SYMMETRIC MATRICES

DR. MATTHIEU R BLOCH

Monday, November 1, 2021

LOGISTICS

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:

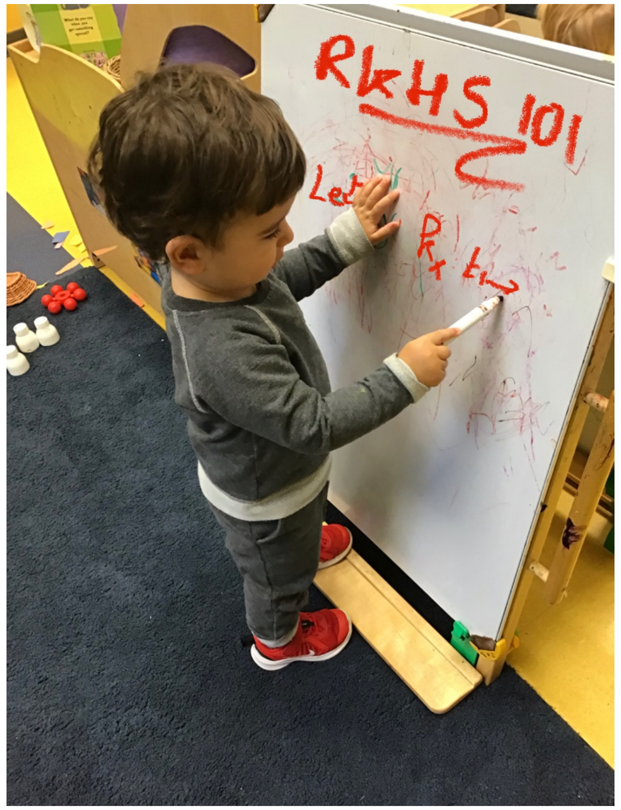
- Moved to **Monday November 8, 2021** (gives you weekend to prepare)
- Coverage: everything since Midterm 1 (dont’ forget the fundamentals though), emphasis on **regression**

WHAT'S ON THE AGENDA FOR TODAY?

Last time:

- Symmetric matrices: more linear algebra
- *Objective:* further understand least-square problems

Reading: lecture notes 12



Toddlers can do it!

SYSTEMS OF SYMMETRIC EQUATIONS

Least square problems involved the normal equations $\mathbf{X}^\top\mathbf{X}\boldsymbol{\theta} = \mathbf{X}^\top\mathbf{y}$

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

SYSTEMS OF SYMMETRIC EQUATIONS

Least square problems involved the normal equations $\mathbf{X}^\intercal\mathbf{X}\boldsymbol{\theta} = \mathbf{X}^\intercal\mathbf{y}$

This is a system of symmetric equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ with $\mathbf{A}^\intercal = \mathbf{A}$

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

Definition.

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

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

SYSTEMS OF SYMMETRIC EQUATIONS

Least square problems involved the normal equations $\mathbf{X}^\intercal\mathbf{X}\boldsymbol{\theta} = \mathbf{X}^\intercal\mathbf{y}$

This is a system of symmetric equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ with $\mathbf{A}^\intercal = \mathbf{A}$

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

> **Definition.**
>
> A real-valued matrix $\mathbf{A}$ is symmetric if $\mathbf{A}^\intercal = \mathbf{A}$
>
> 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 λ is an *eigenvalue* associated to the *eigenvector* $\mathbf{v}$.

If λ is an eigenvalue, there are infinitely many eigenvectors associated to it

if $Av = \lambda v$ then $\forall \alpha \in \mathbb{C}$ $A(\alpha v) = \alpha A v = \alpha \lambda v = \lambda(\alpha v)$ αv is another eigenvector

SYSTEMS OF SYMMETRIC EQUATIONS

Least square problems involved the normal equations $\mathbf{X}^\intercal \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^\intercal \mathbf{y}$

This is a system of symmetric equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ with $\mathbf{A}^\intercal = \mathbf{A}$

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

Definition.

A real-valued matrix $\mathbf{A}$ is symmetric if $\mathbf{A}^\intercal = \mathbf{A}$

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 λ is an *eigenvalue* associated to the *eigenvector* $\mathbf{v}$.

