

What Is a Rank-Revealing Factorization?

Nicholas J. Higham*

May 19, 2021

In many applications a matrix $A \in \mathbb{R}^{m \times n}$ has less than full rank, that is, $r = \text{rank}(A) < \min(m, n)$. Sometimes, r is known, and a full-rank factorization $A = GH$ with $G \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{r \times n}$, both of rank r , is given—especially when $r = 1$ or $r = 2$. Often, though, the rank r is not known. Moreover, rather than being of exact rank r , A is merely close to a rank r matrix because of errors from various possible sources.

What is usually wanted is a factorization that displays how close A is to having particular ranks and provides an approximation to the range space of a lower rank matrix. The ultimate tool for providing this information is the singular value decomposition (SVD)

$$A = U\Sigma V^T, \quad \Sigma = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n},$$

where $p = \min(m, n)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$, and $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal. The Eckart–Young theorem says that

$$\min_{\text{rank}(B)=k} \|A - B\|_q = \begin{cases} \sigma_{k+1}, & q = 2, \\ \left(\sum_{i=k+1}^r \sigma_i^2 \right)^{1/2}, & q = F, \end{cases}$$

and that the minimum is attained at

$$A_k = U\Sigma_k V^T, \quad \Sigma_k = \text{diag}(\sigma_1, \dots, \sigma_k, 0, \dots, 0),$$

so A_k is the best rank- k approximation to A in both the 2-norm and the Frobenius norm.

Although the SVD is expensive to compute, it may not be significantly more expensive than alternative factorizations. However, the SVD is expensive to update when a row or column is added to or removed from the matrix, as happens repeatedly in signal processing applications.

Many different definitions of a rank-revealing factorization have been given, and they usually depend on a particular matrix factorization. We will use the following general definition.

Definition 1. *A rank-revealing factorization (RRF) of $A \in \mathbb{R}^{m \times n}$ is a factorization*

$$A = XDY^T, \quad X \in \mathbb{R}^{m \times p}, \quad D \in \mathbb{R}^{p \times p}, \quad Y \in \mathbb{R}^{n \times p},$$

where $p \leq \min(m, n)$, D is diagonal and nonsingular, and X and Y are well conditioned.

*Department of Mathematics, University of Manchester, Manchester, M13 9PL, UK (nick.higham@manchester.ac.uk).

An RRF concentrates the rank deficiency and ill condition of A into the diagonal matrix D . An RRF clearly exists, because the SVD is one, with X and Y having orthonormal columns and hence being perfectly conditioned. Justification for this definition comes from a version of Ostrowski's theorem, which shows that

$$\sigma_i(A) = \theta_i \sigma_i(D), \quad i = 1: \min(m, n), \quad (1)$$

where $\sigma_p(X)\sigma_p(Y) \leq \theta_i \leq \sigma_1(X)\sigma_1(Y)$. Hence as long as X and Y are well conditioned, the singular values are good order of magnitude approximations to those of A up a scale factor.

Without loss of generality we can assume that

$$D = \text{diag}(d_i), \quad |d_1| \geq |d_2| \geq \dots \geq |d_p|$$

(since $XDY^T = XP \cdot P^T DP \cdot P^T Y^T$ for any permutation matrix P and the second expression is another RRF). For $\tilde{A}_k = X \text{diag}(d_1, \dots, d_k, 0, \dots, 0) Y^T$ we have

$$\|A - \tilde{A}_k\| \leq \|X\| \|Y\| \|\text{diag}(0, \dots, 0, d_{k+1}, \dots, d_p)\|,$$

so A is within distance of order $|d_{k+1}|$ from the rank- k matrix \tilde{A}_k , which is the same order as the distance to the nearest rank- k matrix if $|d_{k+1}| \approx \sigma_{k+1}$.

Definition 2 is a strong requirement, since it requires all the singular values of A to be well approximated by the (scaled) diagonal elements of D . We will investigate below how it compares with another definition of RRF.

Numerical Rank

An RRF helps to determine the numerical rank, which we now define.

Definition 2. *For a given $\epsilon > 0$ the numerical rank of A is the largest integer k such that $\sigma_k > \epsilon$.*

By the Eckart–Young theorem, the numerical rank is the smallest rank attained over all $A + E$ with $\|E\|_2 \leq \epsilon$. For the numerical rank to be meaningful in the sense that it is unchanged if ϵ is perturbed slightly, we need ϵ not to be too close to σ_k or σ_{k+1} , which means that there must be a significant gap between these two singular values.

