

# Pedantic derivations of the equations in the svGPFA paper ([Duncker and Sahani, 2018](#))

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 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. [1](#).

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

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  $\mathbf{t}_P^{(r)} := \{t_1^{(r)}, \dots, t_P^{(r)}\}$ . That is,

$$\mathbf{x}_{kr} := [x_{kr}(t_1^{(r)}), \dots, x_{kr}(t_P^{(r)})]^\top \tag{2}$$

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:

$$p\left(\begin{bmatrix} \mathbf{u}_{kr} \\ \mathbf{x}_{kr} \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \mathbf{u}_{kr} \\ \mathbf{x}_{kr} \end{bmatrix} \middle| \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} K_{zz}^{(kr)} & K_{zt}^{(kr)} \\ K_{tz}^{(kr)} & K_{tt}^{(r)} \end{bmatrix}\right) \tag{3}$$

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. 3, we obtain

$$p(\mathbf{x}_{kr}|\mathbf{u}_{kr}) = \mathcal{N}\left(\mathbf{x}_{kr} \left| K_{\mathbf{tz}}^{(kr)} \left(K_{zz}^{(kr)}\right)^{-1} \mathbf{u}_{kr}, K_{\mathbf{tt}}^{(r)} - K_{\mathbf{tz}}^{(kr)} \left(K_{zz}^{(kr)}\right)^{-1} K_{\mathbf{zt}}^{(kr)} \right.\right) \quad (4)$$

Because Eq. 4 is valid for any  $\mathbf{t}_P^{(r)}$ , and any  $\mathbf{x}_{kr}$  derived from it, it follows that

$$\begin{aligned} p(x_{kr}(\cdot)|\mathbf{u}_{kr}) &= \mathcal{GP}(\tilde{\mu}_{kr}(\cdot), \tilde{\kappa}_{kr}(\cdot, \cdot)) \\ \tilde{\mu}_{kr}(t) &= \kappa_k(t, \mathbf{z}_{kr}) \left(K_{zz}^{(kr)}\right)^{-1} \mathbf{u}_{kr} \\ \tilde{\kappa}_k(t, t') &= \kappa_k(t, t') - \kappa_k(t, \mathbf{z}_{kr}) \left(K_{zz}^{(kr)}\right)^{-1} \kappa_k(\mathbf{z}_{kr}, t') \end{aligned} \quad (5)$$

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. 3 and with Eq. 4, yielding

$$p(\mathbf{x}_{kr}) = \mathcal{N}\left(\mathbf{x}_{kr} \left| \mathbf{0}, K_{\mathbf{tt}}^{(r)} \right.\right) \quad (6)$$

Because Eq. 4 is valid for any  $\mathbf{t}_P^{(r)}$ , and any  $\mathbf{x}_{kr}$  derived from it, it follows that

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

which shows that the svGPFA prior on the latent processes is unrelated to the inducing points.

## 2 Approximate posterior on 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

$$q(x_{kr}(\cdot), \mathbf{u}_{kr}) = p(x_{kr}(\cdot)|\mathbf{u}_{kr})q(\mathbf{u}_{kr}) \quad (7)$$

with  $p(x_{kr}(\cdot)|\mathbf{u}_{kr})$  given in Eq. 5 and

$$q(\mathbf{u}_{kr}) = \mathcal{N}(\mathbf{u}_{kr}(\cdot)|\mathbf{m}_{kr}, S_k) \quad (8)$$

Thus, for any random vector  $\mathbf{x}_{kr}$ , as in Eq. 2, we have

$$q(\mathbf{x}_{kr}, \mathbf{u}_{kr}) = p(\mathbf{x}_{kr}|\mathbf{u}_{kr})q(\mathbf{u}_{kr}) \quad (9)$$

with  $p(\mathbf{x}_{kr}|\mathbf{u}_{kr})$  given in Eq. 4. Using the marginal formula for the linear Gaussian model (Bishop, 2016, Eq. 2.115) with Eqs. 4 and 8 we obtain

$$q(\mathbf{x}_{kr}) = \mathcal{N}\left(\mathbf{x}_{kr} \left| K_{tz}^{kr} (K_{zz}^{kr})^{-1} \mathbf{m}_{kr}, K_{tt}^k + K_{tz}^{kr} \left( (K_{zz}^{kr})^{-1} S_{kr} (K_{zz}^{kr})^{-1} - (K_{zz}^{kr})^{-1} \right) K_{zt}^{kr} \right.\right) \quad (10)$$

