UMIACS-TR-91-79 May 1991
CS-TR 2685

Error Analysis of QR Updating with Exponential Windowing

G. W. Stewart*

ABSTRACT

Exponential windowing is a widely used technique for suppressing the effects of old data as new data is added to a matrix. Specifically, given an $n \times p$ matrix X_n and a "forgetting factor" $\beta \in (0, 1)$, one works with the matrix $\mathrm{diag}(\beta^{n-1}, \beta^{n-2}, \ldots, 1)X_n$. In this paper we examine an updating algorithm for computing the QR factorization of $\mathrm{diag}(\beta^{n-1}, \beta^{n-2}, \ldots, 1)X_n$ and show that it is unconditionally stable in the presence of rounding errors.

*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-87-0188.

Error Analysis of QR Updating with Exponential Windowing

G. W. Stewart*

1. Introduction

In many application (e.g., signal processing, time series) one needs the QR factorization of an $n \times p$ matrix

$$X_n = \begin{pmatrix} x_1^{\mathrm{H}} \\ x_2^{\mathrm{H}} \\ \vdots \\ x_n^{\mathrm{H}} \end{pmatrix}.$$

The rows of X_n represent data that arrives at regular intervals, with x_1^{H} the oldest data and x_n^{H} the most recent.

If the series x_n^{H} is not stationary, it is necessary to suppress the older data so that they do not contaminate more recent information. One widely used method for accomplishing this is called *exponential windowing.* Let $\beta \in (0,1)$ be a "forgetting factor," and let

$$D_n = \mathrm{diag}(\beta^{n-1}, \beta^{n-2}, \ldots, 1).$$

Instead of computing the QR factorization of X_n, one computes the QR factorization of $D_n X_n$; i.e., one computes

$$D_n X_n = Q_n R_n, \tag{1.1}$$

where Q_n has orthonormal columns and R_n is upper triangular. The effect of exponential windowing is to weight x_i^{H} by β^{n-i}, so that it has less and less influence as n increases.

*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-87-0188.

As a rule, only the R-factor R_n in (1.1) is needed in applications. It can be computed efficiently by the following updating procedure. Let $R_0 = 0$. Given R_n, compute the QR decomposition:

$$\begin{pmatrix} U_n & v_n \\ w_n^{\mathrm{H}} & \xi_n \end{pmatrix} \begin{pmatrix} \beta R_n \\ x_{n+1}^{\mathrm{H}} \end{pmatrix} = \begin{pmatrix} R_{n+1} \\ 0 \end{pmatrix}. \tag{1.2}$$

Here

$$\begin{pmatrix} U_n & v_n \\ w_n^{\mathrm{H}} & \xi_n \end{pmatrix} \tag{1.3}$$

is unitary and R_{n+1} is upper triangular. Is is easily seen that the sequence of triangular matrices R_n so generated are the R-factors of the matrices X_n. The details of this updating algorithm may be found in [1]. It requires $O(p^2)$ arithmetic operations.

Exponential windowing and updating allows us to look at the local behavior of an arbitrarily long sequence of data. However, the fact that n is effectively unbounded raises the possibility that rounding error will accumulate to the point where it overwhelms the data. The purpose of this paper is to show that this does not happen: exponential windowing damps old rounding errors along with old data.

In the next section we will present the rounding error analysis. Although the results of this analysis are sufficient for practical purposes, it is clear that the bounds are an overestimate, at least asymptotically. Consequently, §3 is devoted to producing refined bounds.

Throughout this paper $\|X\|$ will denote the Frobenius norm defined by

$$\|A\|^2 = \sum_{i,j} |a_{ij}|^2.$$

All computations will be assumed to be in floating point arithmetic with rounding unit ϵ_{M}; i.e., $-\log \epsilon_{\mathrm{M}}$ is approximately the number of decimal digits carried in the computation.

2. The Error Analysis

Our error analysis will be a classical backward error analysis; that is, we will show that the computed R_n, whatever its accuracy, comes from very slightly perturbed data. The analysis begins with an a backward error analysis of the single update step (1.2).

