

TOSHIBA

第4回 マテリアルズインフォマティクスセミナー

スパースモデリングとその展開

高田 正彬

株式会社 東芝 研究開発センター

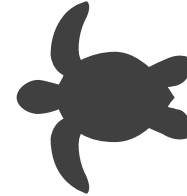
Who Are You?

- 高田 正彬 (たかだ まさあき)
 - (株)東芝 研究開発センター 知能化システム研究所 システムAIラボラトリー 研究主務
 - 統計数理研究所 ものづくりデータ科学研究センター 客員准教授
- 主な研究テーマ
 - 高次元データにおける統計的機械学習
- 主な成果
 - 人工知能学会 現場イノベーション賞 金賞 2016
 - 東京電力ホールディングス株式会社主催 第1回電力需要予測コンテスト 最優秀賞 2017
 - AISTATS2018
 - IJCAI2019 (本日お話しするトピック)
 - NeurIPS2020, IBIS2020優秀発表賞



目次

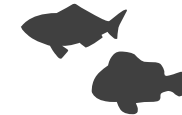
I スパースモデリングの基本を理解する



II スパースモデリングの全体を俯瞰する



III スパースモデリングを深く掘り下げる



研究① 高欠測データのためのスパースモデリング (Takada, Fujisawa & Nishikawa 2019 IJCAI)

研究② 定期解析のためのスパースモデリング (Takada & Fujisawa 2020 NeurIPS)

研究①②は統計数理研究所の藤澤洋徳教授との共同研究です。

L1正則化によりスパースな推定値が得られる



Lasso [Tibshirani 96]

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

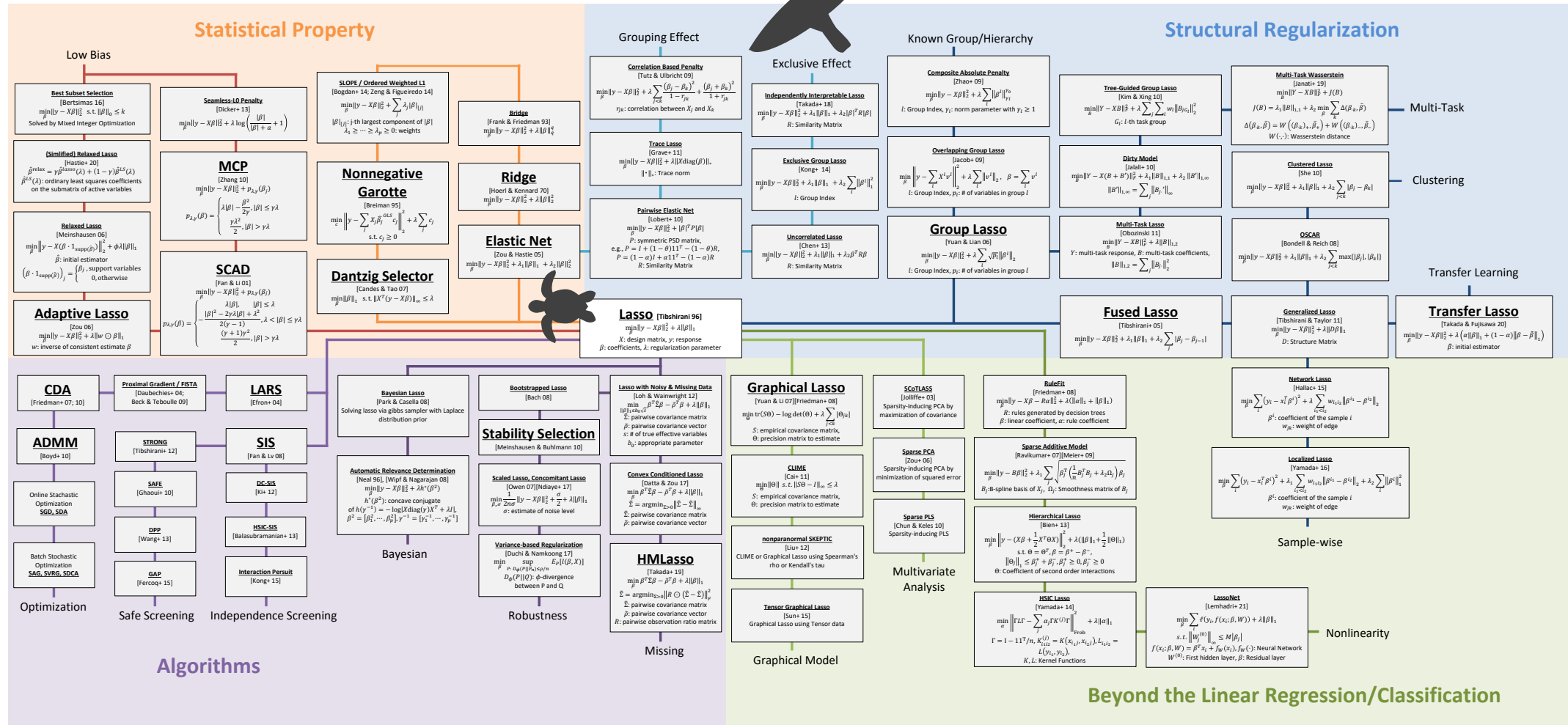
X : design matrix, y : response

β : coefficients, λ : regularization parameter

本日のダイジェスト II : スパースモデリングの全体を俯瞰する

スパースモデリングは多様な方向性で発展している

Sparse Modeling Landscape Map

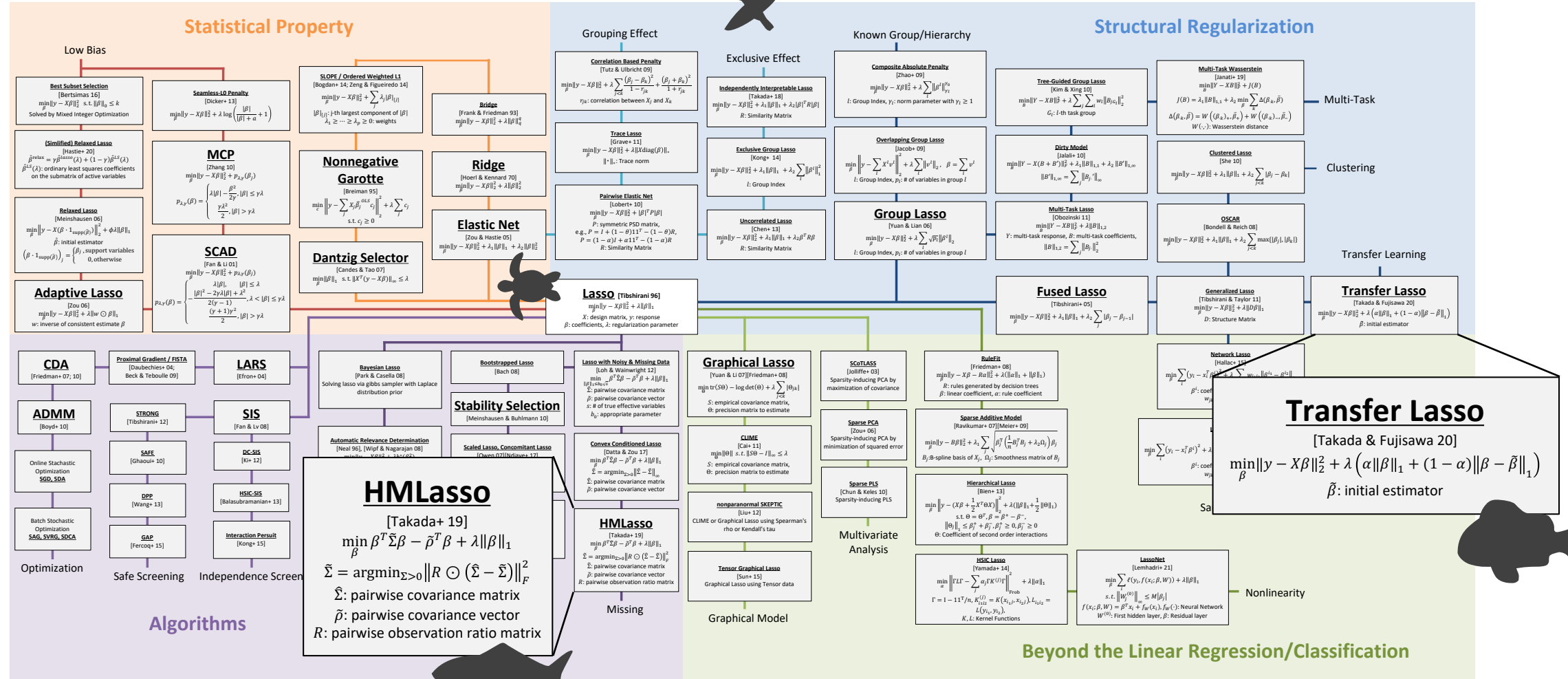


Beyond the Linear Regression/Classification

本日のダイジェストⅢ：スパースモデリングを深く掘り下げる

欠測データや定期解析のための手法を開発した

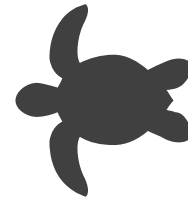
Sparse Modeling Landscape Map



Beyond the Linear Regression/Classification

I

スパースモデリングの基本を理解する



Lasso [Tibshirani 96]

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

X : design matrix, y : response

β : coefficients, λ : regularization parameter

スパースモデリング／スパース推定とは

“パラメータ推定”と“変数選択”を同時に実行する方法論

例：線形回帰モデル

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

スパース推定

↓ ↓ ↓ ↓

推定値： 0 1 0 -0.5 ……値がスパース（疎）

スパースモデル : $y = x_2 - 0.5 x_4$

有効でないパラメータをゼロと推定することで、
パラメータ推定と変数選択を同時に実行する

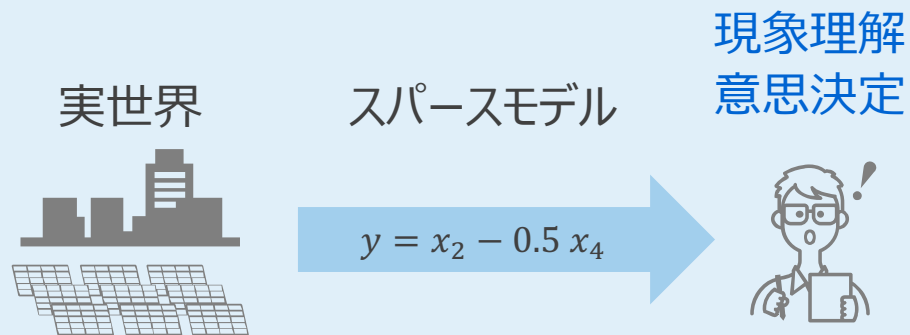
なぜスパースにするとよいのか？

1 モデルの解釈性

ノイズ変数を除去し，重要な変数を同定できる（オッカムの剃刀）

⇒ 少数の変数のみ現れ，解釈しやすい

⇒ 現象理解・意思決定に有効



2 高次元データでの性能

高次元・小標本データは多く存在するが，モデルの推定が一般に難しい

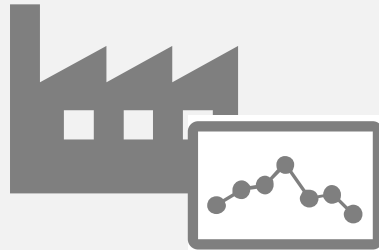
⇒ 真のモデルがスパースと仮定することで，モデルを精度良く推定できる



スパースモデリングが活躍する例

工場・プラントの品質管理

多くのセンサデータを用いて品質特性をモデル化し，不良要因分析／ソフトセンサー／制御／異常検知等を行う。

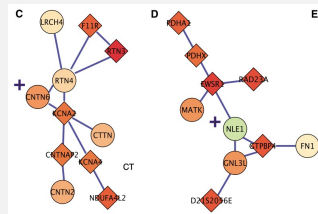


$$\text{品質特性} = \beta_1 * \text{センサ1} + \beta_2 * \text{センサ2} + \dots$$

遺伝子解析

(Jacob+ 2009 ICML)

遺伝子の事前知識を活用しつつ，多くの遺伝子の中から特定の疾患に関連する遺伝子を同定する。



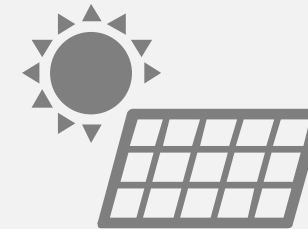
遺伝子のサブネットワーク
(Chuang 2007 Mol. Syst. Biol.)

$$\text{logit(がん有無)} = \beta_1 * \text{遺伝子1} + \beta_2 * \text{遺伝子2} + \dots$$

新材料の開発

(Ishida+ 2016 ACS Photonics)

材料特性やその非線形変換の特徴量の中から，太陽電池の電池効率に寄与する因子を同定する。



$$\text{電池効率} = \beta_1 * \log(\text{特性1} * \text{特性2}) + \beta_2 * \text{特性1} + \dots$$

ブラックホールの撮像

(Akiyama+ 2019 ApJL)

電波干渉計の観測量から画像を復元する。



ブラックホールシャドウの画像
出典：EHT Collaboration

$$\text{観測量} = \text{画素1} * \text{基底1} + \text{画素2} * \text{基底2} + \dots$$

線形回帰モデルの復習

線形回帰モデル

説明変数 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p] \in \mathbb{R}^{n \times p}$, 目的変数 $\mathbf{y} \in \mathbb{R}^n$, 回帰係数 $\beta \in \mathbb{R}^p$, ノイズ $\varepsilon \in \mathbb{R}^n$ とする.

$$\mathbf{y} = \beta_1 \mathbf{x}_1 + \dots + \beta_p \mathbf{x}_p + \varepsilon = \mathbf{X}\beta + \varepsilon$$

一般性を失うことなく, \mathbf{y} は中心化, \mathbf{X} は標準化されているものとする.

