Supplementary Material

In Appendix 1, we provide proofs of Lemmas and Corollaries from our paper. We describe the derivatives of the log marginal likelihood of the Student-t process which is useful for hyperparameter learning in Appendix 2. In Appendix 3 we offer more insights as to why two seemingly different covariance priors for a Gaussian process prior lead to the same marginal distribution.

1 Proofs

Lemma. *[1] The multivariate Student-t is consistent under marginalization.*

Proof. Assume the generative process of equation 3 of the main text. Σ_{11} is $\mathrm{IW}_{n_1}(\nu, K_{11})$ distributed for any principal submatrix of Σ. Futhermore $y_1|\Sigma_{11} \sim \mathrm{N}_{n_1}(0, (\nu-2)\Sigma_{11})$ since the Gaussian distribution is consistent under marginalization. Hence $y_1 \sim \mathrm{MVT}_{n_1}(\nu, \mu_1, K_{11})$. □

Lemma. *[2] Suppose $f \sim \mathcal{TP}(\nu, \Phi, k)$ and $g \sim \mathcal{GP}(\Phi, k)$. Then f tends to g in distribution as $\nu \to \infty$.*

Proof. It is sufficient to show convergence in density for any finite collection of inputs. Let $\boldsymbol{y} \sim \mathrm{MVT}_n(\nu, \boldsymbol{\phi}, K)$ and set $\beta = (\boldsymbol{y}-\boldsymbol{\phi})^\top K^{-1}(\boldsymbol{y}-\boldsymbol{\phi})$ then

$$p(\boldsymbol{y}) \propto \Big(1 + \frac{\beta}{\nu-2}\Big)^{-(\nu+n)/2} \to e^{-\beta/2}$$

an $\nu \to \infty$. Hence the distribution of $\boldsymbol{y}$ tends to a $\mathrm{N}_n(\boldsymbol{\phi}, K)$ distribution as $\nu \to \infty$. □

Lemma. *[3] Suppose $\boldsymbol{y} \sim \mathrm{MVT}_n(\nu, \boldsymbol{\phi}, K)$ and let $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ represent the first n_1 and remaining n_2 entries of $\boldsymbol{y}$ respectively. Then*

$$\boldsymbol{y_2}|\boldsymbol{y_1} \sim \mathrm{MVT}_{n_2}\Big(\nu + n_1, \tilde{\boldsymbol{\phi}}_{\boldsymbol{2}}, \frac{\nu+\beta_1-2}{\nu+n_1-2} \times \tilde{K}_{22}\Big), \tag{1}$$

where $\tilde{\boldsymbol{\phi}}_{\boldsymbol{2}} = K_{21}K_{11}^{-1}(\boldsymbol{y_1}-\boldsymbol{\phi_1}) - \boldsymbol{\phi_2}$, $\beta_1 = (\boldsymbol{y_1}-\boldsymbol{\phi_1})^\top K_{11}^{-1}(\boldsymbol{y_1}-\boldsymbol{\phi_1})$ and $\tilde{K}_{22} = K_{22} - K_{21}K_{11}^{-1}K_{12}$.

Proof. Let $\beta_2 = (\boldsymbol{y_2}-\tilde{\boldsymbol{\phi}}_{\boldsymbol{2}})^\top \tilde{K}_{22}^{-1}(\boldsymbol{y_2}-\tilde{\boldsymbol{\phi}}_{\boldsymbol{2}})$. Note that $\beta_1+\beta_2 = (\boldsymbol{y}-\boldsymbol{\phi})^\top K^{-1}(\boldsymbol{y}-\boldsymbol{\phi})$. We have

$$\begin{aligned} p(\boldsymbol{y_2}|\boldsymbol{y_1}) = \frac{p(\boldsymbol{y_1},\boldsymbol{y_2})}{p(\boldsymbol{y_1})} &\propto \Big(1+\frac{\beta_1+\beta_2}{\nu-2}\Big)^{-(\nu+n)/2}\Big(1+\frac{\beta_1}{\nu-2}\Big)^{(\nu+n_1)/2} \\ &\propto \Big(1+\frac{\beta_2}{\beta_1+\nu-2}\Big)^{-(\nu+n)/2} \end{aligned}$$