QR Factorization

One might attempt to compute an RRF by using a QR factorization $A = QR$, where $Q \in \mathbb{R}^{m \times n}$ has orthonormal columns, $R \in \mathbb{R}^{n \times n}$ is upper triangular, and we assume that $m \geq n$. In Definition 1, we can take

$$X = I, \quad D = \text{diag}(R), \quad Y^T = D^{-1}R. \quad (*)$$

However, it is easy to see that QR factorization in its basic form is flawed as a means for computing an RRF. Consider the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

which is a Jordan block with zero eigenvalue. This matrix is its own QR factorization ($R = A$), and the prescription (*) gives $D = 0$, so $A \neq XDY^T$. The essential problem is that the diagonal of R has no connection with the nonzero singular values of A . What is needed are column permutations: $A\Pi = \text{diag}(1, 1, 1, 0)$ for the permutation matrix Π that reorders $[a_1, a_2, a_3, a_4]$ to $[a_2, a_3, a_4, a_1]$, and this is a perfect RRF with $X = Y = I$.

For a less trivial example, consider the matrix

$$A = \begin{bmatrix} 1 & 1 & \theta & 0 \\ 1 & -1 & 2 & 1 \\ 1 & 0 & 1 + \theta & -1 \\ 1 & -1 & 2 & -1 \end{bmatrix}, \quad \theta = 10^{-8}. \quad (\dagger)$$

Computing the QR factorization we obtain

$$R = \begin{bmatrix} -2.0000\text{e}+00 & 5.0000\text{e}-01 & -2.5000\text{e}+00 & 5.0000\text{e}-01 \\ 0 & 1.6583\text{e}+00 & -1.6583\text{e}+00 & -1.5076\text{e}-01 \\ 0 & 0 & -4.2640\text{e}-09 & 8.5280\text{e}-01 \\ 0 & 0 & 0 & -1.4142\text{e}+00 \end{bmatrix}$$

The (3, 3) element tells us that A is within distance about 4×10^{-9} of being rank deficient and so has a singular value bounded above by this quantity, but it does not provide any information about the next larger singular value. Moreover, in (*), $\kappa_2(Y)$ is of order 10^{16} for this factorization. We need any small diagonal elements to be in the bottom right-hand corner, and to achieve this we need to introduce column permutations to move the “dependent columns” to the end.

QR Factorization With Column Pivoting

A common method for computing an RRF is QR factorization with column pivoting, which for a matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ computes a factorization $A\Pi = QR$, where Π is a permutation matrix, $Q \in \mathbb{R}^{m \times n}$ has orthonormal columns, and $R \in \mathbb{R}^{n \times n}$ is upper triangular and satisfies the inequalities

$$|r_{kk}|^2 \geq \sum_{i=k}^j |r_{ij}|^2, \quad j = k+1 : n, \quad k = 1 : n. \quad (2)$$

In particular,

$$|r_{11}| \geq |r_{22}| \geq \cdots \geq |r_{nn}|. \quad (3)$$

If $|r_{kk}| \geq \epsilon \geq |r_{k+1,k+1}|$ with $\epsilon > 0$ then we can write

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} \begin{matrix} k & n-k \\ n-k & \end{matrix}, \quad (4)$$

with

$$\|R_{22}\|_2 \leq \|R_{22}\|_F \leq 2^{-1/2}(n - k + 1)\epsilon.$$

Hence R is within 2-norm distance $2^{-1/2}(n-k+1)\epsilon$ of the rank- k matrix $\begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix}$. Note that if $Q = [Q_1 \ Q_2]$ is partitioned conformally with Q in (4) then

$$A\Pi = [Q_1 \ Q_2] \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} = Q_1 [R_{11} \ R_{12}] + [0 \ Q_2 R_{22}],$$

so $\|A\Pi - Q_1[R_{11} \ R_{12}]\|_2 \leq \|R_{22}\|_2$, which means that Q_1 provides an $O(\epsilon)$ approximation to the range of A .

