第 55 回自動制御連合講演会
2012 年 11 月 17 日，18 日　京都大学

2B204

スパースモデリングとベイズ統計

○池田思朗 （統計数理研究所）

Sparse Modelling and Bayesian Statisitcs

∗S. Ikeda (Inst. of Statisical Mathematics)

Abstract− Sparse modelling is a new style of information processing which utilizes the sparsity of the information source. The framework is well formulated with the Bayesian statistics. We explain the Bayesian framework with some examples.

Key Words: Sparse modelling, Bayesian statistics

1　はじめに

スパースモデリングとは，情報源の持つ疎性を利用したデータ処理の方法のことである．本稿では，スパースモデリングの方法として代表格である圧縮センシング [1] と LASSO[2] をとりあげ，それぞれの定義と解法について説明する．その後，LASSO とベイズ統計との関係を説明し，スパースモデリングがこれら 2 つの方法だけでなく，より広い概念であることを説明する．

2　圧縮センシングと LASSO

本節ではスパースモデリングの代表的な方法である圧縮センシングと LASSO について説明する．圧縮センシングは情報理論の分野で，LASSO は統計の分野で提案された．共にスパース性を用いる手法であり深く関係しているが，用いるパラメータの記号が伝統的に異なる．以下の説明では比較し易いように LASSO の表記に揃えて説明する．

以下では観測ベクトル $\boldsymbol{y} \in \Re^M$ から情報源のベクトル $\boldsymbol{\beta} \in \Re^N$ を推定する問題を考える．ただし，$M < N$ である．$\boldsymbol{\beta}$ と $\boldsymbol{y}$ の間にはある既知の行列 $X \in \Re^{M\times N}$ を用いて線形の関係があると考える．$\boldsymbol{y}$ は直接観測でき $\boldsymbol{y} = X\boldsymbol{\beta}$ という関係を仮定する．$M < N$ であるから，$\boldsymbol{y} = X\boldsymbol{\beta}$ を満たす $\boldsymbol{\beta}$ は無数に存在し，情報源ベクトルを一意に決めることはできない．そこで，圧縮センシングでは $\boldsymbol{\beta}$ がスパースである，すなわち多くの成分が 0 である，という仮定を置く．

LASSO は回帰分析の方法として提案されたものである．回帰分析では，目的変数 y と説明変数 $x_1, \cdots, x_N$[1] の間に次の関係を仮定する[2]

$$y = \sum_{i=1}^{N} \beta_i x_i + z.$$

ここで $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_N)$ は各説明変数に対する重みであり，Z はノイズであり，正規分布に従うと仮定する．今，y と $\boldsymbol{x} = (x_1, \cdots, x_N)^T$ (本稿では転置を ^{T} で表す) のサンプルが　M　個得られたとする．それらを $(y_1, \boldsymbol{x}_1), \cdots, (y_1, \boldsymbol{x}_M)$ と書くと，

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{z}.$$

となる．ただし，X は $M \times N$ 次の実行列であり，各行は説明変数のサンプル $\boldsymbol{x}_i^T$ である．LASSO ではパラメータ $\boldsymbol{\beta}$ がスパースだと仮定し解析する．

次節以降でそれぞれの解法について簡単に説明するが，準備として本稿で用いるベクトルのノルムを定義しておく．本稿ではベクトル $\boldsymbol{a} = (a_1, \cdots, a_N)^T$ のノルムとして，次のように定義される 0 –ノルム $\|\boldsymbol{a}\|_0$, 1 –ノルム $\|\boldsymbol{a}\|_1$, 2 –ノルム $\|\boldsymbol{a}\|_2$ を用いる．

$$\|\boldsymbol{a}\|_0 = \{0 \text{ でない成分 } a_i \text{ の数 }\}$$
$$\|\boldsymbol{a}\|_1 = \sum_{i=1}^{N} |a_i|, \quad \|\boldsymbol{a}\|_2 = \Big(\sum_{i=1}^{N} |a_i|^2\Big)^{1/2}.$$

