クレジット：
Mathematics and Informatics Center 計算機実験I 2020 藤堂眞治

東京大学 THE UNIVERSITY OF TOKYO

BY NC ND

計算機実験 I (第 8 回)

藤堂眞治

2020/06/17

物理の問題にあらわれる行列演算

- 連立一次方程式・逆行列
  - 偏微分方程式の境界値問題
  - 非線形連立方程式に対するニュートン法
- 対角化・特異値分解
  - 固有値問題・行列関数
  - 最小二乗近似
- 計算機は大規模行列演算が得意
  - 直接法: $\sim 10^4$ 次元
  - 疎行列に対する反復解法: $\sim 10^9$ 次元
- 行列演算についてはライブラリがよく整備されている
  - それぞれの原理とその特徴を理解して正しく使うことが重要
  - 適切なライブラリを使うことで数十倍あるいはそれ以上速くなることとも

一般の非正方行列の場合

- 特異値分解 (SVD: Singular Value Decomposition)
- 任意の $m \times n$ 実行列 A は

$$A = U \Lambda V^T$$

の形に (一意に) 分解できる ($k = \min(m, n)$)

- U: $(m \times k)$ 行列 (列ベクトルは互いに正規直交)
  V: $(n \times k)$ 行列 (列ベクトルは互いに正規直交)
  $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_k) \quad (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0)$ 特異値
- ベクトル表示 (行列をランク 1 の行列で分解)

$$A = \sum_{i=1}^{k} \lambda_i u_i v_i^T$$

特異値分解の例

$$\begin{pmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} -0.14 & -0.62 & -0.05 \\ -0.34 & 0.37 & 0.81 \\ -0.55 & 0.54 & -0.58 \\ -0.75 & -0.44 & 0.06 \end{pmatrix}$$

$$\times \begin{pmatrix} 25.35 & 0 & 0 \\ 0 & 2.15 & 0 \\ 0 & 0 & 1.71 \end{pmatrix} \begin{pmatrix} -0.56 & -0.59 & -0.59 \\ 0.68 & 0.09 & -0.73 \\ 0.48 & -0.81 & 0.35 \end{pmatrix}$$

完全特異値分解 (full SVD)

$$
\begin{pmatrix} 1 & 2 & 3 \\ 6 & 4 & 5 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} -0.14 & -0.62 & -0.05 & \color{red}{-0.77} \\ -0.34 & 0.37 & 0.81 & \color{red}{-0.29} \\ -0.55 & 0.54 & -0.58 & \color{red}{-0.29} \\ -0.75 & -0.44 & 0.06 & \color{red}{0.48} \end{pmatrix}
\times \begin{pmatrix} 25.35 & 0 & 0 \\ 0 & 2.15 & 0 \\ 0 & 0 & 1.71 \\ \color{red}{0} & \color{red}{0} & \color{red}{0} \end{pmatrix} \begin{pmatrix} -0.56 & -0.59 & -0.59 \\ 0.68 & 0.09 & -0.73 \\ 0.48 & -0.81 & 0.35 \end{pmatrix}
$$

- U, V が直交行列になるように、足りない基底ベクトルを追加
- こちらを「SVD」、もともとの分解を「thin SVD」と呼ぶこともある

LAPACK による特異値分解

- 倍精度実行列の特異値分解 dgesvd http://www.netlib.org/lapack/explore-html/d8/d2d/dgesvd_8f.html
- Fortran による関数宣言

```
subroutine dgesvd(character JOBU, character JOBVT,
  integer M, integer N,
  double precision, dimension(lda, *) A,
  integer LDA, double precision, dimension(*) S,
  double precision, dimension(ldu, *) U, integer LDU,
  double precision, dimension(ldvt, *) VT,
  integer LDVT,
  double precision, dimension(*) WORK, integer LWORK,
  integer INFO)
```

- dgesvd の使用例: svd.c (SVD), full_svd.c (完全 SVD)
  - コンパイル方法: `cc -o svd svd.c -llapack -lblas -lm`
  - 実行方法: `./svd matrix2.dat`

