

IBM Research - Tokyo

# 潜在的グラフ構造からの異常検知

IBM東京基礎研究所  
井手 剛

## 内容

---

- やりたいこと
- グラフィカル・ガウシアン・モデルと関連研究
- 疎構造学習の方法
- 相関異常度の定義
- 実験結果
- まとめ

▶ Acknowledgement

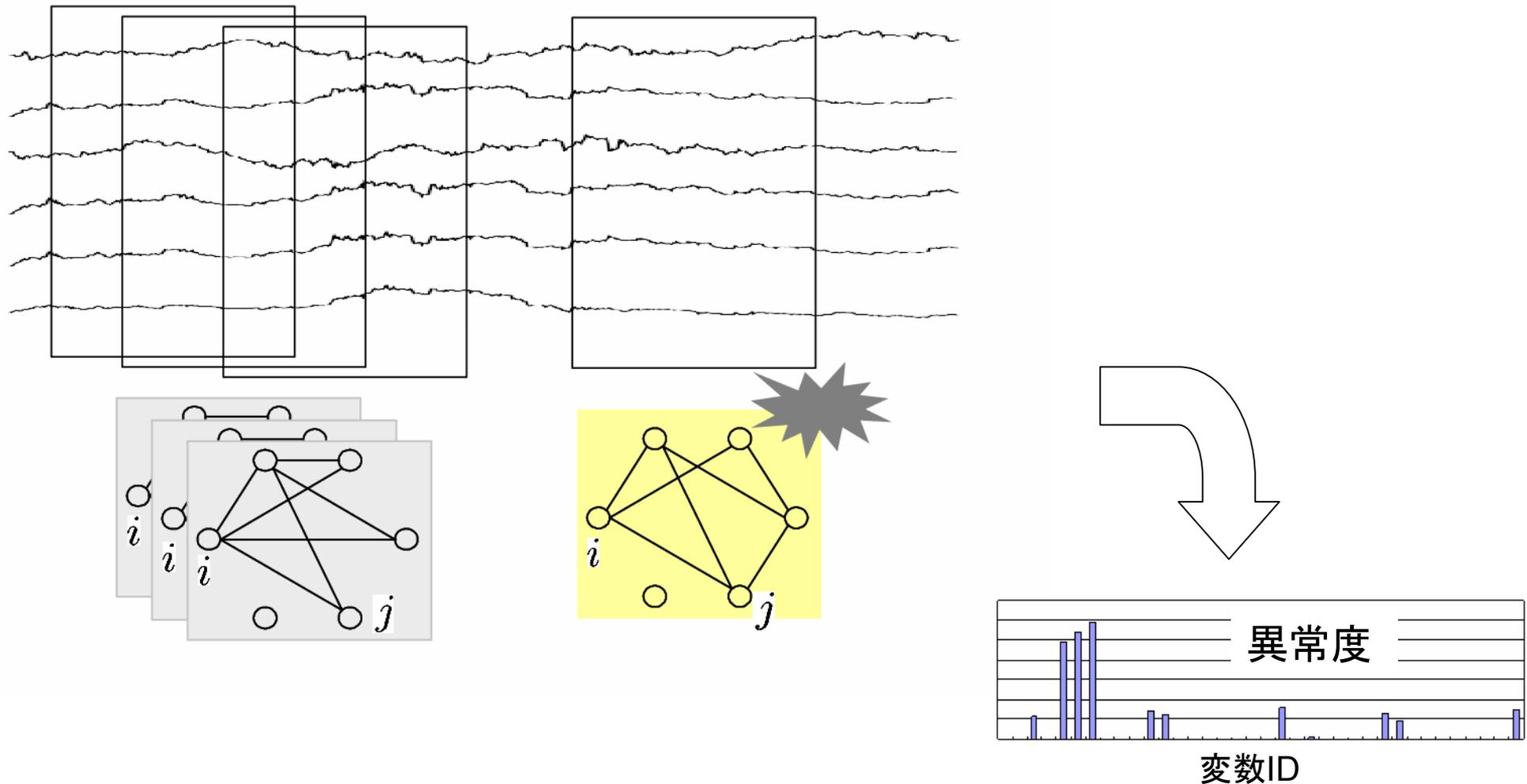
- This is a joint work with Aurelie C. Lozano, Naoki Abe, and Yan Liu (IBM T. J. Watson Research Center).

