d-dimensional Discrete Fourier Transform

The d-dimensional DFT is a linear transformation acting on discrete fields v indexed by $(n_1, \dots, n_d)$ where $n_i \in \{1, \dots, N_i\}$, i.e. $v \in \mathbb{R}^{N_1 \times \dots \times N_d}$ or $v \in \mathbb{C}^{N_1 \times \dots \times N_d}$.

Notation: For $v \in \mathbb{R}^{N_1 \times \dots \times N_d}$ or $v \in \mathbb{C}^{N_1 \times \dots \times N_d}$ and $n_j \in \{1, \dots, N_j\}$, $j = 1, \dots, d$ let

$$v[n_1, \dots, n_d] = \text{"coordinate entry of } v \text{ at index } (n_1, \dots, n_d)\text{"}$$

and let

$$v[:] = \text{"flatten } v \text{ to a vector in } \mathbb{R}^N, \text{ where } N = N_1 \cdots N_d\text{"}$$

$$= \begin{bmatrix} v[:, 1, 1, \dots, 1] \\ v[:, 2, 1, \dots, 1] \\ \vdots \\ v[:, N_1, N_2, \dots, N_d] \end{bmatrix}$$

These are vectors in $\mathbb{R}^{N_1}$

Occationally I will use $[:, \dots, :]$ to denote "un-flattening", i.e.

$$v \in \mathbb{R}^{N_1 \times \dots \times N_d} \Longrightarrow (v[:])[:, \dots, :] = v$$

Notice that the square brackets for indexing is a visual cue that indicates a discrete d-dim array (or field)

$$(v[n_1, \dots, n_d] : n_j \in \{1, \dots, N_j\}, j = 1, \dots, d)$$

discrete d-dimensional field

$$(f(x_1, \dots, x_d) : x_j \in [0, P_j), j = 1, \dots, d)$$

continuous field on $[0, P_1) \times \dots \times [0, P_d)$

Definition:
The d-dimensional DFT (d-DFT for short) of $v \in \mathbb{C}^{N_1 \times \dots \times N_d}$ is written

$$z = W_{N_1, \dots, N_d} v$$

and defined as

$$z[k_1, \dots, k_d] = \sum_{n_1, \dots, n_d = 1}^{N_1, \dots, N_d} \omega_{N_1}^{(k_1 - 1)(n_1 - 1)} \cdots \omega_{N_d}^{(k_d - 1)(n_d - 1)} v[n_1, \dots, n_d]$$

where $\omega_{N_j} = \exp(-\sqrt{-1}\, 2\pi / N_j)$.

The above definition of the d-DFT is for discrete fields $v \in \mathbb{C}^{N_1 \times \dots \times N_d}$.

There is an equivalent definition of d-DFT as a matrix operator which we will use for deriving facts about d-DFT.

Claim: Let $v \in \mathbb{C}^{N_1 \times \dots \times N_d}$ and

$$z = W_{N_1, \dots, N_d} v$$

denote the d-DFT of v. Then

$$z[:] = \underbrace{W_{N_1} \otimes \cdots \otimes W_{N_d}}_{} v[:]$$

each $W_{N_j} \in \mathbb{C}^{N_j \times N_j}$ so this is a $(N_1 \cdots N_d)$-by-$(N_1 \cdots N_d)$ matrix

with inverse given by

$$v[:] = \frac{1}{N_1 \cdots N_d} \overline{W_{N_1}} \otimes \cdots \otimes \overline{W_{N_d}} \, z[:]$$

In these notes we will treat $W_{N_1,\dots,N_d}$ acting on $\mathbb{C}^{N_1\times\cdots\times N_d}$ as equivalent to $W_{N_1}\otimes\cdots\otimes W_{N_d}$ acting on $\mathbb{C}^{N_1\cdots N_d}$ with the understanding that conversion from field $\longleftrightarrow$ vector occurs using reshape $\longleftrightarrow$ flatten.

Now the different scalings of $W_{N_1,\dots,N_d}$ are defined similarly:

$$\mathcal{U}_{N_1,\dots,N_d} := \frac{1}{\sqrt{N_1\cdots N_d}} W_{N_1,\dots,N_d}$$

$$\mathcal{F}_{N_1,\dots,N_d} := \frac{\Delta x_1\cdots\Delta x_d}{(\sqrt{2\pi})^d} W_{N_1,\dots,N_d}$$

where Ωx will sometimes be used to denote pixel volumn $\Delta x_1\cdots\Delta x_d$.

Lets collect a few facts about Kronecker Product:

- If $A\in\mathbb{C}^{N\times M}$ & $B\in\mathbb{C}^{Q\times R}$ then
$$A\otimes B = \begin{bmatrix} A[1,1]B & \cdots & A[1,M]B \\ \vdots & & \\ A[N,1]B & \cdots & A[N,M]B \end{bmatrix} \in \mathbb{C}^{(NQ)\times(MR)}$$
- $(A\otimes B)^T = A^T\otimes B^T$ (same with $(\cdot)^H$)
- $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$
- $A\otimes B\otimes C \equiv \begin{cases}(A\otimes B)\otimes C \\ A\otimes(B\otimes C)\end{cases}$ i.e. associative
- $(A_1\cdots A_d)\otimes(B_1\cdots B_d)$
$$= (A_1\otimes B_1)\cdots(A_d\otimes B_d)$$
- If $A\in\mathbb{C}^{N\times N}$ & $B\in\mathbb{C}^{M\times M}$ then
$$\det(A\otimes B) = (\det(A))^M(\det(B))^N$$

