ℓ_2^2-learning of Gaussian graphical models – miscellanea

Wessel N. van Wieringen

Dept. of Epidemiology & Data Science, Amsterdam UMC
Dept. of Mathematics, Vrije Universiteit Amsterdam
Amsterdam, the Netherlands

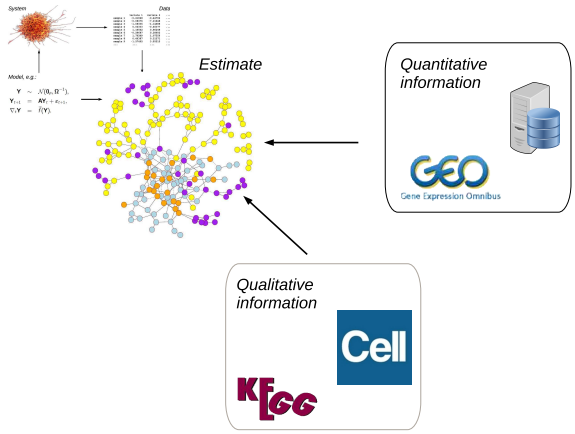
IBC pre-conference course, Riga, 10.07.2022

Network reconstruction

Network
- *CIG* : graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of edges $\mathcal{E}$ that represent CIs among variates $\mathcal{V} = \{1, \ldots, p\}$.
- *GGM* : $\mathsf{Y} \sim \mathcal{N}(0_p, \Omega^{-1})$ with precision matrix Ω.
- *CIs* : $(\Omega)_{1,2} = 0 \iff Y_1 \perp\!\!\!\perp Y_2 \,|\, Y_3, \ldots, Y_p$.

Reconstruction
- Data : $\mathsf{Y}_1, \ldots, \mathsf{Y}_n \sim_{i.i.d.} \mathcal{N}(0_p, \Omega^{-1})$.
- Estimation : $\widehat{\Omega}_{\text{ML}} = \mathsf{S}^{-1}$ with $\mathsf{S} = n^{-1} \sum_{i=1}^n \mathsf{Y}_i \mathsf{Y}_i^\top$.
- Inference : obtain $\hat{\rho}(Y_1, Y_2 \,|\, Y_3, \ldots, Y_p)$ from $\widehat{\Omega}_{\text{ML}}$ and evaluate its likeliness under H_0.

Aim



Problem

The generalized ridge estimator of Ω, combining
i) non-zero target shrinkage, and
ii) element-wise penalization,
is:

$$\widehat{\Omega}(\Lambda, \mathsf{T}) \quad = \quad \arg\max_{\Omega \succ 0} \log(|\Omega|) - \text{tr}(\mathsf{S}\Omega) - \|\sqrt{\Lambda} \circ (\Omega - \mathsf{T})\|_F^2,$$

with
a) target $\mathsf{T} \in \mathcal{S}_+^p$,
b) positive and symmetric penalty matrix Λ, and
c) $\sqrt{\mathsf{B}}$ the Hadamard square root.

If $\Lambda = \lambda 1_{pp}$, we obtain the regular ridge precision estimator.

In general, no analytic expression available. The estimator is evaluated by row-by-row updating.

Illustration

Context:
- Early detection is the best strategy against cancer.
- More information on primary than metastasized cancers.
- Primary cancer is precursor stage to metastasis.

Lung cancer:
- TCGA study,
- 111 cancers / 87 metastases,
- Toll-like receptor signalling pathway of 87 genes,
- Expression and copy number.

Regular ridge estimate of the metastasis precision is identical to the cancer precision estimate, i.e. the target is perfect.

