TECHNICAL NOTES: NOT FOR CIRCULATION

Gaussian Processes for Differential Equation Modelling

Neil D. Lawrence, Guido Sanguinetti and Magnus Rattray

April 23, 2008

1 Linear System

The original inspiration for this system is in [1] other work in this area has been performed by [**?**].

$$\frac{dy_i(t)}{dt} = B_i + S_i f(t) - D_i x(t)$$

Laplace transforms then give

$$sY_i(s) - y_i(0) = B_i + S_i F(s) - D_i X(s)$$

which, ignoring the initial condition, can be rearranged to form

$$Y_i(s) = \frac{y_i(0)}{s + D_i} + \frac{S_i}{s + D_i} F(s)$$

which implies that

$$y_i(t) = y_i(0) e^{-D_i t} + \frac{B_i}{D_i} + S_i \int_0^t e^{-D_i(t-u)} f(u) \, du$$

which can be rewritten as

$$y_i(t) = e^{-D_i t} \left(y_i(0) + S_i \int_0^t e^{D_i u} f(u) \, du \right) + \frac{B_i}{D_i}$$

we will define

$$g_i(t) = y_i(0) + S_i \int_0^t e^{D_i u} f(u) \, du.$$

The key problem is given a covariance function for f, $\mathbf{K}_f$, what is the covariance function for g_i,$\mathbf{K}_{g_i}$. Having got this, the relationship between g_i and y_i is straightforward,

$$k_{y_i}(t, t') = e^{-D_i t} k_{g_i}(t, t') e^{-D_i t'}.$$

The covariance for g_i is given by

$$k_{g_i g_j}(t, t') = \int_0^t e^{D_i u} \int_0^{t'} e^{D_j u'} e^{-\frac{(u-u')^2}{l^2}} du du'$$

We now make the following substitutions, $s = \frac{u}{l}$ and $s' = \frac{u'}{l}$ giving

$$k_{g_i g_j}(t, t') = l^2 \int_0^{\frac{t}{l}} e^{D_i l s} \int_0^{\frac{t'}{l}} e^{D_j l s'} e^{-(s-s')^2} ds ds'$$

Taking the innermost integral over s',

$$\begin{aligned} e^{D_j l s'} e^{-\left(s-s'\right)^2} &= e^{-s^2+2ss'+D_j l s'-s'^2} \\ &= e^{-\left(s'-\left(s+\frac{D_j l}{2}\right)\right)^2} e^{\left(s+\frac{D_j l}{2}\right)^2 - s^2} \\ &= e^{-\left(s'-\left(s+\frac{D_j l}{2}\right)\right)^2} e^{\left(\frac{D_j l}{2}\right)^2} e^{D_j l s} \end{aligned}$$

so we have

$$\begin{aligned} \int_0^{\frac{t'}{l}} e^{D_j l s'} e^{-\left(s-s'\right)^2} &= \frac{\sqrt{\pi}}{2} e^{\left(\frac{D_j l}{2}\right)^2} e^{D_j l s} \left[\operatorname{erf}\left(s' - \left(s + \frac{D_j l}{2}\right)\right)\right]_0^{\frac{t'}{l}} \\ &= \frac{\sqrt{\pi}}{2} e^{\left(\frac{D_j l}{2}\right)^2} e^{D_j l s} \left[\operatorname{erf}\left(\frac{t'}{l} - \left(s + \frac{D_j l}{2}\right)\right) - \operatorname{erf}\left(-\left(s + \frac{D_j l}{2}\right)\right)\right] \\ &= \frac{\sqrt{\pi}}{2} e^{\left(\frac{D_j l}{2}\right)^2} e^{D_j l s} \left[\operatorname{erf}\left(s + \frac{D_j l}{2}\right) - \operatorname{erf}\left(s + \frac{D_j l}{2} - \frac{t'}{l}\right)\right] \end{aligned}$$

where we have used the fact that $\operatorname{erf}(-x) = -\operatorname{erf}(x)$. The integral therefore becomes

$$k_{g_i g_j}(t, t') = \frac{l^2\sqrt{\pi}}{2} e^{\left(\frac{D_j l}{2}\right)^2} \int_0^{\frac{t}{l}} e^{(D_i+D_j)ls} \left[\operatorname{erf}\left(s + \frac{D_j l}{2}\right) - \operatorname{erf}\left(s + \frac{D_j l}{2} - \frac{t'}{l}\right)\right] ds$$

which can be solved by parts.