Theorem 2.1. *Let R_{n+1} denote the result of performing the update (1.2) in floating-point arithmetic with rounding unit ϵ_{M}. Then there is a unitary matrix of the form (1.3), a constant K depending on p, a matrix G and a vector h^{H} satisfying*

$$\left\| \begin{pmatrix} G_n \\ h_n^{\mathrm{H}} \end{pmatrix} \right\| \le K\epsilon_{\mathrm{M}} \left\| \begin{pmatrix} \beta R_n \\ x_{n+1}^{\mathrm{H}} \end{pmatrix} \right\| \tag{2.1}$$

such that

$$\begin{pmatrix} U_n & v_n \\ w_n^{\mathrm{H}} & \xi_n \end{pmatrix} \begin{pmatrix} \beta R_n + G_n \\ x_{n+1} + h_{n+1}^{\mathrm{H}} \end{pmatrix} = \begin{pmatrix} R_{n+1} \\ 0 \end{pmatrix}. \tag{2.2}$$

A proof of this theorem may be found in [4, Ch.3, §§20–24].

The analysis of the updating algorithm with exponential windowing amounts to the recursive application of the bound (2.2). As is typical in backward rounding-error analyses, we let let quantities stand for their *computed* values. The results will be cast in terms of the augmented factorization

$$\begin{pmatrix} 0 \\ D_n X_n \end{pmatrix} = \begin{pmatrix} 0 \\ Q_n \end{pmatrix} R_n.$$

This factorization reflects the actual updating process in which we start with a zero matrix, imagined to lie above X_n, and form R_n in it.

Theorem 2.2. *Let*

$$\|R_n\| \le \rho, \qquad n = 1, 2, \ldots,$$

so that ρ is an upper bound for the norms of the computed R_n. Then there is a matrix

$$\begin{pmatrix} P_n \\ Q_n \end{pmatrix} \tag{2.3}$$

with orthonormal columns and matrices E_n and F_n satisfying

$$\left\| \begin{pmatrix} E_n \\ F_n \end{pmatrix} \right\| \le \frac{K\epsilon_{\mathrm{M}}\rho}{(1-\beta)(1-K\epsilon_{\mathrm{M}})} \tag{2.4}$$

such that

$$\begin{pmatrix} E_n \\ D_n X_n + F_n \end{pmatrix} = \begin{pmatrix} P_n \\ Q_n \end{pmatrix} R_n.$$

Proof. The proof is by induction. The theorem is clearly true for $n = 0$ (take $P_0 = I$ and $R_k = 0$).

Now suppose that the theorem is true for some $n \geq 0$, and suppose that R_n has been updated so that (2.2) holds. Then from (2.1) and (2.2) we have that

$$\left\| \begin{pmatrix} \beta R_n \\ x_n^{\mathrm{H}} \end{pmatrix} \right\| \leq \|R_{n+1}\| + \left\| \begin{pmatrix} G_n \\ h_n^{\mathrm{H}} \end{pmatrix} \right\| \leq \rho + K\epsilon_{\mathrm{M}} \left\| \begin{pmatrix} \beta R_n \\ x_n^{\mathrm{H}} \end{pmatrix} \right\|.$$

Hence

$$\left\| \begin{pmatrix} \beta R_n \\ x_n^{\mathrm{H}} \end{pmatrix} \right\| \leq \frac{\rho}{1 - K\epsilon_{\mathrm{M}}}. \tag{2.5}$$

Now consider the equations

$$\begin{aligned}
\begin{pmatrix} \beta E_n + P_n G_n \\ \beta(D_n X_n + F_n) + Q_n G_n \\ x_{n+1} + h_n^{\mathrm{H}} \end{pmatrix} &= \begin{pmatrix} P_n & 0 \\ Q_n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta R_n + G_n \\ x_{n+1} + h_n \end{pmatrix} \\
&= \begin{pmatrix} P_n & 0 \\ Q_n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U_n^{\mathrm{H}} & w_n \\ v_n^{\mathrm{H}} & \bar{\xi}_n \end{pmatrix} \begin{pmatrix} U_n & v_n \\ w_n^{\mathrm{H}} & \xi_n \end{pmatrix} \begin{pmatrix} \beta R_n + G_n \\ x_{n+1} + h_n \end{pmatrix} \\
&= \begin{pmatrix} P_n U_n^{\mathrm{H}} & P_n w_n \\ Q_n U_n^{\mathrm{H}} & Q_n w_n \\ v_n^{\mathrm{H}} & \bar{\xi}_n \end{pmatrix} \begin{pmatrix} R_{n+1} \\ 0 \end{pmatrix}
\end{aligned}$$

