

Tarun Kathuria

CONTACT INFORMATION

101, Jai Jawan Colony-3,
J.L.N. Marg,
Jaipur - 302018, India

E-mail: tarunkathuria@gmail.com
Homepage: tarunkathuria.github.io
Mobile: +91-8879346633

RESEARCH INTERESTS

Algebraic and Spectral Methods in Combinatorics, Geometry of polynomials and Functional analysis, Hardness of Approximation and Counting, Approximation Algorithms, Markov chains and Probability theory, Randomized Linear Algebra, Iterative methods for Convex and Non-Convex Optimization and their applications to Machine Learning and Database Theory

EMPLOYMENT

Microsoft Research India, Bangalore
Research Fellow, Algorithms and Theory Group

July 2015 - Present
Mentor: Dr. Amit Deshpande

EDUCATION

Indian Institute of Technology - Bombay, Mumbai, India
Bachelor of Technology (Honors)

July 2011 - June 2015

- Major: Computer Science & Engineering Minor: Applied Statistics and Informatics
- CGPA: 9.08/10.00
- Ranked 10th in the department (among 96 students)

PUBLICATIONS

1. Tarun Kathuria, S. Sudarshan. [Efficient and Provable Multi-Query Optimization](#). *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2017)*, to appear.
2. Tarun Kathuria, Amit Deshpande, Pushmeet Kohli. [Batched Gaussian Process Bandit Optimization via Determinantal Point Processes](#). *Advances in Neural Information Processing Systems (NIPS 2016)*
3. L. Elisa Celis, Amit Deshpande, Tarun Kathuria, Nisheeth K. Vishnoi. [How to be Fair and Diverse?](#) *3rd Workshop on Fairness, Accountability, and Transparency in Machine Learning (FATML 2016)*

MANUSCRIPTS

1. L. Elisa Celis, Amit Deshpande, Tarun Kathuria, Damian Straszak, Nisheeth K. Vishnoi. Exact Counting and Sampling with Constraints. *In Submission*
2. Amit Deshpande, Tarun Kathuria, Damian Straszak, Nisheeth K. Vishnoi. [Combinatorial Determinantal Point Processes](#).

(SUB)REVIEWER

Conference on Learning Theory (COLT) 2016
Neural Information Processing Systems (NIPS) 2016
ACM-SIAM Symposium on Discrete Algorithms (SODA) 2017

RESEARCH EXPERIENCE

Research Project, Microsoft Research Dr. Amit Deshpande and Prof. Nisheeth K. Vishnoi, EPFL
Combinatorial Determinantal Point Processes and Kadison-Singer January 2016 - Ongoing

- **Determinantal Point Processes (DPPs)** are popular probabilistic models in mathematics and statistical physics. They are also used in approximation algorithms and for modelling diversity in machine learning. However, efficient sampling algorithms only exist for sampling from unconstrained and cardinality-constrained DPPs
- Devised **exact sampling algorithms** for DPPs constrained to **partition matroids** (and more generally, to “linear constrained set families”). Algorithms based on *coefficients of carefully curated multivariate characteristic polynomials* of a matrix which can be computed in polynomial-time when the size of the partition is constant.
- Showed an *equivalence* between **exact counting/sampling from Partition-DPPs and mixed discriminants**, which generalize permanents
- Making the proof of the **Kadison-Singer** conjecture by [Marcus, Spielman and Srivastava](#) algorithmic is currently an outstanding open problem which would involve approximately computing largest roots of “mixed characteristic polynomials”, which generalize mixed discriminants
- Showed that our algorithms can be used to compute the **higher coefficients of the mixed characteristic polynomial** which directly implies **subexponential time algorithms for Kadison-Singer**.
- Proved a *reduction* between **exact computations of mixed discriminants and mixed characteristic polynomials** allowing for poly-time computation of the latter in special cases.
- Working on extensions and related problems

Undergraduate Dissertation, IIT Bombay
Efficient and Provable Algorithms for Multi-Query Optimization