## 内容

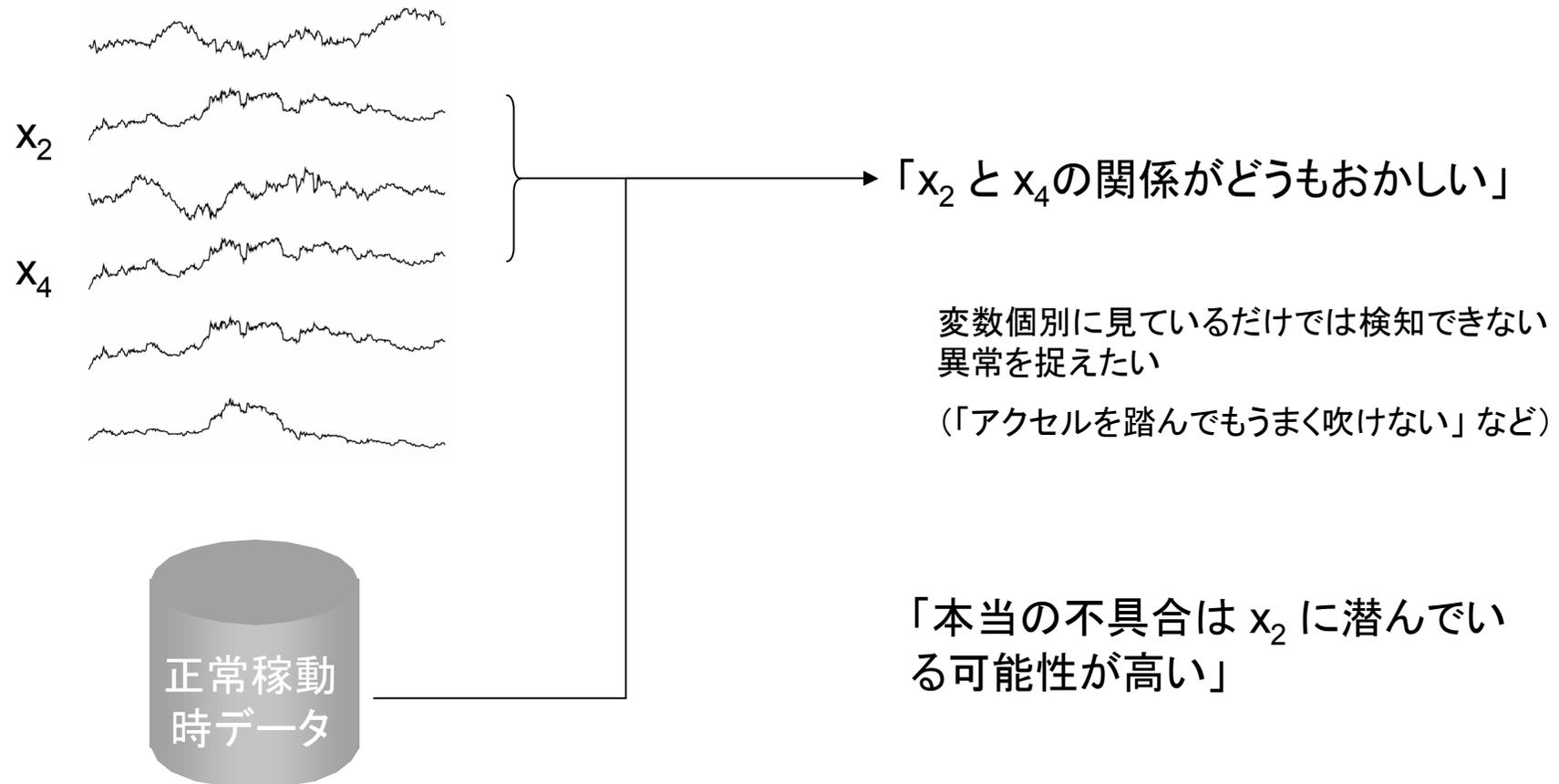
---

- やりたいこと
- グラフィカル・ガウシアン・モデルと関連研究
- 疎構造学習の方法
- 相関異常度の定義
- 実験結果
- まとめ

やりたいこと: 変数同士の「関係の崩れ」を検出したい。  
正常時のデータを元にして、個々の変数の相関異常度を計算したい



やりたいこと: 変数同士の「関係の崩れ」を検出したい。  
正常時のデータを元にして、個々の変数の相関異常度を計算したい



# 何が難しいか: ノイジーなセンサーデータでは変数同士の関係は非常に不安定。 Actual spot rates データの例 (1/2)

- 各国通貨の対ドルレートの変動を表した時系列データ
- ほとんどの相関係数の値は非常に不安定
  - ▶ 経済メカニズム自体は変わっていないはずだが、値は安定していない

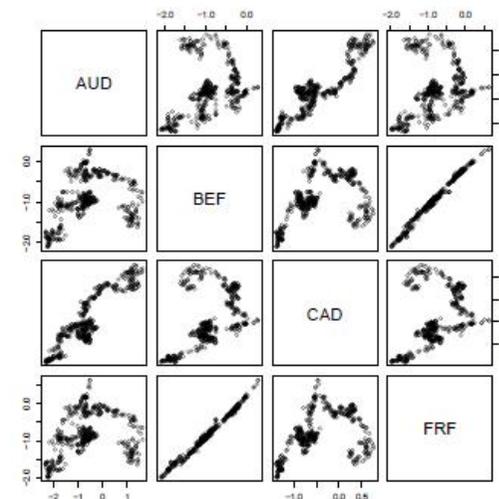
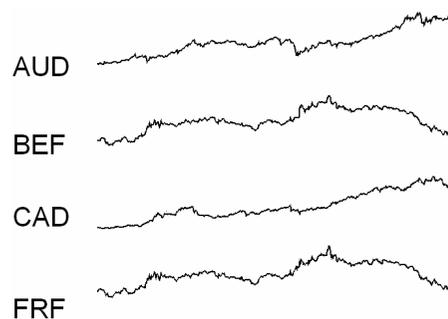
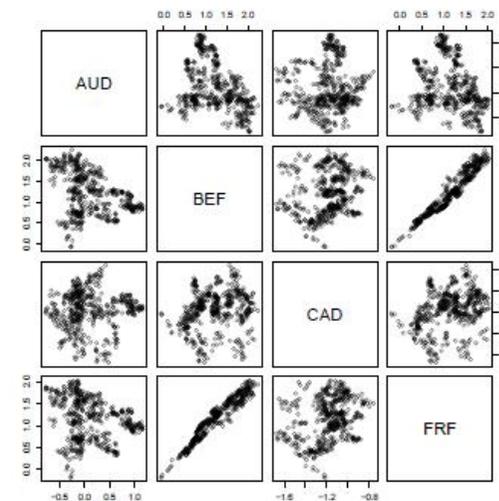
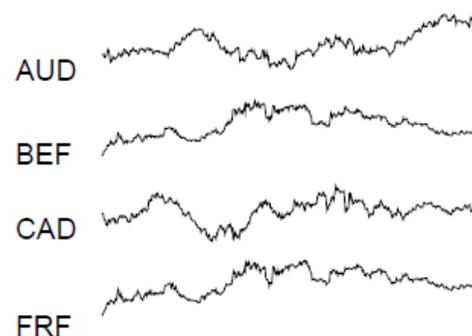


