NSF TRIPODS Phase II NSF 19-604

Collaborative Research: STRIDES: Southeastern Transdisciplinary Research Institute for Data Engineering and Science

The GT&Duke STRIDES Team

January 26, 2020

Summary

[this is from the LOI] We propose creating a new Institute aiming to explore and advance the synergies between mathematics, theoretical computer science (TCS), statistics, and electrical engineering in laying the theoretical foundations for data science. Proposed activities include national interdisciplinary workshops, collaborative cross-disciplinary projects, research labs with participants from all communities, joint seminars, and co-supervision of Ph.D. students and postdocs by multiple mentors and institutes, in order to catalyze and promote a true synergy among all related disciplines. Georgia Tech and Duke University will build on the existing strengths and prominence in all the research areas (mathematics, TCS, statistics, and electrical engineering) and build on synergistic avenues of research. Georgia Tech has renowned faculty in the College of Computing, the College of Engineering, the College of Sciences, and the Scheller College of Business and is considered the preeminent technology university in the Southeast. Duke is known as an institution for its interdisciplinarity, and Duke faculty and Duke faculty have world class expertise and recognition in data science, in addition Duke also has access to rich data repositories, including those of a top medical school and regional hospital as well as a 21st century social science data infrastructure built by the Social Science Research Institute (SSRI). The coupling of the strengths in the technological aspects of data science at Georgia Tech with the liberal arts, social science, and health aspects of data science at Duke offer a rich educational and research foundation in the Southeast. An important component of the proposal is to foster professional development both for junior data scientists as well as domain experts that want to become more involved in data science. Georgia Tech and Duke will develop innovative programs that will reach out to technology centers, medical centers, as well as social science centers with the goal of increasing the interactions between junior data scientists and domain experts. The two institutions will also promote diversity and representation that reflects the community via collaborations with Under-represented Minorities institutions.

Georgia Tech and Duke have already made significant investments. Georgia Tech has a new Interdisciplinary Research Institute on Data Engineering and Science (IDEaS), a new 21-story building (Coda) co-locating data science industry and academia in Midtown Atlanta, and serving as collaborative lead institution to the NSF South Big Data Hub, which spans 16 Southern states and over a hundred partner organizations. Duke University has made considerable investments in data science research and education through the Rhodes Information Initiative at Duke (iiD), initiated in 2013. The iiD has features including labs, team rooms, classrooms, offices, and the 2-story Ahmadieh Family Atrium, all of which are designed to incubate cross-disciplinary interactions among faculty, postdocs, and students. In addition, Duke has a new data science initiative focusing on information and analysis-driven health innovation entitled AI.Health which has a goal of engaging the biotechnology and biomedical industries in the NC Triangle region including the Duke Hospital system with data science expertise and developing the educational framework to train data scientists in this application space. Duke University and Georgia Tech are prominent research institutes in the southeast. Georgia Tech is conveniently located at the midtown of Atlanta, with easy access to public transportation (e.g., MARTA), which routes directly to Atlanta's international airport. The Hartsfield-Jackson Atlanta International Airport serves the largest number of passengers in the world and offers direct flights to most US cities. Many Fortune 500 companies have headquarters in Atlanta. There are many local companies that are data oriented. Duke University is in the Research Triangle in North Carolina with strong local industrial research presence. Both Duke University and Georgia Tech have strengths in all related research communities. The joint institute (by Duke and Georgia Tech) will directly benefit the entire southeast, while also having a positive national impact.

NSF requirement. Project Description, limited to 30 pages total, consists of each of the following topics:

The **intellectual focus of the proposed institute**; the rationale for the proposed institute, its mission and goals, and its expected impact; plans for future growth and resource development; proposed steps toward developing its role as a national resource; and results of prior NSF support of the institute if applicable. This section is not to exceed 20 pages total including results of prior NSF support, which may take up to 5 pages.

A tentative **schedule** of scientific activities, with plans for Year 1 and a provisional schedule for Years 2 and 3.

Plans for human resource development, including the selection and mentoring of a diverse cohort of students and postdoctoral participants, as appropriate, and the selection and involvement of researchers at all career levels.

Plans for outreach and for dissemination of outcomes.

I. Intellectual Focus of the Proposed Institute

(not to exceed 20 pages)

1 Introduction

The Georgia Institute of Technology and the Duke University propose to create **STRIDES:** *the* **S***outheastern* **T***ransdisciplinary* **R***esearch* **I***nstitute for* **D***ata* **E***ngineering and* **S***cience*. STRIDES will integrate research and education in mathematical, statistical, and algorithmic foundations for data science. The initial research topics of focus include **xxx the following will be updated later.** (T1) transcribing data with new models and mathematics, (T2) creating new paradigms for decentralized and scalable inference, and (T3) designing efficient strategies with theoretical guarantees, harnessing the combined perspectives of statistics, optimization, and numerical methods. The research will be carried out in the context of big datasets from multiple domains including biology, design, manufacturing, logistics, and sustainability.

During its lifetime, the institute will develop methods and tools to address high-impact multidisciplinary foundational challenges in data science. The institute will bring together mathematicians, computer scientists, and statisticians to jointly work with domain specialists with data challenges across engineering, science and computing. Data arising from experimental, observational, and/or simulated processes in the natural and social sciences and other areas have created enormous opportunities for understanding the world in which we live. Data science is already a reality in industrial and scientific enterprises and there is ever-increasing demand for both research and training in this field. Virtually every scientific discipline is expected to benefit from advances in theoretical foundations of data science, which we seek to strengthen through the STRIDES institute.

The spectrum of fundamental problems in data science is vast. We broadly categorize them as addressing 1) how the data is collected and interpreted, and 2) how the data is analyzed. To address the former, we propose to study modeling and analysis techniques for data with new features and in nontraditional formats (T1), as well as develop decentralized modeling and processing techniques which would not require the data to be transferred to a data center (T2). To address the latter aspect, we propose to study modeling approaches with optimal statistical and computational tradeoffs, and develop algorithms that can be accelerated, distributed/parallelized, are asynchronous and/or stochastic (T3). Key to fundamental advances in these research topics is to derive theoretical guarantees in both the asymptotic and finite-sample cases.

Intellectual merits. The emergence of massive computational power via cloud computing and supercomputing infrastructure has given theorists an unprecedented opportunity to join the fray of empirical science and make significant impact on applications. STRIDES will be particularly well placed to address the growing challenges in the foundations of data science. STRIDES's intellectual focus is to design and build transdisciplinary research programs that provide an enabling and cross-fertilizing platform of ideas and stakeholders (including theoreticians/scientists from domain sciences and users of technology).

