

Lecture 7

Topics:

I) The conditional distribution of T given the data d & C_e^{TT}

II) Wiener filtering.

We have 2 equivalent ways of writing our idealized data model

Pixel domain) $d(\hat{n}) = \underbrace{q * T(\hat{n})}_{\text{unit white noise}} + \underbrace{\sigma w(\hat{n})}_{Q_{em} = \frac{1}{2\pi} \sqrt{\frac{2\pi k_1}{m}} \exp\left(-\frac{k_2^2}{2\pi k_1} \ell(\hat{n})\right)}$

observed on all $\hat{n} \in S^2 \subset \mathbb{R}^3$.

Harmonic domain) $d_{em} = T_{em} + \underbrace{e_{em}}_{\text{has spectral density } \sigma^2 \exp\left(-\frac{k_2^2}{2\pi k_1} \ell(m)\right)}$

observed for all $\ell = 0, 1, \dots, -L \leq m \leq L$

The previous lectures derived the posteriors

$$P(C_e^{TT} | d) \text{ & } P(G | d)$$

where $\theta = (\lambda_{ek}^2, \lambda_{bm}^2, r, e_s, A_s, n_s)$.

However, the raw temperature map $T(\hat{n})$ is also of scientific interest.

In this lecture we will analyze the map posterior (assuming C_e^{TT} is known)

$$P(T | d, C_e^{TT}). \quad \leftarrow \begin{array}{l} \text{band power} \\ \text{are no longer} \\ \text{sufficient} \\ \text{statistics} \end{array}$$

After this we will combine the two to study the full posterior

$$P(T, C_e^{TT} | d)$$

Finite dimensional Analog

Suppose $d = s + n$

↑
data vector
↑
signal vector
noise vector

where $s \sim N(0, \Sigma)$

$$n \sim N(0, N)$$

are independent where Σ & N are known.

Then the conditional density of s given d is

$$p(s | d) \sim N\left(\underbrace{E(s | d)}_{\text{conditional expected value}}, \underbrace{\text{var}(s | d)}_{\text{conditional var/cov matrix}}\right).$$

A fancy way to derive $E(s | d)$ & $\text{var}(s | d)$ is to notice that

$$E(s | d) = \underset{s}{\operatorname{argmax}} \log p(s | d).$$

$$= \underset{s}{\operatorname{argmax}} \left[\log p(s, d) - \log p(d) \right]$$

$$= \underset{s}{\operatorname{argmax}} \log p(s, d).$$

& similarly

$$\text{var}(s | d)^{-1} = - \nabla_s^2 \log p(s | d). \quad \Big|_{s=E(s | d)}$$

$$= - \nabla_s^2 \log p(s, d). \quad \Big|_{s=E(s | d)}$$

Fisher information.

(Note: we don't need an expected value here since the hessian doesn't depend on s or d)

think of this as the analog to the pixel domain model.

Now for $E(s|d)$ find the stationary points (i.e. $\nabla_s \log p(s,d) = 0$)

$$\begin{aligned}\nabla_s \log p(s,d) &= \nabla_s [\log p(d|s) + \log p(s)] \\ &= -\nabla_s \left[\frac{(d-s)^T N^{-1} (d-s)}{2} + \frac{s^T \Sigma^{-1} s}{2} \right] \\ &= -[-N^{-1}(d-s) + \Sigma^{-1}s] \\ &= N^{-1}d - (N^{-1} + \Sigma^{-1})s \\ &= 0\end{aligned}$$

$$\begin{aligned}\therefore E(s|d) &= \underbrace{\left(N^{-1} + \Sigma^{-1} \right)^{-1} N^{-1} d}_{\text{solves the scale eqn}} \\ &= \left(N^{-1} (\Sigma + N) \Sigma^{-1} \right)^{-1} N^{-1} d \\ \therefore E(s|d) &= \sum \underbrace{(\Sigma + N)^{-1} d}_{\substack{\text{cov matrix of } d \\ \text{cross cov of } d \text{ with } s.}}\end{aligned}$$

$$\begin{aligned}\text{Also } -\nabla_s^2 \log p(s,d) &= N^{-1} + \Sigma^{-1}\end{aligned}$$

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

Note: d represents the pixel space data so Σ represents the spatial cov across pixels.

