LECTURE 13: POWER ITERATIONS

1. Reduction to triangular form

The relevance of eigendecompositions $\boldsymbol{A}=\boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}$ to repeated application of the linear operator $\boldsymbol{A}\in\mathbb{C}^{m\times m}$ as in

$$e^{t\boldsymbol{A}}=\boldsymbol{I}+\frac{1}{1!}t\boldsymbol{A}+\frac{1}{2!}t^2\boldsymbol{A}^2+\cdots=\boldsymbol{X}e^{t\boldsymbol{\Lambda}}\boldsymbol{X}^{-1},$$

suggests that algorithms that construct powers of $\boldsymbol{A}$ might reveal eigenvalues. This is indeed the case and leads to a class of algorithms of wide applicability in scientific computation. First, observe that taking condition numbers gives

$$\mu(\boldsymbol{A})=\mu(\boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1})\leqslant\mu^2(\boldsymbol{X})\,\mu(\boldsymbol{\Lambda})=(|\lambda|_{\max}/|\lambda|_{\min}),$$

where $|\lambda|_{\max}$, $|\lambda|_{\min}$ are the eigenvalues of maximum and minimum absolute values. While these express an intrinsic property of the operator $\boldsymbol{A}$, the factor $\mu^2(\boldsymbol{X})$ is associated with the conditioning of a change of coordinates, and a natural question is whether it is possible to avoid any ill-conditioning associated with a basis set $\boldsymbol{X}$ that is close to linear dependence. The answer to this line of inquiry is given by the following result.

SCHUR THEOREM. *For any $\boldsymbol{A}\in\mathbb{C}^{m\times m}$ there exists $\boldsymbol{Q}$ unitary and $\boldsymbol{T}$ upper triangular such that $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{T}\boldsymbol{Q}^*$.*

Proof. *Proceed by induction, starting from an arbitrary eigenvalue λ and eigenvector $\boldsymbol{x}$. Let $\boldsymbol{u}_1=\boldsymbol{x}/\|\boldsymbol{x}\|$, the first column vector of a unitary matrix $\boldsymbol{U}=[\,\boldsymbol{u}_1\ \ \boldsymbol{V}\,]$. Then*

$$\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}=\begin{bmatrix}\boldsymbol{u}_1^*\\\boldsymbol{V}^*\end{bmatrix}\boldsymbol{A}[\,\boldsymbol{u}_1\ \ \boldsymbol{V}\,]=\begin{bmatrix}\boldsymbol{u}_1^*\\\boldsymbol{V}^*\end{bmatrix}[\,\boldsymbol{A}\boldsymbol{u}_1\ \ \boldsymbol{A}\boldsymbol{V}\,]=\begin{bmatrix}\boldsymbol{u}_1^*\\\boldsymbol{V}^*\end{bmatrix}[\,\lambda\boldsymbol{u}_1\ \ \boldsymbol{A}\boldsymbol{V}\,]=\begin{bmatrix}\lambda_1 & \boldsymbol{b}^*\\\boldsymbol{0} & \boldsymbol{C}\end{bmatrix},$$

with $\boldsymbol{C}\in\mathbb{C}^{(m-1)\times(m-1)}$ that by the inductive hypothesis can be written as $\boldsymbol{C}=\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^*$, with $\boldsymbol{W}$ unitary, $\boldsymbol{S}$ upper triangular. The matrix

$$\boldsymbol{Q}=\boldsymbol{U}\begin{bmatrix}1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{W}\end{bmatrix}$$

is a product of unitary matrices, hence itself unitary. The computation

$$\boldsymbol{Q}^*\boldsymbol{A}\boldsymbol{Q}=\left(\boldsymbol{U}\begin{bmatrix}1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{W}\end{bmatrix}\right)^*\boldsymbol{A}\boldsymbol{U}\begin{bmatrix}1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{W}\end{bmatrix}=\begin{bmatrix}1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{W}^*\end{bmatrix}\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}\begin{bmatrix}1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{W}\end{bmatrix}=\begin{bmatrix}1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{W}^*\end{bmatrix}\begin{bmatrix}\lambda_1 & \boldsymbol{b}^*\\\boldsymbol{0} & \boldsymbol{C}\end{bmatrix}\begin{bmatrix}1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{W}\end{bmatrix}=\begin{bmatrix}\lambda_1 & \boldsymbol{b}^*\\\boldsymbol{0} & \boldsymbol{S}\end{bmatrix}=\boldsymbol{T},$$

then shows that $\boldsymbol{T}$ is indeed triangular. □

The eigenvalues of an upper triangular matrix are simply its diagonal elements, so the Schur factorization is an eigenvalue-revealing factorization.

