Lecture 6
Projectors

NLA Reading Group Spring ’13

by Umut Şimşekli

Definition

A *projector* is a square matrix P that satisfies

$$P^2 = P.$$

- A.k.a ‘idempotent’
- Can be orthogonal or non-orthogonal (oblique)

Properties

- What if v is in range(P) ?

> Observe that if $v \in \text{range}(P)$, then it lies exactly on its own shadow, and applying the projector results in v itself. Mathematically, we have $v = Px$ for some x and
> $$Pv = P^2x = Px = v.$$

- What if v≠Pv? Where does the shine come from?

Properties

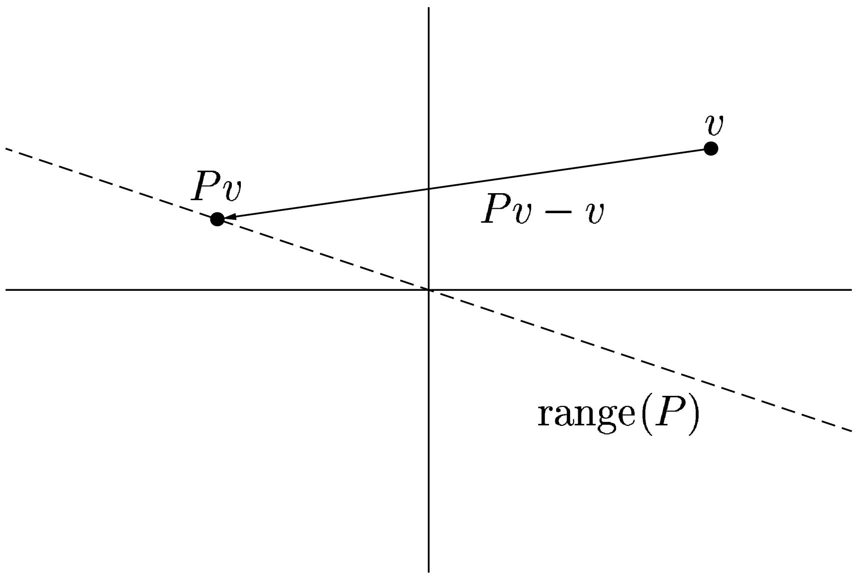


Figure 6.1. *An oblique projection.*

- Draw the line from v to Pv

$$P(Pv - v) = P^2 v - Pv = 0.$$

Complementary Projectors

If P is a projector, $I - P$ is also a projector, for it is also idempotent:

$$(I-P)^2 = I - 2P + P^2 = I - P.$$

The matrix $I - P$ is called the *complementary projector* to P.

Onto what space does $I - P$ project? Exactly the nullspace of P! We know that $\operatorname{range}(I-P) \supseteq \operatorname{null}(P)$, because if $Pv = 0$, we have $(I-P)v = v$. Conversely, we know that $\operatorname{range}(I-P) \subseteq \operatorname{null}(P)$, because for any v, we have $(I-P)v = v - Pv \in \operatorname{null}(P)$. Therefore, for any projector P,

$$\operatorname{range}(I-P) = \operatorname{null}(P). \tag{6.2}$$

By writing $P = I - (I-P)$ we derive the complementary fact

$$\operatorname{null}(I-P) = \operatorname{range}(P). \tag{6.3}$$

We can also see that $\operatorname{null}(I-P) \cap \operatorname{null}(P) = \{0\}$: any vector v in both sets satisfies $v = v - Pv = (I-P)v = 0$. Another way of stating this fact is

$$\operatorname{range}(P) \cap \operatorname{null}(P) = \{0\}. \tag{6.4}$$

Complementary Projectors

Conversely, let S_1 and S_2 be two subspaces of $\mathbb{C}^m$ such that $S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \mathbb{C}^m$, where $S_1 + S_2$ denotes the span of S_1 and S_2, that is, the set of vectors $s_1 + s_2$ with $s_1 \in S_1$ and $s_2 \in S_2$. (Such a pair are said to be *complementary subspaces*.) Then there is a projector P such that $\text{range}(P) = S_1$ and $\text{null}(P) = S_2$. We say that P is the projector *onto* S_1 *along* S_2. This projector and its complement can be seen as the unique solution to the following problem:

$$\boxed{\text{Given } v, \text{ find vectors } v_1 \in S_1 \text{ and } v_2 \in S_2 \text{ such that } v_1 + v_2 = v.}$$

- Eigen decomposition example: What are the coefficients given a particular basis?

