



SVD-Based Screening for the Graphical Lasso (IJCAI'17)

藤原 靖宏, **丸茂 直貴**, Mathieu Blondel,
竹内 孝, 金 秀明, 岩田 具治, 上田 修功

NTT 研究所

スパース精度行列推定 のための 2 つの高速化 手法を提案

スパース精度行列推定:

$$\min_{\Theta \in \mathbb{R}^{p \times p}} \text{tr}(S\Theta) - \log \det \Theta + \rho \|\Theta\|_1$$

- $X \in \mathbb{R}^{n \times p}$: (中心化) データ行列
- $S = X^T X$: 標本共分散行列 (の n 倍)
- $\|\cdot\|_1$: 要素ごとの 1 ノルム
- $\rho \in \mathbb{R}_{>0}$: 正則化パラメータ

スパース精度行列推定:

$$\min_{\Theta \in \mathbb{R}^{p \times p}} \text{tr}(S\Theta) - \log \det \Theta + \rho \|\Theta\|_1$$

$x \sim \mathcal{N}(0, \Sigma)$ のとき,

- 目的関数 = 負の対数尤度 + 正則化項
- 精度行列 $\Sigma^{-1} = (\lambda_{ij})$ に対し,
 $\lambda_{ij} = 0 \iff i, j$ は条件付き独立

スパース精度行列推定:

$$\min_{\Theta \in \mathbb{R}^{p \times p}} \text{tr}(S\Theta) - \log \det \Theta + \rho \|\Theta\|_1$$

スパースな精度行列を推定することで
変数間の (スパースな) 依存関係が推定できる

- 精度行列 $\Sigma^{-1} = (\lambda_{ij})$ に対し,
 $\lambda_{ij} = 0 \iff i, j$ は条件付き独立

既存手法

- Graphical Lasso
- その Screening

スパース精度行列推定:

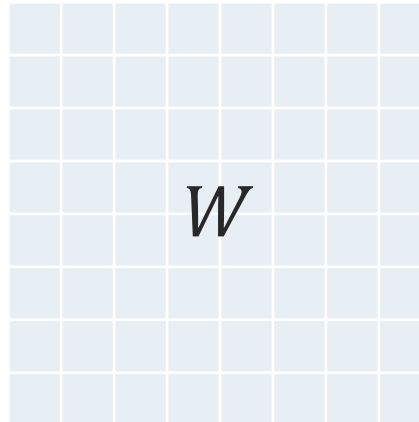
$$\min_{\Theta \in \mathbb{R}^{p \times p}} \text{tr}(S\Theta) - \log \det \Theta + \rho \|\Theta\|_1$$

Graphical Lasso [Friedman (2007)]:

