Simulation Results (For 2 steps RLS)

The running time is around 13.5 hrs (for 50 replicates) [however, it seems not stable running it on the server, might even took 18 hrs sometimes]

I sent you some the .csv files of the results.

* Pred\_err: The accumulative prediction error from 10 folds cross validation.

Increase the range of lambda, after 2.7

Include RRR

* Err: The prediction error using 100 samples to train and have another 100 samples as our validation set. (get rid this one)
* Cong\_iter: the number of iterations it takes to converge using the best tuning lambda choosing from cross validation (report, maybe not)
* Rank\_est: the estimation of rank using the best tuning lambda choosing from cross validation

Interesting

Include Estimation err of B, XB

Estimation T (correlation between true T and estimated T)

Questions about the report:

1. What is the motivation of assuming B is a low rank matrix? Also, a little bit motivation of T in our model.

We might verify B is low rank matrix. Uncertain.

B assume to be low rank, association profile. Why is that interesting? Functional redundant, we are trying to simulate this bio circumstance.

3-step RLS vs 3-step LS, without centering T, it does not make sense, not identifiable. Penalty is on B, not T.

Centering T does make everything identifiable, resolve the issue.

Even it is not true, it does affect the thing we want to estimate, B and XB, it helps address the problem.

1. What simulation we should put in the report?

50 replicates, RRR, 2LS, 3RLS, 2RLS.

Multiply setting:

Not a lot variation about T/variation (high/ low) [how the new method vs RRR]

Signal ratio, increase(high/low)

Noise, increase the variance/or include the correlation(ind/cor1/cor2)

1. Is there any type of graphical you want to suggest displaying some results?

How well the method estimate B and XB

1. Should I explain a little about how adaptive nuclear norm algorithm work? (like using the soft-threshold of singular value decomposition)

In the method part, include a little bit about the background of this method.

1. Will do the 2 steps LS method and 3 steps RLS method as well, should I compare to the original RRR method?

Will compare to it

1. To use RLS, one purpose is to solve the identifiability, but why we set the sum of all the entries from T to 0 again?

It not necessary to be 0, we just to pick a number to solve it.

1. Other particular simulation results you want me to get?

Report: microbiome, data integration, reduced-rank regression

Gets RRR and other two methods going first.

Other setting about T

Writing