

LASSO最適化問題の数値解法： ADMM法

岡山大学 異分野基礎科学研究所

大槻純也

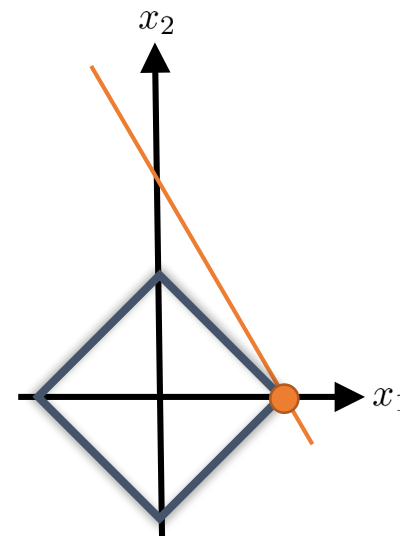


解きたい問題

L1ノルム最小化、Basis Pursuit (BP)

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subj. to} \quad \mathbf{y} = A\mathbf{x}$$

誤差がない場合

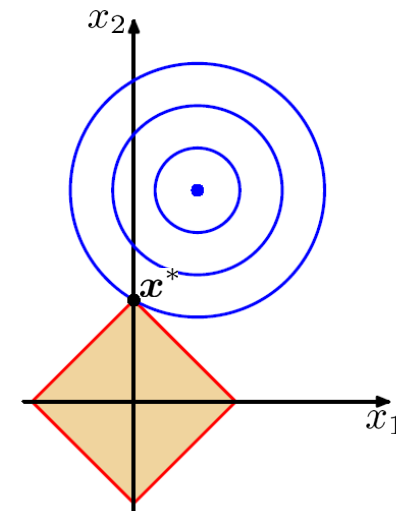


LASSO (Least Absolute Shrinkage of Selection Operators)

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}$$

誤差がある場合

$\mathbf{x}$ がスパースであることを仮定



一般化LASSO

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|B\mathbf{x}\|_1 \right\}$$

$B\mathbf{x}$ がスパースであることを仮定
(B は非正方行列)

