The Arnoldi Process

We now discuss some refinements that allow efficient computation of the approximation lying in $\mathcal{K}_{m+1}$ after already having found the approximation in $\mathcal{K}_m$. Write

$$Q_m = \left[\, q_1 \middle| q_2 \middle| \cdots \middle| q_m \,\right]$$

where q_i are orthonormal column vectors. If we use the modified Gram-Schmidt algorithm to form Q_m then for each $j = 1, \ldots, m$ the span of the first j columns of Q_m is equal to the span of the first j columns of Γ_m. Thus, Q_{m+1} may be formed by simply adding an additional column q_{m+1} to Q_m obtained from $A^m b$ using the Gram-Schmidt algorithm

$$t_0 = A^m b, \qquad t_{k+1} = t_k - (q_k \cdot t_k) q_k, \qquad q_{m+1} = t_{m+1} / \|t_{m+1}\|.$$

Therefore, when moving from $\mathcal{K}_m$ to $\mathcal{K}_{m+1}$ there is no need to recompute the entire reduced QR decomposition of Γ_{m+1} over again. This method of computing Q_{m+1} from Q_m is called the Arnoldi process.

A similar savings can be achieved when solving the least squares problem $AQ_{m+1}z = b$. First note that if $v \in \mathcal{K}_{m+1}$ then $\|Q_{m+1}^T v\| = \|v\|$. To see why this is true, let $v \in K_{m+1}$. Since the column space $C(Q_{m+1}) = K_{m+1}$ it follows that there exists $w \in \mathbf{R}^{m+1}$ such that $Q_{m+1} w = v$. Thus $Q_{m+1}^T Q_{m+1} = I$ implies

$$\|Q_{m+1}^T v\| = \|Q_{m+1}^T Q_{m+1} w\| = \|w\|$$

and also

$$\|v\| = \|Q_{m+1} w\| = \sqrt{w^T Q_{m+1}^T Q_{m+1} w} = \sqrt{w^T w} = \|w\|.$$

Therefore $\|Q_{m+1}^T v\| = \|v\|$.

Since $A\Gamma_m \in \mathcal{K}_{m+1}$ then $AQ_m \in K_{m+1}$ and so $AQ_m z - b \in \mathcal{K}_{m+1}$. Moreover, since $q_1 = b/\|b\|$, then $Q_{m+1}^T b = \|b\| e_1$ where $e_1 \in \mathbf{R}^{m+1}$ is the vector with first component equal 1 and the rest zero. It follows that

$$\big\|AQ_m z - b\big\| = \big\|Q_{m+1}^T A Q_m z - Q_{m+1}^T b\big\| = \big\|H_m z - \|b\| e_1\big\|$$

where $H_m = Q_{m+1}^T A Q_m$. Thus, solving the least squares problem $AQ_m z = b$ is equivalent to solving the least squares problem $H_m z = \|b\| e_1$.

By definition H_m is the $(m+1) \times m$ matrix consisting of the entries $h_{ij} = q_i^T A q_j$. Since q_i is perpendicular to $\mathcal{K}_{i-1}$ then $Aq_j \in \mathcal{K}_{j+1}$ implies that $q_i^T A q_j = 0$ when $i > j+1$. Therefore the left corner of H_m beneath the lower sub-diagonal is zero. Such matrices are called upper Hessenberg.

We now discuss how to obtain the reduced QR decomposition of H_{m+1} from the reduced QR decomposition of H_m. Since we have already used Q_m for the matrix in the Arnoldi process let us denote the reduced QR decomposition of H_m by $P_m U_m$. Thus $H_m = P_m U_m$ where P_m is a $(m+1) \times m$ matrix with orthonormal columns and U_m is a $m \times m$ upper triangular matrix. Using the modified Gram-Schmidt algorithm to form P_m

preserves the zeros in H_m. Thus P_m is also upper Hessenberg. In particular, $H_m = P_m U_m$ has a Wilkinson diagram that looks like

$$\begin{bmatrix} \mathrm{x} & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ \mathrm{x} & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ 0 & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ 0 & 0 & \mathrm{x} & \cdots & \mathrm{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathrm{x} \end{bmatrix} = \begin{bmatrix} \mathrm{x} & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ \mathrm{x} & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ 0 & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ 0 & 0 & \mathrm{x} & \cdots & \mathrm{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathrm{x} \end{bmatrix} \begin{bmatrix} \mathrm{x} & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ 0 & \mathrm{x} & \mathrm{x} & \cdots & \mathrm{x} \\ 0 & 0 & \mathrm{x} & \cdots & \mathrm{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathrm{x} \end{bmatrix}.$$

Now, it is clear that P_{m+1} may be obtained from P_m by first adding a row of zeros at the bottom and then performing the modified Gram-Schmidt algorithm on the vector

$$h_{m+1} = \begin{bmatrix} q_1^T A q_{m+1} \\ \vdots \\ q_{m+2}^T A q_{m+1} \end{bmatrix}$$

to obtain the $m+1$ column of P_{m+1}. In this way the reduced QR decomposition of H_{m+1} may be obtained by extending the reduced QR decomposition of H_m by one additional row and one additional column. A Matlab code for GMRES using these techniques is

Matlab Example 23a

```
function x=krylov(A,b,mmax,tol)
    P=[]; b1=norm(b); e1=1;
    Q(:,1)=b/b1;
    for m=1:mmax
        t=A*Q(:,m);
        for j=1:m
           t=t-(Q(:,j)'*t)*Q(:,j);
        end
        Q(:,m+1)=t/norm(t);
        e1(m+1,1)=0.0;
        P(m+1,:)=0.0;
        t=Q'*A*Q(:,m);
        for j=1:m-1
            U(j,m)=P(:,j)'*t;
            t=t-U(j,m)*P(:,j);
        end
        U(m,m)=norm(t);
        P(:,m)=t/U(m,m);
        rnorm=norm((P*P(1,:)'-e1)*b1);
        disp(sprintf('norm(m=%d)=%g',m,rnorm));
        if rnorm<=tol; break; end
    end
    z=(U\P(1,:)')*b1;
    x=Q(:,1:m)*z;
end
```