

# Gram–Schmidt Orthogonalization: 100 Years and More

Åke Björck

IWMS, June 5–8, 2010, Shanghai

# Definition

## Gram–Schmidt Process:

The process of forming an orthogonal sequence  $\{y_k\}$  from a linearly independent sequence  $\{x_k\}$  of members of an inner-product space.

*James and James, Mathematical Dictionary, 1949*

This process and the related  $QR$  factorization is a fundamental tool of numerical linear algebra.

The earliest linkage of the names Gram and Schmidt to describe this process appears to be in a paper by Y. K. Wong, An application of orthogonalization process to the theory of least squares. *Ann. Math. Statist.*, 6, 53-75, 1935.

## Early History

Classical versus Modified Gram–Schmidt

Loss of Orthogonality

Least Squares Problems

The Householder Connection

Krylov Subspace Methods

# Early History

In 1907 Erhard Schmidt published an orthogonalization algorithm which became popular and widely used.

Given  $n$  linearly independent elements  $a_1, a_2, \dots, a_n$  in an inner-product space. The algorithm computes an orthonormal basis  $q_1, q_2, \dots, q_n$ , such that

$$a_k = r_{1k}q_1 + r_{2k}q_2 + \dots + r_{kk}q_k, \quad k = 1 : n.$$

The matrix interpretation is that  $A = (a_1, a_2, \dots, a_n)$  is factored into a product

$$A = (q_1, q_2, \dots, q_n) \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix} \equiv QR.$$

$R$  is upper triangular with positive diagonal entries. This factorization is **uniquely determined**.

# Early History

Schmidt used what is now known as the **classical** Gram–Schmidt process.



Erhard Schmidt (1876–1959), studied in Göttingen under Hilbert. In 1917 he assumed a position at the University of Berlin, where he started the famous Institute of Applied Mathematics.

Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener. *Math. Ann.* 1907.

# Early History

Schmidt acknowledged that the algorithm was essentially the same as that previously used by Gram.



Jørgen Pedersen Gram (1850–1916), Danish mathematician, Gram worked for Hafnia Insurance Company and made contributions to probability and numerical analysis.

Ueber die Entwicklung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate. *J. Reine Angew. Math.* 1883.

# The Treatise of Laplace

The orthogonalization algorithm had been used much earlier by other mathematicians, e.g., Laplace, Cauchy, and Bienayme!



Pierre-Simon, Marquis de Laplace (1749–1827) professor at École Militaire, Paris. Laplace was one of the most influential scientists of his time and did major work in probability and celestial mechanics.

Théorie Analytique des Probabilités. Troisième Édition. Premier Supplément. Sur l'application du calcul des probabilités à la philosophie naturelle. Paris: Courcier, 1820. Earlier editions 1812 and 1814.

# The Treatise of Laplace

Let  $A$  be an  $m \times n$  matrix ( $m \geq n$ ) with linearly independent columns and  $b$  an  $m$ -vector of observations with normally distributed random errors. Laplace's treats the statistical theory of errors in linear least squares problems

$$\min_x \|Ax - b\|_2.$$

The particular problem he wants to solve is to estimate the mass of Jupiter and Saturn from astronomical data of 6 planets

The method which Laplace introduces consists in successively projecting the system of equations orthogonally to a column of the matrix  $A$ . Ultimately he is left with the residual vector.



# The Treatise of Laplace

This is precisely the main idea behind the Gram–Schmidt process. However, Laplace uses what is now known as the **modified** Gram–Schmidt process.

Laplace used this to prove that the solution is uncorrelated with the residual vector. For the numerical solution of his problem he solved the corresponding  $6 \times 6$  normal equations.

