ON B-ORTHOGONALIZATION

G. W. Stewart
University of Maryland

THE SETTING

- We are given:
  - A positive definite matrix B of order n (with $\|B\| = 1$).
  - An $n \times k$ matrix X with linearly independent columns. We will assume $\|X\| = 1$.
- We want to find U and R such that
  - $X = UR$,
  - $U^*BU = I$.
- We can do this (in theory) by the B-Gram–Schmidt algorithm.

THE B-GRAM–SCHMIDT ALGORITHM

```
U = zeros(n,0);
R = zeros(m);
for i=1:k
    r = U'*B*X(:,i);
    y = X(:,i) - U*r;
    rho = sqrt(y'*B*y);
    U = [U, y/rho];
    R(1:i-1,i) = r;
    R(i,i) = rho;
end
```

- We will call this process B-Orthogonalization.

OUTLINE

In this talk we will consider three topics.

- We will first consider the following problem. Given a projector P, find U and B such that $P = UU^*B$.
  - UU^*B is clearly idempotent and hence a projector.

- Second, we will investigate the properties of the B-projector $P = UU^*B$.

- Finally, we will consider the B-orthogonalization of a Krylov sequence in which B is semi-definite.

PROJECTORS

- An itempotent matrix P is a projector onto $\mathcal{R}(P)$ along the $\mathcal{R}(P^*)^\perp$.

- Projectors have the following useful canonical form.

  Let P be a projector of order n and rank k. Then there is a unitary matrix Q such that

  $$Q^*PQ = \begin{pmatrix} I_k \\ 0 \end{pmatrix} \begin{pmatrix} I_k & T \end{pmatrix} = \begin{pmatrix} I_k & T \\ 0 & 0 \end{pmatrix},$$

  where T_k is an $k\times(n-k)$ diagonal matrix whose diagonal elements are the tangents of the canonical angles between $\mathcal{R}(P)$ and $\mathcal{R}(P^*)$.

- We will use this form to construct our matrix.

THE CONSTRUCTION I

- We will assume that P is in canonical form and that
$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{pmatrix}$$
and
$$U^* = (\hat{U}^* \ \ 0).$$

- Then
$$U^* B U = \hat{U}^* B_{11} \hat{U} = I,$$
and
$$B_{11} = (\hat{U}\hat{U}^*)^{-1}.$$

- We can then use the canonical form to show that
$$B = \begin{pmatrix} B_{11} & B_{11}T \\ T^* B_{11} & B_{22} \end{pmatrix}.$$

THE CONSTRUCTION II

- We may choose B_{22} so that B is positive definite. Equivalently, the Schur complement

$$C = B_{22} - T^* B_{11}^{-1} T$$

must be positive definite.

- If we choose $\hat{U} = I$, then

$$C = B_{22} - T^* T,$$

and we can choose $B_{22} = \mu I$, where $\mu > \|T^* T\|$

- If we normalize B, then as μ grows, U becomes larger and B becomes more ill conditioned.

THE B-PROJECTOR

- We have seen that the matrix $P = UU^*B$ is idempotent and hence is an (oblique) projector onto $\mathcal{R}(X)$.

- If B has small eigenvalues U may be large.

- Despite this fact $\|P\|$ is often near one.

THE QR APPROACH

- Let

$$B^{\frac{1}{2}}X = QR$$

be a QR-factorization of B. Then with

$$U = XR^{-1}$$

we have

$$U^*BU = (U^*B^{\frac{1}{2}})(B^{\frac{1}{2}}U) = Q^*Q = I,$$

so that U is the B-orthonormal basis for $\mathcal{R}(X)$.

- Since $U = XR^{-1} = B^{-\frac{1}{2}}Q$, we have $BU = B^{\frac{1}{2}}Q$, or

$$\|BU\| \leq 1.$$

- In the B Gram–Schmidt, this implied that `r = U'*B*x` has norm no greater than one (no matter how big U is).

THE GRADED STRUCTURE

- The problem is now to show when P is small. WLOG we can assume that B is diagonal with its eigenvalues appearing in descending order.

- If X does not itself have exceptional grading, $B^{\frac{1}{2}}X$ will be graded by rows. We will assume a constant grading ratio (the ratio of the norms of successive rows) of γ.

- The QR factorization of such a matrix tends to have the following properties.
  - The diagonals of R are graded with factor γ.
  - The jth column of Q has an ith component near one and grades downward on either side of that component by a factor of γ.

IMPLICATIONS

The QR factorization of such a matrix tends to have the following properties.

- The diagonals of R are graded with factor γ.
- The jth columns of Q has a kth component near one and grades downward on either side of that component by a factor of k.

