1 Power iteration methods and Krylov subspaces

1.1 Introduction

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

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

The iteration converges to the eigenvector with the largest eigenvalue.

The eigenvalue can be estimated using the *Rayleigh quotient*

$$\lambda[\mathbf{x}_i] = \frac{\mathbf{x}_i^T \mathbf{A}\mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i} = \frac{\mathbf{x}_{i+1}^T \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i}. \tag{2}$$

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

$$\mathbf{x}_{i+1} = \mathbf{A}^{-1}\mathbf{x}_i \ , \tag{3}$$

converges to the smallest eigenvalue of matrix A.

Finally, the *shifted inverse iteration*,

$$\mathbf{x}_{i+1} = (\mathbf{A} - s\mathbf{1})^{-1}\mathbf{x}_i \ , \tag{4}$$

where $\mathbf{1}$ signifies the identity matrix of the same size as $\mathbf{A}$, 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 (4) but rather solve the linear system

$$(\mathbf{A} - s\mathbf{1})\mathbf{x}_{i+1} = \mathbf{x}_i \tag{5}$$

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 $(\mathbf{A} - s\mathbf{1})$. The optimal strategy is probably an update after several iterations.

1.2 Krylov subspaces

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

The set of vectors

$$\mathcal{K}_n = \left\{\mathbf{b}, \mathbf{Ab}, \mathbf{A}^2\mathbf{b}, \ldots, \mathbf{A}^{n-1}\mathbf{b}\right\} , \tag{6}$$

where $n < \text{rank}(\mathbf{A})$, is called the order-$n$ *Krylov matrix*, and the subspace spanned by these vectors is called the order-n *Krylov subspace* [1]. The vectors are not orthogonal but can be made so e.g. by Gram-Schmidt orthogonalisation.

For the same reason that $\mathbf{A}^{n-1}\mathbf{b}$ approximates the dominant eigenvector one can expect that the other orthogonalised vectors approximate the eigenvectors of the n largest eigenvalues.

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.

1.3 Arnoldi iteration

Arnoldi iteration is an algorithm where the order-n orthogonalised Krylov matrix $\mathbf{Q}_n$ for a given matrix $\mathbf{A}$ is built using stabilised Gram-Schmidt process:

start with a set $\mathbf{Q} = \{\mathbf{q}_1\}$ consisting of one random normalised vector $\mathbf{q}_1$;
repeat for $k = 2$ to n :
  make a new vector $\mathbf{q}_k = \mathbf{A}\mathbf{q}_{k-1}$
  orthogonalise $\mathbf{q}_k$ to all vectors $\mathbf{q}_i \in \mathbf{Q}$ storing $\mathbf{q}_i{}^\dagger\mathbf{q}_k \to h_{i,k-1}$
  normalise $\mathbf{q}_k$ storing $\|\mathbf{q}_k\| \to h_{k,k-1}$
  add $\mathbf{q}_k$ to the set $\mathbf{Q}$

By construction the matrix $\mathbf{H}_n$ made of the elements h_{jk} is an upper Hessenberg matrix,

$$\mathbf{H}_n = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} & \cdots & h_{1,n} \\ h_{2,1} & h_{2,2} & h_{2,3} & \cdots & h_{2,n} \\ 0 & h_{3,2} & h_{3,3} & \cdots & h_{3,n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{n,n-1} & h_{n,n} \end{bmatrix} , \tag{7}$$

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

$$\mathbf{H}_n = \mathbf{Q}_n^\dagger \mathbf{A} \mathbf{Q}_n \, . \tag{8}$$

The matrix $\mathbf{H}_n$ can be viewed as a representation of $\mathbf{A}$ in the Krylov subspace $\mathcal{K}_n$. The eigenvalues and eigenvectors of the matrix $\mathbf{H}_n$ approximate the largest eigenvalues of matrix $\mathbf{A}$.

Since $\mathbf{H}_n$ is a Hessenberg matrix of modest size its eigenvalues can be relatively easily computed with standard algorithms.

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

1.4 Lanczos iteration

Lanczos iteration is Arnoldi iteration for Hermitian matrices, in which case the Hessenberg matrix $\mathbf{H}_n$ of Arnoldi method becomes a tridiagonal matrix $\mathbf{T}_n$.

The Lanczos algorithm thus reduces the original hermitian $N \times N$ matrix $\mathbf{A}$ into a smaller $n \times n$ tridiagonal matrix $\mathbf{T}_n$ 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.

1.5 Generalised minimum residual (GMRES)

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

$$\mathbf{A}\mathbf{x} = \mathbf{b} \, , \tag{9}$$

where the exact solution $\mathbf{x}$ is approximated by the vector $\mathbf{x}_n \in \mathcal{K}_n$ that minimises the residual $\mathbf{A}\mathbf{x}_n - \mathbf{b}$ in the Krylov subspace $\mathcal{K}_n$ of matrix $\mathbf{A}$,

$$\mathbf{x} \approx \mathbf{x}_n \leftarrow \min_{\mathbf{x} \in \mathcal{K}_n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \,. \tag{10}$$

The vector $\mathbf{x}_n \in \mathcal{K}_n$ can be represented as $\mathbf{x}_n = \mathbf{Q}_n \mathbf{y_n}$ where $\mathbf{Q}_n$ is the projector on the space $\mathcal{K}_n$ and $\mathbf{y_n}$ is an n-dimensional vector. Substituting $\mathbf{x}_n \in \mathcal{K}_n$ gives an overdetermined system

$$\mathbf{A}\mathbf{Q}_n \mathbf{y_n} = \mathbf{b} \,, \tag{11}$$

which can be solved by the ordinary least-squares method.

One can also project equation (11) onto Krylov subspace $\mathcal{K}_n$ which gives a square system

$$\mathbf{H}_n \mathbf{y_n} = \mathbf{Q}_n^\dagger \mathbf{b} \,, \tag{12}$$

where $\mathbf{H}_n = \mathbf{Q}^\dagger \mathbf{A} \mathbf{Q}_n$.

References

[1] A.N.Krylov. On the numerical solution of equation by which are determined in technical problems the frequencies of small vibrations of material systems. *Izvestiia Akademii nauk SSSR (in Russian)*, 7(4):491–539, 1931.