CERFACS **Technical Report No. TR/PA/01/17**

L. GIRAUD · J. LANGOU

WHEN MODIFIED GRAM-SCHMIDT GENERATES A WELL-CONDITIONED SET OF VECTORS

July 2001

Abstract. In this paper, we show why the modified Gram-Schmidt algorithm applied to a matrix A generates a well-conditioned set of vectors. This result holds under the assumption that A is not "too ill-conditioned". A consequence is that if we perform two iterations of the algorithm, we obtain a matrix whose columns are orthogonal up to machine precision.

1. Previous Results

We consider the Modified Gram-Schmidt (MGS) algorithm applied to a matrix $A \in \mathbb{R}^{m\times n}$ with full rank $n \leq m$ and singular values : $\sigma_1 \geq \ldots \geq \sigma_n > 0$, we define the condition number of A as $\kappa = \sigma_1/\sigma_n$. Using results from Björck and Paige in [1] and [2], we know that MGS computes $\bar{Q}_1 \in \mathbb{R}^{m\times n}$ and $\bar{R} \in \mathbb{R}^{n\times n}$ so that there exists $\bar{E} \in \mathbb{R}^{m\times n}$, $\hat{E} \in \mathbb{R}^{m\times n}$ and $\hat{Q}_1 \in \mathbb{R}^{m\times n}$, where

$$A + \bar{E} = \bar{Q}_1\bar{R} \text{ and } \|\bar{E}\|_2 \leq \bar{c}_1\mathbf{u}\|A\|_2, \tag{1.1}$$

$$\|I - \bar{Q}_1^T\bar{Q}_1\|_2 \leq \bar{c}_2\kappa\mathbf{u}, \tag{1.2}$$

$$\text{if } c\mathbf{u}\kappa < 1 \text{ then } A + \hat{E} = \hat{Q}_1\bar{R}\ ,\ \hat{Q}_1{}^T\hat{Q}_1 = I \text{ and } \|\hat{E}\|_2 \leq c\mathbf{u}\|A\|_2, \tag{1.3}$$

where $\bar{c}_i$ and c are constants depending on m, n and the details of the arithmetic, and $\mathbf{u}$ is the unit round off.
Result (1.1) shows that $\bar{Q}_1\bar{R}$ is a backward-stable factorization of A, that is the product $\bar{Q}_1\bar{R}$ represents accurately A up to machine precision.
Inequality (1.2) indicates that the level of orthogonality in $\bar{Q}_1$ is dependent on κ, if A is well-conditioned then $\bar{Q}_1$ is orthogonal to machine precision.
Equations (1.3) say that $\bar{R}$ solves the QR-factorization problem in a backward stable sense ; that is, there exists an exact orthonormal matrix $\hat{Q}_1$ so that $\hat{Q}_1\bar{R}$ is a QR factorization of a slightly perturbed A. In [2], it is shown that the result (1.3) holds under the assumption that

$$c\mathbf{u}\kappa < 1. \tag{1.4}$$

In fact, (1.4) enables $\bar{R}$ to be singular. Under this assumption and defining

$$\eta = \frac{1}{1 - c\mathbf{u}\kappa}, \tag{1.5}$$

Björck and Paige obtain an upper bound for $\|\bar{R}^{-1}\|_2$ as

$$\|A\|_2\|\bar{R}^{-1}\|_2 \leq \eta\kappa. \tag{1.6}$$

L. Giraud: CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France. e-mail: `Luc.Giraud@cerfacs.fr`
J. Langou: CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France. e-mail: `Julien.Langou@cerfacs.fr`

Assuming $c\mathbf{u}\kappa < 1$, we note that (1.1) and (1.3) are independent of κ while the bound on $\|I - \bar{Q}_1^T \bar{Q}_1\|_2$ in (1.2) is not. This implies that, for an ill-conditioned matrix A, the set of vectors $\bar{Q}_1$ may lose orthogonality. An important question that arises then is whether MGS manages to preserve the full rank of $\bar{Q}_1$ or not. In order to investigate this, we study in the next section the conditioning of $\bar{Q}_1$. For this purpose, we define the singular values of $\bar{Q}_1$, $\sigma_1(\bar{Q}_1) \geq \ldots \geq \sigma_n(\bar{Q}_1)$. When $\bar{Q}_1$ is non singular, $\sigma_n(\bar{Q}_1) > 0$, we also define the condition number $\kappa(\bar{Q}_1) = \sigma_1(\bar{Q}_1)/\sigma_n(\bar{Q}_1)$.

