Reminder: HW #1 due one week from today

OH: today, 11am-noon

Friday, 3pm-4pm

Notes posted on website.

Introduction to Nonlinear Optimization

(Eigenvalues & eigenvectors)

Lecture 02

August 31, 2021

Beck, sections 1.4

Eigenvalues and eigenvectors

$\underline{\underline{A}} \in \mathbb{R}^{n \times n}$

$\lambda \in \mathbb{C}$ is an eigenvalue if $\exists\ \underline{v} \neq 0$ s.t. $\underline{\underline{A}}\,\underline{v} = \lambda \underline{v}$

complex numbers

Any $\underline{v}$ satisfying the above is an eigenvector.

$\lambda = 0$ is OK

$\underline{v} \neq \underline{0}$

Properties of eigenvalues/eigenvectors

- $\underline{\underline{A}}$ <u>always</u> has n (possibly repeated) eigenvalues in $\mathbb{C}$.
- Every <u>distinct</u> eigenvalue has at least 1 eigenvector.

- Eigenvectors corresponding to two distinct eigenvalues are linearly independent.

- Even if $\underline{\underline{A}}$ is real, its eigenvalues and eigenvectors can be complex.

Ex. $\underline{\underline{A}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ $\quad \lambda = \pm\sqrt{-1}$

Ex. Eigenvalues do not measure a type of norm on matrices

$$\underline{\underline{A}} = \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \text{ for any } c \in \mathbb{R}, \quad \lambda(\underline{\underline{A}}) = 1, 1$$

Definition: If an $n \times n$ matrix $\underline{\underline{A}}$ has n linearly independent eigenvectors, then $\underline{\underline{A}}$ is "diagonalizable".

(Because: $\underline{\underline{A}}\,\underline{v}_i = \lambda_i \underline{v}_i$, $i = 1 \ldots n$.

$\Big\downarrow$ concatenate

$$\underline{\underline{A}}\,\underline{\underline{V}} = \underline{\underline{V}}\,\underline{\underline{\Lambda}}, \quad \underline{\underline{V}} = \begin{pmatrix} | & & | \\ \underline{v}_1 & \cdots & \underline{v}_n \\ | & & | \end{pmatrix}, \quad \underline{\underline{\Lambda}} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$\Big\Downarrow$ $\underline{\underline{V}}$ invertible

$$\underbrace{\underline{\underline{V}}^{-1}\,\underline{\underline{A}}\,\underline{\underline{V}}}_{\text{"similarity" transform}} = \underbrace{\underline{\underline{\Lambda}}}_{\text{diagonal}}$$

)

Ex: $\underline{\underline{A}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable.

The spectral theorem

Theorem: Assume $\underline{\underline{A}} \in \mathbb{R}^{n \times n}$ is symmetric $\left((\underline{\underline{A}})_{i,j} = (\underline{\underline{A}})_{j,i} \right)$.
$(\underline{\underline{A}} = \underline{\underline{A}}^T)$

Then:

- All eigenvalues of $\underline{\underline{A}}$ are real-valued.
- There are n linearly independent eigenvectors of $\underline{\underline{A}}$, and they can be chosen to be an orthonormal set.
- $\underline{\underline{A}}$ is orthogonally diagonalizable: $\exists$ orthogonal matrix $\underline{\underline{U}} \in \mathbb{R}^{n \times n}$ s.t. $\underline{\underline{U}}^T \underline{\underline{A}}\, \underline{\underline{U}} = \underline{\underline{\Lambda}}$

(columns of $\underline{\underline{U}}$ are eigenvectors)

(Note if $\underline{\underline{U}}$ is orthogonal $\Rightarrow$ $\underline{\underline{U}}^{-1} = \underline{\underline{U}}^T$)

$\underline{\underline{U}}^T \underline{\underline{U}} = \underline{\underline{I}}$

Consequences of the spectral theorem: (assume $\underline{\underline{A}} \in \mathbb{R}^{n \times n}$ is symmetric):

- $\|\underline{\underline{A}}\|_2 = \max\limits_{\|\underline{x}\|_2 = 1} \|\underline{\underline{A}}\,\underline{x}\|_2$

  $\|A\|_2 = \max\{|\lambda_1|, \dots |\lambda_n|\}$ where $\lambda_1 \dots \lambda_n$ are eigenvalues of $\underline{\underline{A}}$.

- $\underline{\underline{A}}$ has eigenvalues $\lambda_1, \dots, \lambda_n$

  " has eigenvectors $\underline{u}_1, \dots, \underline{u}_n$ that are orthonormal.

  $$\underline{\underline{A}} = \sum_{j=1}^{n} \lambda_j \underbrace{\left(\underline{u}_j\, \underline{u}_j^T\right)}_{n \times n \text{ matrix}}$$

- $\underline{\underline{A}} = \underline{\underline{U}}\, \underline{\underline{\Lambda}}\, \underline{\underline{U}}^T$

  Recall: $\underline{x} \mapsto \underline{\underline{U}}\,\underline{x}$ is a "rotation" of $\mathbb{R}^n$

  $\underline{x} \mapsto \underline{\underline{U}}^T \underline{x}$ is also a rotation.

  Interpretation of $\underline{x} \mapsto \underline{\underline{A}}\,\underline{x}$: 1.) apply a rotation $\underline{\underline{U}}^T$

  2.) Diagonal scaling by $\underline{\underline{\Lambda}}$.

  3.) apply the inverse rotation $\underline{\underline{U}} = (\underline{\underline{U}}^T)^{-1}$.