固有値分解との関係

- 半正定値実対称行列の場合
  「固有値分解」と「特異値分解」は等価
- 一般の (非正方) 実行列 A の場合
  - 完全 SVD $A = U\Lambda V^T$ を考えると
  - $B = A^T A = V\Lambda U^T U\Lambda V^T = V\Lambda^2 V^T$ は実対称行列
    固有値:λ_i^2、固有ベクトル V
  - $C = AA^T = U\Lambda V^T V\Lambda U^T = U\Lambda^2 U^T$ も実対称行列
    固有値:λ_i^2、固有ベクトル U

特異値分解が役に立つ例

- 連立一次方程式の最小二乗解
- 行列の低ランク近似
- 画像圧縮

特異な (特異に近い) 連立一次方程式

- ランク r の $m \times n$ 行列 A の full SVD を考える

$$A = (U_1 U_2) \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} (V_1 V_2)^T = U_1 \Lambda_1 V_1^T$$

U_1: $m \times r$, U_2: $m \times (m-r)$, Λ_1: $r \times r$, V_1: $n \times r$, V_2: $n \times (n-r)$

- $r \neq n$ のとき、方程式 $Ax = b$ は無限個の解をもつ、あるいは解なし
- 無限個の解をもつ場合 $(U_1 U_1^T b = b)$ の一般解

$$x = V_1 \Lambda_1^{-1} U_1^T b + V_2 z$$

z は $(n-r)$ 次元の任意のベクトル

一般化逆行列と最小二乗解

- V_1 と V_2 の列ベクトルは全て直交する。一般解のうちノルム $|x|^2$ が最小となるのは $z = 0$ のとき

$$x = A^{\dagger}b \quad A^{\dagger} = V \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T$$

- $A^{\dagger}$: Moore-Penrose の一般化逆行列 (1/0 を 0 とおくことに相当)
- $Ax = b$ が解を持たない場合にも最良解を与える
- 零に非常に近い特異値がある場合も、それらを零とみなすことで数値的に安定した解がえられる

行列の低ランク近似

- ランク r $(r < k)$ の行列のうち、行列 A を「最も良く近似」する $\tilde{A}$ を選ぶ
- 「最も良く近似」$=$ フロベニウスノルム $||A - \tilde{A}||_{\mathrm{F}}$ を最小化

$$||X||_{\mathrm{F}}^2 \equiv \sum_{ij} x_{ij}^2$$

- 特異値のうち大きなものから r 個とり、残りを零とした

$$\tilde{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0)$$

を使い

$$\tilde{A} = U\tilde{\Lambda}V^T$$

を作れば良い (Eckart-Young の定理)

行列の低ランク近似の例

$$
\begin{pmatrix} -0.14 & -0.62 & -0.05 \\ -0.34 & 0.37 & 0.81 \\ -0.55 & 0.54 & -0.58 \\ -0.75 & -0.44 & 0.06 \end{pmatrix}
\begin{pmatrix} 25.35 & 0 & 0 \\ 0 & 2.15 & 0 \\ 0 & 0 & 0 \end{pmatrix}
$$

$$
\times \begin{pmatrix} -0.56 & -0.59 & -0.59 \\ 0.68 & 0.09 & -0.73 \\ 0.48 & -0.81 & 0.35 \end{pmatrix}
= \begin{pmatrix} 1.04 & 1.93 & 3.03 \\ 5.33 & 5.12 & 4.52 \\ 8.47 & 8.21 & 7.34 \\ 9.95 & 11.1 & 12.0 \end{pmatrix}
$$

- それなりに悪くない近似が得られる
- 誤差 (フロベニウスノルム) $= 1.71$ (無視した特異値の二乗和の平方根)

Eckart-Young の定理の証明

- A を近似する行列 X (ランク $\leq r$) を考えると

$$||A - X||_{\mathrm{F}}^2 = \sum_{ij} (a_{ij} - x_{ij})^2 = \mathrm{tr}[(A - X)(A - X)^T]$$

