# GRAM-SCHMIDT PROJECTIONS

C.R. JOHNSON, D.D. OLESKY

AND

P. VAN DEN DRIESSCHE

DMS-544-IR MAY 1990

# Gram-Schmidt Projections

CHARLES R. JOHNSON(1)
*Department of Mathematics, College of William and Mary,*
*Williamsburg, VA 23185*

D.D. OLESKY(2)
*Department of Computer Science, University of Victoria,*
*Victoria, British Columbia V8W 2Y2*

P. VAN DEN DRIESSCHE(3)
*Department of Mathematics and Statistics, University of Victoria,*
*Victoria, British Columbia V8W 3P4*

**Mathematics Subject Classification Number:** 15A04

(1) The work of this author was supported in part by the National Science Foundation grant DMS88–02836 and a grant from the Office of Naval Research.

(2) The work of this author was supported in part by NSERC grant A–8214 and by the University of Victoria Committee on Faculty Research and Travel.

(3) The work of this author was supported in part by NSERC grant A–8965 and by the University of Victoria Committee on Faculty Research and Travel.

# Gram-Schmidt Projections

Given a linearly independent set in an inner product space, the Gram-Schmidt orthonormalization process is an algorithm that replaces it with an orthonormal set that spans the same subspace. Gram-Schmidt is mentioned in many linear algebra books; see, e.g., [3, 0.6.4]. If $\{x_1,\ldots,x_n\}$ is a linearly independent set in $\mathbb{R}^m$, $m \geq n$, the sequential calculation of the resulting orthonormal set $\{z_1,\ldots,z_n\}$ is given by

$$y_k = x_k - \sum_{i=1}^{k-1} (z_i^T x_k) z_i$$

$$z_k = y_k/(y_k^T y_k)^{\frac{1}{2}}, \qquad k = 1, \ldots, n. \tag{1}$$

If $X$ is the m-by-n matrix with columns $x_1, \ldots, x_n$ and $Z$ the m-by-n matrix with columns $z_1, \ldots, z_n$, this results in the factorization $X = ZR$, with $Z$ orthogonal and $R$ upper triangular. We work with $x_i \in \mathbb{R}^m$, but note that with obvious modifications, our results hold for $x_i \in \mathbb{C}^m$; for example, the matrix $Z$ is then unitary.

The factorization of an m-by-n matrix $A$ into a product $QR$, where $Q$ is an m-by-n matrix with orthonormal columns and $R$ is upper triangular, is called a $QR$ factorization of $A$ (see [3, 2.6]). This factorization is used, for example, in numerical methods for solving least squares problems and computing eigenvalues and singular values (see [2]). When $A$ has full column rank, it has a unique factorization $A = QR$, where $Q$ is an m-by-n matrix with orthonormal columns, and $R$ is an n-by-n upper triangular matrix with positive main diagonal entries.

Here we consider a fixed m-by-n real matrix $A$ having full column rank, and define a sequence of m-by-m matrices $B^{(k)}$, $k = 0, \ldots, n-1$, which we note are the linear transformations that implement the Gram-Schmidt process. This provides an alternate (and we believe novel, though elementary) way of viewing the process and of building up the matrix $Q$ of the $QR$ factorization. This paper is an outgrowth of our detailed study of the combinatorial structure of the matrices $Q$ and $R$ of this factorization [4].

Let $a_i$, $q_i$ denote the ith column of matrix $A$, $Q$, respectively, and let $I_m$ denote the m-by-m identity matrix. Given an m-by-n matrix $A$, we claim that the Gram-Schmidt process applied to the column vectors of $A$ can be written in terms of matrix transformations as follows:

$$y_k = B^{(k-1)} a_k, \qquad q_k = y_k/(y_k^T y_k)^{\frac{1}{2}} \quad \text{for} \quad k = 1, \ldots, n, \tag{2}$$

where

$$B^{(0)} = I_m, \qquad B^{(k)} = B^{(k-1)} - q_k q_k^T \quad \text{for} \quad k = 1, \ldots, n-1. \tag{3}$$

For the first term, $y_1 = a_1$ and $q_1 = a_1/(a_1^T a_1)^{\frac{1}{2}}$. For $k = 1, \ldots, n-1$, from (3) we have

$$B^{(k)} = I_m - \sum_{i=1}^{k} q_i q_i^T. \tag{4}$$

Thus (2) gives $y_k = a_k - \sum_{i=1}^{k-1} (q_i^T a_k) q_i$. Identifying $a_k$ with $x_k$ and $q_i$ with $z_i$, this is exactly the Gram-Schmidt process given by (1) as claimed.

A column of matrix $A$ can be written as the sum of two orthogonal vectors, namely

$$a_k = B^{(k-1)} a_k + \sum_{i=1}^{k-1} (q_i^T a_k) q_i .$$

The first vector $B^{(k-1)} a_k$ is orthogonal to $\mathrm{Span}\left\{q_1, \ldots, q_{k-1}\right\}$, whereas the second vector (involving the summation) is in $\mathrm{Span}\left\{q_1, \ldots, q_{k-1}\right\}$ and is the projection of $a_k$ on this subspace.

