1 26th of September 2018 — F. Poloni

1.1 Orthogonality (II)

In the previous lecture we introduced some sufficent conditions for matrix orthogonality.

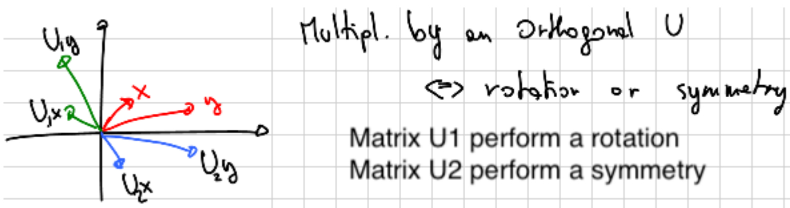
Theorem 1.1. *Let $U \in M(n, \mathbb{R})$ be an orthogonal matrix and let $x \in \mathbb{R}^n$. Then $\|Ux\| = \|x\|$.*

Proof. Instead of proving that $\|Ux\| = \|x\|$ we will prove $\|Ux\|^2 = \|x\|^2$:

$$\|Ux\|^2 = (Ux)^T(Ux) \overset{(1)}{=} x^T U^T U x = x^T I_n x = x^T x = \|x\|$$

where $\overset{(1)}{=}$ follows from the definition of transpose of a product. □

Geometrically an orthogonal preserve the norm, so a matrix A represents a symmetry or a rotations on vector x and these operations do not alter the size of vectors.



Definition 1.1 (Orthogonality)**.** *Let $x, y \in R^n$ we say that x and y are **orthonormal** if $< x, y >= 0$.*

Definition 1.2 (Orthonormality)**.** *Let $x, y \in R^n$ we say that x and y are **orthonormal** if $< x, y >= 0$ and $\|x\| = \|y\| = 1$.*

Fact 1.2. *Let us take $U \in M(n, \mathbb{R})$ such that U is orthogonal. Then its columns $U^1, U^2, \ldots, U^n$ are **orthonormal** and the same holds for its rows.*

$$U^{i^T} U^j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

and

$$U_i {U_j}^T = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Fact 1.3. *Let $U, V \in M(n, \mathbb{R})$, such that U and V are orthogonal, then UV is orthogonal. Orthogonal are closed under the product.*

Proof. $(UV)^T(UV) = V^T U^T U V = V^T I_n V = V^T V = I_n$ □

Fact 1.4. *We will often deal with tall thin rectangular matrices with orthonormal columns:*

$$U_1 = [u_1 \;\; u_2 \;\; \ldots \;\; u_n] \in \mathbb{R}^{m \times n} \quad (m \geq n)$$

There exists a matrix U_2 s.t. $[U_1 \; U_2]$ is square orthogonal.

1.2 Eigenvalues / Eigenvector

Definition 1.3 (Eigenvectors and eigenvalues)**.** *Let $A \in M(n, \mathbb{R})$ and let $x \neq 0 \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.*

*If $Ax = \lambda x$ we say that x is an **eigenvector** of **eigenvalue** λ.*

Fact 1.5. *Let $A \in M(n, \mathbb{R})$ (real triangular matrix). The eigenvalues of A are the scalars on the diagonal.*

NOTE: Eigenvectors and eigenvalues are interesting because we can use it to get a special decomposition of a matrix A.

Eigendecomposition of a matrix:
For almost almost all matrices $A \in \mathbb{R}^{n \times n}$ under some conditions we can decompose A as:

$$A = V \Lambda V^{-1}$$

$$A = V \Lambda V^{-1} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

where v_i, $\forall i = 1, \dots, n$ are eigenvectors of A of eigenvalue λ_i and w_i = rows of V^{-1}.

Another way to see the diagonalized form of A is the following:

$$A = V \Lambda V^{-1} = \sum_{i=1}^{n} v_i \lambda_i w_i^T =$$

$$\boxed{v_1}\ \boxed{\lambda_1}\ \boxed{w_1^T} + \boxed{v_2}\ \boxed{\lambda_2}\ \boxed{w_2^T} + \cdots + \boxed{v_n}\ \boxed{\lambda_n}\ \boxed{w_n^T}$$

Something on Matlab ...