———◇———

- Let $R = DS$, where S is the diagonal of D. Then
$$U = XS^{-1}D^{-1}.$$
If S is well conditioned, then norms of the columns of U can be expected to be graded upward as we progress across U.
- The columns of $B^{\frac{1}{2}}Q = BU$ will be graded downward.
- When we form the product $U(U'B)$, the gradings will cancel out.

EXAMPLE I

- B is a diagonal matrix of order 10 with eigenvalues $1, 10^{-4}, 10^{-8}, \ldots, 10^{-40}$.
- X is an 10×5 orthonormal matrix.
- The matrix R is

-4.9239e-01	2.4990e-01	-8.4315e-02	2.7645e-01
0	1.6609e-03	-9.3237e-04	2.9630e-03
0	0	-2.7165e-05	1.1985e-05
0	0	0	1.3790e-07

EXAMPLE II

- The matrix Q is

-9.9999e-01	-3.2940e-03	-2.9379e-05	9.4494e-08
3.2942e-03	-9.9998e-01	-5.3462e-03	-1.4118e-06
1.1768e-05	5.3463e-03	-9.9998e-01	2.7649e-03
-6.6601e-08	1.6191e-05	-2.7648e-03	-9.9998e-01
1.3695e-09	-5.1989e-07	2.3993e-05	5.0813e-03
-5.1840e-12	-3.9371e-10	-5.1844e-08	4.4815e-05
-1.0187e-14	-4.3995e-12	4.4527e-10	2.0805e-07
-2.6953e-16	4.2853e-14	3.4772e-12	-9.4990e-10
-3.8481e-19	4.6508e-16	-2.3870e-14	-7.2861e-12
-8.7742e-22	-3.5384e-18	-4.2257e-17	8.3744e-14

- The condition of S is 5.5.

EXAMPLE III

- The column norms of U are

```
2.0309e+00   6.7517e+02   4.2422e+04   1.3838e+07,
```

those of BU are

```
9.9999e-01   6.8401e-03   5.6408e-05   2.5521e-07,
```

and their product is

```
2.0309e+00   4.6183e+00   2.3929e+00   3.5315e+00
```

- $\|P\| = \|UU^*B\| = 5.2.$

ROUNDING-ERROR ANALYSIS

- In analyzing the B-Gram–Schmidt step, I found the constants $\|U\|^2$ appearing in the bounds.
  - For our previous example, the former amounts about 10^{14}
  - This is uncomfortable close to the reciprocal of the rounding unit.

- One way it entered was through the norm of $r = U^*Bx$, which we now know be bounded by $\|x\|$.

- It also entered through, $\|P\|$, which, we have seen, may be much smaller than the naive bound $\|U\|^2$.

- The preceding observations support a sharper analysis.

SEMI-DEFINITE B

- When B is semi-definite, then for any vector y in the null space of B, we have $Py = 0$, and hence $(I - P)y = y$.
  - In consequence, the B-GS algorithm cannot purge components in the null space of B.
  - This leads to the possibility that rounding error might cause null-space error to grow.
- For the ordinary method this does not seem to be a problem.
- However, when one is B-orthogonalizing the vectors in a Krylov sequence, the null-space error can grow swiftly.
- This problem has lead to various algorithms for purging the null space error.
  - However, this will be ineffective if the error is allowed to become too large.

THE REDUCED FORM

- The general problem involves B-orthogonalizing the Krylov sequence generated by the matrix AB. By reducing B to the form

$$\begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix},$$

we can transform AB into the form

$$\begin{pmatrix} C & 0 \\ D & 0 \end{pmatrix}.$$

- If $(V_k^* \ \ W_k^*)^*$ is the partition Krylov basis, then we have the relations

$$CV_k = V_k H_k + \eta_k v_{k+1} \mathbf{e}_k^* \quad \text{and} \quad DV_k = W_k H_k + \eta_k w_{k+1} \mathbf{e}_k^*.$$

  - H_k and η_k are determined entirely from the Krylov sequence for C. The second relation follows the leader.

CONVERGING RITZ VECTORS

$$DV_k = W_k H_k + \eta_k w_{k+1} \mathbf{e}_k^*$$

———————◇———————

- Let $H_k s = \theta s$. Then $C_k p_k$ is a Ritz vector of C.

- We have

$$DV_k s = \theta W_k s + \eta_k s_k w_{k+1},$$

where σ_k is the last component of s. It follows that

$$w_{k+1} = \frac{DV_k s - \theta W_k s}{\eta_k s_k}.$$

- Note that s_k is the norm of the residual of the Ritz vector $V_k s$. It is small if the vector is converging. In this case, we can expect w_{k+1} to be large.