2-norm and orthogonality

Many choices of norms — often we favor

$x^T x$

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_m^2} = \sqrt{x^T x}.$$

(The subscript here identifies the norm — not an exponent.)

Reason (at least from my side): there are lots of matrices U such that $\|Ux\|_2 = \|x\|_2$ for each x.

Orthogonal matrices

A square $U \in \mathbb{R}^{m \times m}$ is called orthogonal if:

- $U^T U = I$,
- $UU^T = I$,
- $U^{-1} = U^T$.

(It is sufficient to check one of these.)

Orthogonal matrices

$\langle x, y \rangle \quad x \cdot y$

If U is orthogonal, then $\|Ux\|_2 = \|x\|_2$, and $(Ux)^T(Uy) = x^T y$.

Geometric idea Applying U corresponds to a rotation or a symmetry: lengths and angles are preserved.

The columns $u_1, u_2, \dots, u_m$ of an orthogonal matrix $U = \left[u_1 \mid u_2 \mid \cdots \mid u_m \right]$ are orthonormal (why?):

$$u_i^T u_j = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

orthogonal

$u_i^T u_i = \|u_i\|_2^2 = 1$

$(1+2) =$ orthonormal

...and so are its rows.

Sometimes, vectors $u_1, u_2, \dots, u_m$ such that $u_i^T u_j = 0$ when $i \neq j$, without the second condition, are called orthogonal. I know, it's confusing.

$$(AB)^T = B^T A^T$$

$m \times n \quad n \times p$

$n \times m$

$p \times m \quad p \times n$

If A, B are square,

$$(ABC)^{-1} = C^{-1} B^{-1} A^{-1}$$

$$(AB)^{-1} = B^{-1} A^{-1}$$

(If LHS is invertible, RHS is invertible too)

$$\left\| \begin{bmatrix}1\\2\end{bmatrix} \right\| = \sqrt{1^2 + 2^2} = \sqrt{5} = 2.23\ldots$$

$$U^T U = I$$

$$\|Ux\|_2 = \sqrt{(Ux)^T (Ux)} = \sqrt{x^T \underbrace{U^T U}_{= I} x} =$$

$$= \sqrt{x^T x} = \|x\|_2$$

$$\begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

Product of orthogonal matrices

The product of two orthogonal matrices is an orthogonal matrix:

$$\underbrace{(UV)}_{= V^T U^T}{}^T (UV) = V^T \overbrace{U^T U}^{= I} V = \overbrace{V^T V}^{= I.} = I.$$

Notice: For any two matrices, $(AB)^T = B^T A^T$.

Similarly, for square A, B, $(AB)^{-1} = B^{-1}A^{-1}$ ('shoe-sock identity').

Orthogonal columns

We will often deal with 'tall thin' rectangular matrices with orthonormal columns:

$$U_1 = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \in \mathbb{R}^{m \times n}. \qquad (m \geq n)$$

These can be thought as blocks of orthogonal matrices: $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$.

Some exercises to get used to them:

1. Does $U_1^T U_1 = I$? Does $U_1 U_1^T = I$? (Which sizes would those I's need to be?)
2. Does $\boxed{\|U_1 v\|_2 = \|v\|_2}$ hold for each $v \in \mathbb{R}^n$?
3. Does $\boxed{\|w^T U_1\| = \|w^T\|}$ for each $w \in \mathbb{R}^m$?

$$\begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}^T \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}$$

$$\begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & \cdots & u_1^T u_n \\ u_2^T u_1 & & & \\ \vdots & & \ddots & \\ & & & u_n^T u_n \end{bmatrix}$$

$$= I$$

$$\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} \begin{bmatrix} q_1^T & q_2^T & q_3^T & q_4^T \end{bmatrix} = \begin{bmatrix} q_1 q_1^T & q_1 q_2^T & & \\ \vdots & \vdots & \ddots & \\ & & & q_4 q_4^T \end{bmatrix}$$

Eigenvalues / vectors

Given a square matrix $A \in \mathbb{R}^{m\times m}$, if $A\mathbf{v} = \lambda \mathbf{v}$ for $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^m$, then we call λ eigenvalue and $\mathbf{v}$ eigenvector.

