SYMMETRIC MATRICES

Dr. Matthieu R Bloch

Monday, November 1, 2021
Grades

- Midterm 1 was long… future exams will be better calibrated
- I will curve to get GPA similar to past semesters

My office hours on Tuesdays

- 8am-9am on BlueJeans (https://bluejeans.com/205357142)
- Tomorrow (Tuesday November 02, 2021) will focus on Midterm 1 solution
- I’ll try to record the session

Midterm 2:

- Moved to Monday November 8, 2021 (gives you weekend to prepare)
- Coverage: everything since Midterm 1 (don’t forget the fundamentals though), emphasis on regression
Last time:
- Symmetric matrices: more linear algebra
- **Objective:** further understand least-square problems

**Reading:** lecture notes 12
Least square problems involved the normal equations $X^T X \theta = X^T y$

This is a system of symmetric equations $Ax = y$ with $A^T = A$
Least square problems involved the normal equations $X^\top X\theta = X^\top y$

This is a system of symmetric equations $Ax = y$ with $A^\top = A$

- Ultimately we will talk about the non-symmetric/non square case

**Definition.**

A real-valued matrix $A$ is symmetric if $A^\top = A$ \((a_{ij} = a_{ji} \text{ for } A = [a_{ij}])\)

A complex-valued matrix $A$ is Hermitian if $A^\dagger = A$ (also written $A^H = A$) \(A^\dagger = (A^\top)^* = (A^*)^\top\)
Least square problems involved the normal equations \( \mathbf{X}^\top \mathbf{X} \mathbf{\theta} = \mathbf{X}^\top \mathbf{y} \)

This is a system of symmetric equations \( \mathbf{A} \mathbf{x} = \mathbf{y} \) with \( \mathbf{A}^\top = \mathbf{A} \)

- Ultimately we will talk about the non-symmetric/non square case

**Definition.**

A real-valued matrix \( \mathbf{A} \) is symmetric if \( \mathbf{A}^\top = \mathbf{A} \)

A complex-valued matrix \( \mathbf{A} \) is Hermitian if \( \mathbf{A}^\dagger = \mathbf{A} \) (also written \( \mathbf{A}^H = \mathbf{A} \))

**Definition.**

Given a matrix \( \mathbf{A} \in \mathbb{C}^{n \times n} \), if a vector \( \mathbf{v} \in \mathbb{C}^n \) satisfies \( \mathbf{A} \mathbf{v} = \lambda \mathbf{v} \) for some \( \lambda \in \mathbb{C} \), then \( \lambda \) is an eigenvalue associated to the eigenvector \( \mathbf{v} \).

If \( \lambda \) is an eigenvalue, there are infinitely many eigenvectors associated to it

\[
\text{If } \mathbf{v} = \lambda \mathbf{v} \text{ then } \forall \alpha \in \mathbb{C} \quad \mathbf{A}(\alpha \mathbf{v}) = \alpha \mathbf{A} \mathbf{v} = \alpha (\lambda \mathbf{v}) = \lambda (\alpha \mathbf{v}) \quad \alpha \mathbf{v} \text{ is another eigenvector.}
\]
Least square problems involved the normal equations $X^T X \theta = X^T y$

This is a system of symmetric equations $Ax = y$ with $A^T = A$

- Ultimately we will talk about the non-symmetric/non square case

**Definition.**

A real-valued matrix $A$ is symmetric if $A^T = A$.

A complex-valued matrix $A$ is Hermitian if $A^\dagger = A$ (also written $A^H = A$).

**Definition.**

Given a matrix $A \in \mathbb{C}^{n \times n}$, if a vector $v \in \mathbb{C}^n$ satisfies $Av = \lambda v$ for some $\lambda \in \mathbb{C}$, then $\lambda$ is an *eigenvalue* associated to the *eigenvector* $v$.

If $\lambda$ is an eigenvalue, there are infinitely many eigenvectors associated to it.