Notice that in Matlab the eigenvalues and eigenvectors of a matrix are computed using the command `[V, Lambda] = eig(U)` and this operation has a computational complexity of $O(n^3)$.
We can check that the matrix `A` is equal to the decomposition in this way:
`A - V * Lambda * inv(V)` or `norm(A - V * Lambda * inv(V))` (both should be near to zero).

Notice that not all matrices $A \in M(n, \mathbb{R})$ allow a diagonal decomposition. It may happen that such a matrix is diagonalizable in $\mathbb{C}$ and its eigenvalues are complex.

Fact 1.6. *If this factorization with eigenvalues and eigenvectors holds, then:*

$$A = V\Lambda V^{-1} \Longrightarrow Av_i = v_i\lambda \ , \ \forall i = 1, \dots, n$$

This decomposition tell us the behavior under repeated application of a matrix A to a vector x. This process allow to scale a general vector x.

Something on Matlab ...

e.g. `A = [1 1; 1 1]` and `x = [1 1]`
then `A * x` is equal to `[2 2]'` and `A * A * x` is equal to `[4 4]'`.

Fact 1.7. *If $A \in M(n, \mathbb{R})$ is diagonalizable (aka may be written as $A = V\Lambda V^{-1}$) then:*

$$A^k x = \sum_{i=1}^{n} v_i \lambda_i^k w_i^T$$

Proof.

Linear algebra view point:

$$\begin{aligned}
A^k x &= AA \dots Ax \\
&= V\Lambda \cancel{V^{-1}} \cancel{V} \Lambda \cancel{V^{-1}} \dots \cancel{V} \Lambda V^{-1} x \\
&= V\Lambda^k V^{-1} x \\
&= V \begin{pmatrix} {\lambda_1}^k & & & \\ & {\lambda_2}^k & & \\ & & \ddots & \\ & & & {\lambda_n}^k \end{pmatrix} V^{-1} x
\end{aligned} \tag{1.1}$$

□

NOTE: if A is not square, Av, λv have different sizes and it doesn't make sense to talk about eigenvalues.

What can go wrong with eigenvalue decomposition

1. the eigenvalue decomposition is highly non-unique, we can:
    - reorder eigenvalues/vectors;
    - replace an eigenvector v_i with $2v_i$, $3.5v_i$, $\dots$;
    - for matrices with repeated eigenvalues, even more possibilities:
      e.g. $I = VIV^{-}1$ for every invertible V.

2. some matrices have only complex eigenvalues: e.g. $\begin{pmatrix} 2 & 4 \\ -3 & 3 \end{pmatrix}$;

3. some matrices have fewer eigenvectors than we want and we can't use eigenvalue decomposition: e.g. $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Now, thanks to the eigenvalue decomposition we can prove the following:

Theorem 1.8. *Let $A \in M(n, \mathbb{R})$. If $|\lambda_i| < 1$ for all eigenvalues λ_i of A then $\lim_{k\to\infty} A^k x = 0$.*

Theorem 1.9. *Let $A \in M(n, \mathbb{R})$. If $\forall \lambda_i$ eigenvalues of A $|\lambda_i| < |\lambda_1|$ then $A^k x \approx V^1 {\lambda_1}^k \alpha_1$.*

Fact 1.10. *Let $A \in M(n, \mathbb{R})$ be a diagonalizable matrix and let:*

$$A = V\Lambda V^{-1} = \begin{pmatrix} V^1 & V^2 & \cdots & V^n \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

Let us now consider a reordering of V's columns and apply the same permutations to the "diagonal vector" of Λ such that:

$$\hat{V} = \begin{pmatrix} V^2 & V^1 & V^3 \cdots & V^n \end{pmatrix} \hat{\Lambda} = \begin{pmatrix} \lambda_2 & & & & \\ & \lambda_1 & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ & & & & \lambda_n \end{pmatrix}$$

A can be diagonalized through such $\hat{V}$ and $\hat{\Lambda}$: $A = V\Lambda V^{-1} = \hat{V}\hat{\Lambda}\hat{V}^{-1}$.

Moreover, in the case of repeated eigenvalues.

Fact 1.11. *Let $A \in M(n, \mathbb{R})$ a diagonalizable matrix such that $A = V\Lambda V^{-1}$, where $\lambda_1 = \lambda_2$ (without loss of generality). Then V can be replaced by* $\widetilde{V} = \begin{pmatrix} V^1 + V^2 & V^1 - V^2 & V^3 & \cdots & V^n \end{pmatrix}$.

