Metodi Numerici dell'Informatica

Spectral decomposition

Emanuele Rodolà
rodola@di.uniroma1.it

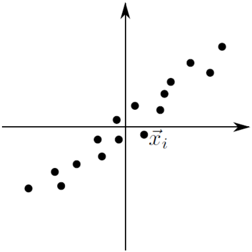
SAPIENZA
UNIVERSITÀ DI ROMA

2nd semester a.y. 2022/2023 · April 4, 2023

Motivation

Principal axis

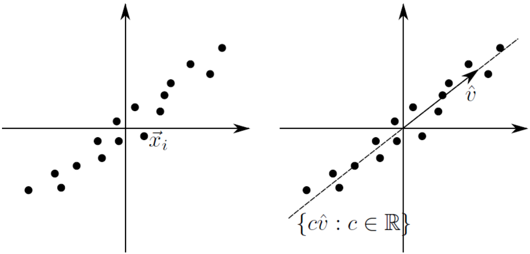
Consider the two-dimensional data in this plot:



(a) Input data

Principal axis

Consider the two-dimensional data in this plot:



(a) Input data (b) Principal axis

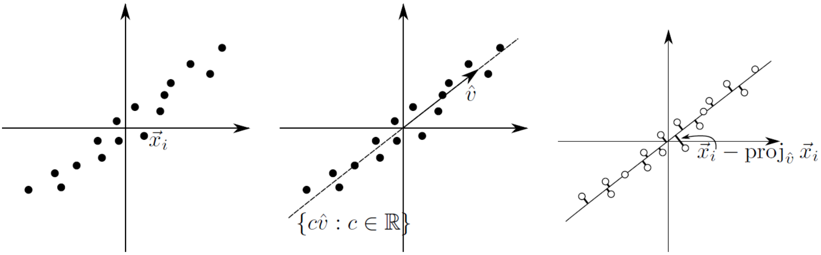
Q: Find the vector $\mathbf{v}$ such that each data point $\mathbf{x}_i$ can be written as

$$\mathbf{x}_i = c_i \mathbf{v}$$

where each $\mathbf{x}_i$ has its own c_i

Principal axis

Consider the two-dimensional data in this plot:



(a) Input data (b) Principal axis (c) Projection error

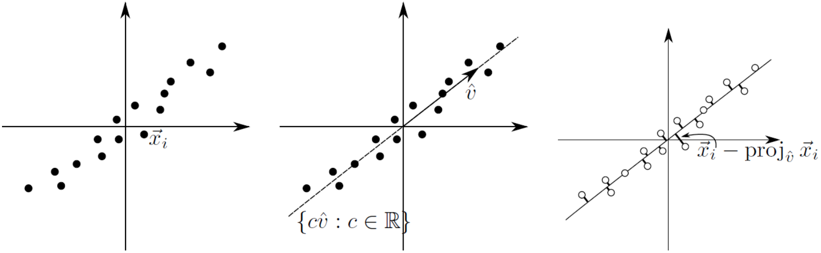
Q: Find the vector $\mathbf{v}$ such that each data point $\mathbf{x}_i$ can be written as

$$\mathbf{x}_i = c_i \mathbf{v}$$

where each $\mathbf{x}_i$ has its own c_i

Principal axis

Consider the two-dimensional data in this plot:



(a) Input data (b) Principal axis (c) Projection error

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \mathrm{proj}_{\mathbf{v}} \mathbf{x}_i\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \mathrm{proj}_{\mathbf{v}}\mathbf{x}_i\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - \frac{\mathbf{x}_i^\top \mathbf{v}}{\mathbf{v}^\top \mathbf{v}} \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i \|\mathbf{x}_i - (\mathbf{x}_i^\top \mathbf{v})\mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \; \sum_i (\mathbf{x}_i - (\mathbf{x}_i^\top \mathbf{v})\mathbf{v})^\top (\mathbf{x}_i - (\mathbf{x}_i^\top \mathbf{v})\mathbf{v})$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - 2(\mathbf{x}_i^\top \mathbf{v})(\mathbf{x}_i^\top \mathbf{v}) + (\mathbf{x}_i^\top \mathbf{v})^2 \|\mathbf{v}\|_2^2)$$