**Definition.**

Given a matrix $A \in \mathbb{C}^{n \times n}$, if a vector $v \in \mathbb{C}^n$ satisfies $Av = \lambda v$ for some $\lambda \in \mathbb{C}$, then $\lambda$ is an *eigenvalue* associated to the *eigenvector* $v$. 
Consider the canonical basis \( \{ e_i \}_{i=1}^{n} \) for \( \mathbb{R}^n \); every vector can be viewed as a vector of coefficients \( \{ \alpha_i \}_{i=1}^{n} \),

\[
\mathbf{x} = \sum_{i=1}^{n} \alpha_i e_i = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]^\top
\]
Consider the canonical basis \( \{e_i\}_{i=1}^n \) for \( \mathbb{R}^n \); every vector can be viewed as a vector of coefficients \( \{\alpha_i\}_{i=1}^n \),

\[
\mathbf{x} = \sum_{i=1}^{n} \alpha_i e_i = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]^T
\]

How do we find the representation of \( \mathbf{x} \) in another basis \( \{v_i\}_{i=1}^n \)? Write \( e_i = \sum_{j=1}^{n} \beta_{ij} v_j \).
Consider the canonical basis \( \{e_i\}_{i=1}^n \) for \( \mathbb{R}^n \); every vector can be viewed as a vector of coefficients \( \{\alpha_i\}_{i=1}^n \),

\[
x = \sum_{i=1}^{n} \alpha_i e_i = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]^\top \quad (\star)
\]

How do we find the representation of \( x \) in another basis \( \{v_i\}_{i=1}^n \)? Write \( e_i = \sum_{j=1}^{n} \beta_{ij} v_j \)

Regroup the coefficients

\[
x = \cdots \left( \sum_{i=1}^{n} \beta_{ij} \alpha_i \right) v_j + \cdots
\]

\[
x = \sum_{i=1}^{n} \alpha_i e_i = \sum_{i=1}^{n} \alpha_i \left( \sum_{j=1}^{n} \beta_{ij} v_j \right) = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} \beta_{ij} \alpha_i \right) v_j
\]

**Indep of \( i \)**
Consider the canonical basis \( \{e_i\}_{i=1}^n \) for \( \mathbb{R}^n \); every vector can be viewed as a vector of coefficients \( \{\alpha_i\}_{i=1}^n \),

\[
\mathbf{x} = \sum_{i=1}^{n} \alpha_i e_i = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]^\top
\]

How do we find the representation of \( \mathbf{x} \) in another basis \( \{v_i\}_{i=1}^n \)? Write \( e_i = \sum_{j=1}^{n} \beta_{ij} v_j \)

Regroup the coefficients

\[
\mathbf{x} = \cdots + \left( \sum_{i=1}^{n} \beta_{ij} \alpha_i \right) v_j + \cdots
\]

In matrix form

\[
\mathbf{x}_{\text{new}} = \begin{bmatrix}
\beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\
\beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{1n} & \beta_{2n} & \cdots & \beta_{nn}
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_n
\end{bmatrix}
\]
SIMILARITY

A change of basis matrix $P$ is full rank (basis vectors are linearly independent)

Any full rank matrix $P$ can be viewed as a change of basis

$P^{-1}$ takes you back to the original basis

Warning: the columns of $P$ describe the old coordinates as a function of the new ones

**Definition.**

If $A, B \in \mathbb{R}^{n \times n}$ then $B$ is similar to $A$ if there exists an invertible matrix $P \in \mathbb{R}^{n \times n}$ such that $B = P^{-1}AP$

Note: $\forall x \in \mathbb{R}^n \quad Bx = P^{-1}APx$

- Back to original basis
- Changing basis of $x$
- Apply transform

$\begin{align*}
\end{align*}$
A change of basis matrix $\mathbf{P}$ is full rank (basis vectors are linearly independent)

