w$) span $\mathbb{R}^n$.

$\therefore$ After n steps the Conj Grad Algorithm is at the exact minimizer x^* (assuming No numerical instability).

Note:

Convergence is fast if the eigenvalues of A are not spread out (i.e. a well conditioned matrix). if A is ill-conditioned one needs to Find T s.t.

$$Ax = b \iff ATT^{-1}x = b$$

$$\iff T^T A T \tilde{x} = \tilde{b}$$

when $\tilde{x} = T^{-1}x$

& $\tilde{b} = T^T b$

and T^TAT is well conditioned.

Finding T is equiv to finding M s.t.

$$M = TT^T \approx A^{-1}$$

Re-arranging terms in a conj gradient algorithm for $T^TAT\tilde{x} = \tilde{b}$ gives

Pre-conditioned Conjugate Gradient

Note: M must be symmetric & pos def.

$x_0 = 0$

$r_0 = b - Ax_0$

$z_0 = M^{-1} r_0$

$p_0 = z_0$

for $k = 0, 1, 2, \ldots$

$$x_{k+1} = x_k + r_k^T z_k \frac{p_k}{p_k^T A p_k}$$

$$r_{k+1} = r_k - r_k^T z_k \frac{A p_k}{p_k^T A p_k}$$

Stop if $\|r_{k+1}\| \leq \varepsilon \|b\|$

$$z_{k+1} = M^{-1} r_{k+1}$$

$$p_{k+1} = r_{k+1} + \frac{r_{k+1}^T \cdot z_{k+1}}{r_k^T \cdot z_k} p_k$$

end

Flat sky Approximation   Ref Wayne Hu's paper "Weak lensing of the CMB: A harmonic Approach" (19)

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

s.t. $s^{FS}(x) \cong s(\underbrace{\theta,\varphi}_{\text{Polar coords}})$, where $x = (\theta\cos\varphi, \theta\sin\varphi)$, $\theta\approx 0$.

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

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

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

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

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

$$\therefore \int_0^{2\pi} e^{-im\varphi}\xi_\ell(\varphi)\frac{d\varphi}{2\pi} = \sum_{n=-\ell}^{\ell}\frac{s_{\ell n}\sqrt{\ell}}{i^n\sqrt{2\pi}}\underbrace{\int_0^{2\pi} e^{-im\varphi}e^{in\varphi}\frac{d\varphi}{2\pi}}_{=\delta(m-n)}$$

$$= \frac{s_{\ell m}\sqrt{\ell}}{i^m\sqrt{2\pi}} I_{\{|m|\le\ell\}}$$

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

$$\therefore \quad s^{FS}(x) = \sum_{\ell=0}^{\infty} \int_0^{2\pi} \xi_\ell(\varphi) \underbrace{\exp(ix\cdot(\ell\cos\varphi, \ell\sin\varphi))} \frac{d\varphi}{2\pi}$$

Jacobi-Anger expansion

$$= \sum_{\ell=0}^{\infty} \int_0^{2\pi} \xi_\ell(\varphi) \left[\sum_{m\in\mathbb{Z}} i^m J_m(\ell\theta_x) e^{im(\varphi_x - \varphi)} \right] \frac{d\varphi}{2\pi}$$

$$= \sum_{\ell=0}^{\infty} \sum_{m\in\mathbb{Z}} i^m J_m(\ell\theta_x) e^{im\varphi_x} \underbrace{\left[\int_0^{2\pi} \xi_\ell(\varphi) e^{-im\varphi} \frac{d\varphi}{2\pi} \right]}_{= \frac{s_{\ell m}\sqrt{\ell}}{i^m\sqrt{2\pi}} I_{\{|m|\le\ell\}}}$$

$$= \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{s_{\ell m}\sqrt{\ell}}{\sqrt{2\pi}} \underbrace{J_m(\ell\theta_x) e^{im\varphi_x}}_{\approx \frac{\sqrt{2\pi}}{\sqrt{\ell}} Y_{\ell m}(\theta_x, \varphi_x)}$$

for $\theta_x \approx 0$ by Gradshteyn & Ryzhik eq. 8.722

$$\approx \sum_{\ell m} s_{\ell m} Y_{\ell m}(\theta_x, \varphi_x)$$

as was to be shown.