These facts allow us to show use our results for 1-DFT:

$$\begin{aligned}
W_{N_1,\dots,N_d}^{-1} &\equiv (W_{N_1}\otimes\cdots\otimes W_{N_d})^{-1} \\
&= W_{N_1}^{-1}\otimes\cdots\otimes W_{N_d}^{-1} \\
&= \frac{1}{N_1\cdots N_d}\overline{W}_{N_1}\otimes\cdots\otimes\overline{W}_{N_d} \\
&= \frac{1}{N_1\cdots N_d} W_{N_1}^H\otimes\cdots\otimes W_{N_d}^H \quad \text{by transpose symmetry} \\
&= \frac{1}{N_1\cdots N_d}(W_{N_1}\otimes\cdots\otimes W_{N_d})^H
\end{aligned}$$

$$\begin{aligned}
\therefore\ I_{N_1\cdots N_d} &= (W_{N_1}\otimes\cdots\otimes W_{N_d})\left[\frac{1}{N_1\cdots N_d}(W_{N_1}\otimes\cdots\otimes W_{N_d})^H\right] \\
&= \frac{(W_{N_1}\otimes\cdots\otimes W_{N_d})}{\sqrt{N_1\cdots N_d}}\frac{(W_{N_1}\otimes\cdots\otimes W_{N_d})^H}{\sqrt{N_1\cdots N_d}} \\
&= (\underbrace{\mathcal{U}_{N_1}\otimes\cdots\otimes\mathcal{U}_{N_d}}_{\text{i.e. this is unitary}})(\mathcal{U}_{N_1}\otimes\cdots\otimes\mathcal{U}_{N_d})^H
\end{aligned}$$

In particular all of our 1-d results extend naturally.

- $$\mathcal{F}_{N_1,\dots,N_d}\mathcal{F}_{N_1,\dots,N_d}^H = \frac{\Omega x}{\Omega \tau} I_{N_1\cdots N_d}$$
where $\Omega x = \Delta x_1\cdots\Delta x_d$, $\Omega\tau = \Delta\tau_1\dots\Delta\tau_d$, $\Delta x_j = \frac{P_j}{N_j}$ and $\Delta\tau_j = \frac{2\pi}{P_j}$.

- $X\sim N(0, \frac{\sigma^2}{\Omega x} I_{N_1\cdots N_d})$
$$\Rightarrow E(zz^H) = \frac{\sigma^2}{\Omega\tau} I \quad \text{for } z = (\mathcal{F}_{N_1}\otimes\cdots\times\mathcal{F}_{N_d})X$$

- For convolution kernel $b[n_1,\dots,n_d]$
$$B_{N_1,\dots,N_d}^b = \mathcal{F}_{N_1,\dots,N_d}^{-1}\operatorname{diag}\left((\sqrt{2\pi})^d\mathcal{F}_{N_1,\dots,N_d}b\right)\mathcal{F}_{N_1\cdots N_d}$$

- $P_{N_1} \otimes \cdots \otimes P_{N_d} = \left(\mathcal{U}_{N_1}^H \otimes \cdots \otimes \mathcal{U}_{N_d}^H \right) \left(\cdots \right) \left(\mathcal{U}_{N_1} \otimes \cdots \otimes \mathcal{U}_{N_d} \right)$

$$\operatorname{diag}(\Gamma_{N_1}) \otimes \cdots \otimes \operatorname{diag}(\Gamma_{N_d})$$

where $\Gamma_{N_j} := (w_{N_j}^0, \ldots, w_{N_j}^{N_j - 1})^T$

Notice that

$$\sum_{n_1, \ldots, n_d = 1}^{N_1 \cdots N_d} v[n_1, \ldots, n_d] \left(\operatorname{diag}(\Gamma_{N_1})^{(n_1 - 1)} \otimes \cdots \otimes \operatorname{diag}(\Gamma_{N_d})^{(n_d - 1)} \right)$$

is a diagonal matrix with diagonal $z[:]$ where $z = W_{N_1, \ldots, N_d} v$.

In particular let $\Sigma \in \mathbb{R}^{(N_1 \cdots N_d) \times (N_1 \cdots N_d)}$ and index the rows and columns with "field" index:

$$\Sigma[\underbrace{(m_1, \ldots, m_d)}_{\text{row } m_1 \cdots m_d}, \underbrace{(n_1, \ldots, n_d)}_{\text{column } n_1 \cdots n_d}]$$

Suppose Σ has nested block circulant with circulant blocks, i.e.

$$\Sigma[:, (n_1, \ldots, n_d)] = P_{N_1}^{(n_1 - 1)} \otimes \cdots \otimes P_{N_d}^{(n_d - 1)} \overbrace{\Sigma[:, (1, \ldots, 1)]}^{\text{first column}}$$

Then

$$\Sigma = \sum_{n_1, \ldots, n_d = 1}^{N_1, \ldots, N_d} \Sigma[(n_1, \ldots, n_d), (1, \ldots, 1)] \, P_{N_1}^{(n_1 - 1)} \otimes \cdots \otimes P_{N_d}^{(n_d - 1)}$$

Therefore

$$\Sigma = \left(\mathcal{U}_{N_1}^H \otimes \cdots \otimes \mathcal{U}_{N_d}^H \right) \Lambda \left(\mathcal{U}_{N_1} \otimes \cdots \otimes \mathcal{U}_{N_d} \right)$$

where Λ is a diagonal matrix with

diagonal of $\Lambda \equiv W_{N_1} \otimes \cdots \otimes W_{N_d} \, \Sigma[:, (1, \ldots, 1)]$