The matrices $B^{(k)}$ have some interesting basic properties which depend only on the fact that they are defined from an orthonormal sequence. We now summarize these (in Theorems 1 and 2) and then (in Theorem 3) prove a result which shows explicitly how zero rows can occur in $B^{(k)}$ and so force zero entries in the output of Gram–Schmidt.

THEOREM 1.

(i) *Given* $k$ $(0 \le k \le n-1)$ , *the matrix* $B^{(k)}$ is *positive semidefinite with rank* $m-k$.

(ii) *If* $0 \le i \le j \le n-1$, *then* $B^{(i)} \ge B^{(j)}$, *that is* $B^{(i)} - B^{(j)}$ *is positive semidefinite.*

(iii) *If* $0 \le i,j \le n-1$, *then* $B^{(i)}B^{(j)} = B^{(j)}B^{(i)} = B^{(q)}$ *where* $q = \max\{i,j\}$.

*Proof.*

(i) The fact that $B^{(k)}$ is symmetric is obvious from (4). As $q_i^Tq_i = 1$ and $q_j^Tq_i = 0$ for $i \neq j$, then $(B^{(k)})^2 = B^{(k)}$; that is $B^{(k)}$ is idempotent. Thus the eigenvalues of $B^{(k)} \in \{0,1\}$, and $B^{(k)}$ is positive semidefinite. (Note that $B^{(0)} = I_m$ is, in fact, positive definite.) Each product $q_iq_i^T$ is symmetric and has rank 1, and $q_iq_i^Tq_jq_j^T = q_jq_j^Tq_iq_i^T = 0$ for all $i \neq j$; so $\left\{q_iq_i^T\right\}$, $i = 1, \ldots, n$, form a commuting family of symmetric matrices. Thus there exists a single orthogonal matrix $U$ such that $U(q_iq_i^T)U^T = D_i$, where $D_i$ is an m-by-m diagonal matrix for $i = 1, \ldots, n$ (see, e.g., [3, 2.5.5]). Each $D_i$ has rank 1, so it has exactly one nonzero entry (equal to 1), and if $i \neq j$, then $D_iD_j = Uq_iq_i^Tq_jq_j^TU^T = 0$. From (4) we see that $UB^{(k)}U^T = I_m - \sum_{i=1}^{k} D_i$, and thus rank $B^{(k)} = m - k$.

(ii) This follows directly from (3). Note that $B^{(i)} \neq B^{(j)}$ for $i \neq j$.

(iii) Suppose $q = i > j$, and consider

$$\begin{aligned} B^{(i)}B^{(j)} &= \left[B^{(j)} - q_{j+1}q_{j+1}^T - \cdots - q_iq_i^T\right]B^{(j)} \\ &= B^{(j)} - \left[q_{j+1}q_{j+1}^T + \cdots + q_iq_i^T\right]\left[I_m - q_1q_1^T - \cdots - q_jq_j^T\right] \\ &= B^{(j)} - \left[q_{j+1}q_{j+1}^T + \cdots + q_iq_i^T\right] = B^{(i)}. \end{aligned}$$

The other products follow in a similar manner. ■

As $B^{(k)}$ is symmetric and idempotent, it follows that it is a projection matrix, hence we call $B^{(k)}$ a *Gram–Schmidt projection.* If the range of $B^{(k)}$ is $S_k$, then $B^{(k)}$ is the orthogonal projection onto $S_k$. We can deduce more about the spectrum of $B^{(k)}$ by using the rank result in theorem 1 (i). In fact, $B^{(k)}$ has eigenvalue 1 with multiplicity $m - k$, and eigenvalue 0 with multiplicity $k$. The spectral properties show immediately that rank $B^{(k)}$ = trace $B^{(k)}$. From theorem 1 (iii) and also the fact that $B^{(i)}B^{(j)}B^{(i)} = B^{(i)}$ for $i > j$, we note that $B^{(j)}$ is a $\{1,3,4,5\}$ generalized inverse of $B^{(i)}$; see, e.g., [1]. Also $B^{(k)}$ is its own $\{1,2,3,4,5\}$ inverse.

We use the properties of $\left\{B^{(k)}\right\}$ to deduce the following.

THEOREM 2. *If a sequence of* $m$-*by*-$m$ *matrices* $B^{(k)}$, $k = 0, \ldots, n-1$, *have the properties of theorem* 1 (i), (ii) *and* (iii), *then they must satisfy* (3) *for some orthonormal set of vectors* $\{q_k\}$.

