1 Power iteration methods and Krylov subspaces

1.1 Power iteration

Power method is an iterative method to calculate an eigenvalue and the corresponding eigenvector of a (real symmetrix) matrix A using the *power iteration*

$$\mathbf{x}_{i+1} = \mathrm{A}\mathbf{x}_i \ . \tag{1}$$

The iteration converges to the eigenvector with the largest eigenvalue. Indeed, according to the spectral theorem the eigenvectors $\mathbf{v}_i$,

$$\mathrm{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \tag{2}$$

of a real symmetric matrix A form an orthogonal basis such that any vector $\mathbf{x}_0$ can be represented as a linear combination of the eigenvectors,

$$\mathbf{x}_0 = \sum_k c_k \mathbf{v}_k \ . \tag{3}$$

Acting on $\mathbf{x}_0$ with the matrix A gives

$$\mathrm{A}^i \mathbf{x}_0 = \sum_k \lambda_k^i c_k \mathbf{v}_k \ . \tag{4}$$

Thus the contribution from the eigenvector corresponding to the largest eigenvalue is amplified with the factor

$$\left(\frac{\lambda_{\text{largest}}}{\lambda_{\text{next largest}}}\right)^i \ . \tag{5}$$

The eigenvalue can be estimated using the *Rayleigh quotient*,

$$\lambda[\mathbf{x}_i] = \frac{\mathbf{x}_i^T \mathrm{A} \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i} = \frac{\mathbf{x}_{i+1}^T \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i}. \tag{6}$$

1.2 Inverse iteration

Alternatively, the *inverse power iteration* with the inverse matrix,

$$\mathbf{x}_{i+1} = \mathrm{A}^{-1} \mathbf{x}_i \ , \tag{7}$$

converges to the smallest (in the absolute value) eigenvalue of matrix A.

Finally, the *shifted inverse power iteration*,

$$\mathbf{x}_{i+1} = (\mathrm{A} - s1)^{-1} \mathbf{x}_i \ , \tag{8}$$

where 1 signifies the identity matrix of the same size as A, converges to the eigenvalue closest to the given number s.

The *inverse iteration method* is a refinement of the inverse power method where the trick is not to invert the matrix in (8) but rather solve the linear system

$$(\mathrm{A} - s1)\mathbf{x}_{i+1} = \mathbf{x}_i \tag{9}$$

using e.g. QR-decomposition.

The better approximation s to the sought eigenvalue is chosen, the faster convergence one gets. However, incorrect choice of s can lead to slow convergence or to the convergence to a different eigenvector. In practice the method is usually used when good approximation for the eigenvalue is known, and hence one needs only few (quite often just one) iteration.

One can update the estimate for the eigenvalue using the Rayleigh quotient $\lambda[\mathbf{x}_i]$ after each iteration and get faster convergence for the price of $O(n^3)$ operations per QR-decomposition; or one can instead make more iterations (with $O(n^2)$ operations per iteration) using the same matrix $(\mathrm{A} - s1)$. The optimal strategy is probably an update after several iterations.

1.3 Krylov subspaces

When calculating an eigenvalue of a matrix A using the power method, one starts with an initial random vector $\mathbf{b}$ and then computes iteratively the sequence $\mathrm{A}\mathbf{b}, \mathrm{A}^2\mathbf{b}, \ldots, \mathrm{A}^{n-1}\mathbf{b}$ normalising and storing the result in $\mathbf{b}$ on each iteration. The sequence converges to the eigenvector of the largest eigenvalue of A.

The set of vectors

$$\mathcal{K}_n = \left\{\mathbf{b}, \mathrm{A}\mathbf{b}, \mathrm{A}^2\mathbf{b}, \ldots, \mathrm{A}^{n-1}\mathbf{b}\right\} , \tag{10}$$

where $n < \mathrm{rank}(\mathrm{A})$, is called the order-$n$ *Krylov matrix*, and the subspace spanned by these vectors is called the order-n *Krylov subspace* [**?**]. The vectors are not orthogonal but can be made so e.g. by Gram-Schmidt orthogonalisation.

Krylov subspaces are the basis of several successful iterative methods in numerical linear algebra, in particular: Arnoldi and Lanczos methods for finding one (or a few) eigenvalues of a matrix; and GMRES (Generalised Minimum RESidual) method for solving systems of linear equations.