2. Conditioning of the set of vectors $\bar{Q}_1$

On the one hand, MGS computes $\bar{Q}_1$; on the other, the matrix $\hat{Q}_1$ has exactly orthonormal columns. It seems natural to study the distance between $\bar{Q}_1$ and $\hat{Q}_1$. We define F as

$$F = \bar{Q}_1 - \hat{Q}_1, \tag{2.1}$$

and are interested in the 2-norm of F. For this, we subtract (1.3) from (1.1) to get

$$(\bar{Q}_1 - \hat{Q}_1)\bar{R} = A + \bar{E} - A - \hat{E},$$
$$F\bar{R} = \bar{E} - \hat{E}.$$

Assuming $c\mathbf{u}\kappa < 1$, $\bar{R}$ is nonsingular and we can therefore write

$$F = (\bar{E} - \hat{E})\bar{R}^{-1}.$$

We bound, in terms of norms, this equality, obtaining

$$\|F\|_2 \leq (\|\bar{E}\|_2 + \|\hat{E}\|_2)\|\bar{R}^{-1}\|_2.$$

Using inequality (1.1) on $\|\bar{E}\|_2$ and inequality (1.1) on $\|\hat{E}\|_2$, we obtain

$$\|F\|_2 \leq (c + \bar{c}_1)\mathbf{u}\|A\|_2\|\bar{R}^{-1}\|_2.$$

Using inequalities (1.6) on $\|A\|_2\|\bar{R}^{-1}\|_2$, we have

$$\boxed{\|F\|_2 \leq (c + \bar{c}_1)\mathbf{u}\eta\kappa.} \tag{2.2}$$

This is the desired bound on $\|F\|_2$. We recall that what interests us is an upper bound on $\kappa(\bar{Q}_1)$, the conditioning of $\bar{Q}_1$ We then look for an upper bound for the largest singular value of $\bar{Q}_1$ and a lower bound for its smallest singular value. In that respect, we consider the following lemma ([3, p. 449]) :

LEMMA :
Taking F, $\bar{Q}_1$ and $\hat{Q}_1 \in \mathbb{R}^{m\times n}$,

$$\text{if } F = \bar{Q}_1 - \hat{Q}_1, \text{ then } \sigma_1(\bar{Q}_1) \leq \sigma_1(\hat{Q}_1) + \|F\|_2$$
$$\text{and } \sigma_n(\bar{Q}_1) \geq \sigma_n(\hat{Q}_1) - \|F\|_2.$$

Since $\hat{Q}_1$ has exactly orthonormal columns, we have $\sigma_1(\hat{Q}_1) = \sigma_n(\hat{Q}_1) = 1$ and the lemma gives

$$\sigma_1(\bar{Q}_1) \leq 1 + \|F\|_2 \text{ and } \sigma_n(\bar{Q}_1) \geq 1 - \|F\|_2.$$

Using the bound (2.2) on $\|F\|_2$, we get

$$\sigma_1(\bar{Q}_1) \leq 1 + (c + \bar{c}_1)\mathbf{u}\eta\kappa \text{ and } \sigma_n(\bar{Q}_1) \geq 1 - (c + \bar{c}_1)\mathbf{u}\eta\kappa.$$

Using (1.5), these inequalities can be written as

$$\begin{aligned}&\sigma_1(\overline{Q}_1) \le \eta(1 - c\mathbf{u}\kappa + (c+\overline{c}_1)\mathbf{u}\kappa) \text{ and } \sigma_n(\overline{Q}_1) \ge \eta(1 - c\mathbf{u}\kappa - (c+\overline{c}_1)\mathbf{u}\kappa),\\ &\sigma_1(\overline{Q}_1) \le \eta(1+\overline{c}_1\mathbf{u}\kappa) \qquad\qquad \text{ and } \sigma_n(\overline{Q}_1) \ge \eta(1-(2c+\overline{c}_1)\mathbf{u}\kappa).\end{aligned}$$

If we assume that

$$(2c+\overline{c}_1)\mathbf{u}\kappa < 1, \tag{2.3}$$

$\sigma_n(\overline{Q}_1) > 0$ so $\overline{Q}_1$ is nonsingular.
Notice that this assumption is slightly stronger than (1.4).
Under this assumption, we have :

$$\boxed{\kappa(\overline{Q}_1) \le \frac{1+\overline{c}_1\mathbf{u}\kappa}{1-(2c+\overline{c}_1)\mathbf{u}\kappa}.} \tag{2.4}$$

To illustrate the behaviour of the upper bound of $\kappa(\overline{Q}_1)$, we plot in Figure 2.1 the upper bound as a function of κ, the condition number of A. In that figure, we arbitrarily fix the constant as

$$c = \overline{c}_1 = 3,\ \mathbf{u} = 1e-16.$$

It can be seen that this upper bound explodes when $(2c+\overline{c}_1)\mathbf{u}\kappa \sim 1$ but in the main part of the domain, $(2c+\overline{c}_1)\mathbf{u}\kappa < 1$, it is small.