Theorem 1.12 (Spectral theorem)**.** *Let $A \in S(n, \mathbb{R})$ (A is a real symmetric matrix). Then A is diagonalizable $A = U\Lambda U^-1$, where eigenvalues are all real numbers and we can take U orthogonal matrix.*

For symmetric matrices, nothing goes wrong: eigenvalues decomposition always exists (Spectral theorem). The matrix will not have complex eigenvalues and not fewer eigenvectors. We can choose an U orthogonal because we said it was possible to reorder eigenvalues/vectors and replace eigenvector v_i.

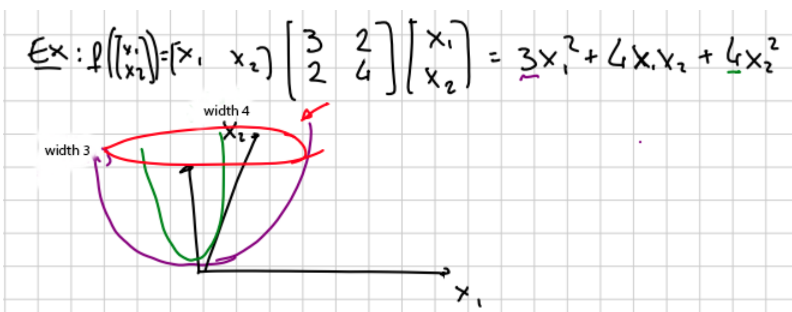
Something on Matlab ...

If we have an symmetric matrix `B` and we compute `[V, D] = eig(B)`, Matlab will always return an orthogonal matrix V.

Quadratic forms: for a fixed symmetric matrix $Q = Q^T$, consider $x \in \mathbb{R}^n$ and $Q \in \mathbb{R}^{n\times n}$ $f(x) = x^T Q x$ (Geometric idea: paraboloids):

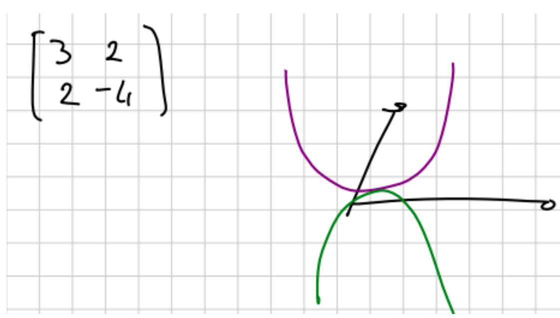
Let's see two example in a Geometric point of view:
Example 1:

$$f(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



Example 2:

$$\text{with a matrix } Q = \begin{pmatrix} 3 & 2 \\ 2 & -4 \end{pmatrix}$$



Fact 1.13. *Let $Q \in S(n, \mathbb{R})$ (For a fixed symmetric matrix) and let $x \in \mathbb{R}^n$. Then:*

$$\lambda_{min}\|x\|^2 \leq x^T Q x \leq \lambda_{max}\|x\|^2$$

where λ_{max} and λ_{min} are respectively the eigenvalue of maximum value and the eigenvalue of minimum value.

Proof.

Easy case with $Q = \Lambda$ diagonal:

$$x^T Q x = x^T \begin{pmatrix} \lambda_2 & & & & \\ & \lambda_1 & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ & & & & \lambda_n \end{pmatrix} x = \lambda_1 {x_1}^2 + \lambda_2 {x_2}^2 + \cdots + \lambda_n {x_n}^2$$

It is obvious that this sum is bounded by:

$$\lambda_{\min}({x_1}^2 + {x_2}^2 + \cdots + {x_n}^2) \leq \lambda_1 {x_1}^2 + \lambda_2 {x_2}^2 + \cdots + \lambda_n {x_n}^2 \leq \lambda_{\max}({x_1}^2 + {x_2}^2 + \cdots + {x_n}^2)$$

The following holds: $\lambda_{\min}({x_1}^2 + {x_2}^2 + \cdots + {x_n}^2) = \lambda_{\min} x^T x = \lambda_{\min}\|x\|^2$ and, on the other hand, $\lambda_{\max}({x_1}^2 + {x_2}^2 + \cdots + {x_n}^2) = \lambda_{\max} x^T x = \lambda_{\max}\|x\|^2$ and this proves the fact in the special case of diagonal matrix Q.