表 2: 図 3 に示す散布図プロットに対応した相関係数。カッコ前の数字は上の図 (カッコの中は下の図) に対応する。

	BEF	CAD	FRF
AUD	0.31 (-0.37)	0.91 (0.04)	0.26 (-0.23)
BEF		0.46 (0.19)	0.99 (0.97)
CAD			0.41(0.30)



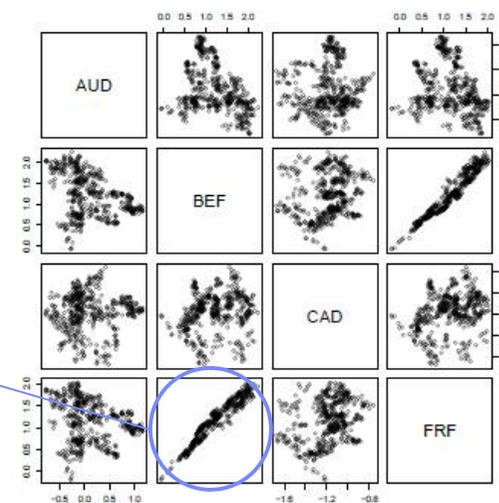
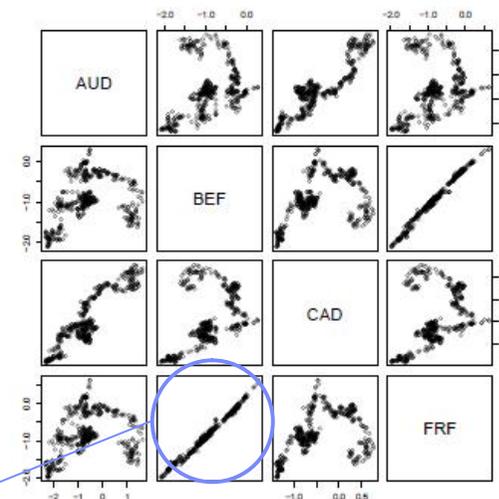
## 何が難しいか: ノイジーなセンサーデータでは変数同士の関係は非常に不安定。 Actual spot rates データの例 (2/2)

- 相関の強いペアについては関係が安定している

個々の変数の「近傍」だけ見たい。  
弱い関連は「適切に」無視したい。

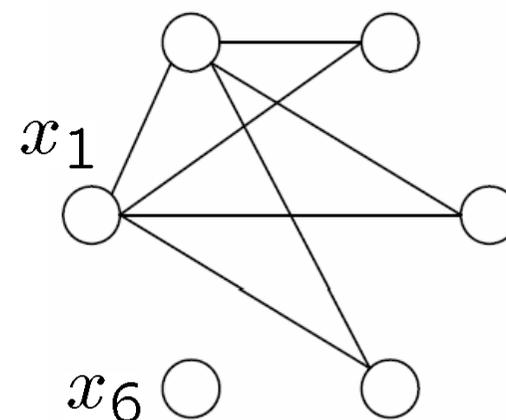
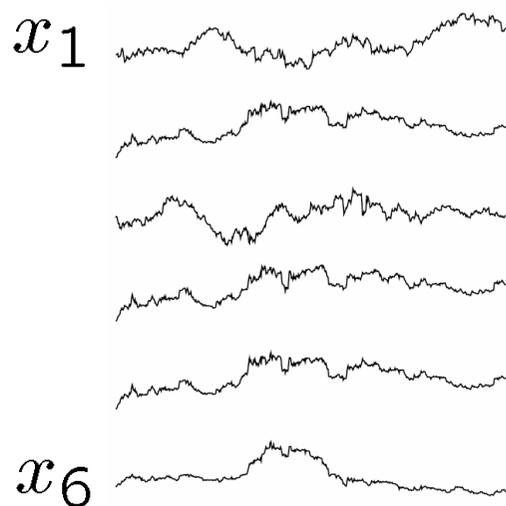
表 2: 図 3 に示す散布図プロットに対応した相関係数。  
カッコ前の数字は上の図 (カッコの中は下の図) に対応する。

	BEF	CAD	FRF
AUD	0.31 (-0.37)	0.91 (0.04)	0.26 (-0.23)
BEF		0.46 (0.19)	0.99 (0.97)
CAD			0.41(0.30)