Because Eq. 10 is valid for any  $\mathbf{t}_P^{(r)}$ , and any  $\mathbf{x}_{kr}$  derived from it, it follows that

$$\begin{aligned}
q(x_{kr}(\cdot)) &= \mathcal{GP}(\check{\mu}_{kr}(\cdot), \check{\kappa}_{kr}(\cdot, \cdot)) \\
\check{\mu}_{kr}(t) &= \kappa_k(t, z_{kr})(K_{zz}^{kr})^{-1} \mathbf{m}_{kr}, \\
\check{\kappa}_{kr}(t, t') &= \kappa_k(t, t') + \kappa_k(t, z_{kr}) \left( (K_{zz}^{kr})^{-1} S_{kr} (K_{zz}^{kr})^{-1} - (K_{zz}^{kr})^{-1} \right) \kappa_k(z_{kr}, t')
\end{aligned} \tag{11}$$

### 3 Approximate posterior on the embedding $h_{nr}(\cdot)$

The embedding for neuron  $n$  and trial  $r$  is

$$h_{nr}(\cdot) = \sum_{k=1}^K c_{nk} x_{kr}(\cdot) \tag{12}$$

Affine transformations of Gaussian processes are Gaussian processes, therefore the approximate posterior on the embedding  $h_{nr}(\cdot)$  is a Gaussian process. That is,

$$q(h_{nr}(\cdot)) = q\left(\sum_{k=1}^K c_{nk} x_{kr}(\cdot)\right) = \mathcal{GP}(\tilde{\mu}_{nr}(\cdot), \tilde{\kappa}_{nr}(\cdot, \cdot)) \tag{13}$$

with

$$\begin{aligned}
\tilde{\mu}_{nr}(t) &= E_q\{h_{nr}(t)\} = \sum_{k=1}^K c_{nk} E_q\{x_{kr}(t)\} = \sum_{k=1}^K c_{nk} \check{\mu}_{kr}(t) \\
\tilde{\kappa}_{nr}(t, t') &= E_q\{(h_{nr}(t) - \tilde{\mu}_{nr}(t))(h_{nr}(t') - \tilde{\mu}_{nr}(t'))\} \\
&= E_q\left\{\left(\sum_{k=1}^K c_{nk}(x_{kr}(t) - \check{\mu}_{kr}(t))\right)\left(\sum_{k'=1}^K c_{nk'}(x_{k'r}(t') - \check{\mu}_{k'r}(t'))\right)\right\} \\
&= \sum_{k=1}^K \sum_{k'=1}^K c_{nk} c_{nk'} E_q\{(x_{kr}(t) - \check{\mu}_{kr}(t))(x_{k'r}(t') - \check{\mu}_{k'r}(t'))\} \\
&= \sum_{k=1}^K c_{nk}^2 E_q\{(x_{kr}(t) - \check{\mu}_{kr}(t))(x_{kr}(t') - \check{\mu}_{kr}(t'))\} \\
&= \sum_{k=1}^K c_{nk}^2 \check{\kappa}(t, t')
\end{aligned} \tag{14}$$

The second-to-last line in Eq. 14 follows from the independence assumption of the latent processes,  $x_{kr}(\cdot)$  and  $x_{k'r}(\cdot)$ ,  $k \neq k'$ , by the variational distribution  $q$  (Duncker and Sahani, 2018, supplementary).

### 4 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<sup>1</sup>. In (Duncker and Sahani, 2018, supplementary) the authors used the former, which requires the non-trivial proof of Eq. 15.

