1 MAP solution with correlated responses

a)

Using Bishop p.78:

$$p(\mathcal{D} \mid \boldsymbol{\theta}) = \mathcal{N}(\mathbf{t} \mid \boldsymbol{\Psi}\mathbf{w}, \boldsymbol{\Omega}) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\boldsymbol{\Omega}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{t} - \boldsymbol{\Psi}\mathbf{w})^\top \boldsymbol{\Omega}^{-1}(\mathbf{t} - \boldsymbol{\Psi}\mathbf{w})\right\} \tag{1}$$

b)

Substituting $\boldsymbol{\Omega} = \mathbf{A}^\top \mathbf{D}\mathbf{A}$ gives:

$$p(\mathcal{D} \mid \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\mathbf{A}^\top\mathbf{D}\mathbf{A}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{t} - \boldsymbol{\Psi}\mathbf{w})^\top (\mathbf{A}^\top\mathbf{D}\mathbf{A})^{-1}(\mathbf{t} - \boldsymbol{\Psi}\mathbf{w})\right\} \tag{2}$$

Using the given properties of determinants and the matrix $\mathbf{A}$:

$$|\mathbf{A}^\top\mathbf{D}\mathbf{A}| = |\mathbf{A}^\top||\mathbf{D}||\mathbf{A}| = |\mathbf{A}^{-1}||\mathbf{D}||\mathbf{A}| = |\mathbf{D}||\mathbf{A}|^{-1}|\mathbf{A}| = |\mathbf{D}| \tag{3}$$

$$(\mathbf{A}^\top\mathbf{D}\mathbf{A})^{-1} = \mathbf{A}^{-1}\mathbf{D}^{-1}(\mathbf{A}^\top)^{-1} = \mathbf{A}^\top\mathbf{D}^{-1}(\mathbf{A}^{-1})^{-1} = \mathbf{A}^\top\mathbf{D}^{-1}\mathbf{A} \tag{4}$$

Eq. (2) then becomes:

$$p(\mathcal{D} \mid \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\mathbf{D}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{t} - \boldsymbol{\Psi}\mathbf{w})^\top \mathbf{A}^\top\mathbf{D}^{-1}\mathbf{A}(\mathbf{t} - \boldsymbol{\Psi}\mathbf{w})\right\} \tag{5}$$

$$= \frac{1}{(2\pi)^{N/2}} \frac{1}{|\mathbf{D}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{A}\mathbf{t} - \mathbf{A}\boldsymbol{\Psi}\mathbf{w})^\top \mathbf{D}^{-1}(\mathbf{A}\mathbf{t} - \mathbf{A}\boldsymbol{\Psi}\mathbf{w})\right\} \tag{6}$$

$$= \frac{1}{(2\pi)^{N/2}} \frac{1}{|\mathbf{D}|^{1/2}} \exp\left\{ -\frac{1}{2}\underbrace{(\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w})^\top \mathbf{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w})}_{=\boldsymbol{\Delta}^2}\right\}, \tag{7}$$

where $\boldsymbol{\tau} = \mathbf{A}\mathbf{t}$ and $\boldsymbol{\Phi} = \mathbf{A}\boldsymbol{\Psi}$.

c)

Define $\mathbf{y} := (\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w})$ with $y_i = \tau_i - \mathbf{w}^\top\phi_i$. Then we see that $\boldsymbol{\Delta}^2$ in Eq. (7) becomes

$$\boldsymbol{\Delta}^2 = \mathbf{y}^\top\mathbf{D}^{-1}\mathbf{y} = \frac{y_1^2}{d_1} + \cdots + \frac{y_N^2}{d_N} = \sum_{i=1}^N \frac{y_i^2}{d_i}, \tag{8}$$

where $d_1, \ldots, d_N$ are the eigenvalues of $\boldsymbol{\Omega}$. We can therefore write

$$\exp\left\{-\frac{1}{2}\boldsymbol{\Delta}^{\mathbf{2}}\right\} = \exp\left\{-\frac{1}{2}\sum_{i=1}^N \frac{(\tau_i - \mathbf{w}^\top\phi_i)^2}{d_i}\right\} = \prod_{i=1}^N \exp\left\{-\frac{(\tau_i - \mathbf{w}^\top\phi_i)^2}{2d_i}\right\}. \tag{9}$$

