Workshop on High-Dimensional Learning and Computation in Physics

Date: Thursday, 20 June 2019

Time: 9:00 – 17:00

Venue: S17-04-06

Organizers:

Yang, Haizhao (NUS)

SPEAKERS

XU, Zhiqin (New York University)
DAI, Xiaowu (UC Berkeley)
FENG, Jiashi (NUS)
LI, Lei (Shanghai Jiao Tong University)
TONG, Xin (NUS)
LIU, Qiang (NUS)
GAO, Tingran (University of Chicago)
CAI, Zhenning (NUS)
LI, Yingzhou (Duke University)
CAI, Yongqiang (NUS)
TAN, Sandra (NUS)
LIANG, Senwei (NUS)

NUS
National University of Singapore

Department of Mathematics
Faculty of Science

Programme

09.00am – 09.30am	**The F-Principle in Deep learning** *XU, Zhiqin* *New York University*
09.30am – 10.00am	**Statistical Calibration for Learning Physics under Limited Data** *DAI, Xiaowu* *University of California, Berkeley*
10.00am – 10.30am	**Understanding Generalization and Optimization Performance of Deep CNNs** *FENG, Jiashi* *National University of Singapore*
10.30am – 10.50am	Tea Break @ Math Lounge
10.50am – 11.20am	**On validity of diffusion approximations for Stochastic Gradient Descent** *LI, Lei* *Shanghai Jiao Tong University*
11.20am – 11.50am	**Sampling High-Dimensional Distributions with Gibbs Sampler** *TONG, Xin* *National University of Singapore*
11.50am – 12.20pm	**Accelerating Metropolis-within-Gibbs sampler with localized computations of differential equations** *LIU, Qiang* *National University of Singapore*
12.20pm – 14.00pm	*Lunch*

Programme

14.00pm – 14.30pm	**Multi-Frequency Phase Synchronization and Class Averaging for Three-Dimensional Cryo-Electron Microscopy** *GAO, Tingran* *University of Chicago*
14.30pm – 15.00pm	**Gauge cooling technique for complex Langevin** *CAI, Zhenning* *National University of Singapore*
15.00pm – 15.30pm	**Approximate Tensor Ring Decomposition** *LI, Yingzhou* *Duke University*
15.30pm – 16.00pm	Tea Break @ Math Lounge
16.00pm – 16.30pm	**A Quantitative Analysis of the Effect of Batch Normalization on Gradient Descent** *CAI, Yongqiang* *National University of Singapore*
16.30pm – 17.00pm	**Analysis of Optimization Algorithms using Sum-of-Squares** *TAN, Sandra* *National University of Singapore*
17.00pm – 17.30pm	**Regularization Methods of Deep Learning for Image Classification** *LIANG, Senwei* *National University of Singapore*

Abstract

The F-Principle in Deep Learning

Xu, Zhiqin, New York University

Abstract: We show the universality of an implicit bias --- the F-Principle that deep neural nets (DNNs) often fit target functions from low to high frequencies --- in both theory and simulation. Second, we propose a Linear F-Principle (LFP) model that accurately predicts the output of DNNs with an extremely large width. Based on the LFP model, we use an optimization framework to explicitize the implicit bias of the F-Principle as an FP norm penalty, in which higher frequencies of feasible solutions are more heavily penalized, underlying the training dynamics of two-layer DNNs. We then provide an a prior estimate of the generalization error, which is bounded by the FP-norm of the target function and is independent of the number of parameters. Overall, our work makes a step towards a quantitative understanding of the learning and generalization of DNNs.

Statistical Calibration for Learning Physics under Limited Data

Dai, Xiaowu, University of California, Berkeley

Abstract: We provide another look at the statistical calibration problem in computer models. This viewpoint is inspired by two overarching practical considerations of computer models: (i) many computer models are inadequate for perfectly modelling physical systems, even with the best-tuned calibration parameters; (ii) only a finite number of data points are available from a physical experiment to calibrate a related computer model. Following this line of thinking, we provide a non-asymptotic theory and derive a prediction-oriented calibration method. Our calibration method minimizes the predictive mean squared error for a finite sample size with statistical guarantees. We introduce an algorithm to perform the proposed calibration method and connect it to existing Bayesian calibration methods. Synthetic and real examples are provided to corroborate the derived theory and illustrate some advantages of the proposed calibration method.