2. Power iteration for real symmetric matrices

When the operator $\boldsymbol{A}$ expresses some physical phenomenon, the principle of action and reaction implies that $\boldsymbol{A}\in\mathbb{R}^{m\times m}$ is symmetric, $\boldsymbol{A}=\boldsymbol{A}^T$ and has real eigenvalues. Componentwise, symmetry of $\boldsymbol{A}=[a_{ij}]$ implies $a_{ij}=a_{ji}$. Consider $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$, and take the adjoint to obtain $\boldsymbol{x}^T\boldsymbol{A}^T=\bar{\lambda}\boldsymbol{x}^T$, or $\boldsymbol{x}^T\boldsymbol{A}=\bar{\lambda}\boldsymbol{x}^T$ since $\boldsymbol{A}$ is symmetric. Form scalar products $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}^T\boldsymbol{x}$, $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{x}=\bar{\lambda}\boldsymbol{x}^T\boldsymbol{x}$, and subtract to obtain

$$0=(\lambda-\bar{\lambda})\boldsymbol{x}^T\boldsymbol{x}\Rightarrow\lambda=\bar{\lambda}\Rightarrow\lambda\in\mathbb{R},$$

since $\boldsymbol{x}\neq\boldsymbol{0}$, an eigenvector.

Example. Consider a linear array of identical mass-springs. The i^{th} point mass obeys the dynamics

$$m\ddot{x}_i=k(x_{i+1}-x_i)-k(x_i-x_{i-1})=k(x_{i+1}-2x_i+x_{i-1}),$$

expressed in matrix form as $\ddot{\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}$, with $\boldsymbol{A}$ symmetric.

For a real symmetric matrix the Schur theorem states that

$$\boldsymbol{A}=\boldsymbol{A}^T \Rightarrow (\boldsymbol{Q}\boldsymbol{T}\boldsymbol{Q}^T)=\boldsymbol{Q}\boldsymbol{T}^T\boldsymbol{Q}^T \Rightarrow \boldsymbol{T}=\boldsymbol{T}^T,$$

and since a symmetric triangular matrix is diagonal, the Schur factorization is also an eigendecomposition, and the eigenvector matrix $\boldsymbol{Q}$ is a basis, $C(\boldsymbol{Q})=\mathbb{R}^m$.

2.1. The power iteration idea

Assume initially that the eigenvalues are distinct and ordered $|\lambda_1|>|\lambda_2|>\cdots>|\lambda_m|$. Repeated application of $\boldsymbol{A}$ on an arbitrary vector $\boldsymbol{v}=\boldsymbol{Q}\boldsymbol{c}\in\mathbb{R}^m=C(\boldsymbol{Q})$ is expressed as

$$\boldsymbol{A}^n\boldsymbol{v}=(\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T)^n\boldsymbol{Q}\boldsymbol{c}=(\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T)(\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T)\ldots(\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T)\boldsymbol{Q}\boldsymbol{c}=\boldsymbol{Q}\boldsymbol{\Lambda}^n\boldsymbol{c},$$

a linear combination of the columns of $\boldsymbol{Q}$ (eigenvectors of $\boldsymbol{A}$) with coefficients $\boldsymbol{\Lambda}^n\boldsymbol{c}=[\ \lambda_1^n c_1\ \ \lambda_2^n c_2\ \ \ldots\ \ \lambda_m^n c_m\ ]^T$.

- For large enough n, $|\lambda_1|>|\lambda_k|$, $k=2,\ldots,n$, leads to a dominant contribution along the direcion of eigenvector $\boldsymbol{q}_1$

  $$\boldsymbol{A}^n\boldsymbol{v}=\boldsymbol{Q}\boldsymbol{\Lambda}^n\boldsymbol{c}=\lambda_1^n c_1\boldsymbol{q}_1+\cdots+\lambda_m^n c_m\boldsymbol{q}_m\cong\lambda_1^n c_1\boldsymbol{q}_1.$$

  This gives a procedure for finding one eigenvector of a matrix, and the Schur theorem proof suggests a recursive algorithm to find all eigenvalues can be defined.

The sequence of normalized eigenvector approximants $\boldsymbol{v}_n=\boldsymbol{A}^n\boldsymbol{v}/\|\boldsymbol{A}^n\boldsymbol{v}\|$ is linearly convergent at rate $r=|\lambda_2/\lambda_1|$.