Also,

$$\frac{1}{(2\pi)^{N/2}} = \frac{1}{(2\pi)^{1/2} \times \cdots \times (2\pi)^{1/2}} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}}, \tag{10}$$

$$\frac{1}{|\mathbf{D}|^{1/2}} = \prod_{i=1}^N \frac{1}{\sqrt{d_i}}. \tag{11}$$

So we can write:

$$p(\mathcal{D} \mid \boldsymbol{\theta}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi d_i}} \exp\left\{-\frac{(\tau_i - \mathbf{w}^\top\phi_i)^2}{2d_i}\right\} \tag{12}$$

$$= \prod_{i=1}^N \mathcal{N}(\tau_i \mid \mathbf{w}^\top\phi_i, \sqrt{d_i}) \tag{13}$$

d)

Given:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w} \mid \mathbf{0}, \alpha\mathbf{I}) \tag{14}$$
$$= \frac{1}{(2\pi)^{N/2}} \frac{1}{|\alpha\mathbf{I}|^{1/2}} \exp\Big\{ -\frac{1}{2}\mathbf{w}^\top(\alpha\mathbf{I})^{-1}\mathbf{w}\Big\} \tag{15}$$

Using that $|\alpha\mathbf{I}| = \alpha^N$ and $(\alpha\mathbf{I})^{-1} = \alpha^{-1}\mathbf{I}$, we get:

$$p(\mathbf{w}) = \frac{1}{(2\pi\alpha)^{N/2}} \exp\Big\{ -\frac{1}{2\alpha}\mathbf{w}^\top\mathbf{w}\Big\} \tag{16}$$

Taking the logarithm:

$$\ln p(\mathbf{w}) = \ln\frac{1}{(2\pi\alpha)^{N/2}} + \ln\exp\Big\{ -\frac{1}{2\alpha}\mathbf{w}^\top\mathbf{w}\Big\} \tag{17}$$
$$= \underbrace{-\frac{N}{2}\ln(2\pi\alpha)}_{=C} - \frac{1}{2\alpha}\mathbf{w}^\top\mathbf{w} \tag{18}$$
$$= -\frac{1}{2\alpha}\mathbf{w}^\top\mathbf{w} + C \tag{19}$$

e)

Using Bayes' rule:

$$p(\mathbf{w} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \mathbf{w})p(\mathbf{w})}{p(\mathcal{D})} \tag{20}$$
$$= \frac{\mathcal{N}(\mathbf{t} \mid \mathbf{\Psi}\mathbf{w}, \mathbf{\Omega})\mathcal{N}(\mathbf{w} \mid \mathbf{0}, \alpha\mathbf{I})}{\int_{\text{all }\mathbf{w}} \mathcal{N}(\mathbf{t} \mid \mathbf{\Psi}\mathbf{w}, \mathbf{\Omega})\mathcal{N}(\mathbf{w} \mid \mathbf{0}, \alpha\mathbf{I})\mathrm{d}\mathbf{w}} \tag{21}$$
$$= \frac{\mathcal{N}(\mathbf{t} \mid \mathbf{\Psi}\mathbf{w}, \mathbf{\Omega})\mathcal{N}(\mathbf{w} \mid \mathbf{0}, \alpha\mathbf{I})}{p(\mathbf{t} \mid \mathbf{\Psi}, \alpha, \mathbf{\Omega})} \tag{22}$$

f)

$$\ln p(\mathbf{w} \mid \mathcal{D}) = \ln\frac{p(\mathcal{D} \mid \mathbf{w})p(\mathbf{w})}{p(\mathcal{D})} = \ln p(\mathcal{D} \mid \mathbf{w}) + \ln p(\mathbf{w}) - \ln p(\mathcal{D}) \tag{23}$$

First we derive $\ln p(\mathbf{w} \mid \mathcal{D})$ in <u>matrix form</u> (here we continue with earlier derived Eqs. (7) and (19)):

$$\ln p(\mathcal{D} \mid \mathbf{w}) = \ln\left(\frac{1}{(2\pi)^{N/2}|\mathbf{D}|^{1/2}}\right) - \frac{1}{2}(\boldsymbol{\tau} - \mathbf{\Phi}\mathbf{w})^\top\mathbf{D}^{-1}(\boldsymbol{\tau} - \mathbf{\Phi}\mathbf{w}) \tag{24}$$
$$\tag{25}$$