Orthogonal Projectors

An *orthogonal projector* (Figure 6.2) is one that projects onto a subspace S_1 along a space S_2, where S_1 and S_2 are orthogonal. (Warning: orthogonal projectors are not orthogonal matrices!)

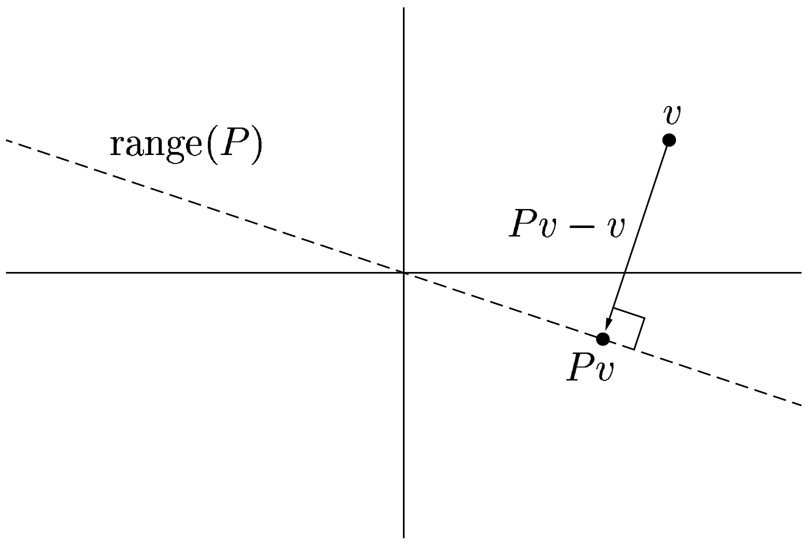


Figure 6.2. *An orthogonal projection.*

Algebraic Definition of Orthogonal Projectors

Theorem 6.1. *A projector P is orthogonal if and only if $P = P^*$.*

Proof. If $P = P^*$, then the inner product between a vector $Px \in S_1$ and a vector $(I - P)y \in S_2$ is zero:

$$x^* P^* (I - P) y = x^* (P - P^2) y = 0.$$

For "only if," we can use the SVD. Suppose P projects onto S_1 along S_2, where $S_1 \perp S_2$ and S_1 has dimension n. Then an SVD of P can be constructed as follows. Let $\{q_1, q_2, \ldots, q_m\}$ be an orthonormal basis for $\mathbb{C}^m$, where $\{q_1, \ldots, q_n\}$ is a basis for S_1 and $\{q_{n+1}, \ldots, q_m\}$ is a basis for S_2.

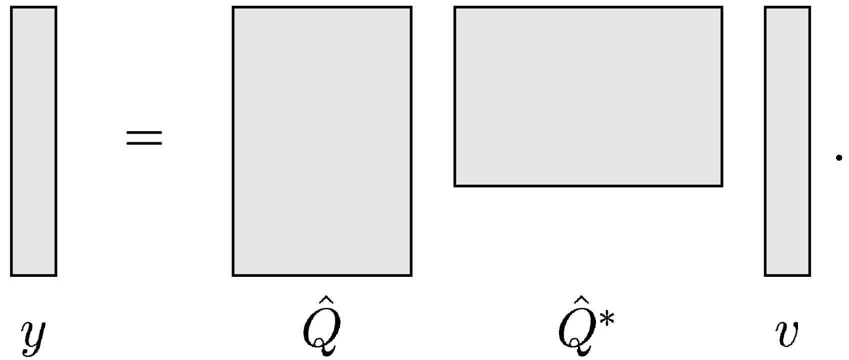
Proof Cont.

$$PQ = \left[\begin{array}{c|c|c|c|c} & & & & \\ q_1 & \cdots & q_n & 0 & \cdots \\ & & & & \end{array}\right],$$

$$Q^*PQ = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} = \Sigma,$$

$P = Q\Sigma Q^*$. => $P^* = (Q\Sigma Q^*)^* = Q\Sigma^* Q^* = Q\Sigma Q^* = P$

Projection with an Orthonormal Basis



The complement of an orthogonal projector is also an orthogonal projector (proof: $I - \hat{Q}\hat{Q}^*$ is hermitian). The complement projects onto the space orthogonal to range$(\hat{Q})$.

