5374 Fall '22

Numerical Linear Algebra Lecture 21

Today: * Eigenvectors: power iteration / QR iteration
* The Singular Value Decomposition

Last time we talked about eigenvectors and the eigenvector decomposition of a square matrix A:

$$A = VDV^{-1}$$

where V is an invertible $n \times n$ matrix and D is a diagonal $n \times n$ matrix.
The columns of V are eigenvectors of A and the elements along the diagonal of D are the eigenvalues of A.

We also learned that if A is symmetric then A has an eigenvector decomposition with V an orthogonal matrix.

Today: How can we compute either an eigenvector of A or a full decomposition?

Literature: Solomon's chapter on eigenvectors has a good practical review of up-to-date algorithms.

Bau-Trefethen: A more in-depth discussion of the underlying theory, and in particular of an important topic called Krylov subspace methods.

Now let's review two basic methods.

1. Power Iteration

Assume A admits a basis of eigenvectors:

$v_1, \ldots, v_n$

with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ s.t.

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$$

Def $\mu := \dfrac{|\lambda_2|}{|\lambda_1|}$ (so $\mu \in (0,1)$)

The power iteration method produces a fast approximation to v_1, and it works as follow:

Given $x \in \mathbb{R}^n$, we know that we can write

$$x = c_1 v_1 + \cdots + c_n v_n \quad \left(\begin{array}{c} \text{for some} \\ c_1, \ldots, c_n \end{array} \right)$$

We know this is so, even if we don't know anything about $v_1, \ldots, v_n$ (other than they exist). Let's assume we have an x such that $c_1 \neq 0$. Then, observe that

$$Ax = \lambda_1 c_1 v_1 + \lambda_2 c_2 v_2 + \cdots + \lambda_n c_n v_n$$

$$\vdots$$

$$A^k x = \lambda_1^k c_1 v_1 + \lambda_2^k c_2 v_2 + \cdots + \lambda_n^k c_n v_n$$

We can write the RHS as:

$$\lambda_1^k \left(c_1 v_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k c_2 v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k c_n v_n \right)$$

The idea is that $\left|\frac{\lambda_2}{\lambda_1}\right|^k \to 0$ as $k \to \infty$ and the same for $\left|\frac{\lambda_j}{\lambda_1}\right|^k$ for $j = 3, 4, \ldots, n$. So for large k, the above vector has almost the same direction as v_1

Observe:

$$\| A^k x - \lambda_1^k c_1 v_1 \|_2 = |\lambda_1|^k \left\| \left(\frac{\lambda_2}{\lambda_1}\right)^k c_2 v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k c_n v_n \right\|$$

$$\le |\lambda_1|^k \sum_{j=2}^{n} \left|\frac{\lambda_j}{\lambda_1}\right|^k |c_j|$$

$$\le |\lambda_1|^k \left|\frac{\lambda_2}{\lambda_1}\right|^k \sum_{j=2}^{n} |c_j|$$

Remark: While we don't know $\sum_{j=2}^{n} |c_j|$ directly, it is independent of k. Moreover, if A is symmetric so that the v_j's are orthonormal, we have

$$\sum_{j=2}^{n} |c_j| \overset{\text{(Cauchy-Schwarz)}}{\le} \left(\sum_{j=2}^{n} 1^2\right)^{1/2} \left(\sum_{j=2}^{n} |c_j|^2\right)^{1/2}$$

$$\le \sqrt{n-1}\, \|x\|_2$$

On the other hand, we have the following basic inequality: if $x_1, x_2 \neq 0$ then

$$\left\| \frac{x_1}{\|x_1\|} - \frac{x_2}{\|x_2\|} \right\| \le \frac{2\|x_1 - x_2\|}{\max\{\|x_1\|, \|x_2\|\}}$$

$$\left(\quad \frac{x_1}{\|x_1\|} - \frac{x_2}{\|x_2\|} = \frac{x_1}{\|x_1\|} - \frac{x_2}{\|x_1\|} + \frac{x_2}{\|x_1\|} - \frac{x_2}{\|x_2\|} \right.$$

$$\left\| \frac{x_1}{\|x_1\|} - \frac{x_2}{\|x_2\|} \right\| \le \frac{\|x_1 - x_2\|}{\|x_1\|} + \|x_2\| \left| \frac{1}{\|x_1\|} - \frac{1}{\|x_2\|} \right|$$

$$\leq \frac{\|x_1 - x_2\|}{\|x_1\|} + \|x_2\| \left| \frac{\|x_2\| - \|x_1\|}{\|x_1\| \cdot \|x_2\|} \right|$$

Since $\big| \|x_2\| - \|x_1\| \big| \leq \|x_2 - x_1\|$ (by the triangle ineq)

$$\leq \frac{\|x_1 - x_2\|}{\|x_1\|} + \frac{\cancel{\|x_2\|} \, \|x_1 - x_2\|}{\|x_1\| \, \cancel{\|x_2\|}} = \frac{2 \|x_1 - x_2\|}{\|x_1\|})$$