So

$$\Sigma = \mathcal{F}_{N_1}^{-1} \otimes \cdots \otimes \mathcal{F}_{N_d}^{-1} \left(\cdots \right) \mathcal{F}_{N} \otimes \cdots \otimes \mathcal{F}_{N_d}$$

$$\operatorname{diag}\left(\frac{(2\pi)^{d/2}}{\sqrt{2\pi}} \mathcal{F}_{N_1} \otimes \cdots \otimes \mathcal{F}_{N_d} \, \Sigma[:, (1, \ldots, 1)] \right)$$

and so

$$\mathcal{F}_{N_1} \otimes \cdots \otimes \mathcal{F}_{N_d} \; \Sigma \; \mathcal{F}_{N_1}^H \otimes \cdots \otimes \mathcal{F}_{N_d}^H$$

$$= \operatorname{diag}\left(\frac{(2\pi)^{d/2}}{\sqrt{2\pi}} \mathcal{F}_{N} \otimes \cdots \otimes \mathcal{F}_{N_d} \, \Sigma[:, (1, \ldots, 1)] \right)$$

Field notation for spatial vrs Fourier indexing

For a field $f(x_1,\dots,x_d)$ indexed by spatial coordinates

$$(x_1,\dots,x_d)^T \in [0,P_1)\times\cdots\times[0,P_d)$$

let $f[x_1,\dots,x_d]$ be short hand for

$$f[x_1,\dots,x_d] \equiv f\left((n_1-1)\frac{P_1}{N_1},\dots,(n_d-1)\frac{P_d}{N_d}\right)$$

square brackets with $x_i\in\mathbb{R}$ denoted by english letters

with flattened coordinate slicing denoted with ":".

Notice that $N_1,\dots,N_d$ are implicit from context.

Example: $f[:]\in\mathbb{R}^{(N_1\cdots N_d)}$

$f[x_1,:,x_3,\dots,x_d]\in\mathbb{R}^{N_2}$

$f[x_1,:,:,\dots,x_d]\in\mathbb{R}^{N_2\times N_3}$

The indexing of the output of d-DFT is done with greek letters + square brackets

$$\tau_j = (k_j-1)\frac{2\pi}{P_j}$$

$$f[\tau_1,\dots,\tau_d] = \left(\mathcal{F}_{N_1,\dots,N_d} f[:,\dots,:]\right)\left[1+\frac{P_1}{2\pi}\tau_1,\dots,1+\frac{P_d}{2\pi}\tau_d\right]$$

square bracket indexing with $\tau_i\in\mathbb{R}$ using greek letters

with flattened coordinate slicing of a index τ_j written as "!", i.e.

$$\underbrace{f[!,\dots,!]}_{\text{Fourier transformed array}} := \left(\mathcal{F}_{N_1,\dots,N_d} \underbrace{f[:,\dots,:]}_{\text{Real space array}}\right)$$

Example:

$$\mathcal{F}_{N_1}\otimes\cdots\otimes\mathcal{F}_{N_d}(f[:]) = \left(\mathcal{F}_{N_1,\dots,N_d} f[:,\dots,:]\right)[!]$$

equality here, not just '$\equiv$'.

We can also mix spatial/Fourier indexing

Example:

For $d=3$, let $x_j=(n_j-1)\frac{P_j}{N_j}$ and $\tau_j=(k_j-1)\frac{2\pi}{P_j}$. Then

$$f[x_1,\tau_2,x_3] = \left(I_{N_1}\otimes\mathcal{F}_{N_2}\otimes I_{N_3} f[:]\right)[n_1,k_2,n_3]$$

at the unflattened coordinate $(n_1,k_2,n_3)^T$

$$f[!,x_2,:] = \left(\mathcal{F}_{N_1}\otimes I_{N_2}\otimes I_{N_3} f[:]\right)[:,n_2,:]$$

$$f[!,!,!] = \Big[\left(\mathcal{F}_{N_1}\otimes I_{N_2}\otimes I_{N_3}\right) \times \left(I_{N_1}\otimes \mathcal{F}_{N_2}\otimes I_{N_3}\right) \times \left(I_{N_1}\otimes I_{N_2}\otimes\mathcal{F}_{N_3}\right) f[:]\Big][:,:,:]$$

$$= \Big[\left(\mathcal{F}_{N_1}\otimes\mathcal{F}_{N_2}\otimes\mathcal{F}_{N_3}\right) f[:]\Big][:,:,:]$$

New field indexing notation

Old array indexing notation

Field operator notation for operators diagonal in Fourier or Spatial coordinates

When working with random field data one has a choice of working in the spatial domain or the fourier domain.

Example

Suppose field observations are taken on grid $\left((n_1-1)\frac{P_1}{N_1},\dots,(n_d-1)\frac{P_d}{N_d}\right) \in [0,P_1)\times\cdots\times[0,P_d)$ with $n_i \in \{1,\dots,N_i\}$.

Let $\mathcal{F} \equiv \mathcal{F}_{N_1}\otimes\cdots\otimes\mathcal{F}_{N_d}$.

Now the generic data model

$$\underbrace{d}_{\text{data field}} = \overbrace{M}^{\text{Pixel mask}}\ \overbrace{B}^{\text{Beam}}\ \underbrace{f}_{\text{signal field}} + \underbrace{\varepsilon}_{\text{noise field}} \qquad (*)$$

can be understood in spatial coordinates

$$d[:] = M\,B\,f[:] + \varepsilon[:]$$

where

$$\left.\begin{aligned} M &= \mathrm{diag}(m[:]) \\ B &= \mathcal{F}^{-1}\,\mathrm{diag}(b[!])\,\mathcal{F}. \end{aligned}\right\} (1)$$

Remark:

Often m and b are given by something like...

$$m[x_1,\dots,x_d] = \begin{cases} 1 & \text{if pixel } x \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$

$$b[\tau_1,\dots,\tau_d] = \exp\left(-\frac{\sigma_b^2}{2}|\tau|^2\right)$$

$$\sigma_b^2 = \frac{\sigma_{FWHM}^2}{8\ln(2)}$$

Note: $Bf(y) = \int \frac{1}{(2\pi)^{d/2}}\, b(x-y) f(y)\, dy_1\dots dy_d$ since $b[k_1,\dots,k_d]$ is defined with $(2\pi)^{d/2}$ pre-multiplied.