If λ is an eigenvalue, there are infinitely many eigenvectors associated to it

~~**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 λ is an *eigenvalue* associated to the *eigenvector* $\mathbf{v}$.~~

CHANGE OF BASIS

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

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i e_i = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{bmatrix}^\mathsf{T}$$

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{th position}$$

CHANGE OF BASIS

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

$$\mathbf{x} = \sum_{i=1}^n \alpha_i e_i = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{bmatrix}^\mathsf{T}$$

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

CHANGE OF BASIS

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

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i e_i = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]^\mathsf{T} \qquad (*)$$

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

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

$$x = \sum_{i=1}^{n} \alpha_i e_i = \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} \beta_{ij} v_j \right) = \sum_{j=1}^{n} \underbrace{\left(\sum_{i=1}^{n} \beta_{ij} \alpha_i \right)}_{\text{indep of } i} v_j$$

CHANGE OF BASIS

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

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i e_i = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{bmatrix}^\mathsf{T}$$

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

Regroup the coefficients

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

In matrix form

$$\mathbf{x}_{\text{new}} = \underbrace{\begin{array}{cccc} e_1 & e_2 & \cdots & e_n \\ \downarrow & \downarrow & & \downarrow \end{array} \begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{n1} \\ \beta_{12} & \beta_{22} & \cdots & \beta_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{1n} & \beta_{2n} & \cdots & \beta_{nn} \end{bmatrix}}_{P} \mathbf{x} \qquad \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

SIMILARITY

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

Note: $\forall x \in \mathbb{R}^n \quad Bx = P^{-1} A \underbrace{Px}_{\text{Changing basis of } x}$

Back to original basis

Apply transform

SIMILARITY

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

$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$ $\{d_i\}_{i=1}^n \in \mathbb{C}^n$ $\qquad Dx = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1x_1 \\ \vdots \\ d_nx_n \end{pmatrix}$

SIMILARITY

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 $R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

SPECTRAL THEOREM

> **Lemma (Existence of eigenvector)** Every complex matrix $\mathbf{A}$ has at least one complex eigenvector and every real symmetrix matrix has real eigenvalues and at least one real eigenvector.

Proof: Let $A \in \mathbb{C}^{n\times n}$

$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$

① Fact: $\lambda \in \mathbb{C}$ is an eigenvalue of A corresponding to $\underline{x} \in \mathbb{C}^n_*$ iff λ is a root of $p(t) \triangleq \det(A - tI)$

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 λ is a root of p

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

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

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

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

$$\text{Then} \quad x^+ A x = x^+ \lambda x = \lambda \|x\|_2^2 \qquad \text{and} \quad \lambda = \lambda^+ \quad \text{so that } \lambda \in \mathbb{R}$$

$$\parallel$$

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

By backtracing our Fact ①, we can find a real-valued eigenvector

SPECTRAL THEOREM

Lemma (Existence of eigenvector) Every complex matrix $\mathbf{A}$ has at least one complex eigenvector and every real symmetrix matrix has real eigenvalues and at least one real eigenvector.

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

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^\dagger = V \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} V^\dagger$$

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

Proof: Let $A \in \mathbb{C}^{n\times n}$; we know $\exists$ eigenvalue λ_1 w/ eigenvector v_1 (non-zero)

$$V_1 = \begin{bmatrix} v_1 & U_1 \end{bmatrix} \quad U_1 \in \mathbb{C}^{n\times n-1} \text{ so that } U_1^+ U_1 = I \text{ and } U_1^+ v_1 = 0 \quad ; \text{ assume wlog that } \|v_1\| = 1$$

$$\text{Hence } AV_1 = A\begin{bmatrix} v_1 & U_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & AU_1 \end{bmatrix} \text{ and } V_1^+ A V_1 = \begin{bmatrix} -v_1^+- \\ U_1^+ \end{bmatrix} \begin{bmatrix} \lambda_1 v_1 & AU_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \\ 0 & V_1^+ A U_1 \\ | & \\ 0 & \end{bmatrix}$$

Assume we can write $A \in \mathbb{C}^{n\times n}$ as $A_p = \begin{pmatrix} \Delta_p & W_p \\ 0 & M_p \end{pmatrix}$ w/ $p \in [\![1, n-1]\!]$ $\underbrace{W_p}_{p\times n-p}$, $\underbrace{M_p}_{n-p\times n-p}$ arbitrary