Rayleigh quotients

Definition: Given $\underline{\underline{A}} \in \mathbb{R}^{n\times n}$, the Rayleigh quotient is a function $R_{\underline{\underline{A}}} : \cancel{\mathbb{R}^n} \to \mathbb{C}$ defined as

$\mathbb{R}^n \setminus \{\underline{0}\}$

$$R_{\underline{\underline{A}}}(\underline{x}) = \frac{\langle \underline{\underline{A}}\underline{x}, \underline{x} \rangle}{\|\underline{x}\|_2^2} = \frac{\underline{x}^T \underline{\underline{A}} \underline{x}}{\underline{x}^T \underline{x}}$$

Goal: Given $\underline{\underline{A}} \in \mathbb{R}^{n\times n}$ that is symmetric, maximize and minimize $R_{\underline{\underline{A}}}$. (Recall that $R_{\underline{\underline{A}}}$ is real-valued here.)

Solve: find $\underline{x}$ s.t. $R_{\underline{\underline{A}}}(\underline{x}) \geq R_{\underline{\underline{A}}}(\underline{y}) \quad \forall\, \underline{y} \in \mathbb{R}^n$

find $\underline{x}$ s.t. $R_{\underline{\underline{A}}}(\underline{x}) \leq R_{\underline{\underline{A}}}(\underline{y}) \quad \forall\, \underline{y} \in \mathbb{R}^n$.

Optimality of Rayleigh quotients

What values (in $\mathbb{R}$) can $R_{\underline{\underline{A}}}$ achieve?

Lemma: Let $\underline{\underline{A}}$ be symmetric, and let $\lambda_{min}(\underline{\underline{A}})$ and $\lambda_{max}(\underline{\underline{A}})$ be the minimum and maximum eigenvalues respectively. Then:

$$\lambda_{min}(\underline{\underline{A}}) \leq R_{\underline{\underline{A}}}(\underline{x}) \leq \lambda_{max}(\underline{\underline{A}}) \quad \forall\, \underline{x} \in \mathbb{R}^n \setminus \{\underline{0}\}$$

Proof: $\underline{\underline{A}} = \underline{\underline{U}}\, \underline{\underline{\Lambda}}\, \underline{\underline{U}}^T$

$$\underline{x}^T \underline{\underline{A}}\, \underline{x} = \underline{x}^T \underline{\underline{U}}\, \underline{\underline{\Lambda}}\, \underline{\underline{U}}^T \underline{x} = (\underline{\underline{U}}^T \underline{x})^T \underline{\underline{\Lambda}}\, \underbrace{(\underline{\underline{U}}^T \underline{x})}_{\underline{y}}$$

$$\underline{x}^T \underline{A} \underline{x} = \underline{y}^T \underline{\Lambda} \underline{y} \quad , \quad \underline{y} = \underline{U}^T \underline{x}.$$

also: $$\underline{x}^T \underline{x} = \underline{x}^T \underbrace{\underline{U}\ \underline{U}^T}_{\underline{I}} \underline{x} = (\underline{U}^T \underline{x})^T (\underline{U}^T \underline{x}) = \underline{y}^T \underline{y}$$

$$R_{\underline{A}}(\underline{x}) = \frac{\underline{x}^T \underline{A} \underline{x}}{\underline{x}^T \underline{x}} = \frac{\underline{y}^T \underline{\Lambda} \underline{y}}{\underline{y}^T \underline{y}} = \frac{\sum_{j=1}^{n} \lambda_j (y_j)^2}{\sum_{j=1}^{n} (y_j)^2}$$

$$\Rightarrow R_{\underline{A}}(\underline{x}) \leq \frac{\sum_{j=1}^{n} \lambda_{max}(\underline{A}) (y_j)^2}{\sum_{j=1}^{n} (y_j)^2} = \lambda_{max}(\underline{A})$$

assuming $\sum_{j=1}^{n} (y_j)^2 \neq 0$,

$\underline{y}^T \underline{y} \neq 0$

$\underline{x}^T \underline{x} \neq 0$

$\underline{x} \neq 0$.

Similarly: $R_{\underline{A}}(\underline{x}) \geq \lambda_{min}(\underline{A})$ since $\lambda_j \geq \lambda_{min}(\underline{A})$. ∎

Note: if $\underline{u}$ is a unit-norm eigenvector of $\underline{A}$ corresponding to $\lambda_{max}(\underline{A})$, then:

$$R_{\underline{A}}(\underline{u}) = \frac{\underline{u}^T \underline{A} \underline{u}}{\underline{u}^T \underline{u}} = \frac{\lambda_{max}(\underline{A}) \cdot \underline{u}^T \underline{u}}{\underline{u}^T \underline{u}} = \lambda_{max}(\underline{A})$$

Theorem: Let $\underline{\underline{A}}$ be symmetric.

Then: **any** eigenvector $\underline{u}$ associated to $\lambda_{max}(\underline{\underline{A}})$ solves

"find $\underline{x}$ s.t. $R_{\underline{\underline{A}}}(\underline{x}) \geq R_{\underline{\underline{A}}}(\underline{y}) \ \forall \underline{y} \in \mathbb{R}^n$".

And: **any** eigenvector $\underline{u}$ associated to $\lambda_{min}(\underline{\underline{A}})$ solves:

"find $\underline{x}$ s.t. $R_{\underline{\underline{A}}}(\underline{x}) \leq R_{\underline{\underline{A}}}(\underline{y}) \ \forall \underline{y} \in \mathbb{R}^n$".

Note: All this leverages symmetry of $\underline{\underline{A}}$.
There is a solution to the optimization problems, but the solution is (terribly) non-unique.