本質的なつながりだけを残すように、変数の依存関係を表すグラフを学習したい  
→ 疎な構造を学習したい

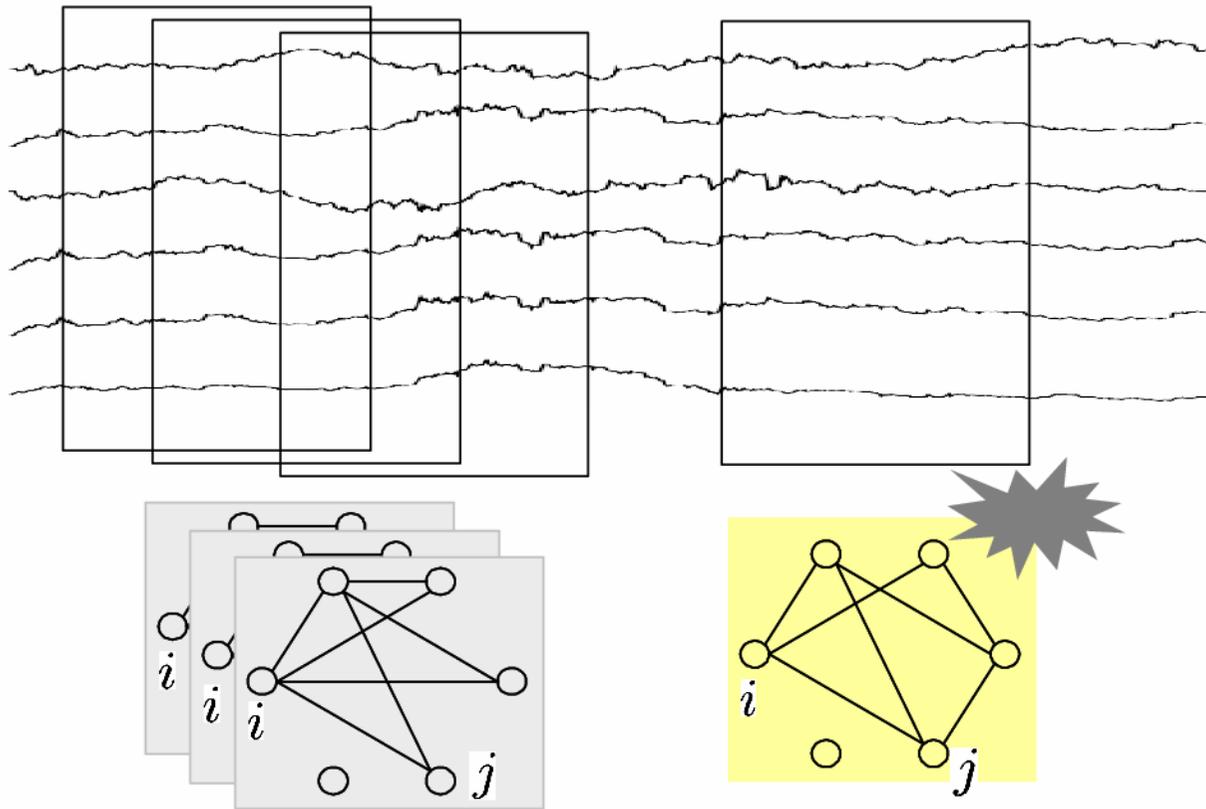
- 入力: (今回は)実数値の、多次元データ
- 出力: つながりを表す重み付きグラフ



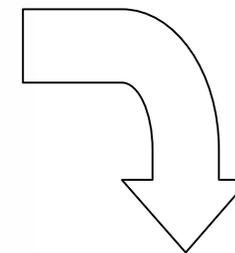
- 頂点は各変数
- 辺は変数間の関連
- 2つの頂点間に辺がない = 他を与えた時に両者は独立

## 2つの課題がある

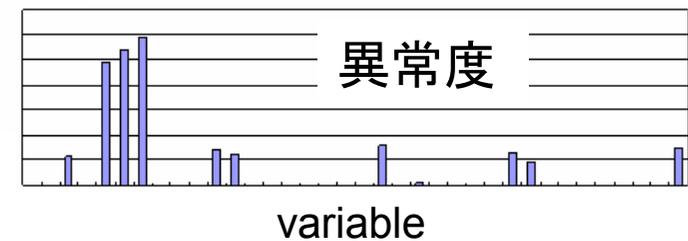
### (1) スパース構造学習、(2) 変数ごとの異常度の計算



(1) スパース構造学習



(2) 相関異常度のスコアリング



## 内容

---

- やりたいこと
- グラフィカル・ガウシアン・モデルと関連研究
- 疎構造学習の方法
- 相関異常度の定義
- 実験結果
- まとめ

## Graphical Gaussian Model (GGM) におけるグラフの定義: 「精度行列の行列要素がゼロなら辺なし」

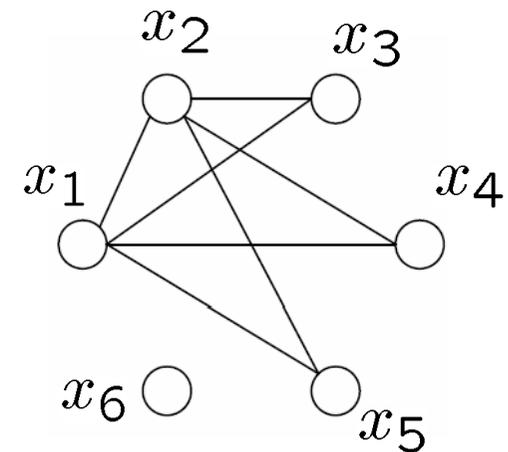
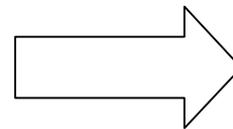
▪ 精度行列  $\Lambda$  = 共分散行列  $S$  の逆行列

▪ 例:  $\Lambda_{1,2} = 0$  なら  $x_1$  と  $x_2$  は条件付き独立で、頂点1と2の間には辺はない

▶ なぜなら exp の部分が因子化されるから:  $\mathcal{N}(x|\mathbf{0}, \Lambda^{-1}) = \frac{\det(\Lambda)^{1/2}}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2}x^\top \Lambda x\right)$

▪ 例2: 6変数の場合の例

$$\Lambda = \begin{pmatrix} * & * & * & * & * & 0 \\ * & * & * & * & * & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & 0 & * & 0 & 0 \\ * & * & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & * \end{pmatrix}$$



## 疎な精度行列を得るための手法 (1/2): 伝統的には共分散構造選択という手法が使われてきた

- **素朴な方法: 共分散行列の逆行列を求めて、ある閾値以下の要素をゼロとしてしまう**
  - ▶ 確率モデルではなくなってしまう。例えば、そういう精度行列は正定値ではなくなる
  - ▶ 閾値の設定が実用上簡単ではない
  
- **伝統的な方法: 共分散構造選択 (Dempster 1972)**
  - ▶ 簡単に言えば以下の繰り返し
    - 小さい行列要素をひとつゼロにする
    - それを拘束として、確率モデルを推定しなおす
    - その上で小さい行列要素をひとつゼロにする
    - それを拘束として、確率モデルを推定しなおす
    - . . . .

