Meeting 14

Duration: 75 Minutes. Present: Rui (First 60 minutes), Alex, and Martin.

**Thesis**

Wrote a large part of introduction and problem setting. Will send to Rui and Alex to have a look, but of course not a full version of anything.

**OMP**

Discussed about the OMP rewriting, results are good. However, the OMP algorithm does not enforce DAGness. When the data is generated according to a VAR(1) with DAG, then we do not have to worry.

However, another question then is, when we have a sparse DAG, when do we stop? We can continue adding edges until we violate the DAG assumption; then we stop.

Alternatively, we can continue adding edges. When we see that we are about to add an edge that violates the DAGness, we discard this edge, and continue. We fix this constant to zero. We continue until we have a dense DAG (or similarly, until we have visited all nodes). This is a more principled approach, as we now also have an *ordering of importance* of the edges in our DAG. Then, when we want to prune our DAG, we can have a more principled approach than simply removing the smallest non-zero edge weight. Removing the smallest non-zero edge weight is not a principled approach, as a small edge weight does not imply a small importance. Using this OMP approach, we have a principled way of ordering the edges by importance. We can then prune by removing the edges from least important to most important.

However, when do we stop pruning? This is a difficult choice. We can e.g. prune all edges who do not significantly improve the model fit (e.g. BIC / AIC). We can e.g. prune all edges that do not contribute enough to the loss value (e.g. by “curvature” in the loss function). The less you prune, the more predictive your model is.

**NOTEARS**

NOTEARS was investigated a bit more. Proposition 3 of the papers was quite meaningless. No scale invariance, see notebooks of Alex.

Also interesting: Stationary points of h(W) where h(W) =/= 0. Furthermore, plots of h(W) as a function of w\_21 and w\_12.

**Benchmarks**

We discussed that it is good to have some *concrete set of benchmarks.* E.g., fix *p = 10*, and vary some other parameters, such as the number of samples *T*, the number of edges *s*. Get some good comparisons of all methods; OMP, NOTEARS, LASSO, OLS.

Ideal: Find instances where each of the methods is not that good, to see the pros and cons.

**Causal or Predictive**

2D-Example where one time series had a much larger range, such that we were unable to detect the underlying *generating* matrix. However, we found a *better* matrix, in the sense that it had a higher predictive power. This showcases that our models are focused on getting the best *predictive* model under our constraints, and we do not necessarily focus on the true generating graph.