<sup>1</sup>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  $\{h_{nr}(\cdot)\}_{n=1, r=1}^{N, R}$

$$\mathbb{E}_{q(\{x_{kr}(\cdot)\})} \log p(\{\mathbf{y}_{nr}\}|\{x_{kr}(\cdot)\}) = \mathbb{E}_{q(\{h_{nr}(\cdot)\})} \log p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}) \quad (15)$$

Here we use the latter, which avoids the need of Eq. 15. Then the complete data likelihood is given in Eq. 16 and a variational lower bound in Eq. 17.

$$p(\{\mathbf{y}_{nr}\}, \{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) = p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\})p(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\})p(\{\mathbf{u}_{kr}\}) \quad (16)$$

$$\log p(\{\mathbf{y}_{nr}\}) \geq \mathbb{E}_{q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})} \log (p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) - KL \{q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})||p(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})\} \quad (17)$$

As in (Duncker and Sahani, 2018, supplementary), we choose the approximating variational distribution in Eq. 18.

$$\begin{aligned} q(\{x_{kr}(\cdot)\}, \{\mathbf{u}_{kr}\}) &= q(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\})q(\{\mathbf{u}_{kr}\}) \\ q(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\}) &= p(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\}) \\ q(\{\mathbf{u}_{kr}\}) &= \prod_{r=1}^R \prod_{k=1}^K q(\mathbf{u}_{kr}) \\ q(\mathbf{u}_{kr}) &= \mathcal{N}(\mathbf{u}_{kr}|\mathbf{m}_{kr}, S_{kr}) \end{aligned} \quad (18)$$

Because  $h_{nr}(\cdot)$  is a function of  $\{x_{kr}(\cdot)\}$  only (Eq. ??), and because  $q(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\}) = p(\{x_{kr}(\cdot)\}|\{\mathbf{u}_{kr}\})$  (Eq. 18), then

$$q(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\}) = p(\{h_{nr}(\cdot)\}|\{\mathbf{u}_{kr}\}) \quad (19)$$

The first term in Eq. 17 can be rewritten as

$$\begin{aligned}
& E_{q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})} \log(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})) = \\
& \int \int q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) \log(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})) d\{\mathbf{u}_{kr}\} d\{h_{nr}(\cdot)\} = \\
& \int \int q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) \log(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\})) d\{\mathbf{u}_{kr}\} d\{h_{nr}(\cdot)\} = \\
& \int \left( \int q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) d\{\mathbf{u}_{kr}\} \right) \log(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\})) d\{h_{nr}(\cdot)\} = \\
& \int q(\{h_{nr}(\cdot)\}) \log(p(\{\mathbf{y}_{nr}\}|\{h_{nr}(\cdot)\})) d\{h_{nr}(\cdot)\} = \\
& \int q(\{h_{nr}(\cdot)\}) \log \left( \prod_{r=1}^R \prod_{n=1}^N p(\mathbf{y}_{nr}|h_{nr}(\cdot)) \right) d\{h_{nr}(\cdot)\} = \\
& \sum_{r=1}^R \sum_{n=1}^N \int q(\{h_{nr}(\cdot)\}) \log(p(\mathbf{y}_{nr}|h_{nr}(\cdot))) d\{h_{nr}(\cdot)\} = \\
& \sum_{r=1}^R \sum_{n=1}^N \int \int q(\{h_{n'r'}(\cdot)\}_{n' \neq n, r' \neq r} | h_{nr}(\cdot)) q(h_{nr}(\cdot)) \log(p(\mathbf{y}_{nr}|h_{nr}(\cdot))) d\{h_{n'r'}(\cdot)\}_{n' \neq n, r' \neq r} dh_{nr}(\cdot) = \\
& \sum_{r=1}^R \sum_{n=1}^N \int \left( \int q(\{h_{n'r'}(\cdot)\}_{n' \neq n, r' \neq r} | h_{nr}(\cdot)) d\{h_{n'r'}(\cdot)\}_{n' \neq n, r' \neq r} \right) q(h_{nr}(\cdot)) \log(p(\mathbf{y}_{nr}|h_{nr}(\cdot))) dh_{nr}(\cdot) = \\
& \sum_{r=1}^R \sum_{n=1}^N \int q(h_{nr}(\cdot)) \log(p(\mathbf{y}_{nr}|h_{nr}(\cdot))) dh_{nr}(\cdot) = \\
& \sum_{r=1}^R \sum_{n=1}^N E_{q(h_{nr}(\cdot))} \log(p(\mathbf{y}_{nr}|h_{nr}(\cdot)))
\end{aligned} \tag{20}$$

