Supplementary Material for the article: Bayesian Neural Networks for 6G Tracking

Bernardo Camajori Tedeschini, *Student Member, IEEE*, Girim Kwon, *Member, IEEE*, Monica Nicoli, *Senior Member, IEEE* and Moe Z. Win, *Fellow, IEEE*

In the following sections, we provide detailed mathematical derivations and explanations to complement the primary content of the manuscript "Bayesian Neural Networks for 6G Tracking", submitted and under consideration in IEEE Journal on Selected Areas in Communications, special issue on Positioning and Sensing Over Wireless Networks.

Appendix A
Proof of (16)

To prove (16), we start by writing the function $A(\boldsymbol{w}|\boldsymbol{x})$ as:

$$\begin{aligned} A(\boldsymbol{w}|\boldsymbol{x}) &= \mathrm{KL}\big(p(t|\boldsymbol{x},\mathcal{D})\|p(t|\boldsymbol{x},\boldsymbol{w})\big) \\ &= \int_{t'} p(t|\boldsymbol{x},\mathcal{D}) \log \frac{p(t|\boldsymbol{x},\mathcal{D})}{p(t|\boldsymbol{x},\boldsymbol{w})} dt' \\ &\backsimeq -\mathbb{E}_{p(t|\boldsymbol{x},\mathcal{D})}\Big\{\log p(t|\boldsymbol{x},\boldsymbol{w})\Big\} + \mathrm{Const}(\not{\boldsymbol{w}}), \end{aligned} \tag{1}$$

where the last approximation comes from the removal of constant terms which do no depend on $\boldsymbol{w}$. Therefore, we can write:

$$\begin{aligned} A(\boldsymbol{w}|\boldsymbol{x}) &\backsimeq -\int_{t'} \left[\int_{\boldsymbol{\theta}'} p(t|\boldsymbol{x},\boldsymbol{\theta}')p(\boldsymbol{\theta}'|\mathcal{D})d\boldsymbol{\theta}'\right] \log p(t'|\boldsymbol{x},\boldsymbol{w})dt' \\ &= -\int_{\boldsymbol{\theta}'} p(\boldsymbol{\theta}'|\mathcal{D}) \left[\int_{t} p(t|\boldsymbol{x},\boldsymbol{\theta}')\log p(t'|\boldsymbol{x},\boldsymbol{w})dt'\right] d\boldsymbol{\theta}' \\ &= -\int_{\boldsymbol{\theta}'} p(\boldsymbol{\theta}'|\mathcal{D}) \left[\mathbb{E}_{p(t|\boldsymbol{x},\boldsymbol{\theta}')}\Big\{\log p(t|\boldsymbol{x},\boldsymbol{w})\Big\}\right] d\boldsymbol{\theta}' \\ &\backsimeq -\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{p(t|\boldsymbol{x},\boldsymbol{\theta}_\ell)}\Big\{\log p(t|\boldsymbol{x},\boldsymbol{w})\Big\} \\ &= \frac{1}{L}\sum_{\ell=1}^{L} A(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{\theta}_\ell), \end{aligned} \tag{2}$$

where we adopted the Monte Carlo (MC) approximation for approximating the integral through the samples $\boldsymbol{\theta}_\ell$. For regression, $p(t|\boldsymbol{x},\boldsymbol{\theta}_\ell) = \mathcal{N}\left(t; y^{(\mathrm{T})}(\boldsymbol{x},\boldsymbol{\theta}_\ell), \sigma_{\boldsymbol{\varepsilon}^{(\mathrm{T})}}{}^2\right)$, whereas $p(t|\boldsymbol{x},\boldsymbol{w}) = \mathcal{N}(t; y^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w}), \sigma_{\boldsymbol{\varepsilon}^{(\mathrm{S})}}(\boldsymbol{x})^2)$. Thus, we can write $A(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{\theta}_\ell)$ without considering the constant values in $\boldsymbol{w}$ as:

$$\begin{aligned} A(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{\theta}_\ell) &= -\int_{t'} p(t'|\boldsymbol{x},\boldsymbol{\theta}_\ell)\log p(t'|\boldsymbol{x},\boldsymbol{w})dt' \\ &\backsimeq -\int_{t'} p(t'|\boldsymbol{x},\boldsymbol{\theta}_\ell)\times \\ &\left[-\frac{1}{2}\log\left(\sigma_{\boldsymbol{\varepsilon}^{(\mathrm{S})}}(\boldsymbol{x})^2\right) + \frac{\|y^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w}) - t'\|_2^2}{2\sigma_{\boldsymbol{\varepsilon}^{(\mathrm{S})}}(\boldsymbol{x})^2}\right] dt'. \end{aligned} \tag{3}$$