疎な精度行列を得るための手法 (2/2):  
 $L_1$ 正則化によりスパース性を得る手法が近年発展している。

### Covariance selection

Doesn't work for rank-deficient situations

### Graphical Lasso [Friedman et al. 08]

- Can handle rank-deficient situations
- Stable even under collinearities

### (Independent) Lasso [Meinshausen & Bühlmann 06]

- Can handle rank-deficient situations
- Quite instable under collinearities

この2つについて後で比較する

## その他の関連研究

---

- **2標本検定: ふたつのデータセット同士の相違を仮説検定する**
  - ▶ 問題が違う: 個々の変数のスコアリングまではしない
  - ▶ 伝統的には漸近分布での仮説検定: ノイジーで小標本なデータだと正当化しにくい
  
- **相関係数の検定**
  - ▶ Wishart 分布論に基づく検定の手法がある
    - たとえば Anderson, “An Introduction to Multivariate Statistical Analysis”, Willy 参照
  - ▶ が、ノイジーで小標本なデータには使い物にならない
  
- **非線形への拡張は今後の課題**
  - ▶ GGMに基づく以上、今回は線形な相関異常のみに着目している
  - ▶ 理論的には可能だと思われるが、うまい実例が見つかるかが(論文的には)カギ

## グラフィカル・ガウシアン・モデルについてのコメント

- 一般に、データは多変量ガウス分布にはまったく従わないが、異常検知の文脈ではGGMは非常に有用
  - ▶ 現在の問題は密度推定ではない
  - ▶ 2次キュムラント以上の高次の統計量は安定した計算が困難で、その解釈も難しい
  
- GGMでは因果性が表現できない。因果グラフの学習は応用上も非常に重要であるが、ノイズにロバストな計算手法は今後の課題
  
- ダイナミックスの取り込みは非常に興味深い研究課題だが、模索中
  - ▶ c.f.
    - Y. Liu et al., “*Learning dynamic temporal graphs for oil-production equipment monitoring system,*” KDD 2009
  - ▶ 今回は、滑走窓の意味でのみ時間変動を考える

## 内容

---

- やりたいこと
- グラフィカル・ガウシアン・モデルと関連研究
- **疎構造学習の方法**
- 相関異常度の定義
- 実験結果
- まとめ

## ラプラス事前分布を付したMAP推定を行い精度行列を求める

- 観測モデルが正規分布

$$p_G(\mathbf{x}|\Lambda) = \mathcal{N}(\mathbf{x}|\mathbf{0}, \Lambda^{-1}) = \frac{\det(\Lambda)^{1/2}}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \Lambda \mathbf{x}\right)$$

- 精度行列  $\Lambda$  についての事前分布

$$p(\Lambda) = \prod_{i,j=1}^M \frac{\rho}{2} \exp\left(-\rho|\Lambda_{i,j}|\right)$$

- MAP (Maximum a posteriori) 推定で求める

$$\Lambda^* = \arg \max_{\Lambda} \left\{ \ln p(\Lambda) \prod_{n=1}^N p_G(\mathbf{x}^{(n)}|\Lambda) \right\}$$

## MAP方程式は、 $L_1$ 正則化項付きの最適化問題に帰着される

### ▪ 入力: 共分散行列 $S$

- ▶ 平均ゼロ、分散1に標準化したデータが前提
- ▶ 普通、ランク落ちしているなので逆は存在せず

$$S_{i,j} \equiv \frac{1}{N} \sum_{n=1}^N x_i^{(n)} x_j^{(n)}$$

### ▪ 出力: スパースな精度行列 $\Lambda$

- ▶ 精度行列 = 共分散行列の逆行列

### ▪ 方法: $L_1$ 正則化項付きの最尤方程式を解く

$$\Lambda^* = \arg \max_{\Lambda} \left\{ \ln \det \Lambda - \text{tr}(S\Lambda) - \rho \|\Lambda\|_1 \right\}$$

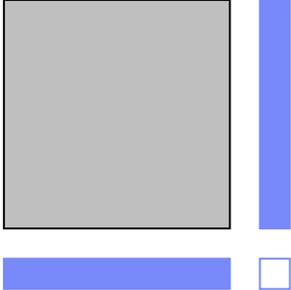
対数尤度  $\ln \prod_{t=1}^N \mathcal{N}(\mathbf{x}^{(t)} | \mathbf{0}, \Lambda^{-1})$

正則化項

## Graphical Lasso algorithm: 各列に着目して、行列についての最適化問題をベクトルに対する問題に直す(ブロック勾配法)

### ■ 精度行列を1列(1行)ずつ最適化

- ▶ 灰色部分を定数だと思って、青色部分についての最適化問題を導く

$$\Lambda = \begin{pmatrix} L & l \\ l^\top & \lambda \end{pmatrix}$$


- ▶ 青色ベクトルについての最適化問題は、 $L_1$ 正則化項付きの2次計画問題になる
  - 劣勾配法により効率のよい固定点方程式を導ける(Friedman et al. 2008)