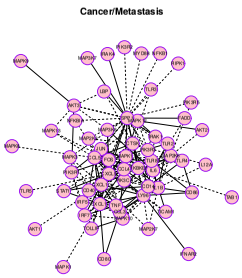


Illustration: structure of Λ

Incorporate molecular biology assumptions in penalty matrix:

$$\Lambda \quad = \quad \begin{pmatrix} 0 & \lambda_{\text{ge}} & \ldots & \lambda_{\text{ge}} & \lambda_{\text{cn}} & \infty & \ldots & \infty \\ \lambda_{\text{ge}} & 0 & \ddots & \lambda_{\text{ge}} & \infty & \lambda_{\text{cn}} & \ddots & \infty \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \lambda_{\text{ge}} & \lambda_{\text{ge}} & \ldots & 0 & \infty & \infty & \ldots & \lambda_{\text{cn}} \\ \lambda_{\text{cn}} & \infty & \ldots & \infty & 0 & \infty & \ldots & \infty \\ \infty & \lambda_{\text{cn}} & \ddots & \infty & \infty & 0 & \ddots & \infty \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \infty & \infty & \ldots & \lambda_{\text{cn}} & \infty & \infty & \ldots & 0 \end{pmatrix}$$

Choose λ_{cn} and λ_{ge} by 5-fold cross-validation.

Illustration: comparison

Cross-validated optimal penalty parameters:

- cancer : $(\lambda^{(c)}_{\text{ge, opt}}, \lambda^{(c)}_{\text{cn, opt}}) = (0.001495, 0.000752)$,
- metastasis : $(\lambda^{(m)}_{\text{ge, opt}}, \lambda^{(m)}_{\text{cn, opt}}) = (0.013933, 14.76048)$.

Difference in λ's due to:

- Sample size,
- DNA-RNA interaction more stable than within expression level,
- Target indeed informative?

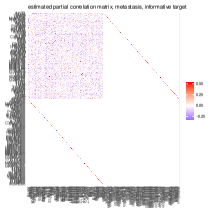
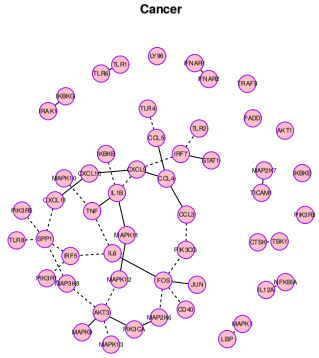


Illustration: results



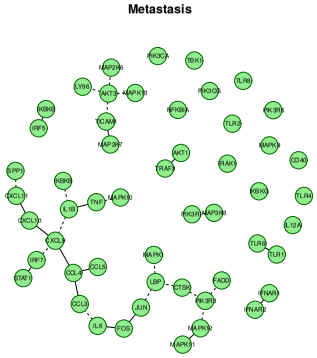
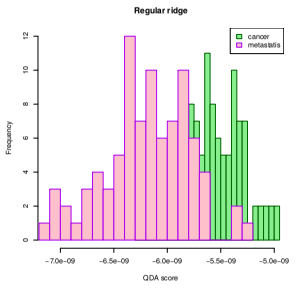
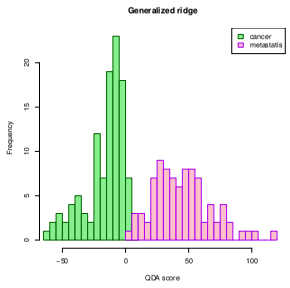


Illustration: results

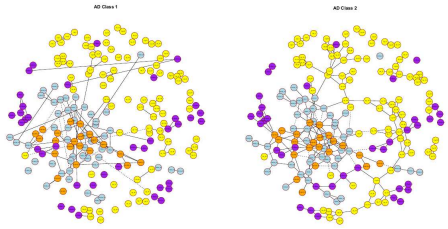
Quadratic discriminant analysis (QDA) scores with Ω estimates:





The Alzheimer metabolite study revisted

Result from fused ridge estimation:



Take-away : compound-type matters.

Sample size : 40 (# AD class 1) vs. 87 (# AD class 2).
Thought : can $\widehat{\Omega}_{\text{AD class 2}}$ inform $\widehat{\Omega}_{\text{AD class 1}}$?

