this is not what we are after, since m is very large an we are only interested in a "small" number of eigenvalues of A.

There is another aspect of Arnoldi iteration, which is that it solves an optimization problem involving polynomials of degree n. Let $\mathcal{P}^n$ denote the set of (complex) monic polynomials of degree n, that is, polynomials of the form

$$p(z) = z^n + c_{n-1}z^{n-1} + \cdots + c_1 z + c_0 \quad (c_i \in \mathbb{C}).$$

For any $m \times m$ matrix A, we write

$$p(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I.$$

The following result is proven in Trefethen and Bau [176] (Lecture 34, Theorem 34.1).

Theorem 18.6. *If Arnoldi iteration run on an $m \times m$ matrix A starting with a nonzero vector b does not have a breakdown at stage $n \leq m$, then there is a unique polynomial $p \in \mathcal{P}^n$ such that $\|p(A)b\|_2$ is minimum, namely the characteristic polynomial $\det(zI - H_n)$ of H_n.*

Theorem 18.6 can be viewed as the "justification" for a method to find some of the eigenvalues of A (say $n \ll m$ of them). Intuitively, the closer the roots of the characteristic polynomials of H_n are to the eigenvalues of A, the smaller $\|p(A)b\|_2$ should be, and conversely. In the extreme case where $m = n$, by the Cayley–Hamilton theorem, $p(A) = 0$ (where p is the characteristic polynomial of A), so this idea is plausible, but this is far from constituting a proof (also, b should have nonzero coordinates in all directions associated with the eigenvalues).

The method known as the *Rayleigh–Ritz method* is to run Arnoldi iteration on A and some $b \neq 0$ chosen at random for $n \ll m$ steps before or until a breakdown occurs. Then run the QR algorithm with shifts on H_n. The eigenvalues of the Hessenberg matrix H_n may then be considered as approximations of the eigenvalues of A. The eigenvalues of H_n are called *Arnoldi estimates* or *Ritz values*. One has to be cautious because H_n is a truncated version of the full Hessenberg matrix H, so not all of the Ritz values are necessarily close to eigenvalues of A. It has been observed that the eigenvalues that are found first are the *extreme* eigenvalues of A, namely those close to the boundary of the spectrum of A plotted in $\mathbb{C}$. So if A has real eigenvalues, the largest and the smallest eigenvalues appear first as Ritz values. In many problems where eigenvalues occur, the extreme eigenvalues are the one that need to be computed. Similarly, the eigenvectors of H_n may be considered as approximations of eigenvectors of A.

The `Matlab` function `eigs` is based on the computation of Ritz values. It computes the six eigenvalues of largest magnitude of a matrix A, and the call is `[V, D] = eigs(A)`. More generally, to get the top k eigenvalues, use `[V, D] = eigs(A, k)`.

In the absence of rigorous theorems about error estimates, it is hard to make the above statements more precise; see Trefethen and Bau [176] (Lecture 34) for more on this subject.