### ■ スパースな精度行列を、明示的な逆行列計算なしに求めることができる

- ▶ 副産物として、精度行列の逆も(逆行列計算なしに)求まる
  - 標本共分散行列Sの修正版のようなもの

(詳しくは: T. Idé et al., "Proximity-Based Anomaly Detection using Sparse Structure Learning," SDM 2009)

## 各列についての最適化問題は、いわゆるLassoと同等になる

- 着目する変数が一番最後の列に来るように変数を並びかえる

$$\Lambda = \begin{pmatrix} L & l \\ l^\top & \lambda \end{pmatrix} \quad \Sigma \equiv \Lambda^{-1} = \begin{pmatrix} W & w \\ w^\top & \sigma \end{pmatrix} \quad S = \begin{pmatrix} S \setminus i & s \\ s^\top & s_{i,i} \end{pmatrix}$$

- 素朴に行列をばらしてゆくことにより(予稿参照)、ベクトル  $\beta \equiv W^{-1}w$  についての次の最適条件が導かれる

$$\min_{\beta} \left\{ \frac{1}{2} \|W^{\frac{1}{2}}\beta - b\|^2 + \rho \|\beta\|_1 \right\} = 0$$

$$b \equiv W^{-1/2}s$$

$L_1$ 正則化項付きの線形回帰と同等=Lasso

- 結局、精度行列のひとつの列は次のように求まる

$$\lambda = \frac{1}{\sigma - \beta^\top W \beta}, \quad l = -\frac{\beta}{\sigma - \beta^\top W \beta}$$

収束するまですべての変数について計算を繰り返す

## Meinshausen & Bühlmann (2006) との比較

### MBの方法はMAP最適性のような大局的最適性を持たない

#### ▪ MBの方法: 各変数に対して独立に、「自分vs他人」、のLasso回帰問題を解く

▶ 例えば変数の数 $M=5$ であれば、

- $x_1$  を  $x_2, x_3, x_4, x_5$  から予測する線形回帰モデルを作る
- $x_2$  を  $x_1, x_3, x_4, x_5$  から予測する線形回帰モデルを作る
- . . .
- $x_5$  を  $x_1, x_2, x_3, x_4$  から予測する線形回帰モデルを作る

$$\min_{\beta} \left\{ \frac{1}{2} \|Z_i \beta - \mathbf{y}_i\|^2 + \mu \|\beta\|_1 \right\} \quad \mathbf{y}_i \equiv (x_i^{(1)}, \dots, x_i^{(N)})^\top$$

$$z_i^{(n)} \equiv (x_1^{(n)}, \dots, x_{i-1}^{(n)}, x_{i+1}^{(n)}, \dots, x_M^{(n)})^\top \in \mathbb{R}^{M-1} \quad Z_i \equiv [z_i^{(1)}, \dots, z_i^{(N)}]^\top$$

#### ▪ この回帰係数は、精度行列の各列に比例しているので(GGMの基本性質)、M回の回帰を行うことで精度行列を求められる

▶ Lassoの解のスプース性から、得られる精度行列もスプースになる

#### ▪ MBの方法は、一般のデータでは、グラフィカルLassoのMAP推定のヒューリスティックに基づく近似と見なされる

## 正規化項の係数 $\rho$ は相関係数の閾値と解釈できる

- 今の問題設定では、異常検知性能を最大化するように  $\rho$  を決める
- $\rho$  は、「相関係数のどの値までを有意な相関とみなすか」の指標と解釈できる
  - ▶  $2 \times 2$ の問題を解析的に解くことで、次の結果を導ける (Idé et al., 2009)
  - ▶ 相関係数  $r$  が  $\rho$  よりも小さいと、対応する偏相関係数はゼロになる
    - つまり、 $\rho$  より小さい |相関係数| はゼロセットされるというような感じ

PROPOSITION 1. *If we consider a  $2 \times 2$  problem defined only by two variables  $x_i$  and  $x_j$  ( $i \neq j$ ), the off-diagonal element of the optimal  $\Lambda$  as the solution to Eq. (4.7) is given by*

$$\Lambda_{i,j} = \begin{cases} -\frac{\text{sign}(r)(|r|-\rho)}{(1+\rho)^2-(|r|-\rho)^2} & \text{for } |r| > \rho \\ 0 & \text{for } |r| \leq \rho, \end{cases}$$

*where  $r$  is the correlation coefficient between the two variables.*

(T. Idé et al., "Proximity-Based Anomaly Detection using Sparse Structure Learning," SDM 2009.)

## 内容

---

- やりたいこと
- グラフィカル・ガウシアン・モデルと関連研究
- 疎構造学習の方法
- **相関異常度の定義**
- 実験結果
- まとめ

## GGMとして学習された確率モデルを使って、 各変数の異常度をKL距離として定義する

- データAとデータBを比べた時の、第*i*番目の変数のスコアの定義

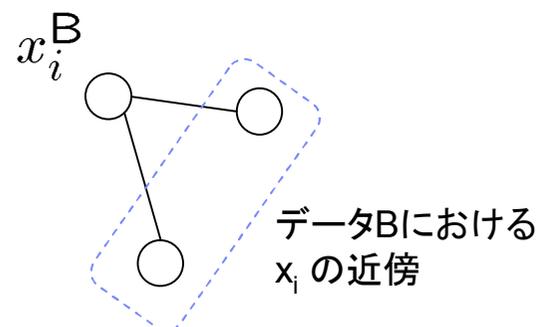
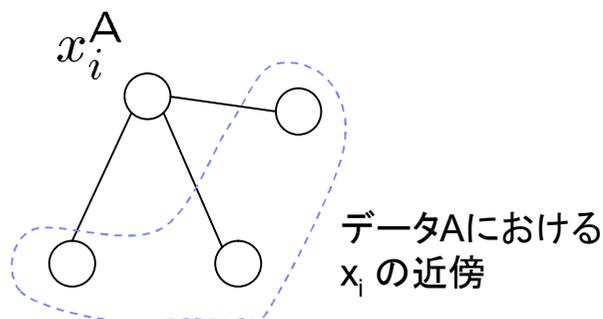
$$d_i^{AB} \equiv \int dz_i p_A(z_i) \int dx_i p_A(x_i|z_i) \ln \frac{p_A(x_i|z_i)}{p_B(x_i|z_i)}$$

$$z_i \equiv (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_M)^T \in \mathbb{R}^{M-1}.$$

