Double Descent LLA KFAC $p = 100.0$

Classical Regime

Variance-Bias Tradeoff

Modern Regime

Larger is Better

← Critical Region

← Interpolation Threshhold

Train/Test loss

Parameters $\times 10^6$
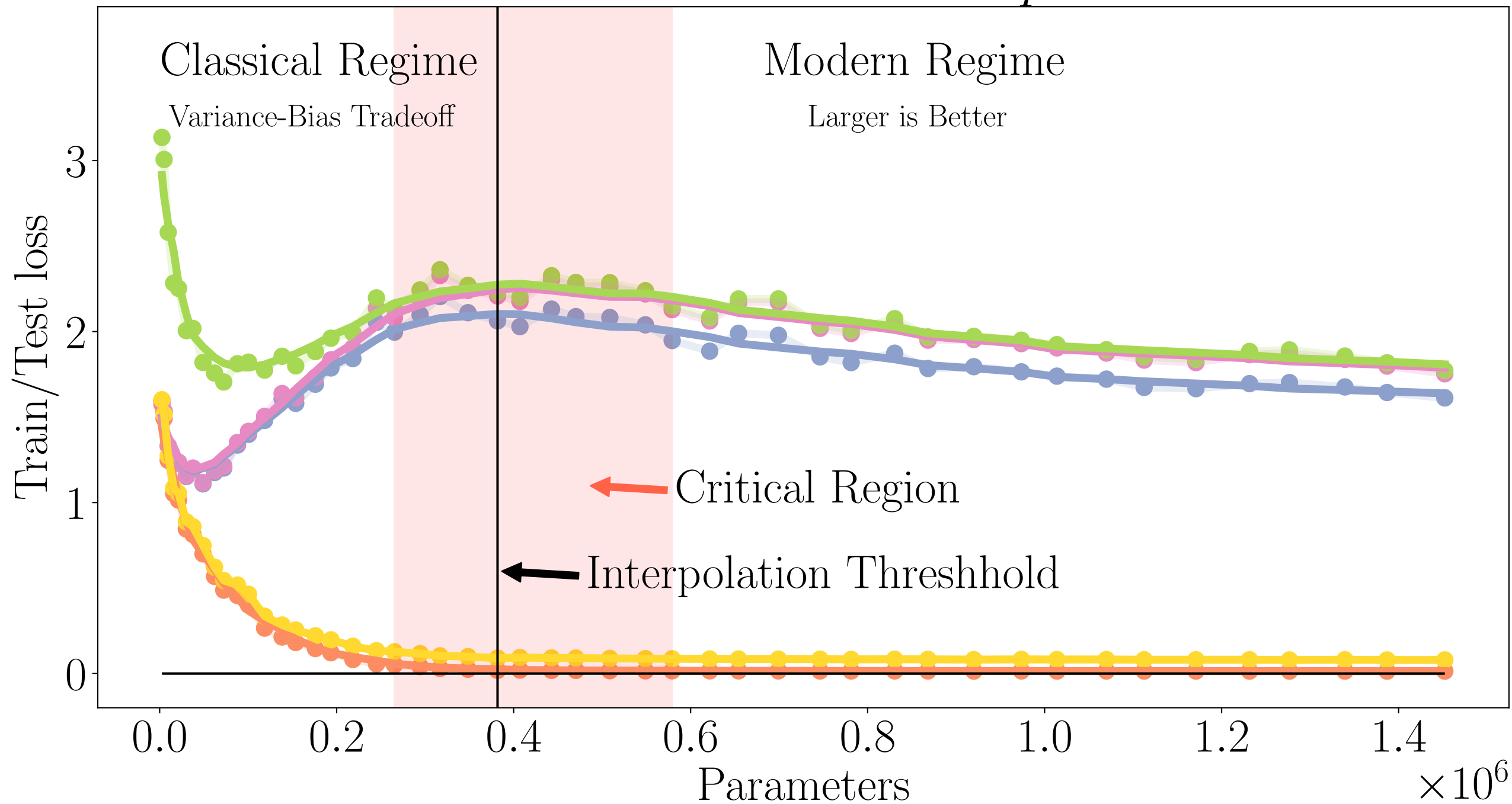
Train loss — Gibbs loss — Neg Log Marginal Laplace — Bayes loss — Bound — Neg Log Marginal Upper Bound