$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - 2(\mathbf{x}_i^\top \mathbf{v})^2 + (\mathbf{x}_i^\top \mathbf{v})^2 \|\mathbf{v}\|_2^2)$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \sum_i (\|\mathbf{x}_i\|_2^2 - (\mathbf{x}_i^\top \mathbf{v})^2)$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \; -\sum_i (\mathbf{x}_i^\top \mathbf{v})^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\min_{\mathbf{v}} \; -\|\mathbf{X}^\top \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where matrix $\mathbf{X}$ contains the vectors $\mathbf{x}_i$ as its columns.

Principal axis

$$\max_{\mathbf{v}} \|\mathbf{X}^\top \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where matrix $\mathbf{X}$ contains the vectors $\mathbf{x}_i$ as its columns.

This can also be written as

$$\max_{\mathbf{v}} \mathbf{v}^\top \underbrace{\mathbf{X}\mathbf{X}^\top}_{\text{symmetric}} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

Principal axis

$$\max_{\mathbf{v}} \|\mathbf{X}^\top \mathbf{v}\|_2^2$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where matrix $\mathbf{X}$ contains the vectors $\mathbf{x}_i$ as its columns.

This can also be written as

$$\max_{\mathbf{v}} \mathbf{v}^\top \underbrace{\mathbf{X}\mathbf{X}^\top}_{\text{symmetric}} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

The global maximizer $\mathbf{v}^*$ of this problem is the principal component of the data contained in the matrix $\mathbf{X}$.

Eigenvectors and eigenvalues

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A} c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda c\mathbf{x}$$

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda c\mathbf{x}$$

For this reason, we can restrict our search to eigenvectors with $\|\mathbf{x}\|_2 = 1$.

Eigenvalue equation

An eigenvector $\mathbf{x}$ of a square matrix $\mathbf{A}$ is any vector satisfying

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for some (possibly complex) number λ that we call eigenvalue.

A few basic facts:

The scale of an eigenvector is not important. In particular:

$$\mathbf{A}c\mathbf{x} = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda c\mathbf{x}$$

For this reason, we can restrict our search to eigenvectors with $\|\mathbf{x}\|_2 = 1$.

Clearly, $\mathbf{x}$ and $-\mathbf{x}$ are both eigenvectors with the same eigenvalue.

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{Tx}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda\mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda \mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda \mathbf{x}$$

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda \mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda \mathbf{x}$$

In other words, $\mathbf{x}$ is an eigenvector of $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ with eigenvalue λ.

Similarity

Eigenvectors can be scaled without changing their eigenvalues:

$$\mathbf{A}c\mathbf{x} = \lambda c\mathbf{x}$$

Instead of just scaling, consider an invertible linear transformation:

$$\mathbf{A}\mathbf{T}\mathbf{x} = \lambda \mathbf{T}\mathbf{x}$$

which is saying that $\mathbf{T}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Is it true that also $\mathbf{x}$ is an eigenvector of $\mathbf{A}$? Let's check:

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{x} = \lambda \mathbf{x}$$

In other words, $\mathbf{x}$ is an eigenvector of $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ with eigenvalue λ.

We say that $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ is a similarity transformation. $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues.

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$\mathbf{Q}\mathbf{x} = \lambda\mathbf{x}$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$\|\mathbf{Q}\mathbf{x}\|_2^2 = \|\lambda\mathbf{x}\|_2^2$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$(\mathbf{Qx})^\top \mathbf{Qx} = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$\mathbf{x}^\top \mathbf{Q}^\top \mathbf{Q}\mathbf{x} = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$\mathbf{x}^\top \mathbf{x} = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$\|\mathbf{x}\|_2^2 = |\lambda|^2 \|\mathbf{x}\|_2^2$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$1 = |\lambda|^2$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  Observe:

$$1 = |\lambda|$$

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  We obtained:

$$\lambda = \pm 1$$

- Diagonal and upper-triangular matrices

  The eigenvalues are the entries along the main diagonal.