Broader impacts. STRIDES will enrich careers of participants ranging from undergraduate students to senior researchers from around the nation in due time. We will make prudent efforts to reach out to diverse communities including participants from smaller colleges and institutions serving under-represented minorities. STRIDES team will actively engage in outreach through public lectures, curricular materials, press releases and dissemination via social channels.

1.1 Key Functions of the Institute

STRIDES will lead and administer a fully integrated program of Research, Education and Outreach in foundations of data science. The institute's research agenda will inform educational programs and will be driven by transdisciplinary solutions to pressing data science challenges. The institute will facilitate interactive interactions between theorists and practitioners with the aim of bringing fundamentally rigorous and broadly applicable solutions and tools to support data-driven discovery in sciences, engineering, and beyond.

Proposed activities include national interdisciplinary workshops, collaborative cross-disciplinary projects, research labs with participants from all communities, joint seminars, and co-supervision of Ph.D. students and postdocs by multiple mentors, in order to catalyze and promote a true synergy among mathematicians, statisticians, and theoretical and algorithmic computer scientists.

STRIDES is particularly well placed to address the growing challenges in data science. Nowadays, analysis of massive, dynamic, and complex data is an area of great importance in many domains. To support true understanding of what is feasible through data-driven approaches and develop broadly applicable and insightful solutions, it is imperative to establish theoretical foundations of data science. Much of the underlying intellectual foundations underpinning data science lies at the intersection between computer science, statistics, and mathematics. In Phase I, STRIDES's main mode of operation will be focused on creating and operating working groups, organizing national and international workshops, and innovation labs. Participating individuals will include senior, mid-career, and junior faculty members, early career researchers including research scientists and postdoctoral fellows, undergraduate and graduate students, and data science practitioners. All activities will be planned to include interdisciplinary researchers rooted in the three foundational disciplines: mathematics, theoretical computer science, and statistics. STRIDES will rapidly deploy information technology and communication infrastructure so that research findings can be quickly and effectively disseminated, while the research community at large can easily access and comment/critique STRIDES's choice of research programs and topics. The utilization of contemporary cyber-infrastructure is likely to lead to international impact and collaboration, which enhances the institute's ability to establish a solid foundation for data science research. The institute will create an intellectual atmosphere that connects theoreticians, practitioners, and scientists from across the nation and the world on a regular basis. Findings in STRIDES's activities will lead to presentations at major conferences and publications in refereed journals.

STRIDES's programs will enrich careers of participants ranging from undergraduate students to senior researchers from around the nation. Postdoctoral fellows and graduate students are introduced to collaborative research in the proposed activities and through workshops. STRIDES personnel will make prudent efforts to reach out to diverse communities including participants from smaller colleges and institutions serving under-represented minorities. STRIDES will conduct public outreach through public lectures, press releases, and dissemination via internet and social media. STRIDES will draw upon the nationally acclaimed Georgia Tech online degree programs (e.g., online masters programs in computer science and in analytics) and curriculum development efforts in machine learning and data science, so that students across the nation can learn about state-of-the-art and interdisciplinary research topics that are not typically covered within traditional campus courses. For Phase II, we propose additional activities (such as customized workshops) that will combine interactive projects and field trips to acquaint undergraduate and/or high-school students nationwide with data-science related techniques and the themes of the STRIDES's year-long programs.

1.2 University Support and Infrastructure

Georgia Institute of Technology (abbreviated Georgia Tech hereafter) is a world-class research university with extensive expertise in data analytics, statistics, theoretical computer science, operations research and simulation, and mathematics. All of its computing and engineering departments are ranked in the top 10 by the U.S. News & World Report, with over half in the top five. Most of its statistics, optimization, and operations research faculty are housed in the School of Industrial Systems and Engineering, ranked as the top such department in the nation. Georgia Tech is the largest producer of engineering degrees awarded to women and underrepresented minorities. Research and education at Georgia Tech is known for its real-world focus, and strong ties to government and industry.

Georgia Tech has shown very strong commitment to Data Science through several major investments in recent years. In 2016, Georgia Tech launched the interdisciplinary Research Institute (IRI) for Data Engineering and Science (IDEaS), charged with facilitating, nurturing, and promoting data science research and data-driven discovery across campus. Georgia Tech has both on-campus and on-line M.S. program in Analytics, and is launching a Ph.D. program in Machine Learning with a sizable data science component. It is investing in a $375 million, 21-story, 750,000+ sq. ft. building (termed Coda) devoted to data science and high performance computing, along with a 80,000 sq. ft. data center to host large-scale computing and data repositories. Planned for early 2019 occupancy, the building will be equally shared by Georgia Tech and relevant data science industry, promoting academia-industry interaction. The STRIDES institute will be administratively structured within IDEaS (led by co-PIs Aluru and Randall), enabling it to benefit from staff, infrastructure, space, and other resources provided through IDEaS, amplifying the impact of TRIPODS Phase I funding.

Georgia Tech has established itself as a national leader in the data sciences. Since 2015, it is serving as the collaborative-lead institution for the NSF South Big Data Regional Innovation Hub (led by co-PI Aluru), one of four such Hubs established to serve the nation. In this role, it supports regional and national-scale efforts in research, industry adoption, and training activities across 16 Southern states from Delaware to Texas, and Washington D.C. The Hub has over 150 partner organizations drawn from academia, industry, government labs, and non-profits. STRIDES team members have been at the forefront of the data science revolution, involved in the White House, NITRID, NSF, NIH, DOE, and DARPA big data initiatives, as well as funding from key programs such as NSF Bigdata, NIH BD2K, and DARPA XDATA.

Georgia Tech is located in Atlanta, the eighth largest economy in the nation and third among cities with the most Fortune 500 companies [34]. With the world's busiest airport and single-hop reachability to many destinations that Atlanta provides, Georgia Tech is uniquely positioned to serve meeting needs of the research community such as workshops and short courses. The proposed STRIDES institute aligns with the university's strategic plan, is synergistic with many new initiatives, and will be housed in the new Coda building alongside IDEaS and the South Hub.

2 Research Program

2.1 Optimal transport

2.2 Optimization related

2.3 Streaming data

2.4 Reinforcement learning

2.5 Partial differential equation identification

2.6 Applications in data science related fields

Symbiotically, data science theory informs new algorithms and methodologies, while the constraints of "real-world" applications define new directions for fundamental exploration. Within Georgia Tech the proximity to top researchers integrating inference and learning algorithms to solve real-world problems will strengthen the impact of our foundational work, finding solutions that address practical barriers.

3 Institute Activities and Community Involvement

The key goal of STRIDES institute is to not only establish a research program in the focus areas at the foundations of data science described earlier, but also nurture and grow a vibrant nationwide community around them, as well as undertake activities for community benefit and training. Georgia Tech's excellent reputation and its established research and educational programs in statistics, theoretical computer science (TCS), and mathematics, along with our vast networks of existing collaborations, will help accomplish this goal.