Computing $(N^{-1} + \Sigma^{-1})^{-1}$ or $(\Sigma + N)^{-1}$ is hard. However if both N & Σ have the same eigen vectors then we get a simplification:

$$\begin{aligned}N &= U \Lambda_n U^T && \text{drag matrix} \\ \Sigma &= U \Lambda_s U^T && \text{with pos. entries}\end{aligned}$$

$$\therefore E(s|d) = U \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n} \right) U^T d$$

$$\text{var}(s|d) = U \left(\frac{1}{\Lambda_s + \Lambda_n} \right) U^T$$

Take the eigen transform of s, n, d :

$$\tilde{s} := U^T s \sim \mathcal{N}(0, \Lambda_s)$$

$$\tilde{n} := U^T n \sim \mathcal{N}(0, \Lambda_n)$$

$$\tilde{d} := U^T d \sim \mathcal{N}(0, \Lambda_s + \Lambda_n)$$

$$\text{So } \tilde{d} = \tilde{s} + \tilde{n} \quad \xleftarrow{\substack{\text{Analogy to the} \\ \text{harmonic domain model}}} \quad \text{where}$$

$$E(\tilde{s}|\tilde{d}) = \frac{\Lambda_s}{\Lambda_s + \Lambda_n} \tilde{d}$$

every thing is done coordinate wise.

$$\text{var}(\tilde{s}|\tilde{d}) = \frac{1}{\Lambda_s + \Lambda_n}$$

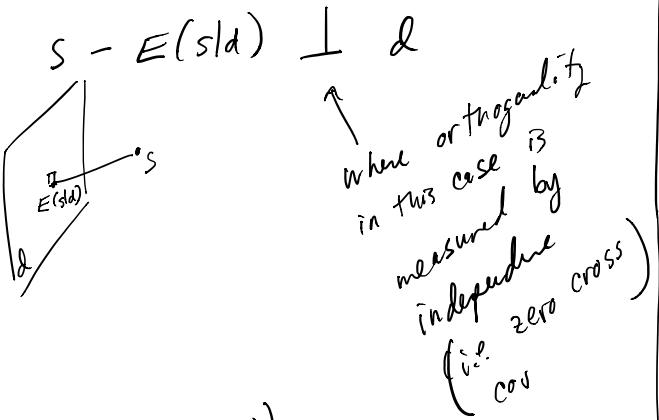
Think of the diag of Λ_s & Λ_n as spectral densities:

$$E(\tilde{s}_i \tilde{s}_j) = \delta_{ij} (\Lambda_s)_{ii}$$

$$E(\tilde{n}_i \tilde{n}_j) = \delta_{ij} (\Lambda_n)_{ii}$$

Remark :

$E(s|d)$ acts like a projection of d onto S so that



$$\therefore \text{var}(s - E(s|d))$$

$$= \text{var}(s - E(s|d)|d)$$

$$(A) = \text{var}(s|d) \quad \begin{matrix} \text{this is constant} \\ \text{when conditioning} \\ \text{on } d \end{matrix}$$

Now to produce a conditional simulation $s^* \sim p(s|d)$ one would normally simulate $\tilde{z}^* \sim N(0, \text{var}(s|d))$ & set

$$s^* = E(s|d) + \underbrace{\tilde{z}^*}_{\text{independent}}$$

Now if you're coding up $E(s|d)$ & want to avoid computing $\text{var}(s|d)$ you can instead produce $s^* \sim p(s|d)$ as follows:

Step 1: simulate a new data set

$$d^0 = s^0 + n^0, \quad \begin{cases} s^0 \sim N(0, \Sigma) \\ n^0 \sim N(0, N) \end{cases}$$

Step 2: Compute the prediction error $\tilde{z}^0 = s^0 - E(s^0|d^0)$

Step 3: Set $s^* = E(s|d) + \tilde{z}^0$

\tilde{s}_i^0 has the right marginal behavior by (A)

(5)

Remark :

If $\tilde{s}^* \sim p(\tilde{s}|\tilde{d})$ then \tilde{s}^* has the right spectral density marginalizing over \tilde{d}

$$E(\tilde{s}_i^* \tilde{s}_j^* | \tilde{d}) = E((E(s|d) + \tilde{z}^*)_i (E(s|d) + \tilde{z}^*)_j | \tilde{d})$$

$$= E(\tilde{s}| \tilde{d})_i E(\tilde{s}| \tilde{d})_j + E(\tilde{z}^*_i \tilde{z}^*_j)$$

$$= \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n} \right)_i \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n} \right)_j \tilde{d}_i \tilde{d}_j + \frac{\delta_{ij}}{(\Lambda_s^{-1} + \Lambda_n^{-1})_{ii}}$$