$$\begin{aligned} k_{g_i g_j}(t, t') &= \frac{\sqrt{\pi} l}{2(D_i + D_j)} e^{\left(\frac{D_j l}{2}\right)^2} \left\{ \left[e^{(D_i+D_j)ls} \left[\operatorname{erf}\left(s + \frac{D_j l}{2}\right) - \operatorname{erf}\left(s + \frac{D_j l}{2} - \frac{t'}{l}\right)\right]\right]_0^{\frac{t}{l}} - I \right\} \\ &= \frac{\sqrt{\pi} l}{2(D_i + D_j)} e^{\left(\frac{D_j l}{2}\right)^2} \left\{ e^{(D_i+D_j)t} \left[\operatorname{erf}\left(\frac{t}{l} + \frac{D_j l}{2}\right) - \operatorname{erf}\left(\frac{D_j l}{2} - \frac{t'-t}{l}\right)\right] \right. \\ &\quad \left. - \left[\operatorname{erf}\left(\frac{D_j l}{2}\right) - \operatorname{erf}\left(\frac{D_j l}{2} - \frac{t'}{l}\right)\right] - I \right\} \end{aligned}$$

$$\begin{aligned} I &= \tfrac{2}{\sqrt{\pi}} \int_0^{\frac{t}{l}} e^{(D_i+D_j)ls} \left[e^{-\left(s+\frac{D_j l}{2}\right)^2} - e^{-\left(s+\frac{D_j l}{2}-\frac{t'}{l}\right)^2} \right] ds \\ &= I_1 - I_2 \end{aligned}$$

where

$$I_1 = \frac{2}{\sqrt{\pi}} \int_0^{\frac{t}{l}} e^{(D_i+D_j)ls} e^{-\left(s+\frac{D_j l}{2}\right)^2} ds$$

and

$$I_2 = \frac{2}{\sqrt{\pi}} \int_0^{\frac{t}{l}} e^{(D_i+D_j)ls} e^{-\left(s+\frac{D_j l}{2}-\frac{t'}{l}\right)^2} ds$$

the first integral is given by

$$\begin{aligned} I_1 &= \frac{2}{\sqrt{\pi}} \int_0^{\frac{t}{l}} e^{D_i ls + D_j ls - s^2 - D_j ls - \left(\frac{D_j l}{2}\right)^2} \\ &= \frac{2}{\sqrt{\pi}} \int_0^{\frac{t}{l}} e^{D_i ls - s^2 - \left(\frac{D_j l}{2}\right)^2} \\ &= \frac{2}{\sqrt{\pi}} e^{-\left(\frac{D_j l}{2}\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} \int_0^{\frac{t}{l}} e^{-\left(s-\frac{D_i l}{2}\right)^2} \\ &= e^{-\left(\frac{D_j l}{2}\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} \left[\operatorname{erf}\left(s - \frac{D_i l}{2}\right)\right]_0^{\frac{t}{l}} \\ &= e^{-\left(\frac{D_j l}{2}\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} \left[\operatorname{erf}\left(\frac{t}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{D_i l}{2}\right)\right] \end{aligned}$$

$$
\begin{aligned}
I_2 &= \int_0^{\frac{t}{l}} e^{D_i ls + D_j ls - s^2 - D_j ls + 2\frac{t's}{l} - \left(\frac{D_j l}{2} - \frac{t'}{l}\right)^2} \\
&= \frac{2}{\sqrt{\pi}} \int_0^{\frac{t}{l}} e^{D_i ls - s^2 + 2\frac{t's}{l} - \left(\frac{D_j l}{2} - \frac{t'}{l}\right)^2} \\
&= \frac{2}{\sqrt{\pi}} e^{-\left(\frac{D_j l}{2} - \frac{t'}{l}\right)^2} e^{\left(\frac{D_i l}{2} + \frac{t'}{l}\right)^2} \int_0^{\frac{t}{l}} e^{-\left(s - \left(\frac{D_i l}{2} + \frac{t'}{l}\right)\right)^2} \\
&= \frac{2}{\sqrt{\pi}} e^{-\left(\frac{D_j l}{2}\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} e^{(D_j + D_i)t'} \int_0^{\frac{t}{l}} e^{-\left(s - \left(\frac{D_i l}{2} + \frac{t'}{l}\right)\right)^2} \\
&= e^{-\left(\frac{D_j l}{2}\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} e^{(D_j + D_i)t'} \left[\operatorname{erf}\left(s - \left(\frac{D_i l}{2} + \frac{t'}{l}\right)\right)\right]_0^{\frac{t}{l}} \\
&= e^{-\left(\frac{D_j l}{2}\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} e^{(D_j + D_i)t'} \left[\operatorname{erf}\left(\frac{t - t'}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{t'}{l} + \frac{D_i l}{2}\right)\right]
\end{aligned}
$$

