

物質科学と機械学習
マテリアルズ・インフォマティクス

目次

- 自己紹介
- マテリアルズ・インフォマティクス(MI)
- MIの目的
- 物質科学と機械学習の関係
- 何故今MIか？
- DFTとMI
- DFTとMIの分業
- MIによるDFTの機能強化・拡張

自己紹介

-2012 阪大基礎工博士課程修了(吉田博先生の下で研究)

第一原理計算によるスピントロニクス材料の探索

2012-2013 阪大基礎工 元素戦略ポスドク(佐藤和則先生の下で研究)

クラスター展開法を用いた磁性半導体の状態図の計算

2013-2016 JASRI/SPring-8 CRESTポスドク(木村滋先生、水牧仁一朗先生の下で研究)

第一原理計算による遷移金属酸化物の吸収分光スペクトルの計算・実験

2016- NIMS、阪大産研 NIMSポスドク(小口多美夫先生の下で研究)

第一原理計算から得られた物性データから有益な物理法則を抽出するため研究

開発・公開コード: LIDG (Linearly independent descriptor generation) program
github.com/Hitoshi-FUJII/LIDG

論文: 執筆中

雑誌: マテリアルズ・インフォマティクスによる材料開発

工業材料(2019年6月号vol67, No.6, p78 日刊工業新聞社)

マテリアルズ・インフォマティクス(MI)

統計学、情報科学、パターン認識、データサイエンス等の分野で発達してきた**機械学習**(統計的学習)の手法を**物質科学**へ応用した研究

様々な機械学習手法

統計学	計算機科学	情報科学	パターン認識
線形回帰分析 最小二乗法(OLS) 主成分分析(PCA)	モンテカルロ法(MC) 遺伝的アルゴリズム(GA) 遺伝的プログラミング(GP) 進化的アルゴリズム(EA) 粒子群最適化(PSO)		サポートベクターマシン(SVM) ニューラルネット(NN) ディープラーニング(DL) ランダムフォレスト(RF) アンサンブル学習
統計的仮説検定 赤池情報量基準(AIC)		決定木(DT) モンテカルロ木探索(MCTS)	
圧縮センシング スパースモデリング	ベイズ推定	ベイズ最適化	
L1正則化(LASSO) L2正則化(Ridge)		ベイズ統計	

MIの目的

1. 物質探索:

良い物性値を持つ物質・材料の発見
予測モデルの構築は必ずしも必要ではない

2. 物性予測:

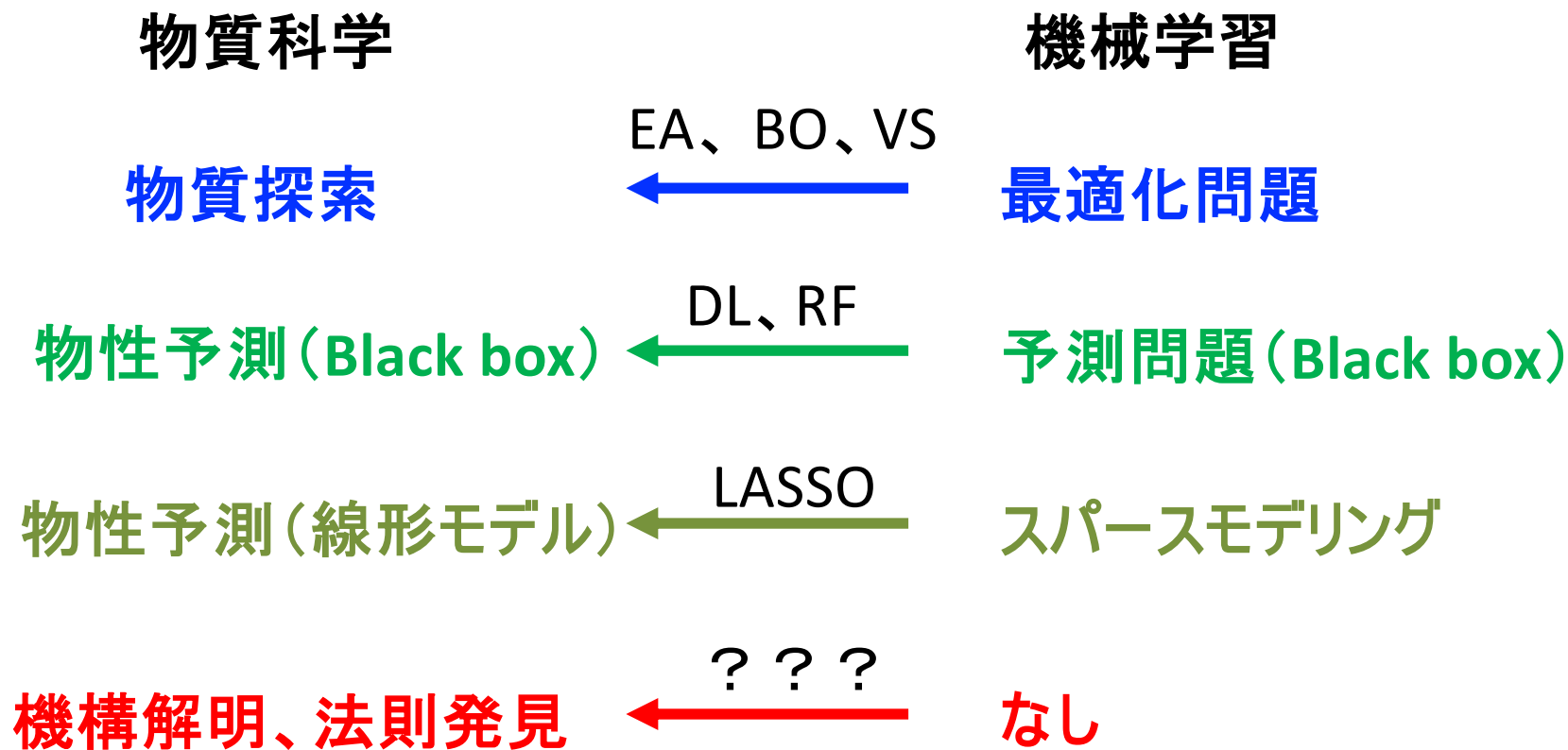
予測モデルを仮定し、未知物質の物性を高精度に予測
非線形モデル(Black box関数)、あるいは線形モデル

3. 法則発見:

物性発現機構の解明
得られた知見から新たな原理、法則の発見

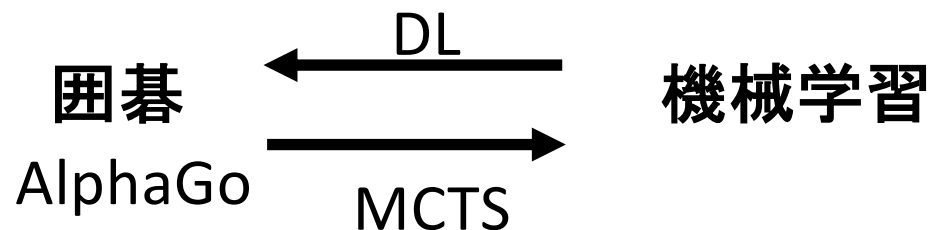
4. 物性データベース構築

物質科学と機械学習との関係



使える統計的手法はない
→ 物質科学者が作るべき

物質科学と機械学習との関係



機械学習の手法を物質科学へ応用

物質科学から新たな機械学習手法の提案

機械学習手法の目的別分類

最適化問題を解く:

(必ずしもモデルを必要としない)

ランダムサンプリング

進化論的アルゴリズム (EA)

遺伝的アルゴリズム (GA)

遺伝的プログラミング (GP)

ベイズ最適化 (BO) (モデル必要)

モンテカルロ木探索 (MCTS)

シミュレーテッドアニーリング

粒子群最適化 (PSO)

予測(識別)問題 (Black box) を解く:

(何らかの複雑で柔軟なモデルを仮定)

サポートベクターマシン (SVM)

ニューラルネット (NN)

ディープラーニング (DL)

ランダムフォレスト (RF)

アンサンブル学習

スパースモデリング (圧縮センシング):

(線形、スパースなモデルを仮定)

最小二乗法 (OLS)

LASSO、Ridge回帰

主成分分析 (PCA)

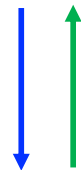
部分最小二乗法 (PLS)

ベイズ推定:

(観測データの従う分布を推定)

最適化問題

モデルパラメータの
最適化



モデルを用いた
最適化

予測・識別問題

何故今MIか？

何故、バイオ・インフォマティクスやケモ・インフォマティクスに比べて、物質科学へのインフォマティクスの導入が遅れたのか？

理由1. 物質科学は基本的には演繹的な研究が可能な分野

理由2. 既に第一原理計算(DFT)という強力なツールが存在

つまりは、単に、帰納的研究という発想がなかっただけ(だと思われる)

しかし、昨今のハイスループット実験、計算による大量の物性データの蓄積(ビッグデータ)や材料開発競争の加速等により、ようやく物質科学の分野にもインフォマティクスが取り入れられるようになった

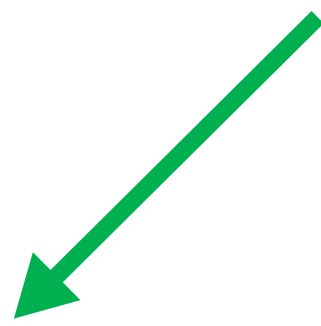
DFTとMI

現在、(計算)物質科学にはDFTとMIという2つの武器がある

DFT

機械学習(+物性データ)

MI



コンピュータシヨナル・マテリアルズ・デザイン(CMD)

これらをどのように活用するかが課題

- ・ DFTとMIの分業
- ・ MIによるDFTの機能強化・拡張

DFTとMIの分業

バイオ・インフォマティクス



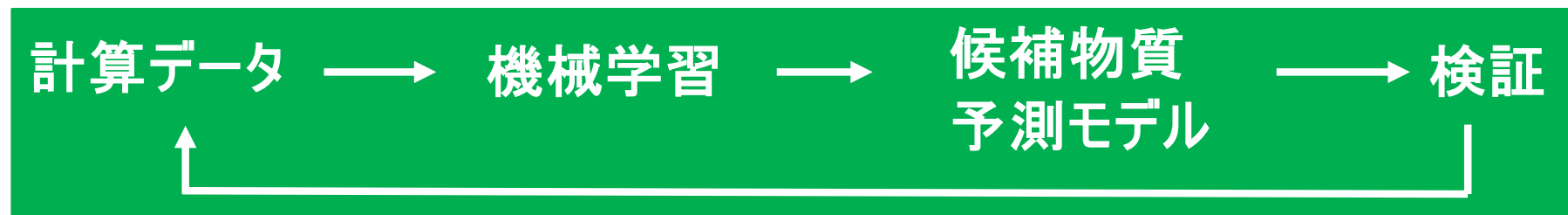
DFTのみ (従来型)

研究者の経験、試行錯誤、勘やひらめきが必要



DFT+MI

in silico



MIによるDFTの機能強化・拡張

例：結晶構造最適化

DFTの構造緩和計算による結晶構造探索には限界がある。
(初期構造としてある程度実験と近いものを用意する必要がある)

機械学習の手法を用いれば、構成元素の原子番号のみを
パラメータとして、最安定構造を得ることができる。
(それにより、未知物質の最安定構造での様々な物性値を
得ることができるようになる。)