- 双対に対する Block Coordinate Descent
- 様々な高速化が提案されている最も有名な解法

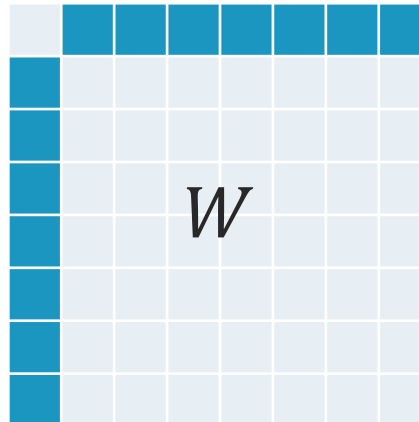
$W \in \mathbb{R}^{p \times p}$: 双対変数

最適性条件を書くと $w_{ii}^* = s_{ii} + \rho$ ($\forall i$) が分かる



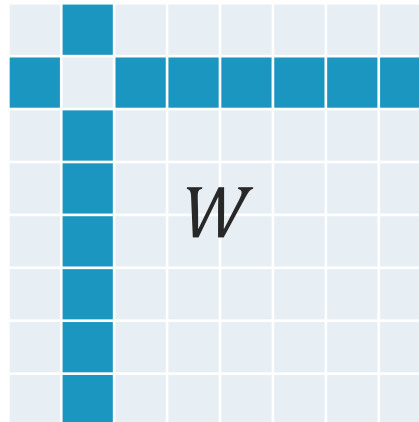
$W \in \mathbb{R}^{p \times p}$: 双対変数

最適性条件を書くと $w_{ii}^* = s_{ii} + \rho$ ($\forall i$) が分かる



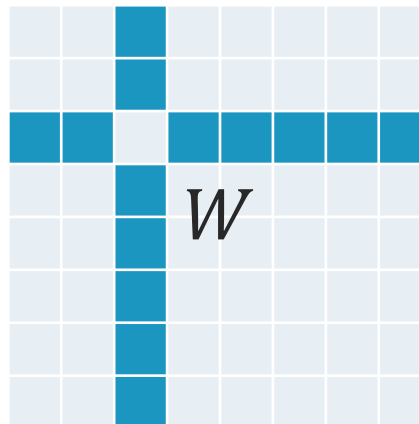
$W \in \mathbb{R}^{p \times p}$: 双対変数

最適性条件を書くと $w_{ii}^* = s_{ii} + \rho$ ($\forall i$) が分かる



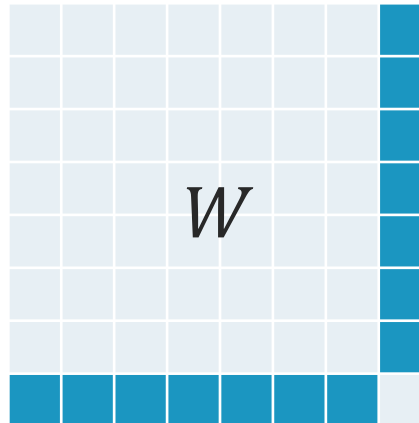
$W \in \mathbb{R}^{p \times p}$: 双対変数

最適性条件を書くと $w_{ii}^* = s_{ii} + \rho$ ($\forall i$) が分かる



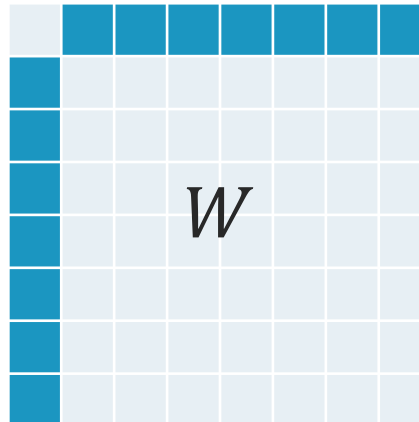
$W \in \mathbb{R}^{p \times p}$: 双対変数

最適性条件を書くと $w_{ii}^* = s_{ii} + \rho$ ($\forall i$) が分かる



$W \in \mathbb{R}^{p \times p}$: 双対変数

最適性条件を書くと $w_{ii}^* = s_{ii} + \rho$ ($\forall i$) が分かる



Graphical Lasso の 1 反復



$$W = \begin{array}{c} \begin{array}{|c|c|} \hline W_A & W_B \\ \hline W_B^T & W_D \\ \hline \end{array} \end{array}$$

w_B のみ動かす

Graphical Lasso の 1 反復



$$W = \begin{array}{c} \begin{array}{|c|c|} \hline W_A & W_B \\ \hline \end{array} \\ \begin{array}{|c|c|} \hline W_B^T & W_D \\ \hline \end{array} \end{array} \quad w_B \text{ のみ動かす}$$

$$1. \hat{\beta} \leftarrow \operatorname{argmin}_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2} \beta^T W_A \beta - s_B^T \beta + \rho \|\beta\|_1 \right\}$$

$$2. w_B \leftarrow W_A \hat{\beta}$$

$$3. \theta_B \leftarrow \frac{\hat{\beta}}{w_D - w_B^T \hat{\beta}}$$

$$4. \theta_D \leftarrow \frac{1}{w_D - w_B^T \hat{\beta}}$$

$$\Theta = \begin{array}{c} \begin{array}{|c|c|} \hline \Theta_A & \theta_B \\ \hline \end{array} \\ \begin{array}{|c|c|} \hline \theta_B^T & \theta_D \\ \hline \end{array} \end{array}$$