- One rank:

$$P_q = qq^*. \qquad P_{\perp q} = I - qq^*.$$

$$P_a = \frac{aa^*}{a^*a}, \quad P_{\perp a} = I - \frac{aa^*}{a^*a}.$$

Projection with Arbitrary Basis

In passing from v to its orthogonal projection $y \in \text{range}(A)$, the difference $y - v$ must be orthogonal to range(A). This is equivalent to the statement that y must satisfy $a_j^*(y - v) = 0$ for every j. Since $y \in \text{range}(A)$, we can set $y = Ax$ and write this condition as $a_j^*(Ax - v) = 0$ for each j, or equivalently, $A^*(Ax - v) = 0$ or $A^*Ax = A^*v$. It is easily shown that since A has full rank, A^*A is nonsingular (Exercise 6.3). Therefore

$$x = (A^*A)^{-1}A^*v. \tag{6.12}$$

Finally, the projection of v, $y = Ax$, is $y = A(A^*A)^{-1}A^*v$. Thus the orthogonal projector onto range(A) can be expressed by the formula

$$P = A(A^*A)^{-1}A^*. \tag{6.13}$$

Lecture 7

QR Factorization

NLA Reading Group Spring '13

by Umut Şimşekli

Reduced QR Factorization

For many applications, we find ourselves interested in the column spaces of a matrix A. Note the plural: these are the *successive* spaces spanned by the columns $a_1, a_2, \ldots$ of A:

$$\langle a_1 \rangle \subseteq \langle a_1, a_2 \rangle \subseteq \langle a_1, a_2, a_3 \rangle \subseteq \ldots.$$

- Find a matrix Q such that:

$$\langle q_1, q_2, \ldots, q_j \rangle = \langle a_1, a_2, \ldots, a_j \rangle, \qquad j = 1, \ldots, n.$$

Reduced QR Factorization

$$\left[\begin{array}{c|c|c|c} & & & \\ a_1 & a_2 & \cdots & a_n \\ & & & \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ q_1 & q_2 & \cdots & q_n \\ & & & \end{array}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix},$$