Comparing this expression to the definition of a MVT density function gives the required result. □

Lemma. *[4] Let $K \in \Pi(n)$, $\boldsymbol{\phi} \in \mathbb{R}^n$, $\nu > 2$, $\rho > 0$ and*

$$\begin{aligned} r^{-1} &\sim \Gamma(\nu/2, \rho/2) \\ \boldsymbol{y}|r &\sim \mathrm{N}_n(\boldsymbol{\phi}, r(\nu-2)K/\rho), \end{aligned} \tag{2}$$

then marginally $\boldsymbol{y} \sim \mathrm{MVT}_n(\nu, \boldsymbol{\phi}, K)$.

Proof. Let $\beta = (\boldsymbol{y} - \boldsymbol{\phi})^\top K^{-1}(\boldsymbol{y} - \boldsymbol{\phi})$. We can analytically marginalize out the scalar r,

$$
\begin{aligned}
p(\boldsymbol{y}) = \int p(\boldsymbol{y}|r)p(r)dr &\propto \int \exp\Big(-\frac{\rho\beta}{2(\nu-2)r}\Big) r^{-\frac{n}{2}} \exp\Big(-\frac{\rho}{2r}\Big) r^{-\frac{(\nu+2)}{2}} dr \\
&\propto \Big(1 + \frac{\beta}{\nu-2}\Big)^{-\frac{(\nu+n)}{2}} \int \exp\Big(-\frac{1}{2r}\Big) r^{-\frac{(\nu+n+2)}{2}} dr \\
&\propto \Big(1 + \frac{\beta}{\nu-2}\Big)^{-\frac{(\nu+n)}{2}}
\end{aligned}
$$

Hence $\boldsymbol{y} \sim \mathrm{MVT}_n(\nu, \boldsymbol{\phi}, K)$. Note the redundancy in ρ. Without loss of generality, let $\rho = 1$. □

Corollary. *[7] Suppose $\mathcal{Y} = \{y_i\}$ is an elliptical process. Any finite collection $\boldsymbol{z} = \{z_1, ..., z_n\} \subset \mathcal{Y}$ has an analytically representable density if and only if $\mathcal{Y}$ is either a Gaussian process or a Student-t process.*

Proof. By Theorem 6, we need to be able to analytically solve $\int p(\boldsymbol{z}|r)p(r)dr$, where $\boldsymbol{z}|r \sim \mathrm{N}_n(\boldsymbol{\mu}, r\Omega\Omega^\top)$. This is possible either when r is a constant with probability 1 or when $r \sim \Gamma^{-1}(\nu/2, 1/2)$, the conjugate prior. These lead to the Gaussian and Student-t processes respectively. □

2 Marginal Likelihood Derivatives

Being able to analytically compute the derivative of the likelihood with respect to the hyperparameters is useful for hyperparameter learning e.g. maximum likelihood or Hamiltonian (Hybrid) Monte Carlo.

$$
\log p(\boldsymbol{y}|\nu, K_\theta) = -\frac{n}{2}\log((\nu-2)\pi) - \frac{1}{2}\log(|K_\theta|) + \log\left(\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})}\right) - \frac{(\nu+n)}{2}\log\Big(1 + \frac{\beta}{\nu-2}\Big),
$$

where $\beta = (\boldsymbol{y} - \boldsymbol{\phi})^\top K_\theta^{-1}(\boldsymbol{y} - \boldsymbol{\phi})$ and its derivative with respect to a hyperparameter is

$$
\frac{\partial}{\partial\theta}\log p(\boldsymbol{y}|\nu, \boldsymbol{\phi}, K_\theta) = \frac{1}{2}\mathrm{Tr}\Big(\Big(\frac{\nu+n}{\nu+\beta-2}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top - K_\theta^{-1}\Big)\frac{\partial K_\theta}{\partial\theta}\Big),
$$

where $\boldsymbol{\alpha} = K_\theta^{-1}(\boldsymbol{y} - \boldsymbol{\phi})$. We may also learn ν using gradient based methods and the following derivative

$$
\begin{aligned}
\frac{\partial}{\partial\nu}\log p(\boldsymbol{y}|\nu, K_\theta) = &-\frac{n}{2(\nu-2)} + \psi\Big(\frac{\nu+n}{2}\Big) - \psi\Big(\frac{\nu}{2}\Big) \\
&-\frac{1}{2}\log\Big(1 + \frac{\beta}{\nu-2}\Big) + \frac{(\nu+n)\beta}{2(\nu-2)^2 + 2\beta(\nu-2)}
\end{aligned} \tag{3}
$$

where ψ is the digamma function.