Graphical Lasso の 1 反復



$$W = \begin{array}{c} \begin{array}{|c|c|} \hline W_A & W_B \\ \hline \end{array} \\ \begin{array}{|c|c|} \hline W_B^T & W_D \\ \hline \end{array} \end{array}$$

w_B のみ動かす

$$1. \hat{\beta} \leftarrow \operatorname{argmin}_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2} \beta^T W_A \beta - s_B^T \beta + \rho \|\beta\|_1 \right\}$$

$$2. w_B \leftarrow W_A \hat{\beta}$$

これも Coordinate Descent で解く

$$3. \theta_B \leftarrow \frac{\hat{\beta}}{w_D - w_B^T \hat{\beta}}$$

$$4. \theta_D \leftarrow \frac{1}{w_D - w_B^T \hat{\beta}}$$

$$\Theta = \begin{array}{c} \begin{array}{|c|c|} \hline \Theta_A & \theta_B \\ \hline \end{array} \\ \begin{array}{|c|c|} \hline \theta_B^T & \theta_D \\ \hline \end{array} \end{array}$$

事実 [Witten+ (2011)]

主最適解 Θ^* , 双対最適解 W^* はともに

$$A \in \mathbb{R}^{p \times p} : a_{ij} = \begin{cases} 1 & (|s_{ij}| > \rho) \\ 0 & (\text{otherwise}) \end{cases}$$

と同じブロック対角構造を持つ

$$\min_{\Theta \in \mathbb{R}^{p \times p}} \text{tr}(S\Theta) - \log \det \Theta + \rho \|\Theta\|_1$$

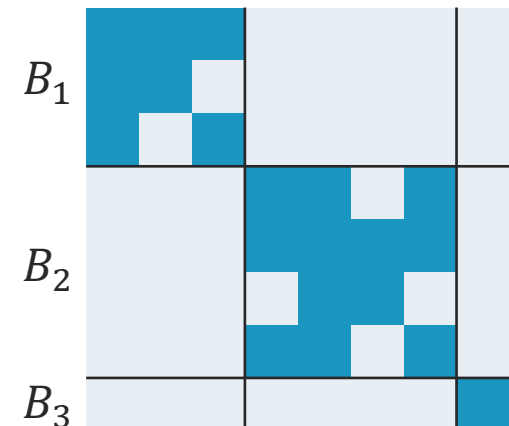
事実 [Witten+ (2011)]

主最適解 Θ^* , 双対最適解 W^* はともに

$$A \in \mathbb{R}^{p \times p} : a_{ij} = \begin{cases} 1 & (|s_{ij}| > \rho) \\ 0 & (\text{otherwise}) \end{cases}$$

と同じブロック対角構造を持つ

小さいサイズの Graphical Lasso を
 m 個適用すればよい!



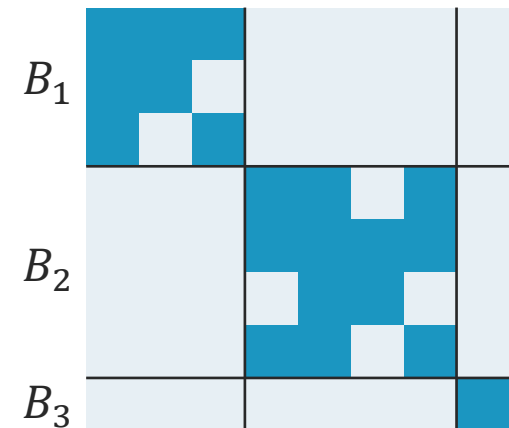
事実 [Witten+ (2011)]

主最適解 Θ^* , 双対最適解 W^* はともに

高速化が期待されるだけで
得られる解は変化しない

と同じブロック対角構造を持つ

小さいサイズの Graphical Lasso を
 m 個適用すればよい!