By multiplying both sides by $\mathcal{F}$ (on the left) we have

$$d[!] = \mathcal{F} M B f[:] + \varepsilon[!]$$

$$= \underbrace{\mathcal{F} M \mathcal{F}^{-1}}_{\substack{\text{models how} \\ \text{a pixel} \\ \text{mask operates} \\ \text{on Fourier coeffs}}}\ \underbrace{\mathrm{diag}(b[!])}_{\substack{\text{beam is} \\ \text{diagonal} \\ \text{on Fourier} \\ \text{coeffs}}}\ \underbrace{\mathcal{F} f[:]}_{f[!]} + \varepsilon[!]$$

Therefore $(*)$ can also be understood as a model in Fourier coordinates

$$d[!] = M\,B\,f[!] + \varepsilon[!]$$

where now

$$\left.\begin{aligned} M &= \mathcal{F}\,\mathrm{diag}(m[:])\,\mathcal{F}^{-1} \\ B &= \mathrm{diag}(b[!]) \end{aligned}\right\} (2)$$

Notation :

In what follows we do not distinguish (1) from (2), treating them as the same operator, just applied differently depending on the coordinates of the field they are acting on.

The notation

$$\mathbb{M} = \mathrm{Diagonal}(m[:])$$
$$\mathbb{B} = \mathrm{Diagonal}(b[!])$$

indicates this equivalence so that

$$(\mathbb{M})(f[:]) = \mathrm{diag}(m[:])\, f[:]$$
$$(\mathbb{M})(f[!]) = \mathcal{F}\, \mathrm{diag}(m[:])\, \mathcal{F}^{-1} f[!]$$

and

$$(\mathbb{B})(f[:]) = \mathcal{F}^{-1} \mathrm{diag}(b[!])\, \mathcal{F} f[:]$$
$$(\mathbb{B})(f[!]) = \mathrm{diag}(b[!])\, f[!]$$

so that

$$d = \mathbb{M}\mathbb{B} f + \varepsilon$$

can be interpreted either in spatial or Fourier coefficients.

Field/operator form of the loglikelihood w.r.t a stationary random field

Lets check that our notation allows consistent evaluation of likelihoods for stationary/isotropic models.

In particular, compare

$$\log P(d[:], f[:]) \quad \text{vrs}$$
$$\log P(d[!], f[!])$$

for data of the form

$$d = \mathbb{M}\mathbb{B} f + \varepsilon$$

with

$$\mathbb{M} = \mathrm{Diagonal}(m[:])$$
$$\mathbb{B} = \mathrm{Diagonal}(b[!])$$

where the covariance models on f and ε are circulant so that the spatial vrs Fourier models are:

$$f[:] \sim N\left(0,\ \mathcal{F}^{-1} \mathrm{diag}\left(\frac{C^{ff}[!]}{\Omega x}\right) \mathcal{F}\right) \quad \text{vrs}$$

$$f[!] \sim N\left(0,\ \mathrm{diag}\left(\frac{C^{ff}[!]}{\Omega k}\right)\right) \quad \text{vrs}$$

where $C^{ff}[x_1,\dots,x_d] = (2\pi)^{d/2} \mathrm{cov}(f[x_1,\dots,x_d], f[y_1,\dots,y_d])$ where $y_1 = \cdots = y_d = 0$

and

$$\varepsilon[:] \sim N\left(0,\ \frac{\sigma^2}{\Omega x} I\right) \quad \text{vrs}$$

$$\varepsilon[!] \sim N\left(0,\ \frac{\sigma^2}{\Omega k} I\right)$$

so that

$$\log P(d[:], f[:])$$

$$= \log P(d[:]/f[:]) + \log P(f[:])$$

diag(m[:])

$$= -\frac{1}{2} \frac{\| d[:] - \mathbb{M}\mathbb{B} f[:] \|^2}{(\sigma^2/\Omega x)} \quad \mathbb{F}^{-1}\mathrm{diag}(b[!])\mathbb{F}$$

$$-\frac{1}{2} f[:]^T \left(\mathbb{F}^{-1} \mathrm{diag}\left(\frac{C^{ff}[!]}{\Omega x}\right) \mathbb{F}\right)^{-1} f[:]$$

+ constant (which I will drop)

$$= -\frac{1}{2} \frac{\| \mathbb{F}^{-1}(d[!] - \mathbb{F}\mathbb{M}\mathbb{B} f[:]) \|^2}{\sigma^2/\Omega x}$$

$$-\frac{1}{2} (\mathbb{F}^{-H} f[:])^H \mathrm{diag}\left(\frac{\Omega x}{C^{ff}[!]}\right)(\mathbb{F} f[:])$$

Notice that since $\mathbb{F}\mathbb{F}^H = \frac{\Omega x}{\Omega k} I$ we have that

$$\| \mathbb{F}^{-1} g[!] \|^2 = g[!]^H \mathbb{F}^{-H} \mathbb{F}^{-1} g[!]$$

$$= \frac{\Omega k}{\Omega x} \| g[!] \|^2$$

and

$$\mathbb{F}^{-H} f[:] = \frac{\Omega k}{\Omega x} \mathbb{F} f[:]$$

Putting these together gives:

$$\log P(d[:], f[:])$$

diag(m[:])

$$= -\frac{1}{2} \frac{\| d[!] - \mathbb{F}\mathbb{M}\mathbb{B} f[:] \|^2}{(\sigma^2/\Omega k)} \quad \mathbb{F}^{-1}\mathrm{diag}(b[!])\mathbb{F}$$

$$-\frac{1}{2} f[!]^H \mathrm{diag}\left(\frac{\Omega k}{C^{ff}[!]}\right) f[!]$$