More basic facts

A few more basic facts that will become useful later:

- Orthogonal matrices

  We obtained:

$$\lambda = \pm 1$$

- Diagonal and upper-triangular matrices

  The eigenvalues are the entries along the main diagonal.

- Commuting matrices

  Consider two matrices $\mathbf{A}$ and $\mathbf{B}$. One can prove:

$$\mathbf{AB} = \mathbf{BA} \quad \Leftrightarrow \quad \mathbf{A} \text{ and } \mathbf{B} \text{ have the same eigenvectors}$$

Big questions

- How to compute eigenvalues and eigenvectors?

Big questions

- How to compute eigenvalues and eigenvectors?
- What to do with them?

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ equals simply scaling $\mathbf{x}$.

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ equals simply scaling $\mathbf{x}$.

But what if $\mathbf{y}$ is some arbitrary vector?

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ equals simply scaling $\mathbf{x}$.

But what if $\mathbf{y}$ is some arbitrary vector?

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ equals simply scaling $\mathbf{x}$.

But what if $\mathbf{y}$ is some arbitrary vector?

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

- If $\mathbf{y}$ is an arbitrary vector, multiplying by $\mathbf{A}$ equals

$$\mathbf{A}\mathbf{y} = \mathbf{A} \sum_i \alpha_i \mathbf{x}_i$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ equals simply scaling $\mathbf{x}$.

But what if $\mathbf{y}$ is some arbitrary vector?

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

- If $\mathbf{y}$ is an arbitrary vector, multiplying by $\mathbf{A}$ equals

$$\mathbf{A}\mathbf{y} = \sum_i \alpha_i \mathbf{A}\mathbf{x}_i$$

A tentative answer

Why do we care about eigenvectors and eigenvalues?

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

A first trivial (but quite valid!) observation:

- If $\mathbf{x}$ is an eigenvector, multiplying by $\mathbf{A}$ equals simply scaling $\mathbf{x}$.

But what if $\mathbf{y}$ is some arbitrary vector?

Suppose the eigenvectors $\{\mathbf{x}_i\}$ form a basis. Then:

$$\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$$

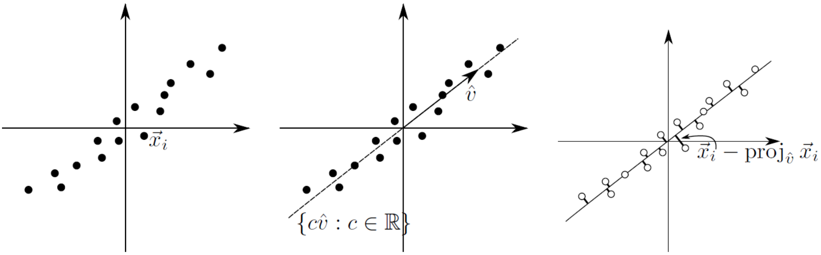
- If $\mathbf{y}$ is an arbitrary vector, multiplying by $\mathbf{A}$ equals

$$\mathbf{A}\mathbf{y} = \sum_i \alpha_i \lambda_i \mathbf{x}_i$$

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \underbrace{\mathbf{X}\mathbf{X}^\top}_{\text{symmetric}} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$



(a) Input data (b) Principal axis (c) Projection error

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A}\mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Theorem If $\mathbf{A}$ is symmetric, then its maximum eigenvalue is given by $\max_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2}$, and $\mathbf{v}$ is the corresponding eigenvector.