They suggest that we should take

$$\begin{aligned}
P_{n+1} &= P_n U_n^{\mathrm{H}}, \\
Q_{n+1} &= \begin{pmatrix} Q_n U_n^{\mathrm{H}} \\ v_n^{\mathrm{H}} \end{pmatrix}, \\
E_{n+1} &= \beta E_n + P_n G_n, \\
F_{n+1} &= \begin{pmatrix} \beta F_n + Q_n G_n \\ h_n^{\mathrm{H}} \end{pmatrix}.
\end{aligned} \tag{2.6}$$

In fact all we need do is verify that E_{n+1} and F_{n+1} so defined satisfy the bound (2.4). From (2.6) and the fact that (2.3) has orthonormal columns, we have

$$\left\| \begin{pmatrix} E_{n+1} \\ F_{n+1} \end{pmatrix} \right\| \leq \beta \left\| \begin{pmatrix} E_n \\ F_n \end{pmatrix} \right\| + \left\| \begin{pmatrix} G_n \\ h_n^{\mathrm{H}} \end{pmatrix} \right\|.$$

Hence from the induction hypothesis (2.4) and from (2.1) and (2.5)

$$\left\| \begin{pmatrix} E_{n+1} \\ F_{n+1} \end{pmatrix} \right\| \leq \frac{\beta K \epsilon_{\mathrm{M}} \rho}{(1-\beta)(1-K\epsilon_{\mathrm{M}})} + \frac{K\epsilon_{\mathrm{M}}\rho}{1-K\epsilon} \leq \frac{K\epsilon_{\mathrm{M}}\rho}{(1-\beta)(1-K\epsilon_{\mathrm{M}})},$$

which establishes the theorem.

One unsatisfactory aspect of this theorem is that it is phrased in terms of an upper-bound on the computed R-factor. This leaves open the possibility that R_n could grow unboundedly, even though the true factors remain bounded. The following corollary shows that this cannot happen.

Corollary 2.3. *Let*

$$\eta = \frac{K\epsilon_{\mathrm{M}}}{(1-\beta)(1-K\epsilon_{\mathrm{M}})}$$

and let $\hat{\rho}$ be an upper bound on the norms of the true R-factors. Then

$$\rho \leq \frac{\hat{\rho}}{1-\eta}.$$

Proof. Since R_n is obtained from $D_n X_n$ by a unitary transformation, $\|D_n X_n\| \leq \hat{\rho}$. From Theorem 2.2, it follows that

$$\rho \leq \hat{\rho} + \eta\rho,$$

from which the corollary follows immediately.

There are three comments to make about this theorem. First, the bound says that the combined effect of all the operations is the same as if we had introduced a relative perturbation in DX_n of approximately norm $\eta \cong -K\epsilon_{\mathrm{M}}/(1-\beta)$. For example, if $\beta = 0.5$, the effect of all the updates is only twice the effect of a single update *whatever the value of n*. Thus, there is never a need to restart the computation to get rid of accumulated rounding errors.

Second, we have focused on the QR factorization for the sake of simplicity. However, the analysis applies *mutatis mutandis* to more complicated decomposition such as the URV and ULV decomposition [2, 3], in which unitary transformations are applied to both sides of $D_n X_n$. The key is to observe that the updating algorithms have backward error analyses in the spirit of Theorem 2.1.

Finally, as we mentioned in the introduction, the bounds are likely to overestimate the error in the long run. The errors do not spread evenly over $D_n X_n$, as the bound seems to imply, but tend to decrease exponentially along with the rows of $D_n X_n$. We will now proceed to analyze this phenomenon.

3. Exponential Decay of the Error

The reason for the weakness of the bounds derived in the last section is that we have ignored the structure of Q_n in passing from the recurrence

$$F_{n+1} = \begin{pmatrix} \beta F_n + Q_n G_n \\ h_n^{\mathrm{H}} \end{pmatrix}$$

to a bound on the backward error. It turns out that the rows of Q_n can decrease exponentially at approximately the same rate as the rows of $D_n X_n$. When this fact is taken into account, we obtain a more realistic bound for the old data.