A translation and modern interpretation of Laplace treatise by Julien Langou is published in Technical Report CCM 280, UC Denver, Denver, Colorado, 2009

# Outline

Early History

**Classical versus Modified Gram–Schmidt**

Loss of Orthogonality

Least Squares Problems

The Householder Connection

Krylov Subspace Methods

# Classical versus Modified Gram–Schmidt

In 1966 John Rice showed by experiments that the two different versions of the Gram–Schmidt orthogonalization, classical (CGS) and modified (MGS) have **very different properties** when executed in finite precision arithmetic. Only for  $n = 2$  are CGS and MGS numerically equivalent:

$$r_{11} = \|a_1\|_2 = (a_1^T a_1)^{1/2}, \quad q_1 = a_1 / r_{11}.$$

Next  $a_2$  is orthogonalized against  $q_1$ , giving

$$r_{12} = q_1^T a_2, \quad a_2^{(2)} = a_2 - r_{12} q_1.$$

Finally, the vector  $a_2^{(2)}$  is normalized, giving

$$r_{22} = \|a_2^{(2)}\|_2, \quad q_2 = a_2^{(2)} / r_{22}.$$

# Classical versus Modified Gram–Schmidt

## Classical Gram–Schmidt algorithm (CGS):

In step  $k$  of CGS, the vector  $a_k$  is orthogonalized against  $q_1, \dots, q_{k-1}$ . The  $k$ th column of  $R$  is computed and only the first  $k$  columns are operated on.

```
function [Q,R] = gschmidt(A);  
[m,n] = size(A);  
Q = A; R = zeros(n);  
for k = 1:n  
    R(1:k-1,k) = Q(:,1:k-1)'*A(:,k);  
    Q(:,k) = A(:,k) - Q(:,1:k-1)*R(1:k-1,k);  
    R(k,k) = norm(Q(:,k));  
    Q(:,k) = Q(:,k)/R(k,k);  
end
```

# Classical versus Modified Gram–Schmidt

At step  $k$  in row-wise MGS applied to  $A = A^{(1)}$  we have computed

$$A^{(k)} = (q_1, \dots, q_{k-1}, a_k^{(k)}, \dots, a_n^{(k)})$$

where  $a_k^{(k)}, \dots, a_n^{(k)}$ , have been made orthogonal to  $q_1, \dots, q_{k-1}$ . In the next step we set

$$q_k = a_k^{(k)} / \|a_k^{(k)}\|_2,$$

and orthogonalize  $a_{k+1}^{(k)}, \dots, a_n^{(k)}$  against  $q_k$ . This gives the  $k$ th row of  $R$ .

## Modified Gram–Schmidt algorithm (MGS):

```
function [Q,R] = mgs(A);  
[m,n] = size(A);  
Q = A; R=zeros(n);  
for k = 1:n  
    R(k,k) = norm(Q(:,k));  
    Q(:,k) = Q(:,k)/R(k,k);  
    R(k,k+1:n) = Q(:,k)' * Q(:,k+1:n);  
    Q(:,k+1:n) = Q(:,k+1:n) - Q(:,k) * R(k,k+1:n);  
end
```

Row-wise MGS has the advantage that a **column pivoting strategy** can be used, giving an upper triangular matrix  $R$  with non-increasing diagonal elements.

# Classical versus Modified Gram–Schmidt

The vector  $q_k$  is computed in CGS as

$$a_k^{(k)} = (I - Q_{k-1} Q_{k-1}^T) a_k, \quad Q_{k-1} = (q_1, \dots, q_{k-1})$$

and in MGS as

$$a_k^{(k)} = (I - q_{k-1} q_{k-1}^T) \cdots (I - q_1 q_1^T) a_k.$$

The crucial difference is that in MGS the projections  $r_{kj} q_k$  are subtracted from  $a_j$  as soon as they are computed.

note that there is a column-wise version of MGS, but no row-wise version of CGS.