$\mathbb{F}\,\mathrm{diag}(m[:])\,\mathbb{F}^{-1}$

$$= -\frac{1}{2} \frac{\| d[!] - \mathbb{M}\mathbb{B} f[!] \|^2}{(\sigma^2/\Omega k)} \quad \mathrm{diag}(b[!])$$

$$-\frac{1}{2} f[!]^H \mathrm{diag}\left(\frac{C^{ff}[!]}{\Omega k}\right)^{-1} f[!]$$

So yes, we have

$$\underbrace{\log P(f[:]/d[:])}_{\text{using one form of } \mathbb{M} \,\&\, \mathbb{B}} = \underbrace{\log P(f[!]/d[!])}_{\text{using the other form of } \mathbb{M} \,\&\, \mathbb{B}.}$$

In fact, it is clear from the above derivation that the whole log-likelihood can be written with Field/operator notation.

Lets write it all out on the next page for refence...

data model

$$d = \mathbb{M}\,\mathbb{B}\,f + \varepsilon$$

Pixel mask

$$\mathbb{M} = \mathrm{Diagonal}(m[:])$$

$$m[x_1,\dots,x_d] = \begin{cases} 1 & \text{if pixel } x \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$

Beam

$$\mathbb{B} = \mathrm{Diagonal}(b[!])$$

$$b[\tau_1,\dots,\tau_d] = \exp\left(-\frac{\sigma_b^2}{2}|\tau|^2\right)$$

$$\sigma_b^2 = \frac{\sigma_{FWHM}^2}{8\ln(2)}$$

white noise model

$$\mathbb{C}^{\varepsilon\varepsilon} = \mathrm{Diagonal}(c^{\varepsilon\varepsilon}[!])$$

$$C^{\varepsilon\varepsilon}[x_1,\dots,x_d] = \begin{cases} (2\pi)^{d/2}\frac{\sigma^2}{\Omega x} & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$C^{\varepsilon\varepsilon}[!] = (\sigma^2, \sigma^2, \dots, \sigma^2)^T$$

$$\varepsilon[:] \sim N\left(0, \frac{1}{\Omega x}\mathbb{C}^{\varepsilon\varepsilon}\right)$$ version that operates on $g[:]$

$$\varepsilon[!] \sim N\left(0, \frac{1}{\Omega\tau}\mathbb{C}^{\varepsilon\varepsilon}\right)$$ version that operates on $g[!]$

Circulant signal model

$$\mathbb{C}^{ff} = \mathrm{Diagonal}(c^{ff}[!])$$

$$C^{ff}[x_1,\dots,x_d] = (2\pi)^{d/2}\,\mathrm{cov}(f[x_1,\dots,x_d], f[1,\dots,1])$$

$$f[:] \sim N\left(0, \frac{1}{\Omega x}\mathbb{C}^{ff}\right)$$ version that operates on $g[:]$

$$f[!] \sim N\left(0, \frac{1}{\Omega\tau}\mathbb{C}^{ff}\right)$$ version that operates on $g[!]$

Fourier log-likelihood

$$\log P(d[!], f[!]) = -\frac{\Omega\tau}{2}\left\|(\mathbb{C}^{\varepsilon\varepsilon})^{-1/2}(d[!] - \mathbb{M}\mathbb{B}f[!])\right\|^2 - \frac{\Omega\tau}{2}\left\|(\mathbb{C}^{ff})^{-1/2}f[!]\right\|^2$$

Spatial Pixel log-likelihood

$$\log P(d[:], f[:]) = -\frac{\Omega x}{2}\left\|(\mathbb{C}^{\varepsilon\varepsilon})^{-1/2}(d[:] - \mathbb{M}\mathbb{B}f[:])\right\|^2 - \frac{\Omega x}{2}\left\|(\mathbb{C}^{ff})^{-1/2}f[:]\right\|^2$$

$$\boxed{\log P(d[!], f[!]) = \log P(d[:], f[:])}$$

Sanity check

Lets unpack the statement

$$(*) \qquad \varepsilon[:] \sim N\left(0, \frac{1}{\Omega x}\mathbb{C}^{\varepsilon\varepsilon}\right)$$

the version that operates on $g[:]$

where

$$\mathbb{C}^{\varepsilon\varepsilon} = \text{Diagonal}\left(c^{\varepsilon\varepsilon}[!]\right)$$

$$c^{\varepsilon\varepsilon}[x_1, \ldots, x_d] = \begin{cases} (2\pi)^{d/2}\dfrac{\sigma^2}{\Omega x} & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

and make sure it is consistent with our circulant matrix theory, i.e.

$$\frac{\sigma^2}{\Omega x} I = \Sigma^{\varepsilon\varepsilon}$$

$$= \mathcal{F}^{-1}\text{diag}\left(\frac{(2\pi)^{d/2}}{\Omega x}\mathcal{F}\,\Sigma^{\varepsilon\varepsilon}[:,1]\right)\mathcal{F}$$

where $\Sigma^{\varepsilon\varepsilon} = E\left(\varepsilon[:]\,\varepsilon[:]^T\right)$

In $(*)$ above $\mathbb{C}^{\varepsilon\varepsilon}$ is the spatial coordinate version, i.e.

$$\mathbb{C}^{\varepsilon\varepsilon} = \mathcal{F}^{-1}\text{diag}\left(c^{\varepsilon\varepsilon}[!]\right)\mathcal{F}$$

where

$$c^{\varepsilon\varepsilon}[:] = (2\pi)^{d/2}\frac{\sigma^2}{\Omega x}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