USPEX

初期構造ランダム生成
10-100

EAによる次世代構造の生成

CrysPY

初期構造ランダム生成
1000~

BOによる効率的なサンプリング
(内部であるモデルを仮定)

少ないDFT計算で
最安定構造の発見



解釈可能モデリング

線形回帰の基礎とスパースモデリング

目次

- 解釈可能性は必要か？
- 解釈可能性の意義
- 解釈可能なモデルとは？
- モデルの「単純さ」と「整合性」の定義
- 線形モデルと非線形モデル
- 線形回帰分析
- 記述子行列と記述子ベクトルの定義
- 記述子生成の例
- LIDGでの記述子生成
- 多重共線性
- 正則化の3つのアプローチ

解釈可能性は必要か？

社会的要請:

物質探索 > 物性予測 (Black box) > 物性予測 (解釈可能モデル)

物質探索には必ずしも物性予測は必要ではない (EA)

物性予測には必ずしも解釈可能性は必要ではない (DL、RF)

また、データ科学者の多くは効率の良い最適化や、
高速・高精度予測にしか興味がない。

解釈可能性と予測精度はある程度トレードオフの関係

70%しか当たらない解釈・説明可能なモデルと、
90%当たるブラックボックス関数では、どちらが価値が高いか？

解釈可能性の意義

- ・ 機構解明、物理法則発見につながる可能性がある
(学術的に新たな概念の発見)

ティコ・ブラーエが集めた大量の惑星運行データ



ケプラーの法則(経験則)

$$\frac{a^3}{T^2} = \text{const.}$$

解釈可能モデル



ニュートンの万有引力の法則

$$F = G \frac{Mm}{r^2}$$

ニュートンが自ら構築した「運動の法則」とケプラーの経験則から、
木からリンゴを落とす力が月をも動かしているという真理を発見

解釈可能性の意義

- ・ 入力(記述子)と出力(物性値)との関係がわかる

ブラック・ボックス関数では入力と出力との間の関係が不明

- どのような入力が良い出力を持つか(予め)わからない
- また、その関数の適用範囲がわからない

一方、解釈可能なモデルであれば、どのパラメータをどれだけ上げれば物性値がどれだけ上がるか、ということがわかる。

- 良い物質候補の見当がつく
- モデルの適用範囲がわかる(可能性がある)
- 限界値がわかる
- その限界値を超えるためのヒントが得られる(可能性がある)

解釈可能モデルの強み

解釈可能なモデルとは？

1. 単純な形である
2. 経験や直感に訴える形
3. 予測精度がそこそこ高い

$$\frac{a^3}{T^2} = \text{const.}$$

よって、

解釈可能モデル(すなわち、物理的意味を抽出可能なモデル)を得るためには、モデルの構築段階で「予測精度」だけでなく、

1. 単純さ
2. 整合性

にも十分配慮する必要がある。

モデルの「単純さ」と「整合性」の定義

「単純さ」

- ・ 説明変数が少数(スパース)である
- ・ 説明変数の形が単純

価値が高いと判断

$$x_i, \frac{x_i}{x_j},$$

$$\frac{|x_i - x_j|}{\sqrt{x_i + x_j}}$$

「整合性」

- ・ 物理的整合性(物理的次元、対称性)
- ・ 事前知識、経験との整合性

これらは通常の機械学習では無視されるが、LIDG法では考慮可能

線形モデルと非線形モデル

線形モデル:

- ・ 計算コスト低い
- ・ 過適合しにくい (low variance)
(サンプル空間の変化に鈍感)
- ・ **シンプル**なモデルが得られる
→ **物理的意味を抽出できる**
- ・ **表現力が低い** (high bias)

非線形モデル:

- ・ 計算コスト高い
- ・ 過適合しやすい (high variance)
(サンプル空間の変化に敏感)
- ・ **殆どブラックボックス**なモデルが得られる
→ **物理的意味の抽出は困難**
- ・ **表現力が高い** (low bias)

表現力とシンプルさとの間のトレードオフをどのように解決するか？

→ 記述子を新たに生成し、その中から良い記述子のみを選ぶ

解釈可能モデリングの基本方針

線形回帰

記述子生成

モデル選択

線形回帰分析

1. 線形モデルの仮定

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \quad \vec{\varepsilon}: \text{誤差項 (ガウス分布)}$$

$$\hat{\vec{y}} = X\hat{\vec{\beta}} \quad \begin{array}{l} \hat{\vec{y}}: \text{推定値} \\ \hat{\vec{\beta}}: \text{回帰係数} \end{array}$$

目的変数
(従属変数)

他の物性値
(特徴量/説明変数/
記述子)

	\vec{y}	\vec{x}_1	\vec{x}_2	...	\vec{x}_n
サンプル 1	y_1	x_{11}	x_{12}	...	x_{1n}
サンプル 2	y_2	x_{21}	x_{22}	...	x_{2n}
...
サンプル m	y_m	x_{m1}	x_{m2}	...	x_{mn}

2. ある条件のもとで回帰係数 $\hat{\vec{\beta}}$ を求める

- 最小二乗法 (OLS)

$$\hat{\vec{\beta}}_{OLS} = \underset{\vec{\beta}}{\operatorname{argmin}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 \right\}$$

- LASSO

$$\hat{\vec{\beta}}_{LASSO} = \underset{\vec{\beta}}{\operatorname{argmin}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\beta}\|_{l_1} \right\}$$

$X (m \times n)$:
記述子行列 (計画行列)

L_p ノルムの定義

$$\|\vec{\beta}\|_{l_p} \equiv \left(\sum_{i=1}^n |\beta_i|^p \right)^{1/p}$$

記述子行列と記述子ベクトルの定義

計画行列
記述子行列
($m \times n + 1$)

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}$$

$\vec{y} = X\vec{\beta} + \vec{\varepsilon}$

パターン認識での記述子(特徴)ベクトルの定義

$$\vec{x}_i = [1, x_{i1}, x_{i2}, \dots, x_{in}]^T$$

$n + 1$ 次元特徴空間上のベクトル
 i 番目のサンプルに対する特徴量の配列

$$X = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m]^T$$

$$y_i = \vec{x}_i^T \vec{\beta} + \varepsilon_i \quad \vec{y} = X\vec{\beta} + \vec{\varepsilon}$$

$$(y_i = \vec{\beta}^T \vec{x}_i + \varepsilon_i)$$

n は通常固定
 m は増加する可能性がある

スパースモデリングでの記述子ベクトルの定義

$$\vec{x}_j = [x_{1j}, x_{2j}, \dots, x_{mj}]^T$$

m 次元サンプル空間上のベクトル
 j 番目の記述子の各サンプルでの値の配列

$$X = [\vec{1}, \vec{x}_1, \vec{x}_2, \dots, \vec{x}_n]$$

$$\vec{y} = \sum_{j=0}^n \vec{x}_j \beta_j + \vec{\varepsilon} \quad \vec{y} = X\vec{\beta} + \vec{\varepsilon}$$

n は増加する可能性がある
 m は通常固定

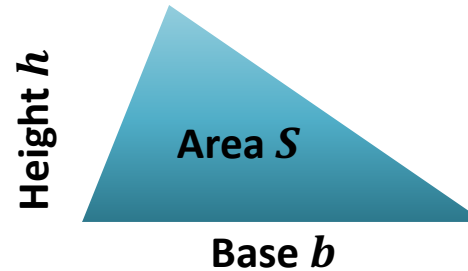
記述子生成の例

20個の様々な三角形のデータ

目的変数 記述子

Area	Base	Height
0.0001	0.2658	0.0007
0.0092	0.9789	0.0187
0.0096	0.0488	0.3931
0.0102	0.3953	0.0514
0.0195	0.0883	0.4416
0.0273	0.2153	0.2536
0.0389	0.3864	0.2012
0.0441	0.1095	0.8048
0.0562	0.8760	0.1283
0.0577	0.2917	0.3958
0.0602	0.4255	0.2832
0.0626	0.7260	0.1725
0.0788	0.1785	0.8830
0.0871	0.7943	0.2192
0.0915	0.3709	0.4936
0.0964	0.6477	0.2976
0.1051	0.5049	0.4163
0.1119	0.7204	0.3108
0.1198	0.7600	0.3154
0.1341	0.5543	0.4839

三角形の面積



底辺と高さを記述子とした線形回帰

$$S = 0.11681 b + 0.13780 h - 0.03876$$

$$Adj. R^2 = 0.5515$$

底辺と高さの積を新しい記述子とした線形回帰

$$S = 0.4998 bh - 2.333 \times 10^{-5}$$

$$Adj. R^2 = 1.0000$$

記述子の数は一つ少ないのに推定精度は良くなっている。
記述子生成の意義は、記述子の数を増加させることではなく、
良い記述子を作ることである。

このモデルから三角形の面積の法則(公式)を再発見

$$S = \frac{1}{2} bh$$

LIDGでの記述子生成1: 基本演算

基本(初期)記述子:

$$\vec{x}_1, \vec{x}_2$$

基本演算:

$$|\vec{x}_1|, |\vec{x}_2|,$$

$$\vec{x}_1^{-1}, \vec{x}_2^{-1},$$

$$|\vec{x}_1|^{-1}, |\vec{x}_2|^{-1},$$

$$|\vec{x}_1 + \vec{x}_2|, |\vec{x}_1 - \vec{x}_2|,$$

$$(\vec{x}_1 + \vec{x}_2)^{-1}, (\vec{x}_1 - \vec{x}_2)^{-1},$$

$$|\vec{x}_1 + \vec{x}_2|^{-1}, |\vec{x}_1 - \vec{x}_2|^{-1},$$

$$|\vec{x}_i| \equiv (|x_{1i}|, |x_{2i}|, \dots, |x_{mi}|)^T$$

$$\vec{x}_i^{-1} \equiv (x_{1i}^{-1}, x_{2i}^{-1}, \dots, x_{mi}^{-1})^T$$

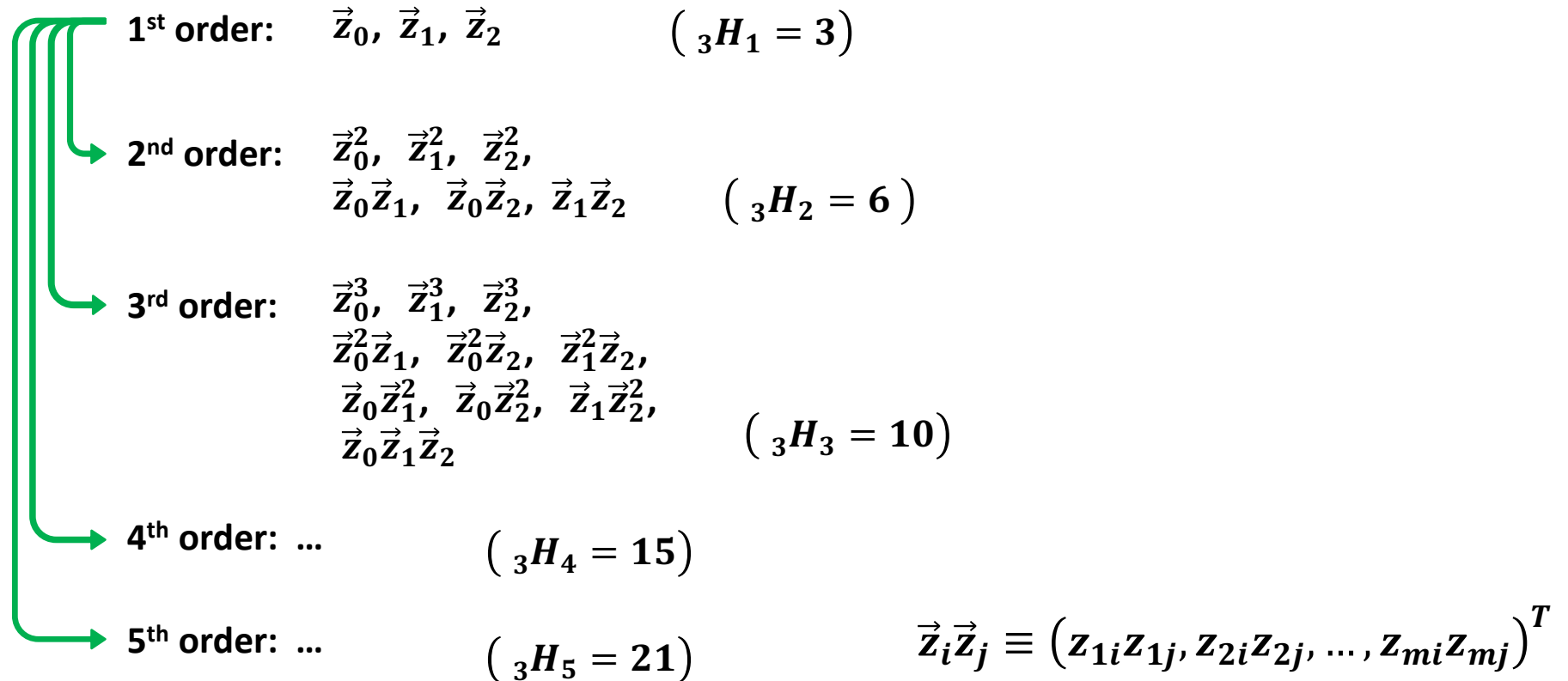
$$|\vec{x}_i|^{-1} \equiv (|x_{1i}|^{-1}, |x_{2i}|^{-1}, \dots, |x_{mi}|^{-1})^T$$

$$|\vec{x}_i + \vec{x}_j| \equiv (|x_{1i} + x_{1j}|, \dots, |x_{mi} + x_{mj}|)^T$$

もちろん、もっと複雑な基本演算(平方根や指数、対数関数など)を考えても良いが、解釈性の観点から、この時点であまりにも複雑な形を考えるのは好ましくない。これだけでも、ある程度の表現力向上が見込める。

LIDGでの記述子生成2: 直積による生成

(\vec{z}_i は基本演算で得られた1次記述子とする。 $\vec{x}_i, \vec{x}_j, 1/\vec{x}_i, |\vec{x}_i|, \dots$)



このようにして、記述子を必要なだけ段階的にシステムティックに生成できる。

記述子の複雑さも段階的に増加させることができる。

(もちろんこれらの記述子の間には後述する多重共線性が生じる可能性がある)

多重共線性

共線性

$$\vec{x}_i = c_j \vec{x}_j + c_0$$

準共線性

$$\vec{x}_i \sim c_j \vec{x}_j + c_0$$

多重共線性

$$\vec{x}_i = c_j \vec{x}_j + c_k \vec{x}_k + \dots + c_0$$

準多重共線性

$$\vec{x}_i \sim c_j \vec{x}_j + c_k \vec{x}_k + \dots + c_0$$

記述子生成を行うと、容易にこのような関係式が生成されてしまう。

何故、多重共線性を検出・除去しなければならないのか？
(何故、記述子空間を線形独立化しなければならないか？)

- ・ 回帰係数の増大、発散
- ・ サンプル空間の変化に対するモデルの不安定性(汎化性能の低下)
- ・ モデル選択の不安定性

最小二乗法 (OLS)

$$X = [\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n]$$

Definition of L_p norm

$$\hat{\vec{\beta}}_{OLS} = \underset{\vec{\beta}}{\operatorname{argmin}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 \right\}$$

$$\|\vec{\beta}\|_{l_p} \equiv \left(\sum_{i=1}^n |\beta_i|^p \right)^{1/p}$$

$$\frac{\partial L(\vec{\beta})}{\partial \vec{\beta}} = -[X^T \vec{y} - X^T X \vec{\beta}] = \vec{0}$$

$$\hat{\vec{\beta}}_{OLS} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \frac{\Delta}{\det|X^T X|}$$

但し、 $X^T X$ は正則行列であるとする。
つまり、 $X^T X$ は逆行列を持つとする。

$$\det|X^T X| \neq 0$$

→ これはつまり、 $\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n$ が線型独立であるということ。

正則化の3つのアプローチ

$X = [\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n]$ において、
 $\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n$ が線形独立でなければ？

1. 罰則項を導入し正則化して、緩和問題を解く(Ridge, LASSO)
2. 基底変換により、直交基底を作る (PCR, PLS)
3. 多重共線性関係を全て見つけて除去し、線型独立化させる(LIDG)

Ridge and LASSO

- **OLS** (ordinary least-squares) method

$$\operatorname{argmin}_{\vec{\beta}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 \right\}$$

- **Ridge regression**

$$\operatorname{argmin}_{\vec{\beta}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\beta}\|_{l_2}^2 \right\}$$

Regularization term (to avoid “over-fitting” problem)

$$\|\vec{\beta}\|_{l_p} \equiv \left(\sum_{i=1}^n |\beta_i|^p \right)^{1/p}$$

- **LASSO** (least absolute shrinkage and **selection** operator)

$$\operatorname{argmin}_{\vec{\beta}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\beta}\|_{l_1} \right\}$$

“selection” means some β_i are to be 0

Ridge and LASSO

$$\|\vec{\beta}\|_{l_2} = \sqrt{|\beta_1|^2 + |\beta_2|^2 + \dots + |\beta_n|^2}$$

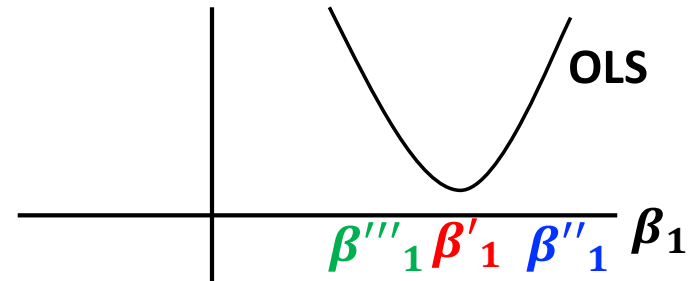
$$\|\vec{\beta}\|_{l_1} = |\beta_1| + |\beta_2| + \dots + |\beta_n|$$

1D:

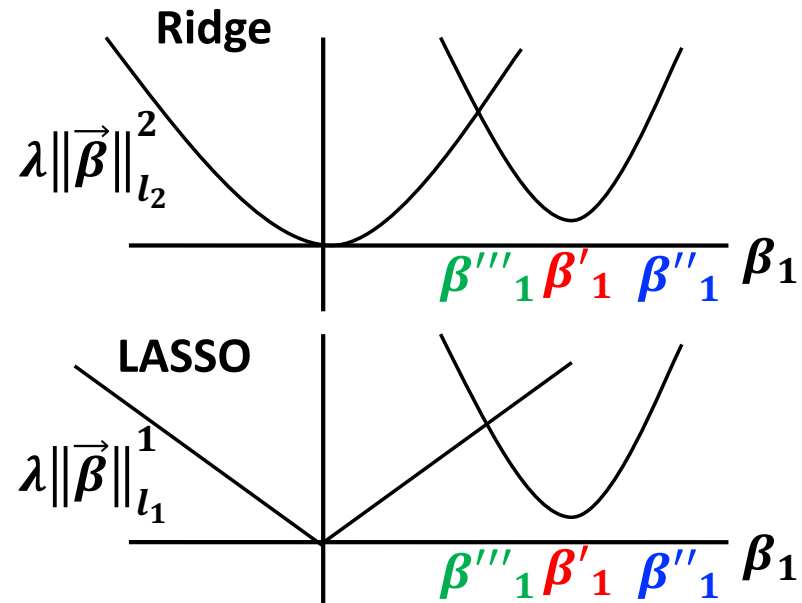
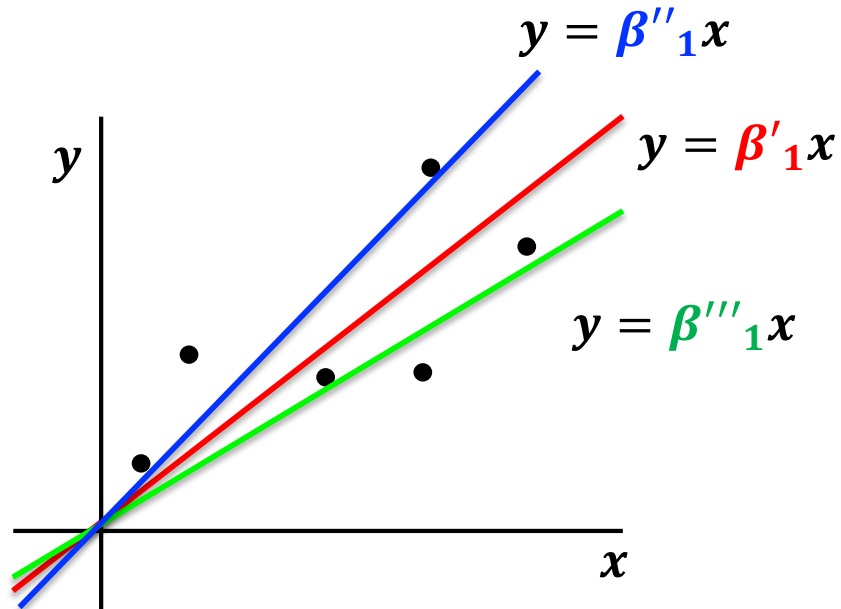
$$\|\vec{\beta}\|_{l_2} = \sqrt{|\beta_1|^2}, \quad \|\vec{\beta}\|_{l_2}^2 = |\beta_1|^2$$

$$\|\vec{\beta}\|_{l_1} = |\beta_1|$$

OLS	$\underset{\vec{\beta}}{\operatorname{argmin}} \left\{ \ \vec{y} - X\vec{\beta}\ _{l_2}^2 \right\}$
Ridge	$\underset{\vec{\beta}}{\operatorname{argmin}} \left\{ \ \vec{y} - X\vec{\beta}\ _{l_2}^2 + \lambda \ \vec{\beta}\ _{l_2}^2 \right\}$
LASSO	$\underset{\vec{\beta}}{\operatorname{argmin}} \left\{ \ \vec{y} - X\vec{\beta}\ _{l_2}^2 + \lambda \ \vec{\beta}\ _{l_1} \right\}$