L1ノルムを扱う難しさ

微分が不連続

$$g(x) = |x|$$

勾配を使用したアルゴリズムが不安定になる

滑らかな関数で近似する

$$g(x) \approx g_s(x)$$

例えば

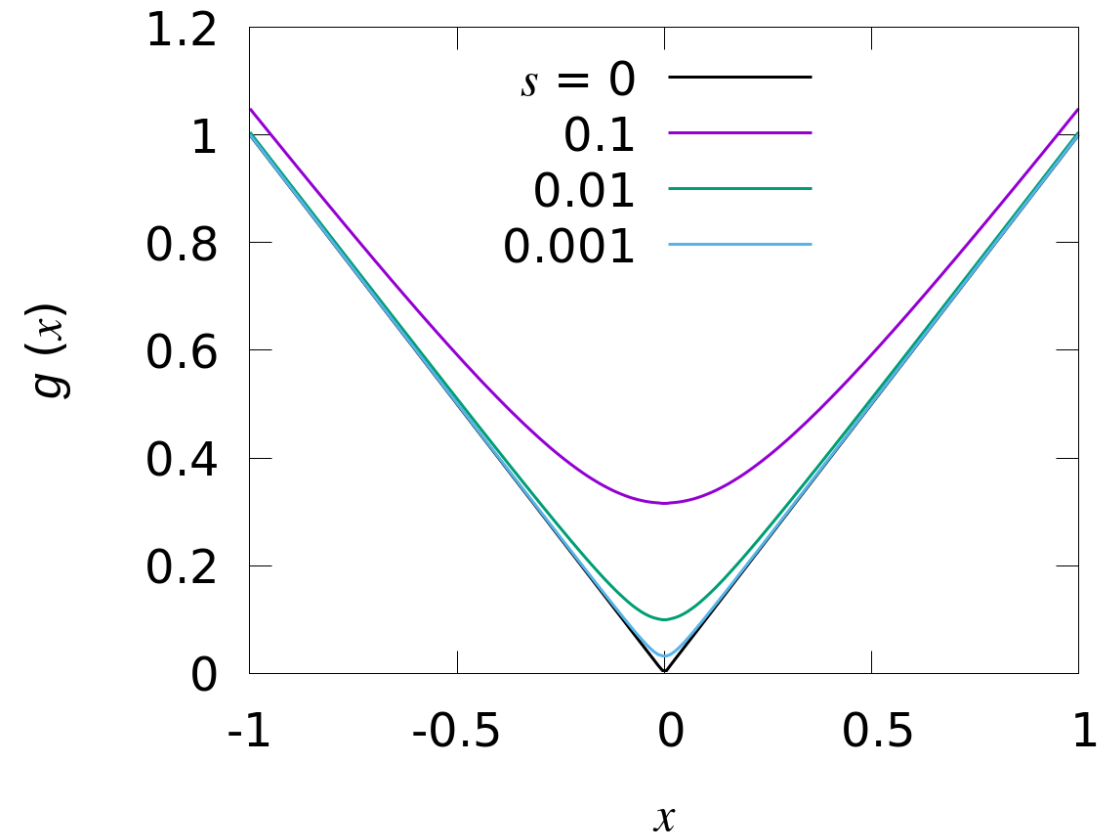
$$g_s(x) = \sqrt{x^2 + s}$$

$$g_s(x) = |x| - s \ln(1 + |x|/s)$$

M. Elad 「スパースモデリング」 共立出版

s が有限であれば、最急降下法など普通の方法で最小化問題を解ける。滑らかさパラメータ s をだんだん小さくしていき、繰り返し解く。

s の減少速度の調整が難しい



overview : LASSO最適化問題の解法

- 凸最適化問題だが、微分が不連続なため、一般的な方法は使えない
- LASSO
 - Coordinate descent (scikit-learnで使われている)
 - Majorization minimization (MM) algorithm; ISTA, FISTA
- generalized LASSO
 - Scikit-learnにはない
 - 行列Bの形式によりアルゴリズム・ライブラリがわかれている
 - 汎用的なものとしてADMM (Alternating Direction Method of Multiplier)がある
Boyd et al. Foundations and Trends in Machine Learning 3, 1 (2011)
拡張ラグランジュ法を利用した方法で非常に汎用的。拘束条件（総和則、非負性など）も課すことができる。ただし、収束は早くはない。

まずはN=M=1の場合

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad \text{LASSO}$$



$$F(x) = \frac{1}{2} (y - x)^2 + \lambda |x|$$

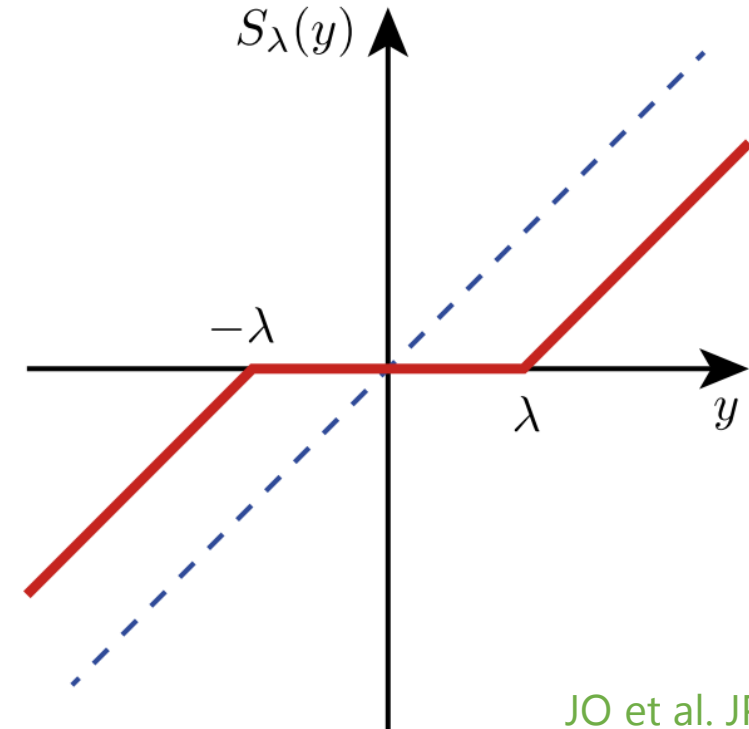
$$x^* = \arg \min_x F(x) \quad \text{1変数の最小化問題}$$