2.1　圧縮センシング

圧縮センシングでは，$\boldsymbol{y} = X\boldsymbol{\beta}$ という線形の関係の下で $\boldsymbol{y}$ から $\boldsymbol{\beta}$ を推定するが，$\boldsymbol{y}$ の次元 M よりも $\boldsymbol{\beta}$ の次元の N の方が大きいため，$\boldsymbol{y} = X\boldsymbol{\beta}$ を満たす $\boldsymbol{\beta}$ が無数にあり，解は不定となる．圧縮センシングでは，そうした数多くある解の中から，$\boldsymbol{\beta}$ がスパースである，すなわち多くの成分が 0 であるものを求める問題を考える．

このアイディアを簡単に表すと，以下の最適化問題となる．

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0 \quad \text{subject to} \quad \boldsymbol{y} = X\boldsymbol{\beta}. \qquad (1)$$

この最適化問題の目的関数は 0 –ノルムで表現されている．目的関数は離散値を取り $\boldsymbol{\beta}$ で微分できないため，最適値を求めるには $\boldsymbol{\beta}$ の成分の一部を取り出し $\boldsymbol{y} = X\boldsymbol{\beta}$ の条件と目的関数 $\|\boldsymbol{\beta}\|_0$ とが満たされるかどうかを調べる，という作業を全ての成分の組み合わせに対して調べなければならない．これは組み合せ最適化となるため，$\boldsymbol{\beta}$ と $\boldsymbol{y}$ の次元が大きいときには効率的に解くことが難しい．

圧縮センシングが大きな注目を集めた理由は，(1) 式で困難さの原因となる 0 –ノルムを 1 –ノルムに変えた以下の問題を考えたことにある．

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = X\boldsymbol{\beta}. \qquad (2)$$

この問題は凸最適化問題であり，線形計画法を用いて効率的に解くことができる．(1) 式と (2) 式とは目的関数が異なるが，そのために問題が解き易くなっている．

[1] 統計学では伝統的に確率変数を大文字で表すが，ここでは区別しやすくするため，小文字で表現する．
[2] ここでは y, x_i の期待値は 0 に正規化されているとする．

このように目的関数を緩和して解くと，最適解も異なるように見える．しかし，(2) 式のように緩和しても解が一致するための条件が示され [3, 4]，多くの応用の場面でそうした条件が満たされると期待できることから，広く研究者が興味を持つようになった．

2.1.1 圧縮センシングと線形計画法

ここでは (2) 式を線形計画法として解くことができることを示す．線形計画法の基本形式は以下の通りである．

$$\max_{\boldsymbol{u}} \ \boldsymbol{c}^T\boldsymbol{u} \quad \text{subject to} \quad A\boldsymbol{u} \leq \boldsymbol{b} \text{ and } \boldsymbol{u} \geq \boldsymbol{0}.$$

ベクトルに対する不等号は成分毎に成り立つとする．(2) 式に関連する以下の問題を考える．

$$\begin{aligned} &\min_{\boldsymbol{\beta}^+,\boldsymbol{\beta}^-} \mathbf{1}^T(\boldsymbol{\beta}^+ + \boldsymbol{\beta}^-), \\ &\qquad \text{subject to} \\ &\boldsymbol{y} = X(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^-), \ \ \boldsymbol{\beta}^+ \geq \boldsymbol{0}, \ \ \boldsymbol{\beta}^- \geq \boldsymbol{0}. \end{aligned} \tag{3}$$

ここで $\mathbf{1}$ は 1 が N 個ならんだベクトルである．上式の $\boldsymbol{\beta}^+$ と $\boldsymbol{\beta}^-$ は (2) 式の $\boldsymbol{\beta}$ の各成分が正の場合と負の場合を分けたものになっており，(3) 式の最適値を与える $\boldsymbol{\beta}^+, \boldsymbol{\beta}^-$ を用いると (2) 式の解は $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^+ - \boldsymbol{\beta}^-$ とかける．倍の数の変数を考えることで，(2) 式を (3) 式と書き直すことができた．