# Outline

Early History

Classical versus Modified Gram–Schmidt

**Loss of Orthogonality**

Least Squares Problems

The Householder Connection

Krylov Subspace Methods



# Loss of Orthogonality

The **condition number** of an  $m \times n$  matrix  $A$  is

$$\kappa_2(A) = \frac{\max_{\|x\|_2=1} \|Ax\|_2}{\min_{\|x\|_2=1} \|Ax\|_2} = \frac{\sigma_1}{\sigma_n},$$

Unless  $c\kappa_2(A)u < 1$ , where  $u$  is the unit roundoff (IEEE double precision  $u = 1.11 \cdot 10^{-16}$ ) and  $c = c(m, n) > 1$  is of moderate size, the matrix  $A$  is **numerically rank deficient**.

The minimum distance from  $A$  to the set of matrices of rank less than  $n$  is

$$\text{dist}_2(A)/\|A\|_2 = \sigma_n/\|A\|_2 = 1/\kappa_2(A),$$

so this means that  $A$  is “close” to a matrix of exact rank less than  $n$ . Then the column space of  $A$  is not “well defined”.

# Loss of Orthogonality

Both CGS and MGS compute factors  $\bar{Q}$  and  $\bar{R}$  such that  $\bar{Q}\bar{R} \approx A$ , but the orthogonality in  $\bar{Q}$  differ substantially.

**MGS consistently produces vectors which are more orthogonal than those generated by CGS.**

For illustration we generated a 50 by 10 matrix

$$A = UDV^T, \quad D = \text{diag}(1, 10^{-1}, \dots, 10^{-9})$$

with  $U$  and  $V$  orthogonal matrices. In the table below the condition number

$$\kappa(A_k) = \sigma_1(A_k)/\sigma_k(A_k), \quad A_k = (a_1, \dots, a_k)$$

and the loss of orthogonality after  $k$  steps are shown.

# Loss of Orthogonality

$k$	$\kappa(A_k)$	$\ I_k - Q_C^T Q_C\ _2$	$\ I_k - Q_M^T Q_M\ _2$
1	1.000e+00	1.110e-16	1.110e-16
2	1.335e+01	2.880e-16	2.880e-16
3	1.676e+02	7.295e-15	8.108e-15
4	1.126e+03	2.835e-13	4.411e-14
5	4.853e+05	1.973e-09	2.911e-11
6	5.070e+05	5.951e-08	3.087e-11
7	1.713e+06	2.002e-07	1.084e-10
8	1.158e+07	1.682e-04	6.367e-10
9	1.013e+08	3.330e-02	8.779e-09
10	1.000e+09	5.446e-01	4.563e-08

The computed vectors  $q_k$  from CGS depart from orthogonality much more rapidly!

# Loss of Orthogonality

Consider the orthogonalization of two vectors  $(a_1, a_2)$ . of unit length. Then  $q_1 = a_1$ . Denote by  $\bar{r}_{12} = fl(q_1^T a_2)$  the **computed** scalar product. Using the standard model for floating point computation, we get

$$|\bar{r}_{12} - r_{12}| < mu + O(u^2).$$

The error in  $\bar{w}_2 = fl(a_2 - fl(\bar{r}_{12}q_1))$  can be bounded by

$$\|\bar{w}_2 - w_2\|_2 < (m + 2)u + O(u^2).$$

Since  $q_1^T w_2 = 0$ , we have  $|q_1^T \bar{w}_2| \lesssim (m + 2)u$ . Assuming that the normalization  $\bar{q}_2 = \bar{w}_2 / \bar{r}_{22}$ ,  $\bar{r}_{22} = \|\bar{w}_2\|_2$ , is carried out without error,

$$|q_1^T \bar{q}_2| < (m + 2)u / \bar{r}_{22}.$$

# Loss of Orthogonality

If  $\bar{r}_{22}$  is small, then **cancellation** has occurred in the orthogonalization. Since  $\|a_2\|_2 = 1$ , we have

$$|q_1^T \bar{q}_2| \approx \frac{(m+2)u}{\sin \angle(a_1, a_2)}.$$

This result is independent of the initial scaling of the vectors  $a_1$  and  $a_2$ . (The **GS algorithms are invariant under column scaling**)

The loss of orthogonality in one step will be propagated and possibly amplified in later steps. Further losses of orthogonality may occur due to cancellations in the computation

$$a_j^{(k+1)} = (I - q_k q_k^T) a_j^{(k)} = a_j^{(k)} - q_k (q_k^T a_j^{(k)}).$$

If  $\|a_j^{(k+1)}\|_2 \ll \|a_j^{(k)}\|_2$ , then cancellation has occurred.