Theorem 3.1. *Let*

$$\xi \geq \|x_i^{\mathrm{H}}\|$$

be an upper bound on the norms of the x_i^{H} *and let*

$$\rho^{(-1)} \geq \|R_i^{-1}\|, \qquad i = p, p+1, \ldots,$$

be an upper bound on the norms of the inverses of the computed R_i. *Let*

$$\begin{aligned} \kappa &= \rho\rho^{(-1)}, \\ \tau &= \frac{K\epsilon_{\mathrm{M}}}{1 - K\epsilon_{\mathrm{M}}}, \\ \tilde{\beta} &= \beta + \kappa\tau. \end{aligned}$$

Then if f_{in}^{H} *denotes the backward error in the* i*th row of* DX_n,

$$\|f_{in}^{\mathrm{H}}\| \leq \tilde{\beta}^{n-i}\tau\Big(\rho + (n-i)\kappa\xi\Big), \qquad i = p, p+1\ldots;\ n = i, i+1, \ldots.$$

Proof. Let q_{in}^{H} denote the ith row of Q_n. Then

$$\beta^{n-i} x_i^{\mathrm{H}} + f_{in}^{\mathrm{H}} = q_{in}^{\mathrm{H}} R_n.$$

Hence

$$\|q_{in}^{\mathrm{H}}\| \leq \rho^{(-1)}(\beta^{n-i}\xi + \|f_{in}^{\mathrm{H}}\|). \tag{3.1}$$

Now from (2.6),

$$\begin{aligned} f_{ii} &= h_i^{\mathrm{H}}, \\ f_{i,n+1}^{\mathrm{H}} &= \beta f_{in}^{\mathrm{H}} + q_{in}^{\mathrm{H}} G_n, \qquad n = i, i+1, \ldots. \end{aligned} \tag{3.2}$$

Hence from the bound on $\|G_n\|$ developed in the proof of Theorem 2.2 and (3.1) we have

$$\begin{aligned} \|f_{ii}^{\mathrm{H}}\| &\leq \tau\rho, \\ \|f_{i,n+1}^{\mathrm{H}}\| &\leq \tilde{\beta}\|f_{in}^{\mathrm{H}}\| + \tilde{\beta}^{n-i}\kappa\tau\xi, \qquad n = i, i+1, \ldots. \end{aligned}$$

Hence if we set

$$\begin{aligned} \varphi_{ii} &= \tau\rho, \\ \varphi_{i,n+1} &= \tilde{\beta}\varphi_{in} + \tilde{\beta}^{n-i}\kappa\tau\xi, \qquad n = i, i+1, \ldots, \end{aligned} \tag{3.3}$$

then $\|f_{in}^{\mathrm{H}}\| \leq \varphi_{in}$. But it is easily verified that

$$\varphi_{in} = \tilde{\beta}^{n-i}\tau\Big(\rho + (n-i)\kappa\xi\Big).$$

The proof of the theorem must be modified for the case $i < n$, since in this case R_i is singular. The key is to use the bound from Theorem 2.2 as an initial condition for the recursion (3.3). The resulting bound exhibits the same exponential decay.

The number κ is an upper bound on the condition of the R_n, and if some of the R_n are very ill-conditioned, the bounds will be large. However, note that even in applications in which rank-degenerate R_n are expected (e.g., direction of arrival estimation), the presence of noise in the data is likely to make the ill-conditioning very mild compared with the rounding unit.

Finally, note that because of the presence of the term $(n-1)\kappa\xi$, the bounds of this section are initially weaker that the bounds of the preceding section. However, as n increases these bounds ultimately become sharper, since they track the decreasing error while the bounds of the preceding section remain constant.

References

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

[2] G. W. Stewart (1990). "An Updating Algorithm for Subspace Tracking." Technical Report CS-TR 2494, Department of Computer Science, University of Maryland. To appear in IEEE Transactions on Signal Processing.

[3] G. W. Stewart (1991). "Updating a Rank-Revealing ULV Decomposition." Technical Report CS-TR 2627, Department of Computer Science, University of Maryland.

[4] J. H. Wilkinson (1965). *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, England.