(DoD) 11.3 Estimating the Likelihood function

(DoD) 11.3.1 Karhunen - Loève Techniques

This technique is not really used prominently these day, nevertheless instructive. Idea is simple: Any experiment will have many modes which are useless, because they are dominated by noise. If it was obvious which ones, we could speed up analysis, because inversion of matrix costs N^3 operations, so if only 10% of modes necessary, this will give ~1000 speed improvement.

If C_S and C_N were diagonal, we would be done, because diagonal entries of C_N and C_S could be compared. And for $(C_S)_{ii} > (C_N)_{ii}$ signal dominates. Karhunen - Loève method achieves this for realistic C_S and C_N!

Assume N_p data points Δ_i ~~and~~ containing signal s_i and noise n_i uncorrelated: $\langle \vec{s}\vec{n}^T \rangle = \langle \vec{n}\vec{s}^T \rangle = 0$

$$\Rightarrow \langle \Delta_i \Delta_i \rangle = C_{ii} = C_{S,ii} + C_{N_{ii}}$$

$$\langle \vec{\Delta}\vec{\Delta}^T \rangle = C = C_S + C_N$$

Use rotated data:

$$\vec{\Delta}' \equiv R\vec{\Delta}$$

$$\Rightarrow C' = \langle \vec{\Delta}'\vec{\Delta}'^T \rangle = \langle (R\vec{\Delta})(R\vec{\Delta})^T \rangle$$

$$= \langle R\vec{\Delta}\vec{\Delta}^T R^T \rangle = R\langle \vec{\Delta}\vec{\Delta}^T \rangle R^T = RCR^T$$

Karhunen-Loève method consists of 3 rotations:

1. R_1 : Diagonalize C_N
2. R_2 : Set $C'_N = \mathbb{1}$
3. R_3 : Diagonalize C_S'

- First step always possible since C_N real, $C_N = C_N^T$
- R_2 trivial, simply choose ~~R~~
  $R_2 = \mathrm{diag}\left([C'_{N,11}]^{-1/2}, [C'_{N,22}]^{-1/2}, \dots\right)$
- Step 3 always possible, since C_S real, $C_S = C_S^T$

Signal matrix becomes:

$$C_S' = R_3 R_2 R_1 C_S R_1^T R_2^T R_3^T$$
$$= R_3 R_2 R_1 C_S R_1^T R_2 R_3^T \qquad ; \quad R_2^T = R_2 \quad (R_2 \text{ diagonal})$$

C_N is $\mathbb{1}$ after step 2. So step 3 leaves it $\mathbb{1}$:

$$R_3 \mathbb{1} R_3^T = R_3 R_3^T \mathbb{1} = \mathbb{1} \quad , \quad R_3 \text{ unitary}$$

$$\Rightarrow \quad C_N' = \mathbb{1}$$

$\Rightarrow$ Elements of diagonal C_S' matrix are a measure of (signal : noise)^2 of modes!

$$\vec{\Delta}' = R_3 R_2 R_1 \vec{\Delta}$$

have diagonal covariance!

$$\langle \Delta_i \Delta_j \rangle = \begin{cases} 1 + C'_{S,ii} & ; i = j \\ 0 & ; i \neq j \end{cases} \quad !$$

$\Rightarrow$ Order these modes according to signal to noise!
$\Rightarrow$ discard modes with $C'_{S,ii} \ll 1$.

EXAMPLE

$$C_N = \begin{pmatrix} \sigma_n^2 & 0 \\ 0 & \sigma_n^2 \end{pmatrix}$$

Two pixel experiment with diagonal noise

Signal matrix has correlations between pixels, say:

$$C_S = \sigma_s^2 \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}$$

$-1 < \varepsilon < 1$ measures correlation

We have to:

1. Diagonalize C_N. Trivial, C_N already diagonal
$\Rightarrow R_1 = \mathbb{1}$

2. Set $C_N \rightarrow \mathbb{1}$: $R_2 = \frac{1}{\sigma_n} \mathbb{1}$

3. Diagonalize C_S, i.e.

$$R_2 R_1 C_S R_1^T R_2 = \frac{\sigma_s^2}{\sigma_n^2} \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}$$

