2　1次ノルム最小化—研究の経緯

データ $\boldsymbol{y}$ が係数 $\boldsymbol{x}$ の線形和 $\boldsymbol{y} = A\boldsymbol{x}$ で得られるとき、少ない情報 $\boldsymbol{y}$ から、より多い情報 $\boldsymbol{x}$ を推定する問題を考えます。$\boldsymbol{x}$ がスパースであると仮定できるとき、単純に考えると $\boldsymbol{y} = A\boldsymbol{x}$ の条件を満たす $\boldsymbol{x}$ の中から、非ゼロ要素の「個数」が最小になる $\boldsymbol{x}$ を探すのが良さそうに思えます。この「非ゼロ要素の個数」は「0 次ノルム」（$\|\boldsymbol{x}\|_0$）と呼ばれます。0 次ノルム最小化の問題を解く単純な方法は $\boldsymbol{x}$ の全ての組み合わせを調べることです。つまり、$\boldsymbol{x}$ から 1 つの要素を取り出したときの組み合わせの数は $\boldsymbol{x}$ の次元を M として、${}_M\mathrm{C}_1$、2 つの組み合わせは ${}_M\mathrm{C}_2$、のように、全組み合わせ ${}_M\mathrm{C}_1 + {}_M\mathrm{C}_2 + \cdots + {}_M\mathrm{C}_M$ について、$\boldsymbol{y} = A\boldsymbol{x}$ をみたす最もスパースな解を探索します。しかし、このような全探索に必要な計算量は M の指数関数的に増大し、実際の高次元 (M の次元が数十以上) の問題に応用するのは現実的ではないことが知られています。

Tibshirani (1996)[1] は線形回帰の問題に 1 次ノルムの項を加える「LASSO」（Least Absolute Shrinkage and Selection Operator）という手法を提案しました。LASSO の問題は前節にでてきた (7) 式に等しいものです。LASSO では、A の列ベクトルの中で不要なものは係数 $\boldsymbol{x}$ をゼロにすることで過適合 (over-fitting) を防ぎ、モデルの予測誤差を小さくすることができます。このため、$\boldsymbol{x}$ の中でデータを説明するのに必要な変数だけを選択する手法（変数選択）としても使われます。

さらに、2000 年代半ばには Candes & Tao[2] や Donoho & Tanner[3] が、$N < M$ の問題で、ある条件下においては $\boldsymbol{y} = A\boldsymbol{x}$ を満たす 1 次ノルム最小解が 0 次ノルム最小解と等しくなることを証明しました。この証明は言い換えれば、元情報の疎性を利用することで、少ない情報からより多い情報が「完全」に再現できることの証明でもあり、多くの分野から注目を集めました。1 次ノルム最小化は従来からよく知られた最適化アルゴリズムである「線形計画法」で解くことができ、計算量の面では 0 次ノルム最小化より有利になります。情報を圧縮して計測（センシング）し、そこから元情報を再構成できるため、この文脈において当該手法・分野は「圧縮センシング」とも呼ばれることもあります。

ℓ_1 項を制約項として使うことによってスパースな解が得られることを直感的に理解するために図 1 の説明がよく使われます。図 1 は変数が 2 つ（$M = 2$）、うち 1 つの係数のみが非ゼロの値をもち（$K = 1$）、観測量 $\boldsymbol{y}$ が 1 つ（$N = 1$）の場合を模式的に示しています。直線は $\boldsymbol{y} = A\boldsymbol{x}$ を表しており、この直線上の点は全て方程式の解となります。図中の「菱形」は 1 次ノルム、すなわち絶対値の和が一定になる線を表しています。1 次ノルム最小化によって問題を解くことは、図中の直線と交わる条件の元で菱形を最小にすることに相当します。直線と菱形の接点では横軸の値がゼロで、縦軸しか非ゼロ値をもっていない、すなわちスパースな解になっています。このような例を高次元に拡張し、N 次元多胞体の M 次元空間への射影を考えることから、Donoho (2006)[3], Donoho & Tanner (2005)[4] はランダム行列 A の場合に 1 次ノルム最小化によって解を完全に再構成できるための条件を調べました。図 2 は $\boldsymbol{x}$ のスパース度 $\rho = K/M$ に対して、完全再構成のために必要なデータ数 $\alpha = N/M$ の閾値が曲線で表されています。ここでは $\boldsymbol{x}$ の値が非負の場合のしきい値を表示しています。図中の曲線よりも上側の観測点数があれば、$\boldsymbol{x}$ は完全に再構成されます。例えば、係数ベクトルの要素の内、2 割ほどしか非ゼロの値を持っていない（$\rho = 0.2$）場合なら、説明変数の数の半分程度のデータ（$\alpha = 0.5$）からでも完全に再構成できることがわかります。この曲線は媒介変数を用いた数式で書き下すことができます。