# Loss of Orthogonality

In 1967 Björck 1967 proved the following error bounds for the computed factors in MGS:

## Theorem

*Let  $\bar{Q}_1$  and  $\bar{R}$  denote the factors computed by the MGS algorithm in floating point arithmetic with unit roundoff  $u$ . Then there are constants  $c_i = c_i(m, n)$ ,  $i = 1, 2$ , depending on  $m$  and  $n$  such that if  $c_1 \kappa(A)u < 1$ , then*

$$\|A - \bar{Q}\bar{R}\|_2 \leq c_2 u \|A\|_2. \quad (1)$$

$$\|I - \bar{Q}^T \bar{Q}\|_2 \leq \frac{c_1 u \kappa_2(A)}{1 - c_1 u \kappa_2(A)}. \quad (2)$$

# Loss of Orthogonality

A sharper upper bound can be obtained by using the invariance of the MGS algorithm under column scaling. For  $\kappa_2(A)$  we can substitute  $\tilde{\kappa}_2 = \min_{D>0} \kappa_2(AD)$ , where  $D$  is a positive diagonal matrix. **Scaling  $A$  so that all columns in  $A$  have equal norm approximately minimizes  $\kappa_2(AD)$**

## Theorem (van der Sluis, 1969)

*Let  $A$  have full column rank and denote by  $\mathcal{D}$  the set of positive diagonal matrices. Then if all columns in  $A$  have equal 2-norm, it holds that*

$$\kappa_2(A) \leq \sqrt{n} \min_{D>0} \kappa_2(AD).$$

# Outline

Early History

Classical versus Modified Gram–Schmidt

Loss of Orthogonality

**Least Squares Problems**

The Householder Connection

Krylov Subspace Methods



# Least Squares Problems

In 1801 Gauss predicted the orbit of the asteroid Ceres using the method of least squares. Since then, the principle of least squares has been the standard procedure for the analysis of scientific data.



Carl Friedrich Gauss (1777–1855), one of the greatest mathematicians of the nineteenth century, spent most of his life in Göttingen.

Gauss (1821, 1823) gave the method a theoretical basis in *Theoria combinationis observationum erroribus minimis obnoxiae*.

# Least Squares Problems

The **first publication** of the method was in 1805 by A. M. Legendre in *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris

*Of all the principles that can be proposed, I think there is none more general, more exact, and more easy of application, than that which consists of rendering the sum of the squares of the errors a minimum.*

Gauss 1809 wrote, much to the annoyance of Legendre,

*Our principle, which we have made use of since 1795, has lately been published by Legendre.*

# Least Squares Problems

One of the most important applications of Gram–Schmidt algorithms is for solving the **linear least squares problem**

$$\min_x \|Ax - b\|_2,$$

It is assumed that the errors in  $b$  are independent and equally distributed. The solution is characterized by  $r \perp \mathcal{R}(A)$ , where  $r = b - Ax$  is the residual vector.

A related problem is the **conditional least squares problem**.

$$\min_y \|y - b\|_2^2 \quad \text{subject to} \quad A^T y = c.$$

A unified treatment of these two problems can be obtained as follows:

# Least Squares Problems

## Theorem

Let the  $m$  by  $n$  matrix  $A$  have full column rank and consider the augmented linear system of  $n + m$  equations

$$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}.$$

Then the system is nonsingular and gives the first order conditions for two least squares problem:

$$\min_x \|Ax - b\|_2^2 + 2c^T x,$$

$$\min_y \|y - b\|_2, \quad \text{subject to } A^T y = c,$$

For  $c = 0$ , the first is the standard least squares problem; For  $b = 0$ , the second is the minimum norm solution of  $A^T y = c$ .