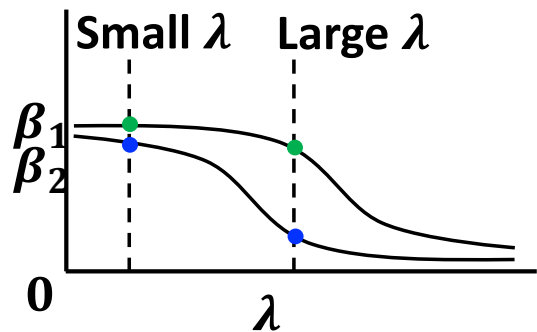
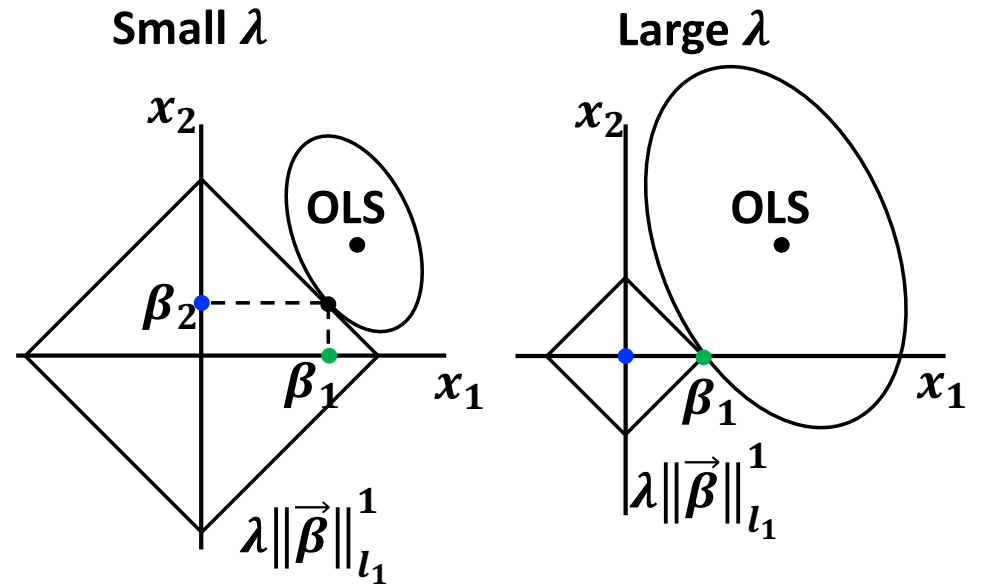
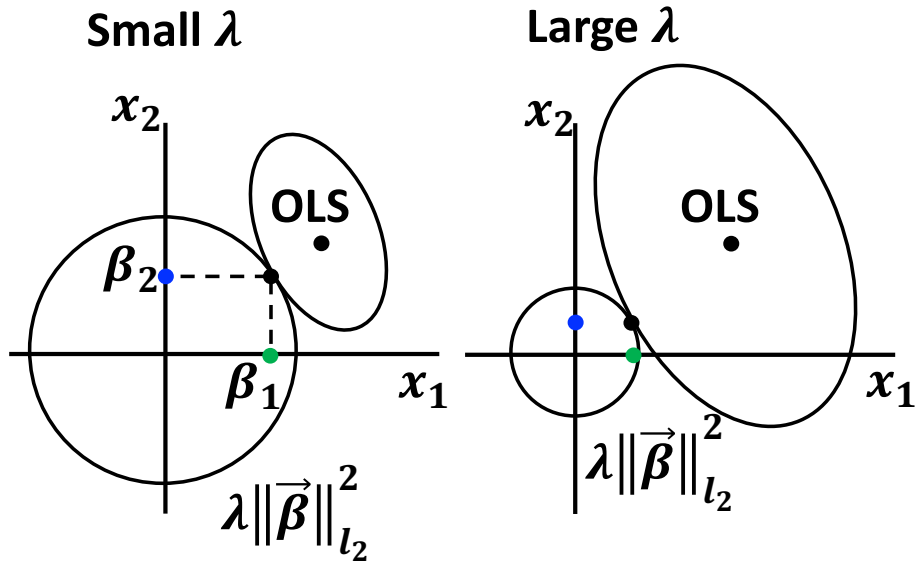
Generally, small $\|\vec{\beta}\|$ are better



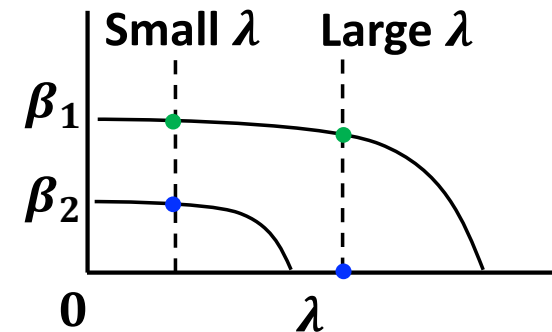
Ridge and LASSO

Ridge $\operatorname{argmin}_{\vec{\beta}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\beta}\|_{l_2}^2 \right\}$

LASSO $\operatorname{argmin}_{\vec{\beta}} \left\{ \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\beta}\|_{l_1}^1 \right\}$



Regularization path
Solution path



Ridge regression

$$L(\vec{\beta}) = f(\vec{\beta}) + f_2(\vec{\beta})$$

$$f(\vec{\beta}) = \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2$$

$$\widehat{\vec{\beta}}_{\text{Ridge}} = \underset{\vec{\beta}}{\operatorname{argmin}} \{L(\vec{\beta})\}$$

$$f_2(\vec{\beta}) = \frac{\lambda}{2} \|\vec{\beta}\|_{l_2}^2$$

$$\frac{\partial f(\vec{\beta})}{\partial \vec{\beta}} = -\frac{1}{m} [X^T \vec{y} - X^T X \vec{\beta}]$$

$$\frac{\partial f_2(\vec{\beta})}{\partial \vec{\beta}} = \lambda \vec{\beta}$$

$$\frac{\partial L(\vec{\beta})}{\partial \vec{\beta}} = -\frac{1}{m} [X^T \vec{y} - X^T X \vec{\beta}] + \lambda \vec{\beta} = \vec{0}$$

$$\widehat{\vec{\beta}}_{\text{Ridge}} = (X^T X + m\lambda I)^{-1} X^T \vec{y}$$

Even if $X^T X$ is not regular, $(X^T X + m\lambda I)$ is always regular.

LASSO

$$L(\vec{\beta}) = f(\vec{\beta}) + f_1(\vec{\beta})$$

$$f(\vec{\beta}) = \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2$$

$$\hat{\vec{\beta}}_{\text{LASSO}} = \underset{\vec{\beta}}{\operatorname{argmin}} \{L(\vec{\beta})\}$$

$$f_1(\vec{\beta}) = \lambda \|\vec{\beta}\|_{l_1}^1$$

$$\frac{\partial f(\vec{\beta})}{\partial \vec{\beta}} = -\frac{1}{m} [X^T \vec{y} - X^T X \vec{\beta}]$$

$$\frac{\partial f_1(\vec{\beta})}{\partial \vec{\beta}} = \times$$

Unlike Ridge regression, we can not find any analytic solution
 f_1 is not continuously differentiable.
→ we can not use any gradient method.

LASSO

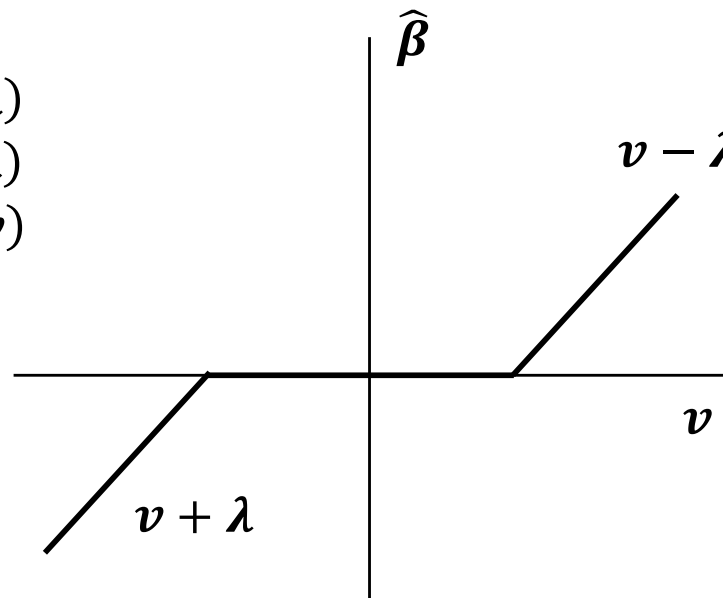
$$\operatorname{argmin}_{\vec{\beta}} \left\{ \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\beta}\|_{l_1} \right\} \quad \text{can not be solved}$$

However, a special form like,

$$\operatorname{argmin}_{\vec{\beta}} \left\{ \frac{1}{2\lambda} \|\vec{v} - \vec{\beta}\|_{l_2}^2 + \|\vec{\beta}\|_{l_1} \right\} \quad \text{can be solved}$$

$$\hat{\beta} = \begin{cases} v - \lambda & (v > \lambda) \\ 0 & (-\lambda \leq v \leq \lambda) \\ v + \lambda & (-\lambda < v) \end{cases}$$

Soft thresholding function



LASSO

$\operatorname{argmin}_{\vec{\beta}} \{f(\vec{\beta}) + f_1(\vec{\beta})\}$ can not be solved

ADMM (Alternating Direction Method of Multipliers)

$\operatorname{argmin}_{\vec{\beta}, \vec{\gamma}} \{f(\vec{\beta}) + f_1(\vec{\gamma}) + g(\vec{\gamma} - \vec{\beta})\}$

$$f(\vec{\beta}) = \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2$$

$$f_1(\vec{\gamma}) = \lambda \|\vec{\gamma}\|_{l_1}^1$$

$$g(\vec{\gamma} - \vec{\beta}) = \vec{h}^T (\vec{\gamma} - \vec{\beta}) + \mu \|\vec{\gamma} - \vec{\beta}\|_{l_2}^2$$

LASSO

Completing the square of g

$$\begin{aligned}g(\vec{\gamma} - \vec{\beta}) &= \vec{h}^T (\vec{\gamma} - \vec{\beta}) + \mu \|\vec{\gamma} - \vec{\beta}\|_{l_2}^2 \\ &= \frac{\mu}{2} \left\| \vec{\gamma} - \left(\vec{\beta} - \frac{1}{\mu} \vec{h} \right) \right\|_{l_2}^2 - \frac{1}{2\mu} \|\vec{h}\|_{l_2}^2\end{aligned}$$

$$\operatorname{argmin}_{\vec{\beta}, \vec{\gamma}} \left\{ f(\vec{\beta}) + f_1(\vec{\gamma}) + g(\vec{\gamma} - \vec{\beta}) \right\}$$

$$\operatorname{argmin}_{\vec{\beta}, \vec{\gamma}} \left\{ \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\gamma}\|_{l_1} + \frac{\mu}{2} \left\| \vec{\gamma} - \left(\vec{\beta} - \frac{1}{\mu} \vec{h} \right) \right\|_{l_2}^2 - \frac{1}{2\mu} \|\vec{h}\|_{l_2}^2 \right\}$$

$$\frac{\partial L(\vec{\beta}; \vec{\gamma})}{\partial \vec{\beta}} = -\frac{1}{m} [X^T \vec{y} - X^T X \vec{\beta}] - \mu \left(\vec{\gamma} - \vec{\beta} + \frac{1}{\mu} \vec{h} \right) = \vec{0}$$

$$\widehat{\vec{\beta}}_{\text{LASSO}} = (X^T X + m\mu I)^{-1} (X^T \vec{y} + m\mu \vec{\gamma} + m\vec{h})$$

