We perform near-**exact inference** and **hyperparameter optimisation** in Bayesian linear models with **millions** of parameters and observations

Take a picture to see the full paper.

Sampling-based inference for large linear models, with application to linearised Laplace

Javier Antorán*, Shreyas Padhy,* Riccardo Barbano, Eric Nalisnick, David Janz, and José Miguel Hernández-Lobato

We consider Bayesian Linear Models of the form $y_i = \phi(x_i)\theta + \eta_i$, where $\theta \sim \mathcal{N}(0, A^{-1})$, $\eta_i \sim \mathcal{N}(0, B_i^{-1})$

$$y_i \in \mathbb{R}^m, \quad \phi(x_i) \in \mathbb{R}^{m\times d}, \quad \theta \in \mathbb{R}^d \quad i \in \{1,\dots,n\}$$

1. Posterior is $\mathcal{N}(\bar{\theta}, \Sigma)$, where $\Sigma^{-1} = \Phi^T B \Phi + A$ ⟵ Inversion of $d \times d$ matrix, $\mathcal{O}(d^3)$!
2. Model Evidence can be tuned for hparams, contains $\operatorname{logdet}(\Sigma^{-1})$ ⟵ Log.det. of $d \times d$ matrix, $\mathcal{O}(d^3)$!

Idea 1: Sample from Posterior with Stochastic Optimisation

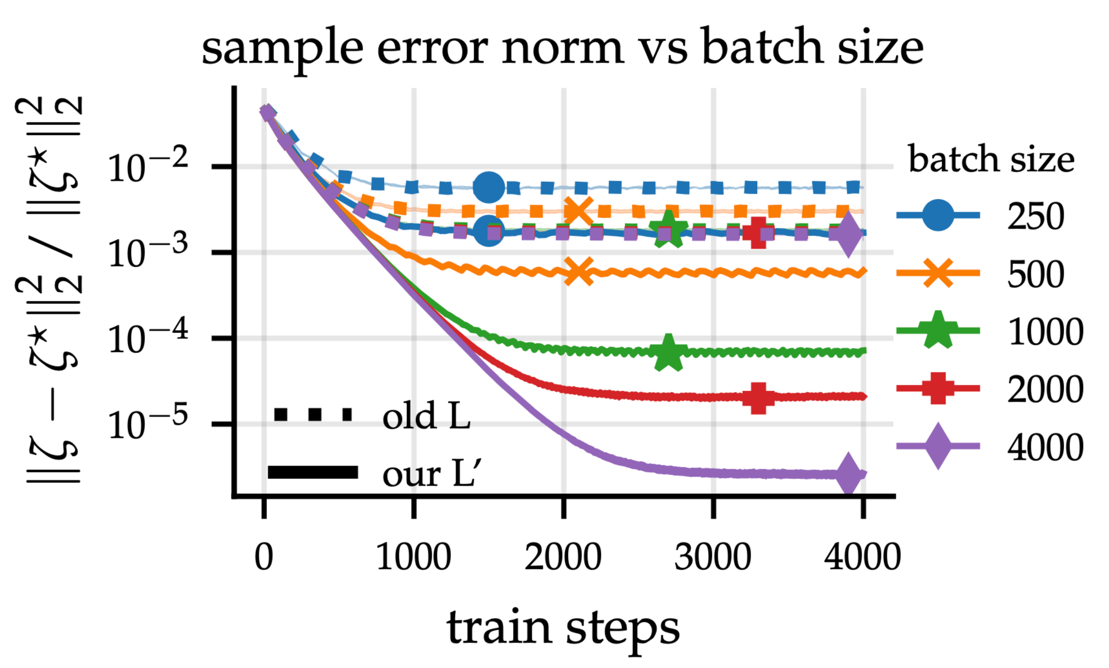
$$z^* \sim \mathcal{N}(0, H^{-1}) \text{ if } z^* = \operatorname{argmin}_z L(z)$$