(3) 式はさらに線形計画法の基本形式に書き換えることができる．$\boldsymbol{c} = -(\mathbf{1}^T, \mathbf{1}^T)^T$, $\boldsymbol{u} = (\boldsymbol{\beta}^{+T}, \boldsymbol{\beta}^{-T})^T$ と定義する．これらを用いて (3) 式は以下のように書き換えられる．

$$\max_{\boldsymbol{u}} \ \boldsymbol{c}^T\boldsymbol{u}, \ \text{ subject to } \ (X, -X)\boldsymbol{u} \leq \boldsymbol{y}, \ \ \boldsymbol{u} \geq \boldsymbol{0}. \tag{4}$$

これは $A = (X, -X)$, $\boldsymbol{b} = \boldsymbol{y}$ とおけば線型計画法の基本形式と等しい．線形計画法は最適化の基本となる問題であり，広く研究されている．有償，無償のパッケージも充実しており，大規模なものであっても比較的効率良く解ける．

2.2 LASSO

統計学の分野では Tibshirani によって LASSO と呼ばれるスパースモデリングの方法が提案され，様々な分野で応用されている [2]．LASSO は以下の問題として定義された．

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_1 \leq t. \tag{5}$$

定義からわかるように LASSO では $\boldsymbol{\beta}$ の 1 –ノルムを制約し，$\boldsymbol{y}$ と $X\boldsymbol{\beta}$ との間の 2 乗距離を最小にするものを求める．(5) 式の問題は (2) 式の問題と異なるように見えるが，Lagrange 未定定数を用いた Lagrange 関数を考えると，2 つの最適化問題の関係が見える．

Lagrange 関数は未定定数 λ を用いて以下の式で表される．

$$L(\boldsymbol{\beta}, \lambda) = \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1. \tag{6}$$

λ をある正の値に定めて $L(\boldsymbol{\beta}, \lambda)$ を $\boldsymbol{\beta}$ について最小化し，その結果得られる $\boldsymbol{\beta}$ を $\boldsymbol{\beta}(\lambda)$ と書く．

$$\boldsymbol{\beta}(\lambda) = \arg\min_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \lambda). \tag{7}$$

このとき，λ は (5) 式の t と対応していることがしられている．すなわち，ある t に対して (5) 式の最適解を求まったとすると，(7) 式の最適解がそれと一致するような λ が存在する．

もし，λ が大きい値をとると，(6) の第 2 項が重要となり，$\hat{\boldsymbol{\beta}}(\lambda)$ は多くの成分が 0 となる．一方，λ が小さいと第 1 項が重要となる．このときには，より多くの $\hat{\boldsymbol{\beta}}(\lambda)$ の成分が 0 でなくなり $X\boldsymbol{\beta}$ が $\boldsymbol{y}$ と近くなり，その上で $\|\boldsymbol{\beta}\|_1$ を小さくする．したがって，圧縮センシングの解に近づく．LASSO では，このように λ の値を変化させることによって解の “スパースさ” を調整できる．

統計学では回帰モデルの説明変数が多くあるとき，その中からどの説明変数を選ぶか，というモデル選択の問題が古くから議論されてきた．LASSO は λ を設定することによって回帰モデルのパラメータを自動的に選択することができ，モデル選択の方法として用いることができる．

2.2.1 LASSO と 2 次計画法

(7) 式の問題は 2 次計画法として解くことができる．2 次計画法の基本形は以下の通りである．

$$\begin{aligned} &\min_{\boldsymbol{u}} \frac{1}{2}\boldsymbol{u}^T H\boldsymbol{u} + \boldsymbol{f}^T\boldsymbol{u} \\ &\text{subject to} \\ &A\boldsymbol{u} \leq \boldsymbol{b}. \end{aligned} \tag{8}$$

