
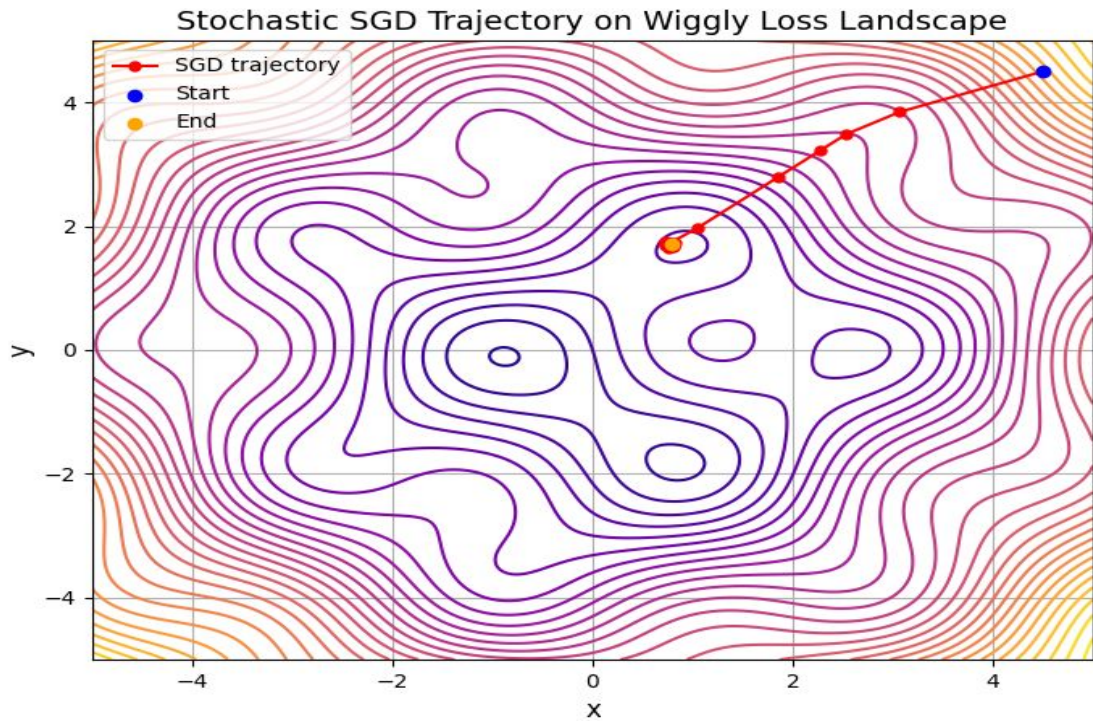


Laplace Approximations for Bayesian Deep Learning

The background is a solid teal color. On the right side, there are several decorative elements: a large, semi-transparent pie chart with a white slice, and several smaller, semi-transparent pie charts of varying sizes. At the bottom right, there is a bar chart with four vertical bars of increasing height, each with a semi-transparent teal top section.