← first index (i.e. @ $x=0$) is the only non-zero

$$\therefore\; c^{\varepsilon\varepsilon}[!] = \mathcal{F}\, c^{\varepsilon\varepsilon}[:] = \sigma^2\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

The first column of $\mathcal{F}_{N_1}\otimes\cdots\otimes\mathcal{F}_{N_d}$ is $\left(\Omega x/(2\pi)^{d/2}\right)\cdot(1,\ldots,1)^T$

Now check this is consistent with our circulant matrix theory.

from circulant theory

$$\Sigma^{\varepsilon\varepsilon} = \mathcal{F}^{-1}\text{diag}\left(\frac{(2\pi)^{d/2}}{\Omega x}\mathcal{F}\,\Sigma^{\varepsilon\varepsilon}[:,1]\right)\mathcal{F}$$

$$= \mathcal{F}^{-1}\text{diag}\Bigg(\frac{1}{\Omega x}\underbrace{\mathcal{F}\underbrace{(2\pi)^{d/2}\Sigma^{\varepsilon\varepsilon}[:,1]}_{= c^{\varepsilon\varepsilon}[:] = \frac{(2\pi)^{d/2}}{\Omega x}(\sigma^2, 0, \ldots, 0)^T}}_{= c^{\varepsilon\varepsilon}[!] = (\sigma^2, \ldots, \sigma^2)^T}\Bigg)\mathcal{F}$$

$$= \frac{1}{\Omega x}\mathbb{C}^{\varepsilon\varepsilon}$$

The spatial coordinate version of this operator

So, indeed, everything checks out.

$E(f|d)$, $var(f|d)$ and $\log P(f|d)$ for field model $d = \mathbb{A}f + \mathbb{B}\varepsilon$

Consider the following generic field model:

$$d = \mathbb{A}f + \mathbb{B}\varepsilon$$

$$f \sim N\left(0, \frac{\mathbb{C}^{ff}}{\Omega}\right) \quad \text{(signal)}$$

$$\varepsilon \sim N\left(0, \frac{\mathbb{C}^{\varepsilon\varepsilon}}{\Omega}\right) \quad \text{(noise)}$$

where

- $\Omega :=$ coordinate (spatial or Fourier) grid volumn
$$= \begin{cases} \Omega_x & \text{if modeling } f[:] \\ \Omega_k & \text{if modeling } f[!] \end{cases}$$

- $\mathbb{C}^{ff} = \text{Diagonal}(c^{ff}[!])$
  ($c^{ff} \equiv$ the spectral density for f)

- $\mathbb{C}^{\varepsilon\varepsilon} = \text{Diagonal}(c^{\varepsilon\varepsilon}[!])$
  ($c^{\varepsilon\varepsilon} \equiv$ the spectral density for ε)

- $\mathbb{A}$ and $\mathbb{B}$ are generic linear operators with $\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^{H}$ inventable

Remark:

Notice that there are multiple ways to write the model for f & ε ...

For example,

$$\begin{cases} f = \text{true CMB} \sim GRF\left(0, \text{auto-cov } \frac{C^{ff}(x_1,x_2)}{2\pi}\right) \\ \varepsilon = \text{instrument noise} \sim GRF\left(0, \text{auto-cov } \frac{C^{\varepsilon\varepsilon}(x_1,x_2)}{2\pi}\right) \end{cases}$$

... or

$$\begin{cases} f[!] = \text{true CMB} \sim N\left(0, \frac{\mathbb{C}^{ff}}{\Omega_k}\right) \\ \varepsilon[!] = \text{instrument noise} \sim N\left(0, \frac{\mathbb{C}^{\varepsilon\varepsilon}}{\Omega_k}\right) \end{cases}$$

or

$\begin{cases} f \text{ and } \varepsilon \text{ are mean zero stationary Gaussian random fields with spectral densities } c^{ff} \text{ and } c^{\varepsilon\varepsilon}\text{, respectively} \end{cases}$

... etc $\square$

The likelihood of d given f is clearly

$$d|f \sim N\left(\mathbb{A}f, \frac{\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^{H}}{\Omega}\right)$$

what we want is

$$f|d \sim N\left(E(f|d), var(f|d)\right).$$

An easy trick to find these is to notice that since $P(d,f)$ is Gaussian

$$\nabla_f \log P(d,f)\Big|_{f=f^\circ} = 0 \iff f^\circ = E(f|d)$$

also note that

$$\underbrace{\left[-\frac{\partial}{\partial f}\nabla_f \log P(d,f)\right]^{-1}}_{\text{this is the same evaluated at any } f \text{ since } \log P(d,f) \text{ is quadratic.}} = var(f|d)$$

- Hessian

Since we are assuming that $\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H$ is invertable $\log P(d,f)$ can be computed as follows:

$$\log P(d,f) = \log P(d|f) + \log P(f)$$
$$= -\frac{\Omega}{2}(d-\mathbb{A}f)^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}(d-\mathbb{A}f)$$
$$-\frac{\Omega}{2}f^H(\mathbb{C}^{ff})^{-1}f$$

Also notice that $\frac{\partial}{\partial f}\log P(d,f)$ is a "row" and operates on displacement "vectors" δf, i.e.