提案手法

- Screening の高速化
- Graphical Lasso の高速化

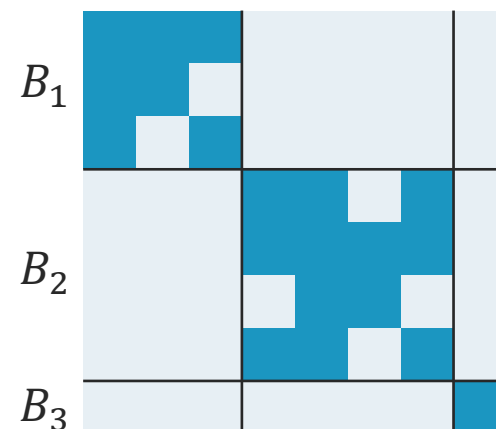


仮定: $X \in \mathbb{R}^{n \times p}$ は与えられているが,
 $S = X^T X \in \mathbb{R}^{p \times p}$ は与えられていない
➤ S の計算に $O(np^2)$ がかかる

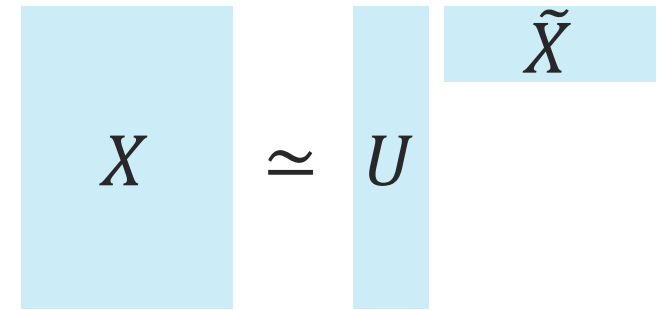
- 仮定:** $X \in \mathbb{R}^{n \times p}$ は与えられているが,
 $S = X^T X \in \mathbb{R}^{p \times p}$ は与えられていない
- S の計算に $O(np^2)$ がかかる

アイデア:

- 既存 Screening では S の全要素を正確に計算 ☹️
- Θ^*, W^* の 0 ブロックに対応する S の値は正確に計算する必要は無い 😊
 - **上下界で代用**



- $X \approx U\tilde{X}$ とランク k 近似
($U \in \mathbb{R}^{n \times k}$: 列直交)



- $$s_{ij} = x_i^\top x_j = \frac{1}{2} \left(\|x_i\|^2 + \|x_j\|^2 - \|x_i - x_j\|^2 \right)$$
$$\leq \frac{1}{2} \left(\|x_i\|^2 + \|x_j\|^2 - \|\tilde{x}_i - \tilde{x}_j\|^2 \right)$$

(下界も同様)

☹ S の厳密計算は $O(np^2)$

☺ S の上下界の計算は $O(np \log k + kp^2)$

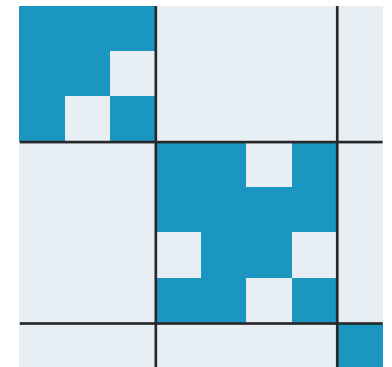
1. $S = X^T X$ の上下界 \bar{s} , \underline{s} を計算

2. $A \in \mathbb{R}^{p \times p} : a_{ij} = \begin{cases} 1 & (\bar{s}_{ij} < -\rho \text{ or } \underline{s}_{ij} > \rho) \\ 0 & (\text{otherwise}) \end{cases}$

