Mathematics for Informatics

Eigenvalues and eigenvectors
(lecture 10 of 12)

Francesco Dolce

francesco.dolce@fjfi.cvut.cz

Czech Technical University in Prague

Fall 2020/2021

created: November 29, 2020, 11:39

Outline

- Eigenvalues and eigenvectors
- Power methods
- QR algoritmus

Eigenvalues and eigenvectors

A complex number λ is called an **eigenvalue** of the matric $M \in \mathbb{C}^{n,n}$, whenever there exists a non-zero vector $u \in \mathbb{C}^n$ such that

$$Mu = \lambda u.$$

The vector u is called an **eigenvector** of the matrix M relative to the eigenvalue λ.

Eigenvalues and eigenvectors

A complex number λ is called an **eigenvalue** of the matric $M \in \mathbb{C}^{n,n}$, whenever there exists a non-zero vector $u \in \mathbb{C}^n$ such that

$$Mu = \lambda u.$$

The vector u is called an **eigenvector** of the matrix M relative to the eigenvalue λ.

The set of eigenvectors of M (relative to the eigenvalues λ and to the zero vector) form a base of the subspace $\ker(M - \lambda E)$.

Eigenvalues and eigenvectors

A complex number λ is called an **eigenvalue** of the matric $M \in \mathbb{C}^{n,n}$, whenever there exists a non-zero vector $u \in \mathbb{C}^n$ such that

$$Mu = \lambda u.$$

The vector u is called an **eigenvector** of the matrix M relative to the eigenvalue λ.

The set of eigenvectors of M (relative to the eigenvalues λ and to the zero vector) form a base of the subspace $\ker(M - \lambda E)$.

The eigenvalues of the matrix M are the roots of the **characteristic polynomial** of the M, that is the polynomial

$$p_M(\lambda) := \det(M - \lambda E).$$

Therefore, each matrix $M \in \mathbb{C}^{n,n}$ has at most n different complex eigenvalues.

Diagonalizability of a matrix

A matrix $M \in \mathbb{C}^{n,n}$ is **diagonalizable** when there exist a diagonal matrix $D \in \mathbb{C}^{n,n}$ and a regular matrix $P \in \mathbb{C}^{n,n}$ such that

$$M = PDP^{-1}.$$

where $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.

Diagonalizability of a matrix

A matrix $M \in \mathbb{C}^{n,n}$ is **diagonalizable** when there exist a diagonal matrix $D \in \mathbb{C}^{n,n}$ and a regular matrix $P \in \mathbb{C}^{n,n}$ such that

$$M = PDP^{-1}.$$

where $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.

Remind: In the previous lecture we saw that $M^k = PD^kP^{-1}$.

Remark: The columns of the matrix P are the eigenvectors of M. These eigenvectors form a basis of $\mathbb{C}^n$. The elements of the diagonal matrix D are the eigenvalues of M (with their multiplicity).

Looking for an eigenvector

Let $M \in \mathbb{C}^{n,n}$. Suppose it is diagonalizable and we can order its eigenvalues as follows

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

Looking for an eigenvector

Let $M \in \mathbb{C}^{n,n}$. Suppose it is diagonalizable and we can order its eigenvalues as follows

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

We are looking for the eigenvector of the eigenvalue λ_1, the so-called **dominant eigenvalue**. It is a vector u_1 such that

$$Mu_1 = \lambda_1 u_1.$$

Looking for an eigenvector

Let $M \in \mathbb{C}^{n,n}$. Suppose it is diagonalizable and we can order its eigenvalues as follows

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

We are looking for the eigenvector of the eigenvalue λ_1, the so-called **dominant eigenvalue**. It is a vector u_1 such that

$$Mu_1 = \lambda_1 u_1.$$

In general, the matrix need not be diagonalizable, but the ideas would be more complicated (actually, we only require to have one eigenvalue which is the greatest in absolute value).

Applications

Eigenvalues play an importan role in several applications:

- Classification of conics and quadratic forms (geometry).
- Quantum computation, quantum mechanics, asymptotic behaviour of dynamical systems (physics).
- PCA, or *Principal Component Analysis* (big data).
- Recognition of 2D and 3D objects using spectral methods (AI).