Any full rank matrix $\mathbf{P}$ can be viewed as a change of basis

$\mathbf{P}^{-1}$ takes you back to the original basis

**Warning:** the columns of $\mathbf{P}$ describe the *old* coordinates as a function of the *new* ones

**Definition.**

If $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ then $\mathbf{B}$ is similar to $\mathbf{A}$ if there exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $\mathbf{B} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$

**Intuition:** similar matrices are the same up to a change of basis

**Definition.**

$\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonalizable if it is similar to a diagonal matrix, i.e., there exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $\mathbf{D} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ with $\mathbf{D}$ diagonal

\[
\mathbf{D} = \begin{pmatrix}
d_1 & 0 & \cdots & 0 \\
0 & d_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & d_n
\end{pmatrix} \quad \mathbf{A}_{\mathcal{L}} \in \mathcal{L}^n \\
\mathbf{D} \mathbf{x} = \begin{pmatrix}
d_1 \mathbf{x}_1 \\
0 \mathbf{x}_2 \\
\vdots \\
0 \mathbf{x}_n
\end{pmatrix} = \begin{pmatrix}
\mathbf{x}_1 \\
\mathbf{x}_2 \\
\vdots \\
\mathbf{x}_n
\end{pmatrix}
\]
A change of basis matrix $\mathbf{P}$ is full rank (basis vectors are linearly independent)

Any full rank matrix $\mathbf{P}$ can be viewed as a change of basis

$\mathbf{P}^{-1}$ takes you back to the original basis

**Warning:** the columns of $\mathbf{P}$ describe the *old* coordinates as a function of the *new* ones

**Definition.**

If $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ then $\mathbf{B}$ is similar to $\mathbf{A}$ if there exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that

$$\mathbf{B} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$$

**Intuition:** similar matrices are the same up to a change of basis

**Definition.**

$\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonalizable if it is similar to a diagonal matrix, i.e., there exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $\mathbf{D} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ with $\mathbf{D}$ diagonal

Not all matrices are diagonalizable!

**Example**

$$\mathbf{R} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$
Lemma (Existence of eigenvector) Every complex matrix $A$ has at least one complex eigenvector and every real symmetric matrix has real eigenvalues and at least one real eigenvector.
Proof: Let $A \in \mathbb{C}^{n \times n}$

1. Fact: $\lambda \in \mathbb{C}$ is an eigenvalue of $A$ corresponding to $x \in \mathbb{C}^n$ iff $\lambda$ is a root of $p(\lambda) = \det(A - \lambda I)$

   Proof: $Ax = \lambda x$ iff $(A - \lambda I)x = 0$ iff $A - \lambda I$ has non-zero kernel iff $\det(A - \lambda I) = 0$ iff $\lambda$ is a root of $p(\lambda)$

2. Fact: every complex matrix $A$ has at least one eigenvalue

   Proof: By the fundamental theorem of algebra (for complex polynomials) applied to $p(\lambda) = \det(A - \lambda I)$, there exists (at least) one complex root $\lambda \in \mathbb{C}$

   By Fact 1, there exists an associated eigenvector $x \in \mathbb{C}^n$

3. Let $A \in \mathbb{R}^{n \times n}$ and symmetric; let $x$ be an eigenvector with eigenvalue $\lambda \in \mathbb{R}$

   Then $x^T A x = x^T \lambda x = \lambda \|x\|_2^2$

   $x^T A^T x = (Ax)^T x = (\lambda x)^T x = \lambda^T \|x\|_2^2$

   and $\lambda = \lambda^T$ so that $\lambda \in \mathbb{R}$

   By backtracking our Fact 1, we can find a real-valued eigenvector
Lemma (Existence of eigenvector) Every complex matrix $A$ has at least one complex eigenvector and every real symmetric matrix has real eigenvalues and at least one real eigenvector.