LASSO

$$\operatorname{argmin}_{\vec{\beta}, \vec{\gamma}} \left\{ f(\vec{\beta}) + f_1(\vec{\gamma}) + \vec{h}^T(\vec{\gamma} - \vec{\beta}) + \mu \|\vec{\gamma} - \vec{\beta}\|_{l_2}^2 \right\}$$

$$\operatorname{argmin}_{\vec{\beta}, \vec{\gamma}} \left\{ \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\gamma}\|_{l_1} + \frac{\mu}{2} \left\| \vec{\gamma} - \left(\vec{\beta} - \frac{1}{\mu} \vec{h} \right) \right\|_{l_2}^2 - \frac{1}{2\mu} \|\vec{h}\|_{l_2}^2 \right\}$$

$$\operatorname{argmin}_{\vec{\gamma}} \left\{ \lambda \|\vec{\gamma}\|_{l_1} + \frac{\mu}{2} \left\| \vec{\gamma} - \left(\vec{\beta} - \frac{1}{\mu} \vec{h} \right) \right\|_{l_2}^2 \right\}$$

$$\operatorname{argmin}_{\vec{\gamma}} \left\{ \|\vec{\gamma}\|_{l_1} + \frac{1}{2\lambda'} \|\vec{\gamma} - \vec{v}\|_{l_2}^2 \right\} \quad \text{Soft thresholding function}$$

$$\hat{\gamma}_{\text{LASSO}} = \begin{cases} v - \lambda' & (v > \lambda') \\ 0 & (-\lambda' \leq v \leq \lambda') \\ v + \lambda' & (-\lambda' < v) \end{cases} \quad \lambda' = \frac{\lambda}{\mu}, \quad \vec{v} = \vec{\beta} - \frac{1}{\mu} \vec{h}$$

$$\hat{\beta}_{\text{LASSO}} = (X^T X + m\mu I)^{-1} (X^T \vec{y} + m\mu \vec{\gamma} + m\vec{h})$$

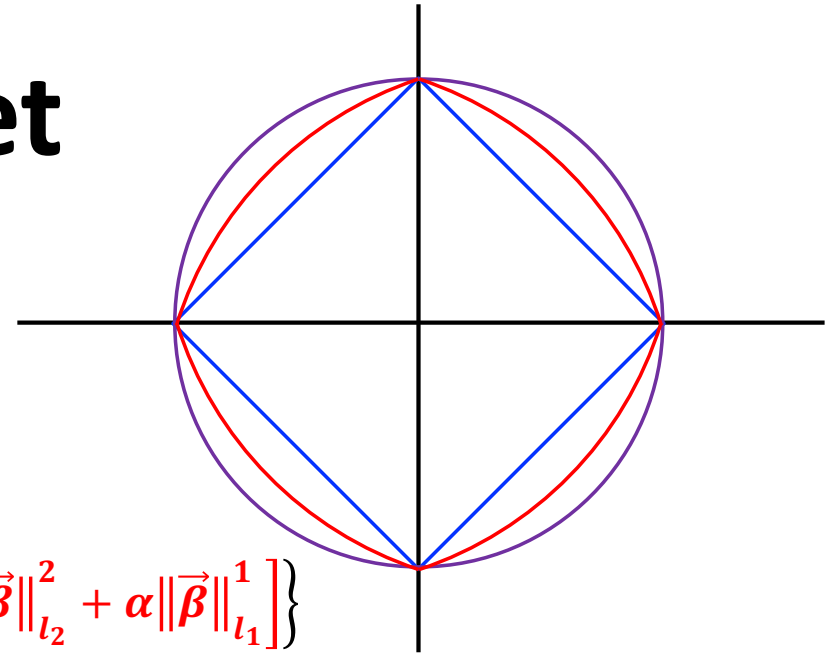
$$\hat{h}_{\text{new}} = \hat{h}_{\text{old}} + \mu (\hat{\gamma}_{\text{LASSO}} - \hat{\beta}_{\text{LASSO}})$$

Elastic net

Ridge $\operatorname{argmin}_{\vec{\beta}} \left\{ \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \frac{\lambda}{2} \|\vec{\beta}\|_{l_2}^2 \right\}$

LASSO $\operatorname{argmin}_{\vec{\beta}} \left\{ \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \|\vec{\beta}\|_{l_1} \right\}$

Elastic net $\operatorname{argmin}_{\vec{\beta}} \left\{ \frac{1}{2m} \|\vec{y} - X\vec{\beta}\|_{l_2}^2 + \lambda \left[\frac{(1-\alpha)}{2} \|\vec{\beta}\|_{l_2}^2 + \alpha \|\vec{\beta}\|_{l_1} \right] \right\}$



$\alpha = 0 \Rightarrow \text{Ridge}$

$\alpha = 1 \Rightarrow \text{LASSO}$

$$\hat{\mathbf{Y}}_{\text{Enet}} = \begin{cases} v - \lambda' & (v > \lambda') \\ \mathbf{0} & (-\lambda' \leq v \leq \lambda') \\ v + \lambda' & (-\lambda' < v) \end{cases} \quad \lambda' = \frac{\lambda\alpha}{\mu} \quad \vec{v} = \vec{\beta} + \frac{1}{\mu} \vec{h}$$

$$\hat{\vec{\beta}}_{\text{Enet}} = (X^T X + m\lambda(1-\alpha) + m\mu I)^{-1} (X^T \vec{y} + m\mu \vec{\gamma} - m\vec{h})$$

$$\hat{\vec{h}}_{\text{new}} = \hat{\vec{h}}_{\text{old}} + \mu (\hat{\vec{\beta}}_{\text{Enet}} - \hat{\vec{Y}}_{\text{Enet}})$$

正則化の3つのアプローチ

$X = [\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n]$ において、
 $\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n$ が線形独立でなければ？

1. 罰則項を導入し正則化して、緩和問題を解く(Ridge, LASSO)
2. 基底変換により、直交基底を作る (PCR, PLS)
3. 多重共線性関係を全て見つけて除去し、線型独立化させる(LIDG)

Ridge, LASSOの具体例、PCR、PLSは実習にて行う予定

Multicollinearity

$$\vec{x}_i = c_1\vec{x}_1 + c_2\vec{x}_2 + \dots + c_n\vec{x}_n$$

$$\vec{x}_i = X^{(-i)}\vec{c}^{(-i)}$$

$$\vec{0} = X\vec{c}$$

i.e., the existence of MCL means the existence of non-trivial solutions of $X\vec{c} = \vec{0}$.

→ for finding MCL, we find the non-trivial solutions.

The condition that $X\vec{c} = \vec{0}$ has non-trivial solutions

1. Under-determined system, i.e., $m < n$.
2. Rank deficient matrix, i.e., $\text{rank}(X) \equiv r < \min(m, n)$

$$m \begin{matrix} & & n & & \\ \begin{bmatrix} 1 & 0 & 0 & 0 & a \\ 0 & 1 & 0 & 0 & b \\ 0 & 0 & 1 & 0 & c \\ 0 & 0 & 0 & 1 & d \end{bmatrix} & & & & \end{matrix}$$

$$m \begin{matrix} & & n & & \\ \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & & & & \end{matrix}$$

The number of non-trivial solutions

$$n - r$$

The number of multicollinearities

Linear span:

$$W = \text{span}(\{\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n\})$$

Design matrix:

$$X = [\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n]$$

The number of non-trivial solutions of $X\vec{c} = \vec{0}$
means the number of extra basis in span W

$$n - \dim(W)$$

$$\dim(W) = \text{rank}(X) \equiv r$$

then, the number of non-trivial (independent) solutions is given by $n - r$

independent one set of solution

RREF method

Make row reduced echelon form (rref) of X by basic operations.

$$RXQ = X_{rref}$$

$R(m \times m)$: row basic transformation operator (regular)

$Q(n \times n)$: column basic transformation operator (regular and orthogonal)
(here suppose that it just change the order of columns)

$$X_{rref} = \begin{bmatrix} 1 & 0 & 0 & a_1 & b_1 & d_1 \\ 0 & 1 & 0 & a_2 & b_2 & d_2 \\ 0 & 0 & 1 & a_3 & b_3 & d_3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$n - r = 6 - 3 = 3$$

$$X_{rref}[\vec{c}_1, \vec{c}_2, \vec{c}_3] = [\vec{0}, \vec{0}, \vec{0}]$$

The simplest solutions:

$$[\vec{c}_1, \vec{c}_2, \vec{c}_3] = \begin{bmatrix} -a_1 & -b_1 & -d_1 \\ -a_2 & -b_2 & -d_2 \\ -a_3 & -b_3 & -d_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & a_1 & b_1 & d_1 \\ 0 & 1 & 0 & a_2 & b_2 & d_2 \\ 0 & 0 & 1 & a_3 & b_3 & d_3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \\ -a_3 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \vec{0}$$

RREF method

The solutions of $X_{rref}\vec{c} = \vec{0}$ are the solutions of $X\vec{c} = \vec{0}$?

$$RXQ = X_{rref} \quad X = R^{-1}X_{rref}Q^T$$

$$X_{rref}\vec{c} = \vec{0}$$

$$R^{-1}X_{rref}\vec{c} = \vec{0}$$

$$R^{-1}X_{rref}Q^T Q\vec{c} = \vec{0}$$

$$XQ\vec{c} = \vec{0}$$

How to obtain Q

We only have to remember the order change

$$XQ = [\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4]Q = [\vec{x}_3, \vec{x}_1, \vec{x}_2, \vec{x}_4]$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} Q = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Subspace selection method

From RREF method,

$$\vec{x}_k = c_i^k \vec{x}_i + c_j^k \vec{x}_j$$

$$\vec{x}_l = c_i^l \vec{x}_i + c_j^l \vec{x}_j$$

$$\vec{x}_m = c_i^m \vec{x}_i + c_j^m \vec{x}_j$$

$$W' = \text{span}(\{\vec{x}_i, \vec{x}_j, \vec{x}_k, \vec{x}_l, \vec{x}_m\})$$

$$\dim(W') = \text{rank}(X') = 2$$

List representation

$$[\vec{x}_k, \vec{x}_l, \vec{x}_m, [\vec{x}_i, \vec{x}_j]]$$

Temporary basis

The number of basis of this span
 $\text{span}(\{\vec{x}_i, \vec{x}_j, \vec{x}_k, \vec{x}_l, \vec{x}_m\})$
is only two.