これは場合分けをすれば簡単に解ける

$$x^* = S_\lambda(y)$$

$$S_\lambda(y) \equiv \begin{cases} y - \lambda & (y > \lambda) \\ 0 & (-\lambda < y < \lambda) \\ y + \lambda & (y < -\lambda) \end{cases}$$

Soft threshold function (軟判定閾値関数)



JO et al. JPSJ 2020

インプット y の絶対値が λ だけ小さくなる
 → λ より小さいインプットはゼロになる

多次元になってもゼロが出てくるのはこの関数から

LASSOに戻る

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} F(\boldsymbol{x})$$

$$F(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$

ベクトル $\boldsymbol{x}$ の成分が混ざるので簡単には解けない
(2^N 通りの場合分けをしなければならない)

そこで次のように書き換える

$$F(\boldsymbol{x}, \boldsymbol{z}) = \frac{1}{2} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{z}\|_1$$

$$\min_{\boldsymbol{x}, \boldsymbol{z}} F(\boldsymbol{x}, \boldsymbol{z}) \quad \text{subj.to} \quad \boldsymbol{x} = \boldsymbol{z}$$

拘束条件付き最小化問題

第1項と第2項が分かれたので、絶対値を場合分けして評価できる。ただし、拘束条件を厳密に課してしまうと元の問題と変わらないので、**拘束条件をゆるく課す**。

以下では、一般的に、元の最小化問題を

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) \quad F(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$$

代わりに最小化問題を

$$\min_{\boldsymbol{x}, \boldsymbol{z}} F(\boldsymbol{x}, \boldsymbol{z}) \quad \text{subj.to} \quad \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}) = \mathbf{0}$$

$$F(\boldsymbol{x}, \boldsymbol{z}) = f(\boldsymbol{x}) + g(\boldsymbol{z})$$

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{x} - \boldsymbol{z}$$

と表す。一般化LASSOの場合は

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}) = B\boldsymbol{x} - \boldsymbol{z}$$

とすればよい。

拘束条件を課す方法

ラグランジュの未定乗数法 (Lagrange multiplier)

$$F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\nu}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \underbrace{\boldsymbol{\nu}^T \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z})}_{\text{multiplier}}$$

拘束条件といえばラグランジュの未定乗数法。
拘束条件が厳密に満たされる

罰金法 (penalty method)

$$F(\boldsymbol{x}, \boldsymbol{z}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \underbrace{\frac{\mu}{2} \|\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z})\|_2^2}_{\text{penalty term}}$$

拘束条件をゆるく課す。 $\mu \rightarrow \infty$ で厳密。
 μ を小さい値からだんだん大きくしていき、
その都度最適化を行う。収束が遅いのがデ
メリット。

拡張 (Augmented) ラグランジュ法

$$F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\nu}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \boldsymbol{\nu}^T \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}) + \frac{\mu}{2} \|\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z})\|_2^2$$

罰金法のゆるさと未定乗数法の厳密さを両立

アルゴリズム

$$\boldsymbol{x} = 0, \quad \boldsymbol{z} = 0, \quad \boldsymbol{\nu} = 0$$

μ は固定

for (収束するまで)

$$\boldsymbol{x}_{\text{new}} = \arg \min_{\boldsymbol{x}} F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\nu})$$

$$\boldsymbol{z}_{\text{new}} = \arg \min_{\boldsymbol{z}} F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\nu})$$

$$\boldsymbol{\nu}_{\text{new}} = \boldsymbol{\nu} + \mu \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z})$$

収束判定

primary

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{z}) = 0 \quad \text{拘束条件が満たされてるか}$$

secondary

$$\boldsymbol{x}_{\text{new}} = \boldsymbol{x}, \quad \boldsymbol{z}_{\text{new}} = \boldsymbol{z} \quad \text{変数が収束してるか}$$