Applications

Eigenvalues play an importan role in several applications:

- Classification of conics and quadratic forms (geometry).
- Quantum computation, quantum mechanics, asymptotic behaviour of dynamical systems (physics).
- PCA, or *Principal Component Analysis* (big data).
- Recognition of 2D and 3D objects using spectral methods (AI).
- More practical example: **PageRank** mesures a relative importance of WWW documents by ispecting links between them.
  - Its values is in fact an eigenvector of the dominant eigenvalues of a modified adjacency matrix of these links. This matrix satisfies requirement of our problem.
  - **PageRank** is calculated using **power methods**.

Introduction and assumptions (1/2)

In its basic variant, the power method is used to find the dominant eigenvalue of a matrix,

Introduction and assumptions (1/2)

In its basic variant, the power method is used to find the dominant eigenvalue of a matrix,

Given a matrix $M \in \mathbb{C}^{n,n}$ let us consider a regular matrix $P \in \mathbb{C}^{n,n}$ such that

$$M = PDP^{-1}$$

where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Let also suppose that the values are ordered:

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Introduction and assumptions (1/2)

In its basic variant, the power method is used to find the dominant eigenvalue of a matrix,

Given a matrix $M \in \mathbb{C}^{n,n}$ let us consider a regular matrix $P \in \mathbb{C}^{n,n}$ such that

$$M = PDP^{-1}$$

where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Let also suppose that the values are ordered:

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Note: We suppose that the dominant eigenvalue λ_1 is not degerate (i.e., that the correspoinding eigenspace has dimension 1).

Introduction and assumptions (2/2)

We are looking for an eigenvector associated to the eigenvalue λ_1, that is a non-zero vector u_1 such that

$$Mu_1 = \lambda_1 u_1.$$

Introduction and assumptions (2/2)

We are looking for an eigenvector associated to the eigenvalue λ_1, that is a non-zero vector u_1 such that

$$Mu_1 = \lambda_1 u_1.$$

The **power method** is an **iterative method**. We will construct a sequence $(x_k)_k$ as follows: x_0 is chosen randomly and the next terms are determined by

$$x_k = Mx_{k-1} \quad \text{for } k > 0.$$

Equivalently, we have

$$x_k = M^k x_0 \quad k \in \mathbb{N}_0.$$

Power method principle (1/4)

If M is regular, thus diagonalizable, there exist eigenvectors $\{u_1, u_2, \ldots, u_n\}$, which form a basis of $\mathbb{C}^{n,1}$.

If M is not regular, then we need to complete the set of eigenvectors by a basis of the kernel of M.

Power method principle (1/4)

If M is regular, thus diagonalizable, there exist eigenvectors $\{u_1, u_2, \ldots, u_n\}$, which form a basis of $\mathbb{C}^{n,1}$.

If M is not regular, then we need to complete the set of eigenvectors by a basis of the kernel of M.

The vector x_0 can be written as $x_0 = \alpha_1 u_1 + \cdots + \alpha_n u_n$.
Suppose that $\alpha_1 \neq 0$.

Power method principle (1/4)

If M is regular, thus diagonalizable, there exist eigenvectors $\{u_1, u_2, \ldots, u_n\}$, which form a basis of $\mathbb{C}^{n,1}$.

If M is not regular, then we need to complete the set of eigenvectors by a basis of the kernel of M.

The vector x_0 can be written as $x_0 = \alpha_1 u_1 + \cdots + \alpha_n u_n$.
Suppose that $\alpha_1 \neq 0$.

Coefficients α_i can be absorbed by the eigenvectors $(u'_i = \alpha_i u_i)$ and we have

$$x_0 = u'_1 + \cdots + u'_n.$$

Power method principle (2/4)

The recurrent definition of x_k implies

$$\begin{aligned} x_k &= M^k x_0 \\ &= M^k u_1 + \cdots + M^k u_n \\ &= \lambda_1^k u_1 + \cdots + \lambda_n^k u_n. \end{aligned}$$

Power method principle (2/4)

The recurrent definition of x_k implies

$$\begin{aligned} x_k &= M^k x_0 \\ &= M^k u_1 + \cdots + M^k u_n \\ &= \lambda_1^k u_1 + \cdots + \lambda_n^k u_n. \end{aligned}$$

