

動的潜在グラフ構造の 動的・非動的成分への分解

原 聡, 鷺尾 隆

大阪大学産業科学研究所

第2回 Latent Dynamics Workshop @東大, 2011/06/22

変数間依存関係の解析

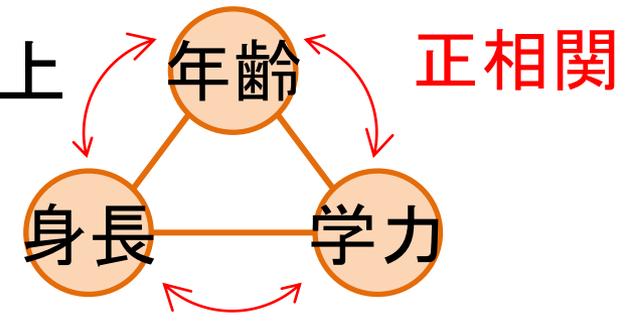
- 例：小学生の「身長，学力，年齢」のデータ
 - 相関の解析 → 小学生の身長と学力には**正の相関**
 - ◆ 「年齢」の影響 – 年齢と共に身長・学力は向上
 - ◆ 第3変数の影響により生じる**見せかけの依存関係**
- 変数間の本質的な依存関係
 - 調査対象を「同年齢の小学生」に限った場合，身長と学力に相関はあるか？
 - 条件を固定した上での比較
 - **本質的な関係の解析**

変数間依存性に基づく潜在グラフ構造

■ 相関 (表面的な関係)

- 年齢とともに身長, 学力は向上

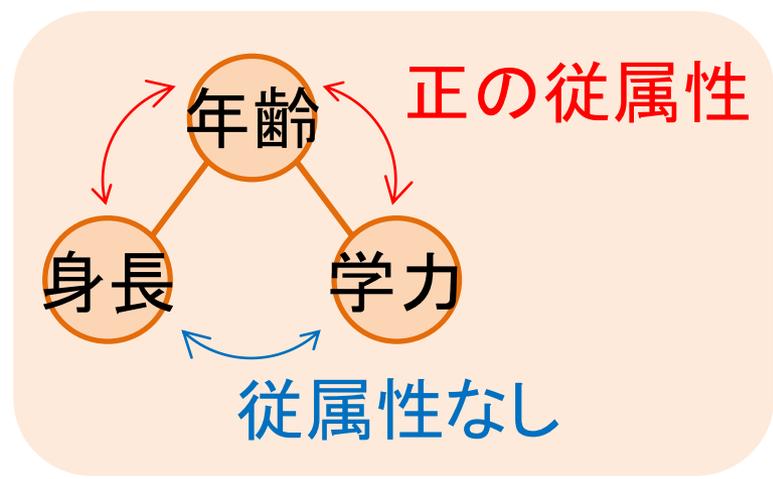
身長, 学力を比較する際に
年齢は可変



■ 条件付き従属性 (本質的な関係)

