SLIM ML4Seismic

Relation

conditional neural density & EIG

Maximizing the expected *posterior density* is *equivalent* to maximizing the expected *information gain*

$$\max_{\mathbf{W}} \; EIG(\mathbf{W}) = \mathbb{E}_{p(\mathbf{y}|\mathbf{W})} \left[D_{KL}(p_\theta(\mathbf{x}\,|\,\mathbf{y})\,||\,p(\mathbf{x})) \right] = \mathbb{E}_{p(\mathbf{y}|\mathbf{W})} \left[\mathbb{E}_{p(\mathbf{x}|\mathbf{y})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) - \log p(\mathbf{x}) \right] \right]$$

$$= \mathbb{E}_{p(\mathbf{y}|\mathbf{W})} \left[\mathbb{E}_{p(\mathbf{x}|\mathbf{y})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) \right] \right]$$ law of total expectation

$$= \mathbb{E}_{p(\mathbf{x},\mathbf{y}|\mathbf{W})} \left[\log p_\theta(\mathbf{x}\,|\,\mathbf{y}) \right]$$ same as neural posterior objective!

Thus optimizing under the posterior density objective will increase the EIG!

Foster, Adam, et al. "A unified stochastic gradient approach to designing bayesian-optimal experiments." PMLR, 2020.
Hoffmann, Till, and Jukka-Pekka Onnela. "Minimizing the Expected Posterior Entropy Yields Optimal Summary Statistics." arXiv preprint arXiv:2206.02340 (2022).

Proposed method

As usual, prepare posterior learning algorithm: $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{i=1}^{N}$

Instead of optimizing only network parameters:

$$\hat{\theta} \ = \underset{\theta}{\arg\max} \ \frac{1}{N}\sum_{n=1}^{N}\left(-\|f_{\theta}(\mathbf{x}^{(n)};\bar{\mathbf{y}}^{(n)})\|_2^2 + \log\left|\det \mathbf{J}_{f_\theta}\right|\right).$$

Jointly optimize experimental design in terms of **density**, $\mathbf{w}$, defining the mask $\mathbf{M}(\mathbf{w})$

$$\hat{\theta},\ \hat{\mathbf{w}} = \underset{\theta,\mathbf{w}}{\arg\max} \ \frac{1}{N}\sum_{i=1}^{N}\left(-\|f_{\theta}(\mathbf{x}^{(n)};(\mathbf{W}[\mathbf{w}]\odot\mathbf{x}^{(n)},\bar{\mathbf{y}}^{(n)}))\|_2^2 + \log\left|\det \mathbf{J}_{f_\theta}\right|\right).$$