Pedantic description of the svGPFA theory

Joaquín Rapela

October 8, 2022

Abstract

Here we provide a pedantic description of the sparse variational Gaussian process factor analysis (svGPFA) theory, expanding on Duncker and Sahani (2018).

1 GPFA model

$$
\begin{aligned}
p(\{x_{kr}(\cdot)\}_{k=1,r=1}^{K,R}) &= \prod_{r=1}^{R}\prod_{k=1}^{K} p(x_{kr}(\cdot)) \\
x_{kr}(\cdot) &\sim \mathcal{GP}(\mu_k(\cdot), \kappa_k(\cdot,\cdot)) \qquad \text{for } k=1,\ldots,K \text{ and } r=1,\ldots,R \\
h_{nr}(\cdot) &= \sum_{k=1}^{K} c_{nk}x_{kr}(\cdot) + d_n \qquad \text{for } n=1,\ldots,N \text{ and } r=1,\ldots,R \\
p(\{\mathbf{y}_{nr}\}_{n=1,r=1}^{N,R}|\{h_{nr}(\cdot)\}_{n=1,r=1}^{N,R}) &= \prod_{r=1}^{R}\prod_{n=1}^{N} p(\mathbf{y}_{nr}|h_{nr}(\cdot))
\end{aligned}
\tag{1}
$$

where $x_{kr}(\cdot)$ is the latent process k in trial r, $h_{nr}(\cdot)$ is the embedding process for neuron n and trial r and $\mathbf{y}_{nr}$ is the activity of neuron n in trial r.

Notes:

- the first equation shows that the latent processes are independent,
- the second equation shows that the latent processes share mean and covariance functions across trials. That is, for any k, the mean and covariance functions of latents processes of different trials, $x_{kr}(\cdot), r = 1,\ldots,R$, are the same ($\mu_k(\cdot)$ and $\kappa_k(\cdot,\cdot)$),
- the fourth equation shows that, given the embedding processes, the responses of different neurons are independent.

2 svGPFA prior

To use the sparse variational framework for Gaussian processes Duncker and Sahani (2018) augmented the GPFA model by introducing inducing points $\mathbf{u}_{kr}$ for each latent process k and trial r. The inducing points $\mathbf{u}_{kr}$ represent evaluations of the latent process $x_{kr}(\cdot)$ at the M_{kr} locations $\mathbf{z}_{kr}$. A joint prior over the latent process $x_{kr}(\cdot)$ and its inducing points $\mathbf{u}_{kr}$ is given in Eq. 2.

$$
\begin{aligned}
p(\{x_{kr}(\cdot)\}_{k=1,r=1}^{K,R},\{\mathbf{u}_{kr}\}_{k=1,r=1}^{K,R}) &= p(\{x_{kr}(\cdot)\}_{k=1,r=1}^{K,R}|\{\mathbf{u}_{kr}\}_{k=1,r=1}^{K,R})p(\{\mathbf{u}_{kr}\}_{k=1,r=1}^{K,R}) \\
p(\{x_{kr}(\cdot)\}_{k=1,r=1}^{K,R}|\{\mathbf{u}_{kr}\}_{k=1,r=1}^{K,R}) &= \prod_{k=1}^{k}\prod_{r=1}^{R} p(x_{kr}(\cdot)|\mathbf{u}_{kr}) \\
p(\{\mathbf{u}_{kr}\}_{k=1,r=1}^{K,R}) &= \prod_{k=1}^{k}\prod_{r=1}^{R} p(\mathbf{u}_{kr}) \\
p(\mathbf{u}_{kr}) &= \mathcal{N}(\mathbf{0}, K_{zz}^{kr})
\end{aligned}
\tag{2}
$$

where $K_{zz}^{(kr)}[i,j] = \kappa_k(\mathbf{z}_{kr}[i], \mathbf{z}_{kr}[j])$.

We next derive the functional form of $p(x_{kr}(\cdot)|\mathbf{u}_{kr})$.

