Tavor Baharav

baharav@broadinstitute.org
www.tavorb.com

Research Interests

Developing provably efficient algorithms for computational genomics.

My recent work has focused on designing statistically valid methods for reference-free analysis of raw sequencing data. More broadly, I have also worked on data-adaptive algorithms for data science problems using multi-armed bandits. My research borrows tools from optimization, information theory, and probability theory to design practically efficient and theoretically grounded algorithms, with a focus in computational genomics.

Education

Broad Institute Sept 2023 - present
Postdoctoral Fellow at the Eric and Wendy Schmidt Center
Advisor: Rafael Irizarry

Stanford University Sept 2018 - Aug 2023
Ph.D., Electrical Engineering
Thesis: *Adaptive algorithms for data science and computational genomics*
Advisors: David Tse and Julia Salzman

University of California, Berkeley Aug 2014 - May 2018
B.S., Electrical and Computer Engineering, Highest Honors
Mathematics minor
Advisor: Kannan Ramchandran

Publications

George Henderson*, Adam Gudys*, **Tavor Z. Baharav**, Punit Sundaramurthy, Marek Kokot, Peter L. Wang, Sebastian Deorowicz, Allison Carey, Julia Salzman. *Ultra-efficient, unified discovery from microbial sequencing with SPLASH and precise statistical assembly*. bioRxiv preprint, 2024.

Marek Kokot*, Roozbeh Dehghannasiri*, **Tavor Z. Baharav**, Julia Salzman, Sebastian Deorowicz. *NOMAD2 provides ultra-efficient, scalable, and unsupervised discovery on raw sequencing reads*. bioRxiv preprint, 2023.

Tavor Z. Baharav, David Tse, Julia Salzman. *OASIS: An interpretable, finite sample valid alternative to Pearson's X2 for scientific discovery*. PNAS, in press 2024.

Roozbeh Dehghannasiri*, George Henderson*, Rob Bierman, Kaitlin Chaung, **Tavor Baharav**, Peter Wang, Julia Salzman. *Unsupervised reference-free inference reveals unrecognized regulated transcriptomic complexity in human single cells*. bioRxiv preprint, 2022.

Kaitlin Chaung*, **Tavor Z. Baharav***, George Henderson, Ivan Zheludev, Peter L. Wang, Julia Salzman. SPLASH: A statistical, reference-free genomic algorithm unifies biological discovery. Cell, 2023.

Tavor Z. Baharav, Tze Leung Lai. *Adaptive Data Depth via Multi-Armed Bandits*. Journal of Machine Learning Research (JMLR), 2023.

Yifei Wang, **Tavor Z. Baharav**, Yanjun Han, Jiantao Jiao, David Tse. *Beyond the Best: Distribution Functional Estimation in Infinite-Armed Bandits*. Advances in Neural Information Processing Systems (NeurIPS), 2022.

Tavor Z. Baharav, Gary Cheng, Mert Pilanci, David Tse. *Approximate Function Evaluation via Multi-Armed Bandits*. International Conference on Artificial Intelligence and Statistics (AISTATS), 2022.

Tavor Z. Baharav, Daniel L. Jiang, Kedarnath Kolluri, Sujay Sanghavi, Inderjit S. Dhillon. *Enabling Efficiency-Precision Trade-offs for Label Trees in Extreme Classification*. ACM

International Conference on Information and Knowledge Management (CIKM), 2021; *oral presentation*.

Vivek Bagaria*, **Tavor Z. Baharav***, Govinda M. Kamath*, David N. Tse. *Bandit-based Monte Carlo Optimization for Nearest Neighbors*. IEEE Journal on Selected Areas of Information Theory, 2021.

Ilai Bistritz, **Tavor Z. Baharav**, Amir Leshem, Nicholas Bambos. *One for All and All for One: Distributed Learning of Fair Allocations with Multi-player Bandits*. IEEE Journal on Selected Areas of Information Theory, 2021.

Govinda M. Kamath*, **Tavor Z. Baharav***, Ilan Shomorony. *Adaptive Learning of Rank-One Models for Efficient Pairwise Sequence Alignment*. Advances in Neural Information Processing Systems (NeurIPS), 2020.

