

# Jerry Chee

---

Department of Computer Science  
Cornell University

JerryChee@cs.cornell.edu  
Jerry-Chee.github.io

Education	<b>Cornell University</b> Ph.D. in Computer Science Advisor: Chris De Sa	Ithaca, NY 2019 - 2024 (expected)
	<b>University of Chicago</b> B.S. in Computational and Applied Mathematics Advisor: Panos Toulis	Chicago, IL 2013 - 2017
Publications	<b>Jerry Chee</b> , Panos Toulis. <i>Convergence Diagnostics for Stochastic Gradient Descent</i> . In <i>AISTATS 2018 (oral presentation)</i>	
	<b>Jerry Chee</b> , Ping Li. <i>Understanding and Detecting Convergence for Stochastic Gradient Descent</i> . In <i>IEEE Big Data 2020</i>	
Talks	<b>Convergence Diagnostics for Stochastic Gradient Descent</b> <i>AISTATS 2018</i> , with Panos Toulis.	Canary I. Apr 2018
	<b>Statistical Properties of Stochastic Gradient Descent</b> <i>Joint Statistics Meeting</i> , with Panos Toulis.	Denver, CO Jul 2019
Projects In Progress	<b>Pruning Neural Networks via Interpolative Decompositions</b> – with Chris De Sa, in submission A principled approach to pruning by low-rank matrix approximation, using a novel application of the interpolative decomposition to approximate the activation output of a layer.	
	<b>Compress Neural Networks via Sampling Methods</b> – with Chris De Sa, in preparation A generic approach to quantization which utilizes the power of sampling methods to operate directly over the discrete problem space.	
	<b>Scalable Inference with Stochastic Gradient Descent (SGD)</b> – with Panos Toulis, in preparation Developed a large scale statistical inference procedure with SGD, with application to an outpatient care analysis with collaborators at UPenn hospital.	
	<b>Parallel Learning with Sublinear Communication</b> – with Chris De Sa, in preparation Building parallel learning algorithms which have distributed learning theoretic guarantees under sublinear communication.	

Industry Experience	<b>Microsoft</b> , IC3-AI	Redmond, WA
	<i>Intern</i>	Jun–Sept 2021
	<ul style="list-style-type: none"> <li>• Sped up inference of deep background noise suppression models used real-time in Teams.</li> <li>• Identified and implemented model compression methods supported by the neural network inference engines ONNX Runtime, CoreML, and TFLite.</li> </ul>	
	<b>Baidu</b> , Cognitive Computing Lab	Bellevue, WA
	<i>Research Intern</i>	Mar–Jul 2019
	<ul style="list-style-type: none"> <li>• Developed statistical convergence tests for variants of stochastic gradient descent with momentum and gradient compression. Published in IEEE Big Data 2020.</li> <li>• Utilized multi-task learning to increase the available training data in order to improve the predictive performance of graph neural networks.</li> </ul>	
	<b>McKinsey &amp; Company</b>	Boston, MA
	<i>Senior Analytics Fellow</i>	Oct 2017 - Feb 2019
	<ul style="list-style-type: none"> <li>• Implemented data science solutions at client organizations, working closely with business leaders and domain experts.</li> <li>• Led several data science initiatives in predictive maintenance for the network technology division of a top telecommunications company. <ul style="list-style-type: none"> <li>– Utilized a cost (of true positive, false positive, etc.) analysis for selecting the prediction target and implementation strategy which maximized business impact and modeling feasibility.</li> <li>– Built classification models for network and customer service use cases.</li> </ul> </li> </ul>	
Teaching	TA, CS 4780/5780: Machine Learning for Intelligent Systems	Fall 2019
	TA, CS 4787: Principles of Large-Scale Machine Learning	Spring 2020
	TA, CS 6787: Advanced Machine Learning Systems	Fall 2020
Outreach	<b>Skype A Scientist Volunteer</b>	Apr 2020-Present
	Video call with classrooms across the country to help educate students about research in computer science and career options as a quantitative scientist.	
Other Information	Programming: Python (PyTorch), R, Julia, C (MPI) Languages: Chinese (Limited oral proficiency)	