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A}\mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Theorem If $\mathbf{A}$ is symmetric, then its maximum eigenvalue is given by $\max_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2}$, and $\mathbf{v}$ is the corresponding eigenvector. More in general:

$$\lambda_{\min} \leq \frac{\mathbf{v}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|_2^2} \leq \lambda_{\max}$$

Min-max theorem

Back to our motivation:

$$\max_{\mathbf{v}} \mathbf{v}^\top \mathbf{A} \mathbf{v}$$
$$\text{s.t. } \|\mathbf{v}\|_2 = 1$$

where $\mathbf{A}$ is symmetric.

Theorem If $\mathbf{A}$ is symmetric, then its maximum eigenvalue is given by $\max_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_2^2}$, and $\mathbf{v}$ is the corresponding eigenvector. More in general:

$$\lambda_{\min} \leq \frac{\mathbf{v}^\top \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_2^2} \leq \lambda_{\max}$$

The ratio $\frac{\mathbf{v}^\top \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_2^2}$ is called Rayleigh quotient.

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\lambda \mathbf{x}^\top \mathbf{x} = (\lambda \mathbf{x})^\top \mathbf{x}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned} \lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\ &= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x}
\end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x})
\end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}}
\end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}}
\end{aligned}$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$
\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}
$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \bar{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

$$(\mathbf{A}\mathbf{x}_i)^\top \mathbf{x}_j = \mathbf{x}_i^\top \mathbf{A} \mathbf{x}_j$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

$$(\lambda_i \mathbf{x}_i)^\top \mathbf{x}_j = \mathbf{x}_i^\top \lambda_j \mathbf{x}_j$$

Symmetric matrices

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$.

Consider an eigenvalue-eigenvector pair $(\lambda, \mathbf{x})$, where $\|\mathbf{x}\|_2^2 = 1$.

All eigenvalues of symmetric matrices are real:

$$\begin{aligned}
\lambda \mathbf{x}^\top \mathbf{x} &= (\lambda \mathbf{x})^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \mathbf{x} \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} \\
&= \mathbf{x}^\top (\mathbf{A}\mathbf{x}) \\
&= \overline{(\mathbf{A}\mathbf{x})^\top \mathbf{x}} \\
&= \overline{(\lambda \mathbf{x})^\top \mathbf{x}} \\
&= \overline{\lambda} \mathbf{x}^\top \mathbf{x}
\end{aligned}$$

Consider distinct eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ with $\lambda_i \neq \lambda_j$:

$$\lambda_i \mathbf{x}_i^\top \mathbf{x}_j = \lambda_j \mathbf{x}_i^\top \mathbf{x}_j$$

For the equality to hold, it must be $\mathbf{x}_i^\top \mathbf{x}_j = 0$, i.e. $\mathbf{x}_i$ and $\mathbf{x}_j$ are orthogonal.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{AX} = \mathbf{X\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

We call it the spectral decomposition of $\mathbf{A}$.

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

We call it the spectral decomposition of $\mathbf{A}$.

Observe the similarity with our motivational problem.

$$\begin{aligned} &\max_{\mathbf{x}} \ \mathbf{x}^\top \mathbf{A} \mathbf{x} \\ &\ \text{s.t.} \ \|\mathbf{x}\|_2 = 1 \end{aligned}$$

Spectral theorem

The set of eigenvalues $\{\lambda_i\}$ of a matrix $\mathbf{A}$ is called the spectrum.

We can write the eigenvalue equation as:

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$$

If $\mathbf{A}$ is symmetric, then $\mathbf{X}$ is an orthogonal matrix of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues.

We call it the spectral decomposition of $\mathbf{A}$.

