

# Why restarting Arnoldi/Lanczos is not trivial

Steven G. Johnson, MIT Applied Mathematics, course 18.335

March 21, 2019

## Overview

The Arnoldi (and Lanczos) algorithms iteratively construct an  $m \times n$  orthonormal basis  $Q_n$  for the Krylov space  $\mathcal{K}_n = \text{span}\{b, Ab, \dots, A^{n-1}b\}$  for a starting vector  $b$  and an  $m \times m$  matrix  $A$ . For very large  $m$  (e.g. huge sparse  $A$ ), however, practical application of this algorithm eventually hits a maximum  $n$  where one runs out of memory for  $Q_n$  [and the computational cost also grows as  $\Theta(mn^2)$ ]. Even Lanczos runs into this problem, because roundoff errors lead to “ghost eigenvalues” if one does not explicitly store  $Q_n$  and periodically re-orthogonalize. The solution is periodic **restarting**: shrink down to a  $k$ -dimensional subspace containing your “best guesses” for the solution vectors, and continue Arnoldi from there. It turns out that the algorithms to perform restarting properly—called **implicitly restarted** Arnoldi or Lanczos—are surprisingly complicated and subtle. The purpose of these notes is not to explain how implicit restarting works. Rather, it is to briefly explain why naive restarting methods *don't* work. That is, why is restarting so hard?

## Restarting in general

In general, restarting means finding a smaller orthonormal basis

$$\underbrace{\hat{Q}_k}_{m \times k} = \underbrace{Q_n}_{m \times n} \underbrace{\hat{Q}}_{n \times k}$$

for a subspace  $\hat{\mathcal{K}}_k \subset \mathcal{K}_n$  ( $k < n$ ), where  $\hat{Q}^* \hat{Q} = I \implies \hat{Q}_k^* \hat{Q}_k = I$ , and then treating this as the  $k$ -th step of an Arnoldi process and continuing from there normally (until you restart again). However, for this to work,  $\hat{Q}_k$  needs to preserve a **key property** of the Arnoldi process:

$$AQ_n = Q_n H_n + r_n e_n^*,$$

where  $H_n = Q_n^* A Q_n$  is upper-Hessenberg  $n \times n$ ,  $r_n = h_{n+1,n} q_{n+1} \perp \mathcal{K}_n$ , and  $e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$ . This is the property that allows subsequent steps to continue the upper-Hessenberg property (which for Lanczos is tridiagonal and

crucial to its three-term recurrence structure). Hence, we would like to obtain the *same* structure for  $A\hat{Q}_k$

To compute  $A\hat{Q}_k$ , it is convenient to define the  $n \times n$  unitary matrix  $Q = \begin{bmatrix} \hat{Q} & \hat{Q}_\perp \end{bmatrix}$  where the  $n \times (n - k)$  matrix  $\hat{Q}_\perp$  is any orthonormal basis for the orthogonal complement of  $\hat{Q}$ 's column space. Then we can write

$$\begin{aligned} A\hat{Q}_k &= AQ_n\hat{Q} = Q_n H_n \hat{Q} + r_n e_n^* \hat{Q} \\ &= Q_n Q Q^* H_n \hat{Q} + r_n e_n^* \hat{Q} \\ &= \begin{bmatrix} \hat{Q}_k & A\hat{Q}_\perp \end{bmatrix} \begin{bmatrix} \hat{Q}^* H_n \hat{Q} \\ \hat{Q}_\perp^* H_n \hat{Q} \end{bmatrix} + r_n e_n^* \hat{Q} \\ &= \boxed{\hat{Q}_k \underbrace{(\hat{Q}^* H_n \hat{Q})}_{\hat{H}_k} + A\hat{Q}_\perp \hat{Q}_\perp^* H_n \hat{Q} + r_n e_n^* \hat{Q}}. \end{aligned}$$

This looks messy, but we can simplify it quite a bit if we make a good choice for  $\hat{Q}$ . With the *right* choice of  $\hat{Q}$ , in fact it *is* possible to have the  $\hat{Q}_k \hat{H}_k + \hat{r}_k e_k^*$  structure, allowing us to restart Arnoldi and Lanczos, but finding such a  $\hat{Q}$  is surprisingly subtle.