- 同年齢の学生間では
身長, 学力は無相関

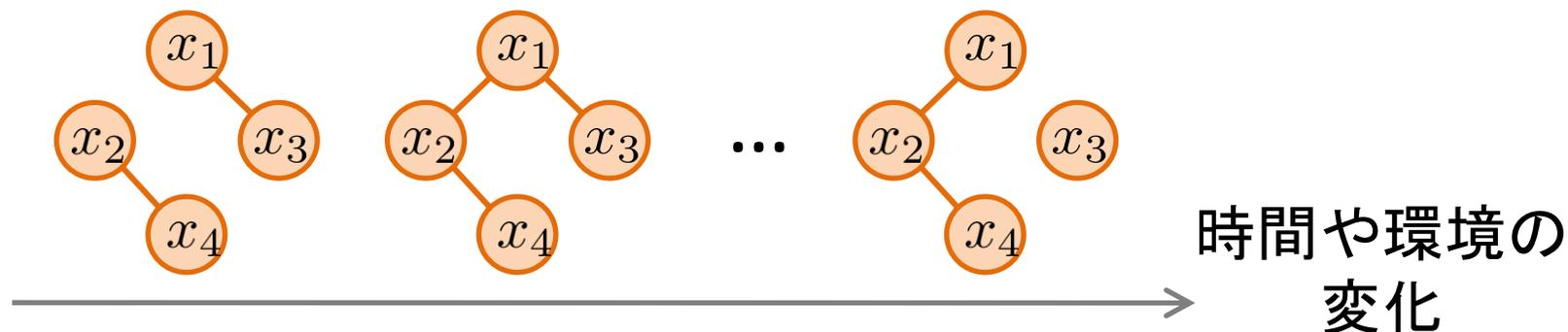
年齢を固定して比較
 → 年齢による影響を除外



潜在グラフ構造

動的な潜在グラフ構造

- データ生成機構の変動(時間発展や環境変化など)
 - 潜在グラフ構造も動的に変化

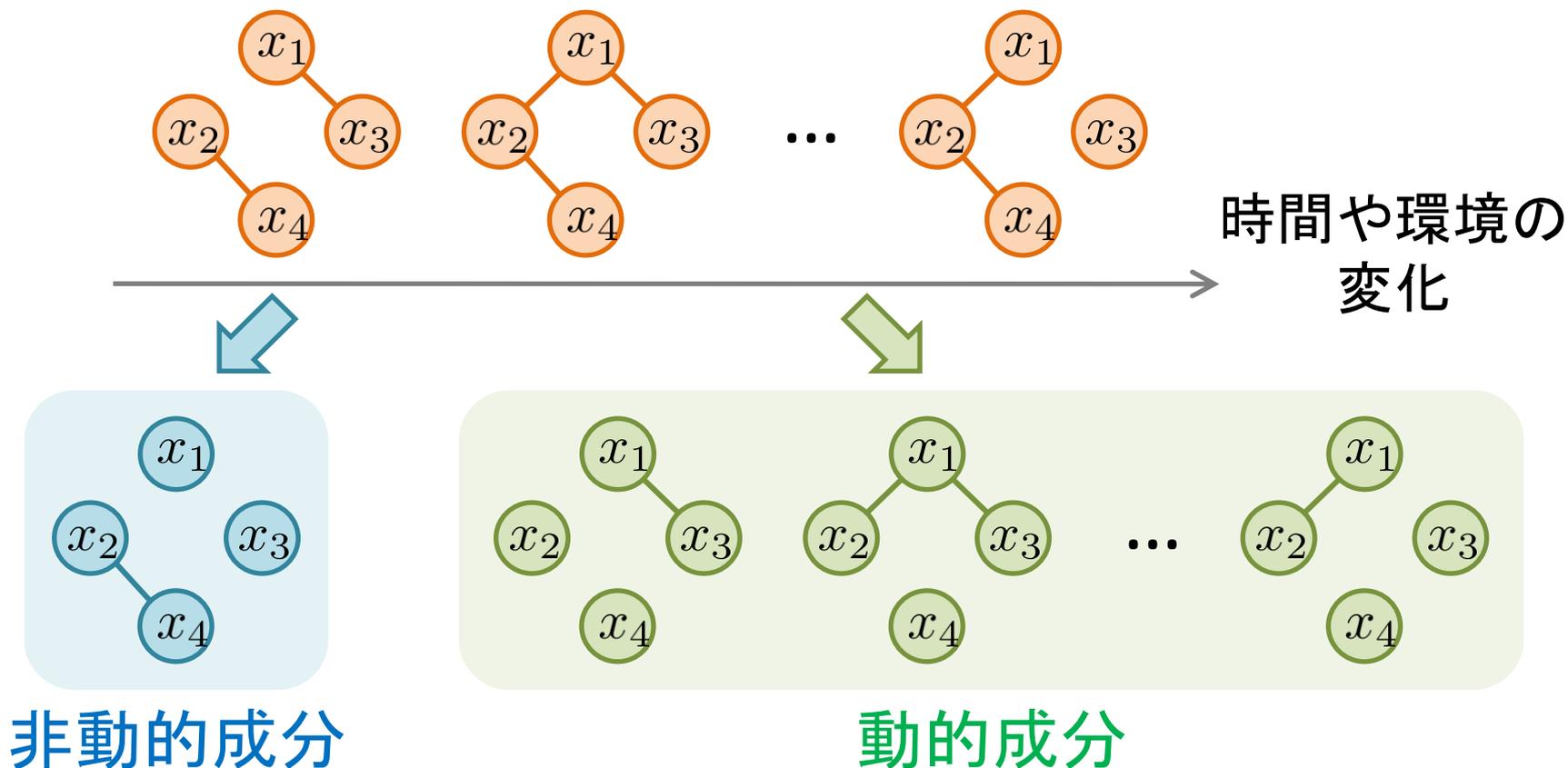


- 構造変化が部分的に起きるシナリオ
 - 工学システム: 部分的な故障が大半
 - ◆ 正常なセンサに関する構造は不変
 - 短い時間単位における変化
 - ◆ 小さな変動を不変構造として近似

本研究の目的

■ 潜在グラフ構造を動的・非動的成分へ分解

- 工学システム: 異常箇所の特定に有用
- データ生成機構の変化に関する情報の抽出



発表内容

- 潜在グラフ構造の動的・非動的成分への分解
- **既存研究（GGMの構造推定）**
- 共通部分構造推定問題
- 相関異常検知への応用
- まとめ

詳細は以下をご参照ください。

- Hara, S. and Washio, T.,
“Common Substructure Learning of Multiple Graphical Gaussian Models”, ECML PKDD 2011 (to appear)

グラフィカル・ガウシアン・モデル (GGM) ⁶

■ 変数間の条件付き独立性の基礎的なモデル

- 連続な確率変数 $\boldsymbol{x} = (x_1, x_2, \dots, x_d)^\top \in \mathbb{R}^d$ が正規分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$ に従う時,

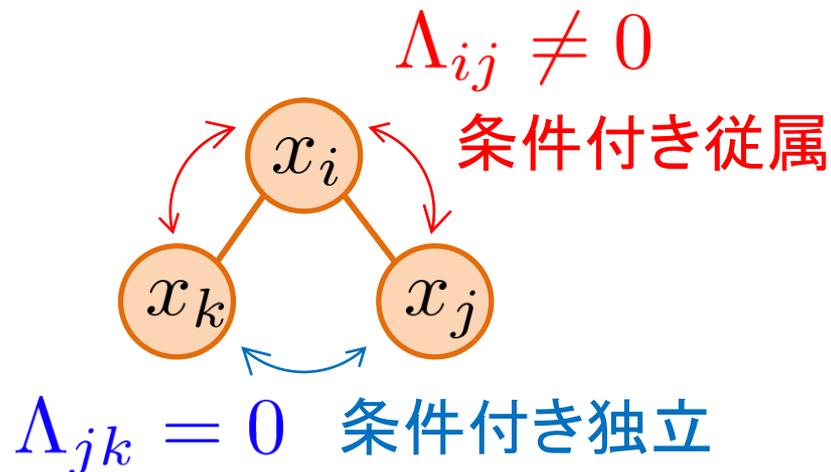
2変数 x_i, x_j が他の変数について条件付き独立

$$\Leftrightarrow \Lambda_{ij} = 0$$

- $\boldsymbol{\Lambda}$: 精度行列 (共分散行列 $\boldsymbol{\Sigma}$ の逆行列)

■ 依存構造のグラフ表現

- 頂点: 各変数
- 辺: 隣接行列 = $\boldsymbol{\Lambda}$



GGMの構造推定(共分散選択)

