Gaussian Processes for Single Input Module Motifs

November 1, 2006

1 Linear System

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

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

Laplace transforms then give

$$sX_i(s) - x_i(0) = B_i + S_i F(s) - D_i X(s)$$

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

$$X_i(s) = \frac{x_i(0)}{s + D_i} + \frac{S_i}{s + D_i} F(s)$$

which implies that

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

which can be rewritten as

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

we will define

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

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 x_i is straightforward,

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

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

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

which can be solved by parts.

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

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

where

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

and

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

the first integral is given by

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

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

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

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

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

where

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

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

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

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

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

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

Cross Covariance between f and x

Cross Covariance between f and x

$$k_{x_if}(t,t') = e^{-D_it}\int_0^t e^{D_iu}e^{-\frac{(u-t')^2}{l^2}}du$$

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

$$k_{x_if}(t,t') = lS_ie^{-D_it}\int_0^{\frac{t}{l}} e^{D_ils}e^{-(s-s')}ds$$

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

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

Gradient of the cross kernels.

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

$$\frac{dk_{x_i f}(t,t')}{dl} = \left(\frac{1}{l}+\frac{lD_i^2}{2}\right)k_{x_i f}(t,t') + \frac{S_i l D_i}{2}\left[e^{-\frac{t'^2}{l^2}-tD_j} - e^{-\left(\frac{t-t'}{l}\right)^2}\right]$$

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

2 The Likelihood

The Gaussian process likelihood we are interested in is

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

where $\mathbf{y} = \left[\mathbf{x}_1^{\mathrm{T}}\dots\mathbf{x}_I^{\mathrm{T}}\right]^{\mathrm{T}}$ 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 $\boldsymbol{\mu} = \left[\frac{B_1}{D_1}\mathbf{1}_N^{\mathrm{T}}\dots\frac{B_I}{D_I}\mathbf{1}_N^{\mathrm{T}}\right]$ is the mean vector. Taking the gradient of the log likelihood with respect to $\mathbf{K}$ we have

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

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

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

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.

$$\mathbf{y}$$

3 Non-Linear Systems

3.1 MAP-Laplace

3.1.1 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. K and W.

3.1.2 Gradient and Hessian

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

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

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

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

$$p(D|f,\{B_j,\theta_j,D_j\}) = -\frac{1}{2}\sum_{i=1}^{T}\sum_{j=1}^{m}\left[\lambda_{ij}\left(x_j(t_i)-y_{ij}\right)^2+\log\left(\lambda_{ij}\right)\right] - \frac{mT}{2}\log(2\pi)\ .$$

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

$$\frac{\delta \log p(D|f)}{\delta f} = -\sum_{i=1}^{T} \Theta(t_i - t) \sum_{j=1}^{m} \lambda_{ij}(x_j(t_i) - y_{ij}) \frac{\partial g(f, \theta_j)}{\partial f} \mathrm{e}^{-D_j(t_i - t)} dt$$

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

$$\begin{aligned} w &= -\frac{\delta^2 \log p(D|f)}{\delta f^2} = \sum_{i=1}^{T} \Theta(t_i - t) \sum_{j=1}^{m} \lambda_{ij}(x_j(t_i) - y_{ij}) g''(f(t), \theta_j) \mathrm{e}^{-D_j(t_i - t)} dt \\ &+ \sum_{i=1}^{T} \Theta(t_i - t)\Theta(t_i - s) \sum_{j=1}^{m} \lambda_{ij} g'(f(t), \theta_j) g'(f(s), \theta_j) \mathrm{e}^{-D_j(2t_i - t - s)} dt ds \end{aligned} \tag{1}$$

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

$$\frac{\delta \log p(f|D)}{\delta f} = \frac{\delta \log p(D|f)}{\delta f} - k^{-1} f \tag{2}$$

$$\frac{\delta^2 \log p(f|D)}{\delta f^2} = -w - k^{-1} \tag{3}$$

3.1.3 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,

$$\nabla_k \log p(D|\mathbf{f}) = \frac{\partial \log p(D|\mathbf{f})}{\partial f_k} = -\tau \sum_{i=1}^{T} \Theta(t_i - t_k) \sum_{j=1}^{m} \lambda_{ij}(x_j(t_i) - y_{ij}) \partial g'(f_k, \theta_j) \mathrm{e}^{-D_j(t_i - t_k)}$$

and the -ve Hessian of the log-likelihood is,

$$\begin{aligned} W_{kl} &= -\nabla_k \nabla_l \log p(D|\mathbf{f}) = \delta_{kl} \tau \sum_{i=1}^{T} \Theta(t_i - t_k) \sum_{j=1}^{m} \lambda_{ij}(x_j(t_i) - y_{ij}) g''(f_k, \theta_j) \mathrm{e}^{-D_j(t_i - t_k)} \\ &+ \tau^2 \sum_{i=1}^{T} \Theta(t_i - t_k)\Theta(t_i - t_l) \sum_{j=1}^{m} \lambda_{ij} g'(f_k, \theta_j) g'(f_l, \theta_j) \mathrm{e}^{-D_j(2t_i - t_k - t_l)} \end{aligned} \tag{4}$$

where δ_{kl} is the Kronecker delta.