To assess how good an RRF this factorization is (with $p = n$) we write it as

$$A = QR\Pi^T = QDY^T, \quad D = \text{diag}(r_{ii}), \quad Y^T = D^{-1}R\Pi^T. \quad (\#)$$

Applying (1) gives

$$\sigma_i(A) = \theta_i \sigma_i(D), \quad i = 1:p, \quad (5)$$

where $\sigma_n(Y) \leq \theta_i \leq \sigma_1(Y)$, since Q has orthonormal columns and so has unit singular values. Now $D^{-1}R$ has unit diagonal and, in view of (2), its off-diagonal elements are bounded by 1. Therefore $\sigma_1(Y) = \|Y\|_2 \leq \|Y\|_F \leq (n(n+1)/2)^{1/2}$. On the other hand, $\sigma_p(Y)^{-1} \leq 2^{n-1}$ by Theorem 1 in Bounds for the Norm of the Inverse of a Triangular Matrix. Therefore

$$2^{1-n} \leq \theta_i \leq (n(n+1)/2)^{1/2}.$$

The lower bound is an approximate equality for small τ for the triangular matrix

$$R_n(\theta) = \text{diag}(1, s, \dots, s^{n-1}) \begin{bmatrix} 1 & -c & -c & \dots & -c \\ & 1 & -c & \dots & -c \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & -c \\ & & & & 1 \end{bmatrix}, \quad c = \cos \tau, \quad s = \sin \tau,$$

devised by Kahan, which is invariant under QR factorization with column pivoting. Therefore QR factorization with column pivoting is not guaranteed to reveal the rank, and indeed it can fail to do so by an exponentially large factor.

For the matrix (\dagger), QR with column pivoting reorders A to $A\Pi = [a_3, a_4, a_2, a_1]$ and yields

$$R = \begin{bmatrix} -3.0000\text{e}+00 & 3.3333\text{e}-01 & 1.3333\text{e}+00 & -1.6667\text{e}+00 \\ 0 & -1.6997\text{e}+00 & 2.6149\text{e}-01 & 2.6149\text{e}-01 \\ 0 & 0 & 1.0742\text{e}+00 & 1.0742\text{e}+00 \\ 0 & 0 & 0 & 3.6515\text{e}-09 \end{bmatrix}$$

This R suggests a numerical rank of 3 for $\epsilon = 10^{-8}$ (say). In fact, this factorization provides a very good RRF, as in ($\#$) we have $\kappa_2(Y) \approx 3.4$.

QR Factorization with Other Pivoting Choices

Consider a QR factorization $A\Pi = QR$ with triangular factor partitioned as

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} \begin{matrix} k & n-k \\ k & n-k \end{matrix}. \quad (6)$$

We have

$$\sigma_{\min}(R_{11}) \leq \sigma_k(A), \quad (7)$$

$$\sigma_{\max}(R_{22}) \geq \sigma_{k+1}(A), \quad (8)$$

where (7) is from singular value interlacing inequalities and (8) follows from the Eckart-Young theorem, since setting R_{22} to zero gives a rank- k matrix. Suppose A has numerical rank k and $\sigma_{k+1} \ll \sigma_k$. We would like to be able to detect this situation from R , so clearly we need

$$\sigma_{\min}(R_{11}) \approx \sigma_k(A), \quad \sigma_{\max}(R_{22}) \approx \sigma_{k+1}(A). \quad (9)$$

In view of the inequalities (7) and (8) this means that we wish to choose Π maximize $\sigma_{\min}(R_{11})$ and minimize $\sigma_{\max}(R_{22})$.

Some theoretical results are available on the existence of such QR factorizations. First, we give a result that shows that for $k = n - 1$ the approximations in (9) can hold to within a factor $n^{1/2}$.

Theorem 1. *For $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ there exists a permutation Π such that A has the QR factorization $A\Pi = QR$ with $|r_{nn}| \leq n^{1/2}\sigma_n(A)$ and $\sigma_{\min}(R_{11}) \geq n^{-1/2}\sigma_{n-1}(A)$, where $R_{11} = R(1:n-1, 1:n-1)$.*

Proof. Let $Av = \sigma_n u$, with $\|v\|_2 = \|u\|_2 = 1$ and let Π^T be such that $\tilde{v} = \Pi^T v$ satisfies $|\tilde{v}_n| = \|\tilde{v}\|_\infty$. Then if $A\Pi = QR$ is a QR factorization,

$$\sigma_n = \|\sigma_n u\|_2 = \|Av\|_2 = \|QR\Pi^T v\|_2 = \|R\tilde{v}\|_2 \geq |r_{nn}\tilde{v}_n| \geq n^{-1/2}|r_{nn}|,$$

since $\|\tilde{v}\|_2 = 1$, which yields the result.