General case: Let us represent Q through its eigendecomposition: $A = U\Lambda U^{-1} = U\Lambda U^T$, where U is an orthogonal matrix.

$$x^T Q x = x^T U \Lambda U^T x \overset{(1)}{=} y^T \Lambda y$$

where $\overset{(1)}{=}$ is due to the change of variable $y = U^T x$ (that implies $y^T = x^T U$).

By the same argument used in the diagonal case:

$$\lambda_{\min}\|y\|^2 \leq y^T \Lambda y \leq \lambda_{\max}\|y\|^2$$

Now the point is that if we can replace $\|y\|^2$ with $\|x\|^2$ we have proved the theorem. In fact this is true, due to the orthogonality of matrix U ($UU^T = U^T U = I$.

$\square$

Corollary 1.14. *Let $Q \in S(n, \mathbb{R})$ and let $x \in \mathbb{R}^n$. If $x \neq 0$, $\lambda_{min} \leq \frac{x^T Q x}{\|x\|^2} \leq \lambda_{max}$, where λ_{max} and λ_{min} are respectively the eigenvalue of maximum value and the eigenvalue of minimum value.*

Definition 1.4 (Positive semidefinite)**.** *Let $Q \in S(n, \mathbb{R})$. If $\lambda_i \geq 0$ for each eigenvalue of Q then $x^T Qx \geq 0$ for each vector x. Q is called positive semidefinite.*

$$x^T Qx \geq 0 \|x\|^2 \geq 0$$

This is Iff, so the reverse holds: if $x^T Qx \geq 0$ for all x, then eigenvalues of Q are ≥ 0.

Definition 1.5 (Positive definite)**.** *Let $Q \in S(n, \mathbb{R})$. If $\lambda_i > 0$ for each eigenvalue of Q then $x^T Qx > 0$ for each vector $x \neq 0$. Q is called positive semidefinite.*

$$x^T Qx \geq \lambda_{min} \|x\|^2 > 0 \|x\|^2 = 0$$

This is Iff, so the reverse holds: if $x^T Qx > 0$ for all x, then eigenvalues of Q are > 0.

Fact 1.15. *Let $Q \in S(n, \mathbb{R})$. Iff Q is* ***positive semidefinite*** *then $\lambda \geq 0 \; \forall \lambda$ eigenvalue of Q iff Q is* ***positive semidefinite****. On the other hand, all eigenvalues are* ***strictly*** *positive iff Q is positive definite.*

Proof.
Let's prove that iff Q is **positive semidefinite** then $\lambda \geq 0 \; \forall \lambda$ by contradiction:
if Q has a eigenvalue $\lambda_i < 0$ then:

$$v_i^T Q v_i = v_i^T \lambda v_i = \lambda_i \|v_i\|^2 < 0$$

for the eigenvector v_i associated to λ_i.

$\square$

Fact 1.16. *Let $B \in M(m, n, \mathbb{R})$ (possibly rectangular), $B^T B \in S(n, \mathbb{R})$ is a valid product and gives a square, symmetric matrix and is positive semidefinite.*

Proof.

Symmetry: $(B^T B)^T = B^T (B^T)^T = B^T B$

Positive definite: $x^T B^T B x = (Bx)^T (Bx) = \|Bx\|^2 \geq 0$

$\square$

Corollary 1.17. *The same holds for BB^T, since we can define $C = B^T$.*

Fact 1.18. *Let $Q \in S(n, \mathbb{R})$. $A \succeq 0$ and A invertible iff Q is* ***strictly positive definite****.*

Something on Matlab . . .

In order to check if a matrix A is positive definite in Matlab we can look at its eigenvalues (cfr. `eig(A)`).

NOTE for complex matrices:
Most of these properties work also for matrices with complex entries, with one change: replace A^T with $\overline{A^T}$ (transpose + entrywise conjugate). Often denoted with A^* or A^H.

The norm of a complex vector:

$$\|x\|_2^2 = x^* x = \overline{x_1} x_1 + \overline{x_2} x_2 + \cdots + \overline{x_n} x_n = |x_1|^2 + \cdots + |x_n|^2 \text{ which is always real} \geq 0$$

Some terminology changes:

- $UU^* = I$: unitary matrix (orthogonal + complex)
- $Q = Q^*$: Hermitian matrix (capital letter, after Charles Hermite).