## Naive restarting

The most obvious way to restart is to use Ritz vectors. Recall the Rayleigh–Ritz procedure: search for  $x \in \mathcal{K}_n$  and  $\nu \in \mathbb{C}$  such that  $Ax - \nu x \perp \mathcal{K}_n$ , or equivalently  $x = Q_n z$  where  $H_n z = \nu z$ . This is how we estimate the eigenvectors and eigenvalues at the  $n$ -th step of Arnoldi. It seems natural that we should want our “restarted” basis  $\hat{Q}_k$  to contain the Ritz vectors  $x = Q_n z$  that are our best estimates so far for the desired eigenvectors. For example, suppose we are looking for the  $k$  biggest- $|\lambda|$  eigenvalues, then a natural choice of restarting basis would be the Ritz vectors  $Q_n Z$  corresponding to the biggest  $|\nu|$ . If we orthogonalize these via QR as  $Z = \hat{Q}R$ , we get

$$\hat{Q} = ZR^{-1}$$

and

$$H_n \hat{Q} = H_n Z R^{-1} = Z \begin{bmatrix} \nu_1 & & & \\ & \nu_2 & & \\ & & \ddots & \\ & & & \nu_k \end{bmatrix} R^{-1} = Z \Lambda R^{-1} = \hat{Q} \Lambda R^{-1}.$$

Two nice things happen! We get

$$\hat{Q}^* H_n \hat{Q} = \hat{Q}^* \hat{Q} \Lambda R^{-1} = \Lambda R^{-1},$$

which is a product of upper-triangular matrices, and hence is upper triangular. Thus, the first boxed term above is

$$\hat{H}_k = \hat{Q}_k (R \Lambda R^{-1}) = \hat{Q}_k (\text{upper triangular}),$$

which is even stronger than requiring upper Hessenberg! Also

$$\hat{Q}_\perp^* H_n \hat{Q} = \hat{Q}_\perp^* \hat{Q} R \Lambda R^{-1} = 0$$

since  $\hat{Q}_\perp^* \hat{Q} = 0$  by construction. So, the second boxed term above disappears! Unfortunately, the third boxed term is

$$r_n e_n^* \hat{Q} = r_n (\text{last row of } \hat{Q}).$$

While  $r_n \perp \mathcal{K}_n \implies r_n \perp \hat{\mathcal{K}}_n$  (that is,  $\hat{Q}_k^* r_n$ ) as desired, in general the last row of  $\hat{Q}$  will **not** be a multiple of  $e_k^*$ . So, this doesn't work.

The same problem arises for naive restarting of the Lanczos case  $A = A^*$ . In this case, the upper-Hessenberg matrix  $H_n$  is Hermitian. Hence the upper-triangular matrix  $\hat{H}_k = \Lambda$  is diagonal ( $R = I$  since the eigenvectors  $Z$  are orthonormal). But there is still no reason why the last row of  $\hat{Q}$  should be a multiple of  $[0 \ 0 \ \cdots \ 0 \ 1]$ , so it doesn't work.

## Implicit restarting

In fact, it is possible to choose a  $\hat{Q}_k$  such that it *mostly* contains the Ritz vectors that we want and *does* preserve the Arnoldi/Lanczos property. One hint of this is that our naive choice above was actually *too good* in two ways:  $\hat{H}_k$  was upper-triangular instead of just upper-Hessenberg, and  $r_n$  was orthogonal to  $\mathcal{K}_n$  and not just  $\hat{\mathcal{K}}_k$ . This gives us “wiggle room:” if we do a little “worse” in making  $\hat{H}_k$  only upper-Hessenberg and  $\hat{r}_k$  only  $\perp \hat{\mathcal{K}}_k$ , we then have enough freedom to make the last row of last row of  $\hat{Q}$  a multiple of  $e_k^*$ .

In particular, instead of taking the Ritz eigenvectors  $Z$  of  $H_n$ , the solution is to do exactly  $n - k$  steps of shifted QR iteration on  $H_n$  and let  $\hat{Q}$  be the resulting eigenvector/Schur-vector estimate. This is a good estimate for the Ritz eigenvectors that we want, and it turns out to be *just right* to preserve the Arnoldi property. Proving that this is true requires care and tedious calculation, but is straightforward using our existing 18.335 tools. However, I won't go through it in detail. If you google “implicitly restarted Arnoldi” or “implicitly restarted Lanczos” you can find a number of reviews that go through the algebra, but in practice you are unlikely to ever need to know the details, since most people use “canned” implementations of Arnoldi and Lanczos such as ARPACK. What is important is to **resist the temptation to do naive restarting!**