The following activities include workshops, innovation labs, short-term visitors, graduate student and post-doc enabled joint research, course development, and a lecture series.

3.1 Innovation Labs/Thinktanks

We propose to organize innovation labs bringing together researchers from TCS, Math, and Stat, taking into account both applied and foundational perspectives, working together to translate applied questions into foundational questions across all areas. The organizers will determine the themes of the innovation labs per the research programs proposed above, and will approach experienced senior researchers in the field with respect to mentoring and supervision. We will utilize the *Idea Lab* framework that has been developed and experimented successfully in other related proposal-fertilization activities. In particular, these innovation labs will take a form similar to study group sessions in industrial problem identification. These events will bring together senior and junior researchers, from different disciplines, to focus on specific data science related questions with the goal of turning the questions into new directions for novel theory and computations, which in turn will advance the foundations of data science. The focus areas are based on themes in data science, and the outcomes of the proposed focus activities can then be translated into subsequent collaboration and co-supervision of PhD and postdoc projects; they will also form the basis for future focused workshops, visitation of external experts, or research proposals. Teams in different areas can work in parallel during the innovation labs. Some groups will likely include industry and government contacts.

Experience. Georgia Tech already has excellent examples of generating think-tank activities. ARC (Algorithms & Randomness Center) was founded to create interdisciplinary expertise tackling cross-disciplinary and real-world problems with experts in foundational needs. GT-MAP (Georgia Tech Mathematics and Applications Portal) facilitates Mathematics as an effective research partner of the broader community at Georgia Tech, and provides a stable entity where researchers from the campus community can present their work and share ideas.

3.2 Short-term visitation

During Phase I, we will approach potential researchers who are interested in spending short periods of time at STRIDES to help enhance the proposed research. Due to budget limitations, we may start with short-term visits, with each researcher staying 1-4 weeks. We will proactively seek opportunities for co-hosting, tapping into other resources for joint sponsorship. We will also encourage and host sabbatical visitors to fully or partially associate with STRIDES.

3.3 Lecture Series in Mathematical Foundations of Data Science

STRIDES will invite high-profile researchers each year to give a series of lectures related to the foundations of data science. Specific examples of distinguished researchers we plan to invite include:

i) Prof. Roman Vershynin (University of Michigan) who has been working for many years on the development of non-asymptotic theory of random matrices and its applications to a number of problems in data science (compressed sensing, covariance estimation, statistics of networks, etc);

ii) Prof. Gabor Lugosi (Pompeu Fabra University, Barcelona, Spain) will be asked to lecture on Elements of Combinatorial Statistics.

iii) Prof. Boaz Klartag (Tel Aviv University, Israel), an expert in convex geometry and analysis, has agreed to give several lectures for a non-expert audience. Galyna Livshyts and Tetali plan to work with the local colleges and help host the lectures at Spelman college in Atlanta.

The institute will throw open these events to other universities/colleges in the area, e.g., Georgia State University, Emory University, Spelman College, as well as webcast the lectures if speakers so permit to reach nationwide audience. Atlanta is home to the largest concentration of colleges and universities in the South, including historically black colleges and universities (HBCUs).

4 Broader Impacts

4.1 Enabling Interdisciplinary Collaboration

One of the main objectives of the cross-disiplinary center will be fostering new collaborations across traditional disciplinary boundaries. Many of the PIs and senior personnel have strong track records for promoting interdisciplinary research and understand the challenges. We also have a lot of experience with data science and the effort required to ensure the theroetical questions pursued are of actual practical significance in the application domains. Educational modules will be integrated into all of STRIDES's activities, such as prefacing all workshops with tutorials and taking extra efforts to enusre their appropriateness for newcomers to various fields. We will provide means for interdisciplinary groups to have regular contact so that research forms a more solid interdisciplinary foundation.

4.2 Communications

An experienced Director of Communications will play a central role in ensuring the success of STRIDES. During Phase I, STRIDES will be supported by IDEaS communication director Jennifer Salazar. She has a record of success working with researchers on large interdisciplinary and multi-institution projects, possess rich experience leading research-related communication campaigns, digital media, and public relations, and is a talented writer. She will work closely with the Executive Leadership team to establish relationships, and to track achievements both internally and externally. Her primary responsibilities will be to proactively remain apprised of important successes, develop a strategic communication plan and product calendar, and drive awareness of project progress both internally across the wider team, and externally to the research community, stakeholders, and general public. We expect sustaining planned Phase II activity will require STRIDES hiring or at least cost-sharing such a position.

4.3 Recruiting students and faculty to STRIDES programs

Georgia Tech's track record with recruiting students to their online and degree programs (e.g., the 25-year old, highly visible, ACO Ph.D. program) assures us that once advertised these courses and programs will be oversubscribed. We will advertise on STRIDES web portal and other Georgia Tech websites as we develop new short courses and webinars. We will actively recruit students at large, including underrepresented minorities and women, and faculty from smaller colleges, to participate in our programs and/or apply for graduate fellowships and assistantships.

4.4 Interaction with domain experts

Due to strong presence of engineering programs and local data science industry, we have an efficient and effective interface to application domains. Domain experts will be made aware of the developments and advances that theoretical foundations of data sciences have to offer. The proposed institute will incorporate deep and frequent interactions between theoreticians and domain experts. Success of theoretical foundations of data sciences will strongly depend on connections between statistical accuracy and quality-of-approximation as a tradeoff of various computational constraints that are imposed by modern computing infrastructure; The fact that STRIDES will also be co-located with the high-performance computing center will enable such connections.

5 Results from prior NSF support

The PIs are each supported by multiple NSF awards during the prior five years. Some reflect existing collaborative strengths among the PIs and senior personnel. We summarize a few.

• **Xiaoming Huo** is supported by DMS-1613152, and DMS-1106940, Achieving spatial adaptation via inconstant penalization: theory and computational strategies, Aug 2011–July 2014, $140,000.

Intellectual Merit. The PI investigated how to achieve adaptive functional estimation when the underlying model has inhomogeneous roughness. Publications that were partially enabled by this project include [108, 107, 43, 5, 103, 45, 44, 46, 47, 112, 104, 40, 24, 95, 40].

Broader Impacts. Huo disseminated the research through multiple external talks. He trained three female Ph.D. students (one graduated and is now an assistant professor).

• **Srinivas Aluru** has been supported by 11 NSF grants, including 6 from CCF. We report on IIS-1247716/1416259, BIGDATA: Genomes Galore - Core Techniques, Libraries, and Domain Specific Languages for High-Throughput DNA Sequencing, Jan 2013–Dec 2017, $2,000,000.