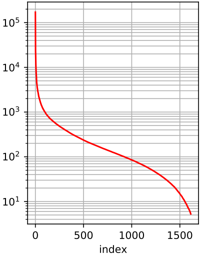
を最小化すればよい。完全 SVD について $UU^T = E_m$, $VV^T = E_n$ より

$$\begin{aligned} ||A - X||_{\mathrm{F}}^2 &= \mathrm{tr}[UU^T(A - X)VV^T(A - X)^T] \\ &= \mathrm{tr}[(\Lambda - G)(\Lambda - G)^T] \quad (G \equiv U^T XV) \\ &= \sum_i^k (\lambda_i - g_{ii})^2 + \sum_i \sum_{j \neq i} g_{ij}^2 \end{aligned}$$

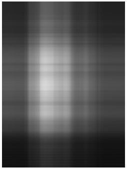
ランク k 以下で、$||A - X||_{\mathrm{F}}^2$ を最小化するには、$g_{ii} = \lambda_i$ $(i = 1 \cdots r)$、それ以外は全て零とすればよい

特異値分解による画像圧縮

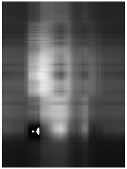
- 特異値の分布とランク r 近似 (1614 × 2178 グレイスケール写真)



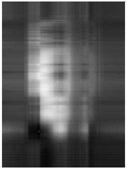
$r = 1$



$r = 2$



$r = 3$



$r = 4$

最小二乗法によるフィッティング

- 説明変数 (例: 電圧): $x_1, x_2, x_3, \cdots, x_n$
- 観測値 (例: 電流): $y_1, y_2, y_3, \cdots, y_n$
- 単回帰モデル: $y = a + bx + \epsilon \quad$ (ϵ: ノイズ)
- 未知母数: a, b
- 最小二乗法: 残差 $R(a, b) = \displaystyle\sum_i^n (y_i - (a + bx_i))^2$ を最小化

$$\frac{\partial R}{\partial a} = -2 \sum_i^n (y_i - (a + bx_i)) = 0$$

$$\frac{\partial R}{\partial b} = -2 \sum_i^n (y_i - (a + bx_i))x_i = 0$$

回帰分析の一般化

- 基底関数: $\phi_j(x) \quad (j = 1 \cdots m)$
- モデル: $y(x) = \sum_j^m \phi_j(x) w_j + \epsilon$
- 残差: $R(\mathbf{w}) = \sum_i^n \left[y_i - \sum_j^m \phi_j(x_i) w_j \right]^2$

$$\frac{\partial R}{\partial w_k} = -2 \sum_i^n \left[y_i - \sum_j^m \phi_j(x_i) w_j \right] \phi_k(x_i) = 0 \qquad (k = 1 \cdots m)$$

- 計画行列 (design matrix) $\Phi_{ij} = \phi_j(x_i)$ を導入すると

$$R(\mathbf{w}) = |\mathbf{y} - \Phi \mathbf{w}|^2$$

- 最小二乗解: $\Phi^{\mathrm{t}} \Phi \mathbf{w} = \Phi^{\mathrm{t}} \mathbf{y} \quad \Rightarrow \quad \mathbf{w} = (\Phi^{\mathrm{t}} \Phi)^{-1} \Phi^{\mathrm{t}} \mathbf{y}$

リッジ回帰 (Ridge Regression)

- 基底関数が線形独立でない (例: $\phi_1(x) = 1$, $\phi_2(x) = x$, $\phi_3(x) = 1 + x$) と、Gram 行列 $(\Phi^{\rm t}\Phi)$ が特異になる
- 基底関数の数 (例: 多項式の次数) を増やしすぎると過学習 (over-fitting)」が生じる
- 正則化最小二乗法 (λ は非負の定数)

$$\begin{aligned} R(\mathbf{w}) &= \sum_i^n \Big[y_i - \sum_j^m \phi_j(x_i) w_j \Big]^2 + \lambda \sum_j^m w_j^2 \\ &= |\mathbf{y} - \Phi\mathbf{w}|^2 + \lambda \mathbf{w}^{\rm t}\mathbf{w} \end{aligned}$$

- 最小二乗解

$$(\Phi^{\rm t}\Phi + \lambda\,\mathrm{I})\mathbf{w} = \Phi^{\rm t}\mathbf{y} \quad \Rightarrow \quad \mathbf{w} = (\Phi^{\rm t}\Phi + \lambda\,\mathrm{I})^{-1}\Phi^{\rm t}\mathbf{y}$$