Let's combine,

$$\|A^k x - \lambda_1^k c_1 v_1\| \leq |\lambda_1|^k \left|\frac{\lambda_2}{\lambda_1}\right|^k \sum_{j=1}^{n} |c_j|$$

and

$$\left\| \frac{x_1}{\|x_1\|} - \frac{x_2}{\|x_2\|} \right\| \leq \frac{\|x_1 - x_2\|}{\max\{\|x_1\|, \|x_2\|\}}$$

We get

$$\left\| \frac{A^k x}{\|A^k x\|} - v_1 \right\| \leq \frac{|\lambda_1|^k \left|\frac{\lambda_2}{\lambda_1}\right|^k \sum_{j=1}^{n} |c_j|}{\|\lambda_1^k c_1 v_1\|} \quad = |\lambda_1|^k |c_1|$$

(since v_1 is a unit vector)

Then,

$$\left\| \frac{A^k x}{\|A^k x\|} - v_1 \right\| \leq \frac{\cancel{|\lambda_1|^k} \left|\frac{\lambda_2}{\lambda_1}\right|^k \sum_{j=1}^{n} |c_j|}{\cancel{|\lambda_1|^k} |c_1|}$$

In conclusion,

$$\left\| \frac{A^k x}{\|A^k x\|} - v_1 \right\| \leq \mu^k \left(\frac{1}{|c_1|} \sum_{j=1}^{n} |c_j| \right)$$

This shows how $\frac{A^k x}{\|A^k x\|}$ can be a good approximation to the "top" eigenvector (v_1) of A.

How good an approximation?, choosing x at random we can be confident — if the v_k's are orthonormal — that

$$\frac{1}{|c_1|} \sum_{j=2}^{n} |c_j|$$

won't be too large. If μ is say, ≤ 0.5, then

$$\mu^k \leq 10^{-15} \quad \text{if} \quad k \geq 50.$$

Even if $\mu = 0.9$, the convergence is still pretty fast! ($\mu^{300} \leq 10^{-13}$)

2. QR Iteration (see Solomon for more details)

Observation: If A and B are such that

$$A = QBQ^t \qquad (Q \text{ orthogonal})$$

Then A and B have the same eigenvalues.

Take A, and compute its QR decomposition

$$A = Q_1 R_1$$

Observe

$$Q_1^{-1} A Q_1 = \cancel{Q_1^{-1}} (\cancel{Q_1} R_1) Q_1$$
$$= R_1 Q_1 =: A_2$$

So $A_2 = Q_1^{-1} A Q_1$, and A_1 and A_2 have the same eigenvalues.

Take the QR decomposition of A_2,

$$A_2 = Q_2 R_2$$

3rd $\quad A_3 := R_2 Q_2 \quad (= Q_2^{-1} A Q_2)$

$\vdots$

and so on.

We end up with sequences A_k, Q_k, R_k where

$$A_k = Q_k R_k \quad (QR \text{ decomposit. of } A_k)$$

$$A_{k+1} := R_k Q_k \quad (= Q_k^{-1} A_k Q_k)$$

The matrices A_k all have the same eigenvalues.

Suppose A_k, Q_k, R_k all converge to matrices $A_\infty, Q_\infty, R_\infty$. Then we have

$$A_\infty = Q_\infty R_\infty = R_\infty Q_\infty$$

Observation: In this situation if v is an eigenvector of R_∞, then $Q_\infty v$ is also an eigenvector of R_∞, and with the same eigenvalue! i.e.

$$R_\infty v = \lambda v_\infty \Rightarrow Q_\infty R_\infty v = \lambda Q_\infty v$$

$\Rightarrow$ (since $Q_\infty R_\infty = R_\infty Q_\infty$) $\quad R_\infty(Q_\infty V) = \lambda(Q_\infty V)$

so $Q_\infty V$ is also an eigenvector of R_∞ with the same eigenvalue.

If the eigenvalues of R_∞ are all different, we would conclude that

$$Q_\infty V = \pm V$$

for any eigenvector of R_∞ $\Rightarrow$ the eigenvectors of R_∞ are the same as those of $R_\infty Q_\infty = A_\infty$

Now if $A_\infty V = \lambda V \Rightarrow \lambda V = (R_\infty Q_\infty) V$
$$= \pm R_\infty V$$

This shows that up to signs the eigenvalues of A_∞ (so, of A) are given by the diagonal elements of R_∞.

Krylov Space Methods ...

It has been found that a powerful approach to study a matrix A is to pick a generic vector x_0 and study the matrix:

$$\left(Ax_0 \quad A^2 x_0 \quad \cdots \quad A^n x_0 \right)$$

(see Bau and Trefethen's latter chapters)

The Singular Value Decomposition

Let A be a $m \times n$ matrix. Then there exists a diagonal $m \times n$ matrix Σ (with non-negative entries), an orthogonal matrices U and V such that

$$A = U \Sigma V^t$$

This generalizes the eigenvector decomposition and it holds for all matrices, the price we pay is having two different orthogonal matrices U and V. Their columns are called the left and right singular vectors, and the diagonal elements of Σ are called the singular values.

Observe (preview): If $A = U \Sigma V^t$, then

$$AA^t = U \Sigma V^t V \Sigma^t U^t = U \Sigma \Sigma^t U^t$$

$$A^t A = V \Sigma^t U^t U \Sigma V^t = V \Sigma^t \Sigma V^t$$