の行・列を置換し, ブロック対角化

3. A の対角ブロックに対応する S の値を厳密計算

4. S の各対角ブロックに対し
Graphical Lasso

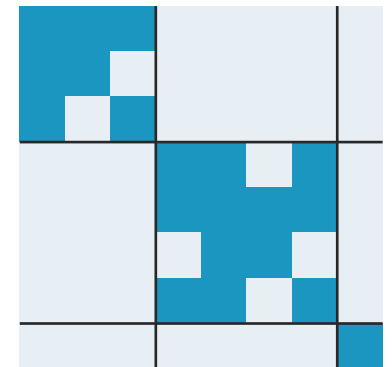


1. $S = X^T X$ の上下界 \bar{s} , \underline{s} を計算

2. $A \in \mathbb{R}^{n \times n}$ の行 i について $A_{ii} > \rho$ の場合、高速化が期待されるだけで得られる解は変化しない

3. A の対角ブロックに対応する S の値を厳密計算

4. S の各対角ブロックに対し Graphical Lasso



提案手法

- Screening の高速化
- Graphical Lasso の高速化

Graphical Lasso の 1 反復 (再掲)



$$W = \begin{array}{c|c} & W_A & W_B \\ \hline W_B^T & & W_D \end{array} \quad w_B \text{ のみ動かす}$$

$$1. \hat{\beta} \leftarrow \operatorname{argmin}_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2} \beta^T W_A \beta - s_B^T \beta + \rho \|\beta\|_1 \right\}$$

$$2. w_B \leftarrow W_A \hat{\beta} \quad \text{これも Coordinate Descent で解く}$$

$$3. \theta_B \leftarrow - \frac{\hat{\beta}}{w_D - w_B^T \hat{\beta}}$$

$$4. \theta_D \leftarrow \frac{1}{w_D - w_B^T \hat{\beta}}$$

$$\Theta = \begin{array}{c|c} & \Theta_A & \theta_B \\ \hline \theta_B^T & & \theta_D \end{array}$$

- 観察:**
- $|s_{ij}| > \rho$ なる (i, j) に対して $\theta_{ij} \neq 0$ となることが多い
 - θ_B は $\hat{\beta}$ のスカラー一倍

$$1. \hat{\beta} \leftarrow \operatorname{argmin}_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2} \beta^\top W_A \beta - s_B^\top \beta + \rho \|\beta\|_1 \right\}$$

$$2. w_B \leftarrow W_A \hat{\beta}$$

これも Coordinate Descent で解く

$$3. \theta_B \leftarrow -\frac{\hat{\beta}}{w_D - w_B^\top \hat{\beta}}$$

$$4. \theta_D \leftarrow \frac{1}{w_D - w_B^\top \hat{\beta}}$$

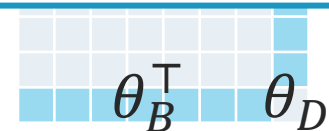
$$\Theta = \begin{array}{c} \begin{array}{|c|c|} \hline \Theta_A & \theta_B \\ \hline \theta_B^\top & \theta_D \\ \hline \end{array} \end{array}$$

- 観察:**
- $|s_{ij}| > \rho$ なる (i, j) に対して $\theta_{ij} \neq 0$ となることが多い
 - θ_B は $\hat{\beta}$ のスカラー倍

$$1. \hat{\beta} \leftarrow \operatorname{argmin}_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2} \beta^\top W_A \beta - s_B^\top \beta + \rho \|\beta\|_1 \right\}$$

1 を解く際に, $|s_{ij}| > \rho$ なる (i, j) に対応する β のみ非ゼロと決め打ちしたらどうか?