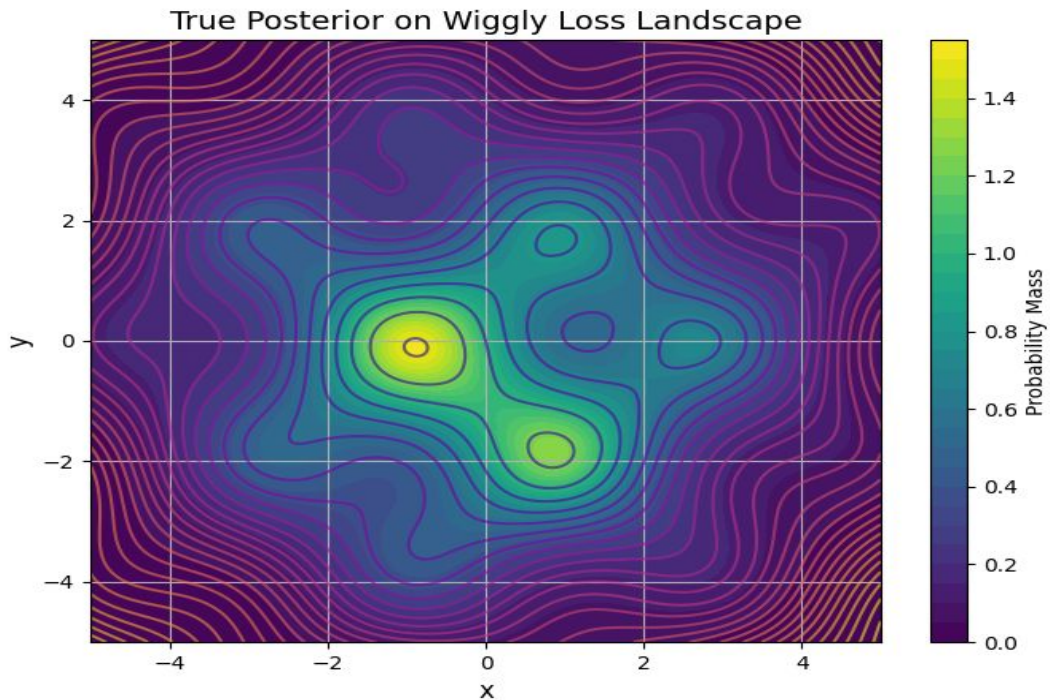


Deep Learning





True Posterior of the Neural Network





Bayesian Deep learning

$$p(y|x, \theta) = \prod_{i=1}^N p(y_i | f_{\theta}(x)) : \textit{Likelihood}$$

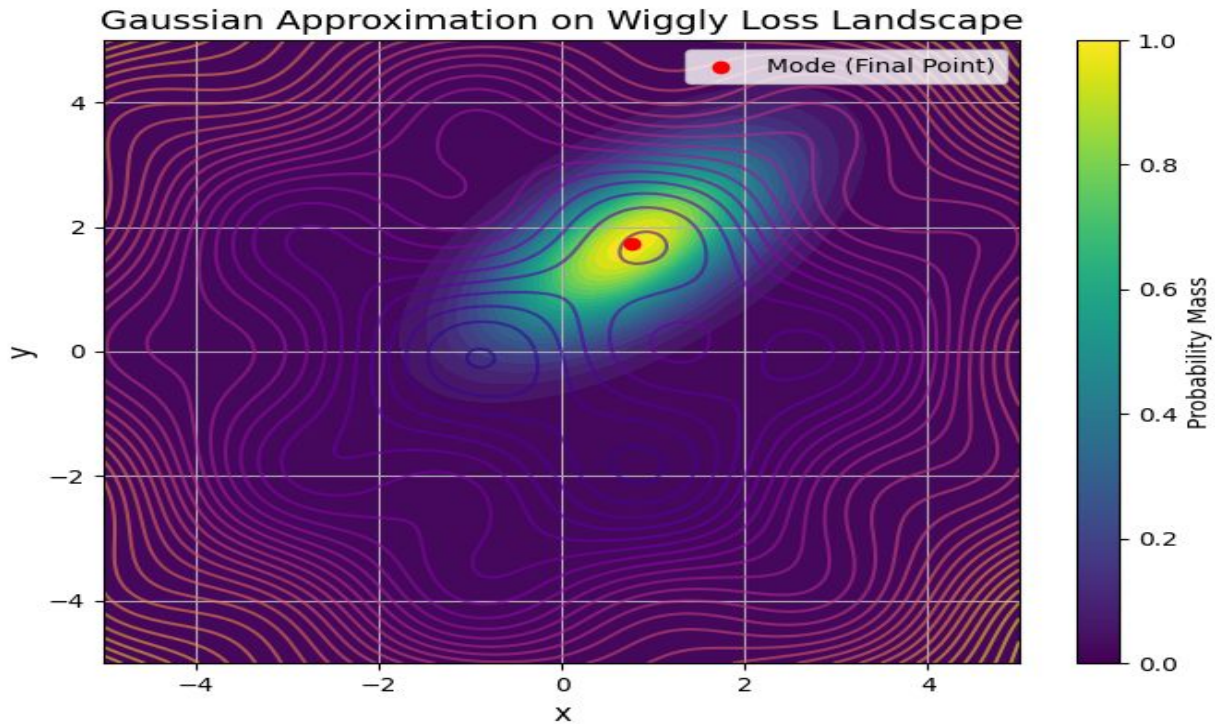
$$p(\theta) : \textit{Prior}$$

Intractable

$$\left\{ \begin{array}{l} p(y|x) = \int_{\theta \in \Theta} p(y|x, \theta) p(\theta) d\theta : \textit{Evidence/Marginal Likelihood} \\ p(\theta|x, y) : \textit{Posterior} \end{array} \right.$$