For any $P \in \mathbb{N}$, define the random vector $\mathbf{x}_{kr}$ as the random process $x_{kr}(\cdot)$ evaluated at times $\mathbf{t}_P^{(r)} := \left\{t_1^{(r)}, \ldots, t_P^{(r)}\right\}$. That is,

$$
\mathbf{x}_{kr} := [x_{kr}(t_1^{(r)}), \ldots, x_{kr}(t_P^{(r)})]^\intercal \tag{3}
$$

Then, because the inducing points $\mathbf{u}_{kr}$ are evaluations of the latent process $x_{kr}(\cdot)$ at $\mathbf{z}_{kr}$, $\mathbf{x}_{kr}$ and $\mathbf{u}_{kr}$ are jointly Gaussian:

$$
p\left(\left[\begin{array}{c}\mathbf{u}_{kr}\\ \mathbf{x}_{kr}\end{array}\right]\right) = \mathcal{N}\left(\left[\begin{array}{c}\mathbf{u}_{kr}\\ \mathbf{x}_{kr}\end{array}\right]\middle|\left[\begin{array}{c}\mathbf{0}\\ \mathbf{0}\end{array}\right], \left[\begin{array}{cc}K_{\mathbf{zz}}^{(kr)} & K_{\mathbf{zt}}^{(kr)}\\ K_{\mathbf{tz}}^{(kr)} & K_{\mathbf{tt}}^{(r)}\end{array}\right]\right) \tag{4}
$$

where $K_{\mathbf{tz}}^{(kr)}[i,j] = \kappa_k(t_i^{(r)}, \mathbf{z}_{kr}[j])$, $K_{\mathbf{zt}}^{(kr)}[i,j] = \kappa_k(\mathbf{z}_{kr}[i], t_j^{(r)})$ and $K_{\mathbf{tt}}^{(r)}[i,j] = \kappa_k(t_i^{(r)}, t_j^{(r)})$.

Now, applying the formula for the conditional pdf of jointly Normal random vectors (Bishop, 2016, Eq. 2.116) to Eq. 4, we obtain

$$
p(\mathbf{x}_{kr}|\mathbf{u}_{kr}) = \mathcal{N}\left(\mathbf{x}_{kr}\middle|K_{\mathbf{tz}}^{(kr)}\left(K_{zz}^{(kr)}\right)^{-1}\mathbf{u}_{kr},\ K_{\mathbf{tt}}^{(r)} - K_{\mathbf{tz}}^{(kr)}\left(K_{zz}^{(kr)}\right)^{-1}K_{\mathbf{zt}}^{(kr)}\right) \tag{5}
$$

Because Eq. 5 is valid for any $\mathbf{t}_P^{(r)}$, and any $\mathbf{x}_{kr}$ derived from it, it follows that

$$
\begin{aligned}
p(x_{kr}(\cdot)|\mathbf{u}_{kr}) &= \mathcal{GP}\left(\tilde{\mu}_{kr}(\cdot), \tilde{\kappa}_{kr}(\cdot,\cdot)\right) \\
\tilde{\mu}_{kr}(t) &= \kappa_k(t, \mathbf{z}_{kr})\left(K_{zz}^{(kr)}\right)^{-1}\mathbf{u}_{kr} \\
\tilde{\kappa}_k(t,t') &= \kappa_k(t,t') - \kappa_k(t,\mathbf{z}_{kr})\left(K_{zz}^{(kr)}\right)^{-1}\kappa_k(\mathbf{z}_{kr}, t')
\end{aligned}
\tag{6}
$$

which is Eq. 3 in Duncker and Sahani (2018).

We obtain the svGPFA prior on the latents by using the marginal formula of the linear Gaussian model (Bishop, 2016, Eq. 2.115) with the last line in Eq. 4 and with Eq. 5, yielding

$$
p(\mathbf{x}_{kr}) = \mathcal{N}\left(\mathbf{x}_{kr}\middle|\mathbf{0},\ K_{\mathbf{tt}}^{(r)}\right) \tag{7}
$$

Because Eq. 5 is valid for any $\mathbf{t}_P^{(r)}$, and any $\mathbf{x}_{kr}$ derived from it, it follows that

$$
p(x_{kr}(\cdot)) = \mathcal{GP}\left(0(\cdot),\ \kappa_k(\cdot,\cdot)\right)
$$

which shows that the svGPFA prior on the latent processes is unrelated to the inducing points and identical to the GPFA one (second line in Eq. 1, with $\mu_k(\cdot) = 0(\cdot)$).