$$4. \theta_D \leftarrow \frac{1}{w_D - w_B^\top \hat{\beta}}$$


$$\theta_B^\top \quad \theta_D$$

1. $S : |s_{ij}| > \rho$ なる (i, j) に対応する β の添字集合
2. S に対応する β の要素のみ非ゼロと仮定して CD
3. $i \notin S$ なる β_i に対し, CD の反復を 1 回ずつ行い, 値が変化しないかを確認
4. 値が変化した i を S に追加し, 2に戻る

注: CD の不動点 \Leftrightarrow 最適解

1. $S : |s_{ij}| > \rho$ なる (i, j) に対応する β の添字集合
2. S に対し **高速化が期待されるだけで** 定して CD
得られる解は変化しない
3. $i \notin S$ なる β_i に対し, CD の反復を 1 回ずつ行い, 値が変化しないかを確認
4. 値が変化した i を S に追加し, 2 に戻る

注: CD の不動点 \Leftrightarrow 最適解

数值実験

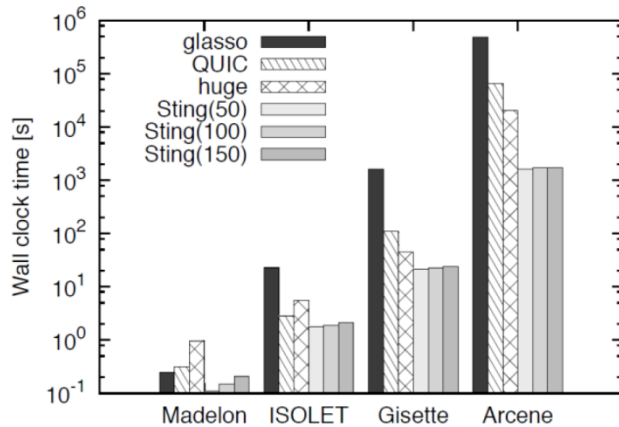
データセット:

- Madelon: 600×500
- ISOLET: 6238×618
- Gisette: 1000×5000
- Arcene: 700×10000

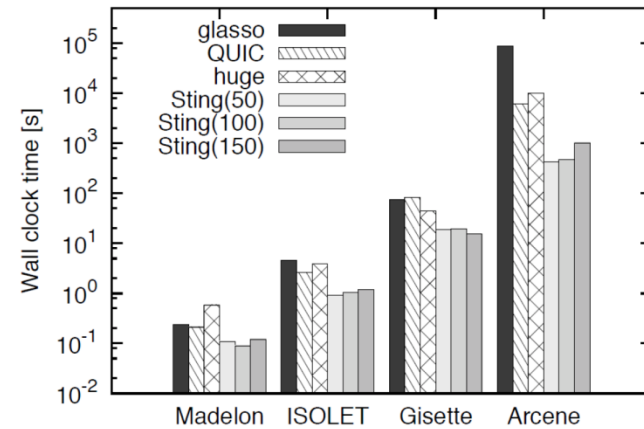
比較手法の実装:

- glasso [Friedman+ (2015)]: Graphical Lasso based
- QUIC [Hsieh+ (2015)]: Newton 法 based
- huge [Zhao+ (2015)]: CD 法 based