Lemma (Schur triangularization lemma) Every matrix $A \in \mathbb{C}^{n \times n}$ is unitarily similar to an upper triangular matrix, i.e.,

$$A = V \Delta V^\dagger = V \begin{pmatrix} \lambda_1 & * \\ 0 & \lambda_2 \end{pmatrix} V^\dagger$$

with $\Delta$ upper triangular and $V^\dagger = V^{-1}$. 
Proof: Let \( A \in \mathbb{C}^{m \times m} \); we know \( \exists \) eigenvalue \( \lambda_i \) \( \forall \) eigenvector \( \mathbf{v}_i \) (non-zero)

\[
\mathbf{V}_i = \begin{bmatrix} \mathbf{v}_i & \mathbf{u}_i \end{bmatrix} \mathbf{U} \in \mathbb{C}^{m \times m} \text{ so that } U_i^T U_i = I \text{ and } U_i^T v_i = 0 \quad \text{; assume wlog that } \| \mathbf{v}_i \| = 1
\]

Hence \( A \mathbf{V}_i = A \begin{bmatrix} \mathbf{v}_i & \mathbf{u}_i \end{bmatrix} = \begin{bmatrix} \lambda_i \mathbf{v}_i & \mathbf{A} \mathbf{v}_i \end{bmatrix} \) and \( \mathbf{V}_i^T A \mathbf{V}_i = \begin{bmatrix} -\mathbf{v}_i^T \mathbf{u}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \lambda_i \mathbf{v}_i & \mathbf{A} \mathbf{v}_i \end{bmatrix} = \begin{bmatrix} \lambda_i & 0 \\ 0 & \mathbf{V}_i^T \mathbf{A} \mathbf{v}_i \end{bmatrix}

Assume we can write \( A \in \mathbb{C}^{m \times m} \) as \( A_p = \begin{pmatrix} \Delta p & \mathbf{W}_p \\ \mathbf{0} & M_p \end{pmatrix} \) \( \forall p \in \mathbb{C}^{1 \times m-1} \) \( \mathbf{W}_p, M_p \) arbitrary

Let \( \mathbf{V}_p \) be an eigenvector (non-zero) of \( M_p \) \( \forall \) eigenvalue \( \lambda_{p+1} \) (assume wlog that \( \| \mathbf{V}_p \| = 1 \))

Construct \( \mathbf{Z}_{p+1} = \begin{bmatrix} \mathbf{v}_p^T & \mathbf{U}_{p+1} \end{bmatrix} \) diagonalizable \( \mathbf{Z}_{p+1}^T = \mathbf{Z}_{p+1}^{-1} \); Set \( \mathbf{V}_{p+1} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{p+1}^{-1} \end{bmatrix} \)

Then \( A_p \mathbf{V}_{p+1} = \begin{bmatrix} \Delta p & \mathbf{W}_p \mathbf{Z}_{p+1} \\ \mathbf{0} & \lambda_{p+1} \mathbf{v}_p^T M_p \mathbf{U}_{p+1} \end{bmatrix} \) and \( \mathbf{V}_{p+1}^T A_p \mathbf{V}_{p+1} = \begin{bmatrix} \Delta p & \mathbf{W}_p \mathbf{Z}_{p+1} \\ \mathbf{0} & \lambda_{p+1} \mathbf{v}_p^T M_p \mathbf{U}_{p+1} \end{bmatrix} \)
Finally

\[ A_1 = V_1^+ A V_1 \]
\[ A_2 = V_2^+ A_1 V_2 = V_2^+ V_1^+ A V_1 V_2 \]
\[ \vdots \]
\[ A_n = V_n^+ \cdots V_1^+ A V_1 \cdots V_n = \Delta \]

So that \[ A = V \Delta V^+ \]
**Lemma (Existence of eigenvector)** Every complex matrix $A$ has at least one complex eigenvector and every real symmetric matrix has real eigenvalues and at least one real eigenvector.

**Lemma (Schur triangularization lemma)** Every matrix $A \in \mathbb{C}^{n \times n}$ is unitarily similar to an upper triangular matrix, i.e.,

$$A = V\Delta V^\dagger$$

with $\Delta$ upper triangular and $V^\dagger = V^{-1}$.