Tavor Z. Baharav*, Govinda M. Kamath*, David N. Tse, Ilan Shomorony. *Spectral Jaccard Similarity: A new approach to estimating pairwise sequence alignments*. Cell Press: Patterns (Patterns), 2020.

Ilai Bistritz, **Tavor Z. Baharav**, Amir Leshem, Nicholas Bambos. *My Fair Bandit: Distributed Learning of Max-Min Fairness with Multi-player Bandits*. International Conference on Machine Learning (ICML), 2020.

Tavor Z. Baharav*, Govinda M. Kamath*, David N. Tse, Ilan Shomorony. *Spectral Jaccard Similarity: A new approach to estimating pairwise sequence alignments*. International Conference on Research in Computational Molecular Biology (RECOMB), 2020.

Tavor Z. Baharav, David Tse. *Ultra Fast Medoid Identification via Correlated Sequential Halving*. Advances in Neural Information Processing Systems (NeurIPS), 2019.

Tavor Z. Baharav, Kangwook Lee, Orhan Ocal, Kannan Ramchandran. *Straggler-proofing massive-scale distributed matrix multiplication with d-dimensional product codes*. IEEE Int. Symp. on Inf. Theory (ISIT) 2018.

Tavor Baharav, Mohini Bariya, Avideh Zakhor. *In Situ Height and Width Estimation of Sorghum Plants from 2.5d Infrared Images*. Electronic Imaging (EI), 2017.

EMPLOYMENT

Amazon Applied Scientist Intern: hosted by Prof. Sujay Sanghavi — Mar 2020 - Aug 2020
Frequency aware hierarchical clustering

Undergraduate researcher (Berkeley): Prof. Kannan Ramchandran — Apr 2016 - July 2018
Straggler mitigation using error correcting codes

Undergraduate researcher (Berkeley): Prof. Avideh Zakhor — May 2015 - Feb 2017
Plant stem width estimation in 2.5d images

HONORS AND AWARDS

Best Poster, DF/HCC Celebration of Early Career Investigators in Cancer Research	2024
SIGIR Student Travel Award for CIKM	2021
RECOMB 2020 travel fellowship	2020
NSF Graduate Research Fellowship (3 years)	2018
Stanford Graduate Fellowship (Alcatel-Lucent Fellow, 3 years)	2018
UC Berkeley EECS Honors Program	2016
Eta Kappa Nu and Tau Beta Pi	2015
UC Berkeley Regents' and Chancellor's Scholar	2014-2018

TEACHING EXPERIENCE

Head Graduate Student Instructor, Stanford EE276a (Information Theory) — Winter 2021-22

- Created material for and taught weekly discussion sections
- Created exams and homework assignments

Head Undergraduate Student Instructor, Berkeley EE126 (Random Processes) — Spring 2018

- Helped scale the course from 70 to 200 students from Spring 2017 to 2018
- Taught weekly discussion sections, coordinated a course staff of 11 TAs and graders
- Served as TA (normal) in Spring 2017

Reader, Berkeley, CS70 (Intro Probability), CS170 (Algorithms) Fall 2015, Spring 2016
- Mentored students, graded assignments, and hosted office hours

TECHNICAL SKILLS

Programming: Python, C++, LaTeX
Languages: English (native), Hebrew (native)

PROFESSIONAL SERVICE

Organizer for Stanford's Information Systems Colloquium (ISL) 2019 - Present

Eta Kappa Nu, Mu Chapter (UC Berkeley): Fall 2015 - May 2018
- *Community Service Officer*, Spring 2016
  - Organized community focused event for students from local Oakland High Schools
  - Secured industry sponsors, organized faculty speaker, student panels, CS and EE labs
- *Treasurer*, Fall 2016
  - Managed budget and allocated finances for club of over 100 active members

Reviewing
- Reviewer for IEEE TPAMI, IEEE TSIPN, IEEE Access, ISIT, NeurIPS, ICML, ICLR, AISTATS, ISMB, RECOMB
- Meta-Reviewer for ICML 2021 ITR3 workshop
- Reviewer award for ICLR 2022, NeurIPS 2022, ICML 2022, AISTATS 2023

INVITED TALKS

Cold Spring Harbor Laboratory: Biological Data Science meeting 2022
Platform presentation

Cornell ORIE Young Researcher's workshop 2021