3 More Insight Into the Inverse Wishart Process and Inverse Gamma Priors

As a reminder, we define a Wishart distribution as follows

Definition. A random $\Sigma \in \Pi(n)$ is *Wishart* distributed with parameters $\nu > n-1$, $K \in \Pi(n)$ and we write $\Sigma \sim \mathrm{W}_n(\nu, K)$ if its density is given by

$$
p(\Sigma) = c_n(\nu, K)|\Sigma|^{(\nu-n-1)/2}\exp\Big(-\frac{1}{2}\mathrm{Tr}\big(K^{-1}\Sigma\big)\Big), \tag{4}
$$

where $c_n(\nu, K) = \Big(|K|^{\nu/2}2^{\nu n/2}\Gamma_n(\nu/2)\Big)^{-1}$.

3.1 The Multivariate Gamma Function

The function in the normalizing constant of the Wishart distribution is called the multivariate gamma function and is defined as follows

Definition. The *multivariate gamma function*, $\Gamma_n(.)$, is a generalization of the gamma function defined as

$$\Gamma_n(a) = \int_{S>0} |S|^{a-(n+1)/2} \exp\big(-\operatorname{Tr}(S)\big) dS \tag{5}$$

where $S > 0$ means S is positive definite.

In the following lemma we illustrate an explicit relationship between the multivariate gamma function and the gamma function.

Lemma. *[A]*

$$\Gamma_n(a) = \pi^{n(n-1)/4} \prod_{j=1}^{n} \Gamma\big(a + (1-j)/2\big) \tag{6}$$

Proof.

$$\begin{aligned}
\Gamma_n(a) &= \int_{S>0} |S|^{a-(n+1)/2} \exp\big(-\operatorname{Tr}(S)\big) dS \\
&= \int_{S>0} S_{11}^{a-(n+1)/2} \exp\big(-S_{11}\big) |S_{22.1}|^{a-(n+1)/2} \exp\big(-\operatorname{Tr}(S_{22.1})\big) \\
&\qquad\qquad\qquad \times \exp\big(-\operatorname{Tr}(S_{21}S_{11}^{-1}S_{12})\big) dS_{11} dS_{12} dS_{22.1} \\
&= \int_{S_{11}>0} (\pi S_{11})^{(n-1)/2} S_{11}^{a-(n+1)/2} \exp\big(-S_{11}\big) dS_{11} \\
&\qquad \times \int_{S_{22.1}} |S_{22.1}|^{a-(n+1)/2} \exp\big(-\operatorname{Tr}(S_{22.1})\big) dS_{22.1} \\
&= \pi^{(n-1)/2} \Gamma(a) \Gamma_{n-1}(a - 1/2)
\end{aligned}$$

This recursive relationship and the fact that $\Gamma_1(b) = \Gamma(b)$ implies

$$\begin{aligned}
\Gamma_n(a) &= \prod_{j=1}^{n} \pi^{(j-1)/2} \Gamma(a - (j-1)/2) \\
&= \pi^{n(n-1)/4} \prod_{j=1}^{n} \Gamma(a + (1-j)/2)
\end{aligned}$$

which is as required. □

A simple corollary of this result will be key later.

Corollary. *[B]*

$$\frac{\Gamma_n(a)}{\Gamma_n(a-1/2)} = \frac{\Gamma(a)}{\Gamma(a-n/2)} \tag{7}$$

3.2 Two Different Covariance Priors

The two generative processes we are interested in are

$$\begin{aligned}
r^{-1} &\sim \Gamma(\nu/2, 1/2) & \qquad \Omega &\sim \mathrm{W}_n(\nu + n - 1, K^{-1}) \\
y_1 &\sim \mathrm{N}_n(0, (\nu-2)rK) & \qquad y_2 &\sim \mathrm{N}(0, (\nu-2)\Omega^{-1})
\end{aligned}$$

where $n \in \mathbb{N}$, $\nu > 2$ and K is a $n \times n$ symmetric, positive definite matrix.