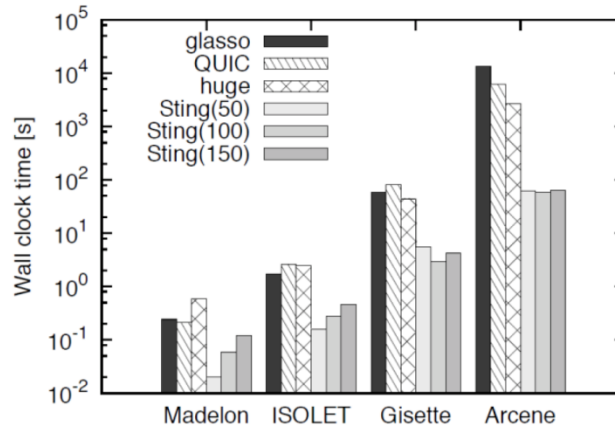
実験結果: 実行時間の比較



$\rho = 0.5$



$\rho = 0.7$



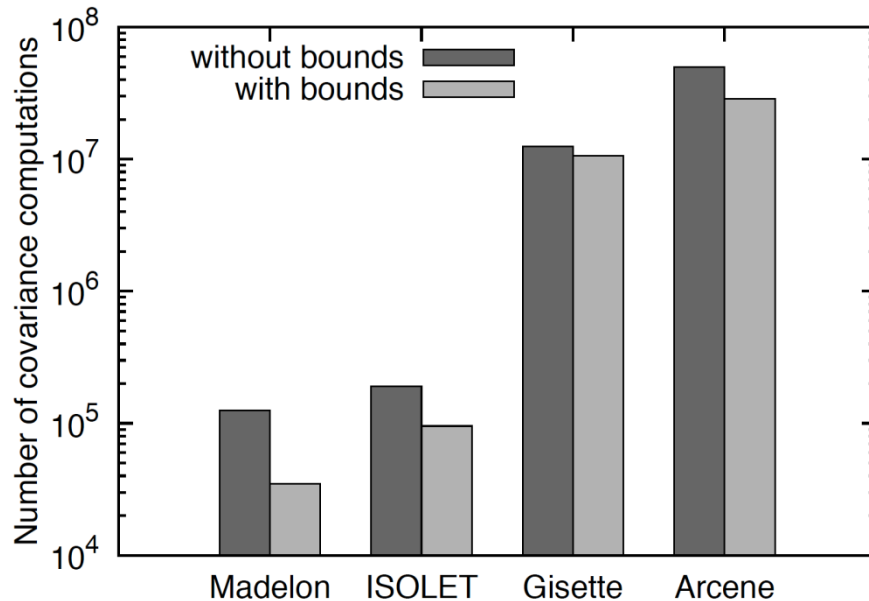
$\rho = 0.9$

提案手法は数十～数百倍高速！

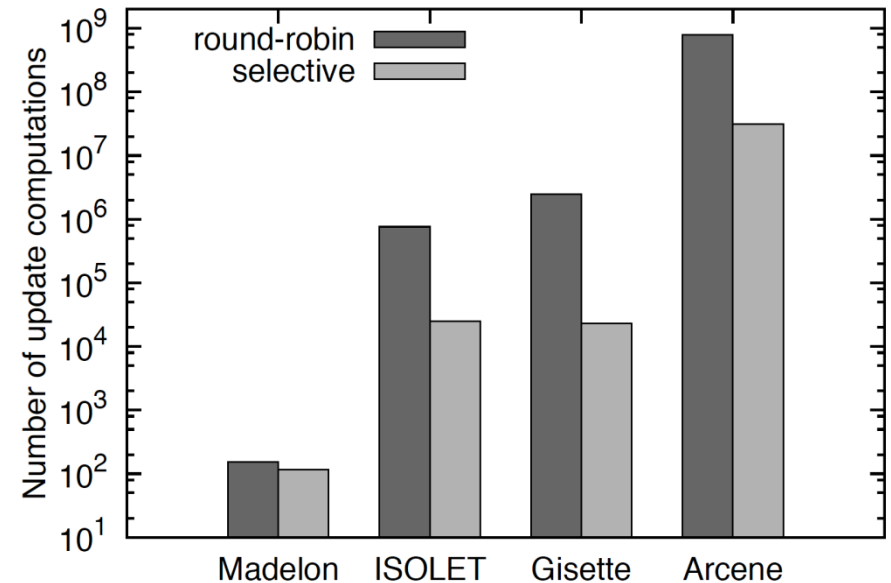
実験結果: 2つの高速化の効果



Sの要素の計算個数



CDの反復数



スクリーニングの高速化も
CDの高速化も効いている

- スパース精度行列推定のための
 - Screening の高速化
 - Graphical Lasso の高速化を提案
- 提案手法で得られる解は既存手法と同じ
- 提案手法の有効性を実験的に確認