# Perturbation Analysis

The **augmented system** was introduced by Lanczos in 1952. It plays an important role, e.g., in the **sensitivity analysis** of least squares problems

From the Schur–Banachiewicz formula it follows that the inverse of the augmented matrix equals

$$\begin{aligned} \begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix}^{-1} &= \begin{pmatrix} (I - A(A^T A)^{-1} A^T) & A(A^T A)^{-1} \\ (A^T A)^{-1} A^T & -(A^T A)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} P_A^\perp & (A^\dagger)^T \\ A^\dagger & -(A^T A)^{-1} \end{pmatrix}, \end{aligned} \quad (3)$$

where  $A^\dagger$  is the pseudoinverse of  $A$  and  $P_A^\perp$  the orthogonal projection onto the nullspace of  $A^T$ .

# Perturbation Analysis

Assume that  $A + \delta A$  and  $b + \delta b$  are perturbed data, that  $\text{rank}(A) = \text{rank}(A + \delta A) = n$ . Then the perturbed solution  $x + \delta x$  and  $r + \delta r$ , satisfies (Björck 1967)

$$\begin{aligned}\|\delta x\|_2 &\lesssim \frac{1}{\sigma_n} \|\delta b\|_2 + \frac{1}{\sigma_n} \|\delta A\|_2 \left( \|x\|_2 + \frac{1}{\sigma_n} \|r\|_2 \right), \\ \|\delta r\|_2 &\lesssim \|\delta b\|_2 + \|\delta A\|_2 \left( \|x\|_2 + \frac{1}{\sigma_n} \|r\|_2 \right).\end{aligned}$$

If  $x \neq 0$  and  $\delta b = 0$ , then an upper bound for the normwise relative perturbation is

$$\frac{\|\delta x\|_2}{\|x\|_2} \leq \kappa_{LS}(A, b) \frac{\|\delta A\|_2}{\|A\|_2}, \quad \kappa_{LS} = \kappa(A) \left( 1 + \frac{\|r\|_2}{\sigma_n \|x\|_2} \right). \quad (4)$$

# Sensitivity Analysis

Note that  $\kappa_{LS}(A, b)$  depends not only on  $A$ , but also on the residual  $r$  and hence on  $b$ .

If  $\|r\|_2 \ll \sigma_n \|x\|_2$ , then  $\kappa_{LS} \approx \kappa(A)$ , but if  $\|r\|_2 > \sigma_n \|x\|_2$  the second term dominates. This can be written as

$$\kappa^2(A) \frac{\|r\|_2}{\|A\|_2 \|x\|_2}.$$

Hence, **the square of the matrix condition number is to some extent relevant to the least squares problem.**

The estimates are sharp, to within a factor of  $\sqrt{2}$ . (Some estimates in current literature can overestimate the error by a factor of  $\kappa_2(A)$ ; see Grcar 2009)!

# Numerical Stability

A method for computing  $y = f(x)$  is **backward stable** if the computed result  $\bar{y}$  equals  $f(x + \Delta x)$ , where  $\|\Delta x\|$  is small. This does not guarantee that the forward error  $\bar{y} - y$  is small. However, if (an upper bound for) the condition number of  $f$  is known a bound for the forward error can be obtained.

A method is said to be **forward stable** if it can be shown to produce forward errors of the same size as a backward stable method.

Backward stability implies forward stability, but not vice versa.