- 精度行列 Λ を推定する問題 (データの正規性を仮定)
 - 問題点: 最尤推定 $\hat{\Lambda} = \hat{\Sigma}^{-1}$ は密行列
 - GGMは完全グラフ ($\hat{\Sigma}$: 共分散の最尤推定)
 - 全ての変数が条件付き従属, 依存構造不明

- l_1 正則化最尤法 (Yuan et. al., Biometrika 2007, Banerjee et. al. JMLR 2008)

$$\max_{\Lambda \succ 0} \ell(\Lambda; \hat{\Sigma}) - \rho \|\Lambda\|_1 \quad l_1 \text{正則化による疎性の獲得}$$

- 正則化パラメータ $\rho > 0$
- $\ell(\Lambda; \hat{\Sigma}) = \log \det \Lambda - \text{Tr} \left(\hat{\Sigma} \Lambda \right)$: 正規分布の対数尤度
- 凸最適化問題, **GLassoアルゴリズム**
(Friedman et. al., Biostatistics 2008)

GGMの同時構造推定

■ マルチタスク構造推定 (Honorio et. al., ICML 2010)

- 共分散 $\hat{\Sigma}_1, \hat{\Sigma}_2, \dots, \hat{\Sigma}_N$ から疎な精度行列を推定
- 仮定: 全てのGGMは同じグラフ構造を持つ.

$$\max_{\{\Lambda_i; \Lambda_i \succ 0\}_{i=1}^N} \sum_{i=1}^N t_i \ell(\Lambda_i; \hat{\Sigma}_i) - \rho \sum_{j \neq j'} \max_{1 \leq i \leq N} |\Lambda_{i,jj'}|$$

- 同時構造が疎

$$\tilde{\Lambda}_{jj'} \equiv \max_{1 \leq i \leq N} |\Lambda_{i,jj'}| = 0 \Leftrightarrow \Lambda_{i,jj'} = 0, \quad \forall i$$

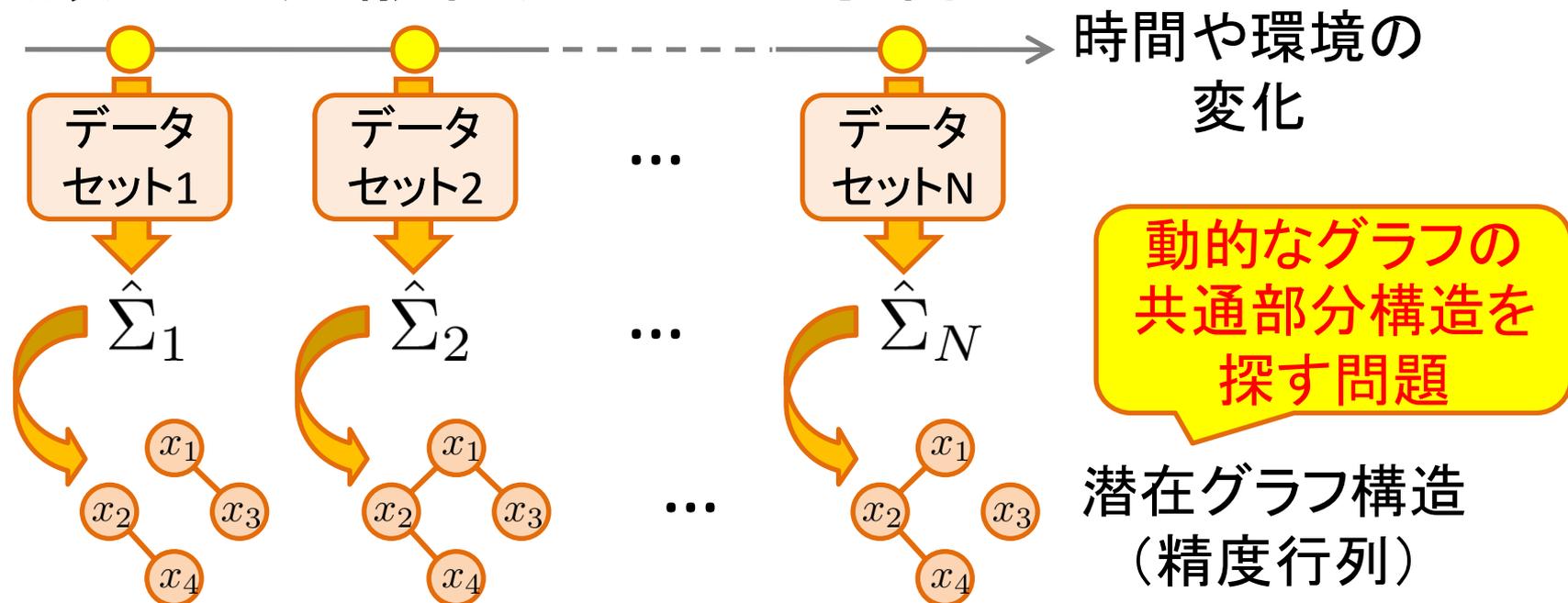
■ 零成分に関する情報を共有することで精度向上

発表内容

- 潜在グラフ構造の動的・非動的成分への分解
- 既存研究（GGMの構造推定）
- **共通部分構造推定問題**
- アルゴリズム
- 相関異常検知への応用
- まとめ

共通部分構造推定問題

■ 複数の共分散行列からの学習



- N 個の共分散行列 $\hat{\Sigma}_1, \hat{\Sigma}_2, \dots, \hat{\Sigma}_N$ から精度行列 $\Lambda_1, \Lambda_2, \dots, \Lambda_N$ の共通部分構造 (非動的な成分) を推定する問題

共通部分構造

■ 精度行列 $\Lambda_1, \Lambda_2, \dots, \Lambda_N$ の共通部分構造 Θ

$$\Theta_{jj'} = \begin{cases} \Lambda_{1,jj'}, & \text{if } \Lambda_{1,jj'} = \Lambda_{2,jj'} = \dots = \Lambda_{N,jj'} \\ 0, & \text{otherwise} \end{cases}$$

