Pergamon

PII: S0893-9659(97)00098-0

The Behavior of the QR-Factorization Algorithm with Column Pivoting

H. ENGLER
Department of Mathematics
Georgetown University
Washington, DC 20057, U.S.A.

Communicated by R. P. Tewarson

Abstract—A bound on the performance of QR-factorization with column pivoting is derived and two classes of matrices are constructed for which the bound is sharp or asymptotically sharp.

1. INTRODUCTION

QR-factorization with column pivoting (QR-CP) is a version of the usual QR-factorization that was proposed in [1] to handle rank-deficient least squares problems. The method is also used to detect rank-deficiency or near rank-deficiency, although it is not entirely reliable for this purpose [2]. Therefore, various modifications have been proposed to obtain "rank-revealing" versions, see [3,4]. Another related use occurs in subset selection or variable selection [2,5], a problem that arises in regression analysis in which a subset of columns of a given matrix is to be selected to form a well-conditioned submatrix. The purpose of this note is to derive a simple performance bound for this last use of the method, to characterize the few cases in which the bound is exact, and to show that the bound is asymptotically sharp in the limit of large matrices.

2. A PERFORMANCE BOUND

Let $A = (a_1, \ldots, a_n) = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$ be a given $m \times n$ matrix with columns a_j. The algorithm as proposed in [1] chooses a sequence of columns $a_{j_1}, a_{j_2}, \ldots, a_{j_k}$ such that each newly chosen column is "as linearly independent" from the previous ones as possible. Here k is the exact rank of A or what is decided to be the numerical rank of A. In some version of QR-CP, the number k is determined along with the sequence $(j_1, j_2, \ldots)$; here I assume that $k \le \min(m, n)$ is given. Then the selection method proceeds as follows. All vector norms below are Euclidean norms.

0. Set $J = \emptyset$, $w_j = \|a_j\|^2$ for $j \in \{1, \ldots, n\}$.
1. For $r = 1, \ldots, k$,
   (a) find $j_r = \arg\max\{w_j : j \notin J\}$;
   (b) replace J with $J \cup \{j_r\}$;

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

(c) for $j \notin J$, replace w_j with $\|a_j - P_r a_j\|^2$, where P_r is the orthogonal projector onto the subspace spanned by $\{a_{j_1}, \ldots, a_{j_r}\}$.

Step 1(c) is equivalent to the following step.

1(c') For $j \notin J$, replace w_j with $\|a_j\|^2 - (a_j^\top a_{j_r})^2 \|a_{j_r}\|^{-2}$.

For definiteness, we assume that ties in Step 1(a) are broken by choosing the smallest j_r. It is then easy to see that the order and the outcome of the selection algorithm remain unchanged if A is replaced by QA, where Q is orthogonal. After relabeling the columns, we can therefore assume that $j_r = r$, $r = 1, \ldots, k$ and that the submatrix $(a_1, \ldots, a_k)$ is upper triangular with positive entries. Indeed, the usual implementation of QR-CP computes precisely such a permutation together with the matrix Q (in factored form) and overwrites A with QA [2]. We therefore have

$$a_{ij} = 0, \qquad (1 \leq j \leq k, \; j < i \leq m), \tag{1.a}$$

$$a_{ii}^2 \geq \sum_{l-i}^{m} a_{lj}^2, \qquad (1 \leq i \leq k, \; i < j \leq n). \tag{1.b}$$

From now on it will be assumed that $k = m < n$. If (1.a) and (1.b) hold for A, we say that A satisfies the **standard assumptions**. Let $J \subset \{1, \ldots, n\}$, we then write A_J for the submatrix of A that consists of the columns of A whose indices are in J (in their natural order). We also write $A_{[j]} = (a_1, \ldots, a_j)$.