最小二乗法

$$\|\mathbf{v}\|_2^2 := \sum_i v_i^2$$

$$\begin{aligned} \hat{\beta} &= \operatorname{argmin}_{\beta} \sum_{i=1}^n (y_i - (\beta_1 x_{i1} + \dots + \beta_p x_{ip}))^2 = \operatorname{argmin}_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 \\ &= \underline{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}} \end{aligned}$$

二乗誤差

- 高次元データの場合 ($n \ll p$)
- 変数間の相関が高い場合 (多重共線性) $\Rightarrow (\mathbf{X}^T \mathbf{X})^{-1}$ が計算できない

高次元データにおける回帰モデル：Ridge と Lasso

Ridge (Hoerl & Kennard 1970)

二乗誤差

L2正則化

$$\begin{aligned}\hat{\beta} &= \operatorname{argmin}_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_2^2 \\ &= (\mathbf{X}^T \mathbf{X} + \lambda I)^{-1} \mathbf{X}^T \mathbf{y}\end{aligned}$$

$$\|\beta\|_2^2 := \sum_j \beta_j^2$$

- ✓ $\lambda > 0$ のとき、 $\mathbf{X}^T \mathbf{X} + \lambda I$ は常に逆行列をもつ。
⇒ 高次元データを扱える。

Lasso (Tibshirani 1996 JRSSB)

二乗誤差

L1正則化

$$\hat{\beta} = \operatorname{argmin}_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1$$

$$\|\beta\|_1 := \sum_j |\beta_j|$$

- ✓ 高次元データに対して、モデルがスパースになる。
- ✓ 凸最適化であるため、大域的最適解が容易に求められる。

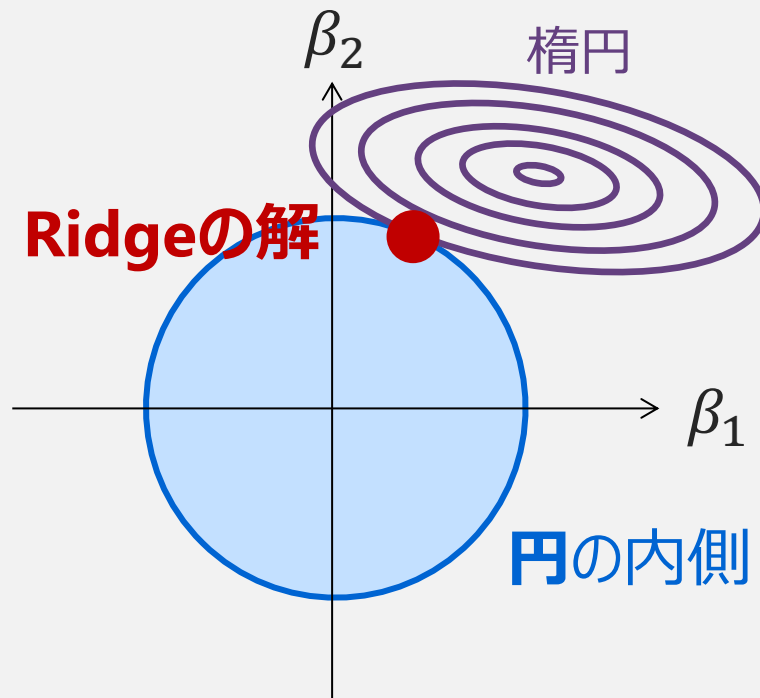
Lasso はなぜスパースになるのか？

L1ノルムが座標軸上で尖るため、モデルがスパースになる

Ridge (2次元の場合の制約条件による定式化)

$$\hat{\beta} = \operatorname{argmin}_{\beta} \|y - \beta_1 \mathbf{x}_1 - \beta_2 \mathbf{x}_2\|_2^2$$

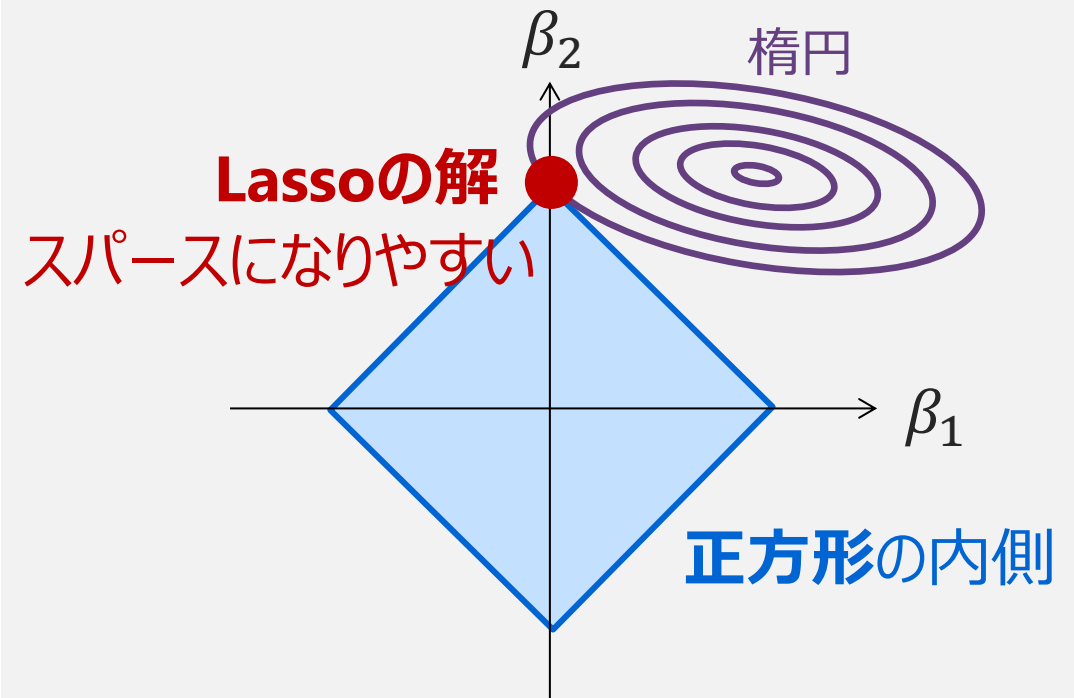
s. t. $\beta_1^2 + \beta_2^2 \leq t$ **L2ノルム**



Lasso (2次元の場合の制約条件による定式化)

$$\hat{\beta} = \operatorname{argmin}_{\beta} \|y - \beta_1 \mathbf{x}_1 - \beta_2 \mathbf{x}_2\|_2^2$$

s. t. $|\beta_1| + |\beta_2| \leq t$ **L1ノルム**



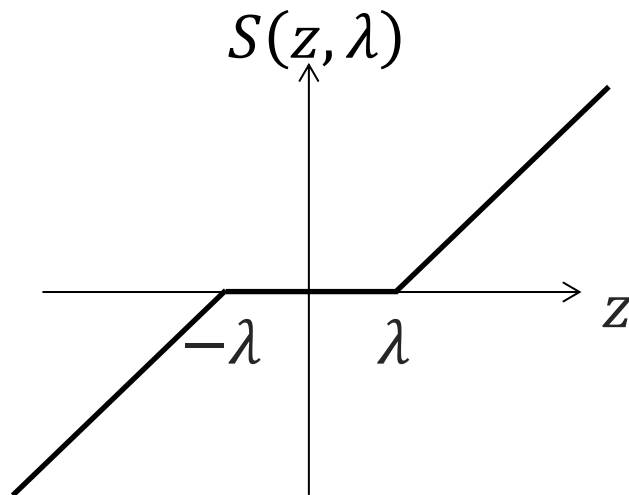
Lasso の最適化アルゴリズム

座標降下法により，大域的最適解が簡単・高速に求められる

座標降下法：

 $j = 1, \dots, p, 1, \dots, p, \dots$ に対して，収束するまで以下を繰り返す。

$$\beta_j \leftarrow S\left(\frac{1}{n} \mathbf{X}_j^T (\mathbf{y} - \mathbf{X}_{-j} \beta_{-j}), \lambda\right)$$



軟閾値関数 S は
一部区間で値0をとる
 \Rightarrow スパースな解が得られる

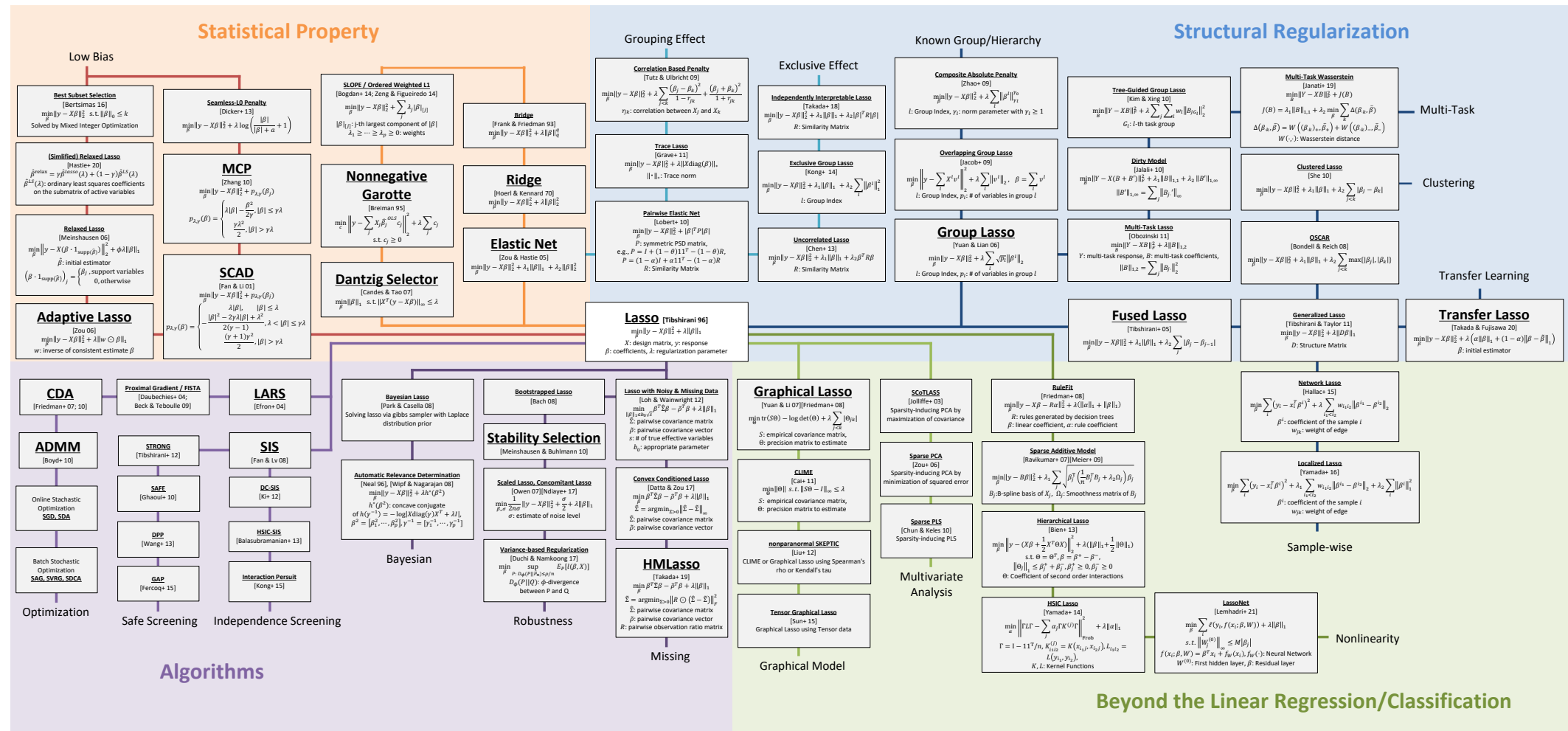
Rによる実行例：

```
> library(glmnet) #パッケージ読み込み
> data(QuickStartExample) #データ読み込み
> cvfit <- cv.glmnet(x,y) #Lasso実行
> coef(cvfit, s="lambda.min") #係数出力
(Intercept) 0.14867414
v1          1.33377821
v2          .
v3          0.69787701
v4          .
v5         -0.83726751
v6          0.54334327
v7          0.02668633
v8          0.33741131
v9          .
v10         .
```