The Alzheimer metabolite study revisted

A re-analysis

```
  # activate packages
library("rags2ridges")
library("porridge")

# load, extract and scale data for AD Class 2
data("ADdata")
ADclass1 <- scale(t(ADmetabolites[,
                   sampleInfo$ApoEClass=="Class 1"]))
ADclass2 <- scale(t(ADmetabolites[,
                   sampleInfo$ApoEClass=="Class 2"]))

# precision matrix estimation with LOOCV
Phat2 <- optPenalty.kCVauto(ADclass2, ...)$optPrec
```

One could sparsify $\widehat{\Omega}_{\text{AD class 2}}$.

The Alzheimer metabolite study revisted

Find optimal group-specific penalty parameters:

```
# make group indicator
groups <- colnames(ADclass1)
groups[grep("Amine", colnames(ADclass1))] <- 1
...
groups <- as.integer(groups)

# find optimal lambdas
optLambdas <- optPenaltyPgen.kCVauto.groups(ADclass1,
                                target       =Phat2,
                                groups       =groups,
                                penalize.diag=FALSE,
                                ...)
```

Warning: takes forever. Spoiler provided later.

The Alzheimer metabolite study revisted

Have a look at the penalty parameter matrix.

```
# construct generalized penalty matrix
lambdaMat <- c(rep(optLambdas[1], sum(groups==1)),
               ...)
lambdaMat <- matrix(lambdaMat,
                    ncol=length(groups),
                    nrow=length(groups))
lambdaMat <- (lambdaMat + t(lambdaMat))/2
diag(lambdaMat) <- 0

# plot generalized penalty matrix
edgeHeat(lambdaMat)
```

The Alzheimer metabolite study revisted

Find $\widehat{\Omega}_{\texttt{AD class 1}}$, sparsify, and plot CIG.

```
# fit precision matrix
Phat1 <- ridgePgen(covML(ADclass1),
                   lambda=lambdaMat,
                   target=Phat2)

# extract the CIG
Phat1 <- sparsify(Phat1, ...)

# visualize the CIG
PcorP1 <- pruneMatrix(Phat1$sparseParCor)
Colors <- rownames(PcorP1)
Colors[grep("Amine", rownames(PcorP1))] <- "lightblue"
...

# plot network
Ugraph(PcorP1, ...)
```

The Alzheimer metabolite study revisted

Q: plot penalty matrix.
$(\lambda^{(\text{opt})}_{\text{amine}}, \lambda^{(\text{opt})}_{\text{org.acid}}, \lambda^{(\text{opt})}_{\text{lipid}}, \lambda^{(\text{opt})}_{\text{ox.stress}}) = (1.7095, 4.4945, 22.1870, 12.3487)$

Q: fit a GGM to the `AD class 1` data by means of the generalized ridge estimate.

Q: sparsify the estimated precision.

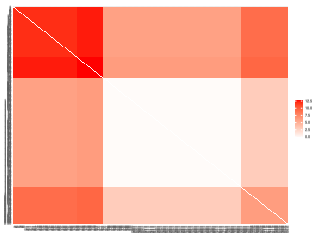
Q: plot the CIG.

Q: compare to the CIGs of the two classes.

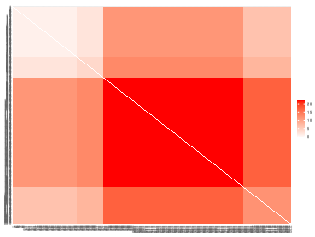
Q: also use a zero target, i.e. $\mathsf{T} = 0_{pp}$.
$(\lambda^{(\text{opt})}_{\text{amine}}, \lambda^{(\text{opt})}_{\text{org.acid}}, \lambda^{(\text{opt})}_{\text{lipid}}, \lambda^{(\text{opt})}_{\text{ox.stress}}) = (11.9672, 12.6888, 0.2882, 6.4313)$

The Alzheimer metabolite study revisted

Penalty matrix, $\mathsf{T} = 0_{pp}$

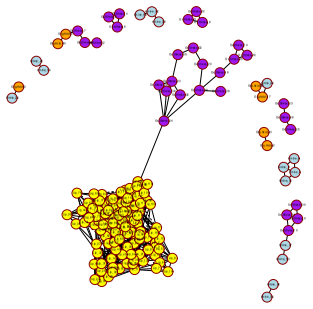


Penalty matrix, $\mathsf{T} = \widehat{\Omega}_{\texttt{AD class 2}}$.