Combining this with $\ln p(\mathbf{w})$ derived in 1d, we get:

$$\ln p(\mathbf{w} \mid \mathcal{D}) = -\frac{1}{2}(\boldsymbol{\tau} - \mathbf{\Phi}\mathbf{w})^\top\mathbf{D}^{-1}(\boldsymbol{\tau} - \mathbf{\Phi}\mathbf{w}) - \frac{1}{2\alpha}\mathbf{w}^\top\mathbf{w} + C, \tag{26}$$

where C contains all terms which are independent of $\mathbf{w}$.
Next we will derive $\ln p(\mathbf{w} \mid \mathcal{D})$ in <u>product form</u> (using Eqs. (12) and again (19), since the matrix and product form are the same for $p(\mathbf{w})$):

$$\ln p(\mathcal{D} \mid \mathbf{w}) = \ln\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi d_i}}\exp\Big\{ -\frac{(\tau_i - \mathbf{w}^\top\phi_i)^2}{2d_i}\Big\} \tag{27}$$
$$= \sum_{i=1}^{N}\left[\ln\frac{1}{\sqrt{2\pi d_i}} - \frac{1}{2d_i}(\tau_i - \mathbf{w}^\top\phi_i)^2\right] \tag{28}$$
$$= -\frac{1}{2}\sum_{i=1}^{N}\ln(2\pi d_i) - \sum_{i=1}^{N}\frac{1}{2d_i}(\tau_i - \mathbf{w}^\top\phi_i)^2 \tag{29}$$

Combining this with $\ln p(\mathbf{w})$ derived in 1d, we get:

$$\ln p(\mathbf{w} \mid \mathcal{D}) = -\sum_{i=1}^{N} \frac{1}{2d_i}(\tau_i - \mathbf{w}^\top \phi_i)^2 - \frac{1}{2\alpha}\mathbf{w}^\top \mathbf{w} + C, \tag{30}$$

where C contains all terms which are independent of $\mathbf{w}$.
If we want to find to full posterior distribution, we would have to work out the integral in Eq. (21), which is not the case if we want to find the MAP. Therefore, finding the MAP is much simpler.

g)

We will derive $\mathbf{w}_{\text{MAP}}$ using the matrix form of $\ln p(\mathbf{w} \mid \mathcal{D})$, given in Eq. (26). Using the properties given in section 5.5 of the book Mathematics for Machine Learning, we see that

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^\top \mathbf{w} = \mathbf{w}^\top(\mathbf{I} + \mathbf{I}^\top) = 2\mathbf{w}^\top \tag{31}$$

$$\frac{\partial}{\partial \mathbf{w}}(\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w})^\top \mathbf{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w}) = -2(\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w})^\top \mathbf{D}^{-1}\boldsymbol{\Phi} \tag{32}$$

So,

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{w}}\ln p(\mathbf{w} \mid \mathcal{D}) &= -\frac{1}{2}\frac{\partial}{\partial \mathbf{w}}(\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w})^\top \mathbf{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w}) - \frac{1}{2\alpha}\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^\top \mathbf{w} && (33)\\
&= (\boldsymbol{\tau} - \boldsymbol{\Phi}\mathbf{w})^\top \mathbf{D}^{-1}\boldsymbol{\Phi} - \frac{1}{\alpha}\mathbf{w}^\top && (34)\\
&= \boldsymbol{\tau}^\top \mathbf{D}^{-1}\boldsymbol{\Phi} - \mathbf{w}^\top \boldsymbol{\Phi}^\top \mathbf{D}^{-1}\boldsymbol{\Phi} - \frac{1}{\alpha}\mathbf{w}^\top && (35)
\end{aligned}$$