Understanding Generalization and Optimization Performance of Deep CNNs

Feng, Jiashi, National University of Singapore

Abstract: TBA

Random Batch Method and its applications

LI, Lei, Shanghai Jiao Tong University

Abstract: A random algorithm for simulating interacting particle systems that reduces the complexity per time step from \$O(N^2)\$ to \$O(N)\$, called Random Batch Method (RBM), will be introduced in this talk. The algorithm is motivated by the mini-batch idea in machine learning and statistics. Under some special conditions, we show the convergence of RBM for the first marginal distribution under Wasserstein distance. Compared with traditional tree code and fast multipole expansion algorithms, RBM works for kernels that do not necessarily decay. This is a joint work with Prof. Shi Jin and Prof. Jian-Guo Liu.

Sampling High-Dimensional Distributions with Gibbs Sampler

Tong, Xin, National University of Singapore

Abstract: We investigate how ideas from covariance localization in numerical weather prediction can be used in Markov chain Monte Carlo (MCMC) sampling of high-dimensional posterior distributions arising in Bayesian inverse problems. To localize an inverse problem is to enforce an anticipated "local" structure by (i) neglecting small off-diagonal elements of the prior precision and covariance matrices; and (ii) restricting the influence of observations to their neighbourhood. For linear problems we can specify the conditions under which posterior moments of the localized problem are close to those of the original problem. We explain physical interpretations of our assumptions about local structure and discuss the notion of high dimensionality in local problems, which is different from the usual notion of high dimensionality in function space MCMC. The Gibbs sampler is a natural choice of MCMC algorithm for localized inverse problems and we demonstrate that its convergence rate is independent of dimension for localized linear problems. Nonlinear problems can also be tackled efficiently by localization and, as a simple illustration of these ideas, we present a localized Metropolis-within-Gibbs sampler. Several linear and nonlinear numerical examples illustrate localization in the context of MCMC samplers for inverse problems.

Accelerating Metropolis-within-Gibbs sampler with localized computations of differential equations

LIU, Qiang, National University of Singapore

Abstract: Bayesian inverse problem is widely encountered when quantifying uncertainty for underlying parameters in practice. For high dimensional spatial models, classical Markov chain Monte Carlo (MCMC) methods are usually slow

to be applied, while it has been shown that Metropolis-within-Gibbs (MwG) sampling works when the parameters are locally dependent. The problem is that its implementation requires $O(n^2)$ calculation, where n is the number of parameters. Our target in this paper is to reduce the computation cost to an optimal scalability of $O(n)$, in the framework of stochastic differential equation (SDE) with local dependence structure. The key is that MwG proposal is only different from the original at local entries, and the difference caused also evolves locally. This inspires us to approximate the solution for the proposal with a surrogate updated only within a local domain, which brings down the computation to our targeting level. Both theoretically and numerically, we prove that the induced errors can be controlled by the local domain size. Implementions of our computation scheme by using Euler-Maruyama and 4th order Runge-Kutta method are also discussed. We demonstrate the finite sample performance of our method in numerical examples of Lorenz 96 and a linear stochastic flow model.

Multi-Frequency Phase Synchronization and Class Averaging for Three-Dimensional Cryo-Electron Microscopy

GAO, Tingran, University of Chicago

Abstract: We propose a novel formulation for phase synchronization—the statistical problem of jointly estimating alignment angles from noisy pairwise comparisons—as a nonconvex optimization problem that enforces consistency among the pairwise comparisons in multiple frequency channels. Inspired by harmonic retrieval in signal processing, we develop a simple yet efficient two-stage algorithm that leverages the multi-frequency information. We demonstrate in theory and practice that the proposed algorithm significantly outperforms state-of-the-art phase synchronization algorithms, at a mild computational costs incurred by using the extra frequency channels. We also extend our algorithmic framework to general synchronization problems over compact Lie groups.