PROPOSITION 1. *Suppose A satisfies the standard assumptions. Then for $j = 2, 3, \ldots, m$*

$$\det\left(A_{[j]}^\top A_{[j]}\right) = \max\left\{\det\left(\tilde{A}_j^\top \tilde{A}_j\right) : \tilde{A}_j = (A_{[j-1]}, a_r), \; j \leq r \leq n\right\}.$$

The proof follows from the observation that if $\tilde{A}_j = (A_{[j-1]}, a_r)$, then

$$\det\left(\tilde{A}_j^\top \tilde{A}_j\right) = \det\left(A_{[j-1]}^\top A_{[j-1]}\right) \cdot \sum_{l=j}^{m} a_{lr}^2 \leq \det\left(A_{[j-1]}^\top A_{[j-1]}\right) \cdot a_{jj}^2 = \det\left(A_{[j]}^\top A_{[j]}\right).$$

Therefore, QR-CP is the greedy algorithm for maximizing $|\det(A_J)|$ among all $J \subset \{1, \ldots, n\}$ with $|J| = m$. It is clear that a submatrix A_J with large determinant can still be poorly conditioned; this is why modifications of the column pivoting strategy above have been developed to detect near rank-deficiency.

The question now arises whether this determinant can be bounded below *a priori* in terms of the original A if J is chosen by QR-CP. It turns out that this is indeed possible, if A has orthonormal rows.

THEOREM 2. *Assume that A has orthonormal rows and that the m-element subset $J \subset \{1, \ldots, n\}$ is selected by QR-CP. Then*

$$|\det(A_J)| \geq \frac{1}{\sqrt{\binom{n}{m}}}. \tag{2}$$

PROOF. Note that QA still has orthonormal rows for any orthogonal Q and that this property also is unchanged if the columns of A are permuted. Thus, we may assume that A satisfies the standard assumptions. Since $\text{trace}(AA^\top) = m$, it follows that $\sum_{j=1}^{n} \|a_j\|^2 = m$. Since the norm of a_1 is maximal, we obtain $a_{11} \geq m/n$. If the first r columns and rows of A are deleted, the resulting $(m - r) \times (n - r)$ matrix still has orthonormal rows and satisfies the standard assumptions; thus also $a_{r+1,r+1}^2 \geq (m - r)/(n - r)$. Therefore

$$\det\left(A_J A_J^\top\right) = \det\left(A_{[m]} A_{[m]}^\top\right) = \prod_{i=l}^{m} a_{ii}^2 \geq \frac{1}{\binom{n}{m}}.$$

By the Cauchy-Binet formula [6],

$$1 = \det\left(AA^\top\right) = \sum_J \det\left(A_J A_J^\top\right)$$

where the sum extends over all $\binom{n}{m}$ subsets $J \subset \{1, \ldots, n\}$ with m elements. Thus, Theorem 2 says that QR-CP leads to a choice J that is a least as good as an average choice for the purpose of maximizing $\det(A_J A_J^\top)$. If the matrix A is arbitrary, it can be factored as $A = LB$, where B has orthonormal rows. Applying now QR-CP to B and noting that $A_J = LB_J$ and $\det(AA^\top) = \det(LL^\top)$, we obtain a selection J such that also $|\det(A_J)| \geq |\det(A)|/\sqrt{\binom{n}{m}}$; i.e., QR-CP applied to any right orthonormal factor B of A results in a selection that is at least as good as the average. It does not matter which right orthonormal factor is used since they are all related by left-multiplication with an orthogonal matrix, which does not affect the course of QR-CP. Applying QR-CP to an orthonormal factor of A and not to A itself is suggested in [2] for the purpose of subset selection.

3. SHARPNESS OF THE PERFORMANCE BOUND

Let us now ask the question whether the bound of Theorem 2 is sharp. If $m = 2$ and $n > 2$ is odd, then the following $2 \times n$ matrix A satisfies the standard assumptions and has orthonormal rows: set $a_{11} = \sqrt{2/n}$, $a_{21} = 0$, $a_{1j} = \sqrt{(n-2)/n(n-1)}$ and $a_{2j} = (-1)^j \sqrt{1/(n-1)}$ for $j = 2 \ldots, n$. Then $|\det(A_{[2]})| = \sqrt{2/n(n-1)}$. Thus the bound is attained in this case. There is essentially only one other situation where this happens.