Intellectual Merit: This project is led by the PI in partnership with Stanford and Virginia Tech. The PI's group developed parallel algorithms for a variety of string and graph based index and data structures prevalent in bioinformatics. The work received multiple recognitions: best student paper (Supercomputing 2015); first selected paper by ACM SIGHPC under the scientific reproducibility initiative; and selection as a benchmark for Student Cluster Competition (Supercomputing 2016).

Broader Impacts: Research results are disseminated as software libraries (github.com/ParBLiSS). The PI ran three international workshops to lead community efforts for high-throughput sequence analytics. He also assisted NSF in organizing a U.S.-Japan Big Data PI meeting.

• **Prasad Tetali** was supported by DMS-1407657 (July 2014–June 2018, $288,000), CCF-1415496 and CCF-1415498 (Mar 2014–Feb 2017, $600,000).

Intellectual Merit. The projects funded novel directions in optimal transport and discrete and continuous optimization, inspired by concrete real-world problems [35, 36]. The team modeled industrial challenges arising from current-day needs, and identified the precise algorithmic and optimization tools needed to solve the corresponding issues [21].

Broader Impacts. Tetali and team trained six Ph.D. students (one female) and two postdocs, and co-hosted an Industry Day for theoreticians. The team developed a pilot expert system *Ask Minmax*, to help non-experts diagnose and identify commonly encountered optimization problems.

• **Jianfeng Lu** has been supported by several NSF projects in the past. The most recent and relevant one is NSF DMS Grant 1454939, "CAREER: Research and training in advanced computational methods for quantum and statistical mechanics" (09/01/2015–08/31/2020).

Intellectual Merits: The research goal of the project is to innovate and analyze efficient algorithms based on advanced computational mathematics for electronic structure theory and computational statistical mechanics, which will greatly advance the scope of ab initio simulations with applications in chemistry, materials science, and related fields. Under the support, Lu has been highly productive and have already accomplished 83 published or accepted journal publications [25, 67, 60, 89, 56, 110, 84, 90, 48, 50, 88, 91, 22, 59, 86, 98, 100, 57, 29, 80, 30, 106, 68, 85, 62, 51, 55, 94, 23, 75, 111, 53, 61, 79, 32, 10, 49, 27, 39, 92, 7, 26, 16, 87, 6, 11, 76, 4, 28, 109, 97, 72, 105, 73, 15, 78, 9, 74, 3, 13, 14, 19, 38, 42, 66, 64, 69, 70, 81, 82, 99, 113, 12] and submitted manuscripts [8, 41, 63, 71, 83, 17, 77, 58, 18, 101, 1, 96, 20, 65, 31, 2, 52, 37, 54, 93, 102, 33].

Broader Impacts: During the project, the PI has advised 10 undergraduate students for summer research projects (4 of them female) and is currently advising 3 graduate students working on this project. In addition, the PI has advised a year-long research project with a team consists of 5 high school students (2 of them female) from North Carolina School of Science and Mathematics. Moreover, the PI has co-organized and taught a MSRI-LBNL Summer School on Mathematical Introduction to Electronic Structure Theory and taught in an IPAM Summer School on Putting Back the Theory into the Density Functional Theory. These summer schools attracted in total around 100 graduate students and early career researchers from US and abroad.

The PI has presented the results obtained during the project in many colloquium talks, seminar talk and invited conference presentations. The research results are further disseminated through interactions and collaborations with computational chemists and materials scientists. The PI has also organized many workshops and conferences related with the project, including the KI-Net Conference on Mathematical and Computational Methods in Quantum Chemistry, May 2016; SAMSI Workshop on Trends and Advances in Monte Carlo Sampling Algorithm, December 2017; the 42nd SIAM Southeastern Section Conference, March 2018; CSCAMM Workshop on Mathematical and Numerical Aspects of Quantum Dynamics, June 2018. These conferences and workshops bring together experts and young researchers across disciplines.

II. Schedule

(The next three parts should be no more than 10 pages.)

A tentative **schedule** of scientific activities, with plans for Year 1 and a provisional schedule for Years 2 and 3.

III. Plans for Human Resource Development

Plans for human resource development, including the selection and mentoring of a diverse cohort of students and postdoctoral participants, as appropriate, and the selection and involvement of researchers at all career levels.

5.1 Graduate students and Postdocs

In Phase I, STRIDES has budgeted two graduate research assistantships and a partially sponsored postdoctoral fellowship. We believe that collaboration can be more productive through cross-supervision; the funding will foster cross-collaborations that take a future-looking approach to recruitment and research training.

Cross-disciplinary graduate/postdoc co-supervision. The graduate research assistantship will support students to be co-advised across academic units on campus. Supervisors must commit to be involved in co-supervision. Students will work on trans-disciplinary projects. The students learn a new field, bring that knowledge/connection back to home group, and bring new expertise to the other sponsoring group. Fellows with advanced preparation and degrees will be admitted in a postdoc position. The joint advising structure with be similar, with additional responsibilities.

Georgia Tech already has several interdisciplinary graduate programs. A new Ph.D. program in Machine Learning is recently approved and will start running in 2018. Another interdisciplinary Ph.D. program that has thrived for 25 years is the Algorithms, Combinatorics and Optimization (ACO) program, run jointly by the schools of mathematics, computer science and industrial and systems engineering, cutting across the Colleges of Sciences, Computing and Engineering. STRIDES participants have ample experience with cross-disciplinary training. They have a strong track record of mentoring and placement: e.g., Ruta Mehta and Jugal Garg (ARC postdocs, now faculty in CS and OR at UIUC, Illinois), Phillipe Rigollet (Math postdoc, now faculty at MIT), Will Perkins (NSF postdoc, faculty at Birmingham, U.K., moving to University of Waterloo, Canada), Kevin Costello (NSF postdoc, faculty at UC-Riverside); the first batch of NSF-funded IMPACT postdocs Maryam Yashtini and Christina Frederick soon to start faculty appointments at Georgetown and NJIT, respectively; Stas Minsker (Math/Stats student, faculty at USC, LA), Nayantara Bhatnagar, Adam Marcus, Amin Saberi and Emma Cohen (all ACO students, now faculty at Delaware, Princeton and Stanford, and a researcher at the Center for Communications Research in Princeton, respectively).

5.2 Curriculum Development

STRIDES will support the development of transdisciplinary courses for foundations of data science. Justin Romberg and Mark Davenport will design a new data science related machine learning course. Tetali and others will develop “Mathematics of Data Science” at the undergraduate level. Planned topics include fundamentals from high-dimensional subspaces, singular value decomposition, Markov chains, and machine learning. The institute will support activities whose benefits transcend Georgia Tech through development of tutorials, on-line courses, course modules, lecture notes, tutorials, and sharable slide sets.

