Research Experience for Undergraduates

Summer School on Mathematical Foundation of Data Science

June 6, 2022 --- July 15, 2022

Join Virtual Zoom Program

https://us06web.zoom.us/j/84400970067?pwd=R2Rpb2ZnSldESmJGT2NzMW1XMlNpdz09

Meeting ID: 844 0097 0067

Passcode: 718281

Sponsored by

Department of Mathematics
University of South Carolina

National Sciences Foundation
RTG award DMS 2038080

Organized by

Prof. Linyuan Lu,

Prof. Wuchen Li,

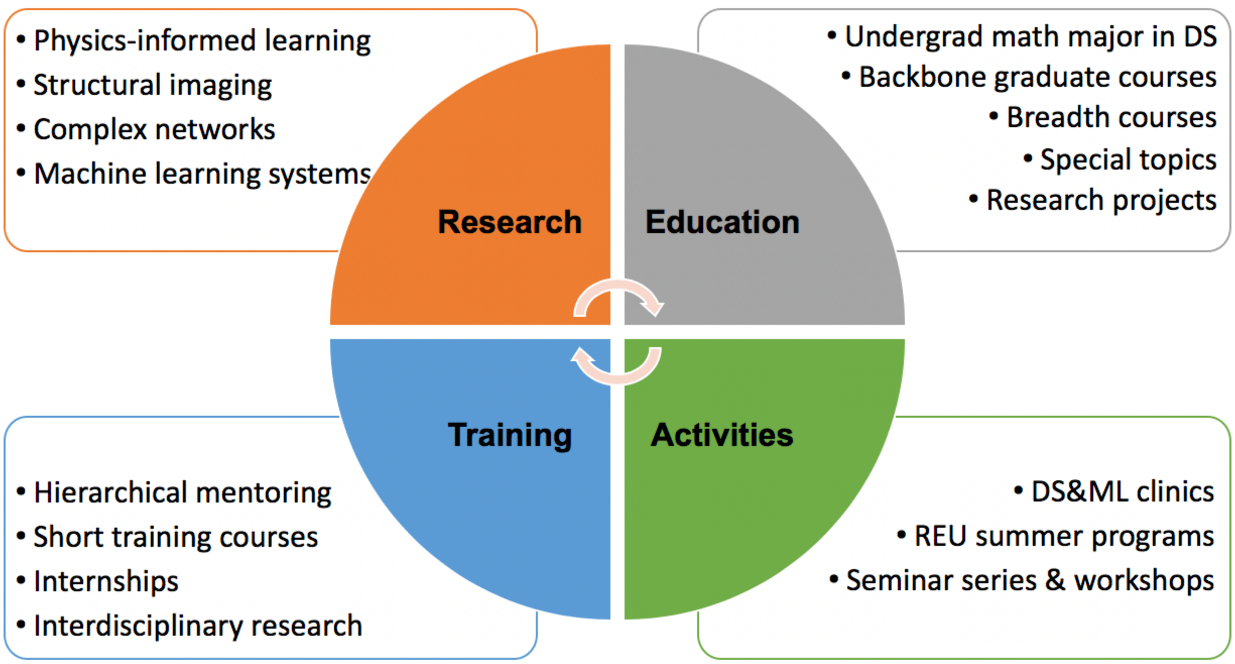
Prof. Qi Wang,

Prof. Zhu Wang

Table of Contents

Section 1: Program Overview

This REU summer program is part of the NSF RTG project **"RTG: Mathematical Foundation of Data Science at University of South Carolina"**, which aims to develop a multi-tier Research Training Program at the University of South Carolina (UofSC) designed to prepare the future workforce in a multidisciplinary paradigm of modern data science. The education and training models will leverage knowledge and experience already existing among the faculty and bring in new talent to foster mathematical data science expertise and research portfolios through a vertical integration of post-doctoral research associates, graduate students, undergraduate students, and advanced high school students. A primary focus of this project is to recruit and train U.S. Citizens, females, and underrepresented minority (URM) among undergraduate and graduate students, and postdocs through research led training in Data Science.



For more information on the NSF RTG project, please visit us at the following URL:
https://sc.edu/study/colleges_schools/artsandsciences/mathematics/my_mathematics/rtg/index.php

The REU summer program of this year runs virtually from June 6 to July 15. In the first two weeks, we teach four short course modules in Mathematical Foundation of Data Science to prepare undergraduate students for the basic level of research projects. Starting from the third week, students will be divided into four groups to work on research projects. Some guest speakers are invited to give talks on the latest development in the Mathematical Foundation of Data Science. On the last day of the program, students will present their research findings.

