Ch 9.3: The power method

Thursday, November 20, 2025 8:09 PM

The power method

Use case is fairly specific/limited but it's a part of more sophisticated algorithms, and it's simple, which is why we teach it.

Simple... but limited. It finds the largest eigenvalue in magnitude (and its associated eigenvector), approximately

It's an iterative algo.

Let's take $A \in \mathbb{C}^{n \times n}$ w/ eigenvalues (sorted)

$$|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n| \geq 0$$

and assume we have a basis of eigenvectors (i.e., diagonalizable)
$\{\vec{v}_1, \ldots, \vec{v}_n\}$

So any $\vec{x} \in \mathbb{C}^n$ can be written $\vec{x} = \sum_{i=1}^{n} \beta_i \vec{v}_i$

thus

$$A \cdot \vec{x} = A \cdot \left(\sum \beta_i \vec{v}_i\right) = \sum_i \beta_i A \vec{v}_i = \sum \beta_i \lambda_i \vec{v}_i$$

and

$$A^k \vec{x} = \ldots = \sum_i \beta_i \lambda_i^k \vec{v}_i$$

i.e. $A^k \vec{x} = \lambda_1^k \cdot \left(\sum_{i=1}^{n} \beta_i \underbrace{\left(\frac{\lambda_i}{\lambda_1}\right)^k}_{} \cdot \vec{v}_i \right)$

For $i=1$, this is 1
For $i \neq 1$, $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$

so $\left|\frac{\lambda_i}{\lambda_1}\right|^k \to 0$

So

$$A^k \vec{x} \approx \underbrace{\lambda_1^k}_{} \cdot \beta_1 \vec{v}_1 \quad \text{for } k \text{ large}$$

probably goes to 0 or ∞ or oscillates, so we normalize to fix that

Simple normalization

Initialize at $\vec{x}_0$ w/ unit norm
For $k = 1, 2, 3, \ldots$

$$\vec{y}_k = A \cdot \vec{x}_{k-1}$$
$$\mu_{k-1} = \vec{x}_{k-1}^T \cdot \vec{y}_k$$
$$\vec{x}_k = \vec{y}_k / \|\vec{y}_k\|_2$$

If $\vec{x}$ is an eigenvector,
$A\vec{x} = \lambda \vec{x}$
so $\vec{x}^T A \vec{x} = \lambda \cdot \vec{x}^T \vec{x}$
or $\lambda = \underbrace{\dfrac{\vec{x}^T A \vec{x}}{\|\vec{x}\|^2}}_{\text{Rayleigh quotient}}$

then μ_k converges to λ_1 at rate $O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$

the vector $\vec{x}_k$ "kind of" converges to the eigenvector $\vec{v}_1$:

really, $\text{dist}(\vec{x}_k, \text{span}(\vec{v}_1)) \to 0$ (in fact, at rate $\left|\frac{\lambda_2}{\lambda_1}\right|^k$)

(why this distinction? First, $\vec{v}_1$ may not be normalized but $\vec{x}_k$ is. Second, $\vec{x}_k$ might flip sign,

ex: $\lambda_1 = -3$, $A = \begin{bmatrix} -3 & 0 \\ 0 & 1 \end{bmatrix}$, $\vec{x}_k = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$\vec{y}_k = A\vec{x} = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$, $\vec{x}_{k+1} = \vec{y}_k / \|\vec{y}_k\| = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

sign alternates. That's ok, both are eigenvalues, but you can't literally say $\vec{x}_k$ converges.

To avoid this flip-flopping, the book introduces a ...

complicated normalization scheme

subroutine signed_ℓ^∞_normalization($\vec{x}$)

Compute $\|\vec{x}\|_\infty = \max_{i \le n} |x_i|$

Let p be the index, $1 \le p \le n$, such that $|x_p| = \|\vec{x}\|_\infty$. If there's more than 1 such index, choose the first index

Return $\vec{x}/x_p$, x_p

So this has $\|\cdot\|_\infty = 1$ but the sign is consistent

power_iter($\vec{x}_0$)

$\vec{x}_0, \mu_0 \leftarrow$ signed_ℓ^∞_norm($\vec{x}_0$)

for $k = 1, 2, 3, \ldots$

$\vec{y}_k = A \cdot \vec{x}_k$

$\vec{y}_k, \mu_k \leftarrow$ signed_ℓ^∞_norm($\vec{y}_k$)

if $\|\vec{y}_k - \vec{x}_k\| <$ tolerance

break

else

$\vec{x}_{k+1} \leftarrow \vec{y}_k$

return $\vec{y}_k, \mu_k$ // estimates of eigenvector, eigenvalue

... but it's still the same basic iteration.

distance($\vec{y}_k$, span($\vec{v}_1$)) converges the same as w/ our other normalization.