Observe the similarity with our motivational problem. We can modify it to solve for all eigenvectors and eigenvalues:

$$\min_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X})$$
$$\text{s.t. } \mathbf{X}^\top \mathbf{X} = \mathbf{I}$$

Finding eigenvalues

Power iteration

Very simple algorithm to find the largest eigenvalue/eigenvector:

function NORMALIZED-ITERATION(A)
  $\vec{v} \leftarrow$ ARBITRARY(n)
  for $k \leftarrow 1, 2, 3, \ldots$
    $\vec{w} \leftarrow A\vec{v}$
    $\vec{v} \leftarrow \vec{w}/\|\vec{w}\|$
  return $\vec{v}$

Power iteration

Very simple algorithm to find the largest eigenvalue/eigenvector:

```
function NORMALIZED-ITERATION(A)
    v⃗ ← ARBITRARY(n)
    for k ← 1, 2, 3, ...
        w⃗ ← A v⃗
        v⃗ ← w⃗/‖w⃗‖
    return v⃗
```

The normalization is needed to reduce the numerical error.

Without normalization, it will still converge to the principal eigenvector (but with a very large scale).

Inverse iteration

To find the smallest eigenvalue/eigenvector, we first observe that:

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x} \quad \Longrightarrow \quad \mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Inverse iteration

To find the smallest eigenvalue/eigenvector, we first observe that:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad \mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Then, we can just apply the power method to $\mathbf{A}^{-1}$:

$$\boxed{\begin{array}{l}
\textbf{function } \textsc{Inverse-Iteration}(A) \\
\quad \vec{v} \leftarrow \textsc{Arbitrary}(n) \\
\quad \textbf{for } k \leftarrow 1, 2, 3, \ldots \\
\qquad \vec{w} \leftarrow A^{-1}\vec{v} \\
\qquad \vec{v} \leftarrow \vec{w}/\|\vec{w}\| \\
\quad \textbf{return } \vec{v}
\end{array}}$$

Inverse iteration

To find the smallest eigenvalue/eigenvector, we first observe that:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad \mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Then, we can just apply the power method to $\mathbf{A}^{-1}$:

function $\textsc{Inverse-Iteration}(A)$
  $\vec{v} \leftarrow \textsc{Arbitrary}(n)$
  for $k \leftarrow 1, 2, 3, \ldots$
    $\vec{w} \leftarrow A^{-1}\vec{v}$
    $\vec{v} \leftarrow \vec{w}/\|\vec{w}\|$
  return $\vec{v}$

In practice, you don't invert $\mathbf{A}$ but apply LU decomposition.

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma \mathbf{I})\mathbf{x}_i =$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i =$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i =$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

We can use this fact to estimate portions of the spectrum:

- Provide a guess σ for an eigenvalue

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

We can use this fact to estimate portions of the spectrum:

- Provide a guess σ for an eigenvalue
- Compute the shifted matrix $\mathbf{B} = \mathbf{A} - \sigma\mathbf{I}$

Shifting

For a matrix $\mathbf{A}$, we have eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\mathbf{x}_i\}$.

Then:

$$(\mathbf{A} - \sigma\mathbf{I})\mathbf{x}_i = \mathbf{A}\mathbf{x}_i - \sigma\mathbf{x}_i = \lambda_i\mathbf{x}_i - \sigma\mathbf{x}_i = (\lambda_i - \sigma)\mathbf{x}_i$$

which means that the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$, i.e., by shifting the matrix, we get a corresponding shift in the eigenvalues.

If we think that σ is near an eigenvalue of $\mathbf{A}$, then $\mathbf{A} - \sigma\mathbf{I}$ has an eigenvalue close to 0.

We can use this fact to estimate portions of the spectrum:

- Provide a guess σ for an eigenvalue
- Compute the shifted matrix $\mathbf{B} = \mathbf{A} - \sigma\mathbf{I}$
- Apply the inverse iteration on $\mathbf{B}$

Finding multiple eigenvalues at once

Recall that if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ then $\mathbf{A}$ and $\mathbf{B}$ have the same spectra.

Finding multiple eigenvalues at once

Recall that if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ then $\mathbf{A}$ and $\mathbf{B}$ have the same spectra.

For example, we could first compute

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

and then overwrite:

$$\mathbf{A} \leftarrow \mathbf{Q}^\top \mathbf{A}\mathbf{Q}$$

and still have the guarantee that the eigenvalues stay the same.