圧縮センシングの場合と同様に，$\boldsymbol{\beta}^+, \boldsymbol{\beta}^-$ を用いて (7) 式を書き直すと，

$$\begin{aligned} &\min_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \lambda) \\ =&\boldsymbol{y}^T\boldsymbol{y} + \min_{\boldsymbol{\beta}}(\boldsymbol{\beta}^T X^T X\boldsymbol{\beta} - 2\boldsymbol{y}^T X\boldsymbol{\beta} + \lambda\|\boldsymbol{\beta}\|_1) \\ =&\boldsymbol{y}^T\boldsymbol{y} + \min_{\boldsymbol{\beta}}\Big((\boldsymbol{\beta}^+ - \boldsymbol{\beta}^-)^T X^T X(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^-) \\ &\qquad - 2\boldsymbol{y}^T X(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^-) + \lambda\mathbf{1}^T(\boldsymbol{\beta}^+ + \boldsymbol{\beta}^-)\Big) \\ &\text{subject to } \ \boldsymbol{\beta}^+ \geq \boldsymbol{0}, \ \boldsymbol{\beta}^- \geq \boldsymbol{0}. \end{aligned}$$

したがって，$\boldsymbol{u}$, H, $\boldsymbol{f}$, A, $\boldsymbol{b}$ を以下のように定めれば，2 次計画法の基本形と一致する．

$$\begin{aligned} &\boldsymbol{u} = \begin{pmatrix} \boldsymbol{\beta}^+ \\ \boldsymbol{\beta}^- \end{pmatrix}, \ H = \begin{pmatrix} X^T X & -X^T X \\ -X^T X & X^T X \end{pmatrix} \\ &\boldsymbol{f} = \begin{pmatrix} -2X^T\boldsymbol{y} + \lambda\mathbf{1} \\ 2X^T\boldsymbol{y} + \lambda\mathbf{1} \end{pmatrix}, \ A = -E_{2N}, \ \boldsymbol{b} = \boldsymbol{0}. \end{aligned}$$

ここで E_{2N} は $2N$ 次の単位行列，$\mathbf{0}$ は N 次の零ベクトルとする．

2 次計画法も最適化の分野で広く研究されており，パッケージも充実している．ただし，線型計画法に比べれば問題が難しいため，計算機で解ける問題の規模，あるいは解を求める速さは線型計画法に比べると小さく，遅くなる．

2.2.2 Solution Path

λ を変化させたときに $\hat{\boldsymbol{\beta}}(\lambda)$ がどのように変化するのかということは，どの成分が λ に寄らずに $\boldsymbol{y}$ に寄与しているかを示しており，重要な意味を持つ．λ を

変化させたときの $\hat{\boldsymbol{\beta}}(\lambda)$ の軌跡を solution path と呼ぶ [5, 6]．

ここでは具体的には示さないが，LASSO の場合，solution path を求めるアリゴリズムは比較的簡単である．その理由は，λ のある区間で $\boldsymbol{\beta}$ のうち零でない成分の組み合せ，及びそれぞれ成分の正負が変化しないならば，その区間で $\hat{\boldsymbol{\beta}}(\lambda)$ が λ に関する線形の関数になるからだ．したがって，$\hat{\boldsymbol{\beta}}(\lambda)$ の各成分の符号が変化せず，0 であった成分が変化しない λ の区間を求め，それぞれの区間で λ の線形関数となる $\hat{\boldsymbol{\beta}}(\lambda)$ の関数を求めればよい．Solution path を求める際のポイントは，λ を大きな値から始めることである．大きな値の λ に対しては $\hat{\boldsymbol{\beta}}(\lambda)$ のほとんどの成分が 0 となるため，計算が簡単になる．その $\hat{\boldsymbol{x}}(\lambda)$ を初期値として，λ を減らしていきながら solution path を求めていく．

ただし，こうした解析は説明変数の次元が非常に大きい場合には実現が難しい．

3 ベイズ統計とスパースモデリング

3.1 ベイズ統計と LASSO

LASSO に関する (7) 式の問題は，ベイズ統計の枠組で定式化することができる．2 節の最初に示したように，$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{z}$ と観測にノイズが加わり，ノイズ $\boldsymbol{z}$ の各成分が平均 0 分散 σ^2 の正規分布にしたがうとする．$\boldsymbol{\beta}$ が与えられたときの $\boldsymbol{y}$ の分布は尤度関数と呼ばれるが，この場合には次の形になる．

$$p(\boldsymbol{y}|\boldsymbol{\beta}, X) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\Big(-\frac{\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2}{2\sigma^2}\Big).$$