so we have $I = I_1 - I_2$

$$
\begin{aligned}
I &= e^{-\left(\frac{D_j l}{2}\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} \left\{\left[\operatorname{erf}\left(\frac{t}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{D_i l}{2}\right)\right]\right. \\
&\quad \left. - e^{(D_j + D_i)t'} \left[\operatorname{erf}\left(\frac{t - t'}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{t'}{l} + \frac{D_i l}{2}\right)\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
k_{g_i g_j}(t, t') &= \frac{\sqrt{\pi} l}{2(D_i + D_j)} \left[e^{\left(\frac{D_j l}{2}\right)^2} \left\{e^{(D_i + D_j)t} \left[\operatorname{erf}\left(\frac{t}{l} + \frac{D_j l}{2}\right) - \operatorname{erf}\left(\frac{D_j l}{2} - \frac{t' - t}{l}\right)\right]\right.\right. \\
&\quad \left. - \left[\operatorname{erf}\left(\frac{D_j l}{2}\right) - \operatorname{erf}\left(\frac{D_j l}{2} - \frac{t'}{l}\right)\right]\right\} \\
&\quad + e^{\left(\frac{D_i l}{2}\right)^2} \left\{e^{(D_i + D_j)t'} \left[\operatorname{erf}\left(\frac{t - t'}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{t'}{l} + \frac{D_i l}{2}\right)\right]\right. \\
&\quad \left.\left. - \left[\operatorname{erf}\left(\frac{t}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{D_i l}{2}\right)\right]\right\}\right]
\end{aligned}
$$

$$
\begin{aligned}
k_{g_i g_j}(t, t') &= \frac{\sqrt{\pi} l}{2(D_i + D_j)} \left[e^{\left(\frac{D_j l}{2}\right)^2} \left\{e^{(D_i + D_j)t} \left[\operatorname{erf}\left(\frac{t' - t}{l} - \frac{D_j l}{2}\right) + \operatorname{erf}\left(\frac{t}{l} + \frac{D_j l}{2}\right)\right]\right.\right. \\
&\quad \left. - \left[\operatorname{erf}\left(\frac{t'}{l} - \frac{D_j l}{2}\right) + \operatorname{erf}\left(\frac{D_j l}{2}\right)\right]\right\} \\
&\quad + e^{\left(\frac{D_i l}{2}\right)^2} \left\{e^{(D_i + D_j)t'} \left[\operatorname{erf}\left(\frac{t - t'}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{t'}{l} + \frac{D_i l}{2}\right)\right]\right. \\
&\quad \left.\left. - \left[\operatorname{erf}\left(\frac{t}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{D_i l}{2}\right)\right]\right\}\right]
\end{aligned}
$$

so we have

$$k_{y_i y_j}(t, t') = S_i C_j e^{-D_i t} k_{g_i g_j}(t, t') e^{-D_j t'}$$

$$k_{y_i y_j}(t, t') = S_i C_j \frac{\sqrt{\pi} l}{2} \left[h_{ji}(t', t) + h_{ij}(t, t')\right]$$

where

$$
\begin{aligned}
h_{ji}(t', t) &= \frac{e^{\left(\frac{D_j l}{2}\right)^2}}{(D_i + D_j)} \left\{e^{-D_j(t' - t)} \left[\operatorname{erf}\left(\frac{t' - t}{l} - \frac{D_j l}{2}\right) + \operatorname{erf}\left(\frac{t}{l} + \frac{D_j l}{2}\right)\right]\right. \\
&\quad \left. - e^{-\left(D_i t + D_j t'\right)} \left[\operatorname{erf}\left(\frac{t'}{l} - \frac{D_j l}{2}\right) + \operatorname{erf}\left(\frac{D_j l}{2}\right)\right]\right\}
\end{aligned}
$$

2 Limit as $t \to \infty$and $t' \to \infty$

In this limit, keeping $t' - t$ finite we loose the effect of the initial conditions,

$$h_{ji}(t', t) = \frac{e^{\left(\frac{D_j l}{2}\right)^2}}{(D_i + D_j)} \left\{ e^{-D_j(t'-t)} \left[\text{erf}\left(\frac{t'-t}{l} - \frac{D_j l}{2} \right) + 1 \right] \right\}.$$