スパースモデリング マップ

スパースモデリングの研究方向性は、大きく4領域に分かれる

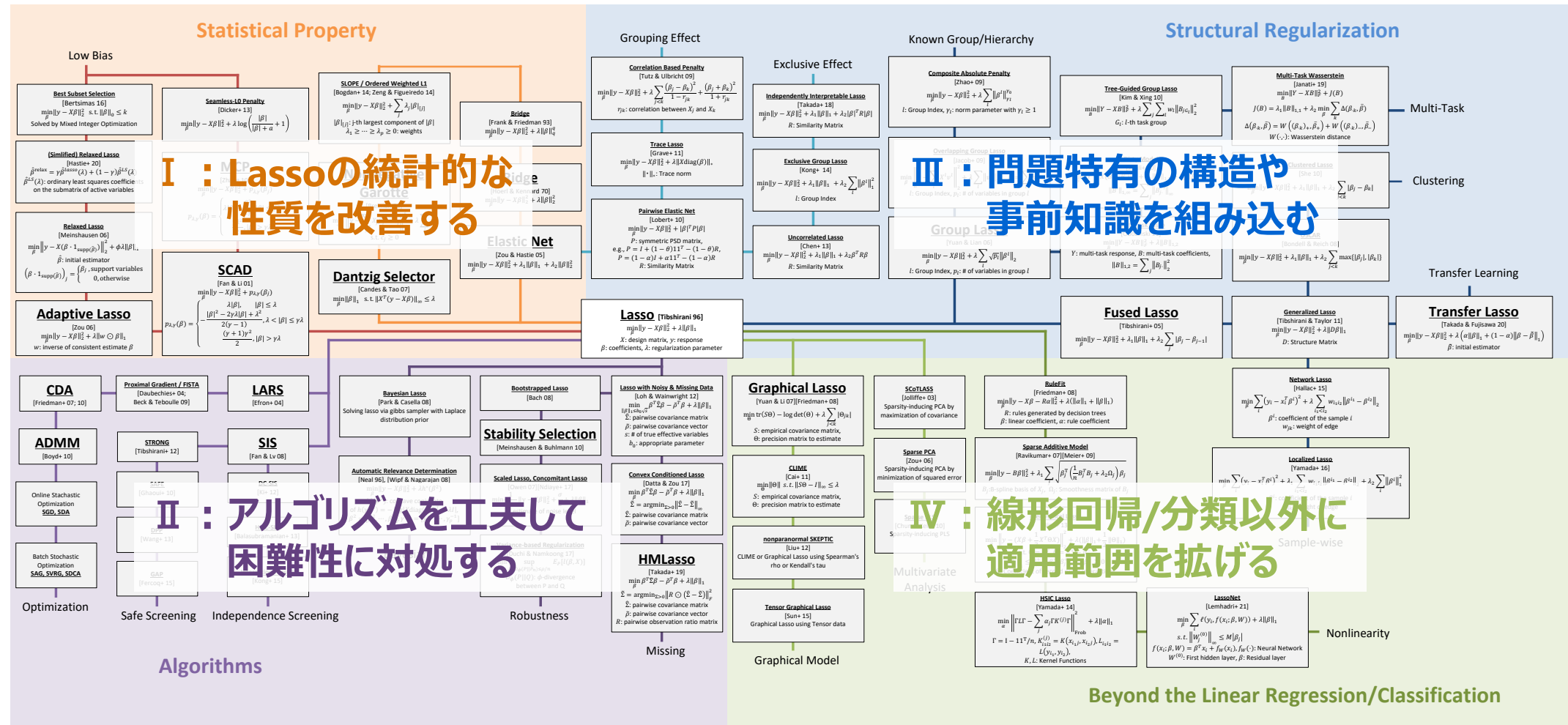
Sparse Modeling Landscape Map



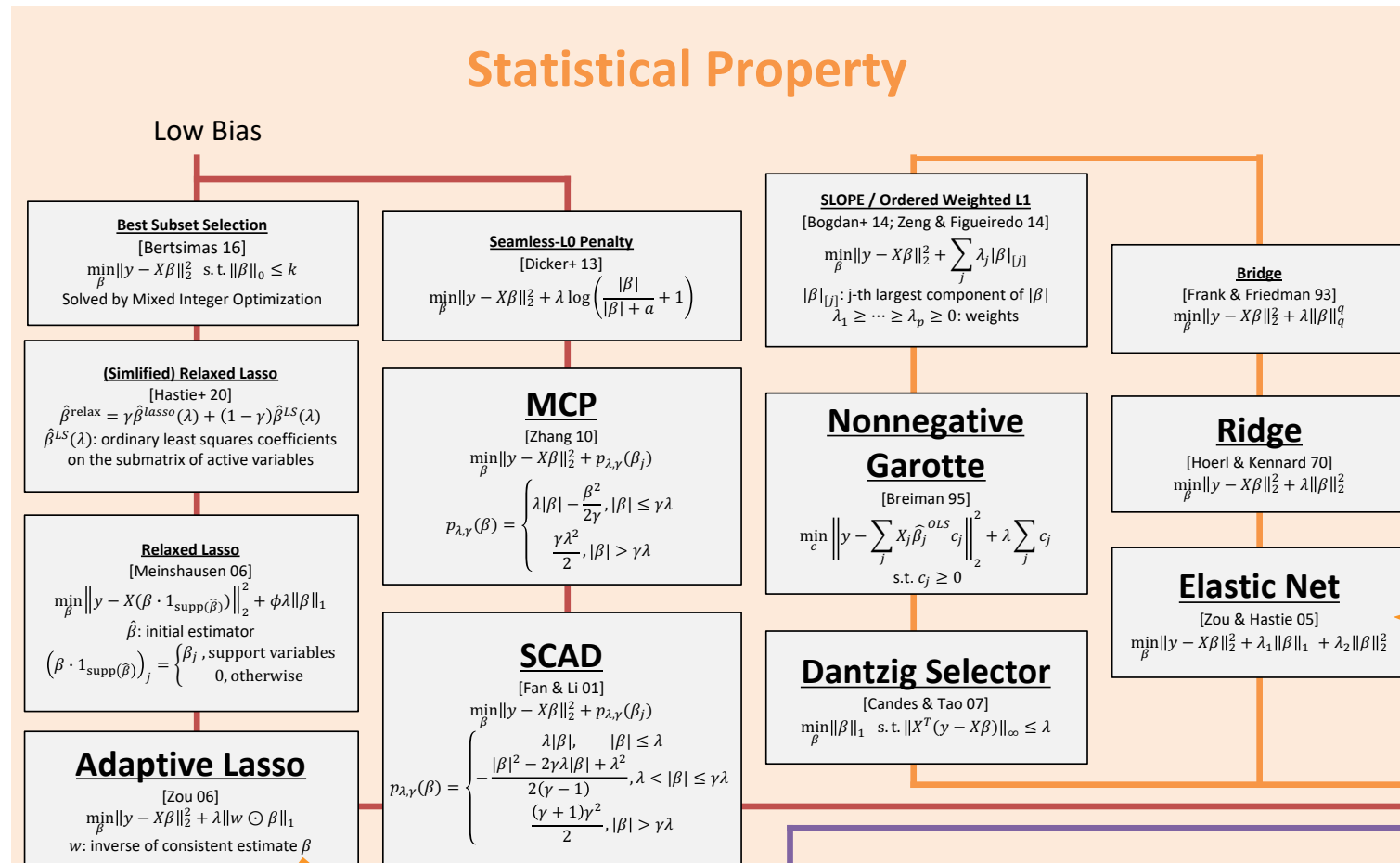
スパースモデリング マップ

スパースモデリングの研究方向性は、大きく4領域に分かれる

Sparse Modeling Landscape Map

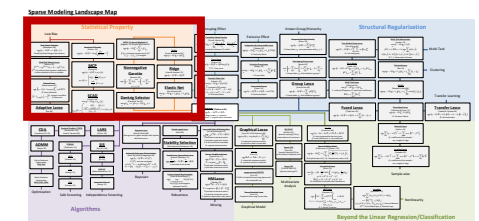


領域 I : Lassoの統計的な性質を改善する (バイアス低減や安定化など)



Lasso推定量の縮小バイアスを低減する

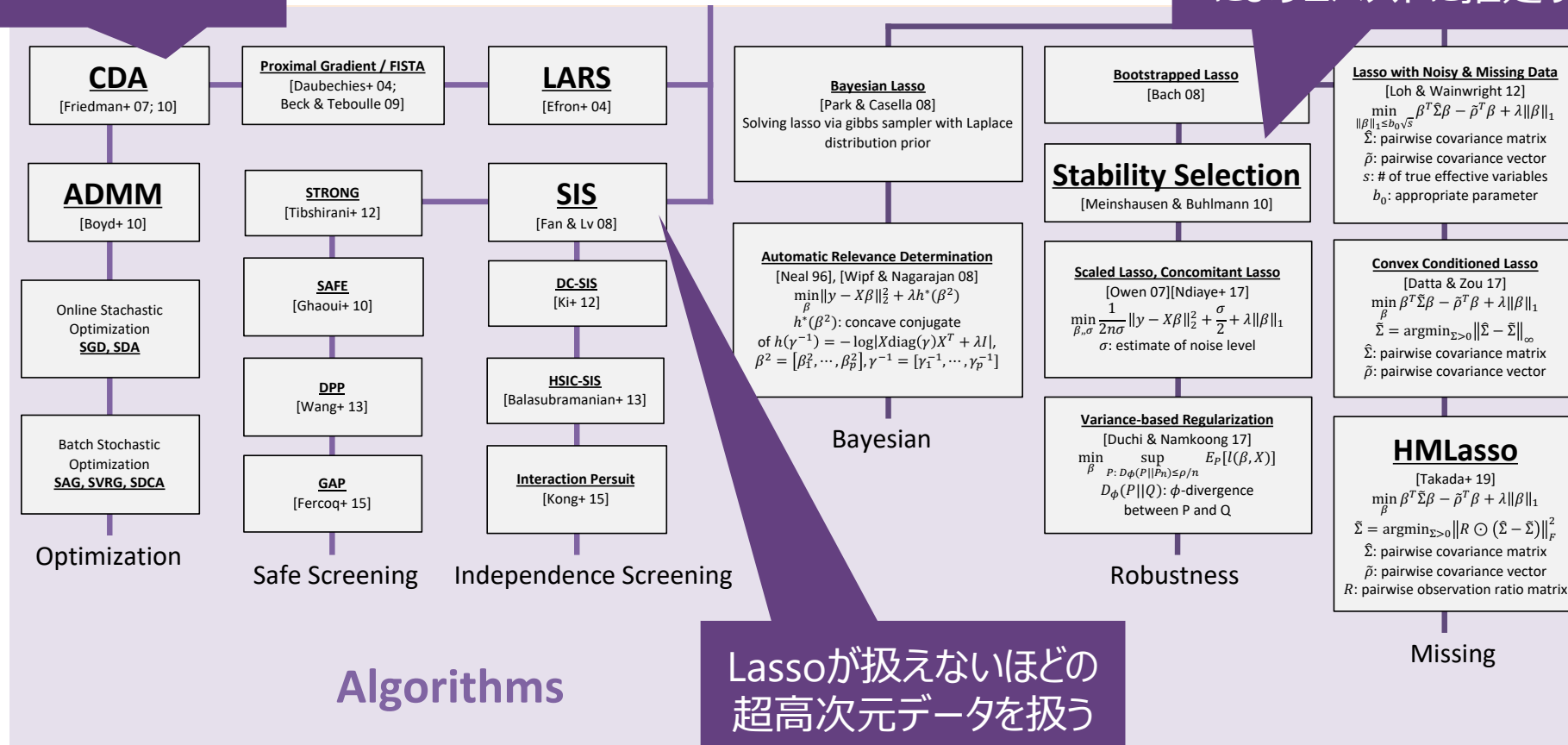
Lasso推定量の共線性による不安定性を低減する



領域Ⅱ：アルゴリズムを工夫して，困難性に対処する（超高次元性や欠測値など）

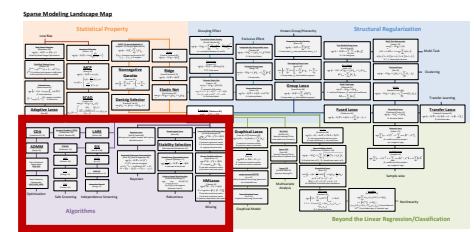
高速な最適化アルゴリズム

ブートストラップサンプリングによりロバストに推定する



Lassoが扱えないほどの超高次元データを扱う

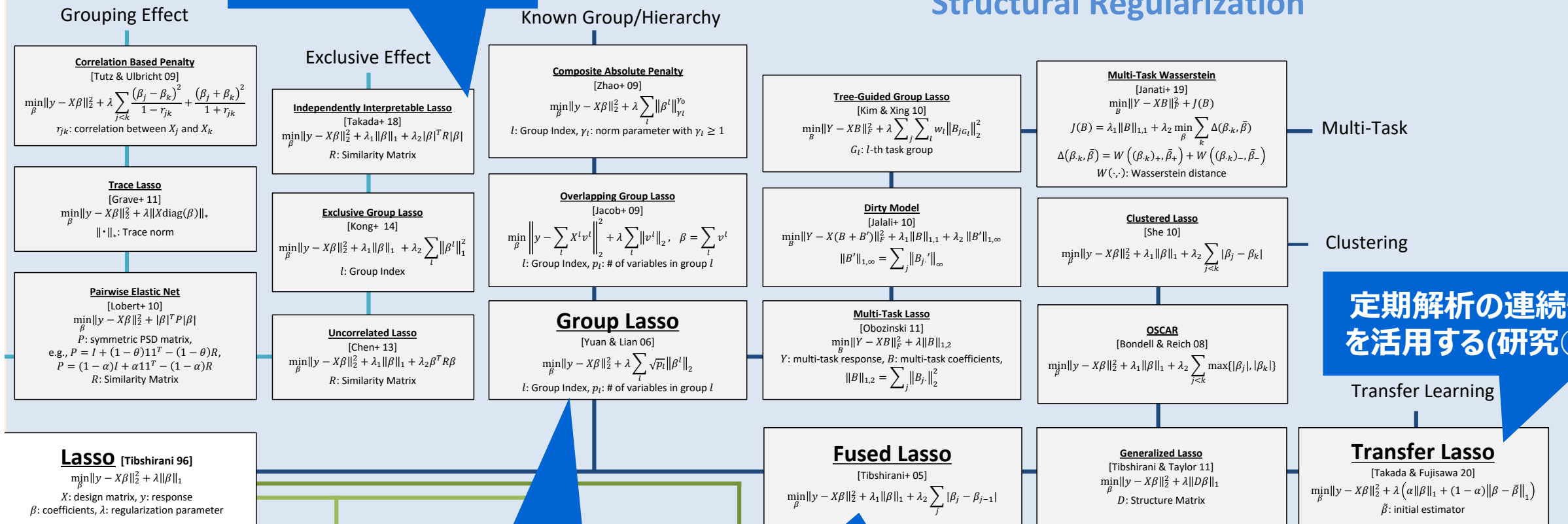
欠測値にうまく対処する (研究①)



