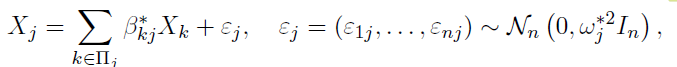
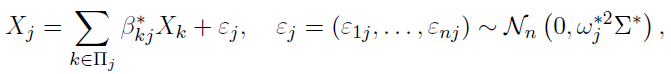
Learning Gaussian DAGs from network data

Learning the structure of a DAG. A new method is proposed for **dependent data**, where we assume the observations are correlated. A method is developed to estimate the DAG structure given a topological ordering of the nodes.

We assume the data matrix is as follows (1):



The goal is to estimate B\*. (1) Assumes the rows of X are jointly **independent**. This is not realistic, however, so we need to develop a method that can take dependencies into account: a method that simultaneously **infers B\*** and the **sample dependencies**. Additionally, we want the method to learn a **sparse** DAG. This is the model (2):

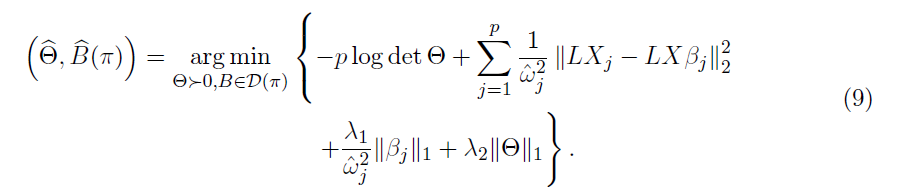


We need to infer B\*, Omega\*, and Sigma\*. This allows the epsilonij to be dependent among i. Then X satisfies



With Psi\* = (I – B\*)-T Omega\* (I – B\*)-1. To promote sparsity, constraints on B are imposed.

# Method

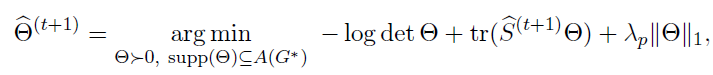
The estimator is the minimizer of a score function

Where lambda1 and lambda2 are regularizers and the first part is the log likelihood given a toplogical ordering pi of the vertices. The algorithm:

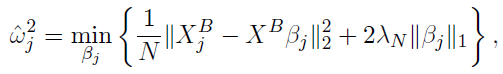
1. Estimate Omega.
2. Estimate B and Theta = Sigma-1 by minimizing a biconvex score function.

Lambda1 promotes sparsity on B^\* and prevents overfitting on variables with small variance. Lambda2 promotes uniqueness and accuracy of Theta.

Given Omega, B and Theta are estimated by



Omega is estimated by assuming an N block structure on Theta.



If there is no known ordering, then we can pick a random permutation. The estimated Theta\* can be used to **decorrelate** X and make the data less dependent, thereby greatly improving results of methods that require independent data.

# Theoretical Results

Error bounds and proofs of consistency are given for B\* Omega\* and Theta\* when assuming a true topological ordering.

# Numerical Experiments

BCD (Block coordinate descent) achieves better performance on simulated and real data for the four different covariance structures analyzed.

BCD also helped in significantly increasing the log likelihood of the data when using BCD to decorrelate the data.