Section 2: Course Modulars

Course module 1: Linear Algebra

Instructor: Zhu Wang

Total hours: 10

Course Contents: Understand fundamental concepts in linear algebra, such as subspaces, projections, least squares, eigenvalue decomposition, and singular value decomposition, etc. Apply these concepts in solving the following central problems in linear algebra: n by n linear system Ax = b; m by n linear system Ax=b; n by n linear system Ax=λx; and m by n linear system Av=σu. The connection of linear algebra with many applications will be discussed as well.

Course module 2: Probability Theory and Optimization

Instructor: Wuchen Li

Total hours: 10

Course Contents: Study basic concepts in probability, statistics and optimizations: Probability distributions. Cumulative distributions. Moments. Mean. Variance. Covariance. Gaussian distribution. Samples. Fisher information matrix. Optimal conditions. Convexities. Gradient descent. Newton's method. Lagrange multiplier. KKT conditions.

Course module 3: Introduction to Complex Networks

Instructor: Linyuan Lu

Total hours: 10

Course Contents:

Graphs, trees, subgraphs, graph isomorphisms, paths, walks, cycles, graph product, planar graphs, Euler formula, Kuratowski's theorem, adjacency matrix, spectrum of special graphs, combinatorial Laplacian, matrix tree theorem, normalized Laplacian, Power law graphs, random graphs, Erdos-Renyi random graphs, random graphs for power law graphs, spectrum of random graphs, transportation distance, Ricci curvature of graphs, Concetration of Lipschitz functions over positive curvature graphs.

Course module 4: Machine Learning

Instructor: Qi Wang

Total hours: 10

Introduce the basic concept in machine learning, especially, to make a distinction between machine learning and optimization of an objective function or loss function. Discuss how to define the loss function using maximum likelihood estimation and Bayesian estimation. Introduce some basic machine learning algorithms such as logistic regression, k-means clustering, k-nearest neighbors, support vector machines, and decision trees. Introduce neural networks and deep learning: machine learning using deep neural networks, including fully connected convolutional and recurrent neural networks. Discuss some deep learning methods of learning dynamical systems underlying given time-series.

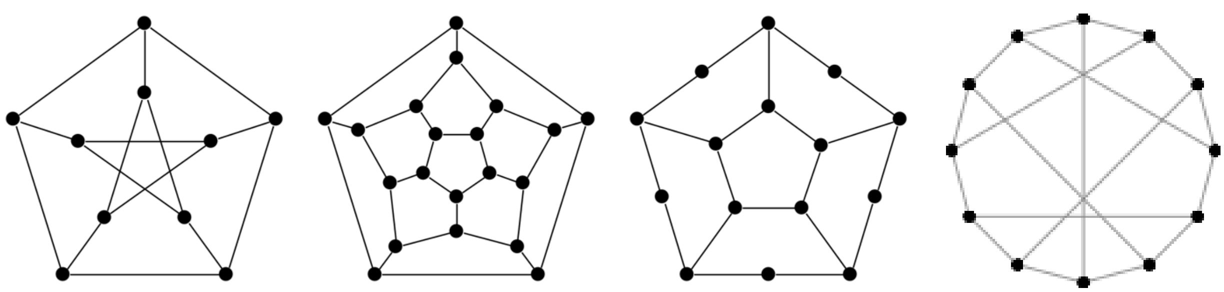
Section 3: Research Projects

Research projects in data-driven reduced order modeling

1. Dimensionality reduction in the parameter space. Study the classic linear dimensionality reduction approaches such as principal component analysis (PCA) and active subspace (AS), and recently developed deep learning methods for reducing the parameter space such as the nonlinear level-set learning (NLL) method. Compare their performances by considering high-dimensional function approximation problems and the numerical simulations of differential equations.
2. Data-driven reduced order modeling. Study the traditional model reduction approaches such as proper orthogonal decomposition (POD) or reduced basis method (RBM), and latest developments on deep learning-based nonlinear model reductions for overcoming the Kolmogorov barrier, such as those based on autoencoders. Compare their performances when simulating convection-dominated phenomena.