両方収束させる。 μ が極端に大きい／小さいと一方しかゼロに向かわない

ADMM for LASSO

LASSO

更新則は

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2$$

$$g(\mathbf{z}) = \lambda \|\mathbf{z}\|_1$$

$$h(\mathbf{x}, \mathbf{z}) = \mathbf{x} - \mathbf{z}$$

$$\mathbf{x}_{\text{new}} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2\lambda} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \frac{\mu}{2} \left\| \mathbf{x} - \mathbf{z} + \frac{\boldsymbol{\nu}}{\mu} \right\|_2^2 \right\}$$

2次形式なので簡単に解ける

$$\mathbf{z}_{\text{new}} = \arg \min_{\mathbf{z}} \left\{ \|\mathbf{z}\|_1 + \frac{\mu}{2} \left\| \mathbf{x} - \mathbf{z} + \frac{\boldsymbol{\nu}}{\mu} \right\|_2^2 \right\}$$

成分ごとに評価できる

$$\boldsymbol{\nu}_{\text{new}} = \boldsymbol{\nu} + \mu(\mathbf{x} - \mathbf{z})$$

argminを評価すると ($\mathbf{u} = \boldsymbol{\nu}/\mu$ とおく)

$$\mathbf{x}_{\text{new}} = \left(\mu \mathbf{I} + \frac{1}{\lambda} A^T A \right)^{-1} \left(\frac{1}{\lambda} A^T \mathbf{y} + \mu(\mathbf{z} - \mathbf{u}) \right)$$

$$\mathbf{z}_{\text{new}} = S_{1/\mu}(\mathbf{x} + \mathbf{u})$$

element-wise
soft threshold function

$$\mathbf{u}_{\text{new}} = \mathbf{u} + \mathbf{x} - \mathbf{z}$$

- Soft threshold functionにより $\mathbf{z}$ の一部の要素がゼロになる。これによりスパース解が得られる。
- 逆行列はループの外で計算しておく。実際には、Cholesky分解（またはLU分解）を保持する。
- ADMM法は一般化LASSOにも容易に適用できる。
詳細は [JO et al. JPSJ 2020](#)

ADMM for BP

BP

$$f(\mathbf{x}) = \boldsymbol{\rho}^T(\mathbf{y} - A\mathbf{x}) \quad \text{\textcolor{brown}{$y = Ax$ を厳密に満たすように未定乗数 $\boldsymbol{\rho}$ を導入}}$$

$$\left. \begin{aligned} g(\mathbf{z}) &= \|\mathbf{z}\|_1 \\ h(\mathbf{x}, \mathbf{z}) &= \mathbf{x} - \mathbf{z} \end{aligned} \right\} \quad \text{\textcolor{brown}{λ がないこと以外はLASSOと同じ}}$$

$\mathbf{x}$ についてだけ計算する

$$\begin{aligned} \mathbf{x}_{\text{new}} &= \arg \min_{\mathbf{x}} \left\{ \boldsymbol{\rho}^T(\mathbf{y} - A\mathbf{x}) + \frac{\mu}{2} \|\mathbf{x} - \mathbf{z} + \mathbf{u}\|_2^2 \right\} \\ &= \frac{1}{\mu} A^T \boldsymbol{\rho} + \mathbf{z} - \mathbf{u} \end{aligned}$$

$\mathbf{y} = A\mathbf{x}$ を満たすように未定乗数 $\boldsymbol{\rho}$ を決めると

$$\boldsymbol{\rho} = \mu(AA^T)^{-1}[\mathbf{y} + A(\mathbf{z} - \mathbf{u})]$$

まとめると

$$\mathbf{x}_{\text{new}} = A^T(AA^T)^{-1}\mathbf{y} + [\mathbf{I} - A^T(AA^T)^{-1}A](\mathbf{z} - \mathbf{u})$$

$$\mathbf{z}_{\text{new}} = S_{1/\mu}(\mathbf{x} + \mathbf{u}) \quad \begin{array}{l} \text{element-wise} \\ \text{soft threshold function} \end{array}$$

$$\mathbf{u}_{\text{new}} = \mathbf{u} + \mathbf{x} - \mathbf{z}$$

$\mathbf{y} = A\mathbf{x}$ は常に満たされている