Meta Variational Monte Carlo

Tianchen Zhao[1], James Stokes[2], Oliver Knitter[1], Brian Chen[1], Shravan Veerapaneni[1,2]

[1]Department of Mathematics, University of Michigan
[2]Flatiron Institute, Simons Foundation

Introduction

- Given a local Hamiltonian H, the Variational Monte Carlo (VMC) algorithm returns an estimate of the minimal eigenvalue $\lambda_{\min}(H)$ and an associated eigenvector.
- State-of-the-art results [1] are achieved through use of neural networks (e.g. Boltzmann machines) as trial wavefunctions, parameterized by θ.
- However, accuracy comes at the expense of significant computational time, due to steepest descent calculations of Rayleigh quotient estimator:

$$L(\theta) := \frac{\langle\psi_\theta|H|\psi_\theta\rangle}{\langle\psi_\theta|\psi_\theta\rangle} = \mathop{\mathbb{E}}_{x\sim|\psi_\theta(\cdot)|^2}\left[\frac{\langle x|H|\psi_\theta\rangle}{\psi_\theta(x)}\right] \geq \lambda_{\min}(H),$$

- We wish to reduce the time needed for VMC to optimize new Hamiltonia by leveraging information obtained from prior Hamiltonia.

Meta-VMC

- We propose Meta-VMC that learns an initialization for θ from which VMC may quickly converge to the ground energy of any one of an entire ensemble $\mathcal{T}$ of Hamiltonia.
- Meta-VMC optimizes the ensemble task loss

$$L_{ML}(\theta) = \mathbb{E}_{\tau\sim\mathcal{T}}\left[L_\tau\left(U_\tau^t(\theta)\right)\right],$$

where L_τ is the task loss on an individual task τ in the ensemble, and $U_\tau^t(\theta) : \mathbb{R}^d \to \mathbb{R}^d$ is the t-fold application of a task adaptation operator U_τ, such as gradient descent.
- This learning-to-learn formulation [2] maintains a distinct advantage over a standard multi-task learning approach (which depends only on the mean of τ), since it leverages information from higher-order moments of τ.

Training

- Meta-training is done using a variant of model-agnostic meta-learning (MAML) [3], which makes gradient updates using the estimator

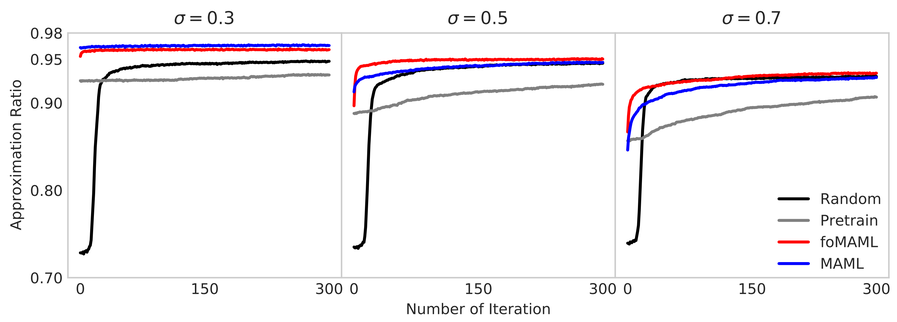
$$\nabla L_{ML}(\theta) = \mathbb{E}_{\tau\sim\mathcal{T}}\left[(U_\tau^t)'(\theta)\nabla L_\tau\left(U_\tau^t(\theta)\right)\right],$$

where $(U_\tau^t)'(\theta)$ is the Jacobian matrix of $U_\tau^t(\theta)$.
- In a practical algorithm, the intermediate variables θ_τ and ∇_τ are estimated stochastically using independent batches of data generated by the same task τ.
- The computation of the Jacobian involves an expensive back-propagation. First-order MAML (foMAML) is a simplification of MAML in which the Jacobian matrix is approximated by the identity matrix.

Algorithm

Input: Hamiltonia ensemble $\mathcal{T}$, adaptation operator U_τ, adaptation steps t.
Initialize θ
While not done **do**
  Sample batch of disorder parameters $B \overset{\text{iid}}{\sim} \mathcal{T}$
  for each disorder parameter $\tau \in B$ **do**
    $\theta_\tau = U_\tau^t(\theta)$
    $\nabla_\tau = (U_\tau^t)'(\theta)\nabla L_\tau(\theta_\tau)$
  end for
  $\nabla = \frac{1}{|B|}\Sigma_{\tau\in B}\nabla_\tau$
  $\theta \leftarrow \textsc{Optimizer}(\theta, \nabla)$
End While

Experiments



- Our experiments focus on the Max-Cut instances with 50 nodes. Ensemble $\mathcal{T}$ for meta-learning was chosen by fixing an adjacency matrix and applying rounded Gaussian $\mathcal{N}(0, \sigma^2)$ to its entries.
- For each task distribution, 32 graph instances are sampled for testing. We train RBMs with various initialization methods on all testing graphs for 300 VMC iterations and measure with the averaged approximation ratio (cut number divided by optimal value of the SDP relaxation).
- Models initialized from MAML and foMAML can discover near-optimal solutions within very few iterations, and outperform the baselines (e.g. Random and Pretrain) in the long run.

References

[1] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325): 602–606, 2017.

[2] Sebastian Thrun and Lorien Pratt. *Learning to learn*. Springer Science & Business Media, 2012.

[3] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.