This line of activities is aligned with many existing efforts on Georgia Tech campus. We describe a select few. Michael Lacey taught a special topics course on “Mathematics of Compressive Sensing” during Fall 2016, which had a large (50+) attendance. Jeff Wu and Arkadi Nemirovski are currently co-teaching a course on topics at the interface of statistics and optimization. Arkadi Nemirovski, Vladimir Koltchinskii, and others are writing a book on optimization methods in statistics. Georgia Tech also has outstanding on-line degree programs hailed as a national model.

IV. Plans for Outreach and for Dissemination of Outcomes

Plans for outreach and for dissemination of outcomes.

5.3 Workshops

STRIDES will hold about three workshops each year, along the lines of the three research themes the institute focuses on. The workshops will be national gatherings of junior and senior researchers, early career researchers such as postdocs, and graduate and senior undergraduate students, with additional international participants. These will be week-long events which will include (besides research presentations) tutorials, poster sessions, panels, working groups, open problem sessions, and tea-time gatherings to encourage collaboration among researchers. STRIDES will use advanced information technology to disseminate findings from these workshops online.

Our team has planned a number of initial workshops to hit the ground running as soon as the institute is formed:

1. With the help of the Algorithms & Randomness Center (ARC), we propose a workshop on *Randomness in Data Science and Optimization*, and feature top experts in TCS and optimization. This will be co-organized by Singh, Vempala, Vigoda and Tetali.

2. Koltchinskii, Nemirovski, and Romberg will co-organize a workshop addressing current challenges in optimization algorithms for modern datasets, high-dimensional statistics and non-convex inference problems, bringing together relevant statisticians and experts in continuous optimization and signal processing.

3. Huo and his colleagues (e.g., Le Song) will organize a workshop on decentralized and scalable statistical inference, as well as deep learning related methodologies.

Topics for other workshops will be decided by the STRIDES team, taking community feedback and evolving data science landscape into account for maximum impact. External researchers will be welcomed and involved in both organizing the scientific program of the workshops as well as plan topic areas and activities. By leveraging support from synergistic institutions at Georgia Tech (IDEaS, ARC, South Hub, etc.), we will be able to organize or co-organize more workshops than the nine budgeted. We will also hold one or two workshops focused on education in foundations of data science, vital to creating future workforce in this economically important area.

Prior experience: The PIs have track record of hosting interdisciplinary workshops, some of which have launched new subfields in recent years. ARC hosted several workshops bringing experts in randomized algorithms, MCMC and phase transitions, submodular optimization and network science. Externally, Tetali co-organized a thematic workshop *Graphical Models, Statistical Inference and Algorithms* (GRAMSIA) twice at the NSF-funded math institutes IPAM (January 2012) and IMA (May 2015). Koltchinskii co-organized a workshop at the NSF-funded institute SAMSI in 2014 on *Geometric Aspects of High-Dimensional Inference*. Huo organized similar events at SAMSI, and is currently involved in the forthcoming Joint Statistics Meeting, and a Banff workshop on data science related topics. These workshops have been instrumental in bringing relevant experts to come together and exchange novel ideas and breakthrough algorithms. The themes have evolved along with the frontier topics at the heart of the foundations of data science.

References

[1] Andrea Agazzi and Jianfeng Lu. Temporal-difference learning for nonlinear value function approximation in the lazy training regime, 2019. preprint, arXiv:1905.10917.

[2] Dong An, Teresa Head-Gordon, Lin Lin, and Jianfeng Lu. Convergence of stochastic-extended Lagrangian molecular dynamics method for polarizable force field simulation, 2019. preprint, arXiv:1904.12082.

[3] Jing An, Jianfeng Lu, and Lexing Ying. Stochastic modified equations for the asynchronous stochastic gradient descent. *Inf. Inference*, in press.

[4] Thomas Barthel and Jianfeng Lu. Fundamental limitations for measurements in quantum many-body systems. *Phys. Rev. Lett.*, 121:080406, 2018.

[5] K. Bastani, Z. (James) Kong, W. Huang, X. Huo, and Y. Zhou. Fault diagnosis using an enhanced relevance vector machine (RVM) for partially diagnosable multi-stationassembly processes. *IEEE Transactions on Automation Science and Engineering*, 10(1):124–136, January 2013.

[6] Zhenning Cai and Jianfeng Lu. A quantum kinetic Monte Carlo method for quantum many-body spin dynamics. *SIAM J. Sci. Comput.*, 40:B706–B722, 2018.

[7] Zhenning Cai and Jianfeng Lu. A surface hopping gaussian beam method for high-dimensional transport systems. *SIAM J. Sci. Comput.*, 40:B1277–B1301, 2018.

[8] Zhenning Cai, Jianfeng Lu, and Kevin Stubbs. On discrete Wigner transforms, 2018. preprint, arXiv:1802.05834.

[9] Zhenning Cai, Jianfeng Lu, and Siyao Yang. Inchworm Monte Carlo method for open quantum systems. *Comm. Pure Appl. Math.*, in press.

[10] Yu Cao and Jianfeng Lu. Lindblad equation and its semiclassical limit of the Anderson-Holstein model. *J. Math. Phys.*, 58:122105, 2017.

[11] Yu Cao and Jianfeng Lu. Stochastic dynamical low-rank approximation method. *J. Comput. Phys.*, 372:564–586, 2018.

[12] Yu Cao and Jianfeng Lu. Tensorization of the strong data processing inequality for quantum chi-square divergences. *Quantum*, in press.

[13] Yu Cao, Jianfeng Lu, and Yulong Lu. Exponential decay of Rényi divergence under Fokker-Planck equations. *J. Stat. Phys.*, 176:1172–1184, 2019.

[14] Yu Cao, Jianfeng Lu, and Yulong Lu. Gradient flow structure and exponential decay of the sandwiched Renyi divergence for primitive Lindblad equations with gns-detailed balance. *J. Math. Phys.*, 60:052202, 2019.

[15] Hongxu Chen, Qin Li, and Jianfeng Lu. Numerical coupling methods of BGK model and Euler equation through linearized Knudsen layer. *J. Comput. Phys.*, in press.

[16] Huajie Chen, Jianfeng Lu, and Christoph Ortner. Thermodynamic limit of crystal defects with finite temperature tight binding. *Arch. Ration. Mech. Anal.*, 230:701–733, 2018.

[17] Ke Chen, Qin Li, Jianfeng Lu, and Stephen J. Wright. Random sampling and efficient algorithms for multiscale PDEs, 2018. preprint, arXiv:1807.08848.

[18] Ke Chen, Qin Li, Jianfeng Lu, and Stephen J. Wright. A low-rank Schwarz method for radative transport equation with heterogeneous scattering coefficient, 2019. preprint, arXiv:1906.02176.

