Appendix

Derivation of Laplace Approximation

We approximated the posterior as a Gaussian distribution by the Laplace approximation, i.e., $p(\mathbf{f}|\mathbf{r},\theta)|_{\mathbf{f}=\mathbf{f_{map}}} \approx \mathcal{N}(\mathbf{f}|\mathbf{f_{map}},\Lambda)$. Here, $\mathbf{f_{map}}$ is the maximum of the log-posterior in eq.(5) and Λ is defined by

$$
\begin{aligned}
\Lambda^{-1} &= -\frac{\partial^2}{\partial \mathbf{f}^2} \log p(\mathbf{f}|\mathbf{r},\theta)|_{\mathbf{f}=\mathbf{f_{map}}} \\
&= -\frac{\partial^2}{\partial \mathbf{f}^2} \left[\mathbf{r}^\top \log(g(\mathbf{f})) - \mathbf{1}^\top g(\mathbf{f}) - \tfrac{1}{2}(\mathbf{f}-\boldsymbol{\mu}_\mathbf{f})^\top K^{-1}(\mathbf{f}-\boldsymbol{\mu}_\mathbf{f})\right]|_{\mathbf{f}=\mathbf{f_{map}}} \\
&= \operatorname{diag}\left\{\mathbf{r}./(g(\mathbf{f})\circ g(\mathbf{f}))\circ\left(\frac{\partial g(\mathbf{f})}{\partial \mathbf{f}}\right)^2 - [\mathbf{r}./g(\mathbf{f})-\mathbf{1}]\circ\left(\frac{\partial^2 g(\mathbf{f})}{\partial \mathbf{f}^2}\right)\right\}|_{\mathbf{f}=\mathbf{f_{map}}} + K^{-1}
\end{aligned}
$$

where $\circ$ and $./$ are Hadamard multiplication and division respectively.

When K is ill-conditioned

We find the SVD of K to select the eigenvectors corresponding to eigenvalues that are greater than some threshold. Assume each column of an orthogonal matrix B is the eigenvectors we will keep, and project $\mathbf{f}, K, \boldsymbol{\mu}_\mathbf{f}$ onto the lower dimensional space by $\mathbf{w} = B^T\mathbf{f}, K_p = B^T K B, \boldsymbol{\mu}_{\mathbf{f}p} = B^T\boldsymbol{\mu}_\mathbf{f}$.

The log-posterior in the lower dimensional space is given by:

$$
\log p(\mathbf{w}|\mathbf{r},\theta) = \mathbf{r}^T\log(g(B\mathbf{w})) - \mathbf{1}^T g(B\mathbf{w}) - \frac{1}{2}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{f}p})^T K_p^{-1}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{f}p}) + const \tag{1}
$$

$\mathbf{w_{map}}$ is the maximum of eq.(2). Assuming $g(x) = \log(1+\exp(x))$, $\Lambda_\mathbf{w}$ in the lower dimensional space is given by

$$
\begin{aligned}
\Lambda_\mathbf{w}^{-1} &= -\frac{\partial^2 \log p(\mathbf{w}|\mathbf{r},\theta)}{\partial \mathbf{w}^2}|_{\mathbf{w}=\mathbf{w_{map}}} \\
&= B^T\operatorname{diag}(z)B + K_p^{-1}
\end{aligned} \tag{2}
$$

where $z = \left[r\circ\left(g(B\mathbf{w})\frac{\partial^2 g}{\partial \mathbf{w}^2} - (\frac{\partial g}{\partial \mathbf{w}})^2\right) - g^2(B\mathbf{w})\circ\frac{\partial^2 g}{\partial \mathbf{w}^2}\right]./\left(g(B\mathbf{w})\circ g(B\mathbf{w})\right)|_{\mathbf{w}=\mathbf{w_{map}}}$