Local Approximation of the True Posterior





Laplace Approximation

$$\mathcal{L}(\theta|x, y) = -\log p(\mathbf{y}|f_{\theta}(\mathbf{x})) - \log p_{\Lambda}(\theta)$$

$$\mathcal{L}(\theta|x, y) \approx \mathcal{L}(\theta_{\text{MAP}}|x, y) + \frac{1}{2}(\theta - \theta_{\text{MAP}})^T \nabla_{\theta}^2 \mathcal{L} \Big|_{\theta_{\text{MAP}}} (\theta - \theta_{\text{MAP}})$$

$$p(y|x, \theta)p_{\Lambda}(\theta) \approx \exp(-\mathcal{L}(\theta_{\text{MAP}}|x, y)) \exp\left(-\frac{1}{2}(\theta - \theta_{\text{MAP}})^T \nabla_{\theta}^2 \mathcal{L} \Big|_{\theta_{\text{MAP}}} (\theta - \theta_{\text{MAP}})\right)$$



Laplace Approximation

- With this approximation the evidence becomes a Gaussian Integral which is tractable

$$\begin{aligned}\int_{\theta \in \Theta} p(y|x, \theta)p(\theta)d\theta &\approx \int_{\theta \in \Theta} \exp(-\mathcal{L}(\theta_{MAP}|x, y)) \exp\left(-\frac{1}{2}(\theta - \theta_{MAP})^T \nabla_{\theta}^2 \mathcal{L} \Big|_{\theta_{MAP}} (\theta - \theta_{MAP})\right) d\theta \\ &= \exp(-\mathcal{L}(\theta_{MAP}|x, y)) \int_{\theta \in \Theta} \exp\left(-\frac{1}{2}(\theta - \theta_{MAP})^T \nabla_{\theta}^2 \mathcal{L} \Big|_{\theta_{MAP}} (\theta - \theta_{MAP})\right) d\theta\end{aligned}$$



Laplace Approximation

- Under the laplace approximation the True Posterior is approximated by the following Gaussian:

$$q_{\text{LLA}}(\boldsymbol{\theta} | \mathcal{D}) = \mathcal{N}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{\text{MAP}}, (\alpha \mathbb{I} + \text{GGN}_{\boldsymbol{\theta}_{\text{MAP}}})^{-1}\right)$$

$$\text{GGN}_{\boldsymbol{\theta}_{\text{MAP}}} = \mathbf{J}_{\boldsymbol{\theta}_{\text{MAP}}}^{\top} \mathbf{H}_{\boldsymbol{\theta}_{\text{MAP}}} \mathbf{J}_{\boldsymbol{\theta}_{\text{MAP}}} \in \mathbb{R}^{P \times P}.$$



Why would we want to do this?

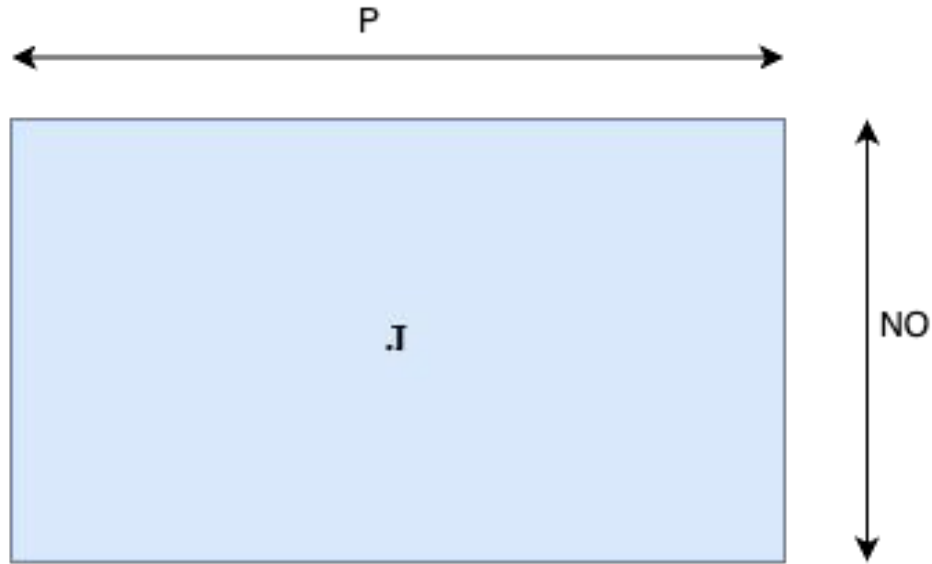
- Bayesian Model Averaging: Marginalize over the posterior for better calibrated predictions
- Principled Uncertainty Quantification
- Marginal Likelihood Optimization: Model Selection without Cross-Validation

Linear Algebra

A decorative pattern at the bottom of the slide consisting of a series of vertical bars of varying heights and widths, all in shades of teal and light blue, creating a rhythmic, bar-like appearance.



