0.1 Bayesian methods

0.1.1 Static learning

Markov chain Monte Carlo

Variational inference

0.1.2 Sequential learning

Sequential Monte Carlo

Kálmán filtering

0.2 Priors

0.2.1 Weight-space priors

0.2.2 Architecture priors

Abstracting to a higher level, we must acknowledge that the choice of architecture reflects a prior on the hyperparameters of the model. If we select an architecture—the number of layers, the number of nodes per each, the activation functions—and lock ourselves into this selection, we have imposed a rather strong prior on our architecture and the form of our function.

0.3 Global behavior of static learning

0.3.1 Covariance over weights

0.3.2 Modality of the posterior distribution

We use Markov chain Monte Carlo to sample from the posterior distribution to investigate the modality of the posterior distribution.

0.4 Dynamic behavior of sequential learning

0.4.1 Convergence results for fixed architecture

We vary the depth of our emulator network and observe the behavior of the posterior distribution of the weights in the sequential learning framework.

0.5 Data assimilation

0.5.1 Convergence results for fixed architecture

We now consider the case that our emulator network acts as the dynamical model for a filter and has access to observations of a subset of the state variables. We again vary the depth of our emulator network and observe the influence on the skill of our predictions. We consider both the strict state estimation problem and the dual estimation problem. In the former, the uncertainty of the emulator model is incorporated in the likelihood:

$$p(x_{k+1}|y_{1:k+1}) \propto \mathcal{L}(y_{k+1}|x_{k+1})\mathcal{L}(x_{k+1}|x_k, w)p(x_k)$$

In the latter, the uncertainty of the emulator model is incorporated in the prior and is updated by the observations:

$$p(x_{k+1}|y_{1:k+1}) \propto \mathcal{L}(y_{k+1}|x_{k+1})\mathcal{L}(x_{k+1}|x_k, w_{k+1})\mathcal{L}(w_{k+1}|w_k)p(x_k, w_k)$$

We find that the dual estimation formulation is robust to out-of-distribution