The last equality gives

$$x_k = \lambda_1^k \left(u_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k u_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k u_n \right).$$

Power method principle (2/4)

The recurrent definition of x_k implies

$$\begin{aligned} x_k &= M^k x_0 \\ &= M^k u_1 + \cdots + M^k u_n \\ &= \lambda_1^k u_1 + \cdots + \lambda_n^k u_n. \end{aligned}$$

The last equality gives

$$x_k = \lambda_1^k \left(u_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k u_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k u_n \right).$$

We rewrite it as

$$x_k = \lambda_1^k \left(u_1 + \varepsilon_k \right).$$

Since for all $j > 1$ we have $\left|\frac{\lambda_j}{\lambda_1}\right| < 1$, then $\lim_{k \to +\infty} \varepsilon_k = 0$.

Power method principle (3/4)

The sequence $\left(\frac{x_k}{\lambda_1^k}\right)_k$ "converges" to the eigenvector of the dominant eigenvalues.

Power method principle (3/4)

The sequence $\left(\frac{x_k}{\lambda_1^k}\right)_k$ "converges" to the eigenvector of the dominant eigenvalues.

We have $\|x_k\| \to +\infty$. Thus we need to control the norm: we may set it to 1 at each step (by *normalizing*, i.e., considering $y_k = \frac{x_k}{\|x_k\|}$).

To have convergence also for the case $\lambda_1 < 0$, we need to pick the right direction for the eigenvector so that it does not oscillate. We may do this by setting the largest entry in absolute value to 1 (and thus use the maximum norm).

Power method principle (3/4)

The sequence $\left(\frac{x_k}{\lambda_1^k}\right)_k$ "converges" to the eigenvector of the dominant eigenvalues.

We have $\|x_k\| \to +\infty$. Thus we need to control the norm: we may set it to 1 at each step (by *normalizing*, i.e., considering $y_k = \frac{x_k}{\|x_k\|}$).

To have convergence also for the case $\lambda_1 < 0$, we need to pick the right direction for the eigenvector so that it does not oscillate. We may do this by setting the largest entry in absolute value to 1 (and thus use the maximum norm).

The speed of convergence is given by λ_2 since $\|\varepsilon_k\| = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$

Power method principe (4/4)

How to find the dominant eigenvalue?

If φ is a linear mapping $\varphi : \mathbb{C}^{n,1} \mapsto \mathbb{C}$ such that $\varphi(u_1) \neq 0$, then

$$\frac{\varphi(x_{k+1})}{\varphi(x_k)} = \frac{\varphi\left(\lambda_1^{k+1}\left(u_1 + \varepsilon_{k+1}\right)\right)}{\varphi\left(\lambda_1^{k}\left(u_1 + \varepsilon_{k}\right)\right)} = \frac{\lambda_1^{k+1}\left(\varphi(u_1) + \varphi(\varepsilon_{k+1})\right)}{\lambda_1^{k}\left(\varphi(u_1) + \varphi(\varepsilon_{k})\right)} \;\to\; \lambda_1 \quad \text{for } k \to +\infty.$$

The mapping φ can be set to the mapping defined for all $x \in \mathbb{C}^{n,1}$ as $\varphi(x) = x_{(1)}$ where $x_{(1)}$ is the first component x (if $\varphi(u_1) \neq 0$)).

Power method - demonstration in $\mathbb{R}^{n,n}$

Let us find the dominant eigenvector of $M = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$, which satisfies the conditions of power method.

The exact solution is $u_1 = (1, \sqrt{2}+1) = \frac{1}{\sqrt{2}+1}(\sqrt{2}-1, 1)$, with eigenvalue $\lambda_1 = 3 + \sqrt{2}$.

k	$\widehat{x}_k$	$\|\widehat{x}_k - \widehat{x}_{k-1}\|_\infty$
0	$(1.0, 1.0)$	-
1	$(0.5999999999999998, 1.0)$	0.4
2	$(0.4782608695652174\,6, 1.0)$	0.121739130435
3	$(0.43689320388349517, 1.0)$	0.0413676656817
4	$(0.4223194748358862\,2, 1.0)$	0.0145737290476
5	$(0.4171202375061851, 1.0)$	0.0051992373297

In the calculations, the maximum entry in absolute value is set to 1 at each step and the convergence criterion $\|\widehat{x}_k - \widehat{x}_{k-1}\|_\infty < 10^{-2}$.