The power method is guaranteed to work if

1) A is diagonalizable
2) $|\lambda_1| > |\lambda_2|$ strict separation

} can be slightly relaxed

3) $\beta_1 \neq 0$, ie. $\vec{x}_0$ not chosen unluckily } due to floating pt. numbers, this isn't a big worry

⚠ The book implies the $\|\cdot\|_2$ normalized version, using $\mu_k = \dfrac{\vec{x}_k^T A \vec{x}_k}{\|\vec{x}_k\|^2}$,

is only applicable for symmetric matrices (or Hermitian, if complex)

That's not true.

But it is true that: $A = A^T$ (or $A = A^*$ if complex) means $O(|\frac{\lambda_2}{\lambda_1}|^{2k})$ convergence

$A \neq A^T$ (or $A \neq A^*$ if complex) means $O(|\frac{\lambda_2}{\lambda_1}|^{k})$ convergence

So it works better for symmetric/Hermitian matrices

proof sketches of convergence rates:

① if $A = A^T$ then $\mu_k \to \lambda_1$ at rate $O(|\frac{\lambda_2}{\lambda_1}|^{2k})$

let's simplify and not give a general proof, but still do enough to convey the idea, so assume $n=3$ and $\vec{x} = \vec{v}_1 + \vec{v}_2 + \vec{v}_3$ (ie. $\beta_1 = \beta_2 = \beta_3$ for simplicity)

eigenvectors

So $A \cdot \vec{x} = \lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 + \lambda_3 \vec{v}_3$

Also assume $|\lambda_1| > |\lambda_2| > |\lambda_3| \geq 0$

Since $A = A^T$ we can choose* $\{\vec{v}_i\}$ to be orthonormal.

(SUBTLE)
* it's not automatically true:
1) normalization isn't automatic
2) orthogonality is possible but only automatic if all eigenvalues are distinct

$A\vec{x}_k = \lambda_1^k \vec{v}_1 + \lambda_2^k \vec{v}_2 + \lambda_3^k \vec{v}_3$

then

$$\mu_k = \frac{\vec{x}_k^T A \vec{x}_k}{\vec{x}_k^T \vec{x}_k} = \frac{\lambda_1^{2k+1} + \lambda_2^{2k+1} + \lambda_3^{2k+1}}{\lambda_1^{2k} + \lambda_2^{2k} + \lambda_3^{2k}}$$

using orthogonality

To show $|\mu_k - \lambda_1| = O(|\frac{\lambda_2}{\lambda_1}|^{2k})$, we'll show $\dfrac{|\mu_k - \lambda_1|}{|\frac{\lambda_2}{\lambda_1}|^{2k}} \leq$ bound (independent of k)

$$\frac{|\mu_k - \lambda_1|}{|\frac{\lambda_2}{\lambda_1}|^{2k}} = \left| \frac{\lambda_1^{2k+1} + \lambda_2^{2k+1} + \lambda_3^{2k+1}}{\lambda_1^{2k} + \lambda_2^{2k} + \lambda_3^{2k}} - \lambda_1 \right| \cdot \left| \frac{\lambda_1^{2k}}{\lambda_2^{2k}} \right|$$

$$= \frac{\left|\lambda_1^{2k+1} + \lambda_2^{2k+1} + \lambda_3^{2k+1} - \lambda_1(\lambda_1^{2k} + \lambda_2^{2k} + \lambda_3^{2k})\right|}{|\lambda_2^{2k}|} \cdot \frac{|\lambda_1^{2k}|}{|\lambda_1^{2k} + \lambda_2^{2k} + \lambda_3^{2k}|}$$

made a common denominator, rearranged products

$$\frac{\left|\cancel{\lambda_1^{2k+1}} + \lambda_2^{2k+1} + \lambda_3^{2k+1} - \lambda_1\left(\cancel{\lambda_1^{2k}} + \lambda_2^{2k} + \lambda_3^{2k}\right)\right|}{|\lambda_2^{2k}|} \cdot \frac{|\lambda_1^{2k}|}{|\lambda_1^{2k}+\lambda_2^{2k}+\lambda_3^{2k}|} =$$