# Least Squares Algorithms

Apply MGS to  $(A, b)$ , where the right-hand side  $b$  is taken as  $(n+1)$ st column. Skipping the normalization of the last column, this gives

$$(A \ b) = (Q_1 \ r) \begin{pmatrix} R & z \\ 0 & 1 \end{pmatrix}.$$

Hence,  $r = b - Q_1 z$ , and further,

$$\|Ax - b\|_2 = \left\| (A \ b) \begin{pmatrix} x \\ -1 \end{pmatrix} \right\|_2 = \|Q_1(Rx - z) - r\|_2.$$

From Pythagoras' theorem follows that if  $Q_1^T r = 0$ , then the minimum of the last expression occurs when  $Rx = z$ .

This algorithm was proved to be **forward stable** by Björck 1967. What about backward stability?

# Outline

Early History

Classical versus Modified Gram–Schmidt

Loss of Orthogonality

Least Squares Problems

**The Householder Connection**

Krylov Subspace Methods

# The Householder Connection

In 1968 Charles Sheffield made the surprising observation that the MGS QR factorization of  $A$  is equivalent to Householder QR algorithm applied to  $A$  with a square matrix of zeros on top. The equivalence holds also in finite precision.



Charles Sheffield (1935–2002) got his PhD in theoretical physics from Cambridge, UK. He later became an award-winning science fiction author. Many of his books (*Cold as Ice*, *The amazing doctor Darwin*, *Georgia on my Mind*) remain in print.

# The Householder Connection

The Householder QR factorization of  $A$  is

$$P_n \cdots P_2 P_1 A = \begin{pmatrix} R \\ O \end{pmatrix},$$

where  $R$  is upper triangular and

$$P_k = I - 2 \frac{u_k u_k^T}{u_k^T u_k}, \quad k = 1 : n.$$

are orthogonal Householder transformations (plane reflections).

The factorization can be written

$$A = Q \begin{pmatrix} R \\ O \end{pmatrix} = Q \begin{pmatrix} I_n \\ O \end{pmatrix} R = Q_1 R,$$

The matrix  $Q = P_1 P_2 \cdots P_n$ , is **implicitly defined by the Householder vectors**  $u_k$ ,  $k = 1 : n$ .

# The Householder Connection

The (normwise) backward stability of Householder QR was proved by J. H. Wilkinson 1965. Note that  $Q$  in the theorem is **not computed by the algorithm**.

## Theorem

*Let  $\bar{R}$  denote the upper triangular matrix computed by the Householder QR algorithm for  $A$ . Then there exists an exactly orthogonal  $m \times m$  matrix  $Q$  such that*

$$A + \Delta A = Q \begin{pmatrix} \bar{R} \\ 0 \end{pmatrix}, \quad \|\Delta a_j\|_2 \leq \frac{c(m, n)u}{1 - c(m, n)u} \|a_j\|_2, \quad j = 1 : n.$$

*Here  $Q = (P_1 P_2 \cdots P_n)^T$ , where  $P_k$  is the Householder matrix that corresponds to the exact application of the  $k$ th step of the algorithm to the computed  $fl(P_{k-1} \cdots P_1 A)$ .*

# The Householder Connection

Sheffield's observation is that for MGS

$$\tilde{A} = \begin{pmatrix} O \\ A \end{pmatrix} = \tilde{Q}R, \quad Q = P_1 P_2 \cdots P_n,$$

where  $P_k = I - v_k v_k^T$ , The Householder vectors are given by

$$v_k = \begin{pmatrix} -e_k \\ q_k \end{pmatrix}, \quad \|v_k\|_2 = 2, \quad k = 1 : n,$$

where  $q_k$  is the  $k$ th column in the MGS factor  $Q$ .

Björck and Paige 1992 used this equivalence to derive

**backward stable MGS least squares algorithms.**

The matrix  $\bar{R}$  computed by MGS satisfies

$$A + E = Q\bar{R}, \quad \|E\|_2 \leq cu\|A\|_2.$$

where  $Q$  is an exactly orthogonal matrix (not computed).