- 弱定常性の定義を偏共分散へと適用したもの

■ (j, j') 成分が共通部分構造

最大変動幅がゼロ

$$\Leftrightarrow \max_{1 \leq i, i' \leq N} |\Lambda_{i,jj'} - \Lambda_{i',jj'}| = 0$$

■ 提案法：最大変動幅を考慮した同時構造推定

問題設定

■ 最大変動幅に制約を加えた同時構造推定問題

$$\max_{\{\Lambda_i\}_{i=1}^N} \sum_{i=1}^N t_i \ell(\Lambda_i; \hat{\Sigma}_i) - \sum_{j \neq j'} \left(\rho \max_{1 \leq i \leq N} |\Lambda_{i,jj'}| + \gamma \max_{1 \leq i, i' \leq N} |\Lambda_{i,jj'} - \Lambda_{i',jj'}| \right)$$

s.t. $\Lambda_1, \Lambda_2, \dots, \Lambda_N \succ 0$

同時構造の
正則化

最大変動幅の
正則化

- 正則化パラメータ $\rho, \gamma > 0$, 非負重み $\sum_{i=1}^N t_i = 1$
- 凸最適化問題

■ 変動幅の正則化により共通部分構造が得られる

■ 変動を直接的にはモデル化しないアプローチ

既存研究との関連

■ 2データセット間での構造変化推定 (Zhang et. al., UAI 2010)

- Lassoを用いた手法 (Meinshausen et. al., Ann. Statist. 2006) に Fused Lasso型の正則化を加えた手法

■ 共通部分構造推定問題との関連

	提案法	Zhangらの方法
目的関数	正規分布の 正則化最尤推定	Lasso + Fused Lasso (近似的定式化)
データセットの数 N	$N \geq 2$	$N = 2$ 限定
アルゴリズム	$N \geq 2$	$N = 2$ 限定

より一般的な枠組み

発表内容

- 潜在グラフ構造の動的・非動的成分への分解
- 既存研究（GGMの構造推定）
- 共通部分構造推定問題
- **アルゴリズム**
- 相関異常検知への応用
- まとめ

アルゴリズム概要 (ブロック座標降下法)¹⁵

- 入力: 共分散行列 $\hat{\Sigma}_1, \hat{\Sigma}_2, \dots, \hat{\Sigma}_N$
正則化パラメータ $\rho, \gamma > 0$
非負重み $t_1, t_2, \dots, t_N, \sum_{i=1}^N t_i = 1$
- 出力: 精度行列 $\Lambda_1, \Lambda_2, \dots, \Lambda_N$

■ 初期化: $\Lambda_i \leftarrow \hat{\Sigma}_i^{-1} \quad (1 \leq i \leq N);$

■ 以下を収束するまで繰り返す;

For $j = 1$ to d , $j' = 1$ to d

$\Lambda_1, \Lambda_2, \dots, \Lambda_N$ の (j, j') 成分を更新;

End For

(j, j') 成分以外を
定数として扱う

ブロック座標降下法

■ 行列の要素を逐次的に最適化

- $\Lambda_1, \Lambda_2, \dots, \Lambda_N$ の (j, j') 成分に関する部分最適化
 ((j, j')成分以外は定数として扱う)

$$\Lambda_i = \begin{bmatrix} & v_i \\ v_i & \omega_i \end{bmatrix}$$

- 対角成分 $\omega = (\omega_1, \omega_2, \dots, \omega_N)^\top$ と
 非対角成分 $v = (v_1, v_2, \dots, v_N)^\top$ とで
 それぞれ異なる最適化問題

■ 共通部分構造推定は凸最適化かつブロック座標降下法により最適化への収束が保証される

(Tseng, JOTA 2001)

対角成分の最適化

■ 解析解

$$\omega_i = z_i^\top Z_i^{-1} z_i + q_i^{-1}$$

$$\Lambda_i = \begin{bmatrix} Z_i & z_i \\ z_i^\top & \omega_i \end{bmatrix} \quad \hat{\Sigma}_i = \begin{bmatrix} P_i & p_i \\ p_i^\top & q_i \end{bmatrix}$$

行列を (j, j) 成分と
それ以外に分割

■ 正定値性の保存

- Λ_i の初期値が正定ならば、ブロック座標降下法の各ステップにおいて常に正定値性は保証される。

今回は $\Lambda_i \leftarrow \hat{\Sigma}_i^{-1}$ により初期化

非対角成分の最適化

- (j, j') 成分のベクトル $\mathbf{v} = (v_1, v_2, \dots, v_N)^\top$ の最適化

$$\min_{\mathbf{v}} \frac{1}{2} \mathbf{v}^\top \text{diag}(\mathbf{a}) \mathbf{v} - \mathbf{b}^\top \mathbf{v} + \rho \|\mathbf{v}\|_\infty + \gamma \max_{1 \leq i, i' \leq N} |v_i - v_{i'}|$$

- $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$: (j, j') 成分以外及び $\hat{\Sigma}_i$ から決まる係数

