# 1　連立一次方程式を解く

連立一次方程式を考えます。

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1M}x_M \\ y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2M}x_M \\ \vdots \\ y_N = a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NM}x_M \end{cases} \tag{1}$$

変数の数は $M$ 個で、方程式の数は $N$ 個です。行列とベクトルを使うともうちょっと簡単に書けます。

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NM} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} \tag{2}$$

$$\boldsymbol{y} = A\boldsymbol{x} \tag{3}$$

ここでベクトル $\boldsymbol{y}$ は $N$ 次元、$\boldsymbol{x}$ は $M$ 次元、行列 $A$ は $N \times M$ 次元 ($N$ 行 $M$ 列) です。

いま、観測値が $\boldsymbol{y}$、$A$ は既知の行列、$\boldsymbol{x}$ が推定したいパラメータだとして、$\boldsymbol{y}$ から $\boldsymbol{x}$ を求める問題を考えましょう。観測点数が推定したいパラメータの数よりも大きければ、つまり $N > M$ なら解くのは簡単です。例えば、測定誤差が正規分布で表されるような場合は最小二乗法を使えばよいでしょう。推定する $\hat{\boldsymbol{x}}$ は、

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_i \left( y_i - \sum_j a_{ij}x_j \right)^2 \tag{4}$$

と表せます。ここで $\arg\min_{\boldsymbol{x}}$ というのは「後の関数が最小値をとるときの $\boldsymbol{x}$」という意味です。この式はノルムを使うことでもう少し簡単に表現できます。

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 \tag{5}$$

ここで、$M$ 次元のベクトル $\boldsymbol{x}$ の $p$ 次ノルムとは以下で定義される量です。

$$\|\boldsymbol{x}\|_p = \left( \sum_i^M |x_i|^p \right)^{1/p} = \sqrt[p]{|x_1|^p + \cdots + |x_M|^p} \tag{6}$$

したがって、$\|\boldsymbol{x}\|_2^2$ というのは「2 次ノルムの 2 乗」です。

さて、$M < N$ のとき、つまり、推定したい変数の数が観測点数よりも多いときはどうなるでしょう。この時、最小二乗法だけでは解は無限に存在してしまい、無限にある解のなかから 1 つの解を選ぶためには最小二乗項 $\|\boldsymbol{y} - A\boldsymbol{x}\|_2^2$ に加えて解を制約する他の条件が必要になります。スパースモデリングの一種である「1 次ノルム最小化」はそのような手法の 1 つです。

　ここで、統計解析に特化したスクリプト言語「`R`」 を使って簡単なデモをしてみましょう。パラメータ ($\boldsymbol{x}$)100 個を観測 ($\boldsymbol{y}$)50 点から推定する問題を考えます。$A$ は 50 行 100 列の行列で、その要素は平均 0.0、標準偏差 1.0 の正規乱数にします。

```
> n <- 50
> m <- 100
> A <- matrix(rnorm(n*m,mean=0.0,sd=1.0),ncol=m)
```

パラメータ $\boldsymbol{x}$ は適当に用意しますが、今は秘密にしましょう。次に、

```
> y <- A%*%x
```

によって 50 個のデータ点を作ります。この 50 個のデータから $\boldsymbol{x}$ の要素 100 個を推定します。具体的には統計処理に特化したスクリプト言語「`R`」 のライブラリ「`glmnet`」[1]を使います。

```
> library(lars)
> library(glmnet)
> glmnet.control(fdev=1e-9,devmax=0.99999999)
> g <- glmnet(A,y,alpha=1,lambda.min.ratio=0.00001)
> xh <- g$beta[,g$dim[2]]
```

最初の 2 行は必要なライブラリを読み込むもので、別途ライブラリをインストールしておく必要があるかもしれません。3 行目はこのデモに限り内部の設定を変えているもので、通常は必要ありません。詳しくは `glmnet` のマニュアルを参照。4 行目で問題を解いています。`xh` が $\boldsymbol{x}$ の推定結果 $\hat{\boldsymbol{x}}$ です。図 1 は推定された $\hat{\boldsymbol{x}}$、元々設定した「正解」である $\boldsymbol{x}$、それらの差 $\Delta\boldsymbol{x} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ を表示しています。$\hat{\boldsymbol{x}}$ が $\boldsymbol{x}$ をよく再現していて、残差 $\Delta\boldsymbol{x}$ をみると、0.1 %以下の精度で推定できていることがわかります。天文学では多くの場合、このような高精度で推定できれば「$\boldsymbol{x}$ は完全に再構成された」とみなして問題ないでしょう。

　このように、少ない観測データからより多くのパラメータを推定するという、魔法のようなことができました。これには 2 つの秘密があります。1 つは図 1 をみてわかるように、$\boldsymbol{x}$ の 100 個の要素のほとんどが実は 0 で、0 でない値をもっている要素（非ゼロ要素）はわずかであることです。このデモでは $\boldsymbol{x}$ は以下のようにして生成されました。

```
> k <- 10
> x <- rep(0,m)
> nz <- floor(runif(k,min=1,max=m+1))
> x[nz] <- runif(k,min=-10,max=+10)
```

なお `m` は最初に設定したものと共通で、このデモでは $m = 100$ です。100 個の要素のうち、$k = 10$ 個のみが非ゼロ要素となるよう設定されています。このようなベクトルを「疎性が高い」「スパース (sparse) である」といいます。もう 1 つの秘密は、`glmnet` ではこの問題の解を決めるために $\boldsymbol{x}$ の 1 次 ($\ell_1$) ノルムの項を制約項として、以下の評価式を最小化するように $\hat{\boldsymbol{x}}$