Finding multiple eigenvalues at once

Recall that if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ then $\mathbf{A}$ and $\mathbf{B}$ have the same spectra.

For example, we could first compute

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

and then overwrite:

$$\mathbf{A} \leftarrow \mathbf{Q}^\top \mathbf{A}\mathbf{Q}$$

and still have the guarantee that the eigenvalues stay the same.

Now observe:

$$\mathbf{Q}^\top \mathbf{A}\mathbf{Q} =$$

Finding multiple eigenvalues at once

Recall that if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ then $\mathbf{A}$ and $\mathbf{B}$ have the same spectra.

For example, we could first compute

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

and then overwrite:

$$\mathbf{A} \leftarrow \mathbf{Q}^\top\mathbf{A}\mathbf{Q}$$

and still have the guarantee that the eigenvalues stay the same.

Now observe:

$$\mathbf{Q}^\top\mathbf{A}\mathbf{Q} = \mathbf{Q}^\top(\mathbf{Q}\mathbf{R})\mathbf{Q} =$$

Finding multiple eigenvalues at once

Recall that if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ then $\mathbf{A}$ and $\mathbf{B}$ have the same spectra.

For example, we could first compute

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

and then overwrite:

$$\mathbf{A} \leftarrow \mathbf{Q}^\top\mathbf{A}\mathbf{Q}$$

and still have the guarantee that the eigenvalues stay the same.

Now observe:

$$\mathbf{Q}^\top\mathbf{A}\mathbf{Q} = \mathbf{Q}^\top(\mathbf{Q}\mathbf{R})\mathbf{Q} = (\mathbf{Q}^\top\mathbf{Q})\mathbf{R}\mathbf{Q} =$$

Finding multiple eigenvalues at once

Recall that if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ then $\mathbf{A}$ and $\mathbf{B}$ have the same spectra.

For example, we could first compute

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

and then overwrite:

$$\mathbf{A} \leftarrow \mathbf{Q}^\top\mathbf{A}\mathbf{Q}$$

and still have the guarantee that the eigenvalues stay the same.

Now observe:

$$\mathbf{Q}^\top\mathbf{A}\mathbf{Q} = \mathbf{Q}^\top(\mathbf{Q}\mathbf{R})\mathbf{Q} = (\mathbf{Q}^\top\mathbf{Q})\mathbf{R}\mathbf{Q} = \mathbf{R}\mathbf{Q}$$

Finding multiple eigenvalues at once

Recall that if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ then $\mathbf{A}$ and $\mathbf{B}$ have the same spectra.

For example, we could first compute

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

and then overwrite:

$$\mathbf{A} \leftarrow \mathbf{Q}^\top \mathbf{A}\mathbf{Q}$$

and still have the guarantee that the eigenvalues stay the same.

Now observe:

$$\mathbf{Q}^\top \mathbf{A}\mathbf{Q} = \mathbf{Q}^\top(\mathbf{Q}\mathbf{R})\mathbf{Q} = (\mathbf{Q}^\top\mathbf{Q})\mathbf{R}\mathbf{Q} = \mathbf{R}\mathbf{Q}$$

In other words, this operation preserves the eigenvalues:

$$\mathbf{A} \leftarrow \mathbf{R}\mathbf{Q}$$

QR iteration

If we keep doing this iteratively, we get the algorithm:

function QR-ITERATION($A \in \mathbb{R}^{n \times n}$)
  for $k \leftarrow 1, 2, 3, \ldots$
    $Q, R \leftarrow$ QR-FACTORIZE(A)
    $A \leftarrow RQ$
  return $\mathrm{diag}(R)$

QR iteration

If we keep doing this iteratively, we get the algorithm:

$$
\begin{array}{l}
\textbf{function } \text{QR-}\textsc{Iteration}(A \in \mathbb{R}^{n \times n}) \\
\quad \textbf{for } k \leftarrow 1, 2, 3, \ldots \\
\quad\quad Q, R \leftarrow \text{QR-}\textsc{Factorize}(A) \\
\quad\quad A \leftarrow RQ \\
\quad \textbf{return } \operatorname{diag}(R)
\end{array}
$$