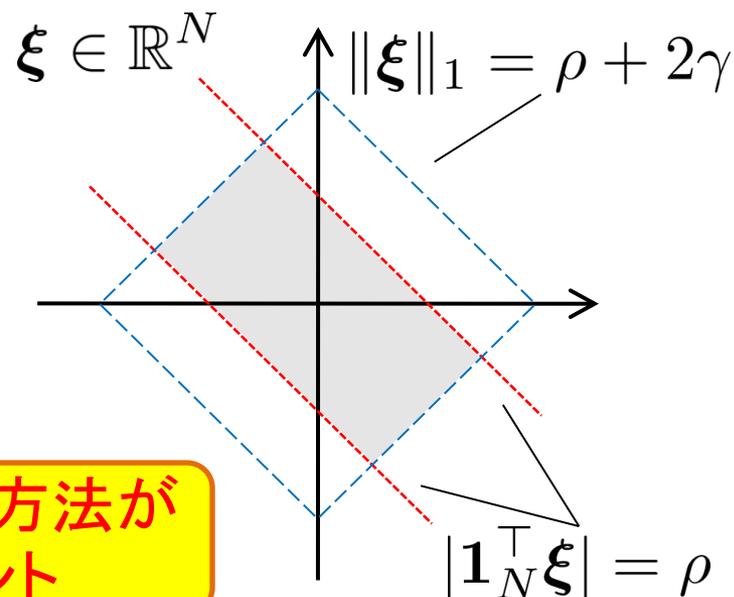
■ 双対問題

$$\min_{\boldsymbol{\xi}} \frac{1}{2} (\mathbf{b} - \boldsymbol{\xi})^\top \text{diag}(\mathbf{a})^{-1} (\mathbf{b} - \boldsymbol{\xi})$$

s.t. $|\mathbf{1}_N^\top \boldsymbol{\xi}| \leq \rho, \quad \|\boldsymbol{\xi}\|_1 \leq \rho + 2\gamma$

- $\boldsymbol{\xi} = \mathbf{b} - \text{diag}(\mathbf{a}) \mathbf{v}$

最適化の方法が
ポイント



双対問題の解

■ 双対問題

$$\min_{\xi} \frac{1}{2} (\mathbf{b} - \xi)^\top \text{diag}(\mathbf{a})^{-1} (\mathbf{b} - \xi)$$

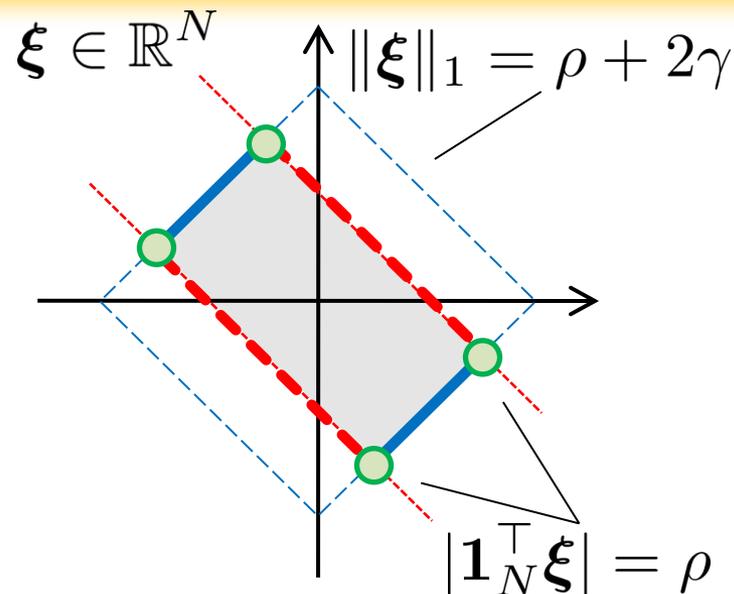
$$\text{s.t. } |\mathbf{1}_N^\top \xi| \leq \rho, \quad \|\xi\|_1 \leq \rho + 2\gamma$$

■ 自明な解

- $\xi = \mathbf{b}$ ($|\mathbf{1}_N^\top \mathbf{b}| \leq \rho, \|\mathbf{b}\|_1 \leq \rho + 2\gamma$ の時)

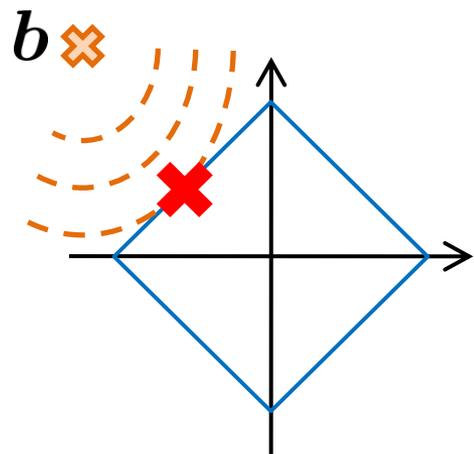
■ 境界上の解の場合分け

- 場合1: $\|\xi\|_1 = \rho + 2\gamma$ (実線部 —)
- 場合2: $|\mathbf{1}_N^\top \xi| = \rho$ (破線部 ---)
- 場合3: $\|\xi\|_1 = \rho + 2\gamma, |\mathbf{1}_N^\top \xi| = \rho$ (交点 ○)



双対問題の解法(概略1)

- 場合1: 解は境界 $\|\xi\|_1 = \rho + 2\gamma$ の上

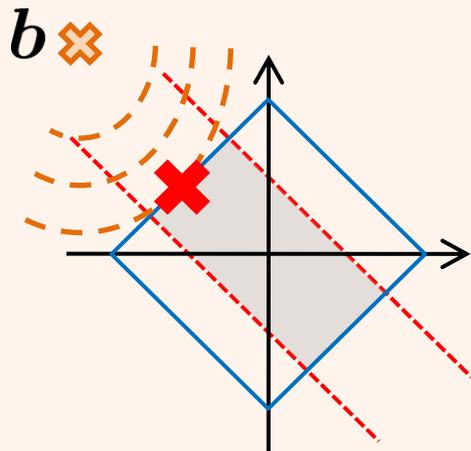


連続2次ナップザック問題 ($\xi_i = \text{sgn}(b_i)y_i$)