Notice also that

$$s_k^{FS} = \int_{\mathbb{R}^2} \frac{dx}{2\pi} e^{-ik\cdot x} s^{FS}(x)$$

$$= \sum_{\ell=0}^{\infty} \int_0^{2\pi} \frac{d\varphi}{2\pi} \int_{\mathbb{R}^2} \frac{dx}{2\pi} e^{-ik\cdot x} \xi_\ell(\varphi) \exp(ix\cdot(\ell\cos\varphi, \ell\sin\varphi))$$

$$= \sum_{\ell=0}^{\infty} \int_0^{2\pi} \frac{d\varphi}{2\pi}\, \xi_\ell(\varphi)\, \delta^{(2\pi)}_{k - \begin{pmatrix}\ell\cos\varphi \\ \ell\sin\varphi\end{pmatrix}}$$

$$= \begin{cases} \frac{d\varphi}{dk}\, \xi_\ell(\varphi) & \text{when } k = (\ell\cos\varphi, \ell\sin\varphi)\ \&\ \ell \in \mathbb{Z}^+ \\ 0 & \text{o.w.} \end{cases}$$

Note $\frac{d\varphi}{dk} = \frac{\cancel{d\varphi}}{\ell\, d\ell\, \cancel{d\varphi}} = \frac{1}{\ell\, d\ell} \approx \frac{1}{\ell}$

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

$$= \left(\frac{d\varphi}{dk}\right)^2 \sum_{n=-\ell}^{\ell} \sum_{n'=-\ell'}^{\ell'} \frac{\sqrt{\ell}\sqrt{\ell'}}{2\pi} \underbrace{E(S_{\ell n}\, S_{\ell' n'})}_{\delta_{\ell\ell'}\,\delta_{nn'}\, C_\ell^{SS}} e^{in\varphi - in'\varphi'}$$

$$= \left(\frac{d\varphi}{dk}\right)^2 \sum_{n=-\ell}^{\ell} \frac{\ell\,\delta_{\ell\ell'}}{2\pi} C_\ell^{SS} e^{in(\varphi-\varphi')}$$

$$= \left(\frac{d\varphi}{dk}\right)^2 \ell\, \delta_{\ell\ell'}\, C_\ell^{SS} \underbrace{\sum_{n=-\ell}^{\ell} \frac{e^{in(\varphi-\varphi')}}{2\pi}}_{\approx\, \delta_{\varphi-\varphi'}}$$

$$\cong I_{k=k'} \left(\frac{1}{\ell}\right)^2 \frac{\ell}{d\varphi}\, C_\ell^{SS}$$

$$\approx \frac{I_{k=k'}}{\ell\, d\ell\, dk}\, C_\ell^{SS} \approx \delta_{k-k'}\, C_\ell^{SS}$$

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

The advantage is that a flat sky model

$$d(x) = s(x) + n(x), \quad x \in \mathbb{R}^2$$

with $C_k^{ss,\mathbb{R}^2} \approx C^{ss}_{[|k|]}$ is often easier to work with and analyze since products, convolutions, derivatives etc are easy to express with FFT etc...

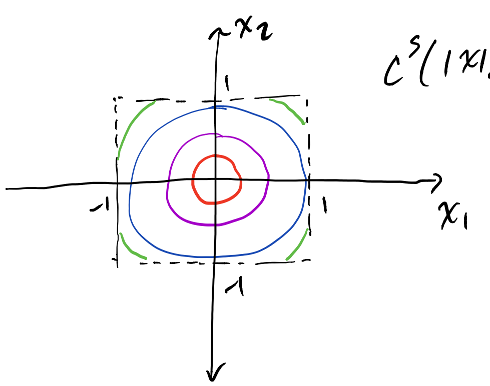
Circulant embedding.

Actually a simulation technique but can be applied for matrix inversion of a cov matrix with special structure.