3 Gradients

We also need gradients of the kernel with respect to the parameters,

$$\frac{dk_{y_i y_j}(t, t')}{dS_i} = \frac{k_{y_i y_j}(t, t')}{S_i}$$

$$\begin{aligned} \frac{dh_{ji}(t', t)}{dD_j} &= \frac{D_j l^2 h_{ji}(t', t)}{2} - \frac{1}{D_i + D_j} h_{ji}(t', t) \\ &+ \frac{e^{\left(\frac{D_j l}{2}\right)^2}}{D_i + D_j} \left\{ -(t' - t) e^{-D_j(t'-t)} \left[\text{erf}\left(\frac{t'-t}{l} - \frac{D_j l}{2} \right) + \text{erf}\left(\frac{t}{l} + \frac{D_j l}{2} \right) \right] \right. \\ &\left. + t' e^{-(D_i t + D_j t')} \left[\text{erf}\left(\frac{t'}{l} - \frac{D_j l}{2} \right) + \text{erf}\left(\frac{D_j l}{2} \right) \right] \right\} \\ &+ \frac{l}{\sqrt{\pi}(D_i + D_j)} \left\{ \left[-e^{-\frac{(t'-t)^2}{l^2}} + e^{-\frac{t^2}{l^2} - D_j t'} \right] \right. \\ &\left. + \left[e^{-\frac{t'^2}{l^2} - D_i t} - e^{-(D_i t + D_j t')} \right] \right\} \end{aligned}$$

$$\frac{dh_{ji}(t', t)}{dD_i} = \frac{e^{\left(\frac{D_j l}{2}\right)^2}}{D_i + D_j} \left\{ t e^{-(D_i t + D_j t')} \left[\text{erf}\left(\frac{t'}{l} - \frac{D_j l}{2} \right) + \text{erf}\left(\frac{D_j l}{2} \right) \right] \right\} - \frac{h_{ji}}{D_i + D_j}$$

$$\begin{aligned} \frac{dh_{ji}(t', t)}{dl} &= \frac{D_j^2 l h_{ji}(t', t)}{2} \\ &+ \frac{2}{\sqrt{\pi}(D_i + D_j)} \left\{ \left[\left(-\frac{(t'-t)}{l^2} - \frac{D_j}{2} \right) e^{\left(\frac{-(t'-t)^2}{l^2} \right)} + \left(-\frac{t}{l^2} + \frac{D_j}{2} \right) e^{\left(-\frac{t^2}{l^2} - D_j t' \right)} \right] \right. \\ &\left. - \left[\left(-\frac{t'}{l^2} - \frac{D_j}{2} \right) e^{\left(-\frac{t'^2}{l^2} - D_i t \right)} + \frac{D_j}{2} e^{-(D_i t + D_j t')} \right] \right\} \end{aligned}$$

3.1 Cross Covariance between f and x

Cross Covariance between f and x

$$k_{y_i f}(t, t') = e^{-D_i t} \int_0^t e^{D_i u} e^{-\frac{(u-t')^2}{l^2}} du$$

substitute $s = \frac{u}{l}$ and $s' = \frac{t'}{l}$ giving

$$k_{y_i f}(t, t') = l S_i e^{-D_i t} \int_0^{\frac{t}{l}} e^{D_i l s} e^{-(s-s')} ds$$

$$
\begin{aligned}
e^{D_i l s} e^{-\left(s-s'\right)} &= e^{D_i l s - s^2 + 2s's - s'^2} \\
&= e^{-\left(s-\left(\frac{D_i l}{2}+s'\right)\right)^2} e^{\left(\frac{D_i l}{2}+s'\right)^2} e^{-s'^2} \\
&= e^{-\left(s-\left(\frac{D_i l}{2}+s'\right)\right)^2} e^{\left(\frac{D_i l}{2}\right)^2} e^{D_i l s'}
\end{aligned}
$$

$$
\begin{aligned}
k_{y_i f}\left(t,t'\right) &= \frac{\sqrt{\pi} l S_i}{2} e^{\left(\frac{D_i l}{2}\right)^2} e^{-D_i\left(t-t'\right)} \left[\operatorname{erf}\left(s - \left(\frac{D_i l}{2} + s'\right)\right)\right]_0^{\frac{t}{l}} \\
&= \frac{\sqrt{\pi} l S_i}{2} e^{\left(\frac{D_i l}{2}\right)^2} e^{-D_i\left(t-t'\right)} \left[\operatorname{erf}\left(\frac{t-t'}{l} - \frac{D_i l}{2}\right) + \operatorname{erf}\left(\frac{t'}{l} + \frac{D_i l}{2}\right)\right].
\end{aligned}
$$

Gradient of the cross kernels.