2.2. Rayleigh quotient

To estimate the eigenvalue revealed by power iteration, formulate the least squares problem

$$\min_c\|\boldsymbol{A}\boldsymbol{v}-\boldsymbol{v}c\|,$$

that seeks the best approximation of one power iteration $\boldsymbol{A}\boldsymbol{v}$ as a linear combination of the initial vector $\boldsymbol{v}$. Of course, if $\boldsymbol{v}=\boldsymbol{q}$ is an eigenvector, then the solution would be $c=\lambda$, the associated eigenvalue. The projector onto $C(\boldsymbol{v})$ is

$$\boldsymbol{P}=\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}},$$

that when applied to $\boldsymbol{A}\boldsymbol{v}$ gives the equation

$$\boldsymbol{P}\boldsymbol{A}\boldsymbol{v}=\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}\boldsymbol{A}\boldsymbol{v}=\frac{\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^T\boldsymbol{v}}\boldsymbol{v}=c\boldsymbol{v}\Rightarrow c=\frac{\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^T\boldsymbol{v}}.$$

The function $r:\mathbb{R}^m\to\mathbb{R}$,

$$r(\boldsymbol{v})=\frac{\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^T\boldsymbol{v}},$$

is known as the Rayleigh quotient which, evaluated for an eigenvector, gives $r(\boldsymbol{q})=\lambda$. To determine how well the eigenvalue is approximated, carry out a Taylor series in the vicinity of an eigenvector $\boldsymbol{q}$

$$r(\boldsymbol{v})=r(\boldsymbol{q})+\frac{1}{1!}[\nabla_{\boldsymbol{v}}r(\boldsymbol{q})]^T(\boldsymbol{v}-\boldsymbol{q})+\mathcal{O}(\|\boldsymbol{v}-\boldsymbol{q}\|^2),$$

where $\nabla_{\boldsymbol{v}}r$ is the gradient of $r(\boldsymbol{v})$

$$\nabla_{\boldsymbol{v}}r=\begin{bmatrix}\dfrac{\partial r}{\partial v_1}\\ \vdots\\ \dfrac{\partial r}{\partial v_m}\end{bmatrix}.$$

Compute the gradient through differentiation of the Rayleigh quotient

$$\nabla_{\boldsymbol{v}}r(\boldsymbol{v})=\frac{\nabla_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v})}{\boldsymbol{v}^T\boldsymbol{v}}-\frac{(\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v})}{(\boldsymbol{v}^T\boldsymbol{v})^2}\nabla_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{v}).$$

Noting that $\nabla_{\boldsymbol{v}}v_i=\boldsymbol{e}_i$, the i^{th} column of $\boldsymbol{I}$, the gradient of $\boldsymbol{v}^T\boldsymbol{v}$ is

$$\nabla_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{v})=\nabla_{\boldsymbol{v}}\sum_{i=1}^m v_i^2=\sum_{i=1}^m\nabla_{\boldsymbol{v}}v_i^2=\sum_{i=1}^m 2v_i\nabla_{\boldsymbol{v}}v_i=2\sum_{i=1}^m v_i\boldsymbol{e}_i=2\boldsymbol{v}.$$

To compute $\nabla_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v})$, let $\boldsymbol{u}=\boldsymbol{A}\boldsymbol{v}$, and since $\boldsymbol{A}$ is symmetric $\boldsymbol{u}^T=\boldsymbol{v}^T\boldsymbol{A}^T=\boldsymbol{v}^T\boldsymbol{A}$, leading to

$$\nabla_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v})=\nabla_{\boldsymbol{v}}(\boldsymbol{u}^T\boldsymbol{v})=\sum_{i=1}^m \nabla_{\boldsymbol{v}}(u_i v_i)=\sum_{i=1}^m u_i \nabla_{\boldsymbol{v}} v_i+\sum_{i=1}^m v_i \nabla_{\boldsymbol{v}} u_i.$$

Use $u_i=\sum_{j=1}^m a_{ij} v_j$ also expressed as $u_j=\sum_{i=1}^m a_{ji} v_i$ by swapping indices to obtain

$$\nabla_{\boldsymbol{v}} u_i=\sum_{j=1}^m a_{ij}\nabla_{\boldsymbol{v}} v_j=\sum_{j=1}^m a_{ij}\boldsymbol{e}_j$$

and therefore

$$\sum_{i=1}^m v_i\nabla_{\boldsymbol{v}} u_i=\sum_{i=1}^m v_i\sum_{j=1}^m a_{ij}\boldsymbol{e}_j=\sum_{j=1}^m\sum_{i=1}^m a_{ij} v_i\boldsymbol{e}_j=\sum_{j=1}^m\sum_{i=1}^m a_{ij} v_i\boldsymbol{e}_j.$$