条件付き分布同士の  
KL距離

- GGMの範囲では解析的に計算ができる

- ▶  $d_i^{AB} =$  ( $x_i$  の近傍グラフの次数の変化を表す項)  
+ ( $x_i$  の近傍グラフの密集度を表す項) + ( $x_i$  それ自身の分散の変化を表す項)



## 内容

---

- やりたいこと
- グラフィカル・ガウシアン・モデルと関連研究
- 疎構造学習の方法
- 相関異常度の定義
- **実験結果**
- まとめ

## 実験1: 共線形性が強いデータでの構造学習 実験の設定

---

- 為替レートデータ@UCR Archive

- ▶ いくつかの変数がほぼ完全相関

- ノイズを入れる前後における構造の変化を測定

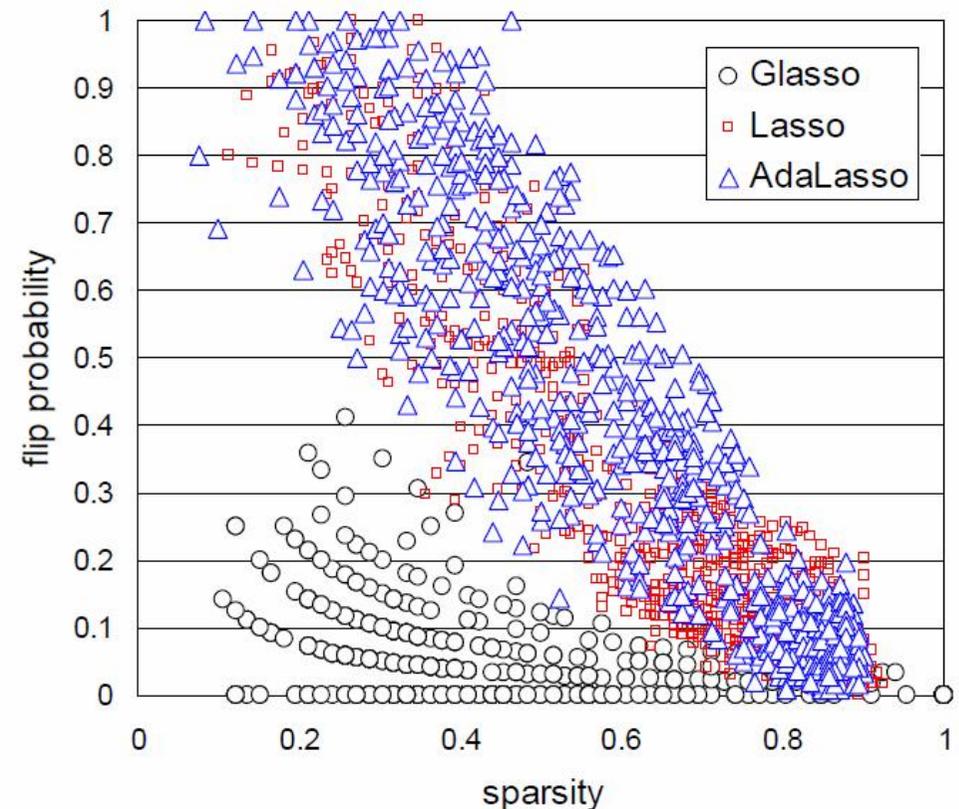
- ▶ データから構造学習
- ▶ 各変数に、標準偏差の10%分のノイズを混ぜてもう一度構造学習

- 比較した手法

- ▶ “Glasso”
  - Friedman, Hastie, & Tibshirani., Biostatistics, 2008
- ▶ “Lasso”
  - Meinshausen & Bühlmann, Ann. Stats. 2006
- ▶ “AdaLasso”
  - 上記のアルゴリズムにおいて、回帰をAdaptive Lasso [H. Zou, JASA, 2006] で行ったもの

## 実験1: 共線形性が強いデータでの構造学習: Graphical lassoアルゴリズムは、Lasso回帰に基づく他の構造学習法に比べて圧倒的にノイズに頑強である

- **sparsity:**
  - ▶ グラフがどれだけスパースか
- **flip prob.:**
  - ▶ ノイズ印加前後でどれだけ辺が変わるかの確率(辺の発生 or 消滅)
- **Meinshausen & Bühlmann の方法は、共線形性の下で結果が不安定**
  - ▶ Dempsterの伝統的な共分散構造選択の欠点を引き継いでいる
  - ▶ これはL1回帰で構造学習をやる際の避けがたい問題
    - 相関が強い変数の中のどれかひとつを強制的に選ぶので、どれが選択されるかはほとんど偶然による



## 実験2: sensor\_error データでの異常度のスコアリング 実験の設定

- **sensor\_error データ**
  - ▶ ある機械システムの実測定データ(M=44変数)
  - ▶ 79個の正常時データと20個の異常データ
  - ▶ 異常データでは、2つの変数が相関異常を呈している(右図)
  