領域Ⅲ：問題特有の構造や事前知識を組み込む（正則化を工夫）

有効変数の低相関性を活用する

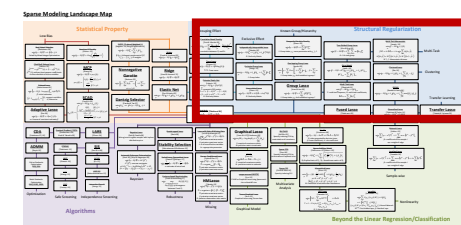
Structural Regularization



変数のグループ構造を活用する

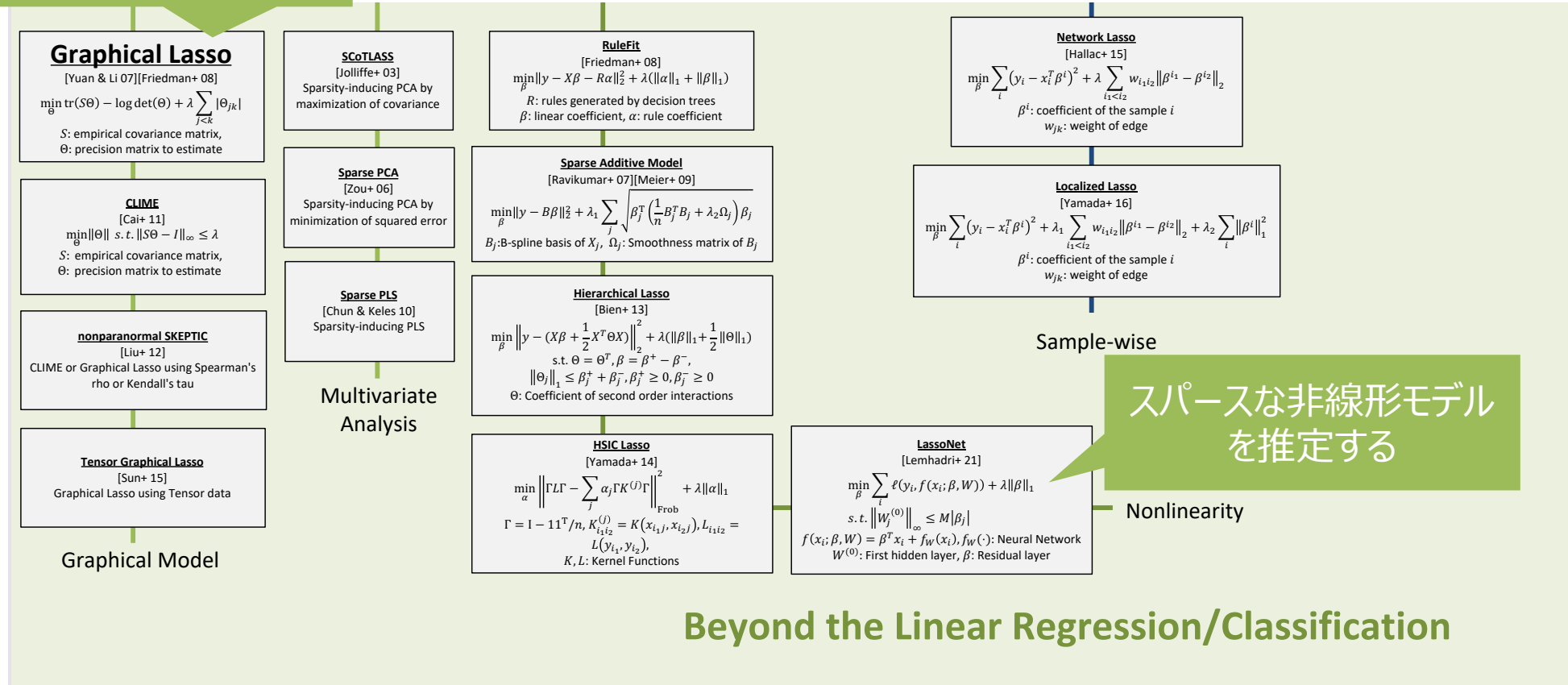
隣り合う変数の連続性を活用する

定期解析の連続性を活用する(研究②)

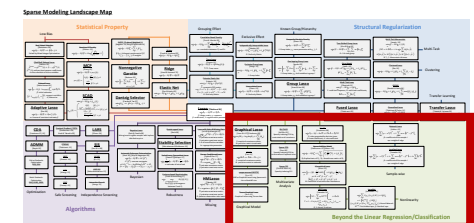


領域IV : 線形回帰/分類以外に適用範囲を広げる (グラフィカルモデルや非線形モデルなど)

変数間の条件付き依存性をスパースに推定する



Beyond the Linear Regression/Classification



III

スパースモデリングを深く掘り下げる (その 1)

研究① 高欠測データのためのスパースモデリング

(Takada, Fujisawa & Nishikawa 2019 IJCAI)



HMLasso

[Takada+ 19]

$$\min_{\beta} \beta^T \tilde{\Sigma} \beta - \tilde{\rho}^T \beta + \lambda \|\beta\|_1$$

$$\tilde{\Sigma} = \operatorname{argmin}_{\Sigma > 0} \|R \odot (\hat{\Sigma} - \Sigma)\|_F^2$$

$\hat{\Sigma}$: pairwise covariance matrix

$\tilde{\rho}$: pairwise covariance vector

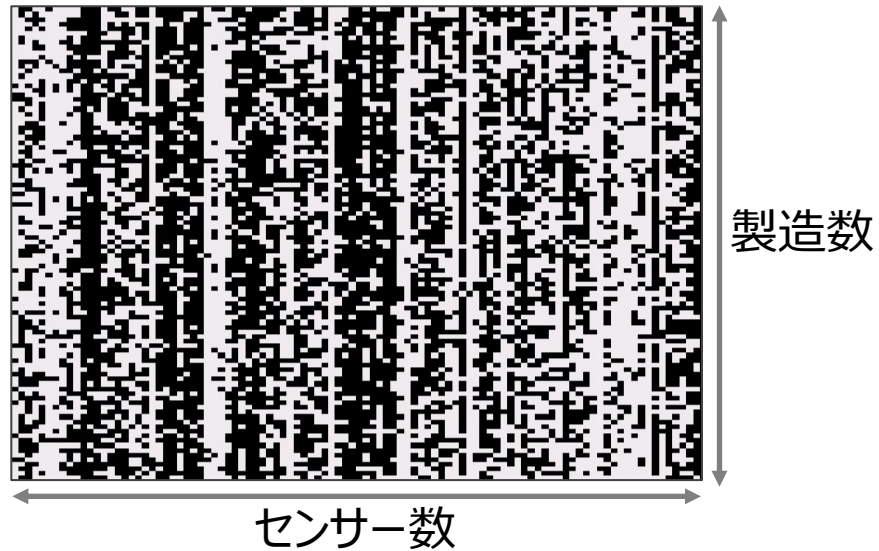
R : pairwise observation ratio matrix

モチベーション：工場における高欠測問題

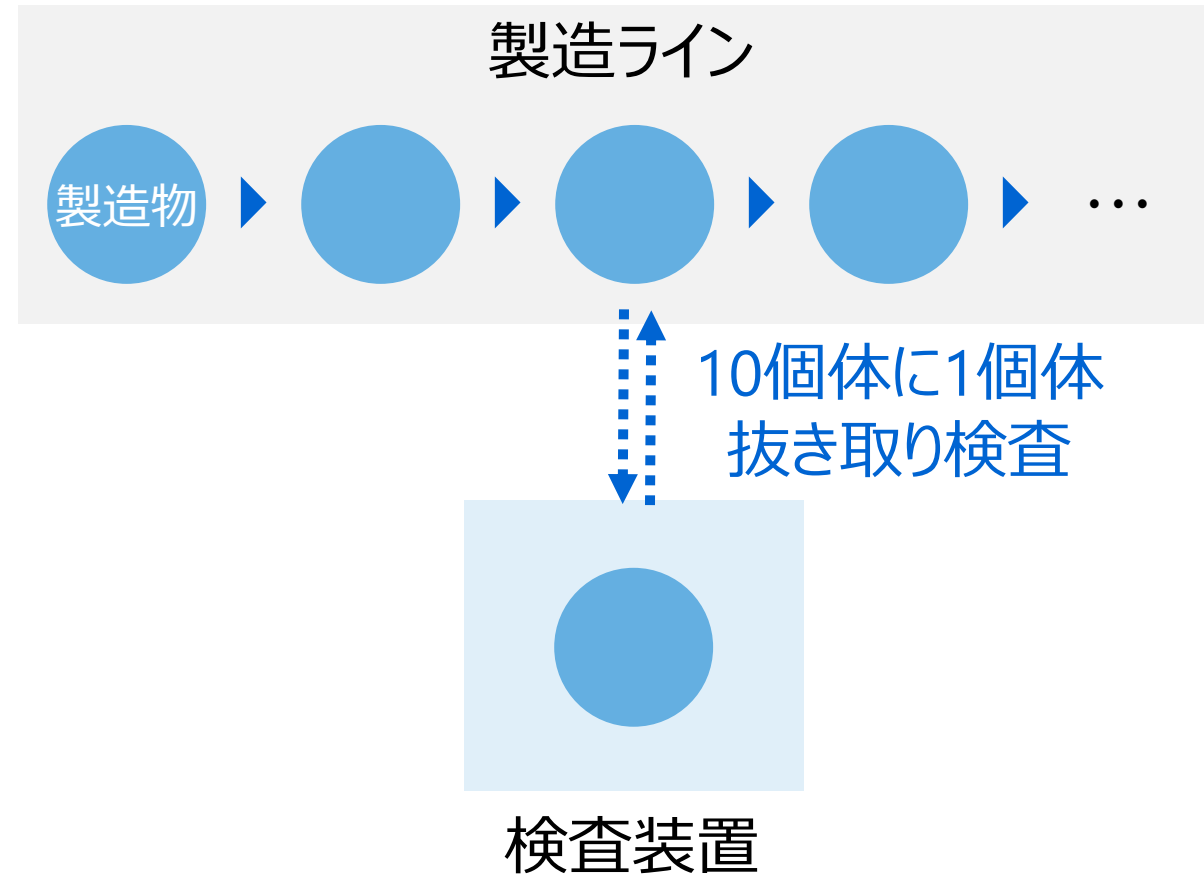
製造履歴データにはたびたび多くの欠測が含まれる

製造履歴データ

データのイメージ（黒：欠測，白：観測）



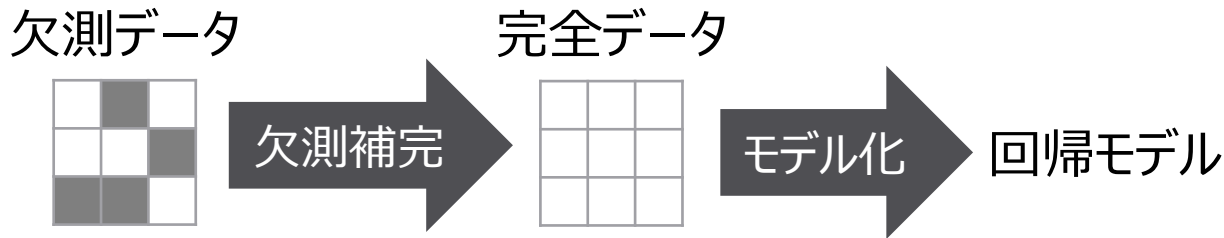
欠測が生じる主要因



- 抜き取り検査で測定される変数が90%欠測
- MCARだが，測定サンプルが非統一で高次元であり，リストワイズ除去不可

欠測データを扱う方法

欠測データから直接モデル化することで、効率的に推定できる



欠測補完 + モデル化

- 例：平均値補完・多重代入法
- 計算速度と精度を両立するのが難しい

直接モデル化

- 例：CoCoLASSO (Datta & Zou 2017 Ann. Stat.)
- 効率的に回帰モデルを推定できる

→ 本研究では、高欠測データも扱える直接モデル化手法を提案

CoCoLassoの基本的な考え方

補完する代わりに、半正定値の共分散行列を推定する

Lassoの定式化を共分散行列を用いた形に変形する. ($\mathbf{X} \in \mathbb{R}^{n \times p}$: 説明変数, $\mathbf{y} \in \mathbb{R}^n$: 目的変数)

$$\hat{\beta} = \operatorname{argmin}_{\beta} \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \quad \leftarrow \text{通常のLassoの定式化}$$

$$= \operatorname{argmin}_{\beta} \frac{1}{2} \beta^\top \mathbf{S} \beta - \rho^\top \beta + \lambda \|\beta\|_1, \mathbf{S} = \mathbf{X}^\top \mathbf{X} / n, \rho = \mathbf{X}^\top \mathbf{y} / n \quad \leftarrow \text{共分散を用いた定式化}$$

\mathbf{S}, ρ の代わりに, $\tilde{\Sigma}, \tilde{\rho}$ を使う.

$$\tilde{\Sigma} = \operatorname{argmin}_{\Sigma \succeq 0} \|\Sigma - \mathbf{S}^{\text{pair}}\|_{\max}, \quad \mathbf{S}_{jk}^{\text{pair}} = \frac{1}{|I_{jk}|} \sum_{i \in I_{jk}} X_{ij} X_{ik} \quad (\text{ペアワイズ共分散行列})$$

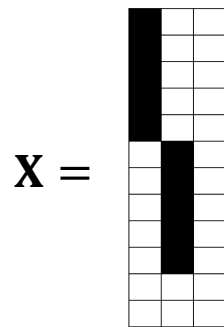
$$\tilde{\rho}_j = \frac{1}{|I_{jj}|} \sum_{i \in I_{jj}} y_i X_{ij}, \quad I_{jk} = \{i: X_{ij} \text{ と } X_{ik} \text{ がともに観測}, i = 1, \dots, n\}$$

* Convex Conditioned Lasso (Datta & Zou 2017 Ann. Stat.)