3.1.4 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,

$$\nabla \Psi(\mathbf{f}) = \nabla \log p(D|\mathbf{f}) - K^{-1} \mathbf{f} , \tag{5}$$

$$\nabla \nabla \Psi(\mathbf{f}) = -(W + K^{-1}) . \tag{6}$$

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

$$\Delta f = -\eta (W + K^{-1})^{-1} (\nabla L - K^{-1} \mathbf{f}) \tag{7}$$

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

$$\mathbf{f} \leftarrow (W + K^{-1})^{-1} (W \mathbf{f} + \nabla L) \tag{8}$$

which converges quickly. The matrix inversion lemma can be used to speed up the inversion of $I+KW$, since 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),

$$\begin{aligned}\log p(D|\{B_j,\theta_j,D_j\},\gamma) &\simeq \log p(D|\hat{\mathbf{f}}) - \tfrac{1}{2}\hat{\mathbf{f}}^T K^{-1}\hat{\mathbf{f}} - \tfrac{1}{2}\log|W+K^{-1}| - \tfrac{1}{2}\log|K| \\ &= \log p(D|\hat{\mathbf{f}}) - \tfrac{1}{2}\hat{\mathbf{f}}^T K^{-1}\hat{\mathbf{f}} - \tfrac{1}{2}\log|I+KW| \qquad (9)\end{aligned}$$

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

3.1.5 Linear case

We first check the results for the simplest case,

$$g(f,S_j) = S_j\, f\ , \qquad k(t,t') = \exp\left(\frac{-\gamma(t-t')^2}{2}\right)\ . \qquad (10)$$

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

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

and then for any parameter z_j,

$$\frac{\partial \log p(D|f)}{\partial z_j} = -\sum_{i=1}^{T}\lambda_{ij}(x_j(t_i)-y_{ij})\frac{\partial x_j(t_i)}{\partial z_j}$$

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

$$W_{kl} = \tau^2\sum_{i=1}^{T}\Theta(t_i-t_k)\Theta(t_i-t_l)\sum_{j=1}^{m}\lambda_{ij}S_j^2\mathrm{e}^{-D_j(2t_i-t_k-t_l)}\ .$$

The gradients with respect to these parameters are,

$$\begin{aligned}\frac{\partial W_{kl}}{\partial S_j} &= 2\tau^2\sum_{i=1}^{T}\Theta(t_i-t_k)\Theta(t_i-t_l)\lambda_{ij}S_j\mathrm{e}^{-D_j(2t_i-t_k-t_l)} \\ \frac{\partial W_{kl}}{\partial D_j} &= -\tau^2\sum_{i=1}^{T}\Theta(t_i-t_k)\Theta(t_i-t_l)\lambda_{ij}S_j^2(2t_i-t_k-t_l)\mathrm{e}^{-D_j(2t_i-t_k-t_l)}\end{aligned}$$

and then for any parameter z,

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

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

$$\frac{\partial K_{kl}}{\partial \gamma} = -\tfrac{1}{2}(t_k-t_l)^2 K_{kl}$$

and we have,

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

So we find,

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

3.1.6 Constraining the TF concentrations to be positive

We constrain the function to be positive,

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

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

$$\nabla_k \log p(D|\mathbf{f}) = \frac{\partial \log p(D|\mathbf{f})}{\partial f_k} = -\tau\sum_{i=1}^{T}\Theta(t_i-t_k)\sum_{j=1}^{m}\lambda_{ij}(x_j(t_i)-y_{ij})S_j\mathrm{e}^{f_k-D_j(t_i-t_k)} \qquad (11)$$

$$\begin{aligned}
W_{kl} = -\nabla_k\nabla_l \log p(D|\mathbf{f}) &= \delta_{kl}\tau\sum_{i=1}^{T}\Theta(t_i-t_k)\sum_{j=1}^{m}\lambda_{ij}(x_j(t_i)-y_{ij})S_j\mathrm{e}^{f_k-D_j(t_i-t_k)} \\
&\quad +\tau^2\sum_{i=1}^{T}\Theta(t_i-t_k)\Theta(t_i-t_l)\sum_{j=1}^{m}\lambda_{ij}S_j^2\mathrm{e}^{f_k+f_l-D_j(2t_i-t_k-t_l)} \\
&= -\delta_{kl}\nabla_k \log p(D|\mathbf{f}) + \tau^2\sum_{i=1}^{T}\Theta(t_i-t_k)\Theta(t_i-t_l)\sum_{j=1}^{m}\lambda_{ij}S_j^2\mathrm{e}^{f_k+f_l-D_j(2t_i-t_k-t_l)} \;. \qquad (12)
\end{aligned}$$

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,

$$\frac{\partial K_{kl}}{\partial\gamma} = -\tfrac{1}{2}(t_k-t_l)^2K_{kl} \;.$$

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

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

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 (9) which involves W. We find,