# Least Squares Algorithms

A backward stable MGS algorithm for  $x$  and  $r$  in the least squares problem. If only the norm of  $r$  is needed, the last loop, which is new, can be deleted.

```
function [x,r,rho] = mgss(Q,R,b);  
[m,n] = size(Q);  
z = zeros(n,1);  
for k = 1:n  
    z(k) = Q(:,k)'*b;  
    b = b - z(k)*Q(:,k);  
end  
x = R\b; r = b;  
for k = n:-1:1  
    w = Q(:,k)'*r;  
    r = r - w*Q(:,k);  
end  
rho = norm(r);
```

# Least Squares Algorithms

A special case of the conditional least squares problem is when  $b = 0$

$$\min \|y\|_2 \quad \text{subject to} \quad A^T y = c.$$

If MGS has computed  $R$  and  $Q_1 = (q_1, \dots, q_n)$ , then the solution satisfies

$$z = R^{-T}c, \quad y = Q \begin{pmatrix} z \\ 0 \end{pmatrix} = Q_1 z.$$

An MGS algorithm to compute  $y$  goes as follows: Solve  $R^T z = c$  for  $z = (\zeta_1, \dots, \zeta_n)^T$ . Set  $y_n = 0$ , and compute

$$y_{k-1} = y_k + q_k(\zeta_k - w_k), \quad w_k = q_k^T y_k, \quad k = n, n-1, \dots, 1.$$

to compute  $y = y_0$ . The corrections  $w_k$  compensate for the lack of orthogonality in  $Q$ .



# Least Squares Algorithms

A backward stable MGS algorithm for the conditional least squares problem.

```
function [y,rho] = mgsc(Q,R,b,c);
[m,n] = size(Q);
h = b;  z = R'\c;
for k = 1:n
    d = Q(:,k)'\*h;
    h = h - d*Q(:,k);
end
for k = n:-1:1
    w = Q(:,k)'\*h;
    h = h + (z(k) - w)*Q(:,k);
end
y = h;  rho = norm(y - b);
```

A multiple purpose orthonormalizing code in 1954 at the National Bureau of Standards (NBS) is described by Davis and Rabinowitz. This used CGS (with a twist). An Algol implementation named **ORTHO** by Walsh 1962 includes reorthogonalization and was much used.

In the late 1950th many computer codes for solving least squares problems used MGS with column pivoting. Björck 1968 published two Algol subroutines for the solution of linear least squares problems based on MGS. They used column pivoting and handled the more general least squares problem with (consistent) linear equality constraints

$$\min_x \|A_2x - b_2\|_2 \quad \text{subject to} \quad A_1x = b_1.$$

The Householder QR least squares algorithm was published by G. H. Golub in 1965. This is slightly more economical, and more flexible than MGS.

G. Peters and J. H. Wilkinson 1970 wrote:

*Evidence is accumulating that the modified Gram–Schmidt method gives better results than Householder. The reasons for this phenomenon appear not to have been elucidated yet.*

The Algol codes of Björck were translated 1979 by Wampler at NBS into Fortran routines **L2A** and **L2B**. They were extended to allow for diagonal weighting of equations and in addition compute the covariance matrix.

# Outline

Early History

Classical versus Modified Gram–Schmidt

Loss of Orthogonality

Least Squares Problems

The Householder Connection

**Krylov Subspace Methods**

# Krylov Subspace Methods

Given a matrix  $A$  and an initial vector  $u_0$ , the **Krylov subspace** of order  $k$  is

$$\mathcal{K}_k(A, u_0) = \text{span}\{u_0, Au_0, A^2u_0, \dots, A^{k-1}u_0\}.$$

Krylov subspace methods (due to Alexei N. Krylov 1931) are used extensively for

1. Computing approximate solutions of large systems of linear equations.
2. Approximating a subset of the eigenvalues and eigenvectors of a large matrix.

As the Krylov subspace vectors are not a good basis for numerical computations. This has contributed significantly to the revival of interest in the Gram-Schmidt process.