Let $\{s(x) : x \in \mathbb{R}^2\}$ be a mean zero, isotropic GRF with cov

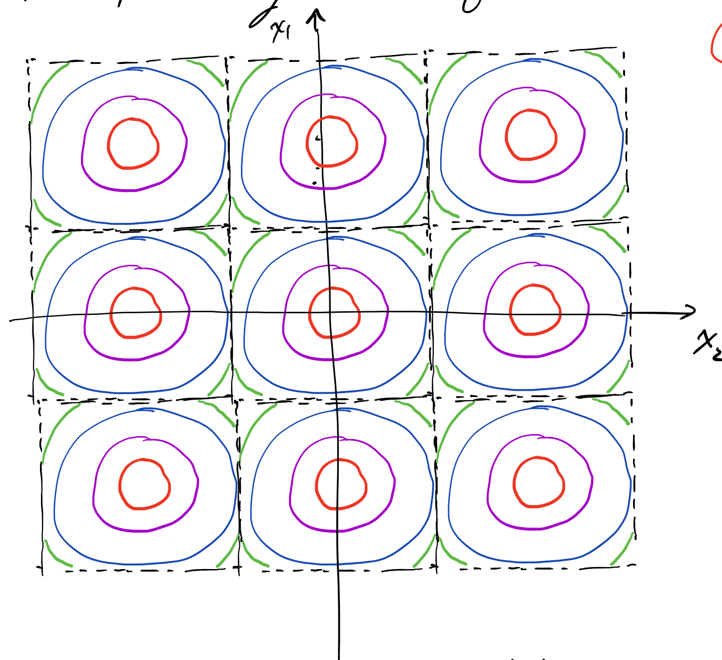
$$cov(s(x), s(y)) = C^s(|x-y|)$$

Restrict $C^s(|x|)$ to $[-1,1]^2$



$C^s(|x|) I_{\{|x| \le 1\}}$

Then periodically tile all of $\mathbb{R}^2$: $PC^s(x)$



In general $PC^s(x)$ will not be positive definite.

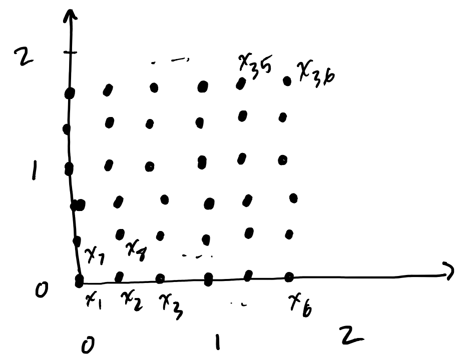
It is easy to see since a positive definite auto covariance $C^s(x)$ must satisfy

$$(\text{smoothness of } C^s(x) \text{ @ } x=0) \le \underbrace{(\text{smoothness of } C^s(x) \text{ @ } x \neq 0.)}$$

...but the periodic tiling of $PC^s(x)$ makes it non-smooth at the edges of the tiles.

Claim: If $C^s(|x|)$ has compact support in $[-1,1]$ then $PC^s(x)$ will be positive definite.

Claim Let $x_1, \dots, x_n$ be spatial points on a equispaced grid in $[0,2)^2$ with an even number of points to a side, ordered as follows



Then the cov matrix $\Sigma = \left(PC^s(x_i, x_j)\right)_{ij}$ is diagonalized as follows

$$\Sigma = c\,(F^T)^* \Lambda F$$

constant

IFFT matrix

FFT matrix

where Λ is a diagonal matrix with positive entries and

$$\operatorname{diag} \Lambda = \text{FFT of } \underbrace{\Sigma[:,1]}_{\text{first colum of } \Sigma.}$$

Proof: The periodicity of PC^s & the grid structure implies Σ is block circulent with circulent blocks. $\blacksquare$

Note: Toeplitz matrix has the form

$$\begin{pmatrix} a_0 & a_{-1} & & a_{-(n-1)} \\ a_1 & a_0 & \cdots & \vdots \\ \vdots & a_1 & & \vdots \\ a_{n-1} & a_{n-2} & & a_0 \end{pmatrix}$$

and a circulent matrix has the form

$$\begin{pmatrix} a_0 & a_{n-1} & & a_{n-1} \\ a_1 & a_0 & \cdots & a_{n-2} \\ \vdots & a_1 & & \vdots \\ a_{n-1} & a_{n-2} & & a_0 \end{pmatrix}$$

So the periodic extension in PC^s converts what would normally be a toeplitz structure to circulent structure.

Note: Now computing Σ^{-1}, $\Sigma^{1/2}$ etc are easily done with FFT, IFFT & never need to construct anything but the first column of Σ.

Note: even if PC^s isn't positive definite, $\operatorname{diag} \Lambda$ may be ≥ 0 for a particular grid. In which case restricting $s = \Sigma^{-1/2} z$ ← $\sim N(0, I)$ to the grid in $[0,1)^2$ results in a exact simulation of a GRF with cove $\underbrace{C^s}_{\text{The Non-periodic } C^s.}$