Prof. S. Sudarshan
July 2014 - May 2015

- Work in the area of multi-query optimization (finding a set of common sub-expressions to materialize in order to find the optimal plan for a batch of queries) relies on *development of heuristics* which work well in practice. *No theoretical guarantees* on the quality of solution obtained by any heuristics exist so far
- Under assumptions of **submodularity** of a reformulation of the problem which is known to work well in practice, proposed an algorithm for **unconstrained normalized, submodular maximization** when the values may be negative, a case which has not been considered before
- **Proved an approximation factor** for the proposed algorithm. Further showed that it is NP-hard to achieve a **better approximation ratio**. Also proposed optimizations and pruning techniques for the algorithm which preserve the theoretical guarantee
- Integrated the above into **PyroJ**, a *Volcano/Cascades*-based query optimizer developed at **IIT Bombay** and observed 15 – 20% improvements over existing algorithms on benchmark queries

Research Project, Microsoft Research
Batched Gaussian Process Bandit Optimization via Determinantal Point Processes

Dr. Amit Deshpande and Dr. Pushmeet Kohli
April 2016 - May 2016

- Optimizing noisy evaluations of an **unknown function** with only *black-box evaluation oracles* is an important problem in machine learning as well as in a variety of other disciplines
- Modelling the function as a **Gaussian process** and employing **stochastic bandit optimization techniques** to optimize it by probing at few points has emerged as a powerful tool. However, it is desirable to run function evaluations *in parallel* while ensuring that the points probed in a batch are “far-apart”
- Showed that an existing method for batched Gaussian process bandit optimization is equivalent to modelling it as *choosing points greedily* to find the subset with maximum probability from a carefully defined **Determinantal Point Process (DPP)**
- Based on this, proposed a **DPP-sampling** based algorithm for this problem and showed **regret bounds better than** previous algorithm
- Implemented our algorithms and obtained *considerable improvements* for optimization of synthetic and real-world functions, e.g. *hyper-parameter optimization for extreme multi-label classification algorithms*

Research Project, Microsoft Research India
Evaluation of Anomaly Detectors in Data Analytics Platforms

Dr. S. Sellamanickam
July 2015 - December 2015

- Anomaly detectors for many modern data analysis platforms are created primarily using *loose domain knowledge*. Many such detectors may not be useful and may be misleading.
- It is, thus, desirable to *rank the detectors based on quality* and prune out bad detectors, in the **absence of accurately labelled data** from past anomalies
- Investigated properties of a restricted set of anomalies and devised an *ensemble-based model for ranking* as well as creating a more accurate detector

Internship, IBM Research Labs, Bangalore
Large Scale Topical Analysis using the Social Network

Dr. Indrajit Bhattacharya
May 2014 - July 2014

- Proposed extensions to existing **non-parametric topic modelling algorithms** to account for various relations amongst users like geographies, friend circles, achieving better accuracy than existing models
- Devised **influence-aware topic models** to capture the effect of *user information cascades* by incorporating the **Independent Cascade Model** in these models
- Implemented a preliminary version of the above in the **Hadoop** framework for large-scale Twitter analysis

Research Project, IIT Bombay
Graph Learning using Orthonormal Representation

Prof. Saketha Nath & Prof. Chiranjib Bhattacharyya
Autumn 2014

- **Proved an equivalence** between *margin complexity* of one-class SVMs and *Lovasz ϑ function* on graphs
- Explored approaches to proving the **statistical consistency** of transductive learning on graphs using *orthonormal representations* using *VC-dimension* and *Rademacher complexity* based risk bounds

ACADEMIC DISTINCTIONS AND AWARDS

Awarded **AP grade** for exceptional performance in *Introduction to Numerical Analysis, Electricity & Magnetism and Introduction to Linear Algebra*; awarded to **top 2 %** students at IIT Bombay

Awarded the **Institute Academic Prize** in 2012 for academic excellence at IIT Bombay

Secured 10.0/10.0 GPA in the 7th semester (Fall 2014)

Amongst the top **300 (0.1%)** to be selected for *Indian National Chemistry Olympiad, 2011*

Awarded **CBSE Merit Certificate** in *Mathematics (100% marks obtained)* & in *Computer Science* for being in the **top 0.1%** of the students taking the *CBSE 2011* high school examinations

Awarded **A* grade** in *Mathematics, Physics, Chemistry and Information Technology (Computer Science)* in the **IGCSE 2009 (University of Cambridge International Examinations)** in the 10th grade

KEY ACADEMIC PROJECTS

Subquery optimization in PyroJ

Guide: Prof. S. Sudarshan

Autumn 2014

- Implemented algorithms for elimination of subqueries in the **PyroJ** query optimizer at IIT Bombay using ideas like the *Apply* operator from **Execution Strategies for SQL Subqueries**, Elhemali et al., **SIGMOD '07**
- Implemented subquery optimization using *Magic sets* as well as by re-ordering and efficient evaluation of semi-joins and anti-joins