CoCoLassoの欠点

高欠測の場合、推定量が大幅に悪化する

もっともシンプルな例: X1, X2の同時観測数が非常に少ない場合



ペアワイズ
共分散行列

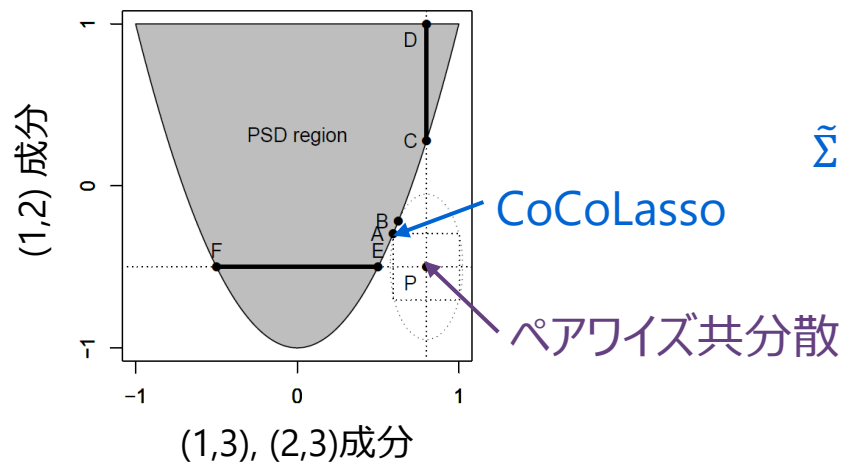
$$\mathcal{S}^{pair} = \begin{bmatrix} 1 & -0.5 & 0.8 \\ -0.5 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}$$

X1, X2の共分散は不確実性が高い
その他の共分散は不確実性が低い



CoCoLassoによる
半正定値化

$$\tilde{\Sigma} = \operatorname{argmin}_{\Sigma \succeq 0} \|\Sigma - \mathcal{S}^{pair}\|_{\max}$$



$$\tilde{\Sigma} = \begin{bmatrix} 1 & -0.3 & 0.6 \\ -0.3 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{bmatrix}$$

共分散の値が(不確実性に関わらず)
全体的に変化してしまう

提案手法：HMLasso

高欠測でも精度が落ちないようにCoCoLassoを改良した

HMLasso (Lasso with High Missing rate) (Takada, Fujisawa & Nishikawa 2019 IJCAI)

$$\hat{\beta} = \operatorname{argmin}_{\beta} \frac{1}{2} \beta^{\top} \tilde{\Sigma} \beta - \tilde{\rho}^{\top} \beta + \lambda \|\beta\|_1$$

$$\tilde{\Sigma} = \operatorname{argmin}_{\Sigma \succeq 0} \left\| R \odot (\Sigma - S^{\text{pair}}) \right\|_{\text{F}}^2, \quad R_{jk} = |I_{jk}|/n, \quad S_{jk}^{\text{pair}} = \frac{1}{|I_{jk}|} \sum_{i \in I_{jk}} X_{ij} X_{ik}$$

↑ 変数ペアの実測率で重み付けて共分散行列を推定 \odot : 要素ごとの積

$$\tilde{\rho}_j = \frac{1}{|I_{jj}|} \sum_{i \in I_{jj}} y_i X_{ij}, \quad I_{jk} = \{i: X_{ij}, \text{と} X_{ik} \text{が同時に観測}, i = 1, \dots, n\}$$

- ✓ HMLassoは共分散の不確実性をバランスさせることで、全体の推定精度を高める。
- ✓ この問題は座標降下法とADMMを用いて解くことができる。

理論解析 (本研究のポイント)

 R で重みづけることは、誤差限界の意味で最適である**定理**

観測行列 M ($M_{ij} = 1$ for 観測, $M_{ij} = 0$ for 欠測)が τ -sub Gaussianを満たすとする。
完全データの共分散行列を S とする。

変数 j, k の同時観測率を r_{jk} として, **重みを r_{jk}^α とする** : $\tilde{\Sigma} = \operatorname{argmin}_{\Sigma \succeq 0} \|R^\alpha \odot (\Sigma - S^{\text{pair}})\|_F^2$

このとき,

$$P\left(\|\tilde{\Sigma} - S\|_F^2/p^2 \leq \varepsilon^2\right) \geq 1 - p^2 C \exp\left(-cn\varepsilon^2 \left(\min_{j,k} r_{jk}\right)^{2\alpha} \left(\min_{j,k} r_{jk}^{2(1-\alpha)}\right) \zeta^{-1}\right),$$

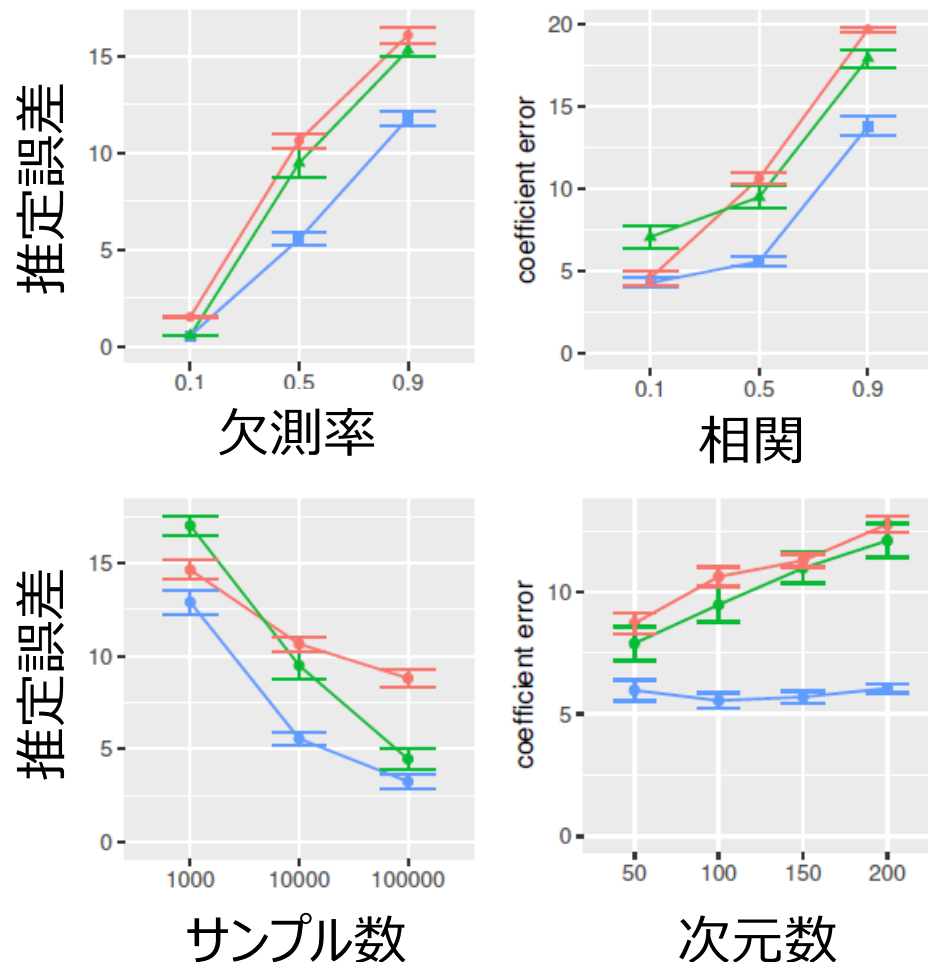
for $\forall \varepsilon \leq c\tau X_{\max}^2 \left(\min_{j,k} r_{jk}^{\alpha-1}\right) / \left(\min_{j,k} r_{jk}\right)^\alpha$ が成立する。

ただし c, C , and ζ は定数。

$\alpha = 1$ のとき, もっとも緩い条件でもっともタイトなバウンドが得られる

HMLassoはほぼ全条件で誤差最小となった

様々な条件下で、平均値補完, CoCoLasso, HMLassoの推定誤差を比較



シミュレーション条件

- 欠測率 = 0.1, 0.5, 0.9
(全体として上記欠測率となるように、各列の欠測率を一様分布から生成)
- 相関 = 0.1, 0.5, 0.9
- サンプル数 = 1000, 10000, 100000
- 次元数 = 50, 100, 150, 200
- 回帰係数: $\beta_1 = 10, \beta_{11} = -9, \dots, \beta_{81} = 2, \beta_{91} = -1, \beta_j = 0$ (otherwise)

- 平均値補完
- CoCoLasso
- HMLasso**

研究①：欠測スパースまとめ

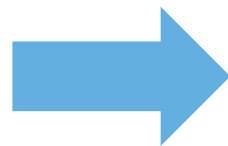
欠測値が非常に多い場合のスパース回帰手法を提案した

- 現実のデータ解析から，高欠測であるという課題を定義した。
- 高欠測でも精度を維持できる手法を提案した。
- 数値実験と理論解析で有効性を確認した。



高欠測データ

HMLasso



$$y = X\beta$$

回帰モデル

論文

Takada, M., Fujisawa, H., & Nishikawa, T. (2019).
"HMLasso: Lasso with High Missing Rate". IJCAI2019.

Rコード

<https://CRAN.R-project.org/package=hmlasso>

III

スパースモデリングを深く掘り下げる (その2)



研究② 定期解析のためのスパースモデリング
(Takada & Fujisawa 2020 NeurIPS)

Transfer Lasso

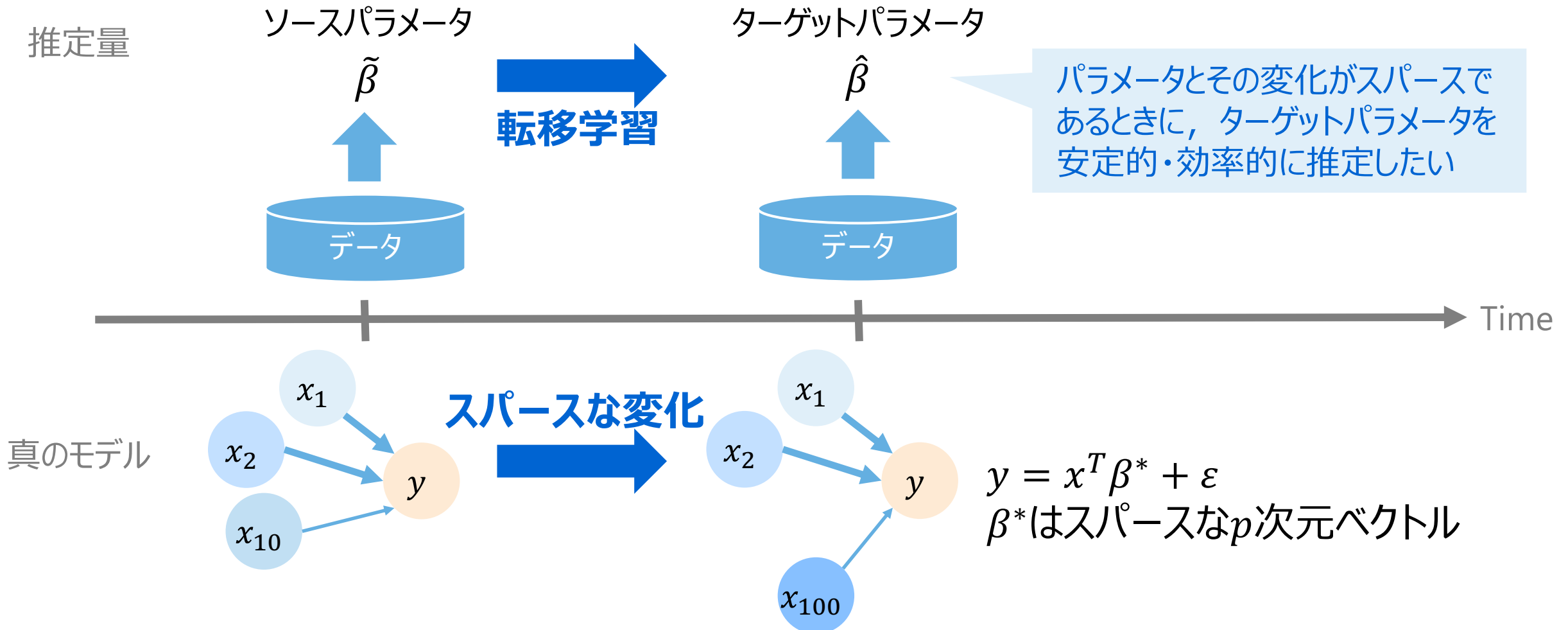
[Takada & Fujisawa 20]

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda (\alpha \|\beta\|_1 + (1 - \alpha) \|\beta - \tilde{\beta}\|_1)$$

