Partial Correlation Network Estimation by Semismooth Newton Methods

Dongwon Kim*[1] Sungdong Lee*[2] Joong-Ho Won[1]

[1]Department of Statistics, Seoul National University
[2]Department of Medicine, National University of Singapore

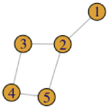
Conference on Neural Information Processing Systems

Background: Partial Correlation Network

From Precision Matrix to a Network

Let $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ be the inverse covariance (precision) matrix of a data distribution.

$$\boldsymbol{\Theta} = \begin{bmatrix} 1 & 0.2 & 0 & 0 & 0 \\ 0.2 & 1 & 0.2 & 0 & 0.2 \\ 0 & 0.2 & 1 & 0.2 & 0 \\ 0 & 0 & 0.2 & 1 & 0.2 \\ 0 & 0.2 & 0 & 0.2 & 1 \end{bmatrix} \Longleftrightarrow$$



- A sparse $\boldsymbol{\Theta}$ induces a graph: $\theta_{ij} = 0 \Leftrightarrow$ no edge.
- The quantity $\rho_{ij} := -\theta_{ij}/\sqrt{\theta_{ii}\theta_{jj}}$ is the **partial correlation**: the degree of association with all the other variables controlled.
- Estimating a sparse precision matrix corresponds to estimating the partial correlation network of the data.

Background: ACCORD estimator

ACCORD estimator (Lee et al., 2025)

$$\hat{\mathbf{\Omega}} = \min_{\mathbf{\Omega} \in \mathbb{R}^{p \times p}} -\log\det \mathbf{\Omega}_D + \frac{1}{2}\mathrm{tr}(\mathbf{\Omega}^\top \mathbf{\Omega} \boldsymbol{S}) + \lambda \|\mathbf{\Omega}_{-D}\|_1,$$

where $\boldsymbol{S}$ is the sample covariance of data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$; $\mathbf{\Omega}_D / \mathbf{\Omega}_{-D}$ denote diagonal/off-diagonal parts.

Why ACCORD?

- Using $\boldsymbol{X}$ from a data distribution with precision matrix $\mathbf{\Theta}^*$, the ACCORD estimator can consistently estimate

  $$\mathbf{\Omega}^* = \mathbf{\Theta}_D^{*-1/2} \mathbf{\Theta}^*,$$

  which is a reparameterization of $\mathbf{\Theta}^*$ preserving its sparsity pattern.
- Pseudolikelihood avoids $\mathbb{R}^{p \times p}$ matrix inversions required in Gaussian graphical models, enhancing scalability.

What We Do: From Row-Separable KKT to Fast Solvers

Key ideas

- Row-wise separability: ACCORD splits into p independent row subproblems.
- Per-row KKT: nonsmooth due to soft-thresholding $T_\lambda(\cdot)$ from the ℓ_1 penalty (kinks at $\pm\lambda$), yet the map is *strongly semismooth*.
- Our method: cast each row's KKT as a root-finding problem and solve via (B-)semismooth Newton plus damping strategy.
- Efficiency: small linear system solves per row; naturally parallel, and GPU-friendly.

Impact

- Rapid convergence in few iterations; locally superlinear/quadratic with stable sparsity recovery.
- Scales partial correlation network estimation to massive p on real multi-omics datasets.

Real data: TCGA LIHC $(p \approx 305k, n = 365)$

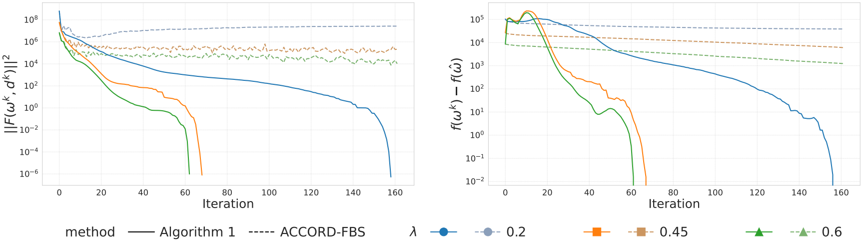


Figure: Convergence comparison on a large-scale multi-omics cancer dataset

- Hardware: $6 \times$ RTX 6000 Ada GPUs; row-wise parallelization over $\mathbf{\Omega}$.
- Convergence: converges in a few dozen iterations across λ values.
- Efficiency: consistently faster wall-clock and much smaller KKT residuals than proximal-only baselines.

Conclusion

Summary

A scalable second-order solver for ℓ_1-regularized pseudolikelihood-based partial correlation network estimation.

Results

Despite the nonsmooth nature of the problem, the proposed methods:

- Convergence: global descent via damping; locally superlinear/quadratic.
- Scalability: row-wise separability $\Rightarrow$ parallel/GPU-friendly updates.
- Efficiency: far fewer iterations than proximal baselines; enables massive-scale analyses.

For details:
Poster (NeurIPS 2025): Thu, Dec 4, 3–6 p.m. AST, Exhibit Hall C,D,E;