Choose two basis in this list

Select \vec{x}_i, \vec{x}_j and remove $\vec{x}_k, \vec{x}_l, \vec{x}_m$

or

select \vec{x}_i, \vec{x}_k and remove $\vec{x}_j, \vec{x}_l, \vec{x}_m$

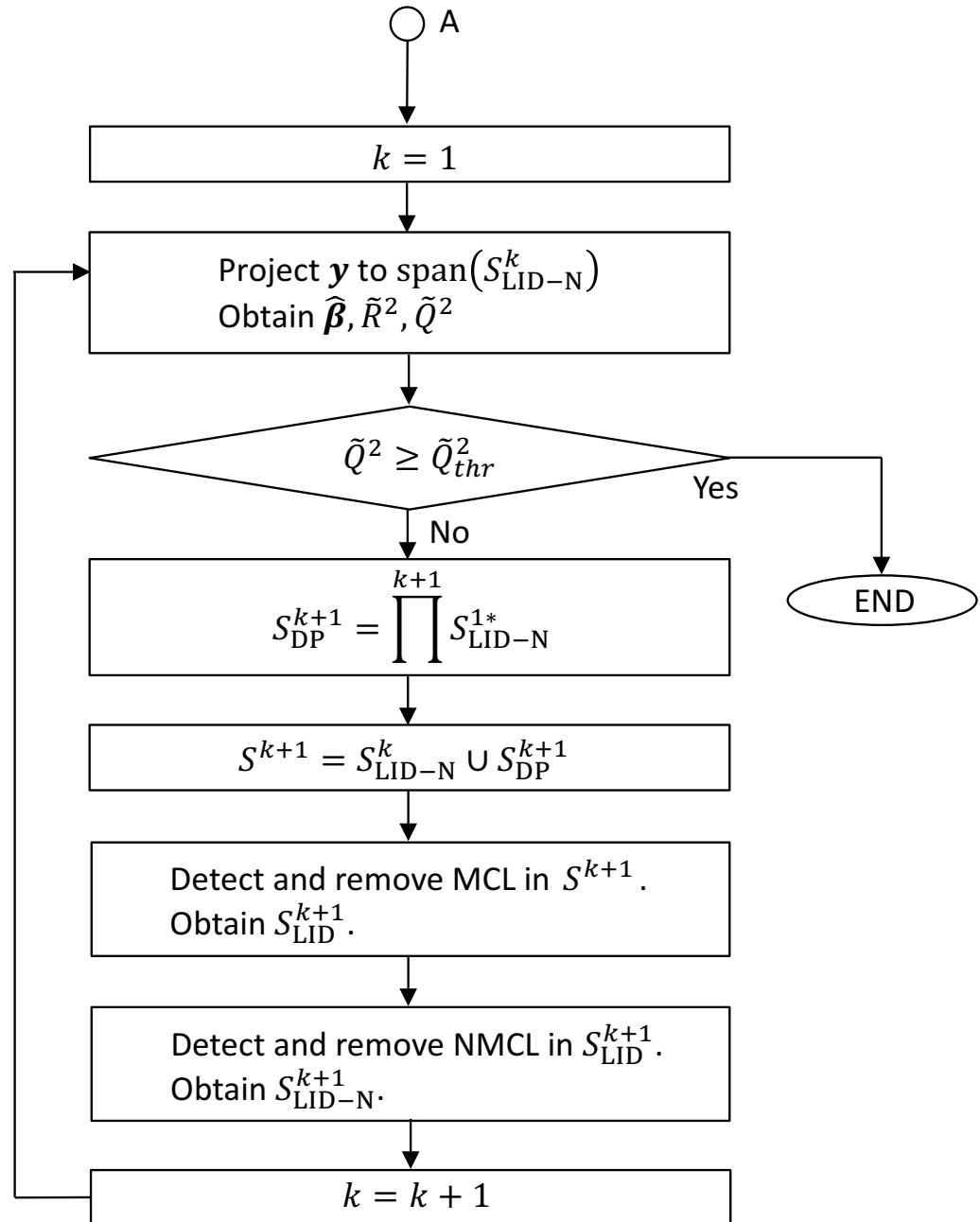
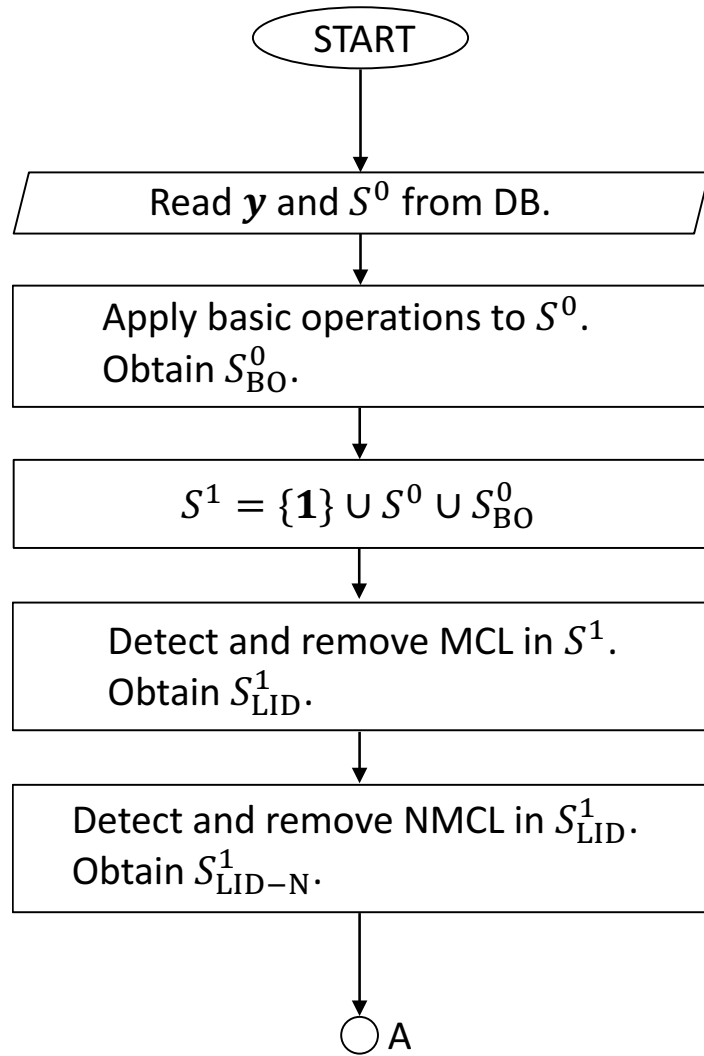
or

select \vec{x}_l, \vec{x}_m and remove $\vec{x}_i, \vec{x}_j, \vec{x}_k$

...

We can reduce the number of descriptors
without any deterioration in regression accuracy

LIDG method



先行研究の紹介
LIDGの適用例、先行研究との比較

先行研究

L. M. Ghiringhelli, *et al.*, PRL 114, 105503 (2015)

1. ある少数の初期記述子に対し、四則演算や和や差の絶対値や指数など、あらゆる演算を施して高次元記述子空間を作る
2. その後LASSO (least absolute shrinkage and selection operator) と呼ばれる罰則項付きの線形回帰によって少数の記述子に絞り込む

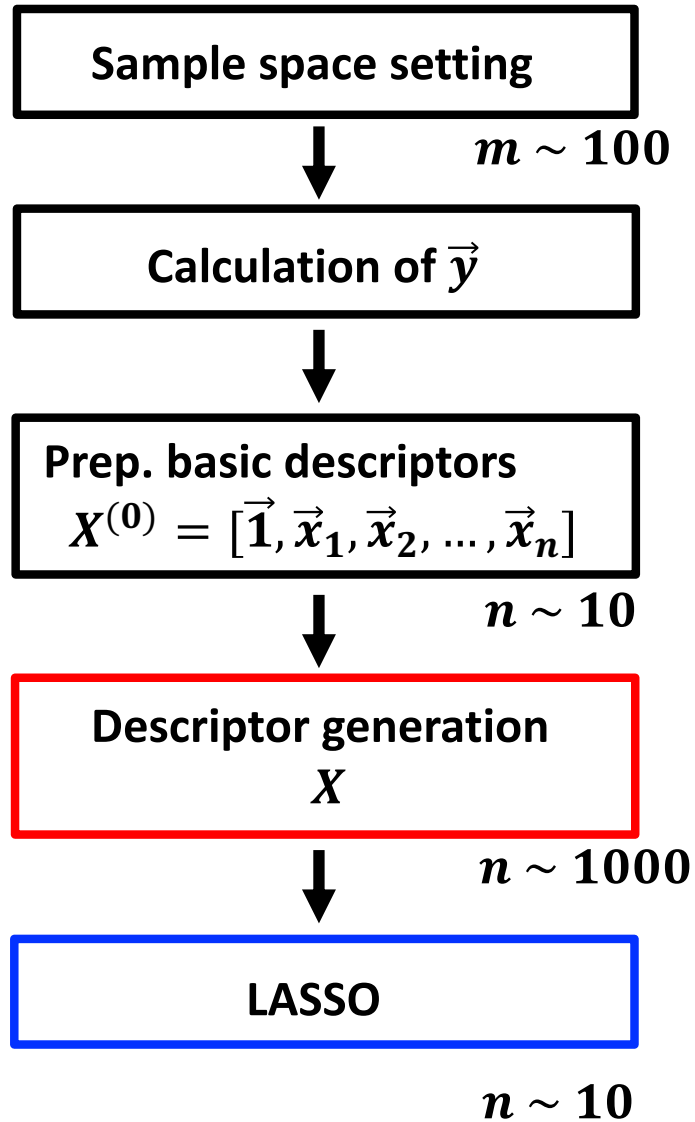
これにより、
表現力、汎化性能が高く、且つ、比較的シンプルな回帰モデルを抽出することに成功

$$\begin{aligned}\Delta E &\equiv E_{AB}(RS) - E_{AB}(ZB) \\ &= 0.108 \frac{EA(B) - IP(B)}{r_p^2(A)} + 1.79 \frac{|r_s(A) - r_p(B)|}{\exp(r_s(A))} + 3.766 \frac{|r_p(B) - r_s(B)|}{\exp(r_d(A))} - 0.0267\end{aligned}$$