Next, we write $\Pi = [\Pi_1 \ \pi]$, where $\pi \in \mathbb{R}^n$, and partition

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}$$

with $R_{11} \in \mathbb{R}^{(n-1) \times (n-1)}$. Then

$$A\Pi_1 = Q \begin{bmatrix} R_{11} \\ 0 \end{bmatrix}$$

implies $\sigma_{\min}(A\Pi_1) = \sigma_{\min}(R_{11})$. On the other hand, if $A = U\Sigma V^T$ is an SVD with $U \in \mathbb{R}^{m \times n}$, $\Sigma = \text{diag}(D_1, \sigma_n) \in \mathbb{R}^{n \times n}$, and $V = [V_1 \ v]$ then

$$A\Pi_1 = U\Sigma V^T \Pi_1 = U \begin{bmatrix} D_1 & 0 \\ 0 & \sigma_n \end{bmatrix} \begin{bmatrix} V_1^T \\ v^T \end{bmatrix} \Pi_1 = U \begin{bmatrix} D_1 V_1^T \\ \sigma_n v^T \end{bmatrix} \Pi_1,$$

so

$$\sigma_{\min}(A\Pi_1) = \sigma_{\min} \left(\begin{bmatrix} D_1 V_1^T \\ \sigma_n v^T \end{bmatrix} \Pi_1 \right) \geq \sigma_{\min}(D_1 V^T \Pi_1) \geq \sigma_{n-1} \sigma_{\min}(V^T \Pi_1).$$

Finally, we note that we can partition the orthogonal matrix $V^T \Pi_1$ as

$$V^T \Pi_1 = \begin{bmatrix} V_1^T \Pi_1 & V_1^T \pi \\ v^T \Pi_1 & v^T \pi \end{bmatrix},$$

and the CS decomposition implies that

$$\sigma_{\min}(V_1^T \Pi_1) = \sigma_{\min}(v^T \pi) = |v^T \pi| = |\tilde{v}_n| \geq n^{-1/2}.$$

Hence $\sigma_{\min}(R_{11}) \geq n^{-1/2}\sigma_{n-1}$, as required. \square

Theorem 1 is a special case of the next result of Hong and Pan (1992).

Theorem 2. For $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and any k there exists a permutation matrix Π such that A has the QR factorization $A\Pi = QR$ where, with R partitioned as in (6),

$$\sigma_{\max}(R_{22}) \leq f(k, n)\sigma_{k+1}(A), \quad \sigma_{\min}(R_{11}) \geq f(k, n)^{-1}\sigma_k(A),$$

where $f(k, n) = (k(n - k) + \min(k, n - k))^{1/2}$.

The proof of Theorem 2 is constructive and chooses Π to move a submatrix of maximal determinant of V_2 to the bottom of V_2 , where V_2 comprises the last $n - k$ columns of the matrix of right singular vectors.

Theorem 2 shows the existence of an RRF up to the factor $f(k, n) \leq (n + 1)/2$, but it does not provide an efficient algorithm for computing one.

Much work has been done on algorithms that choose the permutation matrix Π in a different way to column pivoting or post-process a QR factorization with column pivoting, with the aim of satisfying (9) at reasonable cost. Typically, these algorithms involve estimating singular values and singular vectors. We are not aware of any algorithm that is guaranteed to satisfy (9) and requires only $O(n^3)$ flops.

UTV Decomposition

By applying Householder transformations on the right, a QR factorization with column pivoting can be turned into a *complete orthogonal decomposition* of $A \in \mathbb{R}^{m \times n}$, which has the form

$$A = U \begin{bmatrix} T & 0 \\ 0 & 0 \end{bmatrix} V^T, \quad (10)$$

where $T \in \mathbb{R}^{r \times r}$ is upper triangular and $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal. Stewart (1998) calls (6) with T upper triangular or lower triangular a UTV decomposition and he defines a rank-revealing UTV decomposition of numerical rank r by

$$A = U \begin{bmatrix} T & F \\ 0 & G \end{bmatrix} V^T, \quad T \in \mathbb{R}^{r \times r},$$

$$\sigma_r(T) \approx \sigma_r(A), \quad \|F\|_F^2 + \|G\|_F^2 \approx \sigma_{r+1}^2 + \cdots + \sigma_n^2.$$

The UTV decomposition is easy to update (when a row is added) and downdate (when a row is removed) using Givens rotations and it is suitable for parallel implementation. Initial determination of the UTV decomposition can be done by applying the updating algorithm as the rows are brought in one at a time.