Research projects in complex graphs

1. A graph G is k-existentially closed (k-e.c.) if each k-set of vertices can be extended in all of the possible 2^k ways. Let $m_{ec}(k)$ be the minimum integer n such that a k-e.c. graph on n vertices exists. It is known that $m_{ec}(1) = 4$, $m_{ec}(2) = 9$ and $24 \leq m_{ec}(3) \leq 28$. Improve the bounds of $m_{ec}(3)$.
2. For each integer d, let F(d) (or f(d)) be the maximum integer n such that there exists a connected graph on n vertices with positive curvatures and maximum degree d (or d-regular graph respectively). It is known that $c_1^d \leq f(d) \leq F(d) \leq d^{c_2 d^2}$. Determine the magnitude of F(d) and f(d).
3. Classify all planar d-regular graphs with positive curvatures.



Research projects in transport information learning

Study and understand natural gradient methods from information geometry and optimal transport. Implement the natural gradient algorithms for supervised learning problems, and unsupervised learning problems.

1. In one-dimensional space, compute and implement the Fisher and Wasserstein information matrix for Gaussian and exponential distributions. Then, implement the natural gradient methods to learn the parameters.

2. In discrete graphical models, compute and implement the Wasserstein natural gradient methods for learning parameters in Boltzmann machines.

3. In two-layer neural network models, compute and implement the Wasserstein information matrix and its induced natural gradient dynamics.

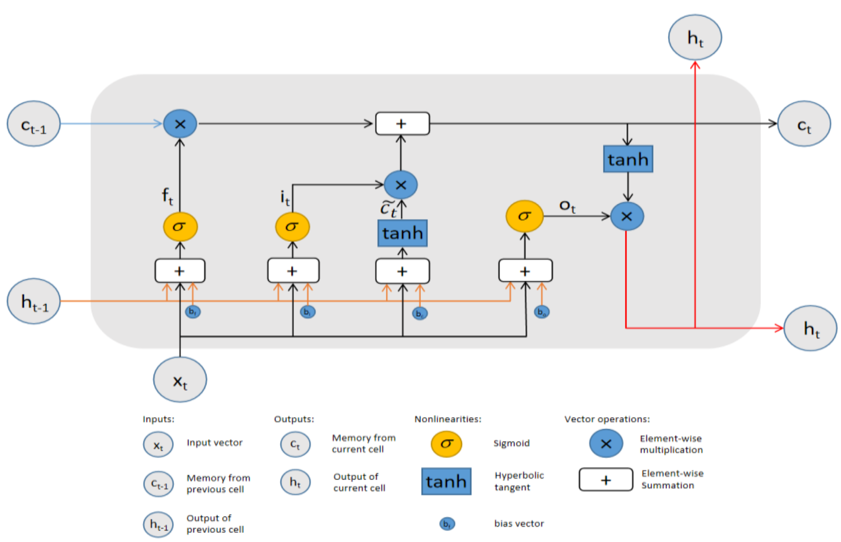
Research projects in dynamical system learning using time-series data

Physical laws and mechanisms in most real-world systems are formulated as time evolutionary equations known as the dynamical systems, which are either given as a discrete or continuous system. Measurements or outputs of the systems are customarily given in time series. Either solving the dynamical systems for a given initial data or learning the dynamical system with given measured dynamical system data (solutions) are important data science and machine learning problems. Here are some simplified projects related to machine learning of dynamical systems.

1. Survey the machine learning methods for solving dynamical systems and then develop more efficient machine learning algorithms for solving simple dynamical systems exploiting the fundamental structure and property of the underlying dynamical systems.

2. Survey model learning using deep neural networks and develop dynamical system models for given time-series data. Example, 1. learning patient-specific metabolic panel dynamics for lung cancer patients with 10 patient data. 2. Design diagnostic models for septic patients based on patient's time series data, etc.

3. Explore the power of dimension reduction in deep learning of dynamical systems. Using order reduction methods such as encoder/decoder to transform time series data to low dimensional latent space and then develop approximate models in the latent space.



This is an LSTM unit for discrete dynamical systems.