Where $L(z) = \underbrace{\sum_{i=1}^{n} \|\epsilon_i - \phi(x_i)z\|^2_{B_i}}_{1} + \underbrace{\|z - \theta^0\|^2_A}_{2}$

$$\epsilon_i \sim \mathcal{N}(0, B_i^{-1}) \qquad \theta^0 \sim \mathcal{N}(0, A^{-1})$$

$$L'(z) = \underbrace{\sum_{i=1}^{n} \|\phi(x_i)z\|^2_{B_i}}_{1} + \underbrace{\|z - \theta^n\|_A}_{2} \qquad \theta^n = \theta^0 + A^{-1}\Phi^T B \mathcal{E}$$

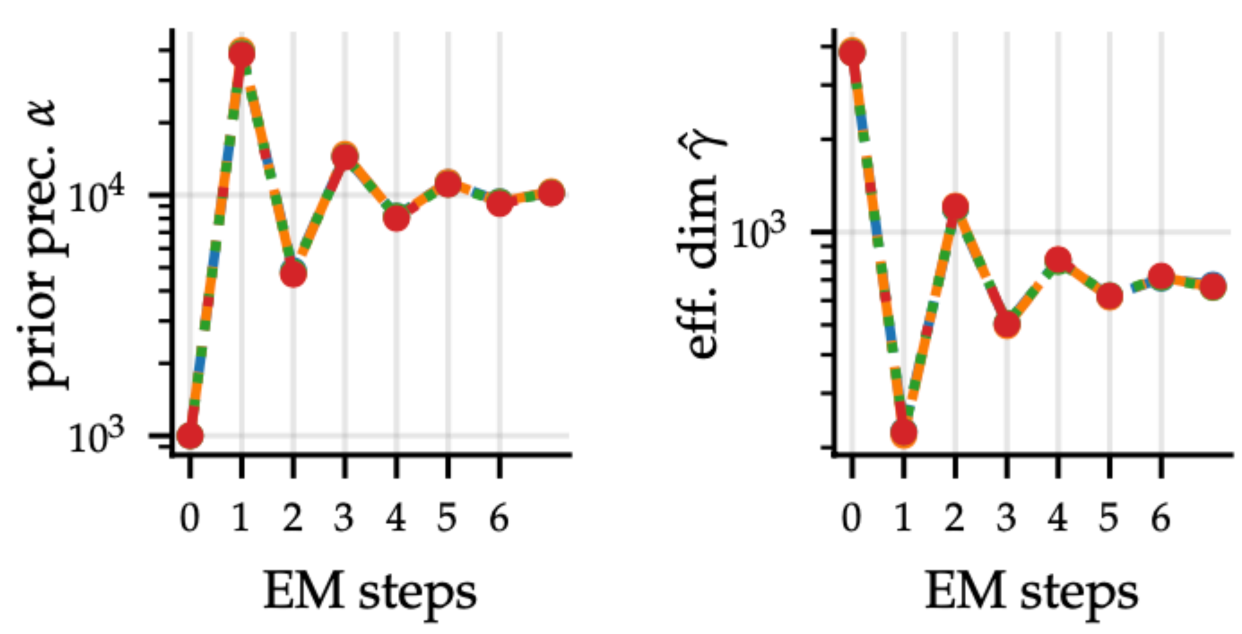
$$\mathcal{E} = [\epsilon_0^T, \dots, \epsilon_n^T]^T \in \mathbb{R}^{nm}$$



Idea 2: Optimise the model evidence using only samples

MacKay proposed update for prior precision $A = \alpha I$, $\alpha = \dfrac{\operatorname{Tr}(\Sigma\ \Phi^T B \Phi)}{\|\bar{\theta}\|^2}$

$$\operatorname{Tr}\{\Sigma M\} = \operatorname{Tr}\left\{\Sigma^{\frac{1}{2}} M \Sigma^{\frac{1}{2}}\right\} = \mathbb{E}\left[z_1^T M z_1\right] \approx \frac{1}{k}\sum_{j=1}^{k} z_j^T \Phi^T\ B \Phi z_j$$



Application: Uncertainty Quantification with very large NNs

Linearised NNs + Laplace Approximation = Bayesian Linear Models

Largest scale exact Laplace inference with ResNet-18 + CIFAR100

$$d = 11M \text{ and } nm = 5M$$