# The Arnoldi Method

The CGS or column-wise MGS algorithm can be used to compute an orthogonal basis.

$$\mathcal{K}_{k+1}(\mathbf{A}, u_0) = \text{span}\{q_0, q_1, \dots, q_k\} = \text{span}\{Q_k\}$$

In the next step the vector  $A^k u_0$  is to be orthogonalized to produce the vector  $q_{k+1}$ .

In the **Arnoldi process** the vector  $A^k u_0$  is replaced by  $Aq_k$ . The orthogonalization step becomes

$$h_{k+1,k} q_{k+1} = Aq_k - \sum_{i=1}^k h_{ik} q_i.$$

The MGS–Arnoldi process computes the factorization

$$AQ_k = Q_{k+1} \tilde{H}_k, \quad k = 0, 1, 2, \dots,$$

where  $\tilde{H}_k$  is the  $(k+1) \times k$  Hessenberg matrix formed by the elements  $h_{ij}$ ,  $i = 1 : k, j \geq i$  and  $h_{k+1,k}$

# The Arnoldi Method

For solving a large, sparse unsymmetric linear system  $Ax = b$ , the Arnoldi process is used with the unit starting vector  $q_0 = b/\|b\|_2$ . We seek an approximate solution of the form

$$x_k = Q_k y_k \in \mathcal{K}_k(A, b),$$

In the **GMRES** (Generalized Minimum Residual) method,  $y_k$  is taken as the solution to the least squares problem

$$\min_{y_k} \|\beta_1 e_1 - \tilde{H}_k y_k\|_2, \quad k = 0, 1, 2, \dots$$

In exact arithmetic this also minimizes  $\|b - Ax_k\|_2$  and the residual norms will not increase.

Paige et al. (2006) have shown that MGS–GMRES produces a backward stable approximate solution and the loss of orthogonality does not affect the convergence.

# The Arnoldi Method

The Arnoldi process is also used to find approximate eigenpairs of an unsymmetric matrix  $A$ . Let  $Q_k = (q_1, \dots, q_k)$  be the orthogonal basis computed at step  $k$ . Then the  $k \times k$  Hessenberg matrix

$$H_k = Q_k^H(AQ_k)$$

is the orthogonal projection of  $A$  onto  $\text{span}(Q_k)$ . Compute the  $k$  eigenvalues and eigenvectors of  $H_k$ ,

$$H_k z_i = \theta_i z_i, \quad i = 1, \dots, k.$$

The **Ritz values**  $\theta_i$  and **Ritz vectors**  $y_i = Q_k z_i$  then are approximate eigenpairs of  $A$ .

In this method *orthogonality to working precision* must be enforced, since otherwise  $H_k = Q_k^H A Q_k$  is not a similarity transformation. This is achieved by reorthogonalization.



# Reorthogonalization

If  $A$  has full numerical column rank, then one reorthogonalization step suffices for CGS and MGS to achieve orthogonality to roundoff levels. That **twice is enough** was proved for MGS and CGS and arbitrary  $n$  by Giraud et al. 2005.

The algorithm CGS2 applied to  $A$  proceeds as follows.

Let  $Q_{k-1} = (q_1, \dots, q_{k-1})$  be computed basis vectors.

Then  $a_k^{(0)} = a_k = Ae_k$  is orthogonalized twice

$$a_k^{(i)} = a_k^{(i-1)} - Q_{k-1}(Q_{k-1}^T a_k^{(i-1)}), \quad i = 1, 2.$$

The new basis vector is then

$$q_k = a_k^{(2)} / \|a_k^{(2)}\|_2.$$

MGS2 is similar, but the column-wise version uses vector operations and is slower.

# Acknowledgements

Steven J. Leon, Åke Björck, and Walter Gander.  
Gram–Schmidt Orthogonalization: 100 Years and More.  
SIAM Review, submitted 2009.