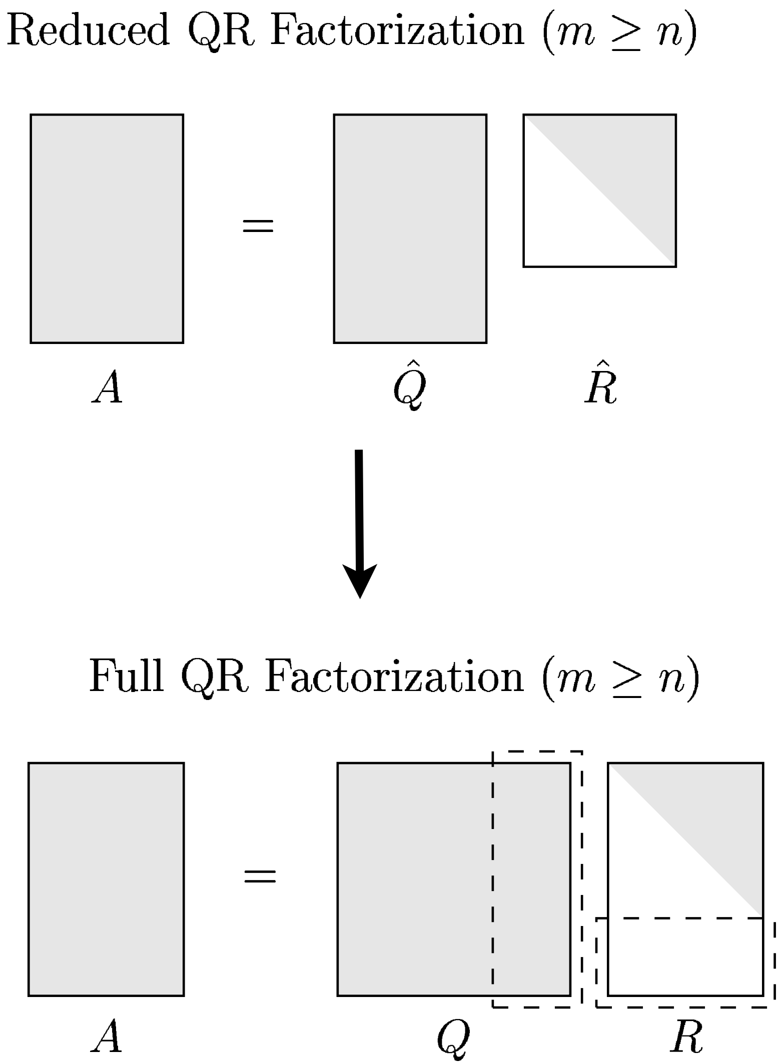
$$\Big\downarrow$$

$$\begin{aligned} a_1 &= r_{11} q_1, \\ a_2 &= r_{12} q_1 + r_{22} q_2, \\ a_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3, \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \cdots + r_{nn} q_n. \end{aligned}$$

$$\Big\downarrow$$

$$\boxed{A = \hat{Q}\hat{R},}$$

Full QR Factorization



Gram-Schmidt Orthogonalization

- From Lecture 2 we know:

$$v_j = a_j - (q_1^* a_j)q_1 - (q_2^* a_j)q_2 - \cdots - (q_{j-1}^* a_j)q_{j-1}$$

- Rewrite

$$\begin{aligned} a_1 &= r_{11}q_1, \\ a_2 &= r_{12}q_1 + r_{22}q_2, \\ a_3 &= r_{13}q_1 + r_{23}q_2 + r_{33}q_3, \\ &\vdots \\ a_n &= r_{1n}q_1 + r_{2n}q_2 + \cdots + r_{nn}q_n. \end{aligned} \longrightarrow \begin{aligned} q_1 &= \frac{a_1}{r_{11}}, \\ q_2 &= \frac{a_2 - r_{12}q_1}{r_{22}}, \\ q_3 &= \frac{a_3 - r_{13}q_1 - r_{23}q_2}{r_{33}}, \\ &\vdots \\ q_n &= \frac{a_n - \sum_{i=1}^{n-1} r_{in}q_i}{r_{nn}}. \end{aligned}$$

$$\begin{aligned} r_{ij} &= q_i^* a_j \qquad (i \neq j). \\ |r_{jj}| &= \left\| a_j - \sum_{i=1}^{j-1} r_{ij}q_i \right\|_2. \end{aligned}$$

Gram-Schmidt Orthogonalization

Algorithm 7.1. Classical Gram–Schmidt (unstable)

for $j = 1$ **to** n

$\quad v_j = a_j$

$\quad$ **for** $i = 1$ **to** $j - 1$

$\quad\quad r_{ij} = q_i^* a_j$

$\quad\quad v_j = v_j - r_{ij} q_i$

$\quad r_{jj} = \|v_j\|_2$

$\quad q_j = v_j / r_{jj}$

Existence

Theorem 7.1. *Every $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$ has a full QR factorization, hence also a reduced QR factorization.*

Proof. Suppose first that A has full rank and that we want just a reduced QR factorization. In this case, a proof of existence is provided by the Gram–Schmidt algorithm itself. By construction, this process generates orthonormal columns of $\hat{Q}$ and entries of $\hat{R}$ such that (7.4) holds. Failure can occur only if at some step, v_j is zero and thus cannot be normalized to produce q_j. However, this would imply $a_j \in \langle q_1, \ldots, q_{j-1}\rangle = \langle a_1, \ldots, a_{j-1}\rangle$, contradicting the assumption that A has full rank.

Now suppose that A does not have full rank. Then at one or more steps j, we shall find that (7.5) gives $v_j = 0$, as just mentioned. At this moment, we simply pick q_j arbitrarily to be any normalized vector orthogonal to $\langle q_1, \ldots, q_{j-1}\rangle$, and then continue the Gram–Schmidt process.

Finally, the full, rather than reduced, QR factorization of an $m \times n$ matrix with $m > n$ can be constructed by introducing arbitrary orthonormal vectors in the same fashion. We follow the Gram–Schmidt process through step n, then continue on an additional $m - n$ steps, introducing vectors q_j at each step.