$\tilde{\beta}$: initial estimator

モチベーション：スパース推定を用いた定期解析

絶えず変わる環境下において、いかに安定的・効率的にモデルを更新するか



提案手法 : Transfer Lasso

ℓ_1 正則化を用いたシンプルかつ効果的な方法を提案した

設定 : $\mathbf{X} \in \mathbb{R}^{n \times p}$: 説明変数, $\mathbf{y} \in \mathbb{R}^n$: 目的変数

Lasso (Tibshirani 1996) : $\hat{\beta} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \right\}$ $\lambda > 0$: 正則化パラメータ

Transfer Lasso (Takada & Fujisawa 2020 NeurlPS) : ソースパラメータ $\tilde{\beta} \in \mathbb{R}^p$ からスパースに変化

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \left(\alpha \|\beta\|_1 + (1 - \alpha) \|\beta - \tilde{\beta}\|_1 \right) \right\}$$

スパース正則化
スパース変化正則化

($0 \leq \alpha \leq 1$: 正則化のバランスを制御するパラメータ)

- ✓ 定常状態のとき, ソースパラメータをベースに更新することで, **前回の知識を転移**できる.
- ✓ 非定常状態のとき, 正則化を制御することで, **前回の知識を捨てる**ことができる.
- ✓ 推定量とその変化がスパースであるため, **モデルを解釈・管理しやすい**

提案手法の直感的理解

正則化の等高線がソースパラメータで尖ることで、スパースな変化を実現する

正則化項 $\alpha\|\beta\|_1 + (1 - \alpha)\|\beta - \tilde{\beta}\|_1$ の等高線 ($\alpha = 1, 3/4, 1/2, \tilde{\beta} = (1, 1/2)^T$)

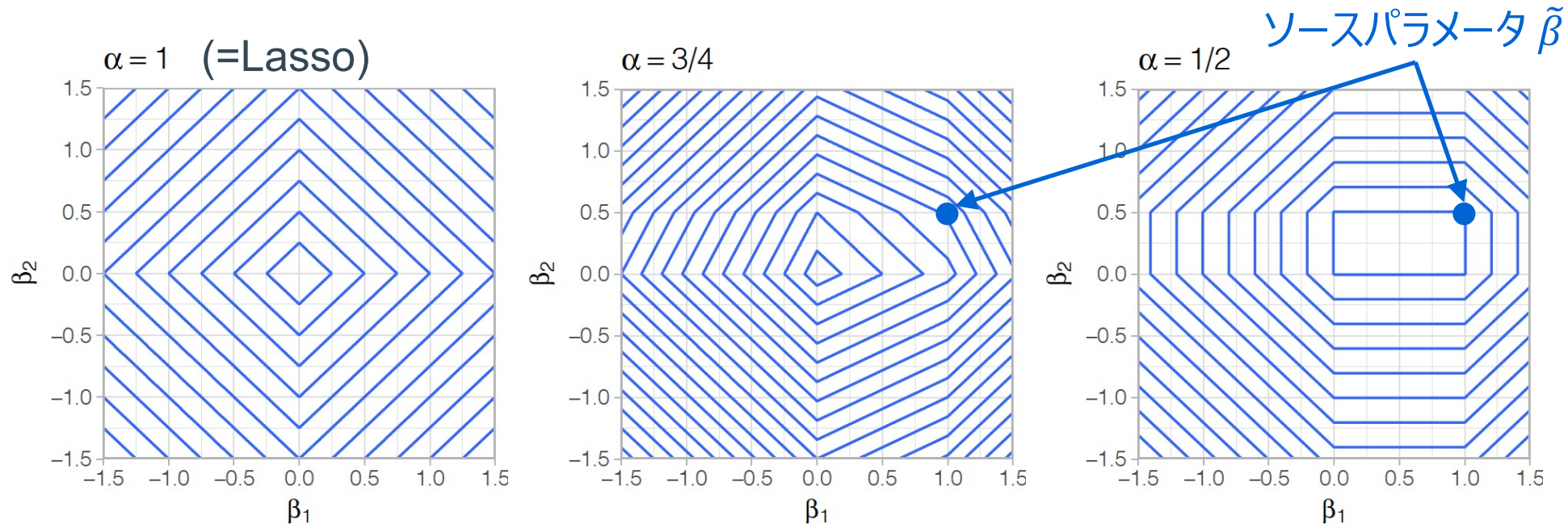


Figure 2: Contours of our regularizer.

$\beta_j = 0$ で尖っている
→ スパースな推定値

$\beta_j = 0$ と $\beta_j = \tilde{\beta}_j$ で尖っている
→ スパースな変化

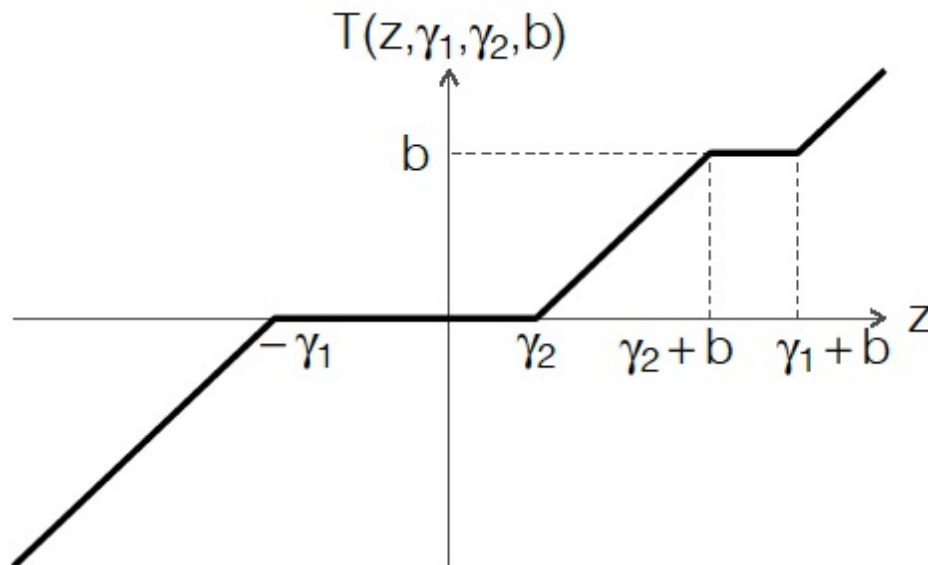
提案手法の最適化アルゴリズム

座標降下法により高速に解を求めることができる

座標降下法：

$j = 1, \dots, p, 1, \dots, p, \dots$ に対して, 収束するまで以下を繰り返す.

$$\beta_j \leftarrow T \left(\frac{1}{n} \mathbf{X}_j^T (\mathbf{y} - \mathbf{X}_{-j} \beta_{-j}), \lambda, \lambda(2\alpha - 1), \tilde{\beta}_j \right)$$



軟閾値関数 T は 0 とソースパラメータ $\tilde{\beta}_j$ の2箇所で一定値をとっている
 \Rightarrow 2つのスパース性をもつ

理論解析(1) 推定誤差の非漸近限界

推定誤差の上限がタイトになることを示した

$$\mathbf{y} = \mathbf{X}\beta^* + \varepsilon, \Delta = \tilde{\beta} - \beta^*, \hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \left(\alpha \|\beta\|_1 + (1 - \alpha) \|\beta - \tilde{\beta}\|_1 \right) \right\}$$

ソースパラメータの誤差

仮定1 : ノイズはサブガウシアン, i.e., $E[\exp(t\varepsilon)] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right), \forall t \in \mathbb{R}$.

仮定2 : 以下の一般化制限固有値条件を満たす. $S = \{j: \beta_j^* \neq 0\}$,

$$\phi = \inf_{v \in \mathcal{B}} \frac{v^T (X^T X / n) v}{\|v\|_2^2} > 0, \mathcal{B} = \left\{ v \in \mathbb{R}^p \mid \begin{array}{l} (2\alpha - 1) \|v_{S^c}\|_1 + 2(1 - \alpha) \|v - \Delta\|_1 \\ \leq (2\alpha + 1) \|v_S\|_1 + 2(1 - \alpha) \|\Delta\|_1 \end{array} \right\}$$

定理 : 上記の仮定のもとで, Transfer Lassoは高い確率で以下を満たす.

$$\|\hat{\beta} - \beta^*\|_2^2 \leq \frac{(2\alpha + 1)^2 \lambda^2 |S|}{4\phi^2} \left(1 + \sqrt{1 + \frac{8(1 - \alpha)\phi \|\Delta\|_1}{(2\alpha + 1)^2 \lambda |S|}} \right)$$

推定誤差

✓ ソースパラメータが正しければ($\Delta = 0$), Transfer Lasso($\alpha < 1$)の誤差はLassoより小さい.

理論解析(2) 推定誤差の収束レート

急激な変化があっても、解に一致性があることを示した

$$\mathbf{y} = \mathbf{X}\beta^* + \varepsilon, \Delta = \tilde{\beta} - \beta^*, \hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \left(\alpha \|\beta\|_1 + (1 - \alpha) \|\beta - \tilde{\beta}\|_1 \right) \right\}$$

ソースパラメータの誤差

仮定1 : ノイズはサブガウシアン, i.e., $E[\exp(t\varepsilon)] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right), \forall t \in \mathbb{R}$.

仮定2 : 以下の一般化制限固有値条件を満たす. $S = \{j: \beta_j^* \neq 0\}$,

$$\phi = \inf_{v \in \mathcal{B}} \frac{v^T (X^T X / n) v}{\|v\|_2^2} > 0, \mathcal{B} = \left\{ v \in \mathbb{R}^p \mid \begin{array}{l} (2\alpha - 1) \|v_{S^c}\|_1 + 2(1 - \alpha) \|v - \Delta\|_1 \\ \leq (2\alpha + 1) \|v_S\|_1 + 2(1 - \alpha) \|\Delta\|_1 \end{array} \right\}$$

定理 : 上記の仮定のもとで, Transfer Lassoは高い確率で以下を満たす.

$$\|\hat{\beta} - \beta^*\|_2^2 = O \left(\left(\alpha + \frac{1}{2} \right)^2 \lambda^2 s + (1 - \alpha) \lambda \|\Delta\|_1 \right)$$

推定誤差

前頁と同じ仮定

✓ ソースパラメータが間違えていても ($\Delta \neq 0$), Transfer Lassoは $\lambda \rightarrow 0$ で一致性をもつ.

理論解析(3) 解の不変性

解が変化しない条件を解析的に示した

定理 : ソースパラメータと同一の解 $\hat{\beta} = \tilde{\beta}$ が存在する必要十分条件は、以下の不等式が成立することである。

$$\left| \frac{1}{n} \mathbf{X}_j^T (\mathbf{y} - \mathbf{X}\tilde{\beta}) \right| \leq \lambda \text{ for } \tilde{\beta}_j = 0,$$
$$-\lambda \left((1 - \alpha) - \alpha \operatorname{sgn} \tilde{\beta}_j \right) \leq \frac{1}{n} \mathbf{X}_j^T (\mathbf{y} - \mathbf{X}\tilde{\beta}) \leq \lambda \left((1 - \alpha) + \alpha \operatorname{sgn} \tilde{\beta}_j \right) \text{ for } \tilde{\beta}_j \neq 0.$$

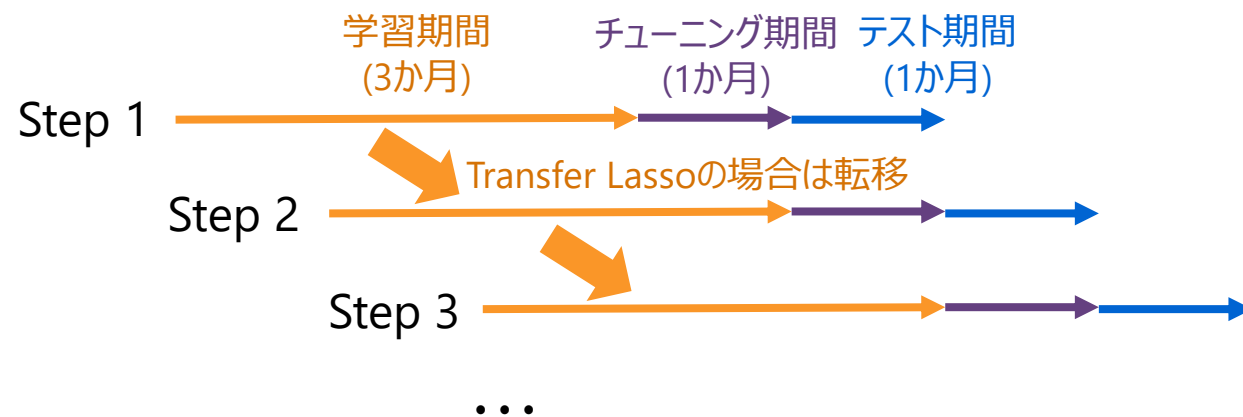
- ✓ α が小さく、 λ が大きければ、Transfer Lassoの解はソースパラメータと一致する。
- ✓ この性質を利用して、 λ の探索範囲の上限を決めることができる。

