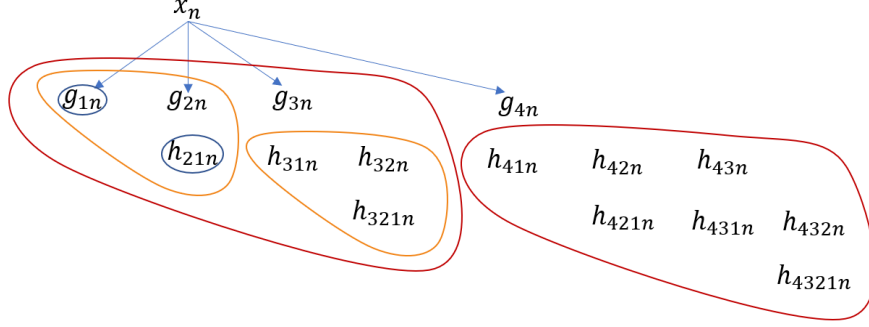


Model 1: Forward direction is in red, backward direction is in black, observations are in bold:

$\mathbf{h}_{1n} :$	$g_1(x_n)$	$h_{13}(g_3(x_n))$	$h_{12}(g_2(x_n))$	$h_{123}(h_{23}(g_3(x_n)))$
$\mathbf{h}_{2n} :$	$g_2(x_n)$	$h_{21}(g_1(x_n))$	$h_{23}(g_3(x_n))$	
$\mathbf{h}_{3n} :$	$g_3(x_n)$	$h_{31}(g_1(x_n))$	$h_{32}(g_2(x_n))$	$h_{321}(h_{21}(g_1(x_n)))$

The number of GPs will explode as the Pascal's Triangle.

For Model 1 forward version:



$$\begin{aligned}
p(\mathbf{y}, g_{1:3}, h_{21,31,32,321}) &= p(g_{1 \neq \mathbf{u}_1} | \mathbf{u}_1) p(g_{2 \neq \mathbf{u}_2} | \mathbf{u}_2) p(g_{3 \neq \mathbf{u}_3} | \mathbf{u}_3) p(\mathbf{u}_1) p(\mathbf{u}_2) p(\mathbf{u}_3) \\
&\quad p(h_{21 \neq \mathbf{v}_1} | \mathbf{v}_1) p(h_{31 \neq \mathbf{v}_2} | \mathbf{v}_2) p(h_{32 \neq \mathbf{v}_3} | \mathbf{v}_3) p(\mathbf{v}_1) p(\mathbf{v}_2) p(\mathbf{v}_3) \\
&\quad p(h_{321 \neq \mathbf{w}_1} | \mathbf{w}_1) p(\mathbf{w}_1) \\
&\quad \prod_n p(y_{1n}; g_1(x_n), \sigma_1^2) \\
&\quad p(y_{2n}; g_2(x_n) + h_{21}(g_1(x_n)), \sigma_2^2) \\
&\quad p(y_{3n}; g_3(x_n) + h_{31}(g_1(x_n)) + h_{32}(g_2(x_n)) + h_{321}(h_{21}(g_1(x_n))), \sigma_3^2)
\end{aligned}$$

$$\begin{aligned}
q(g_{1:3}, h_{21,31,321}) &= p(g_{1 \neq \mathbf{u}_1} | \mathbf{u}_1) p(g_{2 \neq \mathbf{u}_2} | \mathbf{u}_2) p(g_{3 \neq \mathbf{u}_3} | \mathbf{u}_3) q(\mathbf{u}_1) q(\mathbf{u}_2) q(\mathbf{u}_3) \\
&\quad p(h_{21 \neq \mathbf{v}_1} | \mathbf{v}_1) p(h_{31 \neq \mathbf{v}_2} | \mathbf{v}_2) p(h_{32 \neq \mathbf{v}_3} | \mathbf{v}_3) q(\mathbf{v}_1) q(\mathbf{v}_2) q(\mathbf{v}_3) \\
&\quad p(h_{321 \neq \mathbf{w}_1} | \mathbf{w}_1) q(\mathbf{w}_1)
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}_{ELBO} &= \mathbb{E}_q \left[ \frac{p(\mathbf{y}, g_{1:3}, h_{21,31,32,321})}{q(g_{1:3}, h_{21,31,321})} \right] \\
&= \sum_n \int q(g_{1n} | x_n) \log p(y_{1n} | g_{1n}, \sigma_1^2) \\
&\quad + \sum_n \int q(g_{1n} | x_n) q(g_{2n} | x_n) q(h_{21n} | g_{1n}) \log p(y_{2n} | g_{2n}, h_{21n}, \sigma_2^2) \\
&\quad + \sum_n \int q(g_{1n} | x_n) q(g_{2n} | x_n) q(g_{3n} | x_n) \\
&\quad \quad q(h_{21n} | g_{1n}) q(h_{31n} | g_{1n}) q(h_{32n} | g_{2n}) q(h_{321n} | h_{21n}) \log p(y_{3n} | g_{3n}, h_{31n}, h_{32n}, h_{321n}, \sigma_3^2) \\
&\quad - \sum_{\mathbf{u}: \text{all inducing points}} KL(q(\mathbf{u}) || p(\mathbf{u}))
\end{aligned}$$