Setting $\frac{\partial}{\partial \mathbf{w}}\ln p(\mathbf{w} \mid \mathcal{D}) = 0$ gives:

$$\begin{aligned}
\boldsymbol{\tau}^\top \mathbf{D}^{-1}\boldsymbol{\Phi} &= \frac{1}{\alpha}\mathbf{w}_{\text{MAP}}^\top + \mathbf{w}_{\text{MAP}}^\top \boldsymbol{\Phi}^\top \mathbf{D}^{-1}\boldsymbol{\Phi} && (36)\\
&= \mathbf{w}_{\text{MAP}}^\top\left(\frac{1}{\alpha} + \boldsymbol{\Phi}^\top \mathbf{D}^{-1}\boldsymbol{\Phi}\right) && (37)\\
& && (38)
\end{aligned}$$

Now take the transpose of each side:

$$\begin{aligned}
\left(\frac{1}{\alpha} + \boldsymbol{\Phi}^\top \mathbf{D}^{-1}\boldsymbol{\Phi}\right)^\top \mathbf{w}_{\text{MAP}} &= \boldsymbol{\Phi}^\top \mathbf{D}^{-1}\boldsymbol{\tau} && (39)\\
\mathbf{w}_{\text{MAP}} &= \left(\frac{1}{\alpha} + \boldsymbol{\Phi}^\top \mathbf{D}^{-1}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^\top \mathbf{D}^{-1}\boldsymbol{\tau} && (40)
\end{aligned}$$

h)

We can now write $\mathbf{w}_{\text{MAP}}$ in terms of its original quantities:

$$\begin{aligned}
\mathbf{w}_{\text{MAP}} &= (\alpha^{-1} + \boldsymbol{\Psi}^\top \mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^\top \mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\mathbf{t} && (41)\\
&= (\alpha^{-1} + \boldsymbol{\Psi}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^\top \boldsymbol{\Omega}^{-1}\mathbf{t}, && (42)
\end{aligned}$$

where we used

$$\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A} = ((\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A})^{-1})^{-1} = (\mathbf{A}^{-1}\mathbf{D}(\mathbf{A}^\top)^{-1})^{-1} = (\mathbf{A}^\top \mathbf{D}\mathbf{A})^{-1} = \boldsymbol{\Omega}^{-1}. \tag{43}$$

2 ML estimate of angle measurements

The likelihood is given by the product of the probability distributions of c and s.

$$\begin{aligned}
p(\mathcal{D} \mid \theta) = p(c, s \mid \theta) = p(c \mid \theta)p(s \mid \theta) &= \mathcal{N}(c \mid \cos\theta, \sigma^2)\mathcal{N}(s \mid \sin\theta, \sigma^2) && (44)\\
&= \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}(c - \cos\theta)^2\right\}\exp\left\{-\frac{1}{2\sigma^2}(s - \sin\theta)^2\right\} && (45)
\end{aligned}$$

Taking the logarithm:

$$\ln p(\mathcal{D} \mid \theta) = -\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[(c-\cos\theta)^2 + (s-\sin\theta)^2\right] \tag{46}$$

$$\begin{aligned}(c-\cos\theta)^2 + (s-\sin\theta)^2 &= c^2 - 2c\cos\theta + \cos^2\theta + s^2 - 2s\sin\theta + \sin^2\theta &\quad (47)\\ &= 1 + c^2 + s^2 - 2c\cos\theta - 2s\sin\theta &\quad (48)\end{aligned}$$

$$\begin{aligned}\ln p(\mathcal{D} \mid \theta) &= -\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[1 + c^2 + s^2 - 2c\cos\theta - 2s\sin\theta\right] &\quad (49)\\ &= \underbrace{-\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(1 + c^2 + s^2\right)}_{=C} + \frac{c\cos\theta}{\sigma^2} + \frac{s\sin\theta}{\sigma^2} &\quad (50)\\ &= \frac{c\cos\theta}{\sigma^2} + \frac{s\sin\theta}{\sigma^2} + C &\quad (51)\end{aligned}$$

Taking the derivative:

$$\frac{\partial}{\partial\theta}\ln p(\mathcal{D} \mid \theta) = \frac{s\cos\theta}{\sigma^2} - \frac{c\sin\theta}{\sigma^2} \tag{52}$$

