1. The Lanczos process

- If $\mathbf{A}$ is Hermitian, then $\mathbf{H}_j = \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j$ in the Arnoldi process is also Hermitian. Since $\mathbf{H}_j$ is upper Hessenberg, it is tridiagonal:

$$\mathbf{H}_j = \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j = \begin{bmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & b_3 & a_3 & \ddots & \\ & & \ddots & \ddots & b_j \\ & & & b_j & a_j \end{bmatrix} := \mathbf{T}_j.$$

Note that $\mathbf{T}_j \in \mathbb{R}^{j \times j}$. We have the Lanczos relation

$$\mathbf{A}\mathbf{Q}_j = \mathbf{Q}_{j+1} \widetilde{\mathbf{T}}_j, \quad \text{where} \quad \widetilde{\mathbf{T}}_j := \mathbf{Q}_{j+1}^* \mathbf{A} \mathbf{Q}_j.$$

- Compared with the Arnoldi process, we have

$$a_j = h_{jj}, \quad b_{j+1} = h_{j+1,j} = h_{j,j+1}.$$

- We may use the Lanczos relation to derive the following algorithm.

Algorithm: Lanczos process generating the orthonormal basis

$\mathbf{r}$ = arbitrary nonzero vector, $b_1 = 0$, $\mathbf{q}_0 = \mathbf{0}$
$\mathbf{q}_1 = \mathbf{r}/\|\mathbf{r}\|_2$
for $j = 1, 2, 3, \ldots,$
  $\mathbf{v} = \mathbf{A}\mathbf{q}_j$
  $a_j = \mathbf{q}_j^* \mathbf{v}$
  $\mathbf{v} = \mathbf{v} - b_j \mathbf{q}_{j-1} - a_j \mathbf{q}_j$
  $b_{j+1} = \|\mathbf{v}\|_2$
  $\mathbf{q}_{j+1} = \mathbf{v}/b_{j+1}$
end

---

2. Derivation of conjugate gradient iterations

- Note that the matrix

$$\mathbf{T}_j = \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j = \begin{bmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{j-1} & a_{j-1} & b_j \\ & & & b_j & a_j \end{bmatrix}$$

in the Lanczos process is Hermitian positive definite (since $\mathbf{A}$ is HPD). Hence, $\mathbf{T}_j$ can be LU factorized into

$$\mathbf{T}_j = \mathbf{L}_j \mathbf{U}_j = \begin{bmatrix} 1 & & & & \\ c_2 & 1 & & & \\ & \ddots & \ddots & & \\ & & c_{j-1} & 1 & \\ & & & c_j & 1 \end{bmatrix} \begin{bmatrix} d_1 & b_2 & & & \\ & d_2 & b_3 & & \\ & & \ddots & \ddots & \\ & & & d_{j-1} & b_j \\ & & & & d_j \end{bmatrix}$$

with the recurrences for c_j and d_j:

$$c_j = b_j/d_{j-1}, \quad d_j = \begin{cases} a_1 & \text{if } j = 1, \\ a_j - c_j b_j & \text{if } j > 1. \end{cases}$$

- Assume that $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{y}_j$. By $\mathbf{r}_j \perp \mathcal{K}_j$, i.e., $\mathbf{Q}_j^* \mathbf{r}_j = \mathbf{0}$, we have

$$\mathbf{T}_j \mathbf{y}_j = \|\mathbf{r}_0\|_2 \mathbf{e}_1.$$

Rewrite $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{y}_j$ as

$$\mathbf{x}_j = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{T}_j^{-1}(\|\mathbf{r}_0\|_2 \mathbf{e}_1) = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{U}_j^{-1} \mathbf{L}_j^{-1}(\|\mathbf{r}_0\|_2 \mathbf{e}_1).$$

Let

$$\begin{aligned} \mathbf{P}_j &:= \mathbf{Q}_j \mathbf{U}_j^{-1} = \begin{bmatrix} \mathbf{p}_0 & \mathbf{p}_1 & \cdots & \mathbf{p}_{j-1} \end{bmatrix}, \\ \mathbf{z}_j &:= \mathbf{L}_j^{-1}(\|\mathbf{r}_0\|_2 \mathbf{e}_1) = \begin{bmatrix} \zeta_1 & \zeta_2 & \cdots & \zeta_j \end{bmatrix}^\top, \end{aligned}$$

where $\mathbf{p}_0 = \mathbf{q}_1/a_1$, $\zeta_1 = \|\mathbf{r}_0\|_2$ and, for $j \geq 2$,

$$\mathbf{p}_{j-1} = \frac{1}{d_j}(\mathbf{q}_j - b_j \mathbf{p}_{j-2}), \quad \zeta_j = -c_j \zeta_{j-1}.$$

It is now important to observe that (why?)

$$\begin{aligned} \mathbf{P}_j &= \begin{bmatrix} \mathbf{p}_0 & \mathbf{p}_1 & \cdots & \mathbf{p}_{j-1} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{j-1} & \mathbf{p}_{j-1} \end{bmatrix}, \\ \mathbf{z}_j &= \begin{bmatrix} \zeta_1 & \zeta_2 & \cdots & \zeta_j \end{bmatrix}^\top = \begin{bmatrix} \mathbf{z}_{j-1} \\ \zeta_j \end{bmatrix}, \end{aligned} \quad .$$

With this formulation, we arrive at a simple recurrence for $\mathbf{x}_j$:

$$\mathbf{x}_j = \mathbf{x}_0 + \mathbf{P}_j \mathbf{z}_j = \mathbf{x}_0 + \mathbf{P}_{j-1}\mathbf{z}_{j-1} + \zeta_j \mathbf{p}_{j-1} = \mathbf{x}_{j-1} + \zeta_j \mathbf{p}_{j-1}.$$

- The residual $\mathbf{r}_j$ is essentially a multiple of $\mathbf{q}_{j+1}$ (see below for a proof), therefore, all residuals are mutually orthogonal.