Uniqueness

Theorem 7.2. *Each $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$ of full rank has a unique reduced QR factorization $A = \hat{Q}\hat{R}$ with $r_{jj} > 0$.*

Proof. Again, the proof is provided by the Gram–Schmidt iteration. From (7.4), the orthonormality of the columns of $\hat{Q}$, and the upper-triangularity of $\hat{R}$, it follows that any reduced QR factorization of A must satisfy (7.6)–(7.8). By the assumption of full rank, the denominators (7.8) of (7.6) are nonzero, and thus at each successive step j, these formulas determine r_{ij} and q_j fully, except in one place: the sign of r_{jj}, not specified in (7.8). Once this is fixed by the condition $r_{jj} > 0$, as in Algorithm 7.1, the factorization is completely determined. □

Vectors -> Continuous Functions

- Inner product of two functions:

$$(f, g) = \int_{-1}^{1} \overline{f(x)}\, g(x)\, dx.$$

- QR Decomposition

$$\left[\begin{array}{c|c|c|c|c} \\ 1 & x & x^2 & \cdots & x^{n-1} \\ \\ \end{array}\right] = \left[\begin{array}{c|c|c|c} \\ q_0(x) & q_1(x) & \cdots & q_{n-1}(x) \\ \\ \end{array}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}$$

$$A = QR$$

Legendre Polynomials

Solution of Ax = b by QR

$$\boxed{\begin{array}{c} Ax = b \qquad A = QR \\ \big\downarrow \\ Rx = Q^*b. \end{array}}$$

1. Compute a QR factorization $A = QR$.
2. Compute $y = Q^*b$.
3. Solve $Rx = y$ for x.

Lecture 8

Gram-Schimidt Orthogonalization

NLA Reading Group Spring ’13

by Umut Şimşekli

Gram-Schmidt Projections

- Recall the orthogonalization method

$$q_1 = \frac{P_1 a_1}{\|P_1 a_1\|}, \quad q_2 = \frac{P_2 a_2}{\|P_2 a_2\|}, \quad \ldots, \quad q_n = \frac{P_n a_n}{\|P_n a_n\|}.$$

$$P_1 = I$$

$$P_j = I - \hat{Q}_{j-1}\hat{Q}_{j-1}^* \longrightarrow \hat{Q}_{j-1} = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_{j-1} \end{array}\right]$$

Modified Gram-Schmidt Algorithm

- Standard GS Algorithm

$$
\begin{aligned}
v_j &= P_j a_j \\
P_j &= P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1} \\
v_j &= P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1} a_j
\end{aligned}
$$

- Modified GS Algorithm

$$
\begin{array}{ccccccc}
 & & v_j^{(1)} & = & a_j, & & \\
 & & v_j^{(2)} & = & P_{\perp q_1} v_j^{(1)} & = & v_j^{(1)} - q_1 q_1^* v_j^{(1)}, \\
 & & v_j^{(3)} & = & P_{\perp q_2} v_j^{(2)} & = & v_j^{(2)} - q_2 q_2^* v_j^{(2)}, \\
 & & & \vdots & & \vdots & \\
v_j & = & v_j^{(j)} & = & P_{\perp q_{j-1}} v_j^{(j-1)} & = & v_j^{(j-1)} - q_{j-1} q_{j-1}^* v_j^{(j-1)}
\end{array}
$$

  - Numerically more stable

Modified Gram-Schmidt Algorithm

Algorithm 8.1. Modified Gram–Schmidt

for $i = 1$ **to** n

$\qquad v_i = a_i$

for $i = 1$ **to** n

$\qquad r_{ii} = \|v_i\|$

$\qquad q_i = v_i / r_{ii}$

$\qquad$ **for** $j = i + 1$ **to** n

$\qquad\qquad r_{ij} = q_i^* v_j$

$\qquad\qquad v_j = v_j - r_{ij} q_i$

Gram-Schmidt as Triangular Orthogonalization

- First step:

$$\left[\begin{array}{c|c|c|c} & & & \\ v_1 & v_2 & \cdots & v_n \\ & & & \end{array}\right] \overbrace{\begin{bmatrix} \frac{1}{r_{11}} & \frac{-r_{12}}{r_{11}} & \frac{-r_{13}}{r_{11}} & \cdots \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{bmatrix}}^{R_1} = \left[\begin{array}{c|c|c|c} & & & \\ q_1 & v_2^{(2)} & \cdots & v_n^{(2)} \\ & & & \end{array}\right]$$

