Jim Lambers
MAT 610
Summer Session 2009-10
Lecture 11 Notes

These notes correspond to Sections 5.4, 5.5 and 5.7 in the text.

The Rank-Deficient Least Squares Problem

QR with Column Pivoting

When A does not have full column rank, the property

$$\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \text{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$$

can not be expected to hold, because the first k columns of A could be linearly dependent, while the first k columns of Q, being orthonormal, must be linearly independent.

Example The matrix

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -4 & 0 \\ 1 & -2 & 3 \end{bmatrix}$$

has rank 2, because the first two columns are parallel, and therefore are linearly dependent, while the third column is not parallel to either of the first two. Columns 1 and 3, or columns 2 and 3, form linearly independent sets. □

Therefore, in the case where $\text{rank}(A) = r < n$, we seek a decomposition of the form $A\Pi = QR$, where Π is a *permutation matrix* chosen so that the diagonal elements of R are maximized at each stage. Specifically, suppose H_1 is a Householder reflection chosen so that

$$H_1 A = \begin{bmatrix} r_{11} & \\ 0 & \\ \vdots & * \\ 0 & \end{bmatrix}, \quad r_{11} = \|\mathbf{a}_1\|_2.$$

To maximize r_{11}, we choose Π_1 so that in the column-permuted matrix $A = A\Pi_1$, we have $\|\mathbf{a}_1\|_2 \geq \|\mathbf{a}_j\|_2$ for $j \geq 2$. For Π_2, we examine the lengths of the columns of the submatrix of A obtained by removing the first row and column. It is not necessary to recompute the lengths of the columns, because we can update them by subtracting the square of the first component from the square of the total length.

This process is called QR *with column pivoting*. It yields the decomposition

$$Q\begin{bmatrix} R & S \\ 0 & 0 \end{bmatrix} = A\Pi$$

where $Q = H_1 \cdots H_r$, $\Pi = \Pi_1 \cdots \Pi_r$, and R is an upper triangular, $r \times r$ matrix. The last $m - r$ rows are necessarily zero, because every column of A is a linear combination of the first r columns of Q.

Example We perform QR with column pivoting on the matrix

$$A = \begin{bmatrix} 1 & 3 & 5 & 1 \\ 2 & -1 & 2 & 1 \\ 1 & 4 & 6 & 1 \\ 4 & 5 & 10 & 1 \end{bmatrix}.$$

Computing the 2-norms of the columns yields

$$\|\mathbf{a}_1\|^2 = 22, \quad \|\mathbf{a}_2\|^2 = 51, \quad \|\mathbf{a}_3\|^2 = 165, \quad \|\mathbf{a}_4\|^2 = 4.$$

We see that the third column has the largest 2-norm. We therefore interchange the first and third columns to obtain

$$A^{(1)} = A\Pi_1 = A\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 & 1 & 1 \\ 2 & -1 & 2 & 1 \\ 6 & 4 & 1 & 1 \\ 10 & 5 & 4 & 1 \end{bmatrix}.$$

We then apply a Householder transformation H_1 to $A^{(1)}$ to make the first column a multiple of $\mathbf{e}_1$, which yields

$$H_1 A^{(1)} = \begin{bmatrix} -12.8452 & -6.7729 & -4.2817 & -1.7905 \\ 0 & -2.0953 & 1.4080 & 0.6873 \\ 0 & 0.7141 & -0.7759 & 0.0618 \\ 0 & -0.4765 & 1.0402 & -0.5637 \end{bmatrix}.$$

Next, we consider the submatrix obtained by removing the first row and column of $H_1 A^{(1)}$:

$$\tilde{A}^{(2)} = \begin{bmatrix} -2.0953 & 1.4080 & 0.6873 \\ 0.7141 & -0.7759 & 0.0618 \\ -0.4765 & 1.0402 & -0.5637 \end{bmatrix}.$$

We compute the lengths of the columns, as before, except that this time, we update the lengths of the columns of A, rather than recomputing them. This yields

$$\begin{aligned} \|\tilde{\mathbf{a}}_1^{(2)}\|_2^2 &= \|\mathbf{a}_2^{(1)}\|^2 - [a_{12}^{(1)}]^2 = 51 - (-6.7729)^2 = 5.1273, \\ \|\tilde{\mathbf{a}}_2^{(2)}\|_2^2 &= \|\mathbf{a}_3^{(1)}\|^2 - [a_{13}^{(1)}]^2 = 22 - (-4.2817)^2 = 3.6667, \\ \|\tilde{\mathbf{a}}_3^{(2)}\|_2^2 &= \|\mathbf{a}_4^{(1)}\|^2 - [a_{14}^{(1)}]^2 = 4 - (-1.7905)^2 = 0.7939. \end{aligned}$$