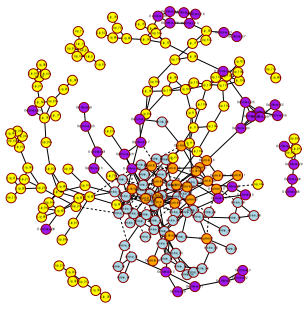


The Alzheimer metabolite study revisted

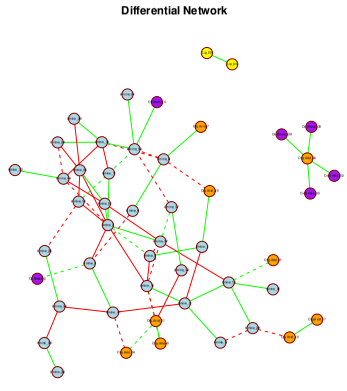
CIG, $\mathsf{T} = 0_{pp}$



CIG, $\mathsf{T} = \widehat{\Omega}_{\texttt{AD class 2}}$.



The Alzheimer metabolite study revisted



Differences appear to be among the amines (the lightblue folks).

Are reconstructed molecular networks reproducible?

Context

Network reconstruction only considers sampling variation, while:

"An element of chance enters into every measurement; hence every set of measurements is inherently a sample of certain more or less unknown conditions. Even in those few instances where we believe that the objective reality being measured is constant, the measurements of this constant are influenced by chance or unknown causes."

– Shewhart, 1931, p. 378.

Additive error model

Replicated data of i-th sample: $\{\mathsf{Y}_{i,k_i}\}_{k_i=1}^{K_i}$.

A 'zignal + noise'-model:

$$\begin{array}{lll} \mathsf{Y}_{i,k_i} & = & \mathsf{Z}_i + \boldsymbol{\varepsilon}_{i,k_i} \\ \mathsf{Z}_i & \sim_{i.i.d.} & \mathcal{N}(0_p, \Omega_z^{-1}), \\ \boldsymbol{\varepsilon}_{i,k_i} & \sim_{i.i.d.} & \mathcal{N}(0_p, \Omega_\varepsilon^{-1}), \end{array}$$

and $\mathsf{Z}_{i'} \perp\!\!\!\perp \boldsymbol{\varepsilon}_{i,k_i}$.

Hence,

$$\mathsf{Y}_{i,k_i} \sim_{i.i.d.} \mathcal{N}(0_p, \Omega_z^{-1} + \Omega_\varepsilon^{-1}).$$

Effect of noise

Ignorance of the noise introduces false positive/negative edges.
For instance, take $\Omega_\varepsilon^{-1} = \mathsf{I}_{33}$ and

$$\Omega_z^{-1} = \begin{pmatrix} 3 & -1 & 2 \\ -1 & 3 & -2 \\ 2 & -2 & 4 \end{pmatrix}.$$

Then, $(\Omega_z)_{1,2} = 0$ but $[(\Omega_z^{-1} + \Omega_\varepsilon^{-1})^{-1}]_{1,2} \neq 0$.

