1 Introduction and Notation

- Eigenvalue Problem: \( Ax = \lambda x \), \( A \in \mathbb{C}^{N \times N} \), \( x \in \mathbb{C}^N \)
- Now \( \lambda \in \mathbb{R} \) since \( A = A^T \).
- Vector \( q_i \) is orthonormal if
  1. \( q_i^T q_i = 1 \),
  2. \( Q^T = Q^{-1} \), \( Q = [q_1, \ldots, q_N] \),
  3. \( ||q_i||_2 = 1 \),
  4. \( q_i^T q_j = 0 \) for \( i \neq j \),

2 A Reminder of the Power Method

- We recall the Power Method is used to find the eigenvector associated with the maximum eigenvalue.
- Simply \( x^{k+1} = cAx^k \), where \( c \) is a normalisation constant to prevent large \( x^{k+1} \).
- As \( k \to \infty \), \( x^{k+1} \to v_1 \), the eigenvector associated with eigenvalue \( \lambda_1 \) where \( \lambda_1 > \lambda_2 \geq \lambda_3 \ldots \lambda_N \).
- We obtain the maximum eigenvalue by the Rayleigh Quotient
  \[
  R(A, x^k) = \frac{(x^k)^T Ax^k}{||x^k||_2^2}
  \]
- Why don’t we just use the QR method? Well if \( A \) is sparse, then applying an iteration of the QR approach does not maintain sparsity of the new matrix. INEFFICIENT.
- Note: We only find ONE eigenvector and eigenvalue. What if we want more?

3 The Idea Behind Lanczos Method

- Lets follow the Power Method, but save each iteration, such that we obtain
  \( v, A v, A^2 v, \ldots, A^{k-1} v \)
- These vectors form the Krylov Space
  \[
  K_k(Av) = \text{span} \{v, Av, A^2 v, \ldots, A^{k-1} v\}
  \]
- So after \( n \) iterations
  \( v, Av, \ldots, A^{n-1} v \)
  are linearly independent and \( x \) can be formed from the space.
- By the Power Method, the n-th iteration tends to an eigenvector hence the sequence becomes linearly dependent but we want a sequence of linearly independent vectors.
- Hence we orthogonalise the vectors, this is the basis of Lanczos Method
4 Lanczos Method

- Assume we have orthonormal vectors $q_1, q_2, \ldots, q_N$

- Simply let $Q = [q_1, q_2, \ldots, q_k]$ hence
  $$Q^TQ = I$$

- We want to change $A$ to a tridiagonal matrix $T$, and apply a similarly transformation:
  $$Q^T AQ = T \text{ or } AQ = QT$$

- So we define $T$ to be
  $$T_{k+1,k} = \begin{bmatrix}
    \alpha_1 & \beta_1 & 0 & \ldots & \ldots & 0 \\
    \beta_1 & \alpha_2 & \beta_2 & 0 & \ldots & 0 \\
    0 & \beta_2 & \alpha_3 & \beta_3 & 0 & \ldots \\
    \vdots & 0 & \ldots & \ldots & \ldots & \vdots \\
    \vdots & \vdots & \ldots & \ldots & \ldots & \beta_{k-1} \\
    0 & \ldots & \ldots & 0 & \beta_{k-1} & \alpha_k \\
    0 & \ldots & \ldots & 0 & 0 & \beta_k
  \end{bmatrix} \in \mathbb{C}^{k+1,k}$$

- After $k$ steps we have $AQ_k = Q_{k+1}T_{k+1,k}$ for $A \in \mathbb{C}^{N,N}$, $Q_k \in \mathbb{C}^{N,k}$, $Q_{k+1} \in \mathbb{C}^{N,k+1}$, $T_{k+1,k} \in \mathbb{C}^{k+1,k}$.

- We observe that
  $$AQ_k = Q_{k+1}T_{k+1,k} = Q_kT_{k,k} + \beta_k q_{k+1}e_k^T$$

- Now $AQ = QT$ hence
  $$A[q_1, q_2, \ldots, q_k] = [q_1, q_2, \ldots, q_k]T_k$$

- The first column of the left hand side matrix is given by
  $$Aq_1 = \alpha_1 q_1 + \beta_1 q_2$$

- The $i$th term by
  $$Aq_i = \beta_{i-1} q_{i-1} + \alpha_i q_i + \beta_i q_{i+1}, \quad i = 2, \ldots$$

- We wish to find the alphas and betas so multiply $q_i^T$ by $q_i^T$ so that
  $$q_i^T Aq_i = q_i^T \beta_{i-1} q_{i-1} + q_i^T \alpha_i q_i + q_i^T \beta_i q_{i+1}$$
  $$= \beta_{i-1} q_i^T q_{i-1} + \alpha_i q_i^T q_i + \beta_i q_i^T q_{i+1}$$
  $$= \alpha_i q_i^T q_i$$

- We obtain $\beta_i$ by rearranging $\beta_i$ from the recurrence formula
  $$r_i = \beta_i q_{i+1} = Aq_i - \alpha_i q_i - \beta_{i-1} q_{i-1}$$

- We assume $\beta_i \neq 0$ and so $\beta_i = ||r_i||_2$.