Finally, by splitting the integral for each of the terms inside the l2 norm and by adopting completion of squares, we obtain:

$$\begin{aligned} A(\boldsymbol{w}|\boldsymbol{x},\boldsymbol{\theta}_\ell) &\backsimeq \frac{1}{2}\log\left(\sigma_{\boldsymbol{\varepsilon}^{(\mathrm{S})}}(\boldsymbol{x})^2\right) + \frac{1}{2}\sigma_{\boldsymbol{\varepsilon}^{(\mathrm{S})}}(\boldsymbol{x})^{-2}\Bigg[\sigma_{\epsilon^{(\mathrm{T})}}{}^2 \\ &\quad + \left\|y^{(\mathrm{T})}(\boldsymbol{x},\boldsymbol{\theta}_\ell) - y^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w})\right\|_2^2\Bigg] \\ &\backsimeq \frac{1}{2}\log\left(y_{\mathrm{al}}^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w})\right) + \frac{1}{2}y_{\mathrm{al}}^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w})^{-1}\Bigg[\sigma_{\epsilon^{(\mathrm{T})}}{}^2 \\ &\quad + \left\|y^{(\mathrm{T})}(\boldsymbol{x},\boldsymbol{\theta}_\ell) - y^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w})\right\|_2^2\Bigg], \end{aligned} \tag{4}$$

where the final approximation arises from the model (17).

Appendix B
Proof of (17)

In this Appendix we provide the loss function block $B(\boldsymbol{w}|\boldsymbol{x})$ that permits the student network to learn the epistemic uncertainty of the teacher. By recalling the model in (14), we start by writing the negative log-likelihood according to the maximum likelihood estimation (MLE) approach:

$$\begin{aligned} B(\boldsymbol{w}|\boldsymbol{x}) &= -\log\left\{p\left(\mathbb{V}\left\{\mathsf{t}|\mathbf{x},\mathsf{D},\boldsymbol{\varepsilon}^{(\mathrm{T})}\right\}\Big|\boldsymbol{x},\boldsymbol{w}\right)\right\} \\ &= -\log\left\{\mathcal{N}\left(\mathbb{V}\left\{\mathsf{t}|\mathbf{x},\mathsf{D},\boldsymbol{\varepsilon}^{(\mathrm{T})}\right\};\, y_{\mathrm{ep}}^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w}), \sigma_{\boldsymbol{\xi}_{\mathrm{ep}}^{(\mathrm{S})}}{}^2\right)\right\}. \end{aligned} \tag{5}$$

Then, we remove constant values not dependent on $\boldsymbol{w}$ and we approximate the epistemic uncertainty with the predictive epistemic uncertainty of the teacher as:

$$\begin{aligned} B(\boldsymbol{w}|\boldsymbol{x}) &\backsimeq \frac{1}{2\sigma_{\boldsymbol{\xi}_{\mathrm{ep}}^{(\mathrm{S})}}{}^2}\left\|\mathbb{V}\left\{\mathsf{t}|\mathbf{x},\mathsf{D},\boldsymbol{\varepsilon}^{(\mathrm{T})}\right\} - y_{\mathrm{ep}}^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w})\right\|_2^2 \\ &\quad + \mathrm{Const}(\not{\boldsymbol{w}}) \\ &\backsimeq \frac{1}{2\sigma_{\boldsymbol{\xi}_{\mathrm{ep}}^{(\mathrm{S})}}{}^2}\Bigg\|\frac{1}{L}\sum_{\ell=1}^{L} y^{(\mathrm{T})}(\boldsymbol{x},\boldsymbol{\theta}_\ell)^2 \\ &\quad - \left\{\frac{1}{L}\sum_{\ell=1}^{L} y^{(\mathrm{T})}(\boldsymbol{x},\boldsymbol{\theta}_\ell)\right\}^2 - y_{\mathrm{ep}}^{(\mathrm{S})}(\boldsymbol{x},\boldsymbol{w})\Bigg\|_2^2, \end{aligned} \tag{6}$$

concluding the derivation.