Generally,

$$\Omega_y = (\Omega_z^{-1} + \Omega_\varepsilon^{-1})^{-1} = \Omega_z - \underbrace{(\mathsf{I}_{pp} + \Omega_z\Omega_\varepsilon^{-1})^{-1}\Omega_z\Omega_\varepsilon^{-1}\Omega_z}_{\propto \text{ edge strengths diff. between } \Omega_y \text{ and } \Omega_z}.$$

Estimation

Penalized maximum likelihood estimators:

$$\widehat{\Omega}_z(\lambda_z), \widehat{\Omega}_\varepsilon(\lambda_\varepsilon) = \arg\max_{\Omega_z, \Omega_\varepsilon} \mathcal{L}(\mathsf{Y}; \Omega_z, \Omega_\varepsilon) - \tfrac{1}{2}\lambda_z\|\Omega_z\|_2^2 - \tfrac{1}{2}\lambda_\varepsilon\|\Omega_\varepsilon\|_2^2,$$

with a diagonal Ω_ε:

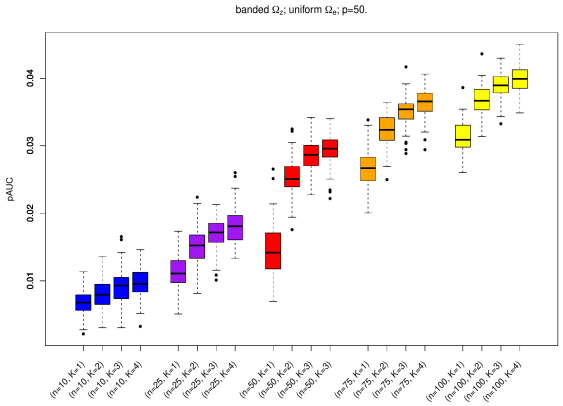
$$\widehat{\Omega}_z(\lambda_z), \widehat{\Omega}_\varepsilon = \arg\max_{\Omega_z, \Omega_\varepsilon} \mathcal{L}(\mathsf{Y}; \Omega_z, \Omega_\varepsilon) - \tfrac{1}{2}\lambda_z\|\Omega_z\|_2^2,$$

a lasso penalty:

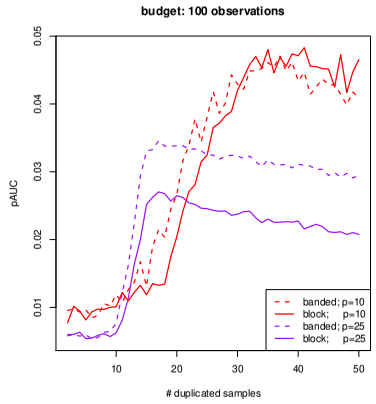
$$\widehat{\Omega}_z(\lambda_z), \widehat{\Omega}_\varepsilon = \arg\max_{\Omega_z, \Omega_\varepsilon} \mathcal{L}(\mathsf{Y}; \Omega_z, \Omega_\varepsilon) - \lambda_z\|\Omega_z\|_1.$$

All found by a penalized EM algorithm (= linear algebra fun).

Simulation: effect of K and n



Fixed # measurements



Illustration

Issues studied:

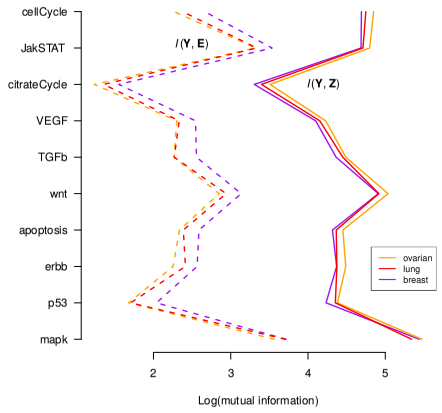
a) Effect on reconstruction of an error-diluted signal.
b) Tenability of the diagonal error assumption.
c) Network differences between replicated and non-replicated data.

