Computational methods for least squares problems

Imre Pólik, PhD

McMaster University
School of Computational Engineering and Science

January 28, 2008

Outline

1. Introduction
   - Motivation
   - Applications
2. Approaches
   - Normal equations
   - QR decomposition
     - QR with Gram-Schmidt
     - Householder transformation
     - Givens rotation
   - Singular value decomposition
   - Rank deficient problems with SVD
   - Transformation to a linear system
3. Summary

Motivation

- How to solve an unsolvable linear system?

$$Ax = b,\ A \text{ is } m \times n$$

- $m > n$
  - overdetermined system
  - possibly no solutions
  - "best" solution: minimize $\|Ax - b\|$
- $m < n$
  - underdetermined system
  - infinitely many solutions
  - choose minimum norm x

Applications

- Polynomial regression
  Given data points $(y_i, b_i), i = 1, \ldots, m$ find the coefficients $x_1, \ldots x_4$ such that $\sum_{i=1}^{k} (b_i - p(y_i))^2$ is minimal, $p(y) = x_1 + x_2 y + x_3 y^2 + x_4 y^3$

$$A = \begin{pmatrix} 1 & y_1 & y_1^2 & y_1^3 \\ 1 & y_2 & y_2^2 & y_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_m & y_m^2 & y_m^3 \end{pmatrix}$$

  ill-conditioned for high degrees
- Noisy measurements
  - A is exact, b has Gaussian noise
  - least squares solution is the maximum likelyhood estimator

Normal equations

- Solve $A^T A x = A^T b$
  - minimizes $\|Ax - b\|$
  - A must be full rank
  - uses existing technology (Cholesky, refinement, etc.)
- Conditioning
  - $\kappa\left(A^T A\right) = \kappa(A)^2$
  - fewer accurate digits
  - not stable
- Cost
  - $n^2 m + \frac{1}{3} n^3 + \mathcal{O}(n^2) = \mathcal{O}(n^2 m)$, since $m \geq n$
  - $A^T A$ dominates

QR decomposition

- Write $A = QR$, A is $m \times n$, full rank, $m \geq n$
  - Q: orthogonal ($Q^T Q = I$), $m \times n$
  - R: upper triangular, positive diagonal, $n \times n$
    - $Rx = Q^T b$
- Cost
  - $2n^2 m - \frac{2}{3} n^3$
  - same as NE if $m \approx n$
  - twice as much if $m \gg n$
- Lapack `sgels`

QR with Gram Schmidt orthogonalization

- Observation: the columns of A are linear combinations of the first few columns of Q

$$
\begin{aligned}
& i = 1 : n \\
& \quad q_i = a_i \\
& \quad j = 1 : i-1 \\
& \quad\quad r_{ji} = q_j^T a_i \text{ (classical GS), or} \\
& \quad\quad r_{ji} = q_j^T q_i \text{ (modified GS)} \\
& \quad\quad q_i = q_i - r_{ji} q_j \\
& \quad r_{ii} = \|q_i\|_2 \\
& \quad q_i = q_i / r_{ii}
\end{aligned}
$$

- CGS is unstable in floating point arithmetic
- MGS is better but still not good for ill-conditioned A

QR with Householder transformations

- Householder reflection
  - $P = I - 2uu^T$, $\|u\|_2 = 1$
  - reflection to a plane perpendicular to u
- Find u such that $Px = (*, 0, \dots, 0)$

$$u = \frac{x \pm \|x\|_2 e_1}{\left\|x \pm \|x\|_2 e_1\right\|}$$

- Apply these transformations to the columns of A
  - $P_n P_{n-1} \dots P_1 A = R$
  - don't form P_i
  - store u_i in A
  - clever choice of the sign
- Difficult to parallelize efficiently
- Pivoting: always work on the largest norm column

QR with Givens rotations

- Givens rotation (counterclockwise by θ)

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

- Rotate (x_1, x_2) to $\left(\sqrt{x_1^2 + x_2^2}, 0\right)$
- Apply to nonzeros in A
- Storage: $\operatorname{sign}(c)s$ or $\operatorname{sign}(s)/c$ for accuracy
- Better parallelization than Householder
- Costs twice as much
- More efficient for sparse A

Singular value decomposition

- Write $A = U\Sigma V^T$, where
  - U: $m \times n$, orthogonal, $U^T U = I$
  - Σ: $n \times n$, diagonal, nonnegative
  - V: $n \times n$, orthogonal, $VV^T = I$
- $x = V\Sigma^{-1}U^T b$
- Expensive decomposition
- Quick solution:
  - same as QR if $m \gg n$
  - $4n^2m - \frac{4}{3}n^3 + \mathcal{O}(n^2)$ otherwise

Rank deficient problems with SVD

- Multiple solutions
- Minimum norm solution

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1 & V_2 \end{pmatrix}^T = U_1 \Sigma_1 V_1^T$$

$$x = V_1 \Sigma_1^{-1} U_1^T b$$

- $\|x\|_2 \leq \|b\|_2 / \sigma$, σ is the smallest nonzero singular value
- Truncated SVD for uncertain data

Transformation to a linear system

- Overdetermined, full rank system $(m \geq n)$
- Solve

$$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}$$

- Symmetric, nonsingular system $(m \geq n$, A is full rank)
- Existing methods can be used
- Condition number: $c\kappa(A)$
- Iterative refinement: $\Delta x = R^{-1}Q^T b - x$

Summary

- Normal equations
  - cheapest
  - least accurate
  - not for ill-conditioned systems
- QR decomposition
  - twice as expensive
  - standard method
  - pivoting for ill-conditioned problems
- Singular value decomposition
  - more expensive
  - can not exploit sparsity
  - suitable for ill-conditioned problems (A is not full rank)
- Transformation to a linear system
  - allows iterative refinement
  - requires a full rank A

James W. Demmel.
Applied Numerical Linear Algebra.
SIAM, Philadelphia, 1997.