The marginal distribution for y_1 is

$$\begin{aligned}
p(y_1) &= \int p(y_1|r)p(r)dr \\
&= \int (2\pi r(\nu-2))^{-n/2}|K|^{-1/2}\exp\Big(-\frac{y_1^\top K^{-1}y_1}{2(\nu-2)r}\Big)r^{-\nu/2-1}\frac{\exp(-1/(2r))}{2^{\nu/2}\Gamma(\nu/2)}dr \\
&= \frac{(2\pi(\nu-2))^{-n/2}|K|^{-1/2}}{2^{\nu/2}\Gamma(\nu/2)}\int r^{-(\nu+n)/2-1}\exp\Big(-\Big(1+\frac{y_1^\top K^{-1}y_1}{(\nu-2)}\Big)/2r\Big)dr \\
&= \frac{(2\pi(\nu-2))^{-n/2}|K|^{-1/2}}{2^{\nu/2}\Gamma(\nu/2)}\bigg(\Big(1+\frac{y_1^\top K^{-1}y_1}{(\nu-2)}\Big)/2\bigg)^{-(\nu+n)/2}\Gamma\big((\nu+n)/2\big) \\
&= (\pi(\nu-2))^{-n/2}|K|^{-1/2}\Big(1+\frac{y_1^\top K^{-1}y_1}{(\nu-2)}\Big)^{-(\nu+n)/2}\frac{\Gamma\big((\nu+n)/2\big)}{\Gamma(\nu/2)}.
\end{aligned} \tag{8}$$

The marginal distribution for y_1 is

$$\begin{aligned}
p(y_2) &= \int p(y_2|\Omega)p(\Omega)d\Omega \\
&= \int \big(2\pi(\nu-2)\big)^{-n/2}|\Omega|^{1/2}\exp\Big(-\frac{y_2^\top\Omega y_2}{2(\nu-2)}\Big) \\
&\qquad\qquad \times c_n(\nu+n-1,K^{-1})|\Omega|^{(\nu-2)/2}\exp\Big(-\frac{1}{2}\mathrm{Tr}\big(K\Omega\big)\Big)d\Omega \\
&= \big(2\pi(\nu-2)\big)^{-n/2}c_n(\nu+n-1,K^{-1}) \\
&\qquad\qquad \times \int |\Omega|^{(\nu-1)/2}\exp\bigg(-\frac{1}{2}\mathrm{Tr}\Big(\Big(K+\frac{y_2y_2^\top}{\nu-2}\Big)\Omega\Big)\bigg)d\Omega \\
&= \big(2\pi(\nu-2)\big)^{-n/2}c_n(\nu+n-1,K^{-1}) \\
&\qquad\qquad \times c_n\bigg(\nu+n,\Big(K+\frac{y_2y_2^\top}{\nu-2}\Big)^{-1}\bigg)^{-1} \\
&= \big(2\pi(\nu-2)\big)^{-n/2}\Big(|K|^{-(\nu+n-1)/2}2^{(\nu+n-1)n/2}\Gamma_n((\nu+n-1)/2)\Big)^{-1} \\
&\qquad \times |K|^{-(\nu+n)/2}\Big(1+\frac{y_2^\top K^{-1}y_2}{\nu-2}\Big)^{-(\nu+n)/2}2^{(\nu+n)n/2}\Gamma_n((\nu+n)/2) \\
&= (\pi(\nu-2))^{-n/2}|K|^{-1/2}\Big(1+\frac{y_2^\top K^{-1}y_2}{\nu-2}\Big)^{-(\nu+n)/2}\frac{\Gamma_n\big((\nu+n)/2\big)}{\Gamma_n\big((\nu+n-1)/2\big)}.
\end{aligned} \tag{9}$$

Both marginal distributions are equivalent given the result in Corollary B.