Means:

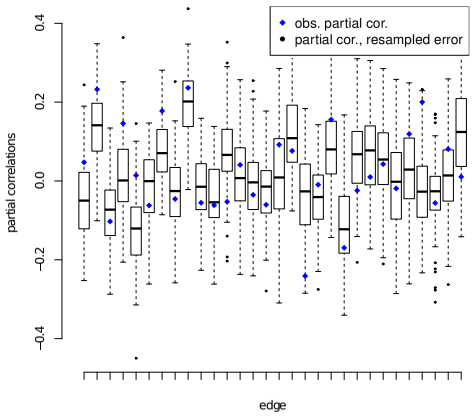
- Three TCGA studies
- Ten pathways: $p \in [29, 247]$
- Technical replicates: microarray and RNAseq

tissue	sample size n
lung	151
breast	526
ovarian	294

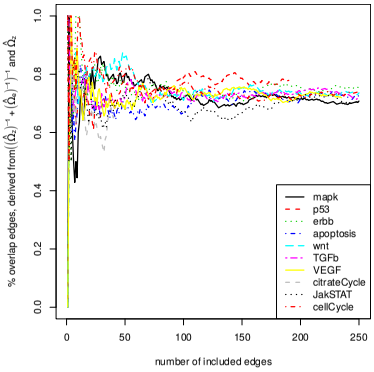
a) Is there any signal?



a) Error effect on partial correlations



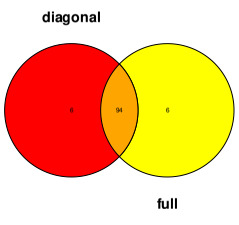
a) Intersection $\widehat{\mathcal{E}}_y$ and $\widehat{\mathcal{E}}_z$

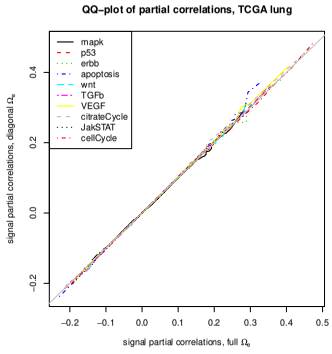


b) Full vs. diag. Ω_ε

QQ-plot of TCGA lung $\longrightarrow$

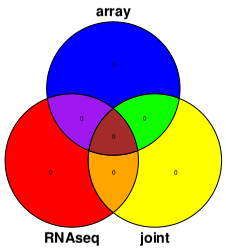
$\downarrow$ top 100 $\widehat{\mathcal{E}}_{\text{full}} \cap$ top 100 $\widehat{\mathcal{E}}_{\text{diag}}$

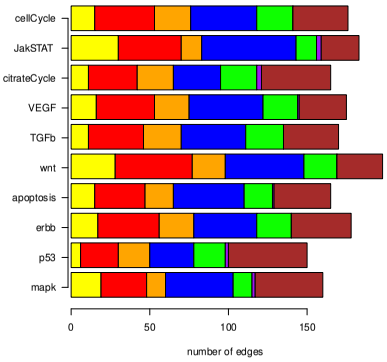




c) Intersection $\widehat{\mathcal{E}}_{\text{seq}}$, $\widehat{\mathcal{E}}_{\text{array}}$ and $\widehat{\mathcal{E}}_{\text{joint}}$

TCGA lung
$\rightarrow$ 100 edges per $\widehat{\mathcal{E}}$.





Conclusion

The generalized ridge precision estimator provides ways to incorporate detailed and structured prior knowledge, both:

- $\rightarrow$ *quantitatively*, via the target matrix, and
- $\rightarrow$ *qualitatively*, by the parametrization of the penalty matrix.

But ...
... reconstructed networks should be taken with reservation,
if measurement error is ignored.

References

These slides are based on:

- van Wieringen, W. N. (2019). The generalized ridge estimator of the inverse covariance matrix *Journal of Computational and Graphical Statistics*, 28(4), 932-942.
- van Wieringen, W.N., Chen, Y. (2021). Penalized estimation of the Gaussian graphical model from data with replicates. *Statistics in Medicine*, 40(19), 4279-4293.
- van Wieringen, W.N. (2022) `porridge`: Ridge-Type Estimation of a Potpourri of Models. R package version 0.3.1. `https://CRAN.R-project.org/package=porridge`.

License