$$
\begin{aligned}
\frac{dk_{y_i f}\left(t,t'\right)}{dD_i} &= \left(\frac{l^2 D_i}{2} - \left(t-t'\right)\right) k_{y_i f}\left(t,t'\right) \\
&\quad + \frac{S_i l^2}{2}\left[\left(-\frac{t-t'}{l^2} - \frac{D_i}{2}\right) e^{-\left(\frac{t-t'}{l^2}\right)^2} + \left(-\frac{t'}{l^2} + \frac{D_i}{2}\right) e^{-\frac{t'^2}{l^2} - tD_i}\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{dk_{y_i f}\left(t,t'\right)}{dl} &= \left(\frac{1}{l} + \frac{l D_i^2}{2}\right) k_{y_i f}\left(t,t'\right) \\
&\quad + \frac{S_i l D_i}{2}\left[e^{-\frac{t'^2}{l^2} - tD_j} - e^{-\left(\frac{t-t'}{l}\right)^2}\right]
\end{aligned}
$$

$$
\frac{dk_{y_i f}\left(t,t'\right)}{dS_i} = \frac{1}{S_i} k_{y_i f}\left(t,t'\right)
$$

3.2 The Likelihood

The Gaussian process likelihood we are interested in is

$$
p\left(\mathbf{y}|\mathbf{t}\right) = \frac{1}{\left(2\pi\right)^{\frac{NG}{2}} \left|\mathbf{K}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}\right)^{\mathrm{T}} \mathbf{K}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)\right)
$$

where $\mathbf{y} = \left[\mathbf{y}_1^{\mathrm{T}} \dots \mathbf{y}_G^{\mathrm{T}}\right]^{\mathrm{T}}$ is a vector which concatanates all the observations of the gene expressions, I is the number of gene targets for the protein of interest, N is the number of observations for each gene and $\boldsymbol{\mu} = \left[\frac{B_1}{D_1}\mathbf{1}_N^{\mathrm{T}} \dots \frac{B_G}{D_G}\mathbf{1}_N^{\mathrm{T}}\right]$ is the mean vector. Taking the gradient of the log likelihood with respect to $\mathbf{K}$ we have

$$
\frac{dL}{d\mathbf{K}} = -\mathbf{K}^{-1} + \mathbf{K}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)\left(\mathbf{y}-\boldsymbol{\mu}\right)^{\mathrm{T}}\mathbf{K}^{-1}
$$

and with respect to $\boldsymbol{\mu}$ we have

$$
\frac{dL}{d\boldsymbol{\mu}} = \mathbf{K}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)
$$

these can be combined with gradients of $\mathbf{K}$ with respect to the parameters (from above) and gradients of $\boldsymbol{\mu}$ with respect to the parameters to find gradients of the log likelihood.

4 Multiple Input Systems

We now consider a linear system driven by multiple functions, so that we have

$$
\frac{dy_i\left(t\right)}{dt} = B_i + \sum_{j=1}^{M} S_{ij} f_j\left(t\right) + \sum_{j=1}^{M}\sum_{k=1}^{M} C_{ijk} f_j\left(t\right) f_k\left(t\right) - D_i y_i\left(t\right)
$$

The solution for $y_i(t)$ is given by

$$y_i(t) = \frac{B_i}{D_i} + e^{-D_i t}\int_0^t e^{D_i u}\left(\sum_{j=1}^{M} S_{ij} f_j(u) + \sum_{j=1}^{M}\sum_{k=1}^{M} C_{ijk} f_j(u) f_k(u)\right) du$$

This implies the covariance for $y_i(t)\, y_l(t) = \int$

5 Non-Linear Systems

5.1 MAP-Laplace

5.1.1 Notation

We use lower case for functions, e.g. k and w, and upper case for the corresponding matrices in the numerical implementation, e.g. $\mathbf{K}$ and $\mathbf{W}$.

5.1.2 Gradient and Hessian

$$\frac{dy_j}{dt} = B_j + g(f(t), \theta_j) - D_j y_j \ .$$

We set the initial conditions $y_j(0) = B_j/D_j$ so that,

$$y_j(t) = \frac{B_j}{D_j} + e^{-D_j t}\int_0^t \mathrm{d}u \ g(f(u), \theta_j) \ e^{D_j u} \ .$$

The log-likelihood of data $D = \{y_{ij}\}$ for gene j at times t_i is,

$$\log p(D|f, \{B_j, \theta_j, D_j\}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{G}\left[\beta_{ij}(y_j(t_i) - y_{ij})^2 - \log(\beta_{ij})\right] - \frac{NG}{2}\log(2\pi) \ .$$

To compute the functional gradient of the log likelihood, we first rewrite

$$\begin{aligned} y_j(t) &= \frac{B_j}{D_j} + e^{-D_j t}\int_0^t \mathrm{d}u \ g(f(u), \theta_j) \ e^{D_j u} \ . \\ &= \frac{B_j}{D_j} + e^{-D_j t}\Delta \sum g(f_k, \theta_j) e^{D_j t_k} \end{aligned}$$