Section 3: Program Calendar

Week 1 (Week of June 6-10): Short Courses

Day		Activity	Instructor/ moderator
Monday June 6	9:00-10:00	Welcome and orientation	Lu
	10:00-12:00	Linear Algebra	Z. Wang
	12:00-2:00	Lunch Break	
	2:00-4:00	Probability Theory and Optimization	Li
	4:00-5:00	Recitation	Tom?
Tuesday June 7	9:00-11:00	Linear Algebra	Z. Wang
	11:00-12:00	Math Programming Lab	
	2:00-4:00	Introduction to Complex Networks	Lu
	4:00-5:00	Recitation	
Wednesday June 8	9:00-11:00	Linear Algebra and Deep Learning	Z. Wang
	11:00-12:00	Math Programming Lab	
	12:00-2:00	Lunch Break	
	2:00-4:00	Probability Theory and Optimization	Li
	4:00-5:00	Recitation	Tom?
Thursday June 9	9:00-11:00	Linear Algebra	Z. Wang
	11:00-12:00	Math Programming Lab	
	12:00-4:00	Lunch Break	
	2:00-4:00	Introduction to Complex Networks	Lu
	4:00-5:00	Recitation	Brooks
Friday June 10	9:00-11:00	Linear Algebra	Z. Wang
	11:00-12:00	Math Programming Lab	Brooks
	12:00-2:00	Lunch Break	
	2:00-4:00	Probability Theory and Optimization	Li
	4:00-5:00	Social Activity Hour	Megan

Week 2 (Week of June 13-17): Short Courses

Day		Activity	Instructor/ moderator
Monday June 13	9:00-11:00	Deep Learning	Q. Wang
	11:00-12:00	Math Programming Lab	
	12:00-2:00	Lunch Break	
	2:00-4:00	Introduction to Complex Networks	Lu
	4:00-5:00	Recitation	Thompson

Tuesday June 14	9:00-11:00	Deep Learning	Q. Wang
	11:00-12:00	Math Programming Lab	
	12:00-2:00	Lunch Break	
	2:00-4:00	Probability Theory and Optimization	Li
	4:00-5:00	Recitation	Tom?
Wednesday June 15	9:00-11:00	Deep Learning	Q. Wang
	11:00-12:00	Math Programming Lab	
	12:00-2:00	Lunch Break	
	2:00-4:00	Introduction to Complex Networks	Lu
	4:00-5:00	Recitation	Brooks
Thursday June 16	9:00-11:00	Deep Learning	Q. Wang
	11:00-12:00	Math Programming Lab	
	12:00-2:00	Lunch Break	
	2:00-4:00	Probability Theory and Optimization	Li
	4:00-5:00	Recitation	McKenzie
Friday June 17	9:00-11:00	Deep Learning	Q. Wang
	11:00-12:00	Math Programming Lab	Brooks
	12:00-2:00	Lunch Break	
	2:00-4:00	Introduction to Complex Networks	Lu
	4:00-5:00	Social Activity Hour	Megan

Week 3 (Week of June 20-24): Introduction of projects, group discussions on research projects, and guest lectures in data sciences

Day		Activity	Instructor/ moderator
Monday June 20	9:00-12:00	Project introduction	Professors
	3:00-5:00	Grouping students into parallel discussions	
Tuesday June 21	9:00-10:00	Parallel research sessions	GAs
	10:00-12:00	Self-research time	
	2:00-3:00	Mentors' Office hours	Professors
	3:00-5:00	Self-research time	
Wednesday June 22	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Guest lecture	
	3:00-5:00	Parallel research sessions	
Thursday June 23	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Mentors' Office hours	Professors
	3:00-5:00	Self-research time	
Friday	9:00-12:00	Parallel and joint research sessions	All

June 24	2:00-4:00	Social Activity Hour	McKay

Week 4 (Week of June 27 - July 1): Introduction of projects, group discussions on research projects, and guest lectures in data sciences

Day		Activity	Instructor/ moderator
Monday June 27	9:00-10:00	Parallel research sessions	GAs
	10:00-12:00	Self-research time	
	2:00-3:00	Guest lecture	
	3:00-5:00	Parallel research sessions	GAs
Tuesday June 28	9:00-10:00	Parallel research sessions	GAs
	10:00-12:00	Self-research time	
	2:00-3:00	Mentors' Office hours	
	3:00-5:00	Self-research time	
Wednesday June 29	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	GAs
	2:00-3:00	Guest lecture	
	3:00-5:00	Parallel research sessions	
Thursday June 30	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	GAs
	2:00-3:00	Mentors' Office hours	
	3:00-5:00	Self-research time	
Friday July 1	9:00-12:00	Parallel and joint research sessions	All
	2:00-4:00	Social Activity Hour	McKay

Week 5 (Week of July 4-8): Introduction of projects, group discussions on research projects, and guest lectures in data sciences