- **79 × 20個の正常-異常ペアで異常検知をしてROC曲線を描かせる**
  - ▶ 2つの異常変数が常にトップ2を占めることを期待
    - この時、AUC (area under curve)はほぼ1となる

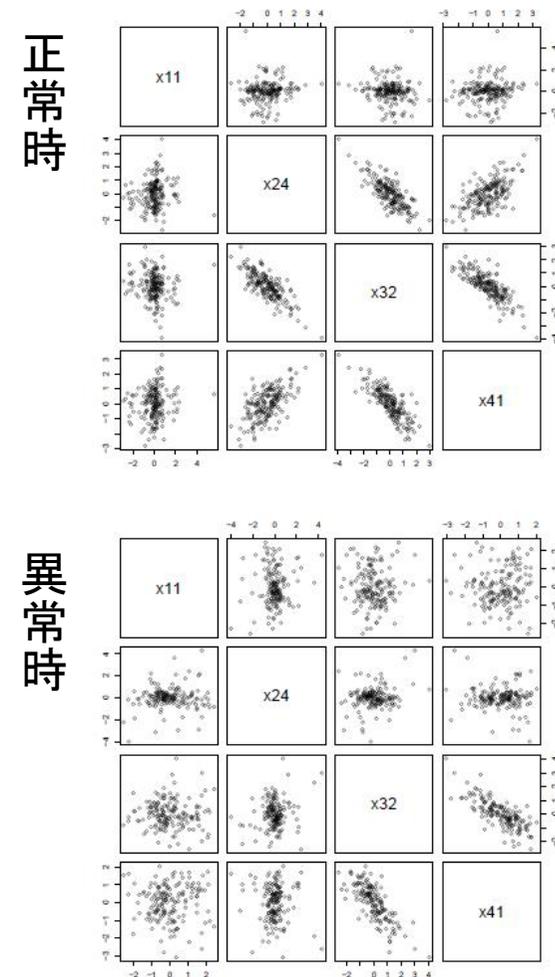


Figure 5: Pairwise scattering plot of *sensor\_error* data. Top: The 10th reference run. Bottom: The third faulty run.

## 実験2: sensor\_error データでの異常度のスコアリング 構造学習による近傍選択を組み込むことで、擬陽性を大幅に減らせる

### ■ 3つの別のスコアと比較

- ▶ 尤度に基づくもの
- ▶ 近傍グラフを素朴に k-NN法で作ったもの
- ▶ あるヒューリスティクスに基づいたスコア定義を用いたもの [Idé et al, ICDM 07]

### ■ KL距離によるスコアが最も良い成績

- ▶ しかも理論的に素性正しい

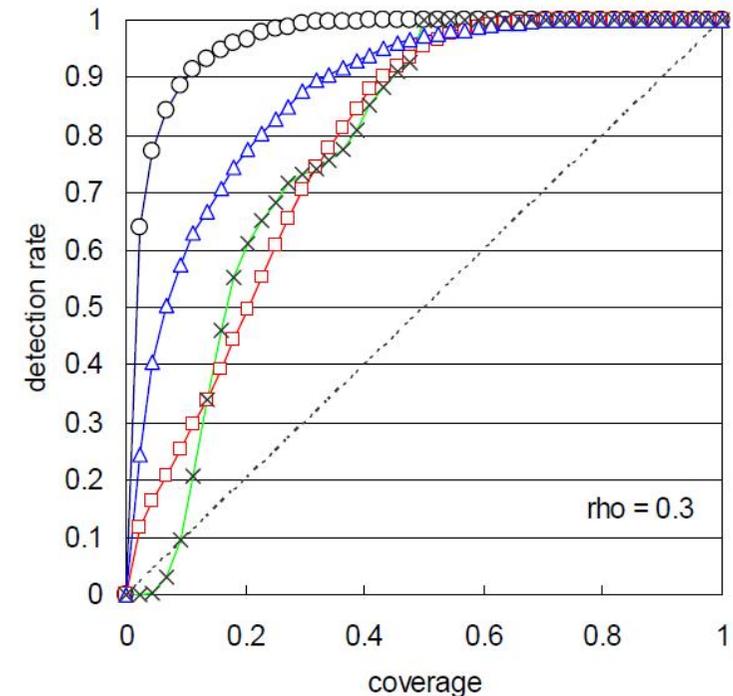


Table 3: Compared anomaly metrics and their best AUC values.

symbol	neighborhood	metric	best AUC
KL	Glasso	Eq. (5.17)	<b>0.96</b> ( $\rho = 0.3$ )
SNG	Glasso	Eq. (6.20)	0.93 ( $\rho = 0.7$ )
SNN	k-NN	Eq. (6.20)	0.87 ( $k = 2$ )
LR	Glasso	Eq. (6.21)	0.81 ( $\rho = 0.5$ )

Figure 6: ROC curves for  $\rho = 0.3$ , comparing KL ( $\circ$ ), SNG ( $\square$ ), SNN ( $\triangle$ ), and LR ( $\times$ ).

## 内容

---

- やりたいこと
- グラフィカル・ガウシアン・モデル
- 関連研究
- 疎構造学習の方法
- 相関異常度の定義
- 実験結果
- まとめ

## まとめ

---

- 相関異常のスコアリングという問題に対して、スパース構造学習を適用した
- 最近提案された疎構造学習の手法の比較検討を行い、代表的な手法と目される Meinshausen-Bühlmann の方法が、共線形性の下では破綻すること、また、精度行列をMAP解として求める方法はそのような弱点を持たないことを示した
- 疎なGGMに対して計算される条件付き期待KL距離を異常度尺度とすることにより、実問題において、相関異常の検知性能を顕著に上げられることを示した