Notes:

- the second line follows from the previous one because, given the embedding process, activities of neurons are independent of inducing points (Eq. 16),
- the fifth line follows from the previous one because, given embedding processes, responses of neurons are independent of each other (last line in Eq. ??),

The second term in Eq. 17 can be rewritten as

$$\begin{aligned}
KL\{q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) || p(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})\} &= \int \int q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) \log \frac{q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})}{p(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\})} d\{h_{nr}(\cdot)\} d\{\mathbf{u}_{kr}\} \\
&= \int \int q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) \log \frac{q(\{h_{nr}(\cdot)\} | \{\mathbf{u}_{kr}\}) q(\{\mathbf{u}_{kr}\})}{p(\{h_{nr}(\cdot)\} | \{\mathbf{u}_{kr}\}) p(\{\mathbf{u}_{kr}\})} d\{h_{nr}(\cdot)\} d\{\mathbf{u}_{kr}\} \\
&= \int \int q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) \log \frac{q(\{\mathbf{u}_{kr}\})}{p(\{\mathbf{u}_{kr}\})} d\{h_{nr}(\cdot)\} d\{\mathbf{u}_{kr}\} \\
&= \int \int q(\{h_{nr}(\cdot)\}, \{\mathbf{u}_{kr}\}) d\{h_{nr}(\cdot)\} \log \frac{q(\{\mathbf{u}_{kr}\})}{p(\{\mathbf{u}_{kr}\})} d\{\mathbf{u}_{kr}\} \\
&= \int q(\{\mathbf{u}_{kr}\}) \log \frac{q(\{\mathbf{u}_{kr}\})}{p(\{\mathbf{u}_{kr}\})} d\{\mathbf{u}_{kr}\} \\
&= \int q(\{\mathbf{u}_{kr}\}) \log \frac{\prod_{r=1}^R \prod_{k=1}^K q(\mathbf{u}_{kr})}{\prod_{r=1}^R \prod_{k=1}^K p(\mathbf{u}_{kr})} d\{\mathbf{u}_{kr}\} \\
&= \int q(\{\mathbf{u}_{kr}\}) \sum_{r=1}^R \sum_{k=1}^K \log \frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})} d\{\mathbf{u}_{kr}\} \\
&= \sum_{r=1}^R \sum_{k=1}^K \int q(\{\mathbf{u}_{kr}\}) \log \frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})} d\{\mathbf{u}_{kr}\} \\
&= \sum_{r=1}^R \sum_{k=1}^K \int \prod_{k'=1}^K \prod_{r'=1}^R q(\mathbf{u}_{k'r'}) \log \frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})} d\{\mathbf{u}_{kr}\} \\
&= \sum_{r=1}^R \sum_{k=1}^K \int q(\mathbf{u}_{kr}) \log \frac{q(\mathbf{u}_{kr})}{p(\mathbf{u}_{kr})} d\mathbf{u}_{kr} \\
&= \sum_{r=1}^R \sum_{k=1}^K KL(q(\mathbf{u}_{kr}) || p(\mathbf{u}_{kr}))
\end{aligned} \tag{21}$$

Notes:

- the third line follows from the previous one by Eq. 19,
- 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. 20 and Eq. 21 into Eq. 17 we obtain

$$\log p(\{\mathbf{y}_{nr}\}) \geq \sum_{r=1}^R \sum_{n=1}^N E_{q(h_{nr}(\cdot))} \log(p(\mathbf{y}_{nr} | h_{nr}(\cdot))) - \sum_{r=1}^R \sum_{k=1}^K KL(q(\mathbf{u}_{kr}) || p(\mathbf{u}_{kr})) \tag{22}$$

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.