The second column is the largest, so there is no need for a column interchange this time. We apply a Householder transformation $\tilde{H}_2$ to the first column of $\tilde{A}^{(2)}$ so that the updated column is a multiple of $\mathbf{e}_1$, which is equivalent to applying a 4×4 Householder transformation $H_2 = I - 2\mathbf{v}_2\mathbf{v}_2^T$, where the first component of $\mathbf{v}_2$ is zero, to the *second* column of $A^{(2)}$ so that the updated column is a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$. This yields

$$\tilde{H}_2\tilde{A}^{(2)} = \begin{bmatrix} 2.2643 & -1.7665 & -0.4978 \\ 0 & -0.2559 & 0.2559 \\ 0 & 0.6933 & -0.6933 \end{bmatrix}.$$

Then, we consider the submatrix obtained by removing the first row and column of $H_2\tilde{A}^{(2)}$:

$$\tilde{A}^{(3)} = \begin{bmatrix} -0.2559 & 0.2559 \\ 0.6933 & -0.6933 \end{bmatrix}.$$

Both columns have the same lengths, so no column interchange is required. Applying a Householder reflection $\tilde{H}_3$ to the first column to make it a multiple of $\mathbf{e}_1$ will have the same effect on the second column, because they are parallel. We have

$$\tilde{H}_3\tilde{A}^{(3)} = \begin{bmatrix} 0.7390 & -0.7390 \\ 0 & 0 \end{bmatrix}.$$

It follows that the matrix $\tilde{A}^{(4)}$ obtained by removing the first row and column of $H_3\tilde{A}^{(3)}$ will be the zero matrix. We conclude that $\text{rank}(A) = 3$, and that A has the factorization

$$A\Pi = Q \begin{bmatrix} R & S \\ 0 & 0 \end{bmatrix},$$

where

$$\Pi = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} -12.8452 & -6.7729 & -4.2817 \\ 0 & 2.2643 & -1.7665 \\ 0 & 0 & 0.7390 \end{bmatrix}, \quad S = \begin{bmatrix} -1.7905 \\ -0.4978 \\ -0.7390 \end{bmatrix},$$

and $Q = H_1H_2H_3$ is the product of the Householder reflections used to reduce $A\Pi$ to upper-triangular form. □

Using this decomposition, we can solve the linear least squares problem $A\mathbf{x} = \mathbf{b}$ by observing that

$$\|\mathbf{b} - A\mathbf{x}\|_2^2 = \left\| \mathbf{b} - Q \begin{bmatrix} R & S \\ 0 & 0 \end{bmatrix} \Pi^T\mathbf{x} \right\|_2^2$$

$$
\begin{aligned}
&= \left\| Q^T\mathbf{b} - \begin{bmatrix} R & S \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \right\|_2^2 \\
&= \left\| \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} - \begin{bmatrix} R\mathbf{u} + S\mathbf{v} \\ \mathbf{0} \end{bmatrix} \right\|_2^2 \\
&= \|\mathbf{c} - R\mathbf{u} - S\mathbf{v}\|_2^2 + \|\mathbf{d}\|_2^2,
\end{aligned}
$$

where

$$
Q^T\mathbf{b} = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}, \quad \Pi^T\mathbf{x} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix},
$$

with $\mathbf{c}$ and $\mathbf{u}$ being r-vectors. Thus $\min \|\mathbf{b} - A\mathbf{x}\|_2^2 = \|\mathbf{d}\|_2^2$, provided that $R\mathbf{u} + S\mathbf{v} = \mathbf{c}$. A *basic solution* is obtained by choosing $\mathbf{v} = \mathbf{0}$. A second solution is to choose $\mathbf{u}$ and $\mathbf{v}$ so that $\|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2$ is minimized. This criterion is related to the pseudo-inverse of A.

Complete Orthogonal Decomposition

After performing the QR factorization with column pivoting, we have

$$
A = Q \begin{bmatrix} R & S \\ 0 & 0 \end{bmatrix} \Pi^T
$$

where R is upper triangular. Then

$$
A^T = \Pi \begin{bmatrix} R^T & 0 \\ S^T & 0 \end{bmatrix} Q^T
$$

where R^T is lower triangular. We apply Householder reflections so that

$$
Z_r \cdots Z_2 Z_1 \begin{bmatrix} R^T & 0 \\ S^T & 0 \end{bmatrix} = \begin{bmatrix} U & 0 \\ 0 & 0 \end{bmatrix},
$$

where U is upper-triangular. Then

$$
A^T = Z \begin{bmatrix} U & 0 \\ 0 & 0 \end{bmatrix} Q^T
$$

where $Z = \Pi Z_1 \cdots Z_r$. In other words,

$$
A = Q \begin{bmatrix} L & 0 \\ 0 & 0 \end{bmatrix} Z^T
$$

where L is a lower-triangular matrix of size $r \times r$, where r is the rank of A. This is the *complete orthogonal decomposition* of A.