Day		Activity	Instructor/ moderator
Monday July 4	Holiday	No Activity	
Tuesday July 5	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Mentors' Office hours	
	3:00-5:00	Parallel research sessions	
Wednesday	9:00-10:00	Parallel research sessions	

July 6	10:00-12:00	Self-research time	
	2:00-3:00	Guest lecture	
	3:00-5:00	Parallel research sessions	
Thursday July 7	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Mentors' Office hours	
	3:00-5:00	Parallel research sessions	
Friday July 8	9:00-12:00	Parallel and joint research sessions	
	2:00-4:00	Social Activity Hour	McKay

Week 6 (Week of July 11-15): Introduction of projects, group discussions on research projects, and guest lectures in data sciences

Day		Activity	Instructor/ moderator
Monday July 11	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Guest lecture	
	3:00-5:00	Parallel research sessions	
Tuesday July 12	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Mentors' Office hours	
	3:00-5:00	Self-research time	
Wednesday July 13	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Guest lecture	
	3:00-5:00	Parallel research sessions	
Thursday July 14	9:00-10:00	Parallel research sessions	
	10:00-12:00	Self-research time	
	2:00-3:00	Mentors' Office hours	
	3:00-5:00	Self-research time	
Friday July 15	10:00-11:00	Plenary lecture	
	11:00-12:00	Plenary lecture	
	1:00-3:00	Group Reporting and presentation	
	3:00-4:00	Assessment	

Confirmed speakers:

Stanley Osher UCLA sjo@math.ucla.edu https://www.math.ucla.edu/~sjo/

Yunan Yang ZTH/Cornell yyn0410@gmail.com

Peng Chen UT Austin peng@oden.utexas.edu https://users.oden.utexas.edu/~peng/

Alex Lin	UCLA	atlin@math.ucla.edu	https://www.alextonglin.com/
Siting Liu	UCLA	siting6@math.ucla.edu	https://sites.google.com/view/siting6ucla/home
Levon Nurbekyan	UCLA	levonnurbekian@gmail.com	http://www.math.ucla.edu/~lnurbek
Samy Wu Fung	Colorado School of Mines	swufung@mines.edu	https://swufung.github.io/

Section 4: Participants contact information

Faculty

Linyuan Lu	Lu@math.sc.edu	https://people.math.sc.edu/lu/
Wuchen Li	wcli@math.ucla.edu	https://people.math.sc.edu/wuchen/
Qi Wang	qwang@math.sc.edu	https://people.math.sc.edu/wangq/
Zhu Wang	wangzhu@math.sc.edu	https://people.math.sc.edu/wangzhu/

Postdoc

William LInz	wlinz2@illinois.edu

Graduate Assistants

George Brooks	GHBROOKS@email.sc.edu
Alec Helm	AH191@email.sc.edu
Thomas Hamori	thamori@email.sc.edu
Megan McKay	MAM80@email.sc.edu
Black McKenzie	MMBLACK@email.sc.edu
Joshua Thompson	joshuact@email.sc.edu

Undergraduate students

Bryson Boast	bryson.boast@gmail.com	Berry College
Cade Stanley	cadets@email.sc.edu	University of South Carolina
David Liu	davdliu@umich.edu	University of Michigan
Dezmon Patten	dpatten@email.sc.edu	University of South Carolina
Jackson Ginn	jrginn47@gmail.com	University of South Carolina
Jacob Rottenberg	jrottenberg@umass.edu	University of Massachusetts
Jasdeep Singh	singhj72004@gmail.com	University of South Carolina
Jillian Garzarella	garzarej@email.sc.edu	University of South Carolina
John Ryan	Johnryan6465@gmail.com	University of South Carolina
Leah Mangano	lmangano@email.sc.edu	University of South Carolina
Luke Hammer	luke_hammer@brown.edu	Brown University
Malcolm Gaynor	malcolm.t.gaynor@gmail.com	Kenyon College
Peter Luo	pluo@college.harvard.edu	Harvard University
ROHIT SWAIN	roswain2002@gmail.com	University of South Carolina

Sabrina Barrat	sbarrat@yahoo.com	Saginaw Valley State University
TOBIN OTTO	totto@oberlin.edu	Oberlin College
Xingcheng Ren	xren@email.sc.edu	University of South Carolina
Zelong Li	lizelong831@ucla.edu	University of California, Los Angeles
Zhiyuan Li	zfl5150@psu.edu	Pennsylvania State University