In fact, we have $\mathbf{r}_0 = \|\mathbf{r}_0\|_2 \mathbf{q}_1$ and, for $j \geq 1$,

$$\begin{aligned}
\mathbf{r}_j &= \mathbf{b} - \mathbf{A}\mathbf{x}_j = \mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{Q}_j \mathbf{y}_j) \\
&= \mathbf{r}_0 - \mathbf{A}\mathbf{Q}_j \mathbf{y}_j = \mathbf{r}_0 - \mathbf{Q}_{j+1} \widetilde{\mathbf{T}}_j \mathbf{y}_j \\
&= \mathbf{r}_0 - \mathbf{Q}_j \mathbf{T}_j \mathbf{y}_j - b_{j+1}(\mathbf{e}_j^* \mathbf{y}_j)\mathbf{q}_{j+1} \\
&= \|\mathbf{r}_0\|_2 \mathbf{q}_1 - \mathbf{Q}_j(\|\mathbf{r}_0\|_2 \mathbf{e}_1) - b_{j+1}(\mathbf{e}_j^* \mathbf{y}_j)\mathbf{q}_{j+1} \\
&= -b_{j+1}(\mathbf{e}_j^* \mathbf{y}_j)\mathbf{q}_{j+1}.
\end{aligned}$$

- If we allow $\mathbf{p}_{j-1}$ to scale and compensate for the scaling in the scalars, we potentially can have simpler recurrences of the form: $\mathbf{p}_0 = \mathbf{r}_0$ and for $j \geq 1$,

$$\begin{aligned}
\mathbf{x}_j &= \mathbf{x}_{j-1} + \alpha_j \mathbf{p}_{j-1}, \\
\mathbf{r}_j &= \mathbf{r}_{j-1} - \alpha_j \mathbf{A}\mathbf{p}_{j-1}, \\
\mathbf{p}_j &= \mathbf{r}_j + \beta_j \mathbf{p}_{j-1}.
\end{aligned}$$

- Note that at present we have

$$\mathbf{P}_j = \begin{bmatrix} \mathbf{p}_0 & \mathbf{p}_1 & \cdots & \mathbf{p}_{j-1} \end{bmatrix} = \mathbf{Q}_j \mathbf{U}_j^{-1} \mathbf{D}_j,$$

where $\mathbf{D}_j$ is diagonal with scaling parameters as diagonal entries. We now derive the $\mathbf{A}$-conjugacy of $\mathbf{p}_j$, i.e., for $i < j$,

$$\mathbf{p}_i^* \mathbf{A} \mathbf{p}_j = 0.$$

It suffices to show that $\mathbf{P}_j^* \mathbf{A} \mathbf{P}_j$ is diagonal. Since

$$\begin{aligned} \mathbf{P}_j^* \mathbf{A} \mathbf{P}_j &= \mathbf{D}_j^* \mathbf{U}_j^{-*} \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j \mathbf{U}_j^{-1} \mathbf{D}_j \\ &= \mathbf{D}_j^* \mathbf{U}_j^{-*} \mathbf{T}_j \mathbf{U}_j^{-1} \mathbf{D}_j \\ &= \mathbf{D}_j^* \mathbf{U}_j^{-*} \mathbf{L}_j \mathbf{D}_j \end{aligned}$$

is Hermitian and lower triangular simultaneously, then $\mathbf{P}_j^* \mathbf{A} \mathbf{P}_j$ must be diagonal.

- Now we can derive the scalar factors α_j and β_j by solely imposing the orthogonality of $\mathbf{r}_j$ and $\mathbf{A}$-conjugacy of $\mathbf{p}_j$. Due to the orthogonality of $\mathbf{r}_j$, it is necessary that

$$\mathbf{r}_{j-1}^*\mathbf{r}_j = \mathbf{r}_{j-1}^*(\mathbf{r}_{j-1} - \alpha_j \mathbf{A}\mathbf{p}_{j-1}) = 0.$$

As a result,

$$\alpha_j = \frac{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}{\mathbf{r}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}} = \frac{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}{(\mathbf{p}_{j-1} - \beta_{j-1}\mathbf{p}_{j-2})^*\mathbf{A}\mathbf{p}_{j-1}} = \frac{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}{\mathbf{p}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}}.$$

Similarly, due to the $\mathbf{A}$-conjugacy of $\mathbf{p}_j$, it is necessary that

$$\mathbf{p}_j^*\mathbf{A}\mathbf{p}_{j-1} = (\mathbf{r}_j + \beta_j\mathbf{p}_{j-1})^*\mathbf{A}\mathbf{p}_{j-1} = 0.$$

As a result,

$$\beta_j = -\frac{\mathbf{r}_j^*\mathbf{A}\mathbf{p}_{j-1}}{\mathbf{p}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}} = -\frac{\mathbf{r}_j^*(\mathbf{r}_{j-1} - \mathbf{r}_j)}{\alpha_j\mathbf{p}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}} = \frac{\mathbf{r}_j^*\mathbf{r}_j}{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}.$$