And finally setting the derivative to zero gives:

$$\begin{aligned}\frac{s\cos\theta_{\text{ML}}}{\sigma^2} - \frac{c\sin\theta_{\text{ML}}}{\sigma^2} &= 0 &\quad (53)\\ c\sin\theta_{\text{ML}} &= s\cos\theta_{\text{ML}} &\quad (54)\\ \frac{\sin\theta_{\text{ML}}}{\cos\theta_{\text{ML}}} &= \tan\theta_{\text{ML}} = \frac{s}{c} &\quad (55)\\ \theta_{\text{ML}} &= \arctan\frac{s}{c} &\quad (56)\end{aligned}$$

3 ML and MAP solution of a Poisson distribution fit

a)

For measurements $\{x_i\}_{i=1}^N$, where $x_i \in \{0, 1, 2, 3, \dots\}$, the likelihood is given by the product of each of the individual Poisson distributions:

$$p(\mathcal{D} \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \tag{57}$$

The log likelihood is:

$$\begin{aligned}\ln p(\mathcal{D} \mid \lambda) &= \sum_{i=1}^{N} \ln\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} &\quad (58)\\ &= \sum_{i=1}^{N}\left[x_i \ln\lambda - \lambda - \ln(x_i!)\right] &\quad (59)\end{aligned}$$

Then we take the derivative w.r.t. λ:

$$\frac{\partial}{\partial\lambda}\ln p(\mathcal{D} \mid \lambda) = -N + \sum_{i=1}^{N}\frac{x_i}{\lambda} = -N + \frac{1}{\lambda}\sum_{i=1}^{N} x_i \tag{60}$$

Now, set the derivative to zero to get λ_{ML}.

$$0 = -N + \frac{1}{\lambda_{\text{ML}}}\sum_{i=1}^{N} x_i \tag{61}$$

$$\frac{1}{\lambda_{\text{ML}}} = N\Big(\sum_{i=1}^{N} x_i\Big)^{-1} \tag{62}$$

$$\lambda_{\text{ML}} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{63}$$

b)

Prior: $p(\lambda) \propto \exp(-\frac{\lambda}{a})$
Posterior: $p(\lambda \mid \mathcal{D}) \propto \underbrace{p(\mathcal{D} \mid \lambda)}_{\text{likelihood}} p(\lambda)$

$$\lambda_{\text{MAP}} = \arg\max_{\lambda} p(\lambda \mid \mathcal{D}) = \arg\max_{\lambda} p(\mathcal{D} \mid \lambda) p(\lambda) \tag{64}$$

$$= \arg\max_{\lambda} \ln p(\mathcal{D} \mid \lambda) + \ln p(\lambda) \tag{65}$$

Assume

$$p(\lambda) = C' \exp(-\frac{\lambda}{a}), \tag{66}$$

for any unknown constant C'. Then

$$\ln p(\lambda) = \ln C' - \frac{\lambda}{a}. \tag{67}$$

Since C' is only present in a separate term which does not depend on λ, it will not influence our final result. So we will continue by finding the solution for Eq. (65).

$$\ln p(\mathcal{D} \mid \lambda) + \ln p(\lambda) = -N\lambda + \sum_{i=1}^{N} x_i \ln \lambda - \frac{\lambda}{a} + C, \tag{68}$$

with $C = \ln C' - \sum_{i=1}^{N} \ln(x_i!)$. Taking the derivative gives:

$$\frac{\partial}{\partial \lambda}(\ln p(\mathcal{D} \mid \lambda) + \ln p(\lambda)) = -N + \frac{1}{\lambda} \sum_{i=1}^{N} x_i - \frac{1}{a} \tag{69}$$

And our final result is given by setting the derivative to zero:

$$-N + \frac{1}{\lambda_{\text{MAP}}} \sum_{i=1}^{N} x_i - \frac{1}{a} = 0 \tag{70}$$

$$\frac{1}{\lambda_{\text{MAP}}} \sum_{i=1}^{N} x_i = N + \frac{1}{a} \tag{71}$$

$$\lambda_{\text{MAP}} = \frac{1}{N + 1/a} \sum_{i=1}^{N} x_i \tag{72}$$

c)

The prior results in a lower estimate of λ, since $\frac{1}{N+1/a} < \frac{1}{N}$ for $a > 0$. The limits are given by:

$$\lim_{a \to \infty} \lambda_{\text{MAP}} = \lambda_{\text{ML}} \tag{73}$$

$$\lim_{a \to 0} \lambda_{\text{MAP}} = 0 \tag{74}$$