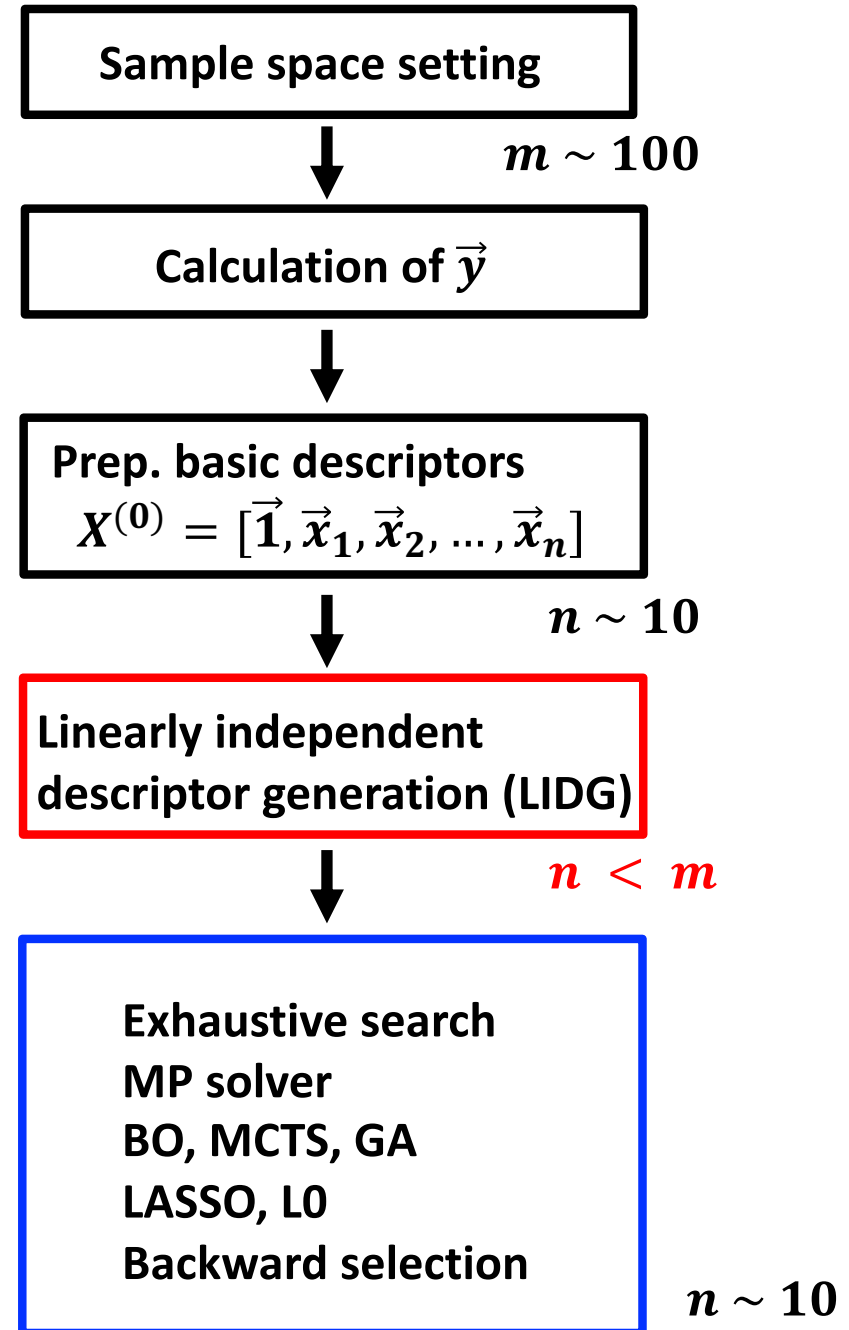
先行研究の問題点

1. 得られた回帰式がまだ複雑
分母に指数関数が含まれていたりして、直ぐには物理的意味を抽出することができない
2. LASSOによるモデル選択においては、記述子の間に強い相関がある場合、どの記述子を選ばれるかコントロールできない
3. (記述子生成の部分がシステムティックでない)

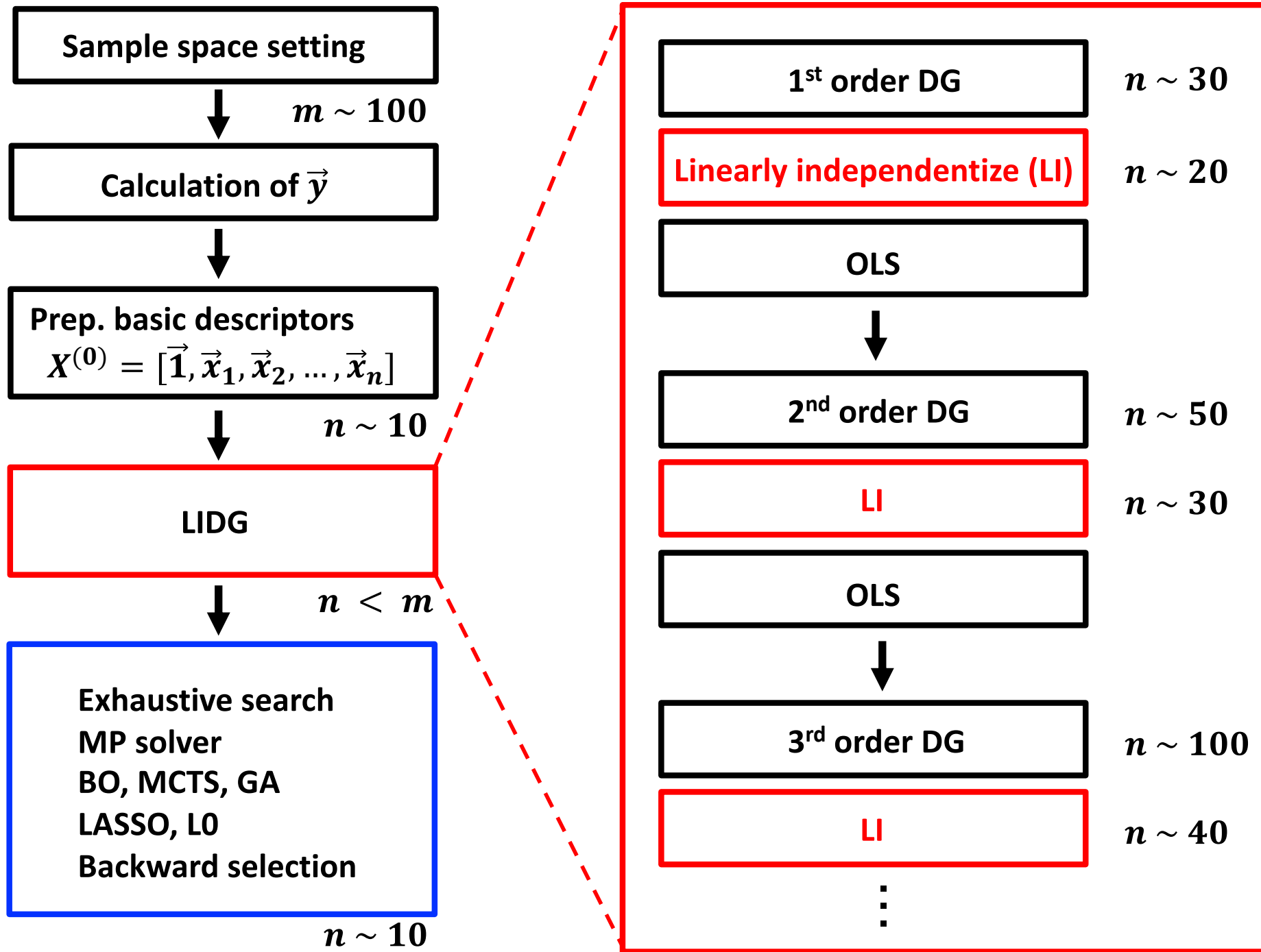
Previous work



Our approach



Linearly independent descriptor generation method

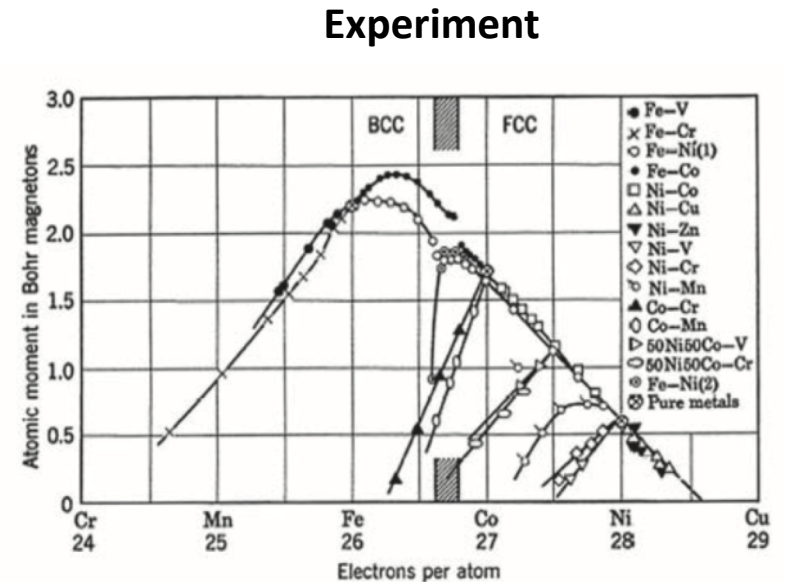
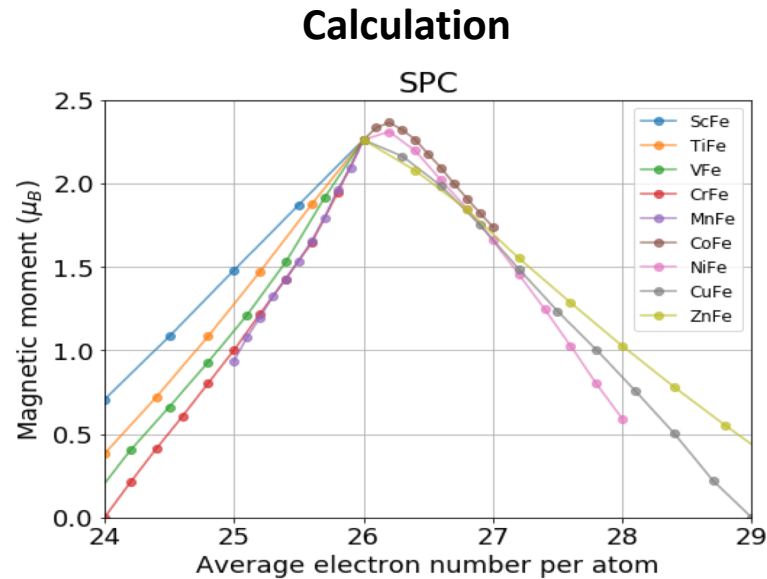


先行研究との比較、LIDG法の優位性

- ・ 記述子を段階的に増加させることができる。
また、記述子を段階的に複雑なものにすることができる。
- ・ 記述子を必要な数だけ生成することができる。(後のモデル選択が楽になる)
- ・ 記述子選択に研究者の知識・経験をある程度導入することができる。
(部分空間リスト)
- ・ 多重共線性や強い準多重共線性が取り除かれているので、
 - 回帰係数が無用に大きくなり、有意な回帰分析を行うことができる
 - 安定したモデル選択が可能
(最も単純な後退選択法も、LASSOも同様のモデルを選択する)
- ・ 完全な多重共線性関係が得られるので、それを使って、最終的に得られたモデルの式変形を行うことができる。

Example

AkaiKKR (KKR-CPA)
 LDA (MJW)
 Crystal structure: BCC
 Lattice constant = 2.86 Å
 Scalar relativistic
 BZ quality = 10 (nk=256)
 edelta = 0.0001
 ewidth = 1.2
 Complex energy mesh = 85
 L max = 3



H. Saito: Physics and Applications of Invar Alloys, Maruzen, Tokyo, (1978), 18.

Everyone may image that there is a simple model on the back of this curve
 We try to find the simple model by using LIDG method

Target property ($m = 99$):

Magnetic moment of binary alloy $A_{1-x}B_x$

$$M(A, B, x)$$

Initial descriptors ($n = 7$):