Gauge cooling technique for complex Langevin

Cai, Zhenning, National University of Singapore

Abstract: We study the mechanism of the gauge cooling technique to stabilize the complex Langevin method in the one-dimensional periodic setting. In this case, we find the exact solutions for the gauge transform which minimizes the Frobenius norm of link variables. Thereby, we derive the underlying stochastic differential equations by continuing the numerical method with gauge cooling, and thus provide a number of insights on the effects of gauge cooling. A specific case study is carried out for the Polyakov loop model in SU(2) theory, in which we show that the gauge cooling may help form a localized distribution to guarantee there is no excursion too far away from the real axis.

Approximate Tensor Ring Decomposition

LI, Yingzhou, Duke University

Abstract: Tensor ring decomposition has been used to compress high dimensional tensors in both deep learning and computational physics. In this work, we study the tensor ring decomposition and its associated numerical algorithms. We establish a sharp transition of algorithmic difficulty of the optimization problem as the bond dimension increases: On one hand, we show the existence of spurious local minima for the optimization energy landscape even when the tensor ring format is much over-parameterized, i.e., with bond dimension much larger than that of the true target tensor. On the other hand, when the bond dimension is further increased, we establish one-loop convergence for alternating least square algorithm for tensor ring decomposition. The theoretical results are complemented by numerical experiments for both local minimum and one-loop convergence for the alternating least square algorithm.

A Quantitative Analysis of the Effect of Batch Normalization on Gradient Descent

GAI, Yongqiang, National University of Singapore

Abstract: Despite its empirical success and recent theoretical progress, there generally lacks a quantitative analysis of the effect of batch normalization (BN) on the convergence and stability of gradient descent. In this paper, we provide such an analysis on the simple problem of ordinary least squares (OLS). Since precise dynamical properties of gradient descent (GD) is completely known for the OLS problem, it allows us to isolate and compare the additional effects of BN. More precisely, we show that unlike GD, gradient descent with BN (BNGD) converges for arbitrary learning rates for the weights, and the convergence remains linear under mild conditions. Moreover, we quantify two different sources of acceleration of BNGD over GD -- one due to over-parameterization which improves the effective condition number and another due having a large range of learning rates giving rise to fast descent. These phenomena set BNGD apart from GD and could account for much of its robustness properties. These findings are confirmed quantitatively by numerical experiments, which further show that many of the uncovered properties of BNGD in OLS are also observed qualitatively in more complex supervised learning problems.

Analysis of Optimization Algorithms using Sum-of-Squares

TAN, Sandra, National University of Singapore

Abstract: In this work, we introduce a new framework for unifying and systematizing the performance analysis of first-order black-box optimization algorithms for unconstrained convex minimization over finite-dimensional Euclidean spaces. The low-cost iteration complexity enjoyed by this class of algorithms renders them particularly relevant for applications in machine learning and large-scale data analysis. However, existing proofs of convergence of such optimization algorithms consist mostly of ad-hoc arguments and case-by-case analyses. On the other hand, our approach is based on sum-of-squares optimization and puts forward a promising framework for unifying the convergence analyses of optimization algorithms. Illustrating the usefulness of our approach, we recover several known convergence bounds for four widely-used first-order algorithms in a unified manner, and also derive one new convergence result for gradient descent with Armijo-terminated line search.

Regularization Methods of Deep Learning for Image Classification

LIANG, Senwei, National University of Singapore

Abstract: Applying regularization methods and adjusting the structure of models are two possible ways to improve generalization in the image classification. The gradually increasing depth and width can improve the nonlinear capability of a network but may lead to a risk of overfitting. We introduce a regularization method called Drop-Activation which drops nonlinear activation functions randomly during training to reduce nonlinearity. The network structure is important to network capacity. Motivated by the emerging of self-attention structure, we propose a Dense-and-Implicit-Attention (DIA) unit that enhances generalization by repeatedly fusing the information throughout different network layers. A modified Long Short Term Memory module in the DIA unit is built in parallel with the DNN to link multi-scale features from different depth levels of the network implicitly and densely.