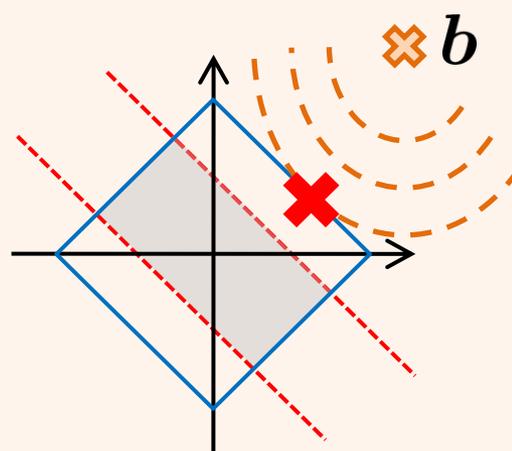
$$\begin{aligned} \min_{\mathbf{y}} \quad & \sum_{i=1}^N \frac{1}{2a_i} (|b_i| - y_i)^2 \\ \text{s.t.} \quad & \mathbf{y} \geq 0, \mathbf{1}_N^\top \mathbf{y} = \rho + 2\gamma \end{aligned}$$

効率的
解法が
存在

- 最適解



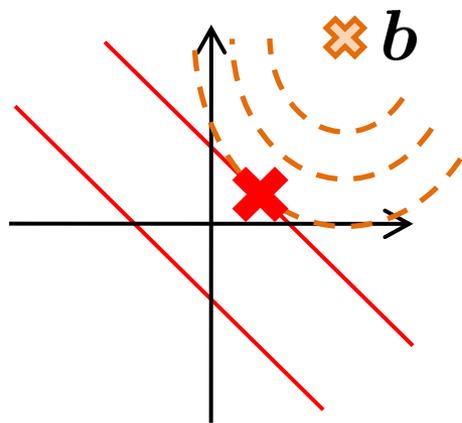
- 最適解でない場合



$$\begin{aligned} |\mathbf{1}_N^\top \xi| &\geq \rho \\ \rightarrow \text{場合2へ} \end{aligned}$$

双対問題の解法(概略2)

■ 場合2: 解は境界 $|\mathbf{1}_N^\top \boldsymbol{\xi}| = \rho$ の上



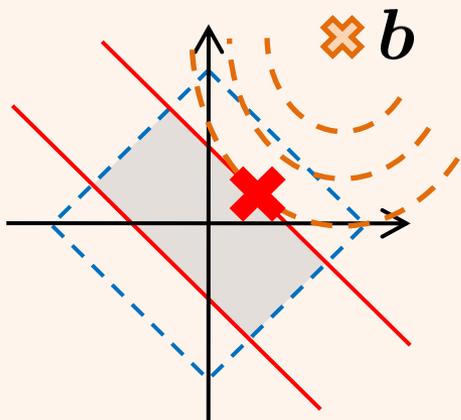
1変数Lasso ($\boldsymbol{\xi} = \mathbf{b} - v_0 \mathbf{a}$)

$$\min_{v_0} \frac{a}{2} v_0^2 - b v_0 + \rho |v_0|$$

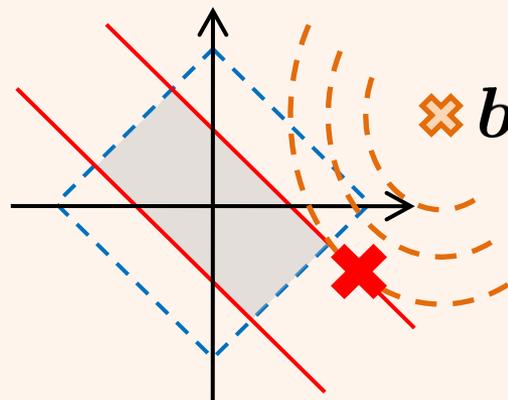
$$a = \mathbf{1}_N^\top \mathbf{a}, \quad b = \mathbf{1}_N^\top \mathbf{b}$$

解析解が
存在

■ 最適解



■ 最適解でない場合

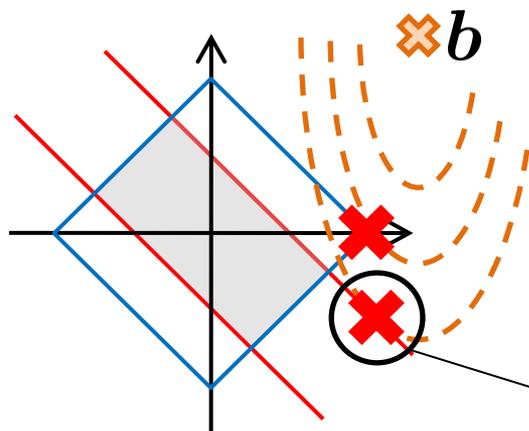


$$\|\boldsymbol{\xi}\|_1 \geq \rho + 2\gamma$$

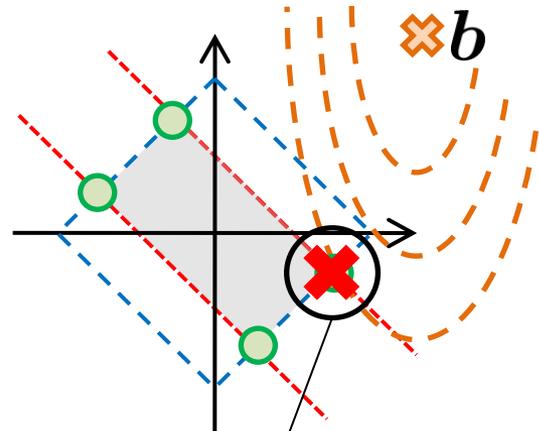
→ 場合3へ

双対問題の解法 (概略3-1)

- 場合3: 解は境界 $\|\xi\|_1 = \rho + 2\gamma$, $|\mathbf{1}_N^\top \xi| = \rho$ の上



場合1, 2共に
解ではない



解の位置は異なるが符号は同じ

- 場合2の解 $\tilde{\xi}$ を用いて $\tilde{\xi}_i \geq 0$ と $\tilde{\xi}_i < 0$ に関する2つの問題へ分解

双対問題の解法 (概略3-2)