THEOREM 3. *Let A satisfy the standard assumptions. Assume that A has orthonormal rows and that $m > 2$. Then $|\det(A_{[m]}| = 1/\sqrt{\binom{n}{m}}$ if and only if $n = m + 1$. In this case,*

$$a_{ii} = \sqrt{\frac{m-i+1}{m-i+2}}, \qquad (1 \leq i \leq m), \tag{3.a}$$

$$a_{ij} = \epsilon_{ij}\sqrt{\frac{1}{(m-i+2)(m-i+1)}}, \quad (1 \leq i \leq m, \; i < j \leq m+1), \tag{3.b}$$

with $\epsilon_{ij} \in \{-1, 1\}$.

PROOF. Suppose that $\det(A_{[m]}A_{[m]}^\top) = 1/\binom{n}{m}$. By the argument used in the proof of Theorem 2, we have for $1 \leq i \leq m$, $i < j \leq n$

$$a_{ii}^2 = \sum_{l=i}^{m} a_{lj}^2 = \frac{m-i+1}{n-i+1}.$$

Thus (3.a) is true. It follows by induction that

$$a_{ij} = \epsilon_{ij}\sqrt{\frac{n-m}{(n-i+1)(n-i)}}, \qquad (1 \leq i \leq m, \; i < j \leq n) \tag{4}$$

with $\epsilon_{ij} \in \{-1, 1\}$. Now the orthogonality of rows # 1 and # $m - 1$ of A implies that

$$\sum_{l=m}^{n} \epsilon_{1,l}\epsilon_{m-1,l}\sqrt{\frac{n-m}{n(n-1)}}\sqrt{\frac{n-m}{(n-m+2)(n-m+1)}} + \epsilon_{1,m-1}\sqrt{\frac{n-m}{n(n-1)}}\sqrt{\frac{2}{(n-m+2)}} = 0$$

and therefore

$$\sqrt{\frac{2(n-m+1)}{n-m}} \in \mathbf{Z}.$$

This is only possible if $n = m + 1$, and then (4) implies (3.b).

Suppose on the other hand that $n = m + 1$. We define a_{ij} by (3.a),(3b.), with

$$\epsilon_{i,m+1} = 1, \quad (1 \le i \le m) \quad \text{and} \quad \epsilon_{ij} = -1, \quad (i < j < m - 1, \ 1 \le i \le m).$$

Then A satisfies the standard assumptions, and direct computations show that A has orthonormal rows and that $\det(A_{[m]}A_{[m]}^{\top}) = 1/(m+1) = 1/\binom{n}{m}$.

Any matrix A that satisfies the assumptions of this theorem has $m + 1$ columns of equal norm $\|a_j\| = (m+1)^{-1/2}$. Deleting any of these columns results in a square matrix with determinant $(m+1)^{-1/2}$, and it is not hard to see that all these matrices have the same set of singular values $\{1, \ldots, 1, (m+1)^{-1/2}\}$. Thus all m-element subsets $J \subset \{1, \ldots, n\}$ are essentially equally good selections, and this is why QR-CP must have an "average" performance. On the other hand, for the $2 \times n$ matrices of the beginning of the section, it is easy to see that for $n > 3$, any selection $J = \{j, k\}$ with $j, k \ge 2$ and $j + k$ odd results in $|\det(A_J)| = 2\sqrt{(n-2)/n(n-1)^2} > \sqrt{2/n(n-1)} = |\det(A_{[2]})|$.

4. ASYMPTOTIC SHARPNESS

We now look at the question whether there are situations in which QR-CP picks an "average" choice, although better choices are available. The class of $2 \times n$ matrices from the last section furnishes such an example. The following construction results in a wider class of matrices for which this happens.