Concentration, x

Magnetic moment of pure bulk, $M_P(A), M_P(B)$

The number of valence electrons, $Z(A), Z(B)$

Magnetic moment of impurity atom, $M_I(A, B)$

$$M_I(B, A)$$

Raw data

Label	$M(A, B, x)$	x	$M_P(A)$	$Z(A)$	$M_P(B)$	$Z(B)$	$M_I(A, B)$	$M_I(B, A)$
Sc100Fe000	0.0000	0.0	0.0000	3	2.2604	8	-0.32667	0.00000
Sc090Fe010	0.0000	0.1	0.0000	3	2.2604	8	-0.32667	0.00000
Sc080Fe020	0.0000	0.2	0.0000	3	2.2604	8	-0.32667	0.00000
Sc070Fe030	0.0000	0.3	0.0000	3	2.2604	8	-0.32667	0.00000
Sc060Fe040	0.0000	0.4	0.0000	3	2.2604	8	-0.32667	0.00000
Sc050Fe050	0.3157	0.5	0.0000	3	2.2604	8	-0.32667	0.00000
Sc040Fe060	0.7056	0.6	0.0000	3	2.2604	8	-0.32667	0.00000
Sc030Fe070	1.0831	0.7	0.0000	3	2.2604	8	-0.32667	0.00000
Sc020Fe080	1.4793	0.8	0.0000	3	2.2604	8	-0.32667	0.00000
Sc010Fe090	1.8709	0.9	0.0000	3	2.2604	8	-0.32667	0.00000
Sc000Fe100	2.2604	1.0	0.0000	3	2.2604	8	-0.32667	0.00000
Ti100Fe000	0.0000	0.0	0.0000	4	2.2604	8	-0.69382	0.00000
Ti090Fe010	0.0000	0.1	0.0000	4	2.2604	8	-0.69382	0.00000
Ti080Fe020	0.0000	0.2	0.0000	4	2.2604	8	-0.69382	0.00000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Zn040Fe060	1.2862	0.6	0.0000	12	2.2604	8	0.03119	0.00000
Zn030Fe070	1.5524	0.7	0.0000	12	2.2604	8	0.03119	0.00000
Zn020Fe080	1.8440	0.8	0.0000	12	2.2604	8	0.03119	0.00000
Zn010Fe090	2.0763	0.9	0.0000	12	2.2604	8	0.03119	0.00000
Zn000Fe100	2.2604	1.0	0.0000	12	2.2604	8	0.03119	0.00000

Descriptors

M_p : Magnetic moment of pure bulk A

Z: # of valence electrons

A	$M_p(A)$	Z(A)
Sc	0.000	3
Ti	0.000	4
V	0.005	5
Cr	0.000	6
Mn	0.937	7
Fe	2.260	8
Co	1.740	9
Ni	0.591	10
Cu	0.000	11
Zn	0.000	12

$M_I(A,B)$:

Magnetic moment of impurity A (B)
in bulk B (A)

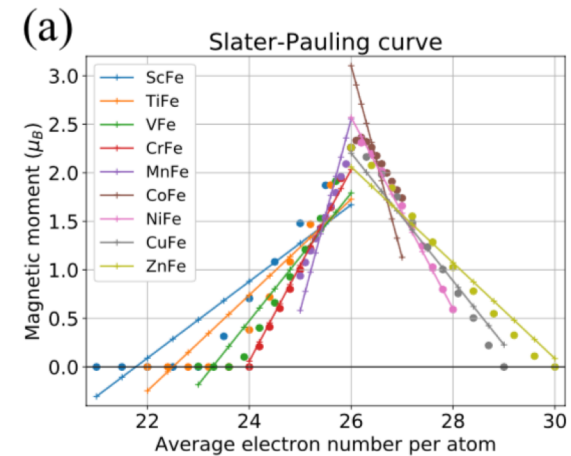
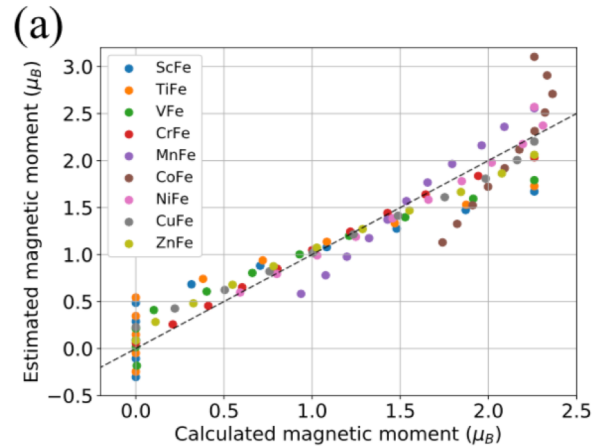
A	B	$M_I(A, B)$	$M_I(B, A)$
Sc	Fe	-0.327	0.000
Ti	Fe	-0.694	0.000
V	Fe	-1.165	-0.005
Cr	Fe	-1.644	1.825
Mn	Fe	-1.769	2.387
Co	Fe	1.846	2.673
Ni	Fe	1.075	2.797
Cu	Fe	0.268	2.332
Zn	Fe	0.031	0.000

The results of OLS with 1st, 2nd, and 3rd order descriptors

1st order (8)

$$R^2 = 0.906$$

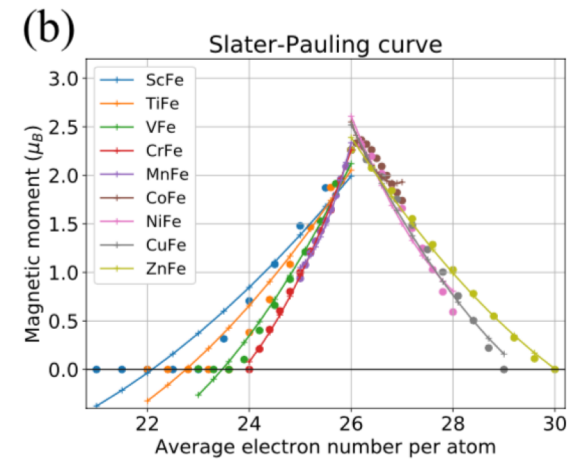
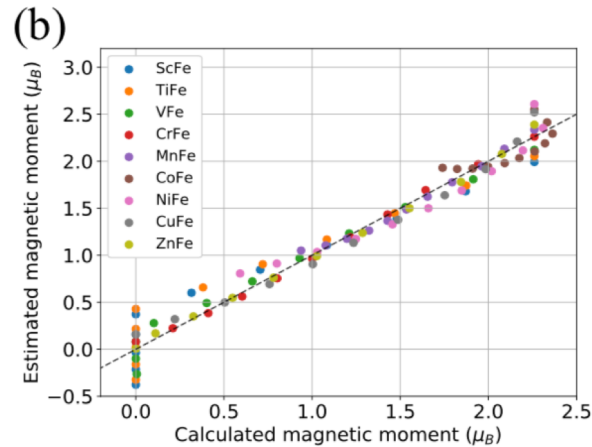
$$Q^2 = 0.887$$



Up to 2nd order (17)

$$R^2 = 0.969$$

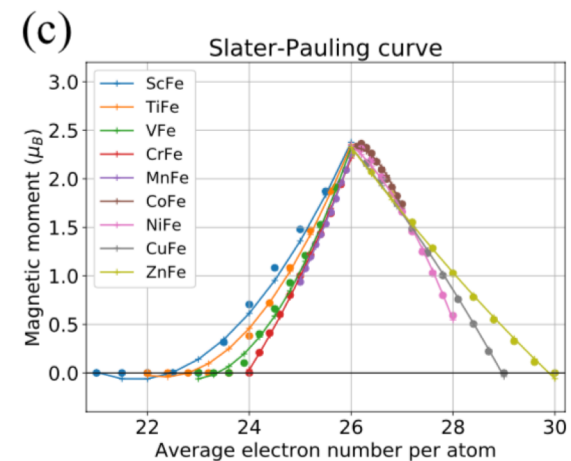
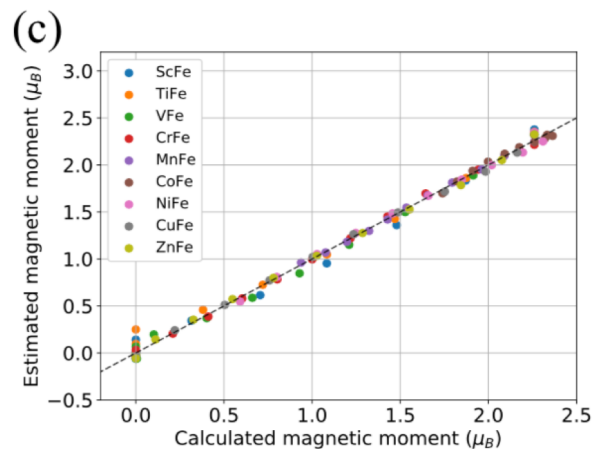
$$Q^2 = 0.950$$



Up to 3rd order (24)

$$R^2 = 0.996$$

$$Q^2 = 0.991$$



Obtained model by backward selection

$$M(A, B, x) \sim (1 - x)M_P(A) + xM_P(B) \\ + x(1 - x)[-4.03 + 0.28Z(A) + 0.62M_I(B, A) + 0.20|M_P(A) + M_I(A, B)|]$$

An interpretation

by analogy with the regular solution approximation for binary compound system

$$M(A, B, x) = \bar{M}(A, B, x) + \Delta M(A, B, x)$$

$$\bar{M}(A, B, x) \equiv (1 - x)M_P(A) + xM_P(B) \quad \text{Averaged magnetic moment term} \\ \text{(linear to } x)$$

$$\Delta M(A, B, x) \equiv x(1 - x)\Omega(A, B)$$

Mixing (excessive) magnetic moment term generated by alloying

$$\Omega(A, B) = -4.03 + 0.28Z(A) + 0.62M_I(B, A) + 0.20|M_P(A) + M_I(A, B)|$$

Interaction parameter (does not depend on x)

Summary

- **We propose RREF method for detecting MCL**
- **The subspace list is useful for breaking the detected MCL relationships**
- **This is a new approach to solve MCL problem in linear regression analysis**
- **By combining these methods, LIDG method was proposed**
- **LIDG method was applied to analyze SPCs and we could obtain a simple (interpretable) model with high generalization capability.**

Use of symmetry in target property

If we can find symmetries,

1. we can increase samples which are symmetrically equivalent.
2. we can construct descriptors to satisfy the above regulations in advance

Magnetic moment of alloys expected to have a symmetry like,

$$M(A, B, x) = M(B, A, 1 - x)$$

$$M(A, B, x_A, x_B) = M(B, A, x_B, x_A)$$

A, B, x_A, x_B : identifiers (primary key)
(not descriptor)

$$\mathbf{E}M(A, B, x_A, x_B) = M(A, B, x_A, x_B)$$

$$\boldsymbol{\sigma}M(A, B, x_A, x_B) = M(B, A, x_B, x_A) = M(A, B, x_A, x_B)$$

Symmetrization operator (projection operator):

$$\mathbf{S} = \mathbf{E} + \boldsymbol{\sigma}$$

Symmetrization of descriptors

Symmetrization operator:

$$\mathbf{S} = \mathbf{E} + \sigma$$

$$\mathbf{S}M_P(A) = M_P(A) + M_P(B)$$

$$\mathbf{S}M_P(B) = M_P(B) + M_P(A)$$

$$\mathbf{S}x_A = x_A + x_B$$

$$\mathbf{S}x_B = x_B + x_A$$

$$\mathbf{S}M_I(A, B) = M_I(A, B) + M_I(B, A)$$

$$\mathbf{S}M_I(B, A) = M_I(B, A) + M_I(A, B)$$

$$\mathbf{S}x_A M_P(A) = x_A M_P(A) + x_B M_P(B)$$

$$M(A, B, x) = \overline{M}(A, B, x) + \Delta M(A, B, x),$$

$$\overline{M}(A, B, x) \equiv (1 - x)M_P(A) + xM_P(B),$$

$$\Delta M(A, B, x) \equiv x(1 - x)\Omega(A, B).$$

**If we use these symmetrical descriptors,
we can get a more reasonable model.**