一方，$\boldsymbol{\beta}$ の事前分布を各成分が独立なラプラス分布にとる．ラプラス分布は正の定数 ϕ をパラメータとして以下のように定義される．

$$p(\boldsymbol{\beta}) = \frac{1}{2\phi} \exp\Big(-\frac{\|\boldsymbol{\beta}\|_1}{\phi}\Big).$$

X, $\boldsymbol{y}$ が得られたもとでの $\boldsymbol{\beta}$ の事後分布を考える．

$$\begin{aligned} p(\boldsymbol{\beta}|\boldsymbol{y}, X) &\propto p(\boldsymbol{y}|\boldsymbol{\beta}, X)p(\boldsymbol{\beta}) \\ &= \frac{1}{2\phi(2\pi\sigma^2)^{N/2}} \exp\Big(-\frac{\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2}{2\sigma^2} - \frac{\|\boldsymbol{\beta}\|_1}{\phi}\Big). \end{aligned}$$

対数をとると，以下のようになる

$$\log p(\boldsymbol{\beta}|\boldsymbol{y}, X) = \mathrm{Const} - \frac{\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2}{2\sigma^2} - \frac{\|\boldsymbol{\beta}\|_1}{\phi}$$

事後確率を最大とする $\boldsymbol{\beta}$ を推定値とする方法は MAP 推定と呼ばれる．具体的には以下に示す形となる．

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \arg\max_{\boldsymbol{\beta}} \ \log p(\boldsymbol{\beta}|\boldsymbol{y}, X) \\ &= \arg\min_{\boldsymbol{\beta}} \big(\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\big) \end{aligned}$$

ただし，$\lambda = 2\sigma^2/\phi$ である．これは (7) 式の問題と等しい．このように，LASSO はベイズ統計と関係していることがわかる．

3.2 尤度関数

前節では正規分布の観測ノイズがある場合，MAP 解と LASSO とが結びつくことを示した．しかし，確率モデルには様々なものがあり，正規分布ではない尤度関数を用いる場合も多い．そのような確率モデルに対してもスパースモデリングを使うことができる [7]．例えば y が 0 または 1 を取る logistic 回帰モデルを考えると，その尤度関数はパラメータ $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_N)^T$ を用いて次のようにかける．

$$p(y = 1|\boldsymbol{\beta}, \boldsymbol{x}) = \frac{1}{1 + \exp(-\boldsymbol{x}^T\boldsymbol{\beta})}.$$

観測された M 個のサンプル $(y_1, \boldsymbol{x}_1), \cdots, (y_M, \boldsymbol{x}_M)$ から $\boldsymbol{\beta}$ を推定する際の尤度は $p(\boldsymbol{y}|\boldsymbol{\beta}, X) = \prod_{i=1}^M p(y_i|\boldsymbol{\beta}, \boldsymbol{x}_i)$ となるが，これは LASSO の場合と違って 2 乗誤差とはならない．この場合も，$\boldsymbol{\beta}$ に対してラプラス分布を仮定し，MAP 推定を行なうことができる．そうして得られた $\boldsymbol{\beta}$ は一般にスパースとなる．このモデルに対するスパースな解の効率的な計算法も詳しく調べられている [8]．また，こうしたモデルに対して solution path を求める方法も研究されている [9]．

また，ニューラルネットワークモデルのパラメータの学習に際して石川が提案した忘却項は，スパースモデリングと同じものである [10]．この場合も学習後に得られるパラメータは 0 を多く含み，スパースとなる．

このように，様々な確率モデルに対し，パラメータにの事前分布をラプラス分布として与えて MAP 推定を行なう，という手法を考えることができる．尤度がパラメータに関して連続で滑らかであれば，このような方法でスパースな推定値が得られると期待できる．

我々も光学の分野で昔から知られている位相回復の問題に対してスパースモデリングを用いた手法を提案している [11]．

4 まとめ

本稿ではスパースモデリングで広く用いられている 2 つの方法，圧縮センシングと LASSO について説明した．こうした方法はスパースモデリングの標準的な問題であるが，スパース性を用いる解析手法はこれらの 2 つの手法に限られるものではない．他の尤度関数を考えることによって，様々な確率モデルに対してスパースモデリングを用いることができる．

