CSE 494 CSE/CBS 598 (Fall 2007): Numerical Linear Algebra for Data Exploration— QR, least squares, linear regression

Instructor: Jieping Ye

1 Matrix Decompositions

- We wish to decompose the matrix A by writing it as a product of two or more matrices:

$$\begin{aligned} A &= BC, \text{ where } A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times k}, C \in \mathbb{R}^{k\times n}, \text{ or} \\ A &= BCD, \text{ where } A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times k}, C \in \mathbb{R}^{k\times r}, D \in \mathbb{R}^{r\times n}. \end{aligned}$$

- This is done in such a way that the right side of the equation yields some useful information or insight to the nature of the data matrix A.
- Or is in other ways useful for solving the problem at hand.
- There are numerous examples of useful matrix decompositions:
  - Eigendecomposition: $A = U\Lambda U^T$ (A symmetric).
  - QR Decomposition: $A = QR$, Q has orthonormal columns and R is upper triangular.
  - Singular Value Decomposition: $A = U\Sigma V^T$, U and V are orthogonal, and Σ is diagonal with nonnegative diagonal entries.
- Matrix factorization is the same thing as matrix decomposition.
- How to compute these decompositions?
  - Manipulate A by multiplying it by intelligently chosen, fairly simple (orthogonal) matrices from both sides:

$$V_k \cdots V_2 V_1 A W_1 W_2 \cdots W_s = B \quad (\text{ until } B \text{ has the desirable form.})$$

  - The key is how to choose V_i's and W_j's.

1.1 Givens (plane) rotations

$$G = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}, \quad c^2 + s^2 = 1.$$

- We can choose c and s so that $c = \cos(\theta)$, $s = \sin(\theta)$ for some θ.
- Then multiplication of a vector x by G means that we rotate the vector in $\mathbb{R}^2$ by an angle θ.
- For a given vector $x = (x_1, x_2)^T$, we can form matrix G defined as above with $c = x_1/||x||_2$ and $s = x_2/||x||_2$. Show that $Gx = (||x||_2, 0)^T$.

- Let x be a vector. The parameters c and s ($c^2 + s^2 = 1$) can be chosen so that multiplication of x by

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & c & 0 & s \\ 0 & 0 & 1 & 0 \\ 0 & -s & 0 & c \end{pmatrix}$$

will zero the 4-th element of vector x by a rotation.

- By applying several Givens rotations in succession, we can transform x to $(\alpha, 0, 0, \cdots, 0)^T$.

1.2 Householder Transformations

- For a given nonzero vector $v \in \mathbb{R}^n$, we can define the following Householder matrix:

$$P = I_n - \frac{2}{v^T v} v v^T.$$

- Show that P is symmetric and orthogonal.
- Question: Given two vectors x and y of the same length, $||x||_2 = ||y||_2$, can we determine a Householder transformation P such that $Px = y$?
- It follows from $Px = y$ that

$$y \quad = \quad Px = \left(I_n - \frac{2}{v^T v} v v^T \right) x = x - \frac{2 v^T x}{v^T v} v.$$

Thus v is a multiple of $x - y$.

- We can simply choose $v = x - y$, as both v and $\alpha\, v$ for any scalar α result in the same Householder transformation.
- In practice, the matrix P should not be formed explicitly, since it can be represented much more compactly by the vector v. Multiplication by P can be done according to

$$Px = x - \frac{2 v^T x}{v^T v} v.$$

If we normalize the vector v to have a unit length, $||v||_2 = 1$ then

$$Px = x - (2 v^T x) v.$$

1.3 QR Transformation

- $A = QR$, where Q is orthogonal and R is upper triangular.
  - Any matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ can be transformed to an upper triangular form by an orthogonal matrix.
  - If the columns of A are linearly independent, then the matrix R is non-singular.

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$$

- In MATLAB, use $[Q, R] = qr(A)$ to compute the QR decomposition.

- Skinny QR decomposition
  - Partition $Q = (Q_1, Q_2)$, where $Q_1 \in \mathbb{R}^{m \times n}$

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix} = (Q_1, Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R.$$

  - In MATLAB, use $[Q, R] = qr(A, 0)$ to compute the skinny QR decomposition.

- Classical Gram-Schmidt
  - $z_1 = a_1$, $q_1 = z_1/||z_1||$.
  - $z_k = a_k - \sum_{i=1}^{k-1} (q_i^T a_k) q_i$, $q_k = z_k/||z_k||$.
  - We thus obtain $A = QR$, where $Q = [q_1, q_2, \cdots, q_n]$ has orthonormal columns, and $R = (r_{ij})$ is upper triangular with

$$r_{ij} = \begin{cases} q_i^T a_j & \text{if } i < j \\ 0 & \text{if } i > j \\ ||z_i|| & \text{if } i = j \end{cases}$$

- QR decomposition can be used for solving the least squares problem.

2 Linear Regression

- Consider the following problem: Given a number of measurements $X = [x_1, \cdots, x_n]$ and an outcome $y = [y_1, \cdots, y_n]$, build a linear model

$$y = b_0 + \sum_{j=1}^{n} x_j b_j.$$

  Here X is used to predict the outcome y.

- It can be written in the following form:

$$y = X^T b, \text{ where } b = [b_1, \cdots, b_n, b_0], X = [x_1, \cdots, x_n, e],$$

  and $e = [1, \cdots, 1]$ is the vector of all ones.

- Fitting a linear model to data is usually done using the method of least squares.

3 The least squares problem

- Solve $Ax = b$, where $A \in \mathbb{R}^{m \times n}$ and $m \geq n$.

- This is an overdetermined system: more equations than unknown. Usually such a system has no solution.

- Approach: make the residual vector
$$r = b - Ax$$
as small as possible.

- Minimize the length of r by solving the following minimization problem:
$$\min_x ||b - Ax||_2^2.$$

- Normal equations:
$$A^T Ax = A^T b.$$
  - If the column vectors of A are linearly independent, then $A^T A$ are non-singular and the normal equations have a unique solution.
  - Limitation: the condition number of $A^T A$ is the square of that of A
$$\kappa(A^T A) = \kappa(A)^2.$$

- Solving least squares problem using QR
  - Let $A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$ be the QR decomposition of A.
  - Partition $Q = (Q_1, Q_2)$, where $Q_1 \in \mathbb{R}^{m\times n}$. Denote
$$Q^T b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \equiv \begin{pmatrix} Q_1^T b \\ Q_2^T b \end{pmatrix}.$$
  - Based on the properties of the 2-norm, we have
$$\begin{aligned} ||b - Ax||_2^2 &= \left\| b - Q \begin{pmatrix} R \\ 0 \end{pmatrix} x \right\|^2 \\ &= \left\| Q \left(Q^T b - \begin{pmatrix} R \\ 0 \end{pmatrix} x \right) \right\|^2 \\ &= \left\| Q^T b - \begin{pmatrix} R \\ 0 \end{pmatrix} x \right\|^2 \\ &= \left\| \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} - \begin{pmatrix} R \\ 0 \end{pmatrix} x \right\|^2 \\ &= \|b_1 - Rx\|^2 + \|b_2\|^2. \end{aligned}$$
  - The second term is a constant. The optimal x is given by solving
$$Rx = b_1.$$