



Graduate School Event

Thesis Defense: Compression and Inference in High Dimensions: An Approximate Message Passing Perspective

Kevin Kögler (Mondelli Group)

Mondelli Group

Host: Krzysztof Pietrzak

Modern datasets often contain as many variables as observations, or many more, making classical statistical tools difficult to apply and standard machine-learning models hard to understand. To address this problem, this thesis develops methods for inference and compression in this high-dimensional regime, centered on approximate message passing (AMP), a family of iterative inference algorithms which are conjectured to be optimal for many of these problems. First, we introduce rotationally invariant generalized AMP (RI-GAMP), extending AMP beyond independent Gaussian design matrices to a broad class of matrices with complex correlations and arbitrary limiting spectra. We characterize its performance rigorously through state evolution and show that it approaches the performance of Vector AMP while avoiding its computationally expensive singular-value decomposition. We then study nonlinear autoencoders when the input and compressed dimensions grow proportionally. For Gaussian data, we derive fundamental limits for one-bit compression by shallow two-layer autoencoders, characterize their optimal solutions, and prove that gradient methods reach them. For structured data, however, a shallow linear decoder may ignore information such as sparsity and behave exactly as it does for Gaussian inputs. Adding nonlinearities or depth overcomes this limitation: deeper decoders can learn computations closely related to AMP and approach optimal reconstruction for a fixed encoder. Finally, we apply these ideas to predict plasma-protein abundance from UK Biobank individual-level genetic data, using approximately 8.4 million genetic variants from roughly 40,000 individuals. A two-stage genetic Vector AMP pipeline first selects a small set of predictive genomic regions and then fits more detailed linear or tree-based models. Combined with a one-hot representation of the genotype data, this procedure closes the previously reported average gap between linear and deep-learning prediction and improves prediction for many proteins in a proteome-wide comparison.

Monday, August 17, 2026 04:00pm - 05:00pm

Office Bldg West / Ground floor / Heinzl Seminar Room (I21.EG.101) and Zoom



This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station.
Please find a schedule of the ISTA Shuttle on our webpage:
<https://ista.ac.at/en/campus/how-to-get-here/> The ISTA Shuttle bus is marked ISTA Shuttle
(#142) and has the Institute Logo printed on the side.

www.ista.ac.at | Institute of Science and Technology Austria | Am Campus 1 | 3400 Klosterneuburg