*Proof*. First note that $C^{(k)} \equiv B^{(k-1)} - B^{(k)} \geq 0$ from theorem 1 (ii), so if $x^TB^{(k-1)}x = 0$ then $x^TB^{(k)}x = 0$. With $S_k$ denoting the subspace onto which $B^{(k)}$ projects, we have $S_k \subset S_{k-1}$. Letting $S_k^\perp$ denote the orthogonal complement of $S_k$, if $x \in S_k \cup S_{k-1}^\perp$ then $C^{(k)}x = 0$ (as $B^{(k-1)}x = B^{(k)}x = 0$ if $x \in S_{k-1}^\perp$; and $B^{(k)}x = x = B^{(k-1)}x$ if $x \in S_k$). Thus, by theorem 1 (i), the null space of $C^{(k)}$ has dimension at least $(m-k) + (k-1) = m-1$. But

$C^{(k)}$ is not the zero matrix, so the dimension of the null space of $C^{(k)}$ is exactly $m-1$, which implies that $C^{(k)}$ has rank one. We can therefore write $C^{(k)} = f_k f_k^T$ for some vector $f_k \in \mathbb{R}^m$. As trace $C^{(k)} = 1$, we must have $f_k^T f_k = 1$. Also, by theorem 1 (iii), $C^{(i)}C^{(j)} = 0$ for $i \neq j$, so $f_i f_i^T f_j f_j^T = (f_i^T f_j) f_i f_j^T = 0$, which implies $f_i^T f_j = 0$, and $\{f_k\}$ is an orthonormal set of vectors which we can identify with $\{q_k\}$. ∎

Given $k$ $(2 \le k \le n)$, $t$ $(1 \le t \le k-1)$ and indices $1 \le j_1 < j_2 < \dots < j_t \le k-1$, note that $B^{(r)}$, $r \ge k-1$, projects into the orthogonal complement of $\mathrm{Span}\left\{a_{j_1},\dots,a_{j_t}\right\}$, since it projects onto the orthogonal complement of $\mathrm{Span}\{a_1,\dots,a_r\}$. In particular, $\mathrm{Span}\left\{a_{j_1},\dots,a_{j_t}\right\}$ is in the null space of $B^{(r)}$. In the event that $a_{j_1}, \dots, a_{j_t}$ collectively have nonzero entries in *only* the $t$ rows $i_1, \dots, i_t$, then $\mathrm{Span}\left\{a_{j_1},\dots,a_{j_t}\right\}$ is exactly the coordinate subspace of $\mathbb{R}^m$ in which there are arbitrary entries in positions $i_1, \dots, i_t$ and 0 entries elsewhere. For $B^{(r)}$ to have such a subspace in its null space, each row of $B^{(r)}$ must have zeros in the positions $i_1, \dots, i_t$. We have thus proved the following combinatorial result.

THEOREM 3. *If columns* $a_{j_1}, \dots, a_{j_t}$ *for* $1 \le j_1 < \dots < j_t \le k-1$ *and* $2 \le k \le n$ *collectively have nonzero entries only in rows* $i_1, \dots, i_t$, *then rows and columns* $i_1, \dots, i_t$ of $B^{(r)}$ *are zero for* $k-1 \le r \le n-1$. ∎

It follows from (2) that if some subset of the columns of $A$ has the property stated in Theorem 3, then the entries $i_1, \dots, i_t$ of vectors $q_{r+1}, \dots, q_n$ are zero.

The following example illustrates the construction and some properties of the Gram-Schmidt projections.

Consider $A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 0 & 1 \\ 2 & 4 & 2 \end{bmatrix}$. Then (2) and (3) give:

$$B^{(0)} = I_3, \qquad q_1 = \frac{1}{\sqrt{5}}(1,\ 0,\ 2)^T,$$

$$B^{(1)} = I_3 - q_1q_1^T = \begin{bmatrix} \frac{4}{5} & 0 & -\frac{2}{5} \\ 0 & 1 & 0 \\ -\frac{2}{5} & 0 & \frac{1}{5} \end{bmatrix}, \ q_2 = \frac{1}{\sqrt{5}}(2,\ 0,\ -1)^T,$$

$$B^{(2)} = B^{(1)} - q_2q_2^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad q_3 = (0,1,0)^T.$$

Here columns $a_1$, $a_2$ have nonzero entries only in rows 1, 3; thus rows and columns 1 and 3 of $B^{(2)}$ are zero and the 1 and 3 entries of $q_3$ are zero. ■

## REFERENCES

1. Ben-Israel, A. and Greville, T.N.E., Generalized inverses: theory and applications, 1974, Wiley.

2. Golub, G.H. and Van Loan, C.F., Matrix Computations, 1989, Johns Hopkins University Press.

3. Horn, R.A. and Johnson, C.R., Matrix Analysis, 1985, Cambridge University Press.

4. Johnson, C.R., Olesky, D.D., and van den Driessche, P., Sparsity analysis of the QR factorization, 1990, University of Victoria DMS-536-IR Technical Report.