Power method - demonstration in $\mathbb{C}^{n,n}$ (1/2)

Let us consider the matrix

$$M = \begin{pmatrix} 36408 + 16769i & -5412 - 2481i & 107256 + 49397i & -492 - 214i \\ -10656 - 5164i & 1584 + 762i & -31392 - 15210i & 144 + 66i \\ -12876 - 5954i & 1914 + 881i & -37932 - 17539i & 174 + 76i \\ 4329 - 262i & -643 + 39i & 12753 - 771i & -58 + 6i \end{pmatrix}$$

The eigenvalues are $-2i$, $-i$, $3i/2$ and $3/2$.

Power method - demonstration in $\mathbb{C}^{n,n}$ (1/2)

Let us consider the matrix

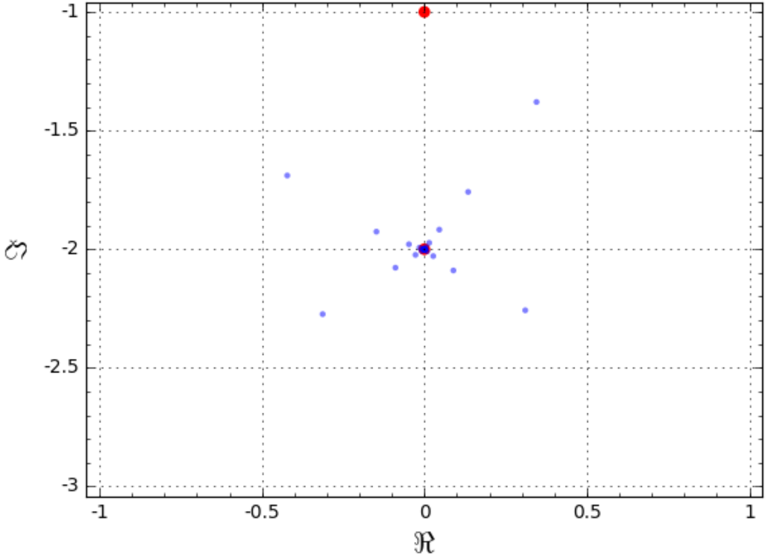
$$M = \begin{pmatrix} 36408+16769i & -5412-2481i & 107256+49397i & -492-214i \\ -10656-5164i & 1584+762i & -31392-15210i & 144+66i \\ -12876-5954i & 1914+881i & -37932-17539i & 174+76i \\ 4329-262i & -643+39i & 12753-771i & -58+6i \end{pmatrix}$$

The eigenvalues are $-2i$, $-i$, $3i/2$ and $3/2$.

Let us fix the accuracy at $\varepsilon = 10^{-6}$. The last 7 iterations of $\lambda_1^{(k)}$ are:

0.000047758815096087 - 1.99991424541241 *i*
-0.0000479821875446196 - 1.99998019901599 *i*
-0.0000272650944159076 - 2.00002375338328 *i*
0.0000271520045767515 - 2.00002973125038 *i*
0.0000154506695115737 - 1.99997272532314 *i*
-0.0000152424622193764 - 1.99999349337182 *i*

Power method - demonstration in $\mathbb{C}^{n,n}$ (2/2)



Power method: other eigenvalues

Suppose that using the power method we found the dominant eigenvalue λ_1 and its correspoding (normalized) eigenvector u_1. How can we find the other eigenvalues?

Power method: other eigenvalues

Suppose that using the power method we found the dominant eigenvalue λ_1 and its correspoding (normalized) eigenvector u_1. How can we find the other eigenvalues?

Suppose that the matrix M is normal (i.e., that $MM^* = M^*M$, where M^* is the conjugate transpose of M). Then its eigenvectors are orthogonal.

Power method: other eigenvalues

Suppose that using the power method we found the dominant eigenvalue λ_1 and its correspoding (normalized) eigenvector u_1. How can we find the other eigenvalues?