Chess Titans

Guide: Prof. Amitabha Sanyal

Spring 2013

- Developed a one player chess game, in PLT Scheme using the DrRacket in-built GUI Toolkit
- Implemented the **Minimax algorithm** with $\alpha - \beta$ **pruning** and various clever heuristics for move determination

Twitter Sentiment Analysis

Guide: Prof. Pushpak Bhattacharya

Spring 2014

- Implemented a feed-forward with back-propagation neural network and applied it for sentiment analysis of about 50k Twitter annotated tweets
- Implemented various syntactic, semantic and stylistic features for better feature selection. Achieved the highest accuracy in the class for the same

Detecting and Classifying Geometric Shapes

Guide : Prof. Varsha Apte

Autumn 2012

- Developed an application to detect various shapes like *conics*, *lines*, *polygons* in a given image
- Implemented well-known techniques like **Hough Transforms**, **Bresenham Line Drawing** and **Harris Corner Detection** along with novel approaches based on regression analysis and our own heuristics

Estimation of Gamma Parameters with Censored Samples

Guide: Prof. Siuli Mukhopadhyay

Autumn 2013

- Studied methods for estimation of gamma distribution parameters in censored data
- Implemented line search optimization techniques and Newton's numerical approximation methods to efficiently find the point and interval **Maximum Likelihood** estimates in Matlab

On the Complexity of Linear Prediction

Autumn 2014

Surveyed and presented literature on *risk*, *margin* and *covering number bounds* of linear classification algorithms using **Rademacher** & **Gaussian complexities** as part of an *Statistical Learning Theory* course

Computational Humor

Autumn 2014

Surveyed recent literature on **Computational Humor Recognition and Generation** and presented the same to a class of 90 students as part of an *AI* course, earning the **highest mark** for the same

TEACHING EXPERIENCE

Undergraduate Teaching Assistant

Summer 2013, Spring 2014, Spring 2015

Course : *Introduction to Numerical Analysis*

Prof. S. Baskar

Assisted the professor in *setting question papers and model solutions* for examinations, conducting *problem solving* sessions and invigilating for examinations for **three offerings** of the course

Undergraduate Teaching Assistant

Autumn 2012

Course : *Electricity & Magnetism*

Prof. Tomy C.V.

Assisted the professor in *setting question papers and model solutions* for examinations, conducting *problem solving* sessions and invigilating for examinations of the course

Teaching Volunteer, S.B.S. High School

Winter 2013

Course : *Basic Programming in Java*

Taught a class of 30 students in high school about fundamentals of *Java programming* and *algorithmic thinking*

TECHNICAL SKILLS

Programming

C++, Java, Python, Julia, Ruby, Haskell, Scheme (Lisp)

Databases

PostgreSQL, Hive, Hyracks, HBase, Neo4j

Web Development

JavaScript, Rails, Django

Software Packages

Hadoop, Mahout, Matlab, OpenCV

RELEVANT
COURSES
UNDERTAKEN

Core: Statistical Techniques in Data Mining, Implementation of Relational Database Systems, Advanced Databases, Foundations of Machine Learning, Topics in Machine Learning, Markov Decision Processes, Game Theory, Artificial Intelligence, Linear Optimization, Convex Optimization, Operating Systems, Networks, Computer Architecture, Automata Theory, Data Structures, and Algorithms, Algorithm Design

Breadth: Calculus, Linear Algebra, Differential Equations, Numerical Analysis, Electricity and Magnetism, Chemistry, Psychology

INDEPENDENT
STUDY

I have studied the following courses on [coursera.org](https://www.coursera.org)

- Machine Learning
- Probabilistic Graphical Models
- Algorithms - I & II
- Mining Massive Datasets

EXTRA
CURRICULAR
ACTIVITIES

- Participated in **Mozilla's MozBoot 2014**, an overnight code contribution sprint to bootstrap developer contribution to Mozilla's large open source projects like *Firefox* & *Servo*
- Completed a 1 year **Guitar** course offered by *National Sports Organization* at IIT Bombay
- Solved Rubik's Cube as part of IIT Bombay's successful Guinness World Record attempt for maximum number of people simultaneously solving the Rubik's Cube
- Served as Editor of my high-school magazine
- Won various national debate competitions at the high school level
- Playing Chess and solving game puzzles like Sudoku & Kakuro