A_{k+1} is similar to A_k, as before. If A_k is upper Hessenberg, then it is easy to see that A_{k+1} is also upper Hessenberg.

If A is upper Hessenberg and if σ_i is exactly equal to an eigenvalue, then $A_k - \sigma_k I$ is singular, and forming the QR-factorization will detect that R_k has some diagonal entry equal to 0. Assuming that the QR-algorithm returns $(R_k)_{nn} = 0$ (if not, the argument is easily adapted), then the last row of $R_k Q_k$ is 0, so the last row of $A_{k+1} = R_k Q_k + \sigma_k I$ ends with σ_k (all other entries being zero), so we are in the case where we can deflate A_k (and σ_k is indeed an eigenvalue).

The question remains, what is a good choice for the shift σ_k?

Assuming again that H is in upper Hessenberg form, it turns out that when $(H_k)_{nn-1}$ is small enough, then a good choice for σ_k is $(H_k)_{nn}$. In fact, the rate of convergence is quadratic, which means roughly that the number of correct digits doubles at every iteration. The reason is that shifts are related to another method known as inverse iteration, and such a method converges very fast. For further explanations about this connection, see Demmel [48] (Section 4.4.4) and Trefethen and Bau [176] (Lecture 29).

One should still be cautious that the QR method with shifts does not necessarily converge, and that our convergence proof no longer applies, because instead of having the identity $A^k = P_k \mathcal{R}_k$, we have

$$(A - \sigma_k I) \cdots (A - \sigma_2 I)(A - \sigma_1 I) = P_k \mathcal{R}_k.$$

Of course, the QR algorithm loops immediately when applied to an orthogonal matrix A. This is also the case when A is symmetric but not positive definite. For example, both the QR algorithm and the QR algorithm with shifts loop on the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In the case of symmetric matrices, Wilkinson invented a shift which helps the QR algorithm with shifts to make progress. Again, looking at the lower corner of A_k, say

$$B = \begin{pmatrix} a_{n-1} & b_{n-1} \\ b_{n-1} & a_n \end{pmatrix},$$

the *Wilkinson shift* picks the eigenvalue of B closer to a_n. If we let

$$\delta = \frac{a_{n-1} - a_n}{2},$$

it is easy to see that the eigenvalues of B are given by

$$\lambda = \frac{a_n + a_{n-1}}{2} \pm \sqrt{\delta^2 + b_{n-1}^2}.$$