- Then:

$$R_2 = \begin{bmatrix} 1 & & & \\ & \frac{1}{r_{22}} & \frac{-r_{23}}{r_{22}} & \cdots \\ & & 1 & \\ & & & \ddots \end{bmatrix}, \quad R_3 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \frac{1}{r_{33}} & \cdots \\ & & & \ddots \end{bmatrix}, \cdots.$$

$$A \underbrace{R_1 R_2 \cdots R_n}_{\hat{R}^{-1}} = \hat{Q}.$$

Lecture 9
MATLAB

NLA Reading Group Spring ’13

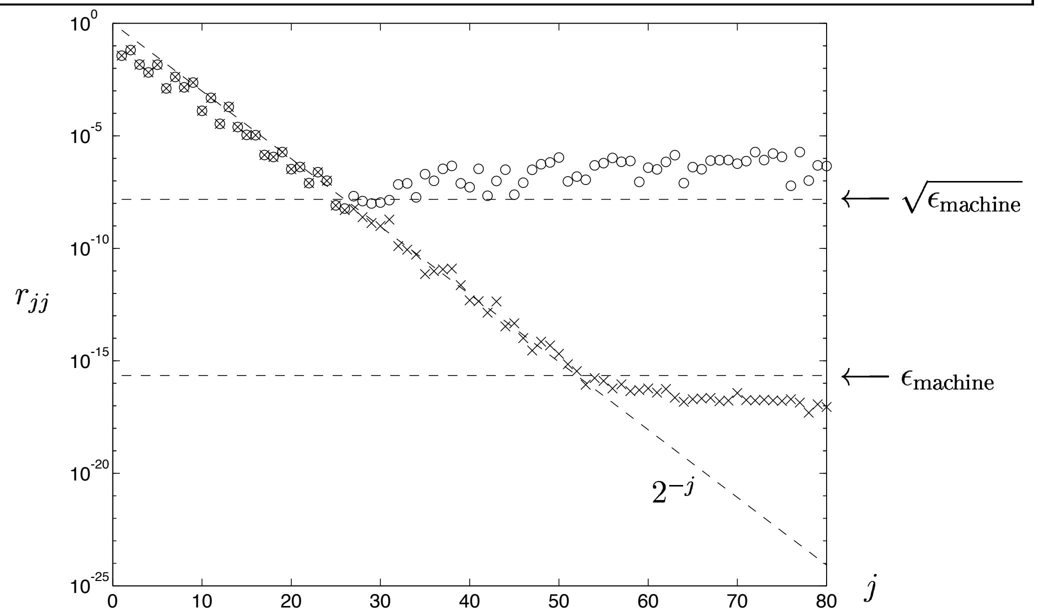
by Umut Şimşekli

Modified vs Standard Gram-Schmidt

- Numerical stability

`[U,X] = qr(randn(80));`	Set U to a random orthogonal matrix.
`[V,X] = qr(randn(80));`	Set V to a random orthogonal matrix.
`S=diag(2.^(-1:-1:-80));`	Set S to a diagonal matrix with exponentially graded entries.
`A = U*S*V;`	Set A to a matrix with these entries as singular values.

`[QC,RC] = clgs(A);`	Compute a factorization $Q^{(c)}R^{(c)}$ by classical Gram–Schmidt.
`[QM,RM] = mgs(A);`	Compute a factorization $Q^{(m)}R^{(m)}$ by modified Gram–Schmidt.



Another numerical stability example

$$A = \begin{bmatrix} 0.70000 & 0.70711 \\ 0.70001 & 0.70711 \end{bmatrix}$$

- By hand (5 digit precision) $Q = \begin{bmatrix} 0.70710 & 1.0000 \\ 0.70711 & 0.0000 \end{bmatrix}$
- MATLAB

`A = [.70000 .70711; .70001 .70711];`	Define A.
`[Q,R] = qr(A);`	Compute factor Q by Householder.
`   norm(Q'*Q-eye(2))`	Test orthogonality of Q.
`[Q,R] = mgs(A);`	Compute factor Q by modified G–S.
`   norm(Q'*Q-eye(2))`	Test orthogonality of Q.

The lines without semicolons produce the following printed output:

```
ans = 2.3515e-16,    ans = 2.3014e-11.
```