提案された時期や背景・分野は異なりますが、LASSO も圧縮センシングも 1 次ノルム最小化を使うことは共通しています。これらは $\boldsymbol{x}$ も $\boldsymbol{y}$ も共に高次元である問題において、データ $\boldsymbol{y}$ から $\boldsymbol{x}$ に関するスパース性を利用して、私たちが欲しい低次元の情報を高精度に抽出する

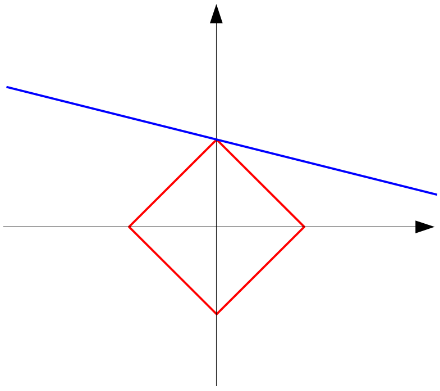

図 1: $N = 2$、$K = 1$、$M = 1$ の場合の模式図。直線が $\boldsymbol{y} = A\boldsymbol{x}$ の線で、菱形の線は 1 次ノルム（=絶対値の和）一定の線。この図では縦軸上の 1 点で交わっており、スパースな解（つまり、片方のみが値をもち、もう片方はゼロ）が得られている。

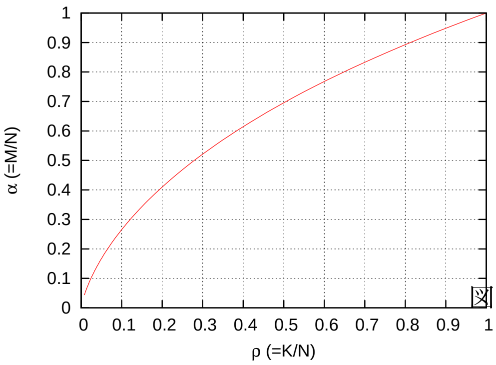

図 2: 非負の場合の弱しきい値。

手法と言うことができます。ただし、実際の問題が圧縮センシングの完全再構成の条件を全て満たしているとは限りません。1 次ノルム最小化で得られた解が 0 次ノルム最小の解と異なることもあり、問題ごとに適切に結果を検証する必要があります。0 次ノルム最小化の問題についても、全組み合わせを探索するのは計算困難ですが、効率的な探索法や追跡アルゴリズムは提案されています。本稿では 1 次ノルム最小化の応用のみを扱いますが、問題によっては 1 次ノルム最小化よりもこれらの手法がより優れた結果を与えることがあるでしょう。これらについてはより専門的な文献を参考にしてください。次章からは天文・宇宙物理分野への応用例について紹介します。

参考文献

[1] Tibshirani, R.: Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B*, 58, 267 (1996)

[2] Candes, E. J., Tao, T.: Near-Optimal Signal Recovery From Random Projections: Universal Encodin g Strategies?, *IEEE Transactions, Information Theory*, 52, 5406 (2006)

[3] Donoho, D. L.: Compressed sensing, *IEEE Transactions, Information Theory*, 52, 1289 (2006)

[4] Donoho, D. L., Tarner, J.: Neighborliness of randomly projected simplices in high dimensions, *Proc. Academy of Sciences of USA*, 102, 9452 (2005)