**Theorem (Spectral theorem)** Every hermitian matrix is unitarily similar to a real-valued diagonal matrix.

$$A^\dagger = A$$

$$A = V\Lambda V^\dagger$$

$\Lambda$ diagonal and $V^\dagger V = I$
Proof: By the Schur triangularization lemma

\[ A = V \Delta V^+ = A^+ = (V \Delta V^+)^+ = V \Delta^+ V^+ \]

Hence \( \Delta = \Delta^+ \) (\( V^+AV = \Delta = \Delta^+ \))

\[ \Delta = \begin{pmatrix} \text{diag}(\mathbb{R}) & 0 \\ 0 & 0 \end{pmatrix} = \Delta^+ \begin{pmatrix} \text{diag}(\mathbb{R}) & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \text{diag}(\mathbb{R}) & 0 \\ 0 & 0 \end{pmatrix} \]

\( \lambda_i \in \mathbb{R} \)  \( \square \)
**Lemma (Existence of eigenvector)** Every complex matrix $A$ has at least one complex eigenvector and every real symmetric matrix has real eigenvalues and at least one real eigenvector.

**Lemma (Schur triangularization lemma)** Every matrix $A \in \mathbb{C}^{n \times n}$ is unitarily similar to an upper triangular matrix, i.e.,

$$A = V \Delta V^\dagger$$

with $\Delta$ upper triangular and $V^\dagger = V^{-1}$.

**Theorem (Spectral theorem)** Every hermitian matrix is unitarily similar to a real-valued diagonal matrix. Note that if $A = V \Delta V^\dagger$ then

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^\dagger$$

How about real-valued matrices $A \in \mathbb{R}^{n \times n}$
Definition.

A symmetric matrix $A$ is positive definite if it has positive eigenvalues, i.e., $\forall i \in \{1, \cdots, n\}$, $\lambda_i > 0$.

A symmetric matrix $A$ is positive semidefinite if it has nonnegative eigenvalues, i.e., $\forall i \in \{1, \cdots, n\}$, $\lambda_i \geq 0$.

Convention: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

Variational form of extreme eigenvalues for symmetric positive definite matrices $A$:

\[
\lambda_1 = \max_{\mathbf{x} \in \mathbb{R}^n: \|\mathbf{x}\|_2 = 1} \mathbf{x}^\top A \mathbf{x} = \max_{\mathbf{x} \in \mathbb{R}^n} \frac{\mathbf{x}^\top A \mathbf{x}}{\|\mathbf{x}\|_2^2}
\]

\[
\lambda_n = \min_{\mathbf{x} \in \mathbb{R}^n: \|\mathbf{x}\|_2 = 1} \mathbf{x}^\top A \mathbf{x} = \min_{\mathbf{x} \in \mathbb{R}^n} \frac{\mathbf{x}^\top A \mathbf{x}}{\|\mathbf{x}\|_2^2}
\]

Theorem (Sylvester theorem)

For any analytic function $f$, we have

\[
f(A) = \sum_{i=1}^{n} f(\lambda_i) \mathbf{v}_i \mathbf{v}_i^\top
\]
Consider the system $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\mathbf{A}$ symmetric positive definite

**Proposition.**

Let $\{\mathbf{v}_i\}$ be the eigenvectors of $\mathbf{A}$.

$$\mathbf{x} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \langle \mathbf{y}, \mathbf{v}_i \rangle \mathbf{v}_i$$

Assume that there exists some observation error $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$

- $\mathbf{e}$ is unknown
- we try to reconstruct $\mathbf{x}$ as $\tilde{\mathbf{x}}$ by applying $\mathbf{A}^{-1}$

**Proposition.**

$$\frac{1}{\lambda_1} \| \mathbf{e} \|_2^2 \leq \| \mathbf{x} - \tilde{\mathbf{x}} \|_2 \leq \frac{1}{\lambda_n} \| \mathbf{e} \|_2^2.$$