3 Variational distribution of the latent process $x_{kr}(\cdot)$

From Eq. 3 in Duncker and Sahani (2018, supplementary), the approximate joint posterior for the latent process $x_{kr}(\cdot)$ and the inducing points $\mathbf{u}_{kr}$ is

$$q(x_{kr}(\cdot), \mathbf{u}_{kr}) = p(x_{kr}(\cdot)|\mathbf{u}_{kr})q(\mathbf{u}_{kr}) \tag{8}$$

with $p(x_{kr}(\cdot)|\mathbf{u}_{kr})$ given in Eq. 6 and

$$q(\mathbf{u}_{kr}) = \mathcal{N}(\mathbf{u}_{kr}(\cdot)|\mathbf{m}_{kr}, S_k) \tag{9}$$

Thus, for any random vector $\mathbf{x}_{kr}$, as in Eq. 3, we have

$$q(\mathbf{x}_{kr}, \mathbf{u}_{kr}) = p(\mathbf{x}_{kr}|\mathbf{u}_{kr})q(\mathbf{u}_{kr}) \tag{10}$$

with $p(\mathbf{x}_{kr}|\mathbf{u}_{kr})$ given in Eq. 5. Using the marginal formula for the linear Gaussian model (Bishop, 2016, Eq. 2.115) with Eqs. 5 and 9 we obtain

$$q(\mathbf{x}_{kr}) = \mathcal{N}\left(\mathbf{x}_{kr}|K_{tz}^{kr}(K_{zz}^{kr})^{-1}\mathbf{m}_{kr},\ K_{tt}^{k} + K_{tz}^{kr}\left((K_{zz}^{kr})^{-1}S_{kr}(K_{zz}^{kr})^{-1} - (K_{zz}^{kr})^{-1}\right)K_{zt}^{kr}\right) \tag{11}$$

Because Eq. 11 is valid for any $\mathbf{t}_P^{(r)}$, and any $\mathbf{x}_{kr}$ derived from it, it follows that

$$\begin{aligned}
q(x_{kr}(\cdot)) &= \mathcal{GP}\left(\breve{\mu}_{kr}(\cdot), \breve{\kappa}_{kr}(\cdot,\cdot)\right)\\
\breve{\mu}_{kr}(t) &= \kappa_k(t, z_{kr})(K_{zz}^{kr})^{-1}\mathbf{m}_{kr},\\
\breve{\kappa}_{kr}(t,t') &= \kappa_k(t,t') + \kappa_k(t, z_{kr})\left((K_{zz}^{kr})^{-1}S_{kr}(K_{zz}^{kr})^{-1} - (K_{zz}^{kr})^{-1}\right)\kappa_k(z_{kr}, t')
\end{aligned} \tag{12}$$

4 Variational distribution of $h_{nr}(\cdot)$

Finally, because affine trasformations of Gaussians are Gaussians, $h_{nr}(\cdot)$ is an affine transformation of $\{x_{kr}(\cdot)\}$ (which are Gaussians, Eq. 12), then the approximate posterior of $h_{nr}(\cdot)$ is the Gaussian process in Eq. 13.

$$\begin{aligned}
q(h_{nr}(\cdot)) &= \mathcal{GP}\left(\tilde{\mu}_{nr}(\cdot), \tilde{\kappa}_{nr}(\cdot,\cdot)\right)\\
\tilde{\mu}_{nr}(t) &= \sum_{k=1}^{K} c_{nk}\breve{\mu}_{kr}(t)\\
\tilde{\kappa}_{nr}(t,t') &= \sum_{k=1}^{K} c_{nk}^2\breve{\kappa}_{kr}(t,t')
\end{aligned} \tag{13}$$

which is Eq. 5 in Duncker and Sahani (2018).

5 Derivation of the svGPFA variational lower bound

To derive the sparse variational lower bound we have the freedom of using $\{x_{kr}(\cdot)\}_{k=1,r=1}^{K,R}$ or $\{h_{nr}(\cdot)\}_{n=1,r=1}^{N,R}$ as latent processes in the complete-data likelihood[1]. In (Duncker and Sahani, 2018, supplementary) the

[1] To simplify notation below we write $\{x_{kr}(\cdot)\}_{k=1,r=1}^{K,R}$ as $\{x_{kr}(\cdot)\}$ and similarly for $\{\mathbf{u}_{kr}\}_{k=1,r=1}^{K,R}$ and $\{h_{nr}(\cdot)\}_{n=1,r=1}^{N,R}$

authors used the former, which requires the non-trivial proof of Eq. 14.