$$\therefore E(\tilde{s}_i^* \tilde{s}_j^*) = \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n} \right)_i \left(\frac{\Lambda_s}{\Lambda_s + \Lambda_n} \right)_{jj} \delta_{ij} (\Lambda_s + \Lambda_n)_{ii} + \frac{\delta_{ij}}{(\Lambda_s^{-1} + \Lambda_n^{-1})_{ii}}$$

$$= \left(\frac{\Lambda_s^2}{\Lambda_s + \Lambda_n} \right)_{ii} \delta_{ij} + \left(\frac{\Lambda_s \Lambda_n}{\Lambda_s + \Lambda_n} \right)_{ii} \delta_{ij}$$

$$= (\Lambda_s)_{ii} \delta_{ij}$$

\therefore if someone looked at \tilde{s}^* but didn't see \tilde{d} they would think \tilde{s}^* was simply generated from $N(0, \Lambda_s)$.

Remark : Interpretation:

$$E(\tilde{s}|\tilde{d}) = \underbrace{\frac{\Lambda_s}{\Lambda_s + \Lambda_n} \tilde{d}}_{\text{represents the proportion of signal power in } \tilde{s}_i \text{ compared to total power in } \tilde{d}_i}$$

represents the proportion of signal power in \tilde{s}_i compared to total power in \tilde{d}_i

\therefore estimated $\tilde{s}_i = \begin{cases} 0 & \text{if proportion is small} \\ \tilde{d}_i & \text{if proportion is large} \end{cases}$

$$\text{var}(\tilde{s}_i|\tilde{d}) = \frac{1}{(\Lambda_s^{-1} + \Lambda_n^{-1})_{ii}} < \min((\Lambda_s)_{ii}, (\Lambda_n)_{ii})$$

$$\text{var}(\tilde{s}_i|\tilde{d}) = \left(\frac{\Lambda_s \Lambda_n}{\Lambda_s + \Lambda_n} \right)_{ii} = \left(\frac{\text{Proportion of signal power}}{\text{noise power}} \right) \cdot \text{noise power}$$

(6)

Wiener Filtering

(7)

Wiener filtering is just the analog of what we just did but for Stationary / isotropic Gaussian random fields.

so the functional eigenvalues of the cov is given by Fourier or spherical Harmonic's

Claim (Wiener filter):

Suppose $s(x) \& n(x)$ are stationary, mean zero, Gaussian Random fields on $x \in \mathbb{R}^d$ (possibly generalized) with spectral densities C_k^{ss}, C_k^{nn} (on \mathbb{R}^{2d}) respectively.

Let $d(x) = s(x) + n(x)$.

Then $\hat{s} := E(s|d)$ satisfies the whole functions

$$\hat{s}_k = \frac{C_k^{ss}}{C_k^{ss} + C_k^{nn}} d_k$$

Claim (Wiener filtering on S^2):

Same as above but now

$d(\hat{n}) = s(\hat{n}) + n(\hat{n})$
isotropic, mean zero,
Gaussian RFs on S^2
with spherical spectral
densities C_e^{ss}, C_e^{nn}

Then $\hat{s} := E(s|d)$

satisfies

$$\hat{s}_{em} = \frac{C_e^{ss}}{C_e^{ss} + C_e^{nn}} d_{em}$$

Remark: In some sense, Wiener filtering is just Gaussian conditional expected value for the special case of stationary / isotropic models when $d = s + n$.

(8)

Another way to put it:

$$E(Y|X) = \Sigma_{xx}^{-1} X$$

= "generalized Wiener filtering".

Remark: To produce a conditional sample from $s^* \sim p(s|d)$

$$\text{on } S^2: S_{em}^* = \frac{C_e^{ss}}{C_e^{ss} + C_e^{nn}} d_{em} + \frac{C_e^{ss} C_e^{nn}}{C_e^{ss} + C_e^{nn}} \varepsilon_{em}$$

$$\text{on } \mathbb{R}^d: S_k^* = \frac{C_k^{ss}}{C_k^{ss} + C_k^{nn}} d_k + \frac{C_k^{ss} C_k^{nn}}{C_k^{ss} + C_k^{nn}} \varepsilon_k$$

When $\varepsilon(n)$ & $\varepsilon(x)$ are unit var white noise.

Remark: Although $s^* \sim p(s|d)$ behave like marginal draws from $p(s)$, the Wiener filters $\hat{s}(x)$ look different as compared to realizations of $s(x)$.