These methods are particularly suitable for large sparse matrices as they avoid matrix-matrix operations but rather multiply vectors by matrices and work with the resulting vectors and matrices in Krylov subspaces of modest sizes.

1.4 Arnoldi iteration

Arnoldi iteration is an algorithm where the order-n orthogonalised Krylov matrix Q_n for a given matrix A is built using stabilised Gram-Schmidt process [**?**]:

```
start with a set Q = {q_1} where q_1 is a random normalised vector;
repeat for k = 2 to n :
   make a new vector q_k = Aq_{k-1}
```

```
orthogonalise q_k to all vectors q_i ∈ Q storing q_i†q_k → h_{i,k-1}
normalise q_k storing ||q_k|| → h_{k,k-1}
add q_k to the set Q
```

By construction the matrix H_n made of the elements h_{jk} is an upper Hessenberg matrix,

$$\mathrm{H}_n = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} & \cdots & h_{1,n} \\ h_{2,1} & h_{2,2} & h_{2,3} & \cdots & h_{2,n} \\ 0 & h_{3,2} & h_{3,3} & \cdots & h_{3,n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{n,n-1} & h_{n,n} \end{bmatrix}, \tag{11}$$

which is a partial orthogonal reduction of A into Hessenberg form,

$$\mathrm{H}_n = \mathrm{Q}_n^{\dagger} \mathrm{A} \mathrm{Q}_n \,. \tag{12}$$

The matrix H_n can be viewed as a representation of A in the Krylov subspace $\mathcal{K}_n$. The eigenvalues and eigenvectors of the matrix H_n approximate the largest (actually, the extreme) eigenvalues of matrix A.

Since H_n is a Hessenberg matrix of modest size one can relatively easily apply to it the standard algorithms of linear algebra.

In practice if the size n of the Krylov subspace becomes too large the method is restarted.

1.5 Lanczos iteration

Lanczos iteration is Arnoldi iteration for Hermitian matrices [**?**], in which case the Hessenberg matrix H_n of Arnoldi method becomes a tridiagonal matrix T_n.

The Lanczos algorithm thus reduces the original hermitian $N \times N$ matrix A into a smaller $n \times n$ tridiagonal matrix T_n by an orthogonal projection onto the order-n Krylov subspace. The eigenvalues and eigenvectors of a tridiagonal matrix of a modest size can be easily found by e.g. the QR-diagonalisation method.

In practice the Lanczos method is not very stable due to round-off errors leading to quick loss of orthogonality. The eigenvalues of the resulting tridiagonal matrix may then not be a good approximation to the original matrix. Library implementations fight the stability issues by trying to prevent the loss of orthogonality and/or to recover the orthogonality after the basis is generated.

1.6 Generalised minimum residual (GMRES)

GMRES is an iterative method for the numerical solution of a system of linear equations,

$$\mathrm{A}\mathbf{x} = \mathbf{b} \,, \tag{13}$$

where the exact solution is approximated by the least-squares solution that minimises the residual $\|\mathrm{A}\mathbf{x} - \mathbf{b}\|^2$ in the Krylov subspace $\mathcal{K}_n$ of matrix A.

The original equation is projected on the Krylov subspace,

$$\mathrm{A}\mathbf{x} = \mathbf{b} \;\;\Rightarrow\;\; \mathrm{Q}_n\mathrm{Q}_n^\dagger \mathrm{A}\mathrm{Q}_n\mathrm{Q}_n^\dagger \mathbf{x} = \mathrm{Q}_n\mathrm{Q}_n^\dagger \mathbf{b} \;, \tag{14}$$

which gives a linear system of size-n,

$$\mathrm{H}_n\tilde{\mathbf{x}} = \tilde{\mathbf{b}} \;, \tag{15}$$

where

$$\mathrm{H}_n = \mathrm{Q}_n^\dagger A \mathrm{Q}_n \;,\;\; \tilde{\mathbf{b}} = \mathrm{Q}_n^\dagger \mathbf{b} \;. \tag{16}$$

The Hessenberg equation (15) can be easily solved by one run of Gauss elimination and then a run of back-substitution (like in cubic splines). The approximate solution to the original equation is then given as

$$\mathbf{x} \approx \mathrm{Q}_n\tilde{\mathbf{x}} \;. \tag{17}$$