

### 3.d

We repeat the same process but now taking the derivative w.r.t  $\Omega = \Sigma^{-1}$ :

$$\frac{\partial}{\partial \Omega} \left( -\frac{1}{2} (NK \log 2\pi + N \log \det \Omega^{-1} + \sum_{i=1}^N (t_i - y(x_i, W))^T \Omega (t_i - y(x_i, W))) \right) =$$

We cancel out  $-\frac{NK}{2} \log 2\pi$  since it does not depend on  $\Omega$ :

$$= \frac{\partial}{\partial \Omega} \left( -\frac{1}{2} (N \log \det \Omega^{-1} + \sum_{i=1}^N (t_i - y(x_i, W))^T \Omega (t_i - y(x_i, W))) \right)$$

Separating both parts and computing them, we have:

$$\frac{\partial}{\partial \Omega} \log \det \Omega^{-1} =$$

By the relationship  $\det X^{-1} = \frac{1}{\det X}$ , we get:

$$= \frac{\partial}{\partial \Omega} \log \frac{1}{\det \Omega} = \frac{\partial}{\partial \Omega} (-\log \det \Omega)$$

Now using the given identity, we have:

$$\frac{\partial}{\partial \Omega} \log \det \Omega^{-1} = -(\Omega^{-1})^T$$

And for the second part:

$$\frac{\partial}{\partial \Omega} \sum_{i=1}^N (t_i - y(x_i, W))^T \Omega (t_i - y(x_i, W)) = \frac{\partial}{\partial \Omega} \sum_{i=1}^N (t_i - W^T \phi(x_i))^T \Omega (t_i - W^T \phi(x_i)) =$$

Now to use the second identity that was given to us, we need to transform  $\Omega$  into  $\Omega^T$ . We can do that simply by taking the transpose, since we know that covariance matrix is symmetric and the inverse of a symmetric matrix is also symmetric. Therefore, changing the equation to get to use the identity and also by the sum rule, we get:

$$= \sum_{i=1}^N \frac{\partial}{\partial \Omega} (t_i - W^T \phi(x_i))^T \Omega^T (t_i - W^T \phi(x_i)) = \sum_{i=1}^N (t_i - W^T \phi(x_i))(t_i - W^T \phi(x_i))^T$$

Now putting everything together, we have:

$$\begin{aligned} & \frac{\partial}{\partial \Omega} \left( -\frac{1}{2} (N \log \det \Omega^{-1} + \sum_{i=1}^N (t_i - y(x_i, W))^T \Omega (t_i - y(x_i, W))) \right) = \\ & = -\frac{1}{2} (-N(\Omega^{-1})^T + \sum_{i=1}^N (t_i - W^T \phi(x_i))(t_i - W^T \phi(x_i))^T) \end{aligned}$$

Finally equaling it zero:

$$-N(\Omega^{-1})^T + \sum_{i=1}^N (t_i - W^T \phi(x_i))(t_i - W^T \phi(x_i))^T = 0$$

Going back to  $\Sigma$ , we have:

$$N\Sigma^T = \sum_{i=1}^N (t_i - W^T \phi(x_i))(t_i - W^T \phi(x_i))^T$$

$$\Sigma^T = \frac{1}{N} \sum_{i=1}^N (t_i - W^T \phi(x_i))(t_i - W^T \phi(x_i))^T$$

And using the symmetry of covariance matrix, we get:

$$\Sigma_{ML} = \frac{1}{N} \sum_{i=1}^N (t_i - W_{ML}^T \phi(x_i))(t_i - W_{ML}^T \phi(x_i))^T$$

So it's shown that the maximum likelihood solution for  $\Sigma$  is given by  $\Sigma_{ML} = \frac{1}{N} \sum_{i=1}^N (t_i - W_{ML}^T \phi(x_i))(t_i - W_{ML}^T \phi(x_i))^T$ .