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ all have the same eigenvalues

QR iteration

If we keep doing this iteratively, we get the algorithm:

$$
\begin{array}{l}
\textbf{function } \text{QR-}\textsc{Iteration}(A \in \mathbb{R}^{n \times n}) \\
\quad \textbf{for } k \leftarrow 1, 2, 3, \dots \\
\quad\quad Q, R \leftarrow \text{QR-}\textsc{Factorize}(A) \\
\quad\quad A \leftarrow RQ \\
\quad \textbf{return } \operatorname{diag}(R)
\end{array}
$$

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \dots$ all have the same eigenvalues
- The iterations converge, i.e. at some point $\mathbf{A}_k \approx \mathbf{A}_{k-1}$

QR iteration

If we keep doing this iteratively, we get the algorithm:

```
function QR-ITERATION(A ∈ ℝ^{n×n})
    for k ← 1, 2, 3, ...
        Q, R ← QR-FACTORIZE(A)
        A ← RQ
    return diag(R)
```

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ all have the same eigenvalues
- The iterations converge, i.e. at some point $\mathbf{A}_k \approx \mathbf{A}_{k-1}$
- At convergence, we get $\mathbf{Q}_\infty^\top \mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{A}_\infty$, which implies $\mathbf{Q}_\infty \mathbf{R}_\infty = \mathbf{R}_\infty \mathbf{Q}_\infty$ and $\mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{Q}_\infty \mathbf{A}_\infty$

QR iteration

If we keep doing this iteratively, we get the algorithm:

```
function QR-ITERATION(A ∈ ℝ^{n×n})
    for k ← 1, 2, 3, . . .
        Q, R ← QR-FACTORIZE(A)
        A ← RQ
    return diag(R)
```

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ all have the same eigenvalues
- The iterations converge, i.e. at some point $\mathbf{A}_k \approx \mathbf{A}_{k-1}$
- At convergence, we get $\mathbf{Q}_\infty^\top \mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{A}_\infty$, which implies $\mathbf{Q}_\infty \mathbf{R}_\infty = \mathbf{R}_\infty \mathbf{Q}_\infty$ and $\mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{Q}_\infty \mathbf{A}_\infty$
- If $\mathbf{A}_\infty \mathbf{x} = \lambda \mathbf{x}$, then we can write

$$\lambda \mathbf{x} = \mathbf{A}_\infty \mathbf{x} =$$

QR iteration

If we keep doing this iteratively, we get the algorithm:

```
function QR-ITERATION(A ∈ ℝ^{n×n})
    for k ← 1, 2, 3, ...
        Q, R ← QR-FACTORIZE(A)
        A ← RQ
    return diag(R)
```

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ all have the same eigenvalues
- The iterations converge, i.e. at some point $\mathbf{A}_k \approx \mathbf{A}_{k-1}$
- At convergence, we get $\mathbf{Q}_\infty^\top \mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{A}_\infty$, which implies $\mathbf{Q}_\infty \mathbf{R}_\infty = \mathbf{R}_\infty \mathbf{Q}_\infty$ and $\mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{Q}_\infty \mathbf{A}_\infty$
- If $\mathbf{A}_\infty \mathbf{x} = \lambda \mathbf{x}$, then we can write

$$\lambda \mathbf{x} = \mathbf{A}_\infty \mathbf{x} = \mathbf{Q}_\infty \mathbf{R}_\infty \mathbf{x} =$$

QR iteration

If we keep doing this iteratively, we get the algorithm:

```
function QR-ITERATION(A ∈ ℝ^{n×n})
    for k ← 1, 2, 3, ...
        Q, R ← QR-FACTORIZE(A)
        A ← RQ
    return diag(R)
```