また，スパース性についても様々な拡張が考えられている．本稿で説明した，ラプラス分布を事前分布に用い，個々のパラメータが独立に 0 となるように調整する方法のみがスパースモデリングなのではない．例えば，多次元のパラメータをグループに分け，それぞれのグループがそろって 0 となること [12] をスパース性と考える Group LASSO と呼ばれる方法や，隣合うパラメータの差の絶対値の和 (total variation) が小さくなる [6] ようにスパース性を考えるといった方法が提案されている．こうしたスパース性を実現するため，計算の都合上 1 –ノルムを用いることが多いが，他の関数を用いる場合もある [13]．

このように，尤度，スパース性を拡張することによって，スパースモデリングは様々な問題に適用できる考

え方である．こうした新しい考え方を有効に利用していくことで，新たな情報処理が可能となる．

本稿では圧縮センシングや LASSO がそれぞれが線形計画法，2 次計画法といった最適化の基本的な方法によって解けることを示した．こうした基本的な方法に関しては汎用のパッケージが存在しているが，尤度関数，あるいはスパース性を拡張した場合には汎用のパッケージは使えない．欧米など，他の国では各々のスパースモデリングの問題に特化した最適化プログラムを開発している場合が多く見られる．最適化法の専門家によるこうしたアルゴリズムの開発は効果的である．

以上のように，スパースモデリングに関する研究は様々な方向に発展している．日本国内においても，スパースモデリングの基礎研究，そして応用研究が今後発展していくことを期待している．

参考文献

1) D. Donoho: “Compressed sensing”, IEEE transaction on Information Theory, **52**, 4, pp. 1289–1306 (2006).
2) R. Tibshirani: “Regression shrinkage and selection via the lasso”, J. R. Stat. Soc., Ser. B, **58**, 1, pp. 267–288 (1996).
3) E. J. Candès and T. Tao: “Near-optimal signal recovery from random projections: Universal encoding strategies?”, IEEE transaction on Information Theory, **52**, 12, pp. 5406–5425 (2006).
4) D. L. Donoho and J. Tanner: “Neighboriness of randomly-projected simplices in high dimensions”, Proceedings of the National Academy of Sciences, **102**, 27, pp. 9452–9457 (2005).
5) B. Efron, T. Hastie, I. Johnstone and R. Tibshirani: “Least angle regression”, Annals of Statistics, **32**, 2, pp. 407–499 (2004).
6) S. Rosset and J. Zhu: “Piecewise linear regularized solution paths”, Annals of Statistics, **35**, 3, pp. 1012–1030 (2007).
7) J. Lokhorst: “The lasso and generalized linear models”, Honors project, University of Adelaide, Adelaide (1999).
8) B. Krishnapuram, L. Carin, M. A. Figueiredo and A. J. Hartemink: “Sparse multinomial logistic regression: fast algorithms and generalization bounds”, IEEE transaction on Pattn Anal. Mach. Intell., **27**, 6, pp. 957–968 (2005).
9) S. Rosset: “Following curved regularized optimization solution paths”, Advances in Neural Information Processing Systems 17, MIT Press, Cambridge, MA, pp. 1153–1160 (2005).
10) M. Ishikawa: “Structural learning with forgetting”, Neural Networks, **9**, 3, pp. 509–521 (1996).
11) S. Ikeda and H. Kono: “Phase retrieval from single biomolecule diffraction pattern”, Optics Express, **20**, 4, pp. 3375–3387 (2012).
12) M. Yuan and Y. Lin: “Model selection and estimation in regression with grouped variables”, J. R. Statist. Soc. B, **68**, pp. 49–67 (2006).
13) Z. Xu, X. Chang, F. Xu and H. Zhang: “$l_{1/2}$ regularization: A thresholding representation theory and a fast solver”, IEEE Trans. on Neural Networks and Learning Systems, **23**, 7, pp. 1013–1027 (2012).