$$\frac{\left|(\lambda_2-\lambda_1)\lambda_2^{2k} + (\lambda_3-\lambda_1)\lambda_3^{2k}\right|}{\lambda_2^{2k}} \cdot \frac{\lambda_1^{2k}}{\lambda_1^{2k}+\lambda_2^{2k}+\lambda_3^{2k}}$$

drop $|\cdot|$ since $\lambda \in \mathbb{R}$ so $\lambda^2 > 0$

$\lambda_1^{2k}+\lambda_2^{2k}+\lambda_3^{2k} \ge \lambda_1^{2k}$ so $\frac{1}{\cdots} \le \frac{1}{\lambda_1^{2k}}$

$$\le \left[|\lambda_2-\lambda_1| \cdot \underbrace{\frac{\lambda_2^{2k}}{\lambda_2^{2k}}}_{=1} + |\lambda_3-\lambda_1| \underbrace{\frac{\lambda_3^{2k}}{\lambda_2^{2k}}}_{\le 1}\right] \cdot \left[\frac{\lambda_1^{2k}}{\lambda_1^{2k}}\right] \le |\lambda_2-\lambda_1| + |\lambda_3-\lambda_1|$$

which is finite and doesn't grow with k. □

(2) If $A \ne A^T$ (or $A \ne A^*$ if complex),

$$|\mu_k - \lambda_1| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right) \text{ but } \underbrace{\text{not } O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right)}_{\text{let's show this}}$$

(via a counterexample)

The difference is that eigenvectors need not be orthogonal:

take $\vec{x} = \vec{v}_1 + \vec{v}_2$, say $\|\vec{v}_1\|_2 = \|\vec{v}_2\|_2$ but $\vec{v}_1^T \vec{v}_2 = c \ne 0$

so

$$\vec{x}_k = \lambda_1^k \vec{v}_1 + \lambda_2^k \vec{v}_2, \text{ as before,}$$

but

$$\mu_k = \frac{\vec{x}_k^T A \vec{x}_k}{\|\vec{x}_k\|_2^2} = \frac{(\lambda_1^k \vec{v}_1 + \lambda_2^k \vec{v}_2)^T(\lambda_1^{k+1}\vec{v}_1 + \lambda_2^{k+1}\vec{v}_2)}{(\lambda_1^k \vec{v}_1 + \lambda_2^k \vec{v}_2)^T(\lambda_1^k \vec{v}_1 + \lambda_2^k \vec{v}_2)}$$

$$= \frac{\lambda_1^{2k+1} + \lambda_2^{2k+1} + \boxed{c\lambda_1^k \lambda_2^{k+1} + c\lambda_1^{k+1}\lambda_2^k}}{\lambda_1^{2k} + \lambda_2^{2k} + 2c\lambda_1^k \lambda_2^k}$$

these cross-terms are problematic

I'm ignoring abs. values

cancel

$$\frac{|\mu_k - \lambda_1|}{|\lambda_2/\lambda_1|^{2k}} = \frac{\cancel{\lambda_1^{2k+1}} + \lambda_2^{2k+1} + c\lambda_1^k\lambda_2^{k+1} + c\lambda_1^{k+1}\lambda_2^k - \lambda_1\left(\cancel{\lambda_1^{2k}} + \lambda_2^{2k} + 2c\lambda_1^k\lambda_2^k\right)}{\lambda_2^{2k}} \cdot \frac{\lambda_1^{2k}}{\lambda_1^{2k}+\lambda_2^{2k}+2c\lambda_1^k\lambda_2^k}$$

$$= \boxed{\frac{\lambda_2^{2k+1} + c\left[\lambda_1^k\lambda_2^{k+1} + \lambda_1^{k+1}\lambda_2^k\right]}{\lambda_2^{2k}}} \cdot \underbrace{\frac{\lambda_1^{2k}}{\lambda_1^{2k}+\lambda_2^{2k}+2c\lambda_1^k\lambda_2^k}}_{\text{OK}}$$

Problematic

$$\lim_{k\to\infty} \frac{\lambda_1^{k+1}\lambda_2^k}{\lambda_2^{2k}} = \lim_{k\to\infty} \lambda_1 \cdot \underbrace{\left(\frac{\lambda_1}{\lambda_2}\right)}_{>1}^{k}$$

So raising ^k means it's not bounded.