The functional gradient of the log-likelihood with respect to f is,

$$\frac{\delta \log p(D|f)}{\delta f} = -\sum_{i=1}^{N}\Theta(t_i - t)\sum_{j=1}^{G}\beta_{ij}(y_j(t_i) - y_{ij})\frac{\partial g(f, \theta_j)}{\partial f} e^{-D_j(t_i - t)} dt$$

The negative Hessian of the log-likelihood with respect to f is,

$$\begin{aligned} w &= -\frac{\delta^2 \log p(D|f)}{\delta f^2} = \sum_{i=1}^{N}\Theta(t_i - t)\sum_{j=1}^{G}\beta_{ij}(y_j(t_i) - y_{ij}) g''(f(t), \theta_j) e^{-D_j(t_i - t)} dt \\ &\quad + \sum_{i=1}^{N}\Theta(t_i - t)\Theta(t_i - s)\sum_{j=1}^{G}\beta_{ij} g'(f(t), \theta_j) g'(f(s), \theta_j) e^{-D_j(2t_i - t - s)} dt ds \end{aligned}$$

where $g'(f) = \partial g/\partial f$ and $g''(f) = \partial^2 g/\partial f^2$. The gradient and Hessian of the unnormalised log posterior are,

$$\begin{aligned} \frac{\delta \log p(f|D)}{\delta f} &= \frac{\delta \log p(D|f)}{\delta f} - k^{-1} f \\ \frac{\delta^2 \log p(f|D)}{\delta f^2} &= -w - k^{-1} \end{aligned}$$

5.1.3 Numerical implementation

We discretize time t_k with $\tau = t_k - t_{k-1}$ constant. We write $\mathbf{f} = [f_k]$ to be the vector realisation of the function f. The gradient of the log-likelihood is then given by,

$$\nabla_k \log p(D|\mathbf{f}) = \frac{\partial \log p(D|\mathbf{f})}{\partial f_k} = -\tau \sum_{i=1}^{N} \Theta(t_i - t_k) \sum_{j=1}^{G} \beta_{ij} (y_j(t_i) - y_{ij}) \partial g'(f_k, \theta_j) e^{-D_j(t_i - t_k)}$$

and the negative Hessian of the log-likelihood is,

$$\begin{aligned} w_{kl} &= -\nabla_k \nabla_l \log p(D|\mathbf{f}) = \delta_{kl} \tau \sum_{i=1}^{N} \Theta(t_i - t_k) \sum_{j=1}^{G} \beta_{ij} (y_j(t_i) - y_{ij}) g''(f_k, \theta_j) e^{-D_j(t_i - t_k)} \\ &\quad + \tau^2 \sum_{i=1}^{N} \Theta(t_i - t_k) \Theta(t_i - t_l) \sum_{j=1}^{G} \beta_{ij} g'(f_k, \theta_j) g'(f_l, \theta_j) e^{-D_j(2t_i - t_k - t_l)} \end{aligned}$$

where δ_{kl} is the Kronecker delta.

5.1.4 MAP solution and Laplace approximation

The gradient and Hessian of the unnormalised log posterior $\Psi(\mathbf{f}) = \log p(D|\mathbf{f}) + \log p(\mathbf{f})$ are,

$$\begin{aligned} \nabla\Psi(\mathbf{f}) &= \nabla \log p(D|\mathbf{f}) - \mathbf{K}^{-1}\mathbf{f} , \\ \nabla\nabla\Psi(\mathbf{f}) &= -\left(\mathbf{W} + \mathbf{K}^{-1}\right) . \end{aligned} \tag{1}$$

We use this matrix to find the Newton direction for optimisation,

$$\Delta f = -\eta \left(\mathbf{W} + \mathbf{K}^{-1}\right)^{-1} \left(\nabla L - \mathbf{K}^{-1}\mathbf{f}\right)$$

where $L = \log p(D|\mathbf{f})$. Doing a full Newton up-date one finds,

$$\mathbf{f} \leftarrow \left(\mathbf{W} + \mathbf{K}^{-1}\right)^{-1} (\mathbf{W}\mathbf{f} + \nabla L)$$

which converges quickly. The matrix inversion lemma can be used to speed up the inversion of $\mathbf{I}+\mathbf{K}\mathbf{W}$, since $\mathbf{W}$ can be written as a sum of outer-products, but I haven't bothered with that yet. The Laplace approximation to the log marginal likelihood is (ignoring terms that do not involve model parameters),

$$\begin{aligned} \log p(D|\{B_j, \theta_j, D_j\}, \gamma) &\simeq \log p\left(D|\hat{\mathbf{f}}\right) - \tfrac{1}{2}\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{K}^{-1}\hat{\mathbf{f}} - \tfrac{1}{2}\log|\mathbf{W} + \mathbf{K}^{-1}| - \tfrac{1}{2}\log|\mathbf{K}| \\ &= \log p\left(D|\hat{\mathbf{f}}\right) - \tfrac{1}{2}\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{K}^{-1}\hat{\mathbf{f}} - \tfrac{1}{2}\log|\mathbf{I} + \mathbf{K}\mathbf{W}| \end{aligned} \tag{2}$$

where $\hat{\mathbf{f}}$ is the MAP solution and γ are the kernel parameters.