Let v_{p+1} be an eigenvector (non-zero) of M_p w/ eigenvalue λ_{p+1} (assume wlog that $\|v_{p+1}\| = 1$)

$$\text{Construct } Z_{p+1} = \begin{bmatrix} v_{p+1} & U_{p+1} \end{bmatrix} \text{ orthonormal } (Z_{p+1}^+ = Z_{p+1}^{-1}) \; ; \text{ Set } V_{p+1} \stackrel{\Delta}{=} \begin{bmatrix} \underbrace{I_p}_{p\times p} & 0 \\ 0 & \underbrace{Z_{p+1}}_{n-p\times n-p} \end{bmatrix}$$

$$\text{Then } A_p V_{p+1} = \begin{bmatrix} \Delta_p & W_p Z_{p+1} \\ 0 & \begin{bmatrix} \lambda_{p+1} v_{p+1} & M_p U_{p+1} \end{bmatrix} \end{bmatrix} \text{ and } V_{p+1}^+ A_p V_{p+1} = \begin{bmatrix} \Delta_p & W_p Z_{p+1} \\ 0 & \begin{bmatrix} \lambda_{p+1} & \\ 0 & Z_{p+1}^+ M_p U_{p+1} \\ | & \\ 0 & \end{bmatrix} \end{bmatrix}$$

Finally $A_1 = V_1^\dagger A V_1$

$$A_2 = V_2^\dagger A_1 V_2 = V_2^\dagger V_1^\dagger A V_1 V_2$$

$$\vdots$$

$$A_n = \underbrace{V_n^\dagger \cdots V_1^\dagger}_{\hat{=} V^\dagger} A \underbrace{V_1 \cdots V_n}_{\hat{=} V} = \Delta$$

So that $A = V \Delta V^\dagger$ $\blacksquare$

SPECTRAL THEOREM

Lemma (Existence of eigenvector) Every complex matrix $\mathbf{A}$ has at least one complex eigenvector and every real symmetrix matrix has real eigenvalues and at least one real eigenvector.

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

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^\dagger$$

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

Theorem (Spectral theorem) Every $\underbrace{\text{hermitian matrix}}_{A^\dagger = A}$ is $\underbrace{\text{unitarily similar to a real-valued diagonal}}_{A = V\Lambda V^\dagger \text{ w/ } \Lambda \text{ diagonal and } V^\dagger V = I}$ matrix.

Proof: By the Shur triangularization lemma

$$A = V \Delta V^{+} = A^{+} = (V\Delta V^{+})^{+} = V \Delta^{+} V^{+}$$

Hence $\Delta = \Delta^{+}$ $\left(V^{+} A V = \Delta = \Delta^{+} \right)$

$$\Delta = \begin{pmatrix} \ddots & * \\ 0 & \ddots \end{pmatrix} = \Delta^{+} \begin{pmatrix} \ddots & 0 \\ * & \ddots \end{pmatrix} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \quad \lambda_i \in \mathbb{R}$$

0

$\in \mathbb{R}$

$\blacksquare$

SPECTRAL THEOREM

> **Lemma (Existence of eigenvector)** Every complex matrix $\mathbf{A}$ has at least one complex eigenvector and every real symmetrix matrix has real eigenvalues and at least one real eigenvector.

> **Lemma (Schur triangularization lemma)** Every matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is unitarily similar to an upper triangular matrix, i.e.,
>
> $$\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^{\dagger}$$
>
> with $\boldsymbol{\Delta}$ upper triangular and $\mathbf{V}^{\dagger} = \mathbf{V}^{-1}$.

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

Note that if $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{\dagger}$ then

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^{\dagger}$$

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

SYMMETRIC POSITIVE DEFINITE MATRICES

Definition.

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

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

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

Variational form of extreme eigenvalues for symmetric positive definite matrices $\mathbf{A}$

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

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

Theorem (Sylvester theorem)

For any analytic function f, we have

$$f(\mathbf{A}) = \sum_{i=1}^{n} f(\lambda_i) \mathbf{v}_i \mathbf{v}_i^\top$$

SYSTEM OF SYMMETRIC DEFINITE EQUATIONS

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.

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