$$\underbrace{\lim_{\varepsilon\to 0}\frac{\log P(d,f+\varepsilon\delta f)-\log P(d,f)}{\varepsilon}}_{\in\mathbb{R}}$$
$$= \nabla_f \log P(d,f)^H\cdot\delta f$$
$$= \left[\frac{\partial}{\partial f}\log P(d,f)\right]\delta f$$

where

$$\frac{\partial}{\partial f}\log P(d,f) = \frac{\partial}{\partial f}\log P(d|f) + \frac{\partial}{\partial f}\log P(f)$$
$$= \frac{\partial}{\partial f}\left(-\frac{\Omega}{2}(d-\mathbb{A}f)^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}(d-\mathbb{A}f)\right)$$
$$+\frac{\partial}{\partial f}\left(-\frac{\Omega}{2}f^H(\mathbb{C}^{ff})^{-1}f\right)$$
$$= +\Omega(d-\mathbb{A}f)^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A}$$
$$-\Omega f^H(\mathbb{C}^{ff})^{-1}$$

Notice this eats δf and spits out a number

So that

$$\nabla_f\log P(d,f) = \left[\frac{\partial}{\partial f}\log P(d,f)\right]^H$$
$$= \Omega\,\mathbb{A}^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}(d-\mathbb{A}f)$$
$$-\Omega(\mathbb{C}^{ff})^{-1}f$$

Notice this is a field.

$$= \Omega\,\mathbb{A}^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}d$$
$$-\Omega\left[\mathbb{A}^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A}+(\mathbb{C}^{ff})^{-1}\right]f$$

Therefore

$$\nabla_f\log P(d,f)\Big|_{f=f^0} = 0$$
$$\Updownarrow$$
$$\left[\mathbb{A}^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A}+(\mathbb{C}^{ff})^{-1}\right]f^0 = \mathbb{A}^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}d$$
$$\Updownarrow$$
$$\underbrace{\left[\mathbb{A}^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A}+(\mathbb{C}^{ff})^{-1}\right]^{-1}\mathbb{A}^H(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}d}_{=E(f|d)}$$

Also notice

- Hessian of $\log P(d,f)$ w.r.t. f

$$= -\frac{\partial}{\partial f} \nabla_f \log P(d,f)$$

$$= \frac{\partial}{\partial f}\Big(-\Omega \cancel{\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1} d} \xrightarrow{} 0$$
$$+ \Omega \left[\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A} + (\mathbb{C}^{ff})^{-1}\right] f\Big)$$

$$= \Omega \left[\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A} + (\mathbb{C}^{ff})^{-1}\right]$$

Invert this to obtain the conditional var/cov of $f \mid d$.

$$\mathrm{var}(f \mid d) = \frac{\left[\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A} + (\mathbb{C}^{ff})^{-1}\right]^{-1}}{\Omega}$$

(Notice that the coordinate pixel volumn is in the correct location)

There is also a trick for simulating

$$f \mid d \sim N\Big(E(f \mid d),\ \mathrm{var}(f \mid d)\Big)$$

in the case above. First simulate from the prior and noise model

$$\left.\begin{aligned} \tilde{f} &\sim N\Big(0, \frac{\mathbb{C}^{ff}}{\Omega}\Big) \\ \tilde{\varepsilon} &\sim N\Big(0, \frac{\mathbb{C}^{\varepsilon\varepsilon}}{\Omega}\Big) \end{aligned}\right\} \text{independent simulations}$$

Notice that

$$\mathrm{var}\Big(\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{B}\tilde{\varepsilon}\Big)$$
$$= \mathbb{A}^H \cancel{(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}} \cancel{\frac{(\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)}{\Omega}}$$
$$\times (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A}$$
$$= \frac{\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A}}{\Omega}$$

and that

$$\mathrm{var}\big((\mathbb{C}^{ff})^{-1}\tilde{f}\big)$$
$$= \cancel{(\mathbb{C}^{ff})^{-1}} \cancel{\frac{\mathbb{C}^{ff}}{\Omega}} (\mathbb{C}^{ff})^{-1}$$
$$= \frac{(\mathbb{C}^{ff})^{-1}}{\Omega}$$

which implies

$$\mathrm{var}\Big(\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{B}\tilde{\varepsilon} + (\mathbb{C}^{ff})^{-1}\tilde{f}\Big)$$
$$= \frac{\left[\mathbb{A}^H (\mathbb{B}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{B}^H)^{-1}\mathbb{A} + (\mathbb{C}^{ff})^{-1}\right]}{\Omega}$$

Now adjust the linear model which characterizes $E(f|d)$

$$\left[A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} A + (C^{ff})^{-1}\right] E(f|d) = A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} d$$

as follows

$$\left[A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} A + (C^{ff})^{-1}\right] f^{sim} = A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} d + A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} B \tilde{\varepsilon} + (C^{ff})^{-1} \tilde{f}$$

which satisfies

$$f^{sim} = E(f|d) + \mathcal{E}(\tilde{f}, \tilde{\varepsilon})$$

with

$$\mathcal{E}(\tilde{f}, \tilde{\varepsilon}) \sim N\left(0, \frac{\left[A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} A + (C^{ff})^{-1}\right]^{-1}}{\Omega}\right)$$

so that

$$f^{sim} \sim N\big(E(f|d), \operatorname{var}(f|d)\big)$$

as desired.