- We may now determine the next orthonormal vector
  $$q_{i+1} = \frac{r_i}{\beta_i}.$$
5 A Little Proof - Omit from Seminar

Lemma: All vectors $q_{i+1}$ generated by the 3-term are orthogonal to all $q_k$ for $k < i$

Proof

- We assume $q_{i+1}^T q_i = 0 = q_{i+1}^T q_{i-1}$ and by induction step $q_i^T q_k$ for $k < i$.
- We prove $q_{i+1}^T q_k$ for $k < i$.
- Multiply $\dagger$ by $q_k$ for $k \leq i - 2$ and we show $q_k, q_i$ are $A$ orthogonal. Hence

$$q_k^T A q_i = (q_k^T A^T) q_i = (AQ_k)^T q_i$$
$$= (Q_k^T q_{k-1} + \alpha_i q_k + \beta_i q_k) q_i$$
$$= \beta_i q_{k-1}^T q_i + \alpha_i q_k^T q_i + \beta_i q_k^T q_i$$
$$= 0 + 0 + 0 = 0$$

- Now multiply $\dagger$ by $q_k$ so that

$$q_k^T A q_i = \beta_i q_{i-1}^T q_i + \alpha_i q_k^T q_i + \beta_i q_k^T q_{i+1}$$

Rearranging we obtain

$$\beta_i q_{i-1}^T q_{i+1} = q_k^T A q_i - \beta_i q_i^T q_{i-1} - \alpha_i q_k^T q_k = 0$$

6 The Lanczos Algorithm

Initialise: choose $r = q_0$ and let $\beta_0 = ||q_0||_2$

Begin Loop: for $j = 1, \ldots$

- $q_j = \frac{r}{\beta_j}$
- $r = Aq_j$
- $r = r - q_{j-1} \beta_{j-1}$
- $\alpha_j = q_j^T r$

Orthogonalise if necessary

$\beta_j = ||r||_2$

Compute approximate eigenvalues of $T_j$

Test Convergence (see remarks)

End Loop

7 Remarks 1: Finding the Eigenvalues and Eigenvectors

- So how do we find the eigenvalues and eigenvectors?
- If $\beta_k = 0$ then
  1. We diagonalise the matrix $T_k$ using simple QR method to find the exact eigenvalues.

$$T_k = S_k \text{diag}(\lambda_1, \ldots, \lambda_k) S_k^T$$

where the matrix $S_k$ is orthonormal $S_k S_k^T = I$. 
2. The exact eigenvectors are given correspondingly in the columns of the matrix $Y$ where

$$S_k^T Q_k^T A Q_k S_k = \text{diag} (\lambda_1, \ldots, \lambda_k)$$

so that $Y = Q_k S_k$.

3. We converge to the $k$ largest eigenvalues. The proof is very difficult and is omitted.

- Now $\beta_k$ is never really zero. Hence we only converge to the eigenvalue.
  - After $k$ steps we have $A Q_k = Q_k T_{k,k} + \beta_k q_{k+1} e_k^T$
  - For $\beta_k$ small we obtain approximations to the eigenvalues $\theta_i \approx \lambda_i$ by
    $$T_k = S_k \text{diag} (\theta_1, \ldots, \theta_k) S_k^T$$

- We multiply $A Q_k$ by $S_k$ from above so that

$$A Q_k S_k = Q_k T_{k,k} S_k + \beta_k q_{k+1} e_k^T S_k$$

$$A Y_k = Y_k \text{diag} (\theta_1, \ldots, \theta_k) + \beta_k q_{k+1} e_k^T S_k$$

$$A y_j = y_j \theta_j + \beta_k q_{k+1} S_{kj}$$

$$\therefore ||A y_j - \theta_j y_j|| = ||\beta_k|| S_{kj}$$

- So if $\beta_k \to 0$ we prove $\theta_j \to \lambda_j$.
- Otherwise $||\beta_k|| S_{kj}$ needs to be small to have a good approximation, hence convergence criterion

$$||\beta_k|| S_{kj} < \epsilon$$

8 Remarks 2: Difficulties with Lanzcos Method

- In practice, the problem is that the orthogonality is not preserved.
- As soon as one eigenvalue converges all the basis vectors $q_i$ pick up perturbations biased toward the direction of the corresponding eigenvector and orthogonality is lost.
- A “ghost” copy of the eigenvalue will appear again in the tridiagonal matrix $T$.
- To counter this we fully re-orthonormalize the sequence by using Gram-Schmidt or even QR.
- However, either approach would be expense if the dimension if the Krylov space is large.
- So instead a selective re-orthonormalization is pursued. More specifically, the practical approach is to orthonormalize half-way i.e., within half machine-recision $\sqrt{\epsilon M}$.
- If the eigenvalues of $A$ are not well separated, then we can use a shift and employ the matrix

$$(A - \sigma I)^{-1}$$

following the shifted inverted power method to generate the appropriate Krylov subspaces.