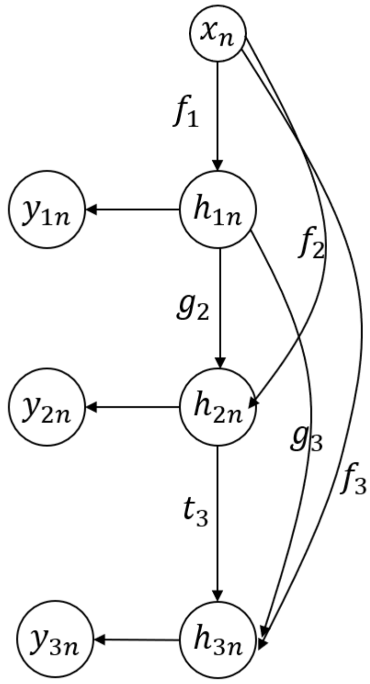


Figure 1: GPLAR with separate GPs

As shown in Figure 1, h_{2n} is generated by sum of two Gaussian process value, which is the following:

$$
\begin{aligned}
p(h_{1n}) &= \mathcal{N}(h_{1n}; f_1(\mathbf{x}_n), \sigma^2) \\
p(h_{2n}) &= \mathcal{N}(h_{2n}; f_2(\mathbf{x}_n) + g_2(h_{1n}), \sigma^2) \\
p(h_{3n}) &= \mathcal{N}(h_{3n}; f_3(\mathbf{x}_n) + g_3(h_{1n}) + t_3(h_{2n}), \sigma^2)
\end{aligned}
$$

Introducing inducing-points into each GP function, we have the approximate posterior and the complete probability of the model as follows,

$$
\begin{aligned}
p(f_{1:3}, g_{2:3}, t_3, \mathbf{h}_{1:3}, \mathbf{y}_{1:3}) =& p(f_{1\neq\mathbf{u}_1}|\mathbf{u}_1)p(\mathbf{u}_1) \\
& p(f_{2\neq\mathbf{u}_2}|\mathbf{u}_2)p(\mathbf{u}_2)p(g_{2\neq\mathbf{v}_2}|\mathbf{v}_2)p(\mathbf{v}_2) \\
& p(f_{3\neq\mathbf{u}_3}|\mathbf{u}_3)p(\mathbf{u}_3)p(g_{3\neq\mathbf{v}_3}|\mathbf{v}_3)p(\mathbf{v}_3)p(t_{3\neq\mathbf{w}_3}|\mathbf{w}_3)p(\mathbf{w}_3)\times \\
& \prod_n p(h_{1n}|f_1, \mathbf{x}_n)p(h_{2n}|f_2, g_2, \mathbf{x}_n, h_{1n})p(h_{3n}|f_3, g_3, t_3, \mathbf{x}_n, h_{1n}, h_{2n}) \\
& \quad p(y_{1n}|h_{1n})p(y_{2n}|h_{2n})p(y_{3n}|h_{3n})
\end{aligned}
$$

$$
\begin{aligned}
q(f_{1:3}, g_{2:3}, t_3, \mathbf{h}_{1:3}) =& p(f_{1\neq\mathbf{u}_1}|\mathbf{u}_1)q(\mathbf{u}_1) \\
& p(f_{2\neq\mathbf{u}_2}|\mathbf{u}_2)q(\mathbf{u}_2)p(g_{2\neq\mathbf{v}_2}|\mathbf{v}_2)q(\mathbf{v}_2) \\
& p(f_{3\neq\mathbf{u}_3}|\mathbf{u}_3)q(\mathbf{u}_3)p(g_{3\neq\mathbf{v}_3}|\mathbf{v}_3)q(\mathbf{v}_3)p(t_{3\neq\mathbf{w}_3}|\mathbf{w}_3)q(\mathbf{w}_3)\times \\
& \prod_n p(h_{1n}|f_1, \mathbf{x}_n)p(h_{2n}|f_2, g_2, \mathbf{x}_n, h_{1n})p(h_{3n}|f_3, g_3, t_3, \mathbf{x}_n, h_{1n}, h_{2n})
\end{aligned}
$$

Lower bound to the marginal log-likelihood becomes,

$$
\begin{aligned}
\mathcal{L}_{ELBO} &= \mathbb{E}_q\left[\log \frac{p(f_{1:3}, g_{2:3}, t_3, \mathbf{h}_{1:3}, \mathbf{y}_{1:3})}{q(f_{1:3}, g_{2:3}, t_3, \mathbf{h}_{1:3})}\right] \\
&= \sum_{l,n} \mathbb{E}_q \log p(y_{ln}|h_{ln}) - \sum_{l=1}^{3} KL\left[q(\mathbf{u}_l)\|p(\mathbf{u}_l)\right] - \sum_{l=2}^{3} KL\left[q(\mathbf{v}_l)\|p(\mathbf{v}_l)\right] - \sum_{l=3}^{3} KL\left[q(\mathbf{w}_l)\|p(\mathbf{w}_l)\right]
\end{aligned}
$$

where the expectation log-term requires computation of the following:

$$
\begin{aligned}
&\int p(f^1_{l\neq \mathbf{u}^1_l}|\mathbf{u}^1_l)q(u^1_l)\dots p(f^l_{l\neq \mathbf{u}^l_l}|\mathbf{u}^l_l)q(u^l_l)p(h_{ln}|f^{1\dots l}_l, \mathbf{x}_n, h_{1n,\dots,(l-1)n})\\
&\quad q(h_{1n})q(h_{2n}|h_{1n})\dots q(h_{(l-1)n}|h_{1n}\dots h_{(l-2)n})\\
&\quad \log p(y_{ln}|h_{ln})d\mathbf{h}
\end{aligned}
$$