In the above derivation we have been assuming $B C^{\varepsilon\varepsilon} B^H$ was invertable which allows us to compute the Gaussian likelihood of the conditional

$$d|f \sim N(Af, B C^{\varepsilon\varepsilon} B^H)$$

It is possible to extend this in the case we know how to compute the

$$(B C^{\varepsilon\varepsilon} B^H)^{-1} \leftarrow \text{Generalized inverse in red.}$$

where Σ^{-1} denotes the generalized inverse which satisfies

$$\Sigma \Sigma^{-1} \Sigma = \Sigma$$

In this case we have

$$E(f|d) = \left[A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} A + (C^{ff})^{-1}\right]^{-1} \times A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} d$$

$$\operatorname{var}(f|d) = \frac{\left[A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} A + (C^{ff})^{-1}\right]^{-1}}{\Omega}$$

$$f^{sim} = \left[A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} A + (C^{ff})^{-1}\right]^{-1} \times \left[A^H (B C^{\varepsilon\varepsilon} B^H)^{-1} (d + B\tilde{\varepsilon}) + (C^{ff})^{-1} \tilde{f}\right]$$

where $\tilde{\varepsilon} \sim N\left(0, \frac{C^{\varepsilon\varepsilon}}{\Omega}\right)$, $\tilde{f} \sim N\left(0, \frac{C^{ff}}{\Omega}\right)$

South Pole Telescope Data set

Using our field notation we can write the model for the data field $d(x_1, x_2)$ indexed by spatial coordinates $(x_1, x_2)^T$ measured in radians and restricted to

$$[0, P_1) \times [0, P_2)$$

and observed on a grid

$$\left\{ \left((n_1 - 1)\tfrac{P_1}{N_1}, (n_2 - 1)\tfrac{P_2}{N_2}\right)^T : n_j = 1, \ldots, N_j, j = 1, 2 \right\}$$

as follows

$$d = \mathbb{M}\mathbb{T}f + \mathbb{M}\mathbb{W}\varepsilon$$

where

$\mathbb{M} = \text{Diagonal}(m[:]) \equiv$ pixel mask
(models which pixels are observed)

$\mathbb{T} = \text{Diagonal}(t[!]) \equiv$ transfer function
(models the instrument, processing/distortion and beam)

$\mathbb{W} = \text{Diagonal}(w[:]) =$ pixel weight
(models pixel noise s.d. due to to things like the time spent observin a pixel)

f = true CMB $= (\mathbb{C}^{ff})^{1/2}$ white noise

ε = instrument noise $= (\mathbb{C}^{\varepsilon\varepsilon})^{1/2}$ white noise

independent & Gaussian

$\mathbb{C}^{ff} = \text{Diagonal}(c^{ff}[!])$
($c^{ff} \equiv$ the spectral density for f)

$\mathbb{C}^{\varepsilon\varepsilon} = \text{Diagonal}(c^{\varepsilon\varepsilon}[!])$
($c^{\varepsilon\varepsilon} \equiv$ the spectral density for ε)

Notice this model has the form

$$d = \underbrace{\mathbb{A}}_{\mathbb{M}\mathbb{T}}f + \underbrace{\mathbb{B}}_{\mathbb{M}\cdot\mathbb{W}}\varepsilon$$

But since $m[x_1 x_2] = 0$ for some pixel values we would need the generalized matrix inverse $(\mathbb{M}\mathbb{W}\mathbb{C}^{\varepsilon\varepsilon}\mathbb{W}\mathbb{M})^{-1}$ to compute $E(f|d)$ and do conditional simulation $f|d$.

This is typically difficult unless $\mathbb{C}^{\varepsilon\varepsilon} = \sigma^2 \mathbb{I}$, in which case

$$(\mathbb{M}\mathbb{W}(\sigma^2\mathbb{I})\mathbb{M}\mathbb{W})^{-1} = \frac{1}{\sigma^2}\text{Diagonal}\left(\frac{1}{(m[:]w[:])^2}\mathbb{1}(m[:]w[:] \neq 0)\right)$$

For the south pole data $\mathbb{C}^{\varepsilon\varepsilon} \neq \sigma^2\mathbb{I}$ but we can add noise to the data to obtain this as follows.

Let $\sigma^2 > c^{\varepsilon\varepsilon}[\tau_1, \tau_2]$ for all (grid) frequencies (τ_1, τ_2).

Let

$$\varepsilon^c \sim N\left(0, \frac{\sigma^2\mathbb{I} - \mathbb{C}^{\varepsilon\varepsilon}}{\Omega}\right)$$

and define

$$d^c = d + \mathbb{M}\mathbb{W}\varepsilon^c$$
$$\overset{\mathcal{D}}{=} \mathbb{M}\mathbb{T}f + \mathbb{M}\mathbb{W}(\varepsilon + \varepsilon^c)$$

where $\varepsilon + \varepsilon^c$ has cov $\dfrac{\sigma^2\mathbb{I} - \mathbb{C}^{\varepsilon\varepsilon} + \mathbb{C}^{\varepsilon\varepsilon}}{\Omega} = \dfrac{\sigma^2}{\Omega}$

Now setting $\mathbb{A} = \mathbb{M}\mathbb{T}$, $\mathbb{B} = \mathbb{M}\mathbb{W}$ and noticing

$$\mathbb{B}\left(\frac{\sigma^2}{\Omega}\right)\mathbb{B}^H = \sigma^2 \mathbb{M}^2 \mathbb{W}^2$$

we have

$$E(f|d^c) = \left[\mathbb{A}^H(\sigma^2\mathbb{M}^2\mathbb{W}^2)^{-1}\mathbb{A} + (\mathbb{C}^{ff})^{-1}\right]^{-1} \times \mathbb{A}^H(\sigma^2\mathbb{M}^2\mathbb{W}^2)^{-1} d^c$$

$$f^{sim} = \left[\mathbb{A}^H(\sigma^2\mathbb{M}^2\mathbb{W}^2)^{-1}\mathbb{A} + (\mathbb{C}^{ff})^{-1}\right]^{-1} \times \left[\mathbb{A}^H(\sigma^2\mathbb{M}^2\mathbb{W}^2)^{-1}(d^c + \mathbb{M}\mathbb{W}\tilde{\varepsilon}) + (\mathbb{C}^{ff})^{-1}\tilde{f}\right]$$

where $\tilde{\varepsilon} \sim N\left(0, \frac{\sigma^2 I}{\Omega}\right)$, $\tilde{f} \sim N\left(0, \frac{\mathbb{C}^{ff}}{\Omega}\right)$