Suppose that the matrix M is normal (i.e., that $MM^* = M^*M$, where M^* is the conjugate transpose of M). Then its eigenvectors are orthogonal.

We can consider a new matrix M' defined as:

$$M' := M - \lambda_1 u_1 \cdot u_1^*$$

Power method: other eigenvalues

Suppose that using the power method we found the dominant eigenvalue λ_1 and its correspoding (normalized) eigenvector u_1. How can we find the other eigenvalues?

Suppose that the matrix M is normal (i.e., that $MM^* = M^*M$, where M^* is the conjugate transpose of M). Then its eigenvectors are orthogonal.

We can consider a new matrix M' defined as:

$$M' := M - \lambda_1 u_1 \cdot u_1^*$$

The matrix M' has u_1 as eigenvector, but the associated eigenvalue is 0, indeed:

$$M'u_1 = Mu_1 - \lambda_1 u_1 \cdot \|u_1\|^2 = \lambda_1 u_1 - \lambda_1 u_1 = 0.$$

Power method: other eigenvalues

Suppose that using the power method we found the dominant eigenvalue λ_1 and its correspoding (normalized) eigenvector u_1. How can we find the other eigenvalues?

Suppose that the matrix M is normal (i.e., that $MM^* = M^*M$, where M^* is the conjugate transpose of M). Then its eigenvectors are orthogonal.

We can consider a new matrix M' defined as:

$$M' := M - \lambda_1 u_1 \cdot u_1^*$$

The matrix M' has u_1 as eigenvector, but the associated eigenvalue is 0, indeed:

$$M'u_1 = Mu_1 - \lambda_1 u_1 \cdot \|u_1\|^2 = \lambda_1 u_1 - \lambda_1 u_1 = 0.$$

We can now apply the power method to the matrix M'. The dominant eigenvalue of M' will be the second largest (in absolute value) eigenvalue of M.

QR factorization and QR algorithm (1/2)

The power method is not suitable to find all eigenvalues of a given matrix M.

QR factorization and QR algorithm (1/2)

The power method is not suitable to find all eigenvalues of a given matrix M.

Other algorithms are based on the fact that similar matrices have the same eigenvalues. The goal of QR algorithm is to construct a sequence $(M_k)_{k=0}^{\infty}$ of similar matrices in the following way:

$$M_0 = M \quad \text{and} \quad M_k = P_k M_{k-1} P_k^{-1} \; k \in \mathbb{N},$$

where each P_k is a regular matrix, $M_k \to M_\infty$ and for M_∞ is easy to find the eigenvalues (for instance, M_∞ is upper triangular).

QR factorization and QR algorithm (2/2)

The **QR factorization** consists in expressing a real (or complex) matrix $M \in \mathbb{R}^{n,n}$ as a product

$$M = Q \cdot R$$

where Q is an orthogonal matrix (unitary if $M \in \mathbb{C}^{n,n}$) and R is upper triangular.

QR factorization and QR algorithm (2/2)

The **QR factorization** consists in expressing a real (or complex) matrix $M \in \mathbb{R}^{n,n}$ as a product

$$M = Q \cdot R$$

where Q is an orthogonal matrix (unitary if $M \in \mathbb{C}^{n,n}$) and R is upper triangular.

There exist several algorithms to compute such a factorization (Gram-Schmidt, LR algorithm, ...)

QR factorization and QR algorithm (2/2)

The **QR factorization** consists in expressing a real (or complex) matrix $M \in \mathbb{R}^{n,n}$ as a product

$$M = Q \cdot R$$

where Q is an orthogonal matrix (unitary if $M \in \mathbb{C}^{n,n}$) and R is upper triangular.

There exist several algorithms to compute such a factorization (Gram-Schmidt, LR algorithm, ...)

The **QR algorithm** consists in applying such a factorization at any step, that is for every k we have

$$M_k = Q_k \cdot R_k$$

and we define

$$M_{k+1} := R_k Q_k = Q_k^{-1} Q_k R_k Q_k = Q_k^{-1} M_k Q_k.$$

We start the iteration with $M_0 = M$. Every matrix M_k is similar to the previous matrix M_{k-1} in the sequence, so that all matrices have the same eigenvalues. Under certain conditions M_k converges to a triangular matrix.