5.1.5 Linear case

We first check the results for the simplest case,

$$g(f, S_j) = S_j\, f , \qquad k(t, t') = \exp\left(\frac{-\gamma (t - t')^2}{2}\right) .$$

We will treat the noise as known and ignore noise propagation for now, so β_{ij} does not have to be estimated. The gradients we have to work out are,

$$\begin{aligned} \frac{\partial y_j(t)}{\partial B_j} &= \frac{1}{D_j} \\ \frac{\partial y_j(t)}{\partial S_j} &= e^{-D_j t} \int_0^t \mathrm{d}u\, f(u)\, e^{D_j u} = \frac{y_j(t) - B_j/D_j}{S_j} \\ \frac{\partial y_j(t)}{\partial D_j} &= -\frac{B_j}{D_j^2} + S_j\, e^{-D_j t} \int_0^t \mathrm{d}u\, f(u)\, (u - t)\, e^{D_j u} \end{aligned}$$

and then for any parameter z_j,

$$\frac{\partial \log p\left(D|f\right)}{\partial z_j} = -\sum_{i=1}^{N} \beta_{ij}\left(y_j\left(t_i\right) - y_{ij}\right)\frac{\partial y_j\left(t_i\right)}{\partial z_j}$$

The Hessian has no dependence on f, it only depends on the parameters D_j and S_j,

$$w_{kl} = \tau^2 \sum_{i=1}^{N} \Theta\left(t_i - t_k\right)\Theta\left(t_i - t_l\right)\sum_{j=1}^{G} \beta_{ij} S_j^2 e^{-D_j\left(2t_i - t_k - t_l\right)} \ .$$

The gradients with respect to these parameters are,

$$\begin{aligned}
\frac{\partial w_{kl}}{\partial S_j} &= 2\tau^2 \sum_{i=1}^{N} \Theta\left(t_i - t_k\right)\Theta\left(t_i - t_l\right)\beta_{ij} S_j e^{-D_j\left(2t_i - t_k - t_l\right)} \\
\frac{\partial w_{kl}}{\partial D_j} &= -\tau^2 \sum_{i=1}^{N} \Theta\left(t_i - t_k\right)\Theta\left(t_i - t_l\right)\beta_{ij} S_j^2 \left(2t_i - t_k - t_l\right) e^{-D_j\left(2t_i - t_k - t_l\right)}
\end{aligned}$$

and then for any parameter z,

$$\frac{\partial}{\partial z}\log\left|\mathbf{I} + \mathbf{KW}\right| = \mathrm{tr}\left[\left(\mathbf{I} + \mathbf{KW}\right)^{-1}\mathbf{K}\frac{\partial \mathbf{W}}{\partial z}\right]$$

The gradient of the kernel with respect to its parameter is,

$$\frac{\partial \mathbf{K}_{kl}}{\partial \gamma} = -\tfrac{1}{2}\left(t_k - t_l\right)^2 \mathbf{K}_{kl}$$

and we have,

$$\begin{aligned}
\frac{\partial}{\partial \gamma}\log\left|\mathbf{I} + \mathbf{KW}\right| &= \mathrm{tr}\left[\left(\mathbf{I} + \mathbf{KW}\right)^{-1}\mathbf{W}\frac{\partial \mathbf{K}}{\partial \gamma}\right] \ , \\
\frac{\partial \mathbf{K}^{-1}}{\partial \gamma} &= -\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \gamma}\mathbf{K}^{-1} \ .
\end{aligned}$$

So we find,

$$\frac{\partial \log p\left(D|\gamma\right)}{\partial \gamma} = -\tfrac{1}{2}\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \gamma}\mathbf{K}^{-1}\hat{\mathbf{f}} - \tfrac{1}{2}\mathrm{tr}\left(\left(\mathbf{W}^{-1} + \mathbf{K}\right)^{-1}\frac{\partial \mathbf{K}}{\partial \gamma}\right)$$