実データ実験：半導体製造データへの適用

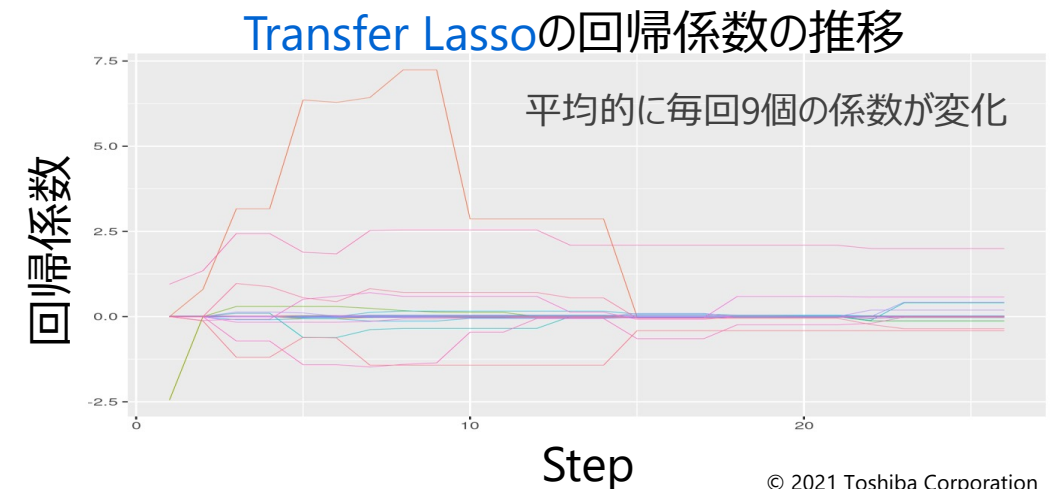
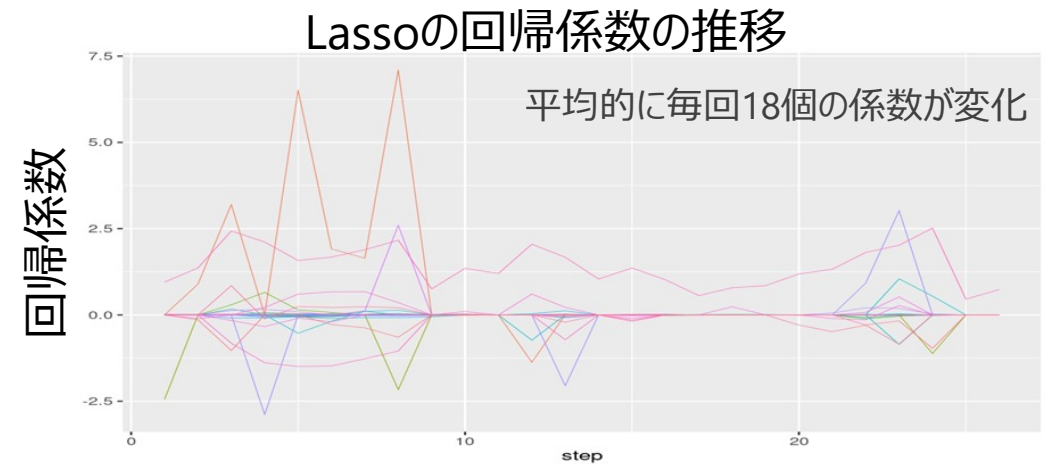
定期解析にTransfer Lassoを適用した結果、解析安定性が向上した

実験設定

- Y: 半導体の品質特性, X: 製造プロセスデータ
- 3か月データで学習, 次の1か月データでチューニング
- 2年半にわたって上記を毎月繰り返す

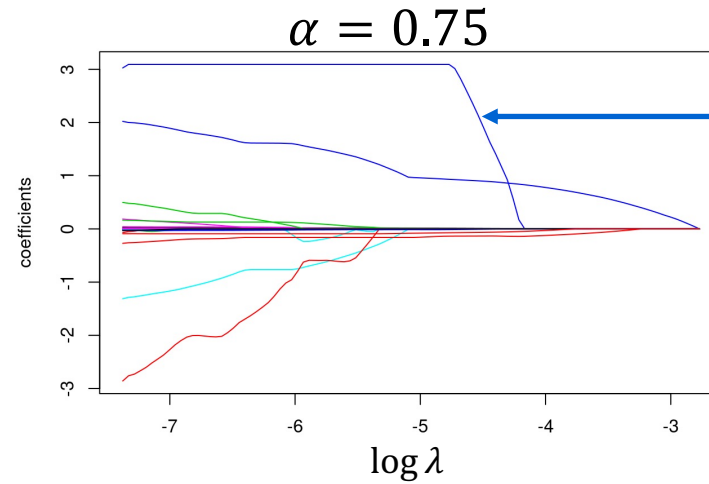
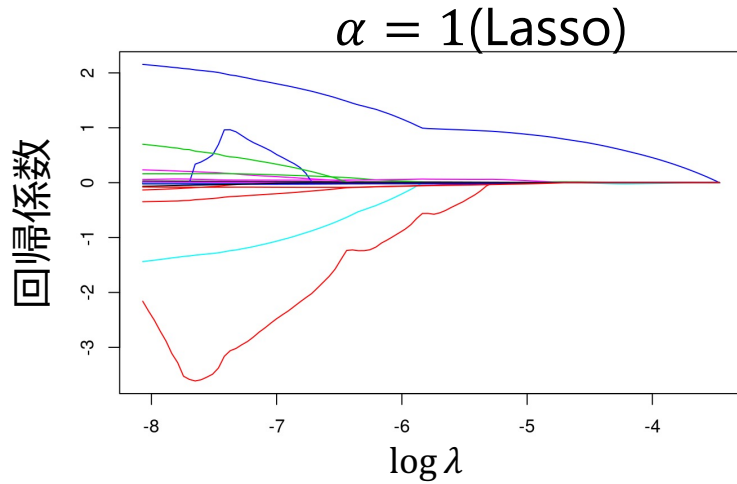


結果



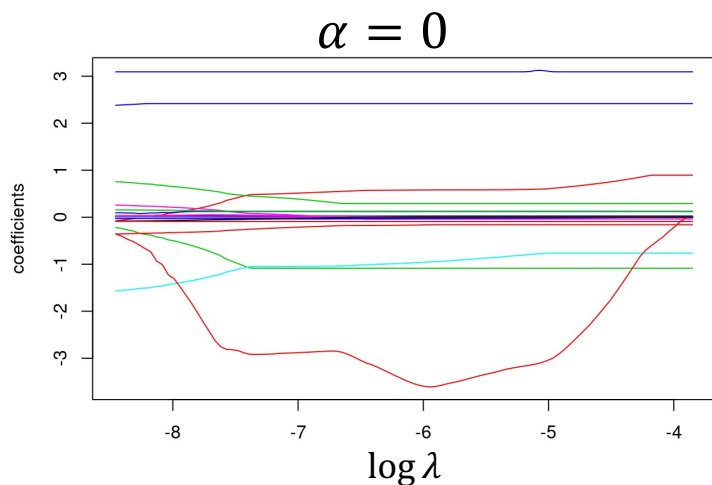
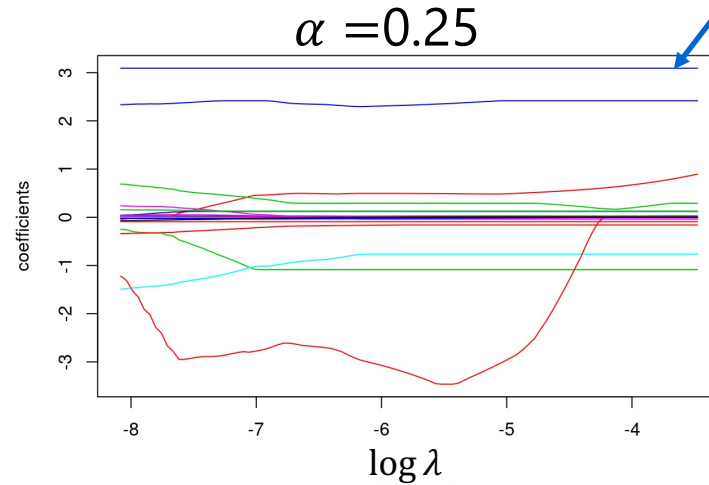
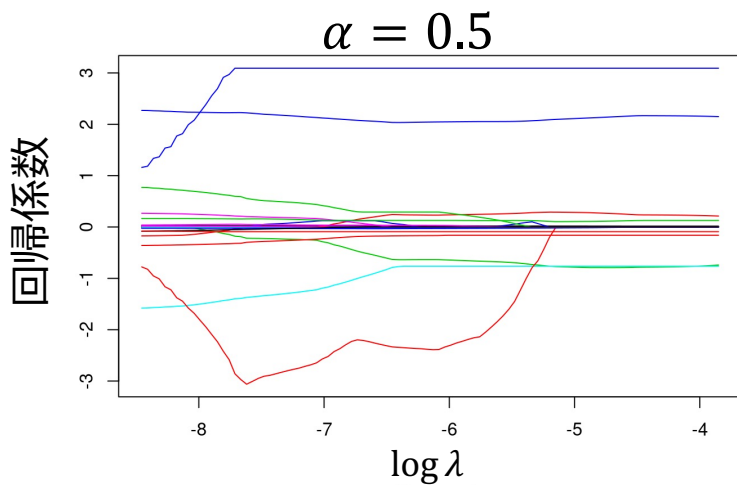
実データ実験：半導体製造データへの適用

ソースパラメータにトラップされやすい解パスとなる



$\alpha > 0.5$ の場合、解パスはゼロからスタート。ソースパラメータで停止することもある。

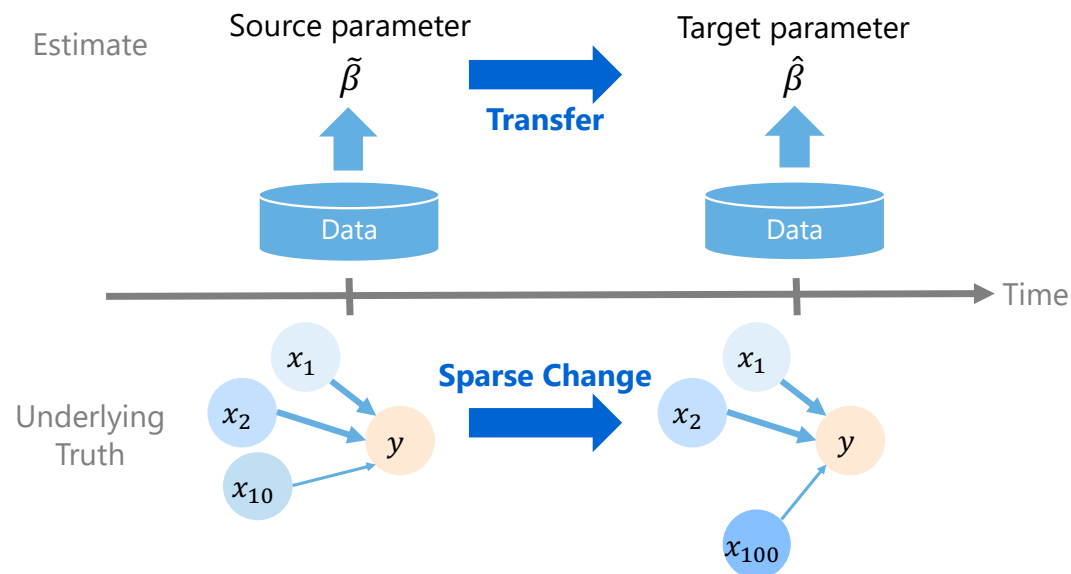
$\alpha < 0.5$ の場合、ソースパラメータからスタート。係数がまったく変わらない変数もあれば、途中で変化する変数もある。



研究②：スパース転移学習まとめ

L1正則化を用いた理論保証付き転移学習法を提案した

- 変わり続ける環境の中で，モデルをうまく更新する方法に関する研究。
- 「**パラメータ自身**」と「**パラメータの変化**」がともに**スパース**になる手法を提案した。
- 定常状態では推定誤差がタイトになり，非定常状態でも**一致性**があることを示した。



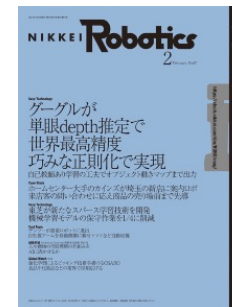
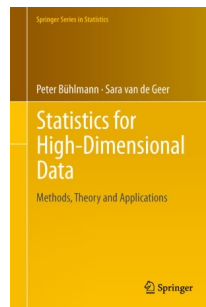
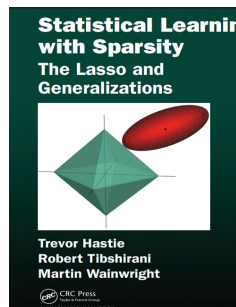
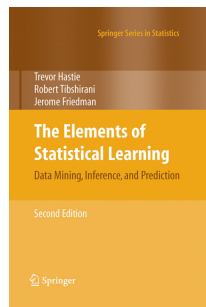
論文

Takada, M., & Fujisawa, H. (2020). "Transfer Learning via ℓ_1 Regularization". NeurIPS2020.