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

At the MAP solution we find,

$$\frac{\partial W_{pq}}{\partial \hat{f}_k} = (\delta_{kp}+\delta_{kq})W_{pq} \;.$$

where we have used the self-consistent condition that the first term in equation (12) vanishes at the MAP solution. This then simplifies to,

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

$>$From equation (12) we have the self-consistent equation $\hat{\mathbf{f}} = K\nabla \log p(D|\hat{\mathbf{f}})$ and differentiating that we get,

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

3.2 Variational

The solution to the differential equation is given by

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

The second moment of the process $x_i\left(t\right) = \left(y_i\left(t\right) - \frac{B_i}{D_i}\right) \frac{e^{D_i t}}{S_i}$ is given by

$$\left\langle \int_0^t \int_0^{t'} g\left(f\left(u\right)\right) g\left(f\left(u'\right)\right) \exp\left(-D_i u - D_j u'\right) du du' \right\rangle_{p(f(t))}$$

$$\left\langle x_i x_i^{\mathrm{T}} \right\rangle = \int_0^t \int_0^{t'} \left\langle g\left(f\left(u\right)\right) g\left(f\left(u'\right)\right) \right\rangle_{p(f)} \exp\left(-D_i u - D_j u'\right) du du'$$

So the relevant integral is $\int g\left(\mathbf{f}\right) g\left(\mathbf{f}^{\mathrm{T}}\right) N\left(\mathbf{f}|\mathbf{0}, \mathbf{K}\right) d\mathbf{f}$ which we transform to

$$\int$$

so the i, jth element of the intergral is

$$\int g\left(f_i\right) g\left(f_j\right) N\left(\left[f_i\, f_j\right] |\mathbf{0}, \mathbf{C}\right)$$

$$\frac{1}{\left(2\pi\right) \left|\mathbf{C}\right|^{\frac{1}{2}}} \int \exp\left(-\frac{1}{2} \left[f_i\, f_j\right] \mathbf{C}^{-1} \left[f_i\, f_j\right]^{\mathrm{T}} + \mathbf{a}^{\mathrm{T}} \left[f_i\, f_j\right]^{\mathrm{T}}\right) df_i df_j = \exp\left(\frac{\alpha^2}{2} \mathbf{1}^{\mathrm{T}} \mathbf{C} \mathbf{1}\right)$$

But the process will also have a mean

$$\begin{aligned} &= \int g\left(f_i\right) N\left(f_i|\mathbf{0}, k_{ii}\right) df_1 \\ &= \int \exp\left(-\frac{1}{2k_{ii}} f_i^2 + \alpha f_i\right) = \exp\left(\frac{1}{2} k_{ii} \alpha^2\right) \end{aligned}$$

where

$$\operatorname{cov}\left(g_i, g_j\right) = \exp\left(\frac{\alpha^2}{2}\left(k_{ii} + k_{jj} + 2k_{ij}\right)\right) - \exp\left(\frac{\alpha^2}{2}\left(k_{ii} + k_{jj}\right)\right)$$

for stationary (normalised) kernels this simplifies to

$$\operatorname{cov}\left(g_i, g_j\right) = \exp\left(\alpha^2\left(1 + k_{ij}\right)\right) - \exp\left(\alpha^2\right)$$

3.2.1 Reversing the Approximation

Let's assume that we have a non-linearity g(f(t)). We require this expectation to minimise the KL

$$\left\langle g\left(f\left(t\right)\right)\right\rangle_{p(f(t))}$$

We conjecture that this will result in an approximation of the form

$$\left\langle g\left(f\left(t\right)\right)\right\rangle_{p(f(t))} = h\left(\left\langle f\left(t\right)\right\rangle_{p(f(t))}\right)$$

If this conjecture is true then to reverse the approximation

$$h^{-1}\left(\langle g\left(f\left(t\right)\right)\rangle_{p(f(t))}\right)=\langle f\left(t\right)\rangle_{p(f(t))}$$

Now substitute

$$x\left(t\right)=g\left(f\left(t\right)\right)$$

But another way of writing this would be

$$h^{-1}\left(\langle x\left(t\right)\rangle\right)=\left\langle g^{-1}\left(x\left(t\right)\right)\right\rangle_{p(f(t))}$$

which can then be re-ordered to give

$$\langle x\left(t\right)\rangle=h\left(\left\langle g^{-1}\left(x\left(t\right)\right)\right\rangle_{p(f(t))}\right)$$

which implies that if $g\left(\cdot\right)=\exp\left(\cdot\right)$

$$h_{\exp}^{-1}\left(\langle y\left(t\right)\rangle\right)=\langle\log\left(y\left(t\right)\right)\rangle_{p(y(t))}$$

similarly we have

$$h_{\log}^{-1}\left(\langle f\left(t\right)\rangle\right)=\langle\exp\left(f\left(t\right)\right)\rangle_{p(f(t))}$$

$$y\left(t\right)$$

is a Gaussian process

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. 1