$$\mathrm{E}_{q(\{x_{kr}(\cdot)\})} \log p(\{\mathbf{y}_{nr}\}|\{x_{kr}(\cdot)\}) = \mathrm{E}_{q(\{h_{nr}(\cdot)\})} \log p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}) \tag{14}$$

Here we use the latter, which avoids the need of Eq. 14. Then the complete data likelihood is given in Eq. 15 and a variational lower bound in Eq. 16.

$$p(\{\mathbf{y}_{nr}\}, \{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) = p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\})p(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\})p(\{\mathbf{u}_{kr}\}) \tag{15}$$

$$\log p(\{\mathbf{y}_{nr}\}) \geq E_{q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})} \log\left(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}\right) - KL\left\{q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})||p(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})\right\} \tag{16}$$

As in (Duncker and Sahani, 2018, supplementary), we choose the approximating variational distribution in Eq. 17.

$$\begin{aligned}
q(\{x_{kr}(\cdot)\}, \{\mathbf{u}_{kr}\}) &= q(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\})q(\{\mathbf{u}_{kr}\}) \\
q(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\}) &= p(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\}) \\
q(\{\mathbf{u}_{kr}\}) &= \prod_{r=1}^{R}\prod_{k=1}^{K} q(\mathbf{u}_{kr}) \\
q(\mathbf{u}_{kr}) &= \mathcal{N}(\mathbf{u}_{kr}|\mathbf{m}_{kr}, S_{kr})
\end{aligned} \tag{17}$$

Because $h_{nr}(\cdot)$ is a function of $\{x_{kr}(\cdot)\}$ only (Eq. 1), and because $q(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\}) = p(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\})$ (Eq. 17), then

$$q(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\}) = p(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\}) \tag{18}$$

The first term in Eq. 16 can be rewritten as

$$
\begin{aligned}
&E_{q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})} \log\left(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\}\right) = \\
&\int\int q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\}) \log\left(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\}\right) d\{\mathbf{u}_{kr}\}d\{h_{nr}(\cdot)\} = \\
&\int\int q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\}) \log\left(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}\right) d\{\mathbf{u}_{kr}\}d\{h_{nr}(\cdot)\} = \\
&\int\left(\int q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})d\{\mathbf{u}_{kr}\}\right) \log\left(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}\right) d\{h_{nr}(\cdot)\} = \\
&\int q(\{h_{nr}(\cdot)\}) \log\left(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}\right) d\{h_{nr}(\cdot)\} = \\
&\int q(\{h_{nr}(\cdot)\}) \log\left(\prod_{r=1}^{R}\prod_{n=1}^{N} p(\mathbf{y}_{nr}|h_{nr}(\cdot))\right) d\{h_{nr}(\cdot)\} = \\
&\sum_{r=1}^{R}\sum_{n=1}^{N}\int q(\{h_{nr}(\cdot)\}) \log\left(p(\mathbf{y}_{nr}|h_{nr}(\cdot))\right) d\{h_{nr}(\cdot)\} = \\
&\sum_{r=1}^{R}\sum_{n=1}^{N}\int\int q(\{h_{n'r'}(\cdot)\}_{n'\neq n,r'\neq r}|h_{nr}(\cdot))q(h_{nr}(\cdot)) \log\left(p(\mathbf{y}_{nr}|h_{nr}(\cdot))\right) d\{h_{n'r'}(\cdot)\}_{n'\neq n,r'\neq r}dh_{nr}(\cdot) = \\
&\sum_{r=1}^{R}\sum_{n=1}^{N}\int\left(\int q(\{h_{n'r'}(\cdot)\}_{n'\neq n,r'\neq r}|h_{nr}(\cdot))d\{h_{n'r'}(\cdot)\}_{n'\neq n,r'\neq r}\right) q(h_{nr}(\cdot)) \log\left(p(\mathbf{y}_{nr}|h_{nr}(\cdot))\right) dh_{nr}(\cdot) = \\
&\sum_{r=1}^{R}\sum_{n=1}^{N}\int q(h_{nr}(\cdot)) \log\left(p(\mathbf{y}_{nr}|h_{nr}(\cdot))\right) dh_{nr}(\cdot) = \\
&\sum_{r=1}^{R}\sum_{n=1}^{N} E_{q(h_{nr}(\cdot))} \log\left(p(\mathbf{y}_{nr}|h_{nr}(\cdot))\right)
\end{aligned}
\tag{19}
$$