Let $m > 1$ be given. Choose a number $L \ge m$ such that an $L \times L$ Hadamard matrix H_L exists, that is a matrix with all entries from $\{-1, 1\}$ such that $L^{-1}H_L H_L^{\top} = I$, the identity matrix. This is always possible; e.g., one can choose any $L = 2^l \ge m$. For $r \ge 1$, set $N = N_r = m - 1 + rL$. Define numbers $c_i, d_i > 0$ for $1 \le i \le m$ by

$$d_i^2 = \frac{(N - m + 2)^{i-1} - (N - m)(N - m + 1)^{i-2}}{(N - m + 2)^{i-1}},$$

$$c_i^2 = \frac{1}{N - m + 1}\left(1 - d_i^2\right) = \frac{(N - m)(N - m + 1)^{i-3}}{(N - m + 2)^{i-1}}, \qquad (i \ge 2), \quad c_1 = 1.$$

Then define the $m \times N$ matrix C with columns c_j and entries c_{ij} by

$$c_{ij} = \begin{cases} d_{m-i+1}\delta_{ij}, & (1 \le i \le m, \ 1 \le j \le m - 1), \\ c_{m-i+1}h_{ik}, & (1 \le i \le m, \ m \le j \le N) \text{ where } k = j - m - 1 \mod L. \end{cases}$$

Here δ_{ij} is Kronecker's delta, and the entries of H_L are h_{ik}. Thus the first $m - 1$ columns of C are a diagonal matrix with diagonal entries d_{m-i+1}, and the last rL columns of C are r identical blocks of the form $\operatorname{diag}(c_m, c_{m-1}, \ldots, c_1)\tilde{H}_L$, where $\tilde{H}_L$ consists of the first m rows of H_L.

THEOREM 4. *The matrix C satisfies the standard assumptions, and $CC^{\top} = I$. As $N \to \infty$,*

$$\sqrt{N}d_i \to \sqrt{i}, \quad \sqrt{N}c_i \to 1,$$

$$\binom{N}{m}\det\left(C_{[m]}C_{[m]}^{\top}\right) \to 1. \tag{5}$$

PROOF. The assertions that C satisfies the standard assumptions and has orthonormal rows follow by direct calculations from the construction. In particular, $c_{m-k+1,m-k+1}^2 = d_k^2 > \sum_{i=1}^{k} c_i^2 = \sum_{i=m-k+1}^{m} c_{il}^2$ for $k \ge 2$. The limiting behavior of d_i and c_i follows by an application of l'Hopital's rule. The assertion about the determinant of $C_{[m]}$ is implied by

$$\binom{N}{m}\det\left(C_{[m]}C_{[m]}^{\top}\right) = \prod_{i=l}^{m}\left(d_i^2 \cdot \frac{N-i}{i}\right) \to 1.$$

For the matrix C, the choice $J = \{1, \ldots, m\}$ usually does *not* produce a submatrix with maximal determinant. This is clear if $m = L$. In this case, the choice $\tilde{J} = \{N - m + 1, \ldots, N\}$ results in $C_{\tilde{J}} = \operatorname{diag}(c_m, c_{m-1}, \ldots, c_1) \cdot H_L$. Thus $\sqrt{N} C_{\tilde{J}} \to H_L$ and $N^{m/2} \det(C_{\tilde{J}}) \to m^{m/2}$, but $N^{m/2} \det(C_{[m]}) \to \sqrt{m}!$. Also in this case, the condition number of $C_{\tilde{J}}$ converges to the condition number of H_L, which is 1, and all singular values of $N^{1/2} C_{\tilde{J}}$ are approximately $\sqrt{m}$. On the other hand, the largest and smallest singular values of $C_\infty = \lim_{N \to \infty} N^{1/2} C_{[m]}$ turn out to be $O(\sqrt{m})$ and $O((\log m)^{-1/2})$. Thus the condition number of $C_{[m]}$ is approximately $\sqrt{m \log m}$. QR-factorization with column pivoting results in a poor choice for such matrices.

REFERENCES

1. P.A. Businger and G.H. Golub, Linear least squares solutions by householder transformation, *Numer. Math.* **7**, 269–276, (1965).

2. G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd edition, Johns Hopkins University Press, Baltimore, MD, (1989).

3. T.F. Chan, Rank revealing QR factorizations, *Linear Algebra Appl.* **88/89**, 67–82, (1987).

4. L.V. Foster, Rank and null space calculations using matrix decomposition without column interchanges, *Linear Algebra Appl.* **74**, 47–71, (1986).

5. A Miller, *Subset Selection in Regression*, Chapman & Hall, New York, (1990).

6. P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd edition, Academic Press, San Diego, CA, (1985).