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ all have the same eigenvalues
- The iterations converge, i.e. at some point $\mathbf{A}_k \approx \mathbf{A}_{k-1}$
- At convergence, we get $\mathbf{Q}_\infty^\top \mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{A}_\infty$, which implies $\mathbf{Q}_\infty \mathbf{R}_\infty = \mathbf{R}_\infty \mathbf{Q}_\infty$ and $\mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{Q}_\infty \mathbf{A}_\infty$
- If $\mathbf{A}_\infty \mathbf{x} = \lambda \mathbf{x}$, then we can write

$$\lambda \mathbf{x} = \mathbf{A}_\infty \mathbf{x} = \mathbf{Q}_\infty \mathbf{R}_\infty \mathbf{x} = \mathbf{R}_\infty \mathbf{Q}_\infty \mathbf{x} =$$

QR iteration

If we keep doing this iteratively, we get the algorithm:

function QR-ITERATION($A \in \mathbb{R}^{n \times n}$)
  for $k \leftarrow 1, 2, 3, \ldots$
    $Q, R \leftarrow$ QR-FACTORIZE(A)
    $A \leftarrow RQ$
  return diag(R)

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ all have the same eigenvalues
- The iterations converge, i.e. at some point $\mathbf{A}_k \approx \mathbf{A}_{k-1}$
- At convergence, we get $\mathbf{Q}_\infty^\top \mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{A}_\infty$, which implies $\mathbf{Q}_\infty \mathbf{R}_\infty = \mathbf{R}_\infty \mathbf{Q}_\infty$ and $\mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{Q}_\infty \mathbf{A}_\infty$
- If $\mathbf{A}_\infty \mathbf{x} = \lambda \mathbf{x}$, then we can write

$$\lambda \mathbf{x} = \mathbf{A}_\infty \mathbf{x} = \mathbf{Q}_\infty \mathbf{R}_\infty \mathbf{x} = \mathbf{R}_\infty \mathbf{Q}_\infty \mathbf{x} = \pm \mathbf{R}_\infty \mathbf{x}$$

QR iteration

If we keep doing this iteratively, we get the algorithm:

$$
\begin{array}{l}
\textbf{function } \text{QR-ITERATION}(A \in \mathbb{R}^{n\times n}) \\
\quad \textbf{for } k \leftarrow 1, 2, 3, \ldots \\
\quad\quad Q, R \leftarrow \text{QR-FACTORIZE}(A) \\
\quad\quad A \leftarrow RQ \\
\quad \textbf{return } \operatorname{diag}(R)
\end{array}
$$

- The generated matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ all have the same eigenvalues
- The iterations converge, i.e. at some point $\mathbf{A}_k \approx \mathbf{A}_{k-1}$
- At convergence, we get $\mathbf{Q}_\infty^\top \mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{A}_\infty$, which implies $\mathbf{Q}_\infty \mathbf{R}_\infty = \mathbf{R}_\infty \mathbf{Q}_\infty$ and $\mathbf{A}_\infty \mathbf{Q}_\infty = \mathbf{Q}_\infty \mathbf{A}_\infty$
- If $\mathbf{A}_\infty \mathbf{x} = \lambda \mathbf{x}$, then we can write

$$\lambda \mathbf{x} = \mathbf{A}_\infty \mathbf{x} = \mathbf{Q}_\infty \mathbf{R}_\infty \mathbf{x} = \mathbf{R}_\infty \mathbf{Q}_\infty \mathbf{x} = \pm \mathbf{R}_\infty \mathbf{x}$$

- In other words, the eigenvalues of $\mathbf{A}_\infty$ (and hence of $\mathbf{A}$) equal the diagonal elements of $\mathbf{R}_\infty$ up to sign.

Suggested reading

Read Sections 6.1.1, 6.2, 6.2.1, 6.3.1, 6.3.2, 6.3.3, 6.4.2 of the book:

J. Solomon, “Numerical Algorithms”