Use symmetry of $\boldsymbol{A}$ to write

$$\sum_{i=1}^m a_{ij} v_i=\sum_{i=1}^m a_{ji} v_i=u_j,$$

and substitute above to obtain

$$\sum_{i=1}^m v_i\nabla_{\boldsymbol{v}} u_i=\sum_{j=1}^m u_j\boldsymbol{e}_j=\boldsymbol{u}=\boldsymbol{A}\boldsymbol{v}.$$

Gathering the above results

$$\nabla_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{v})=2\boldsymbol{v},\ \nabla_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v})=2\boldsymbol{A}\boldsymbol{v},$$

gives the following gradient of the Rayleigh quotient

$$\nabla_{\boldsymbol{v}} r(\boldsymbol{v})=\frac{2}{\boldsymbol{v}^T\boldsymbol{v}}(\boldsymbol{A}\boldsymbol{v}-r(\boldsymbol{v})\boldsymbol{v}).$$

When evaluated at $\boldsymbol{v}=\boldsymbol{q}$, obtain $\nabla_{\boldsymbol{v}} r(\boldsymbol{q})=\boldsymbol{0}$, implying that near an eigenvector the Rayleigh quotient approximation of an eigenvalue is of quadratic accuracy,

$$r(\boldsymbol{v})-\lambda=\mathcal{O}(\|\boldsymbol{v}-\boldsymbol{q}\|^2).$$

2.3. Refining the power iteration idea

Power iteration furnishes the largest eigenvalue. Further eigenvalues can be found by use of the following properties:

- $(\lambda,\boldsymbol{q})$ eigenpair of $\boldsymbol{A}\Rightarrow(\lambda-\mu,\boldsymbol{q})$ eigenpair of $\boldsymbol{A}-\mu\boldsymbol{I}$;
- $(\lambda,\boldsymbol{q})$ eigenpair of $\boldsymbol{A}\Rightarrow(1/\lambda,\boldsymbol{q})$ eigenpair of $\boldsymbol{A}^{-1}$.

If μ is a known initial approximation of the eigenvalue then the inverse power iteration $\boldsymbol{v}_n=(\boldsymbol{A}-\mu\boldsymbol{I})^{-1}\boldsymbol{v}_{n-1}$, actually implemented as successive solution of linear systems

$$(\boldsymbol{A}-\mu\boldsymbol{I})\boldsymbol{v}_n=\boldsymbol{v}_{n-1},$$

leads to a sequence of Rayleigh quotients $\boldsymbol{r}(\boldsymbol{v}_n)$ that converges quadratically to an eigenvalue close to μ. An important refinement of the idea is to change the shift at each iteration which leads to cubic order of convergence

Algorithm (Rayleigh quotient iteration)

Given $\boldsymbol{v},\boldsymbol{A}$
$\mu=\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}/\boldsymbol{v}^T\boldsymbol{v}$
for $i=1$ to $n_{\max}$
  $\boldsymbol{w}=(\boldsymbol{A}-\mu\boldsymbol{I})\backslash\boldsymbol{v}$ (solve linear system)
  $\boldsymbol{v}=\boldsymbol{w}/\|\boldsymbol{w}\|$
  $\lambda=\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}$
  if $|\lambda-\mu|<\varepsilon$ exit
  $\mu=\lambda$
end
return $\lambda,\boldsymbol{v}$

Power iteration can be applied simultaneously to multiple directions at once

Algorithm (Simultaneous iteration)

Given $\boldsymbol{A}$

$\boldsymbol{Q} = \boldsymbol{I}$; $\boldsymbol{\mu} = \text{diag}(\boldsymbol{A})$

for $i = 1$ to n_{max}

  $\boldsymbol{V} = \boldsymbol{A}\boldsymbol{Q}$ (power iteration applied to multiple directions)

  $\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{V}$ (orthogonalize new directions)

  $\boldsymbol{\lambda} = \text{diag}(\boldsymbol{Q}^T \boldsymbol{A}\boldsymbol{Q})$

  if $\|\boldsymbol{\lambda} - \boldsymbol{\mu}\| < \varepsilon$ exit

end

return $\boldsymbol{\lambda}, \boldsymbol{Q}$