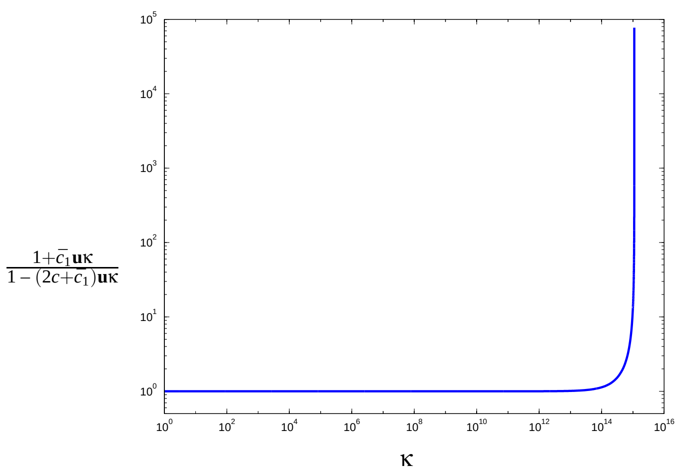

Fig. 2.1. Behaviour of the upper bound on $\kappa(\overline{Q}_1)$ in fonction of κ.

For instance, if we slighly increase the constraint (1.4) used in [2] and assume that

$$c\mathbf{u}\kappa < 0.1 \text{ and } \overline{c}_1\mathbf{u}\kappa < 0.1 \tag{2.5}$$

then

$$\boxed{\kappa(\overline{Q}_1) < 1.6.} \tag{2.6}$$

3. Two remarks

Iterative Modified Gram-Schmidt

If the assumption (2.5) on the condition number of A holds, then we obtain, after a first sweep of MGS, $\bar{Q}_1$ satisfying (2.6). If we run MGS a second time on $\bar{Q}_1$ to obtain $\bar{Q}_2$, we deduce using (1.2) that $\bar{Q}_2$ is such that :

$$\|I - \bar{Q}_2^T \bar{Q}_2\|_2 \leq \bar{c}_2 \kappa(\bar{Q}_1)\mathbf{u},$$

so we get

$$\boxed{\|I - \bar{Q}_2^T \bar{Q}_2\|_2 < 1.6\bar{c}_2\mathbf{u},} \tag{3.1}$$

meaning that $\bar{Q}_2$ has columns orthogonal to machine precision. Two MGS sweeps are indeed enough to have an orthogonal set of vectors Q.

We recover, in a slightly different framework, the famous sentence of Kahan :

Twice is enough.

Based on unpublished notes of Kahan, Parlett in [4] explains that an iterative Gram-Schmidt process on two vectors with a selective criteria (optional) gives two orthonormal vectors up to machine precision. In this paper, we show that *twice is enough* for n vectors under the assumption (2.5) with Modified Gram-Schmidt and full a posteriori reorthogonalization (i.e. no selective criteria).

Importance of the assumptions $c\mathbf{u}\kappa < 0.1$ and $\bar{c}_1\mathbf{u}\kappa < 0.1$

In this part, we want to study what happens when $(2c + \bar{c}_1)\mathbf{u}\kappa \sim 1$. Does there exist matrices in this region ($\kappa\mathbf{u} \sim 1$) so that $\kappa(\bar{Q}_1)$ is high or is the bound (2.6) not so accurate in this region ?

In order to investigate this problem, we developed a Matlab code (available upon email request) that gives as many examples as desired where effectively, if $\kappa\mathbf{u} \sim 1$, $\kappa(\bar{Q}_1)$ is high. For example, we have generated $A \in \mathbb{R}^{50\times 40}$ so that

$$\kappa = 3.140289e+16 \text{ and } \kappa(\bar{Q}_1) = 4.185579e+14.$$

We can verify that in this case, as $\bar{c}_1 > 1$, $c > 1$ and $\mathbf{u}\kappa \sim 3.4864$, assumption (2.5) does not hold anymore. This experiment indicates why the bound on $\kappa(\bar{Q}_1)$ increases rapidly when $\mathbf{u}\kappa \sim 1$.

One can also remark that in this case two MGS sweeps are no longer enough since

$$\|I - \bar{Q}_2^T \bar{Q}_2\|_2 = 6.529765e-02.$$

Acknowledgment

We particularly want to thank Miroslav Rozložník for fruitful discussions on the Modified Gram-Schmidt algorithm that made us understand that the sentence *twice is enough* required the assumption of a not “too ill-conditioned” matrix A.

References

1. A. Björck. Solving linear least squares problems using Gram-Schmidt orthogonalization. *BIT*, 7:1–21, 1967.
2. A. Björck and C. C. Paige. Loss and recapture of orthogonality in the modified Gram-Schmidt Algorithm. *SIAM J. Matrix Analysis and Applications*, 13(1):176–190, January 1992.
3. G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins, 1983.
4. B. Parlett. *The symmetric eigenvalue problem*. Prentice Hall, 1980.