Size of the Jacobian Matrix



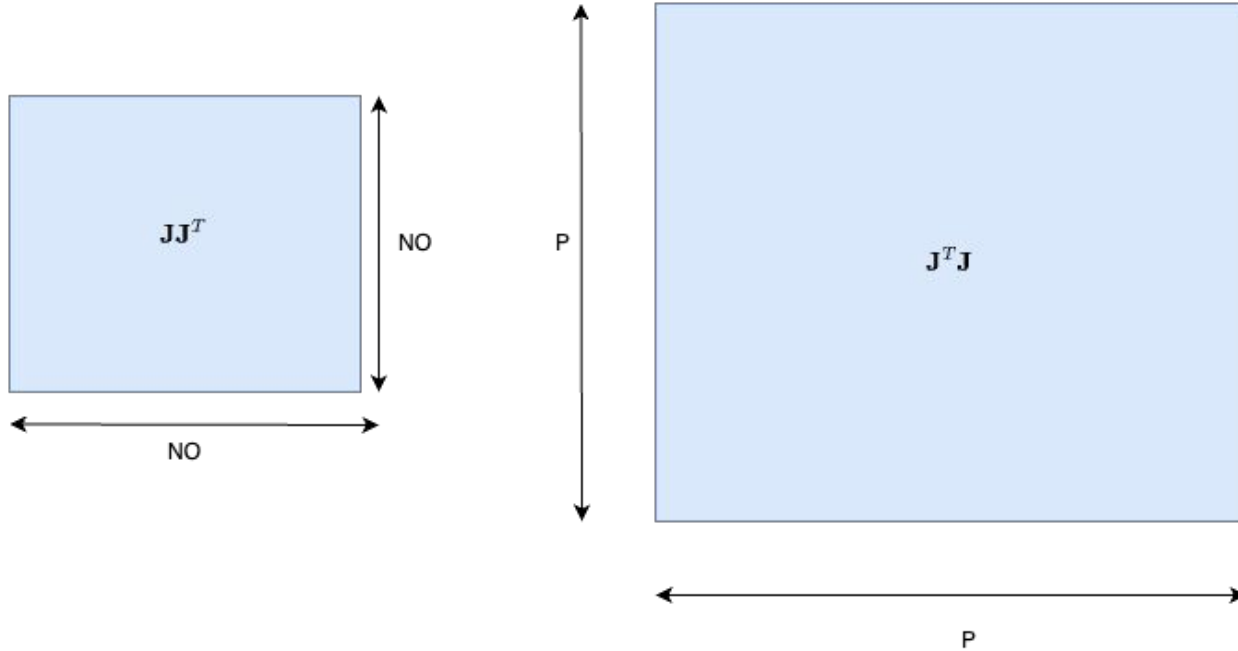


Sampling is hard!





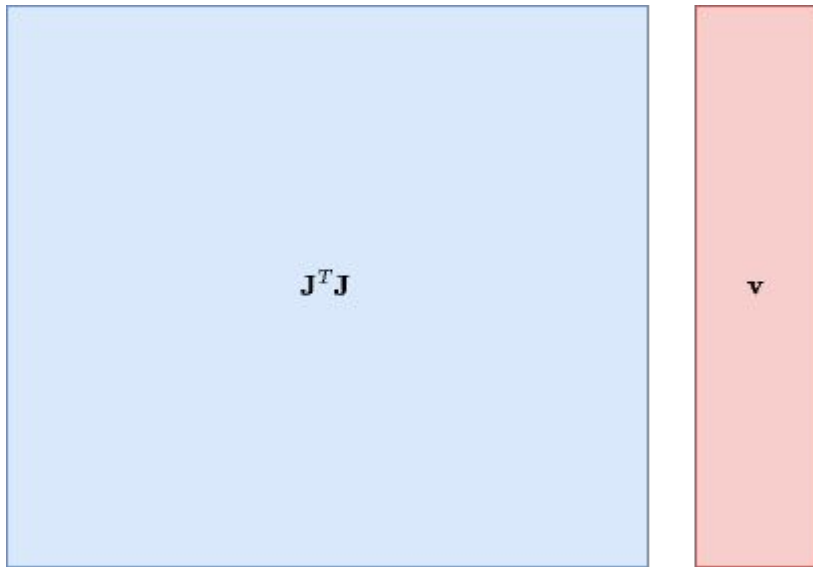
Big Matrices



- Too big to instantiate
- Direct Solvers are impractical



But Matrix-Vector Products are cheap



- Matrix Vector Products only require Jacobian Vector Products and Vector Jacobian Products(Autodiff)
- We can use Iterative Solvers!

Thank You!