[19] Ke Chen, Qin Li, Jianfeng Lu, and Stephen J. Wright. Randomized sampling for basis functions construction in generalized finite element methods. *Multiscale Model. Simul.*, in press.

[20] Ziang Chen, Yingzhou Li, and Jianfeng Lu. Tensor ring decomposition: Energy landscape and one-loop convergence of alternating least squares, 2019. preprint, arXiv:1905.07101.

[21] Henrik I Christensen, Arindam Khan, Sebastian Pokutta, and Prasad Tetali. Approximation and online algorithms for multidimensional bin packing: A survey. *Computer Science Review*, 2017.

[22] Bruno Cornelis, Haizhao Yang, Alex Goodfriend, Noelle Ocon, Jianfeng Lu, and Ingrid Daubechies. Removal of canvas patterns in digital acquisitions of paintings. *IEEE Trans. Image Process.*, 26:160–171, 2017.

[23] Shibin Dai, Bo Li, and Jianfeng Lu. Convergence of phase-field free energy and boundary force for molecular solvation. *Arch. Ration. Mech. Anal.*, 227:105–147, 2018.

[24] Debraj De, Wen-Zhan Song, Mingsen Xu, Diane Cook, and Xiaoming Huo. FindingHuMo:real-time tracking of motion trajectories from anonymous binary sensing in smart environments. In *The 32nd International Conference on Distributed Computing Systems (ICDCS'12)*, 2012. (acceptance ratio 13%: 71 out of 515).

[25] Ricardo Delgadillo, Jianfeng Lu, and Xu Yang. Gauge-invariant frozen Gaussian approximation method for Schrödinger equation with periodic potentials. *SIAM J. Sci. Comput.*, 38:A2440–A2463, 2016.

[26] Ricardo Delgadillo, Jianfeng Lu, and Xu Yang. Frozen Gaussian approximation for high frequency wave propagation in periodic media. *Asymptotic Anal.*, 110:113–135, 2018.

[27] Qiang Du, Xingjie Helen Li, Jianfeng Lu, and Xiaochuan Tian. A quasinonlocal coupling method for nonlocal and local diffusion models. *SIAM J. Numer. Anal.*, 56:1386–1404, 2018.

[28] D. Fang and J. Lu. A diabatic surface hopping algorithm based on time dependent perturbation theory and semiclassical analysis. *Multiscale Model. Simul.*, 16:1603–1622, 2018.

[29] Yuan Gao, Jian-Guo Liu, and Jianfeng Lu. Continuum limit of a mesoscopic model of step motion on vicinal surfaces. *J. Nonlinear Sci.*, 27:873–926, 2017.

[30] Yuan Gao, Jian-Guo Liu, and Jianfeng Lu. Weak solution of a continuum model for vicinal surface in the attachment-detachment-limited regime. *SIAM J. Math. Anal.*, 49:1705–1731, 2017.

[31] Yuan Gao, Jian-Guo Liu, Jianfeng Lu, and Jeremy L. Marzuola. Analysis of a continuum theory for broken bond crystal surface models with evaporation and deposition effects, 2019. preprint, arXiv:1905.00179.

[32] Ludwig Gauckler, Jeremy Marzuola, Jianfeng Lu, Frederic Rousset, and Katharina Schratz. Trigonometric integrators for quasilinear wave equations. *Math. Comp.*, 88:717–749, 2019.

[33] Rong Ge, Holden Lee, and Jianfeng Lu. Estimating normalizing constants for log-concave distributions: Algorithms and lower bounds, 2019. preprint, arXiv:1911.03043.

[34] Geolounge. Fortune 1000 companies list. https://www.geolounge.com/fortune-1000-companies-list-2016/, 2016.

[35] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, Yan Shu, and Prasad Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *arXiv preprint arXiv:1509.04202*, 2015.

[36] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. Kantorovich duality for general transport costs and applications. *arXiv preprint arXiv:1412.7480*, 2014.

[37] Michael Holst, Houdong Hu, Jianfeng Lu, Jeremy L. Marzuola, Duo Song, and John Weare. Symmetry breaking in density functional theory due to Dirac exchange for a Hydrogen molecule, 2019. preprint, arXiv:1902.03497.

[38] Hui Huang, Jian-Guo Liu, and Jianfeng Lu. Learning interacting particle systems: diffusion parameter estimation for aggregation equations. *Math. Models Methods Appl. Sci.*, 29:1–29, 2019.

[39] Yufang Huang, Jianfeng Lu, and Ping-Bing Ming. A concurrent global-local numerical method for multiscale PDEs. *J. Sci. Comput.*, 76:1188–1215, 2018.

[40] Xiaoming Huo and Gabor J. Szekely. Fast computing for distance covariance. *Technometrics*, 2015. Accepted.

[41] Yuehaw Khoo, Jianfeng Lu, and Lexing Ying. Solving parametric pde problems with artificial neural networks, 2017. preprint, arXiv:1707.03351.

[42] Yuehaw Khoo, Jianfeng Lu, and Lexing Ying. Solving for high dimensional committor functions using artificial neural networks. *Res. Math. Sci.*, 6:1, 2019.

[43] H. Kim and X. Huo. Locally optimal adaptive smoothing splines. *Journal of Nonparametric Statistics*, 24(3):665–680, September 2012.

[44] H.-Y. Kim, X. Huo, and J. Shi. A single interval based classifier. *Annals of Operations Research*, 216(1):307–325, 2014. http://www.springer.com/alert/urltracking.do?id=L471c28fMeb6377Sae0acc1.

[45] Heeyoung Kim and Xiaoming Huo. Optimal sampling and curve interpolation via wavelets. *Applied Mathematics Letters*, 26:774–779, 2013.

[46] Heeyoung Kim and Xiaoming Huo. Asymptotic optimality of a multivariate version of the generalized cross validation in adaptive smoothing splines. *Electronic Journal of Statistics*, 8(0):159–183, 2014.

[47] Heeyoung Kim, Xiaoming Huo, M. Shilling, and H. D. Tran. A lipschitz regularity based statistical model, with applications in coordinate metrology. *IEEE Transactions on Automation Science and Engineering*, 11(2, ITASC7):327–337, April 2014.

[48] Rongjie Lai and Jianfeng Lu. Localized density matrix minimization and linear scaling algorithms. *J. Comput. Phys.*, 315:194–210, 2016.

[49] Rongjie Lai and Jianfeng Lu. Point cloud discretization of Fokker-Planck operators for committor functions. *Multiscale Model. Simul.*, 16:710–726, 2018.

[50] Chen Li, Jianfeng Lu, and Weitao Yang. Gentlest ascent dynamics for calculating first excited state and exploring energy landscape of Kohn-Sham density functionals. *J. Chem. Phys.*, 143:224110, 2015.

[51] Chen Li, Jianfeng Lu, and Weitao Yang. On extending Kohn-Sham density functionals to systems with fractional number of electrons. *J. Chem. Phys.*, 146:214109, 2017.

[52] Lei Li, Yingzhou Li, Jian-Guo Liu, Zibu Liu, and Jianfeng Lu. A stochastic version of Stein variational gradient descent for efficient sampling, 2019. preprint, arXiv:1902.03394.

[53] Lei Li, Jian-Guo Liu, and Jianfeng Lu. Fractional stochastic differential equations satisfying fluctuation-dissipation theorem. *J. Stat. Phys.*, 169:316–339, 2017.

[54] Lei Li, Jianfeng Lu, Jonathan Mattingly, and Lihan Wang. Numerical methods for stochastic differential equations based on Gaussian mixture, 2018. preprint, arXiv:1812.11932.

[55] Qin Li and Jianfeng Lu. An asymptotic preserving method for transport equations with oscillatory scattering coefficients. *Multiscale Model. Simul.*, 15:1694–1718, 2017.

[56] Qin Li, Jianfeng Lu, and Weiran Sun. Half-space kinetic equations with general boundary conditions. *Math. Comp.*, 86:1269–1301, 2017.

[57] Qin Li, Jianfeng Lu, and Weiran Sun. Validity and regularization of classical half-space equations. *J. Stat. Phys.*, 166:398–433, 2017.

[58] Wuchen Li, Jianfeng Lu, and Li Wang. Fisher information regularization schemes for Wasserstein gradient flows, 2019. preprint, arXiv:1907.02152.

[59] Xiantao Li, Lin Lin, and Jianfeng Lu. PEXSI-Σ: A Green's function embedding method for Kohn-Sham density functional theory. *Ann. Math. Sci. Appl.*, 3:441–472, 2018.

[60] Xiantao Li and Jianfeng Lu. Traction boundary conditions for molecular static simulations. *Comput. Methods Appl. Mech. Engrg.*, 308:310–329, 2016.

[61] Xiaoou Li, Jingchen Liu, Jianfeng Lu, and Xiang Zhou. Moderate deviation for random elliptic PDEs with small noise. *Ann. Appl. Prob.*, 28:2781–2813, 2018.

[62] Xingjie Helen Li and Jianfeng Lu. Quasinonlocal coupling of nonlocal diffusions. *SIAM J. Numer. Anal.*, 55:2394–2415, 2017.

[63] Yingzhou Li, Xiuyuan Cheng, and Jianfeng Lu. Butterfly-Net: Optimal function representation based on convolutional neural networks, 2018. preprint, arXiv:1805.07451.

[64] Yingzhou Li and Jianfeng Lu. Bold diagrammatic Monte Carlo in the lens of stochastic iterative methods. *Trans. Math. Appl.*, 3:1–17, 2019.

[65] Yingzhou Li, Jianfeng Lu, and Anqi Mao. Variational training of neural network approximations of solution maps for physical models, 2019. preprint, arXiv:1905.02789.

[66] Yingzhou Li, Jianfeng Lu, and Zhe Wang. Coordinate-wise descent methods for leading eigenvalue problem. *SIAM J. Sci. Comput.*, 41:A2681–A2716, 2019.

[67] Lin Lin and Jianfeng Lu. Decay estimates of discretized Green's functions for Schrödinger type operators. *Sci. China Math.*, 59:1561–1578, 2016. Special issue dedicated to ICIAM 2015.

[68] Ling Lin, Jianfeng Lu, and Eric Vanden-Eijnden. A mathematical theory of optimal milestoning (with a detour via exact milestoning). *Comm. Pure Appl. Math.*, 71:1149–1177, 2018.

[69] Jian-Guo Liu, Jianfeng Lu, Dionisios Margetis, and Jeremy Marzuola. Asymmetry in crystal facet dynamics of homoepitaxy in a continuum model. *Phys. D*, 393:54–67, 2019.

[70] Jianfeng Lu, Yulong Lu, and Jim Nolen. Scaling limit of the Stein variational gradient descent: the mean field regime. *SIAM J. Math. Anal.*, 51:648–671, 2019.

[71] Jianfeng Lu and Felix Otto. Optimal artificial boundary condition for elliptic random media, 2018. preprint, arXiv:1803.09593.

[72] Jianfeng Lu, Matthias Sachs, and Stefan Steinerberger. Quadrature points via heat kernel repulsion. *Constr. Approx.*, in press.

[73] Jianfeng Lu, Christopher D. Sogge, and S. Steinerberger. Approximating poinwise products of Laplacian eigenfunctions. *J. Funct. Anal.*, 277:3271–3282, 2019.

[74] Jianfeng Lu and Konstantinos Spiliopoulos. Analysis of multiscale integrators for multiple attractors and irreversible Langevin samplers. *Multiscale Model. Simul.*, 16:1859–1883, 2018.

[75] Jianfeng Lu and Stefan Steinerberger. A variation on the Donsker-Varadhan inequality for the principal eigenvalue. *Royal Soc. Proceedings A.*, 473:20160877, 2017.

[76] Jianfeng Lu and Stefan Steinerberger. Detecting localized eigenstates of linear operators. *Res. Math. Sci.*, 5:33, 2018.

[77] Jianfeng Lu and Stefan Steinerberger. Optimal trapping for Brownian motion: a nonlinear analogue of the torsion function, 2019. preprint, arXiv:1908.06273.

[78] Jianfeng Lu and Stefan Steinerberger. A dimension-free Hermite-Hadamard inequality via gradient estimates for the torsion function. *Proc. Amer. Math. Soc.*, in press.

[79] Jianfeng Lu and Kyle Thicke. Cubic scaling algorithms for RPA correlation using interpolative separable density fitting. *J. Comput. Phys.*, 351:187–202, 2017.

[80] Jianfeng Lu and Kyle Thicke. Orbital minimization method with ℓ^1 regularization. *J. Comput. Phys.*, 336:87–103, 2017.

[81] Jianfeng Lu and Eric Vanden-Eijnden. Methodological and computational aspects of parallel tempering methods in the infinite swapping limit. *J. Stat. Phys.*, 174:715–733, 2019.

[82] Jianfeng Lu and Zhe Wang. The full configuration interaction quantum Monte Carlo method in the lens of inexact power iteration. *SIAM J. Sci. Comput.*, in press.

[83] Jianfeng Lu, Alex Watson, and Michael I. Weinstein. Dirac operators and domain walls, 2018. preprint, arXiv:1808.01378.

[84] Jianfeng Lu, Benedikt Wirth, and Haizhao Yang. Combining $2d$ synchrosqueezed wave packet transform with optimization for crystal image analysis. *J. Mech. Phys. Solids*, 89:194–210, 2016.

