DAG learning with R package `clrdag`

Chunlin Li

School of Statistics, University of Minnesota

April 29th, 2019

DAG learning

Structural equations model:

$$X_k = \sum_{j \in \mathrm{pa}_k} A_{jk} X_j + \varepsilon_k, \quad k = 1, \ldots, p,$$

where $A_{jk} \neq 0$ iff $j \in \mathrm{pa}_k$ and $\varepsilon_k \overset{iid}{\sim} N(0, \sigma^2)$.

The iid errors ε_k implies $A_{p \times p}$ is a (weighted) adjacency matrix of a directed acyclic graph. Denote $A \in \mathbb{D}$.

Assume $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$; $i = 1, \ldots, n$ are iid observations. Let $\mathbb{X}_{n \times p} = (\mathbf{X}_1', \ldots, \mathbf{X}_n')'$.

Goal Estimate and make inference on A.

Likelihood: original

Constrained MLE: difficult!

$$\max_{A \in \mathbb{D}} \quad \ell(A) = \frac{1}{2}\|\mathbb{X} - \mathbb{X}A\|_F^2 + \mu\|A\|_{off,0}.$$

Relaxation?

Likelihood: relaxation

Iterative convex relaxation (DC)[1]:

At $(t+1)$th step, $B^{[t]} = H_\tau(A^{[t]})$; H_τ is hard threshold.

$$(A^{[t+1]}, \Lambda^{[t+1]}) = \arg\max \frac{1}{2}\|\mathbb{X} - \mathbb{X}A\|_F^2 + \mu\tau^{-1}\|B^{[t]} \circ A\|_{off,1}$$

$$\text{s.t. } \lambda_{jk} + \mathbb{I}(i \neq k) - \lambda_{ik} \geq \tau^{-1}|A_{ij}|B_{ij}^{[t]} + (1 - B_{ij}^{[t]}),$$

$$i, j, k = 1, \ldots, p; \quad i \neq j.$$

Iterate until objective converges.

[1]The details are included in the paper and R package vignette.

Computation

Our implementation is based ADMM.

Algorithm Initiate $(A^{[0]}, \Lambda^{[0]})$ satisfying DAG constraint. Set $B^{[0]} = H_\tau(A^{[0]})$.

Set the optimization accuracy $\epsilon > 0$, for $t = 1, \ldots, \infty$

1. Compute $(A^{[t]}, \Lambda^{[t]})$ by ADMM. Set $B^{[t]} = H_\tau(A^{[t]})$.
2. If $B^{[t]}$ has a cycle, for each $|A_{ij}^{[t]}| > 0$ in increasing order, if (i, j) is in a cycle,

$$A_{ij}^{[t]} \leftarrow 0, \quad B_{ij}^{[t]} \leftarrow 0.$$

`clrdag`

C++ and R (`Rcpp`) with OpenMP:

- Friendly API.
- Highly optimized linear algebra libraries (‘Armadillo‘).

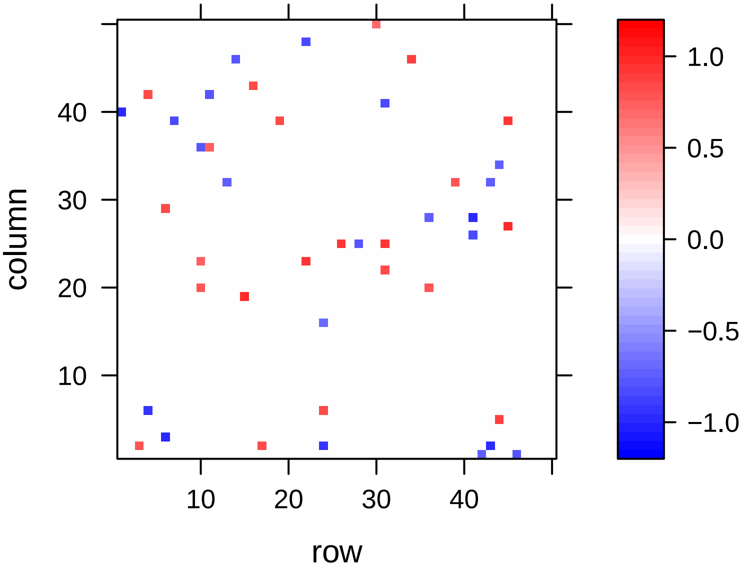
Do one thing and do it well: `MLEdag`

- Compute MLE for DAG.
- Make likelihood inference on DAG edges.

A toy example: setup

- $p = 50, n = 1000$
- Random DAG A with connection probability $1/p$
- Nonzero A_{ij} are assigned $\mathrm{Unif}[-1,-0.7] \cup [0.7,1]$
- $\varepsilon_k \overset{iid}{\sim} N(0,1)$

A toy example: $\|A\|_0 = 43$



A toy example: code

```r
library(clrdag)
t <- proc.time()
out <- MLEdag(X=X,tau=0.3,mu=1,rho=1.2)
```

```
## DC iteration: 0, objective value: 26.8347
## DC iteration: 1, objective value: 25.2819
## DC iteration: 2, objective value: 25.0798
```

```r
proc.time() - t
```

```
##    user  system elapsed
##  59.947   0.940   5.314
```

A toy example: result

$\|\hat{A} - A\|_F^2$:

```r
sum((out$A - A)^2)
```

```
## [1] 0.0257755
```

Hamming$(\mathcal{G}_{\hat{A}}, \mathcal{G}_A)$:

```r
sum(abs((out$A != 0) - (A!=0)))
```

```
## [1] 0
```

Summary

URL: `https://github.umn.edu/li000007/clrdag`

Due to time limit, the inference part is not presented.

TODOs:

- Cross-validation function, diagnostic plots.
- Publish on CRAN.

Thank you

References

- Li, C., Shen, X., & Pan, W. (2019). Likelihood ratio tests of a large directed acyclic graph. Submitted.
- Yuan, Y., Shen, X., Pan, W., & Wang, Z. (2018). Constrained likelihood for reconstructing a directed acyclic Gaussian graph. *Biometrika*, 106(1), 109-125.