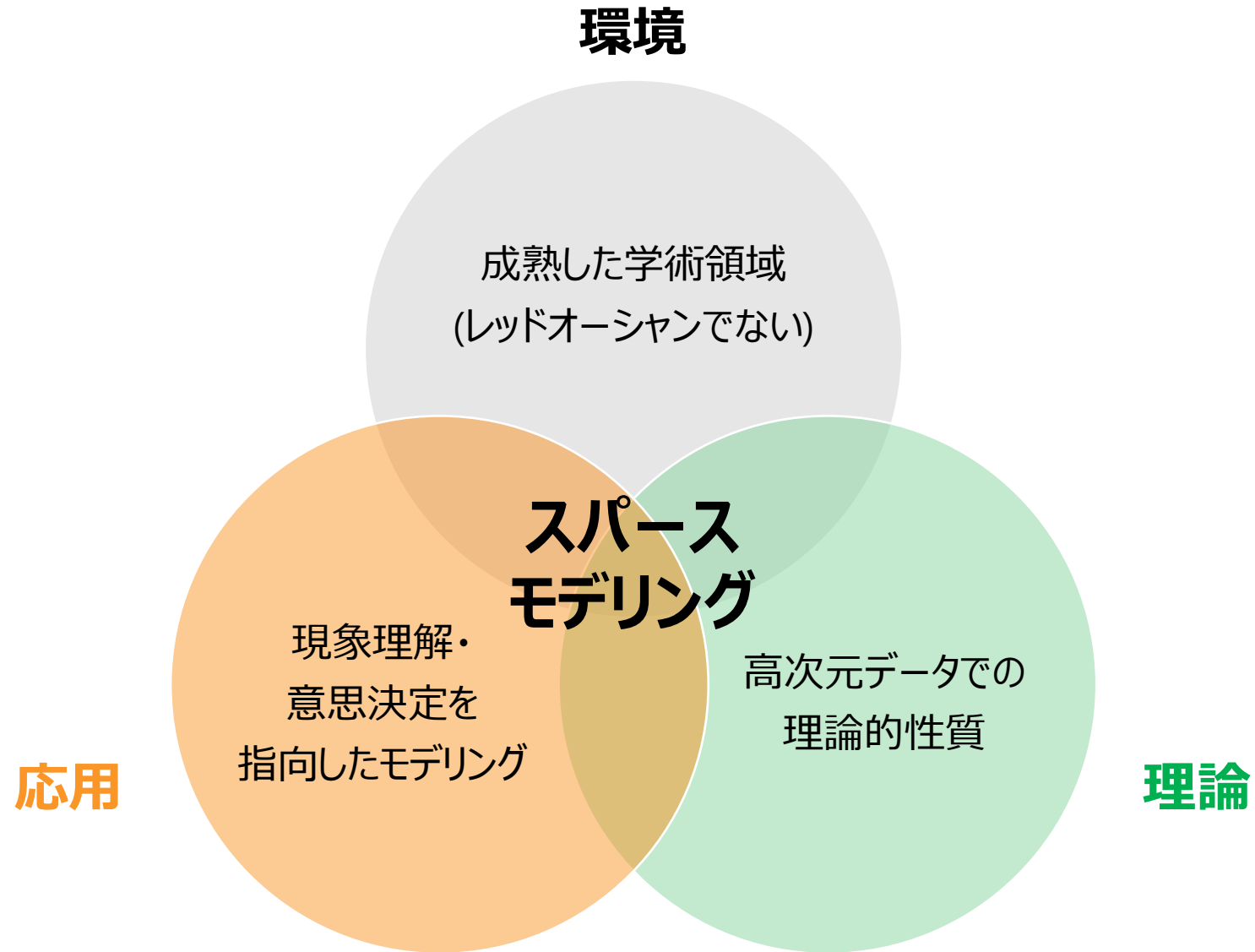
おわりに

さらに知りたい方へ

- スパースモデリングの教科書
 - 概要を知りたい→“The Element of Statistical Learning”, Chapter 18 (Hastie, Tibshirani, Friedman)
 - 全体像を知りたい→“Statistical Learning with Sparsity” (Hastie, Tibshirani, Wainwright)
 - 理論を学びたい→“Statistics for High-Dimensional Data” (Bühlmann, van de Geer)
 - 全体像をコンパクトに知りたい→「スパース推定法による統計モデリング」(川野, 松井, 廣瀬)
- 紹介した研究の詳細
 - Takada, M., Fujisawa, H., & Nishikawa, T. (2019). "HMLasso: Lasso with High Missing Rate". IJCAI2019. (*1)
 - Takada, M., & Fujisawa, H. (2020). "Transfer Learning via ℓ_1 Regularization". NeurIPS2020. (*2)
 - 日経ロボティクス2019年11月号と2021年2月号のSexy Technologyで, 5~6ページにわたる解説記事が掲載 (*3, 4)
 - 統数研ニュース2021年5月号に東芝-統数研の共同研究に関する記事が掲載 (*5)
 - 東芝AI技術カタログに関連情報が掲載 (*6)

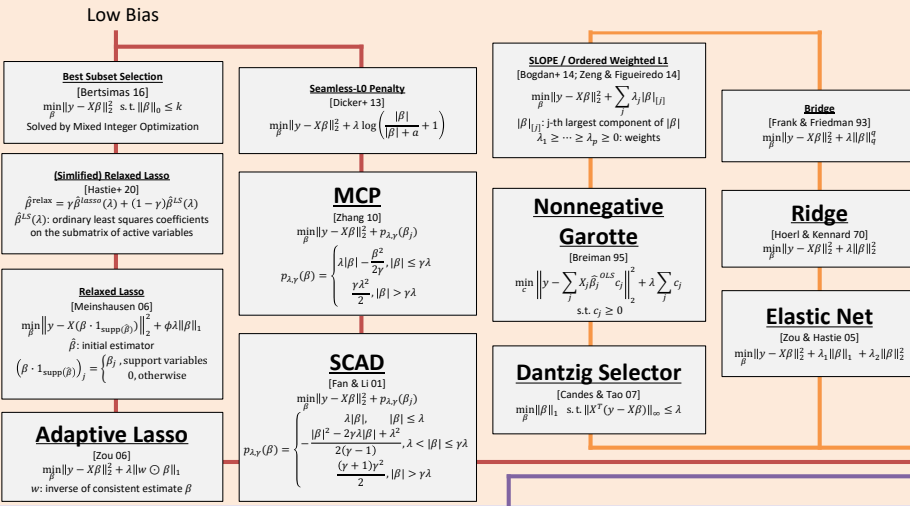


- *1 <https://www.ijcai.org/proceedings/2019/491>
- *2 <https://proceedings.neurips.cc/paper/2020/hash/a4a83056b58ff983d12c72bb17996243-Abstract.html>
- *3 <https://xtech.nikkei.com/atcl/nxt/mag/rob/18/012600001/00038/>
- *4 <https://xtech.nikkei.com/atcl/nxt/mag/rob/18/012600001/00070/>
- *5 <https://www.ism.ac.jp/gaiyo-news/News.html>
- *6 <https://www.global.toshiba/jp/technology/corporate/ai/catalog013.html>

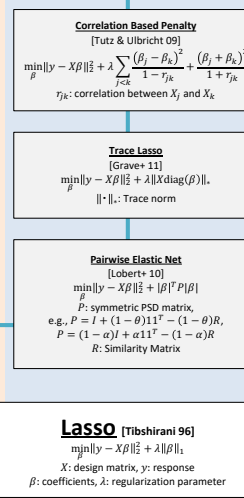


Sparse Modeling Landscape Map

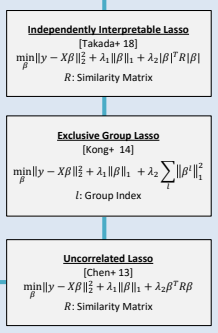
Statistical Property



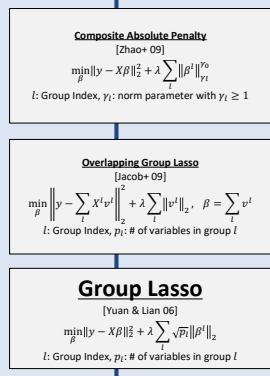
Grouping Effect



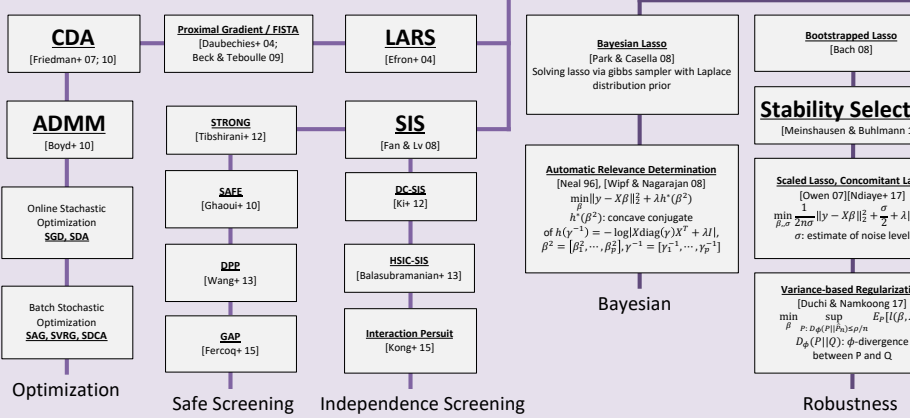
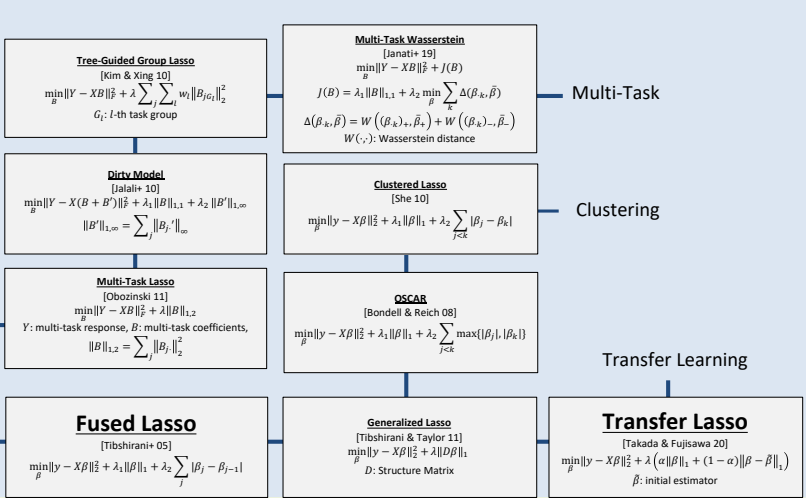
Exclusive Effect



Known Group/Hierarchy

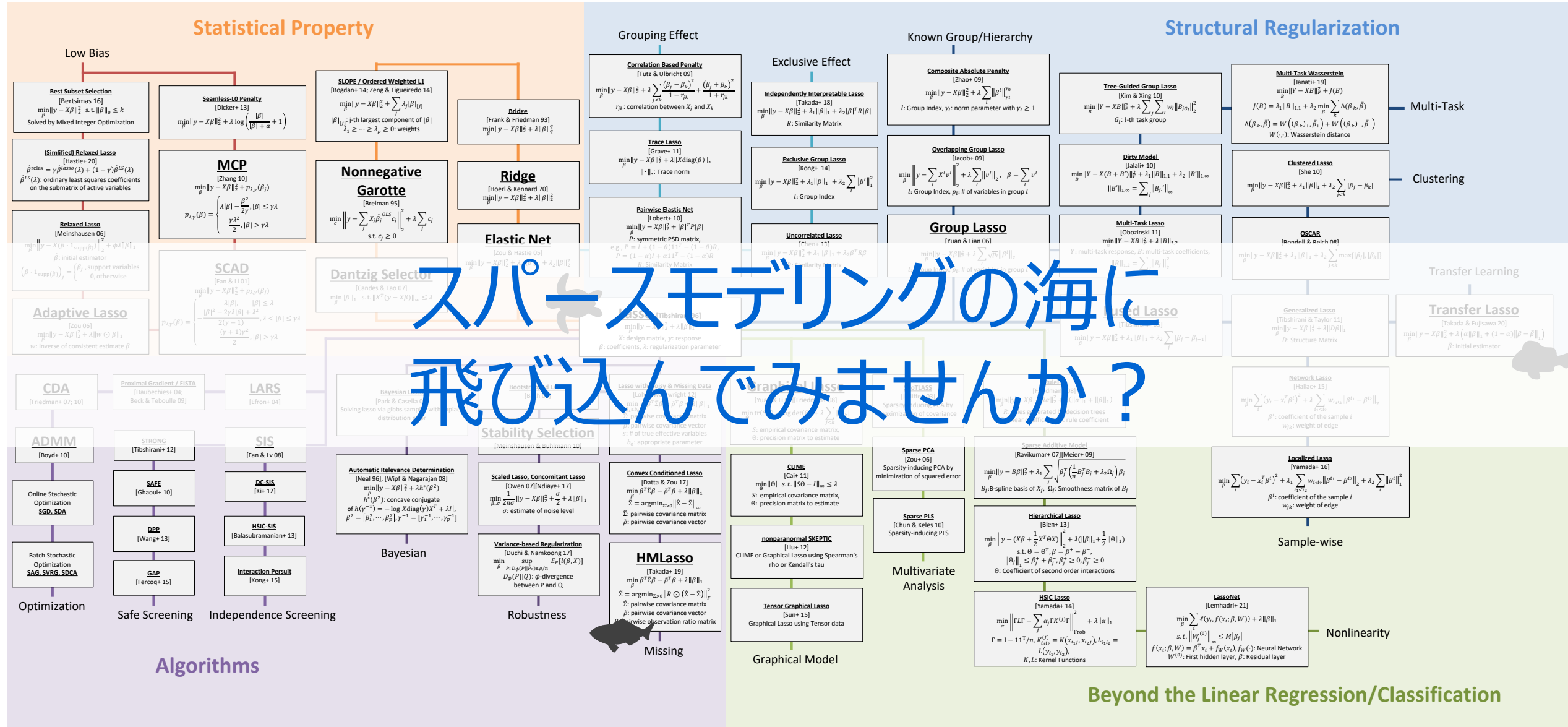


Structural Regularization



Algorithms

Beyond the Linear Regression/Classification



スパースモデリングの海に飛び込んでみませんか?

Beyond the Linear Regression/Classification

TOSHIBA