where  $q(\cdot)$  is the variational posterior with inducing points analytically marginalized out. The following table shows the space where each GP's inducing inputs should lie.

$g_{1:3}$	$\mathcal{X}$
$h_{21,31}$	$g_1(\mathcal{X})$
$h_{32}$	$g_2(\mathcal{X})$
$h_{321}$	$h_{21}(g_2(\mathcal{X}))$

If bi-directional is implemented, the expectation log-term will be:

$$\sum_n \int q(g_{1n}|x_n)q(g_{2n}|x_n)q(g_{3n}|x_n) \left[ \int q(h_{13n}|g_{3n})q(h_{12n}|g_{2n})q(h_{23n}|g_{3n})q(h_{123n}|h_{23n}) \log p(y_{1n}|g_{1n}, h_{13n}, h_{12n}, h_{123n}, \sigma_1^2) \right. \\ \left. \int q(h_{21n}|g_{1n})q(h_{23n}|g_{3n}) \log p(y_{2n}|g_{2n}, h_{21n}, h_{23n}, \sigma_2^2) \right. \\ \left. \int q(h_{21n}|g_{1n})q(h_{31n}|g_{1n})q(h_{32n}|g_{2n})q(h_{321n}|h_{21n}) \log p(y_{3n}|g_{3n}, h_{31n}, h_{32n}, h_{321n}, \sigma_3^2) \right]$$

The above model can also be relaxed as Model 2:

$t_{1n} :$			$g_1(x_n)$	$g_{13}(b_{3n})$	$g_{12}(b_{2n})$	$: b_{1n}$
$t_{2n} :$		$g_{21}(t_{1n})$	$g_2(x_n)$	$g_{23}(b_{3n})$		$: b_{2n}$
$t_{3n} :$	$g_{32}(t_{2n})$	$g_{31}(t_{1n})$	$g_3(x_n)$			$: b_{1n}$

where the hidden variables in two directions share the GP over input as follows,

$$\begin{aligned} t_{1n} &= g_1(x_n) \\ t_{2n} &= g_2(x_n) + g_{21}(t_{1n}) \\ t_{3n} &= g_3(x_n) + g_{32}(t_{2n}) + g_{31}(t_{1n}) \end{aligned}$$

$$\begin{aligned} b_{3n} &= g_3(x_n) \\ b_{2n} &= g_2(x_n) + g_{23}(b_{3n}) \\ b_{1n} &= g_1(x_n) + g_{12}(b_{2n}) + g_{13}(b_{3n}) \end{aligned}$$

$$\begin{aligned} h_{1n} &= g_1(x_n) + g_{12}(b_{2n}) + g_{13}(b_{3n}) \\ h_{2n} &= g_2(x_n) + g_{21}(t_{1n}) + g_{23}(b_{3n}) \\ h_{3n} &= g_3(x_n) + g_{32}(t_{2n}) + g_{31}(t_{1n}) \end{aligned}$$

Just like Model 1, Model 2's backwards direction does not have any duplicated function over time. The number of GPs is  $L^2$ , where  $L$  is number of outputs. However, the dependencies will still run through aggregated hidden variables, except that time and each previous output are explicitly separated.

Here is the connection of all three model:

