

Collaboration and Evaluation Plan

Rayan : Yoav, I edited quite a bit, hope this is OK

Our team consists of University of California San Diego (UCSD) faculty members from the departments of Computer Science (Yoav Freund), Electrical Engineering (Piya Pal, Peter Gerstoft), and Mathematics (Alex Cloninger, Rayan Saab), directly representing three of the four communities of HDR TRIPODS. Additionally, all members of the team are founding members of UCSD’s Halicioglu Data Science Institute (HDSI), a new academic unit at UCSD with a focused mission. Its mission, one that it shares to a large extent with HDR TRIPODS, is to lay the groundwork for the scientific foundations of this emerging discipline, develop new methods and infrastructure, and train students, faculty and industrial partners to use data science in ways that will allow them to solve some of the world’s most pressing problems [?].

Having all our team members located on the same campus and regularly interacting through HDSI will greatly enhance our ability to collaborate closely on the proposed projects. In addition to the direct and frequent interaction of the faculty members involved in this proposal, with the NSF’s support, we intend to jointly supervise PhD students and to co-mentor a postdoctoral fellow.

Additionally, we note that our team includes theoreticians who also have a deep interest in and knowledge of applications, and all **Rayan :** I put this in, but is this true? of us have worked closely and published papers with practitioners. In the context of this proposal, Co-PI Gerstoft’s intimate knowledge of sensors and sensor networks will allow us to account for practical issues that arise in data science, including (among others) its multi-modality and incompleteness. Our awareness of practical issues will hopefully allow us to maximize the impact of the algorithms and methodologies that result from our work. **Rayan :** Peter, please edit!

In what follows, we will describe our expertise, plans for advising joint students, for collaborating, as well as for evaluating the results of our work.

1 Expertise

- **PI Freund’s** expertise is in Computational Learning Theory and Computer Science. Among his theoretical work are Boosting [], online learning [], learning and game theory [] and the RP-tree algorithm []. He has also worked on applications of machine learning to image analysis for biology []. **Rayan :** Yoav, do you want to throw in that you have published in stats journals, as a nod to the missing discipline on our team?
- **Co-PI Cloninger’s** expertise... **Yoav :** please add yourself
- **Co-PI Gerstoft’s** expertise... **Yoav :** please add yourself
- **Co-PI Pal’s** expertise... **Yoav :** please add yourself
- **Co-PI Saab’s** expertise is in applied and computational harmonic analysis, and in mathematical signal processing. He is interested in, and has published extensively on questions regarding efficient acquisition, quantization, representation, and processing of data. He has studied these problems in classical contexts, such as those of band limited functions, but also in modern contexts, such as compressed sensing of structured signals, be they sparse vectors, low-rank matrices, or signals from arbitrary sets. He has also worked on signal and data processing problems related to multiple sensors, including the blind source separation problem. In his work, Saab uses and develops a variety of tools from high dimensional probability theory, applied harmonic analysis, convex analysis and optimization, and quantization theory, among others.

2 Collaboration Plan

Joint Students

Rayan : Shouldn’t we mention the postdoc here or elsewhere? In addition to its focus on the proposed scientific content, a main goal of this proposal is to recruit and train highly qualified personnel for the workforce of tomorrow. Exploiting the fact that our team members are all UCSD faculty with affiliations to HDSI, we will recruit and advise graduate and post-graduate students as a group. We will actively recruit students who have a strong mathematical background and are able and willing to implement algorithms as reusable code. Equally importantly, we will actively seek out students from underrepresented groups in STEM.

Each student will have two advisors, from two of the three disciplines, who will work closely with the student.

From a training point of view, this will help the student gain a broader perspective on data science as a field while acquiring an in-depth understanding of tools and methods from more than one area. From a collaboration point of view, joint advising will provide one of the mechanisms by which our team members will work together.

Collaboration

Rayan : *see if you like these edits.* Each graduate student will have a one hour meeting with each of their advisors each week. The student and both advisors will also have regular bi-weekly meetings to assess progress and ensure continuing synergy on each project. In addition, we will have a one and a half hour weekly group meeting, where each meeting will include all PIs and students, and where one of the students or PIs will present to the others.

Rayan : *not sure about this one, get rid of it if you don't like it* Finally, we will have annual retreats where all the team members meet to assess our progress on the proposed work...

Evaluation

Rayan : *not sure what they want here?* We plan to submit paper to the following conferences and journals: COLT, ICML, AISTAT, JMLR, IEEE Signal Processing, IEEE Information theory, Applied and Computational Harmonic Analysis, Foundations of Computational Mathematics, Information and Inference, SIAM Journal on Mathematics of Data Science, please add.

We will implement our algorithms in python and jupyter notebooks and make them available through GitHub.

We will use several benchmarks to evaluate the accuracy. **Yoav :** *Piya, Peter, can you suggest benchmark data sets?*

We will collect audio data in Prof. Christiansen's instrumented living quarter and evaluate our algorithms on the collected data. **Rayan :** *do we need a collaboration letter from Prof. Christiansen?*

We will evaluate our performance in terms of:

1. Number of publications and quality of publication venues.
2. Number of (open source) implementations.
3. Performance of our methods on benchmark data.

Institute for Learning Sensor Networks

Yoav : *Please suggest names for the institute*

In phase II we plan to

We plan to join forces with researchers in the areas of embedded computers, robotics and medical sensing. All PIs are members of the Haligluou Data Science Institute (HDSI).

Rayan : Below are "Additional Solicitation Specific Review Criteria" from the call that I attempted to address above - need help on this!:

- Does the proposal describe a well-integrated research and training program focused on the theoretical foundations of data science and fostering collaboration and interaction among the targeted (at least three of the four) communities of HDR TRIPODS? **Rayan :** *not sure about the well-integrated, but I think we hit the second point*
- Does the proposal address the "broad themes of the program" listed in the Program Description? **Rayan :** *I tried to hit a few of these points here, but we NEED to go back and make edits to the proposal to make sure we have decent coverage.*
- Does the proposal address strategies for workforce development, including but not limited to novel educational and training activities and efforts toward full participation of groups underrepresented in science, technology, engineering and mathematics (STEM)? **Rayan :** *We should make explicit mention of these things in the Broader Impacts section. For example, I am currently advising two undergrads on a project with industry (through HDSI). We need to capitalize on such things more, particularly in the context of underrepresented groups.*
- Transdisciplinarity/Synergy: Is the project transdisciplinary, bringing together theories and approaches from at least three of electrical engineering, mathematics, statistics, and theoretical computer science? Is there synergy between the different groups? **Rayan :** *I think this is fine*

- Vision: Is there a strong case for the ability to identify and articulate a vision for the foundations of data science? **Rayan : good question**
- Quality and Value of Collaboration: Is the expertise of the PIs complementary and well-suited to the research and training programs developed in this project? Are the specific roles of each collaborating investigator clear? Does the collective team have expertise representing at least three of the four communities (electrical engineering, mathematics, statistics, and theoretical computer science)? **Rayan : fine?**
- Is there a well-developed plan for collaboration and interaction with the domain sciences and industry? **Rayan : help?**
- Does the proposal provide a clear plan and rationale for an investment of the size proposed, including clear plans to develop capacity for potential future Phase II operations? **Rayan : ?**
- Does the Collaboration and Evaluation Plan identify clear measures of success, both for Phase I operations and development of capability for a potential Phase II, along with a plan to evaluate the project with respect to those measures by gathering quantitative and qualitative data? **Rayan : ?**
- Does the Collaboration and Evaluation Plan provide a clear plan for thoughtful, ongoing assessment of all Institute activities? How will the assessment be used to inform and improve both daily Institute operations and long-range planning, aiming toward a successful Phase II Institute? **Rayan : ?**