LU Factorization

Instead of QR factorization we can build an RRF from an LU factorization with pivoting. For $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, let

$$\Pi_1 A \Pi_2 = LU = \begin{bmatrix} L_{11} & 0 \\ L_{12} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix},$$

where Π_1 and Π_2 are permutation matrices, L and U are $m \times n$ lower and $n \times n$ upper triangular, respectively, and L_{11} and U_{11} are $k \times k$. Analogously to (7) and (8), we always have $\sigma_{\min}(L_{11}U_{11}) \leq \sigma_k(A)$ and $\sigma_{\max}(L_{22}U_{22}) \geq \sigma_{k+1}(A)$. With a suitable pivoting strategy we can hope that $\sigma_{\min}(L_{11}U_{11}) \approx \sigma_k(A)$ and $\sigma_{\max}(L_{22}U_{22}) \approx \sigma_{k+1}(A)$.

A result of Pan (2000) shows that an RRF based on LU factorization always exists up to a modest factor $f(k, n)$. This is analogue for LU factorization of Theorem 2.

Theorem 3 *For $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and any k there exist permutation matrices Π_1 and Π_2 such that*

$$\Pi_1 A \Pi_2 = LU = \begin{bmatrix} L_{11} & 0 \\ L_{12} & I_{m-k, n-k} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix} \begin{matrix} k & n-k \\ k & n-k \end{matrix},$$

where L_{11} is unit lower triangular, U_{11} is upper triangular, and

$$\sigma_{\max}(U_{22}) \leq f(k, n)\sigma_{k+1}(A), \quad \sigma_{\min}(L_{11}U_{11}) \geq f(k, n)^{-1}\sigma_k(A),$$

where $f(k, n) = k(n - k) + 1$.

Again the proof is constructive, but the permutations it chooses are too expensive to compute. In practice, complete pivoting often yields a good RRF.

In terms of Definition 1, an RRF has

$$X = \Pi_1^T L D, \quad D = \text{diag}(u_{ii}), \quad Y^T = D^{-1} U \Pi_2. \quad (\ddagger)$$

For the matrix (\ddagger) , the U factor for LU factorization without pivoting is

$$U = \begin{bmatrix} 1.0000\text{e}+00 & 1.0000\text{e}+00 & 1.0000\text{e}-08 & 0 \\ 0 & -2.0000\text{e}+00 & 2.0000\text{e}+00 & 1.0000\text{e}+00 \\ 0 & 0 & 5.0000\text{e}-09 & -1.5000\text{e}+00 \\ 0 & 0 & 0 & -2.0000\text{e}+00 \end{bmatrix}$$

As for QR factorization without pivoting, an RRF is not obtained from (\ddagger) . However, with complete pivoting we obtain

$$U = \begin{bmatrix} 2.0000\text{e}+00 & 1.0000\text{e}+00 & -1.0000\text{e}+00 & 1.0000\text{e}+00 \\ 0 & -2.0000\text{e}+00 & 0 & 0 \\ 0 & 0 & 1.0000\text{e}+00 & 1.0000\text{e}+00 \\ 0 & 0 & 0 & -5.0000\text{e}-09 \end{bmatrix}$$

which yields a very good RRF (\ddagger) with $\kappa_2(X) = 3.5$ and $\kappa_2(Y) = 3.4$.

Notes

QR factorization with column pivoting is difficult to implement efficiently, as the criterion for choosing the pivots requires the norms of the active parts of the remaining columns and this requires a significant amount of data movement. In recent years, randomized RRF algorithms have been developed that use projections with random matrices to make pivot decisions based on small sample matrices and thereby reduce the amount of data movement. See, for example, Martinsson et al. (2019).

References

This is a minimal set of references, which contain further useful references within.

- Shivkumar Chandrasekaran and Ilse C. F. Ipsen, On Rank-Revealing Factorisations, SIAM J. Matrix Anal. Appl. 15 (2), 592–622, 1994
- Y. P. Hong and C.-T. Pan, Rank-Revealing QR Factorizations and the Singular Value Decomposition, Math. Comp. 58 (197), 213–232, 1992.
- P. G. Martinsson, G. Quintana-Orti, and N. Heavner, randUTV: A Blocked Randomized Algorithm for Computing a Rank-Revealing UTV Factorization, ACM Trans. Math. Software 45(1), 4:1–4:26, 2019.
- C.-T. Pan, On the Existence and Computation of Rank-Revealing LU Factorizations, Linear Algebra Appl. 316, 199–222, 2000.
- G. W. Stewart, Matrix Algorithms. Volume I: Basic Decompositions, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998.

Related Blog Posts

- What Is the CS Decomposition? (2020)
- What Is an LU Factorization? (2021)
- What Is a QR Factorization? (2020)
- What Is the Singular Value Decomposition? (2020)

This article is part of the “What Is” series, available from <https://nhigham.com/category/what-is> and in PDF form from the GitHub repository <https://github.com/nhigham/what-is>.