where $p(h_{ln}|f^{1\dots l}_l, \mathbf{x}_n, h_{1n,\dots,(l-1)n}) = \mathcal{N}(h_{ln}; f^1_l(\mathbf{x}_n) + f^2_l(h_{1n}) + \dots + f^l_l(h_{(l-1)n}), \sigma^2)$ and the function $f^{1\dots l}_l$ can be integrated out analytically as follows,

$$
\begin{aligned}
q(h_{ln}|h_{1n},\dots,h_{(l-1)n}) &= \int p(f^1_{l\neq \mathbf{u}^1_l}|\mathbf{u}^1_l)q(u^1_l)\dots p(f^l_{l\neq \mathbf{u}^l_l}|\mathbf{u}^l_l)q(u^l_l)p(h_{ln}|f^{1\dots l}_l, \mathbf{x}_n, h_{1n,\dots,(l-1)n})df\\
&= \mathcal{N}(h_{ln}; \mu_l(\hat{\mathbf{x}}_n), \sigma^2_l(\hat{\mathbf{x}}_n))
\end{aligned}
$$

where,

$$
\begin{aligned}
\mu_l(\hat{\mathbf{x}}_n) =& K^1_l(\mathbf{x}_n, \mathbf{Z}^1_l)K^1_l(\mathbf{Z}^1_l, \mathbf{Z}^1_l)^{-1}\mathbf{m}^1_l + K^2_l(h_{1n}, \mathbf{Z}^2_l)K^2_l(\mathbf{Z}^2_l, \mathbf{Z}^2_l)^{-1}\mathbf{m}^2_l + \dots + K^l_l(h_{(l-1)n}, \mathbf{Z}^l_l)K^l_l(\mathbf{Z}^l_l, \mathbf{Z}^l_l)^{-1}\mathbf{m}^l_l\\
\sigma^2_l(\hat{\mathbf{x}}_n) =& K^1_l(\mathbf{x}_n, \mathbf{x}_n) + K^1_l(\mathbf{x}_n, \mathbf{Z}^1_l)K^1_l(\mathbf{Z}^1_l, \mathbf{Z}^1_l)^{-1}(K^1_l(\mathbf{Z}^1_l, \mathbf{Z}^1_l) - \mathbf{S}^1_l)K^1_l(\mathbf{Z}^1_l, \mathbf{Z}^1_l)^{-1}K^1_l(\mathbf{Z}^1_l, \mathbf{x}_n)\\
&+ K^2_l(h_{1n}, h_{1n}) + K^2_l(h_{1n}, \mathbf{Z}^2_l)K^2_l(\mathbf{Z}^2_l, \mathbf{Z}^2_l)^{-1}(K^2_l(\mathbf{Z}^2_l, \mathbf{Z}^2_l) - \mathbf{S}^2_l)K^2_l(\mathbf{Z}^2_l, \mathbf{Z}^2_l)^{-1}K^2_l(\mathbf{Z}^2_l, h_{1n})\\
&+ \dots\\
&+ K^l_l(h_{(l-1)n}, h_{(l-1)n}) + K^l_l(h_{(l-1)n}, \mathbf{Z}^l_l)K^l_l(\mathbf{Z}^l_l, \mathbf{Z}^l_l)^{-1}(K^l_l(\mathbf{Z}^l_l, \mathbf{Z}^l_l) - \mathbf{S}^l_l)K^l_l(\mathbf{Z}^l_l, \mathbf{Z}^l_l)^{-1}K^l_l(\mathbf{Z}^l_l, h_{(l-1)n})\\
&+ \sigma^2
\end{aligned}
$$

Here, $\mathbf{Z}^2_l \dots \mathbf{Z}^l_l$ will be vectors, since input is univariate.

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.

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,

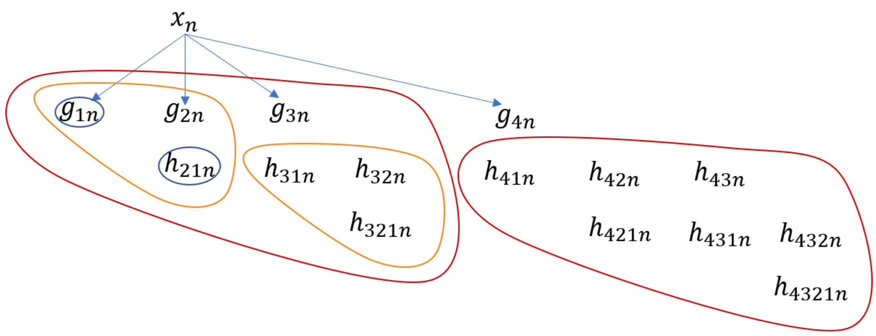
$$
\begin{aligned}
t_{1n} &= g_1(x_n)\\
t_{2n} &= g_2(x_n) + g_{21}(t_{1n})\\
t_{3n} &= g_3(x_n) + g_{32}(t_{2n}) + g_{31}(t_{1n})
\end{aligned}
$$

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

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

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.

For Model 1 forward version:



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

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

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

where $q(\cdot|\cdot)$ 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:

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