5.1.6 Constraining the TF concentrations to be positive

We constrain the function to be positive,

$$g\left(f, S_j\right) = S_j\, e^{f} \ , \qquad k\left(t, t'\right) = \exp\left(\frac{-\gamma\left(t - t'\right)^2}{2}\right) \ .$$

I will only consider optimisation of $\mathbf{f}$ by MAP learning and the kernel parameters by MAP-Laplace. The gradient and Hessian are,

$$\begin{aligned}
\nabla_k \log p\left(D|\mathbf{f}\right) = \frac{\partial \log p\left(D|\mathbf{f}\right)}{\partial f_k} &= -\tau \sum_{i=1}^{N}\Theta\left(t_i - t_k\right)\sum_{j=1}^{G}\beta_{ij}\left(y_j\left(t_i\right) - y_{ij}\right)S_j e^{f_k - D_j\left(t_i - t_k\right)} \\
w_{kl} = -\nabla_k\nabla_l \log p\left(D|\mathbf{f}\right) &= \delta_{kl}\tau \sum_{i=1}^{N}\Theta\left(t_i - t_k\right)\sum_{j=1}^{G}\beta_{ij}\left(y_j\left(t_i\right) - y_{ij}\right)S_j e^{f_k - D_j\left(t_i - t_k\right)} \\
&\quad + \tau^2 \sum_{i=1}^{N}\Theta\left(t_i - t_k\right)\Theta\left(t_i - t_l\right)\sum_{j=1}^{G}\beta_{ij}S_j^2 e^{f_k + f_l - D_j\left(2t_i - t_k - t_l\right)} \\
&= -\delta_{kl}\nabla_k \log p\left(D|\mathbf{f}\right) + \tau^2 \sum_{i=1}^{N}\Theta\left(t_i - t_k\right)\Theta\left(t_i - t_l\right)\sum_{j=1}^{G}\beta_{ij}S_j^2 e^{f_k + f_l - D_j\left(2t_i - t_k - t_l\right)} \qquad (3)
\end{aligned}$$

The term in the Hessian proportional to the gradient vanishes at the MAP solution. As before the gradient of the kernel with respect to its parameter is,

$$\frac{\partial \mathbf{K}_{kl}}{\partial \gamma} = -\tfrac{1}{2}\left(t_k - t_l\right)^2 \mathbf{K}_{kl} \ .$$

The gradient of the log-marginal with respect to γ is,

$$\frac{\partial \log p\left(D|\gamma\right)}{\partial \gamma} = -\tfrac{1}{2}\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \gamma}\mathbf{K}^{-1}\hat{\mathbf{f}} - \tfrac{1}{2}\mathrm{tr}\left(\left(\mathbf{W}^{-1}+\mathbf{K}\right)^{-1}\frac{\partial \mathbf{K}}{\partial \gamma}\right) + \sum_k \frac{\partial \log p\left(D|\gamma\right)}{\partial \hat{f}_k}\frac{\hat{f}_k}{\partial \gamma}$$

where $\hat{\mathbf{f}}$ is the MAP solution and the final term is due to the implicit dependence of $\hat{\mathbf{f}}$ on γ. The only term that contributes to this is the final one in equation (2) which involves $\mathbf{W}$. We find,

$$\frac{\partial \log p\left(D|\gamma\right)}{\partial \hat{f}_k} = -\frac{1}{2}\mathrm{tr}\left(\left(\mathbf{K}^{-1}+\mathbf{W}\right)^{-1}\frac{\partial \mathbf{W}}{\partial \hat{f}_k}\right)$$

At the MAP solution we find,

$$\frac{\partial w_{pq}}{\partial \hat{f}_k} = \left(\delta_{kp} + \delta_{kq}\right) w_{pq} \ .$$

where we have used the self-consistent condition that the first term in equation (3) vanishes at the MAP solution. This then simplifies to,

$$\frac{\partial \log p\left(D|\gamma\right)}{\partial \hat{f}_k} = -\left(\left(\mathbf{W}+\mathbf{K}^{-1}\right)^{-1}\mathbf{W}\right)_{kk} \ .$$

>From equation (3) we have the self-consistent equation $\hat{\mathbf{f}} = \mathbf{K}\nabla \log p\left(D|\hat{\mathbf{f}}\right)$ and differentiating that we get,

$$\frac{\partial \hat{\mathbf{f}}}{\partial \gamma} = \left(\mathbf{W}+\mathbf{K}^{-1}\right)^{-1}\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \gamma}\nabla \log p\left(D|\hat{\mathbf{f}}\right) \ .$$

References

[1] Martino Barenco, Daniela Tomescu, Daniel Brewer, Robin Callard, Jaroslav Stark, and Michael Hubank. Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biology*, 7(3):R25, 2006. 2