■ 最適化問題

$$\min_{\boldsymbol{\xi}} \frac{1}{2} (\mathbf{b} - \boldsymbol{\xi})^\top \text{diag}(\mathbf{a})^{-1} (\mathbf{b} - \boldsymbol{\xi}) \quad \text{s.t.} \quad |\mathbf{1}_N^\top \boldsymbol{\xi}| = \rho, \quad \|\boldsymbol{\xi}\|_1 = \rho + 2\gamma$$

■ 等価な連続2次ナップザック問題

$$\min_{\mathbf{y}^+} \sum_{\tilde{\xi}_i \geq 0} \frac{1}{2a_i} (y_i^+ - b_i)^2 \quad \text{s.t.} \quad \mathbf{y}^+ \geq 0, \quad \mathbf{1}_N^\top \mathbf{y}^+ = \alpha^+$$

$$\min_{\mathbf{y}^-} \sum_{\tilde{\xi}_i < 0} \frac{1}{2a_i} (y_i^- - b_i)^2 \quad \text{s.t.} \quad \mathbf{y}^- \geq 0, \quad \mathbf{1}_N^\top \mathbf{y}^- = \alpha^-$$

- $\xi_i = y_i^+$ ($\tilde{\xi}_i \geq 0$) 及び $\xi_i = -y_i^-$ ($\tilde{\xi}_i < 0$)
- $(\alpha^+, \alpha^-) = (\rho + \gamma, \gamma)$ または $(\alpha^+, \alpha^-) = (\gamma, \rho + \gamma)$

発表内容

- 潜在グラフ構造の動的・非動的成分への分解
- 既存研究（GGMの構造推定）
- 共通部分構造推定問題
- アルゴリズム
- **相関異常検知への応用**
- まとめ

相関異常検知への応用

■ 異常検知

- 各変数が2つのデータセット間の差異にどれだけ寄与しているかを定量的に評価

■ 相関異常 (Idé et al., SDM 2009)

- 各変数の条件付き分布の差異をKLダイバージェンスで評価
- 疎な精度行列の利用を提案 – ノイズの影響を低減

■ 共通部分構造推定の利用

- **正常な変数間の依存構造は共通部分構造**
- 相関異常の検知性能改善が見込める

実データへの適用

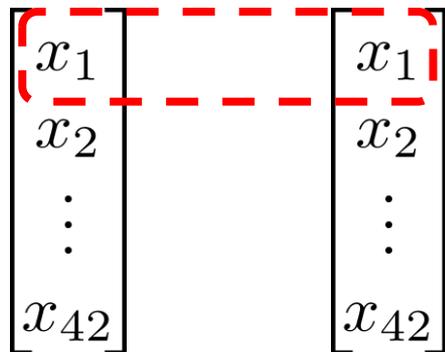
■ 自動車センサデータ (Idé et al., SDM 2009)

- センサ数42
- 正常状態で79試行, 異常状態で20試行
- 相関異常: 24番と25番のセンサが逆に接続

各試行につき
共分散行列1つ

■ 相関異常度スコア (Idé et al., SDM 2009)

- 各変数の条件付き分布の差異 (KLダイバージェンス)



Dataset 1

Dataset 2

$$a_j = \max(d_j^{12}, d_j^{21})$$

$$d_j^{12} = \int D_{\text{KL}}[p_1(x_j | \mathbf{x}_{\setminus j}) || p_2(x_j | \mathbf{x}_{\setminus j})] p_1(\mathbf{x}_{\setminus j}) d\mathbf{x}_{\setminus j}$$

実験設定

- 25個の共分散行列（正常20個，異常5個）で実験

1. 25個の精度行列を推定

- 比較対象
- GLasso (ℓ_1 正則化最尤推定, Friedman)により個別推定
 - 既存の同時構造推定法 (MSL, Honorio)により同時推定
 - 提案法により同時推定

重みは2状態のデータ
セット数を考慮して設定

2. 得られた精度行列について異常度スコアを計算

- 20×5 通りのペアについて計算した結果を平均
- スコアと閾値の比較により異常センサを検出

実験結果（検知性能）

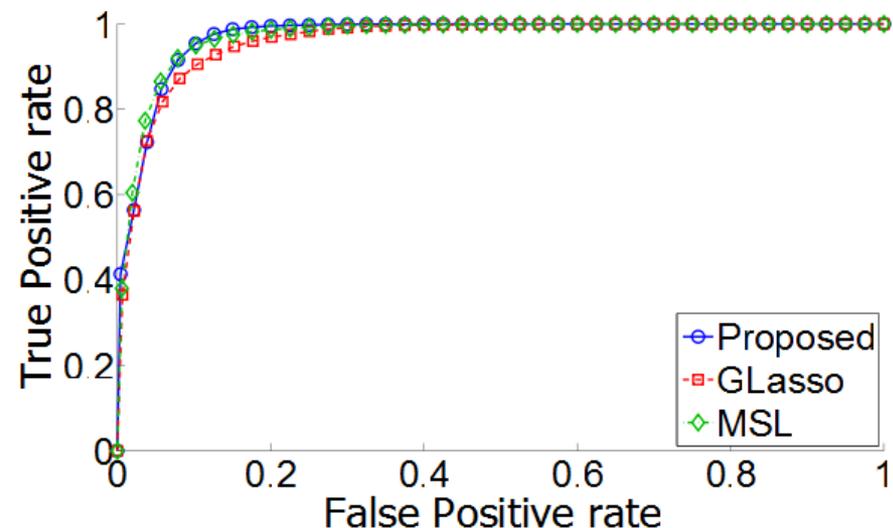
■ 行列の選択をランダムに100回繰り返し

- 正則化パラメータ ρ は0.05~0.30
- 提案法の γ はヒューリスティックにより選択

(Hara et. al., ECML 2011)