Recall that X is the *pseudo-inverse* of A if

1. $AXA = A$
2. $XAX = X$
3. $(XA)^T = XA$
4. $(AX)^T = AX$

Given the above complete orthogonal decomposition of A, the pseudo-inverse of A, denoted A^+, is given by

$$A^+ = Z \begin{bmatrix} L^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q^T.$$

Let $\mathcal{X} = \{\mathbf{x} | \|\mathbf{b} - A\mathbf{x}\|_2 = \min\}$. If $\mathbf{x} \in \mathcal{X}$ and we desire $\|\mathbf{x}\|_2 = \min$, then $\mathbf{x} = A^+\mathbf{b}$. Note that in this case,

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} = \mathbf{b} - AA^+\mathbf{b} = (I - AA^+)\mathbf{b}$$

where the matrix $(I - AA^+)$ is a projection matrix $P^\perp$. To see that $P^\perp$ is a projection, note that

$$\begin{aligned} P &= AA^+ \\ &= Q \begin{bmatrix} L & 0 \\ 0 & 0 \end{bmatrix} Z^T Z \begin{bmatrix} L^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q^T \\ &= Q \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q^T. \end{aligned}$$

It can then be verified directly that $P = P^T$ and $P^2 = P$.

Applications of the SVD

Minimum-norm least squares solution

One of the most well-known applications of the SVD is that it can be used to obtain the solution to the problem

$$\|\mathbf{b} - A\mathbf{x}\|_2 = \min, \quad \|\mathbf{x}\|_2 = \min.$$

The solution is

$$\hat{\mathbf{x}} = A^+\mathbf{b} = V\Sigma^+U^T\mathbf{b}$$

where A^+ is the *pseudo-inverse* of A.

Closest Orthogonal Matrix

Let $\mathcal{Q}_n$ be the set of all $n \times n$ orthogonal matrices. Given an $n \times n$ matrix A, we wish to find the matrix Q that satisfies

$$\|A - Q\|_F = \min\,, \quad Q \in \mathcal{Q}_n, \quad \sigma_i(Q) = 1.$$

Given $A = U\Sigma V^T$, if we compute $\hat{Q} = UIV^T$, then

$$\begin{aligned} \|A - \hat{Q}\|_F^2 &= \|U(\Sigma - I)V^T\|_F^2 \\ &= \|\Sigma - I\|_F^2 \\ &= (\sigma_1 - 1)^2 + \cdots + (\sigma_n - 1)^2 \end{aligned}$$

It can be shown that this is in fact the minimum.

Low-Rank Approximations

Let $\mathcal{M}_{m,n}^{(r)}$ be the set of all $m \times n$ matrices of rank r, and let $A \in \mathcal{M}_{m,n}^{(r)}$. We wish to find $B \in \mathcal{M}_{m,n}^{(k)}$, where $k < r$, such that $\|A - B\|_F = \min$.

To solve this problem, let $A = U\Sigma V^T$ be the SVD of A, and let $\hat{B} = U\Sigma_k V^T$ where

$$\Sigma_k = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}.$$

Then

$$\begin{aligned} \|A - \hat{B}\|_F^2 &= \|U(\Sigma - \Sigma_k)V^T\|_F^2 \\ &= \|\Sigma - \Sigma_k\|_F^2 \\ &= \sigma_{k+1}^2 + \cdots + \sigma_r^2. \end{aligned}$$

We now consider a variation of this problem. We wish to find $\hat{B}$ such that $\|A - \hat{B}\|_F^2 \leq \epsilon^2$, where the rank of $\hat{B}$ is minimized. We know that if $B_k = U\Sigma_k V^T$ then

$$\|A - B_K\|_F^2 = \sigma_{k+1}^2 + \cdots + \sigma_r^2.$$

It follows that $\hat{B} = B_k$ is the solution if

$$\sigma_{k+1} + \cdots + \sigma_r^2 \leq \epsilon^2, \quad \sigma_k^2 + \cdots + \sigma_r^2 > \epsilon^2.$$

Note that

$$\|A^+ - \hat{B}^+\|_F^2 = \left(\frac{1}{\sigma_{k+1}^2} + \cdots + \frac{1}{\sigma_r^2} \right).$$