[85] Jianfeng Lu and Haizhao Yang. A cubic scaling algorithm for excited states calculations in particle-particle random phase approximation. *J. Comput. Phys.*, 340:297–308, 2017.

[86] Jianfeng Lu and Haizhao Yang. Preconditioning orbital minimization method for planewave discretization. *Multiscale Model. Simul.*, 15:254–273, 2017.

[87] Jianfeng Lu and Haizhao Yang. Phase space sketching for crystal image analysis based on synchrosqueezed transforms. *SIAM J. Imaging Sci.*, 11:1954–1978, 2018.

[88] Jianfeng Lu and Lexing Ying. Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *J. Comput. Phys.*, 302:329–335, 2015.

[89] Jianfeng Lu and Lexing Ying. Fast algorithm for periodic density fitting for Bloch waves. *Ann. Math. Sci. Appl.*, 1:321–339, 2016.

[90] Jianfeng Lu and Lexing Ying. Sparsifying preconditioner for soliton calculations. *J. Comput. Phys.*, 315:458–466, 2016.

[91] Jianfeng Lu and Zhennan Zhou. Improved sampling and validation of frozen Gaussian approximation with surface hopping algorithm for nonadiabatic dynamics. *J. Chem. Phys.*, 145:124109, 2016.

[92] Jianfeng Lu and Zhennan Zhou. Accelerated sampling by infinite swapping of path integral molecular dynamics with surface hopping. *J. Chem. Phys.*, 148:064110, 2018.

[93] Jianfeng Lu and Zhennan Zhou. Continuum limit and preconditioned Langevin sampling for the path integral molecular dynamics, 2018. preprint, arXiv:1811.10995.

[94] Jianfeng Lu and Zhennan Zhou. Frozen Gaussian approximation with surface hopping for mixed quantum-classical dynamics: A mathematical justification of fewest switches surface hopping algorithms. *Math. Comp.*, 87:2189–2232, 2018.

[95] Y. Lu, X. Huo, and P. Tsiotras. Beamlet-based graph structure for path planning using multiscale information. *IEEE Trans. Automatic Control*, 57(5):1166–1178, 2012.

[96] Yulong Lu, Jianfeng Lu, and James Nolen. Accelerating Langevin sampling with birth-death, 2019. preprint, arXiv:1905.09863.

[97] Anton Martinsson, Jianfeng Lu, Benedict Leimkuhler, and Eric Vanden-Eijnden. Simulated tempering method in the infinite switching limit with adaptive weight learning. *J. Stat. Mech.*, 2019:013207, 2019.

[98] Christian Mendl, Jianfeng Lu, and Jani Lukkarinen. Thermalization of particle chains with onsite anharmonicity and comparison with kinetic theory. *Phys. Rev. E*, 94:062104, 2016.

[99] Akihiko Nishimura, David Dunson, and Jianfeng Lu. Discontinuous Hamiltonian Monte Carlo for sampling discrete parameters. *Biometrika*, in press.

[100] Xiaohua Niu, Tao Luo, Jianfeng Lu, and Yang Xiang. Dislocation climb models from atomistic scheme to dislocation dynamics. *J. Mech. Phys. Solids*, 99:242–258, 2017.

[101] Deborshee Sen, Matthias Sachs, Jianfeng Lu, and David Dunson. Efficient posterior sampling for high-dimensional imbalanced logistic regression, 2019. preprint, arXiv:1905.11232.

[102] Kyle Thicke, Alex Watson, and Jianfeng Lu. Computing edge states without hard truncation, 2018. preprint, arXiv:1810.07082.

[103] C. Wang and X. Huo. Object tracking under low signal-to-noise-ratio with the instantaneous-possible-moving-position model. *Signal Processing*, 93(5):1044–1055, May 2013.

[104] Huizhu Wang, Seong-Hee Kim, Xiaoming Huo, Youngmi Hur, and James Wilson. Monitoring nonlinear profiles adaptively with a wavelet-based distribution-free CUSUM chart. *International Journal of Production Research*, 53(15):4648–4667, August 2015.

[105] Zhe Wang, Yingzhou Li, and Jianfeng Lu. Cooridnate descent full configuration interaction. *J. Chem. Theory Comput.*, 15:3558–3569, 2019.

[106] Alexander Watson, Jianfeng Lu, and Michael I. Weinstein. Wavepackets in inhomogeneous periodic media: Effective particle-field dynamics and Berry curvature. *J. Math. Phys.*, 58:021503, 2017.

[107] H. Xu, D. Luo, X. Huo, and X. Yang. World expo problem and its mixed integer programming based solution. In *Workshop on Behavior and Social Informatics (BSIUCBCN2013), in conjunction with the 2013 Pacific-Asia Conference on Data Mining and Knowledge Discovery (PAKDD2013)*, Gold Coast, Australia, April 14 2013. acceptance rate 44%: 16 out of 36.

[108] Zhouwang Yang, Huizhi Xie, and Xiaoming Huo. *Perspectives on Big Data Analysis Contemporary Mathematics*, volume 622, chapter Data-driven smoothing can preserve good asymptotic properties, pages 125–139. American Mathematical Society, Providence, RI, 2014.

[109] Zhiyi You, Liying Li, Jianfeng Lu, and Hao Ge. Integrated tempering enhanced sampling method as the infinite switching limit of simulated tempering. *J. Chem. Phys.*, 149:084114, 2018.

[110] Tang-Qing Yu, Jianfeng Lu, Cameron F. Abrams, and Eric Vanden-Eijnden. A multiscale implementation of infinite-swap replica exchange molecular dynamics. *Proc. Natl. Acad. Sci. USA*, 113:11744–11749, 2016.

[111] Victor Yu, Fabiano Corsetti, Alberto Garcia, William Huhn, Mathias Jacquelin, Weile Jia, Bjorn Lange, Lin Lin, Jianfeng Lu, Wenhui Mi, Ali Seifitokaldani, Alvaro Vazquez-Mayagoitia, Chao Yang, Haizhao Yang, and Volker Blum. ELSI: A unified software interface for Kohn-Sham electronic structure solvers. *Comput. Phys. Commun.*, 222:267–285, 2018.

[112] Yuanyuan Zhang, Renfu Li, Dinggen Li, Yang Hu, and Xiaoming Huo. Stabilization of the stochastic jump diffusion systems by state-feedback control. *Journal of the Franklin Institute*, 351(3):1596–1614, March 2014.

[113] Wei Zhu, Qiang Qiu, Bao Wang, Jianfeng Lu, Guillermo Sapiro, and Ingrid Daubechies. Stop memorizing: A data-dependent regularization framework for intrinsic pattern learning. *SIAM J. Math. Data Sci.*, in press.