Eigenvalues of $\begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}$: $\det \begin{pmatrix} 1-\lambda & \varepsilon \\ \varepsilon & 1-\lambda \end{pmatrix} = 0$

$\Rightarrow (1-\lambda)^2 - \varepsilon^2 = 0 \Rightarrow$ ~~$1 - 2\lambda + \lambda^2 - \varepsilon^2 = 0$~~
~~$\Rightarrow \lambda_{1,2} = 1 \pm \sqrt{1}$~~ $\quad \lambda^2 - 2\lambda + (1-\varepsilon^2) = 0$

$$\Rightarrow \lambda_{1,2} = 1 \pm \sqrt{1-(1-\varepsilon^2)} = 1 \pm \sqrt{\varepsilon^2} = 1 \pm \varepsilon$$

Diagonal matrix has eigenvalues as entries, so

$$C_s' = \frac{\sigma_s^2}{\sigma_n^2}\begin{pmatrix} 1+\varepsilon & 0 \\ 0 & 1-\varepsilon \end{pmatrix}$$

and rotation matrix has eigenvectors as column vectors

$$\Rightarrow \begin{pmatrix} 1-1-\varepsilon & \varepsilon \\ \varepsilon & 1-1-\varepsilon \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \varepsilon\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = -\varepsilon\begin{pmatrix} x-y \\ x-y \end{pmatrix} = 0$$

$\Rightarrow \begin{pmatrix} 1 \\ +1 \end{pmatrix}$ e.v. to $\lambda = 1+\varepsilon$

$$\begin{pmatrix} 1-1+\varepsilon & \varepsilon \\ \varepsilon & 1-1+\varepsilon \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \varepsilon\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \varepsilon\begin{pmatrix} x+y \\ x+y \end{pmatrix} = 0$$

$\Rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ e.v. to $\lambda = 1-\varepsilon$

$$\Rightarrow R_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = R_3^T$$

So new modes are

$$\vec{\Delta}' = R_3 R_2 \vec{\Delta}$$

$$= \frac{1}{\sqrt{2}\,G_n}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \frac{1}{\sqrt{2}\,G_n}\begin{pmatrix} \Delta_1+\Delta_2 \\ \Delta_1-\Delta_2 \end{pmatrix}$$

What are our new modes? Consider special cases $\varepsilon = 0$; $\varepsilon = 1$. For $\varepsilon = 0$, C_s' has same entries on diagonal, hence both modes have same signal to noise.

For $\varepsilon = 1$, i.e. maximal correlation, we see that $C_s' \rightarrow \frac{\sigma_s^2}{\sigma_n^2} \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$; i.e. difference

mode $\sim \Delta_1 - \Delta_2$ has no information at all. Yet, the sum-mode $\Delta_1 + \Delta_2$ has signal to noise $\sqrt{2}\,\frac{\sigma_s}{\sigma_n}$, because the two measurements beat down the noise by a factor of $\sqrt{2}$!

- Examples: COBE and CFA2, Fig 11.8 + 11.9
Note: Useful bonus for LSS: small scale modes in which Poisson noise dominates and non-linearities rule automatically removed!

- Karhunen-Loève still not soo useful, because even though new C_s for reduced number of modes is much smaller, still needs to be diagonalized at each point in parameter space to probe likelihood contours.

- Theoretically problematic that one has to choose a "pivot" C_s in the beginning to determine important modes

- Fig 11.10 Karhunen-Loève very useful for consistency check: In the Karhunen-Loève basis, each data point d'_i should be drawn from a gaussian (in case of CMB) with variance $(1 + C'_{s,ii})$. So histogram of $d_i / \sqrt{1 + C_{s,ii}}$ should be gaussian

In case of Python experiment (ignore central spike), too much power in tails, so not a gaussian.
Fig 11.11. Noise model was not perfect? Adjacent points were more correlated than anticipated?