■ スコアの閾値を変化させてROC曲線を描画

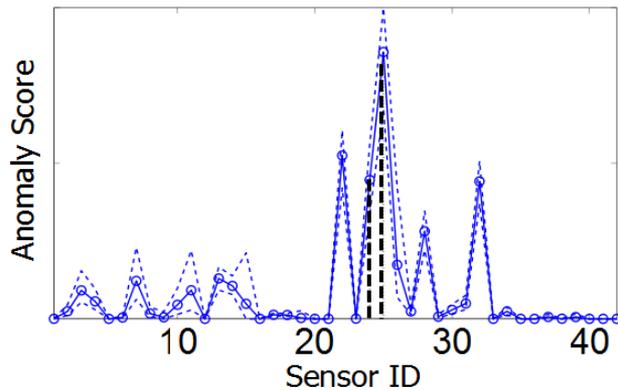
	Best AUC	ρ
提案法	0.97	0.05
GLasso	0.96	0.20
MSL	0.97	0.05



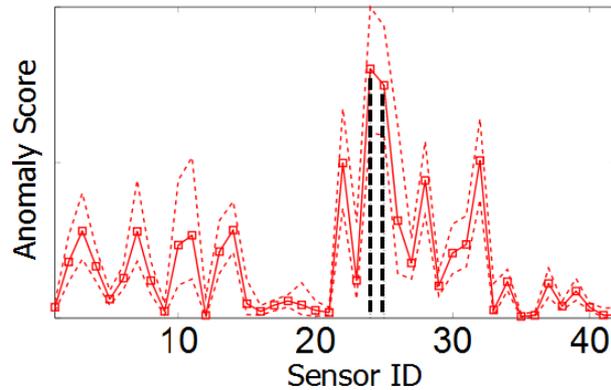
実験結果 (異常スコア1)

- 正常-異常状態間 (100回の中央値, 25%・75%点)

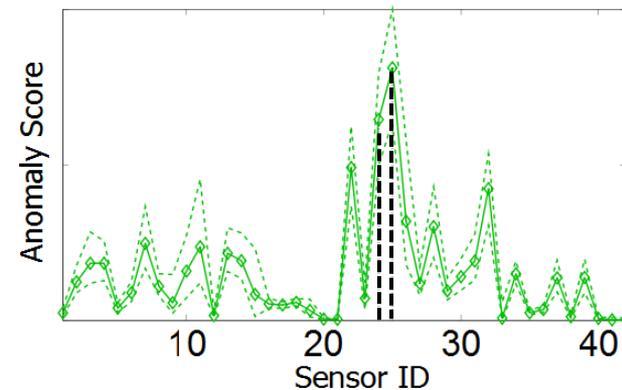
提案法



GLasso



MSL

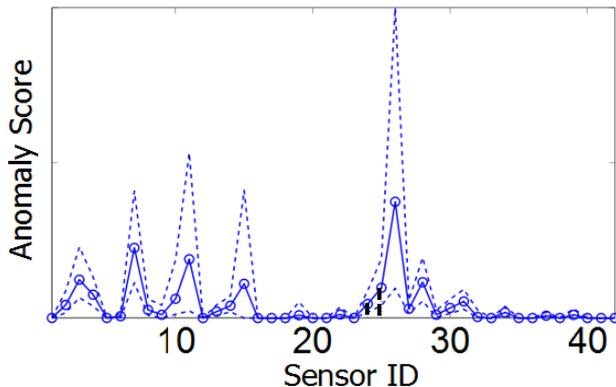


- 共通部分構造推定は正常センサについて既存法に比して**低い異常度スコア**
- 他2手法より**異常度スコアのばらつきも低い**
→ **安定した推定結果**

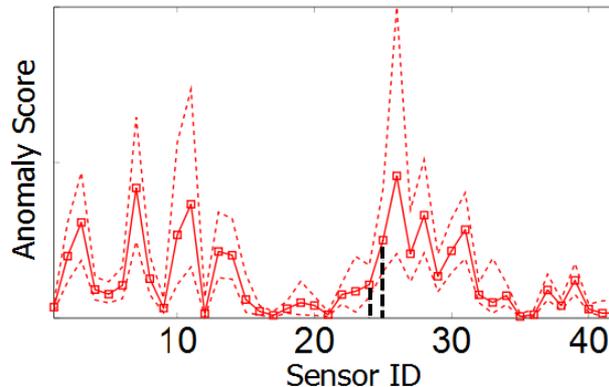
実験結果 (異常スコア2)

- 正常-正常状態間 (100回の中央値, 25%・75%点)

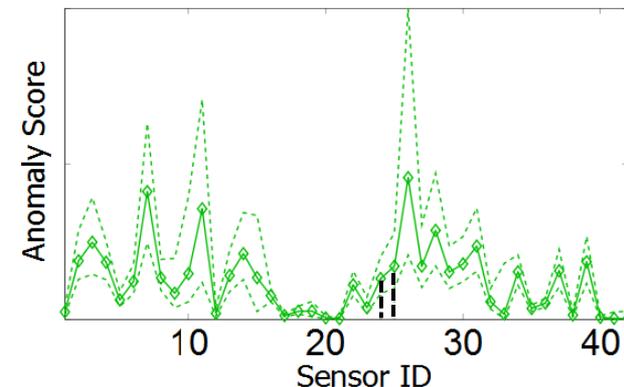
提案法



GLasso



MSL



- 正常-異常間と同等の傾向

- 提案法が他2手法よりも**低スコア, 低分散**

- 正常-正常間では理想的には"全スコア=0"

- 状態非依存のセンサ固有の変動が存在
- 正常-異常間との対比により異常候補の選択が可能

まとめ

- 潜在グラフ構造の動的・非動的成分への分解
 - 共通部分構造推定問題として定式化
 - ブロック座標降下法による最適化アルゴリズム
 - 各部分問題を解の場合分けにより効率的に計算
- 相関異常検知への応用
 - データセット間での異常変数の発見
 - 自動車センサデータへの応用
 - 提案法は従来法同様の検知性能
 - 異常スコアは従来法よりも安定, 解釈性も向上