Notes:

– the second line follows from the previous one because, given the embedding process, activities of neurons are independent of inducing points (Eq. 15),

– the fifth line follows from the previous one because, given embedding processes, responses of neurons are independent of each other (last line in Eq. 1), The second term in Eq. 16 can be rewritten as

$$
\begin{aligned}
KL\left\{q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})||p(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})\right\} &= \int\int q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})\log\frac{q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})}{p(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})}d\{h_{nr}(\cdot)\}d\{\mathbf{u}_{kr}\}\\
&= \int\int q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})\log\frac{q(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\})q(\{\mathbf{u}_{kr}\})}{p(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\})p(\{\mathbf{u}_{kr}\})}d\{h_{nr}(\cdot)\}d\{\mathbf{u}_{kr}\}\\
&= \int\int q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})\log\frac{q(\{\mathbf{u}_{kr}\})}{p(\{\mathbf{u}_{kr}\})}d\{h_{nr}(\cdot)\}d\{\mathbf{u}_{kr}\}\\
&= \int\int q(\{h_{nr}(\cdot)\},\{\mathbf{u}_{kr}\})d\{h_{nr}(\cdot)\}\log\frac{q(\{\mathbf{u}_{kr}\})}{p(\{\mathbf{u}_{kr}\})}d\{\mathbf{u}_{kr}\}\\
&= \int q(\{\mathbf{u}_{kr}\})\log\frac{q(\{\mathbf{u}_{kr}\})}{p(\{\mathbf{u}_{kr}\})}d\{\mathbf{u}_{kr}\}\\
&= \int q(\{\mathbf{u}_{kr}\})\log\frac{\prod_{r=1}^{R}\prod_{k=1}^{K}q(\mathbf{u}_{kr})}{\prod_{r=1}^{R}\prod_{k=1}^{K}p(\mathbf{u}_{kr})}d\{\mathbf{u}_{kr}\}\\
&= \int q(\{\mathbf{u}_{kr}\})\sum_{r=1}^{R}\sum_{k=1}^{K}\log\frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})}d\{\mathbf{u}_{kr}\}\\
&= \sum_{r=1}^{R}\sum_{k=1}^{K}\int q(\{\mathbf{u}_{kr}\})\log\frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})}d\{\mathbf{u}_{kr}\}\\
&= \sum_{r=1}^{R}\sum_{k=1}^{K}\int\prod_{k'=1}^{K}\prod_{r'=1}^{R}q(\mathbf{u}_{k'r'})\log\frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})}d\{\mathbf{u}_{kr}\}\\
&= \sum_{r=1}^{R}\sum_{k=1}^{K}\int q(\mathbf{u}_{kr})\log\frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})}d\mathbf{u}_{kr}\\
&= \sum_{r=1}^{R}\sum_{k=1}^{K}KL(q(\mathbf{u}_{kr})||p(\mathbf{u}_{kr}))
\end{aligned}
\tag{20}
$$

Notes:

- the third line follows from the previous one by Eq. 18,
- the tenth line follows from the previous because, for $k'\neq k$ or $r'\neq r$, the factors $q(\mathbf{u}_{k'r'})$ integrate out to one.

Replacing Eq. 19 and Eq. 20 into Eq. 16 we obtain

$$
\log p(\{\mathbf{y}_{nr}\})\geq\sum_{r=1}^{R}\sum_{n=1}^{N}E_{q(h_{nr}(\cdot))}\log\left(p(\mathbf{y}_{nr}|h_{nr}(\cdot))\right)-\sum_{r=1}^{R}\sum_{k=1}^{K}KL(q(\mathbf{u}_{kr})||p(\mathbf{u}_{kr}))
\tag{21}
$$

which is Eq. 4 in Duncker and Sahani (2018).

References

Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer-Verlag New York.

Duncker, L. and Sahani, M. (2018). Temporal alignment and latent gaussian process factor inference in population spike trains. In *Advances in Neural Information Processing Systems*, pages 10445–10455.