[1] https://cran.r-project.org/web/packages/glmnet/

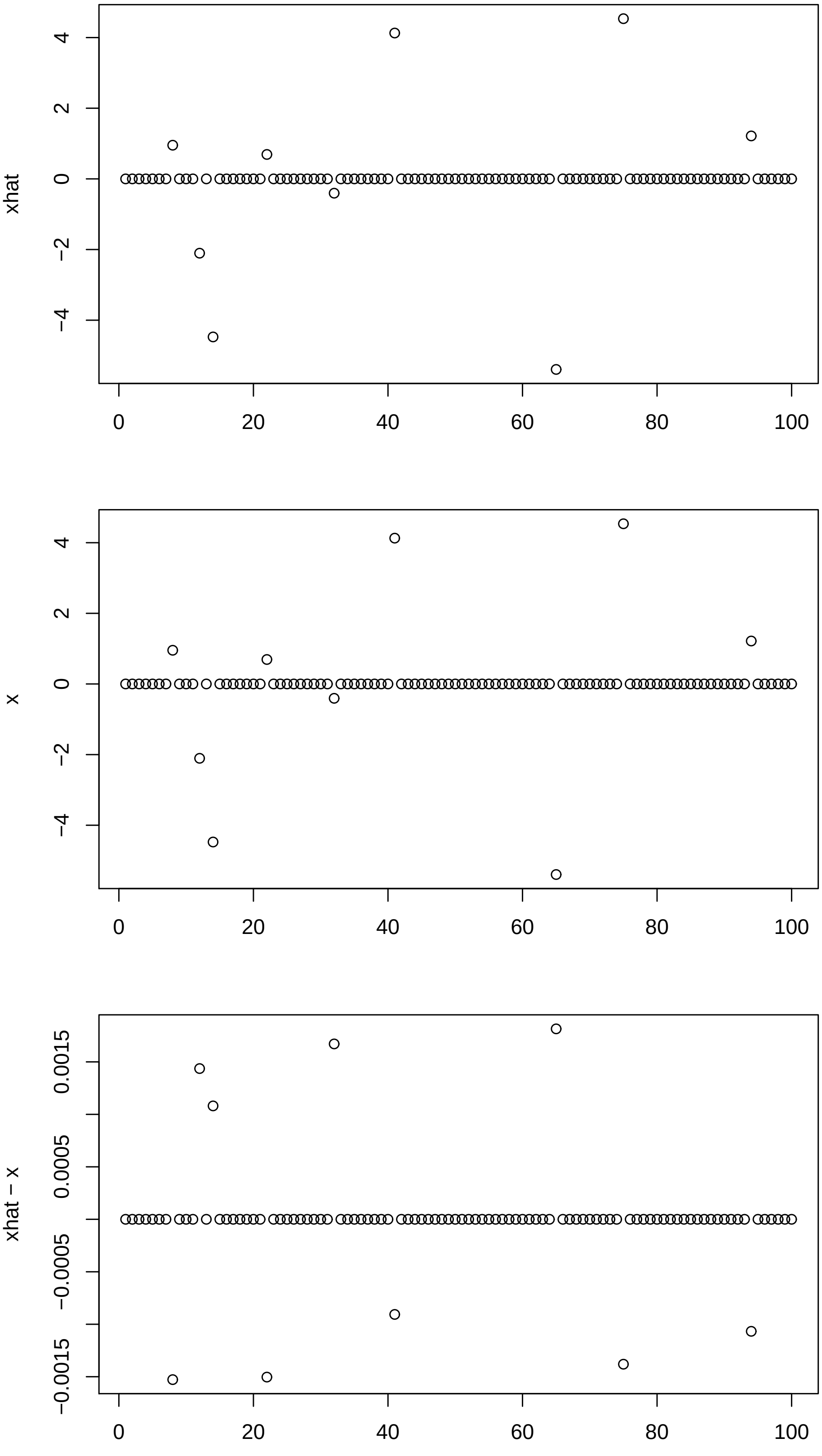


図 1: （上から）推定された $\hat{\boldsymbol{x}}$、本来の $\boldsymbol{x}$、両者の差 $\Delta\boldsymbol{x} = \hat{\boldsymbol{x}} - \boldsymbol{x}$

を決めていることです。

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \tag{7}$$

(6) 式からわかるように、1 次ノルム ($= \ell_1$) とは絶対値の総和（$\|\boldsymbol{x}\|_1 = \sum_i |x_i|$）です。また $\lambda$ は $\ell_1$ 項の「重み」で、この値を変えることで $\boldsymbol{x}$ のスパース性を調整することができ、交差検定 (cross validation) や情報量規準によってデータ $\boldsymbol{y}$ から適切な $\lambda$ を決めることもできます。$\lambda$ の決定については本稿 6 章で触れます。

スパースなベクトルであれば、1 次ノルム最小化を使うことで少ない情報からでも元情報が再構成できます。たとえ $\boldsymbol{x}$ そのものがスパースでなくても、ある線形変換 $\boldsymbol{x} = \Phi\boldsymbol{x}'$ で $\boldsymbol{x}'$ がスパースになるような $\Phi$ があれば、$A' = A\Phi$ として $A'$ と $\boldsymbol{x}'$ について上の式と同じように解けます。$\boldsymbol{x}$ は「画像」とみなすこともでき、画像の再構成にも応用することができます。

次回は 1 次ノルム最小化やその周辺の研究の経緯に関して紹介します。