You should recall from linear algebra: many matrices A can be written as $\quad Av_i = \lambda_i v_i \quad i = 1, 2, \dots m$

$$A = V\Lambda V^{-1} = \begin{bmatrix} \vert & \vert & & \vert \\ v_1 & v_2 & \cdots & v_m \\ \vert & \vert & & \vert \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & & 0 & \lambda_m \end{bmatrix} \begin{bmatrix} \text{---} & w_1 & \text{---} \\ \text{---} & w_2 & \text{---} \\ & \vdots & \\ \text{---} & w_m & \text{---} \end{bmatrix}$$

$$= v_1\lambda_1 w_1^T + v_2\lambda_2 w_2^T + \cdots + v_m\lambda_m w_m^T.$$

(Here, w_i = rows of V^{-1}.)

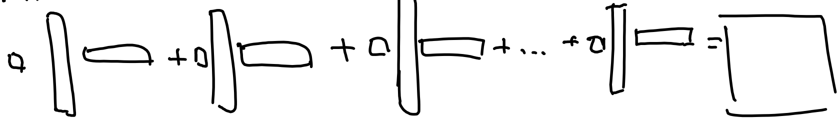
Geometric idea: in a suitable basis, A is diagonal.

`[V, D] = eig(A)` costs $O(n^3)$.

$$\left[\begin{array}{c|c|c} v_1 & v_2 & \cdots v_m \end{array}\right]\begin{bmatrix}\lambda_1 & 0 & 0 & \cdots & 0\\ 0 & \lambda_2 & & & 0\\ \vdots & & \ddots & & 0\\ 0 & \cdots & 0 & & \lambda_m\end{bmatrix} = \left[\begin{array}{c|c|c} \lambda_1 v_1 & \lambda_2 v_2 & \lambda_3 v_3 \cdots \lambda_m v_m \end{array}\right]$$

$$\left[\begin{array}{c|c|c} \lambda_1 v_1 & \lambda_2 v_2 \cdots & \lambda_m v_m \end{array}\right]\begin{bmatrix} w_1^T \\ \hline w_2^T \\ \hline \vdots \\ \hline w_m^T \end{bmatrix} =$$

$$= \lambda_1 v_1 w_1^T + \lambda_2 v_2 w_2^T + \cdots + \lambda_m v_m w_m^T$$



$$\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix}$$

What do eigenvalues tell us

$$\boxed{x \mapsto Ax}$$

Behavior under repeated application of a matrix:

$$A^k\mathbf{x} = (V\Lambda \cancel{V^{-1}})(\cancel{V}\Lambda \cancel{V^{-1}})\cdots(\cancel{V}\Lambda V^{-1}) = V\Lambda^k V^{-1}\mathbf{x}$$

$$= V \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_m^k \end{bmatrix} V^{-1}\mathbf{x}$$

$(-0.5)^k \to 0$

$(0.7)^k \to 0$

for $k \to \infty$

Theorem

If $|\lambda_i| < 1$ for all eigenvalues λ_i, $\lim_{k\to\infty} A^k = 0$

(we can prove it also for non-diagonalizable matrices using the Schur form)

Worst (or best) case: eigenvectors.

(Matlab example — using scripts and `for` cycles).

$$\text{state} = \begin{bmatrix} F_{n-1} \\ F_n \end{bmatrix}$$

$$\text{next}\left(\begin{bmatrix} F_{n-1} \\ F_n \end{bmatrix}\right)$$

What can go wrong

Eigendecomposition is highly non-unique: we can:

- Reorder eigenvalues/vectors
- Replace an eigenvector v_i with $2v_i$, $-3.5v_i, \ldots$
- For matrices with repeated eigenvalues, even more possibilities: e.g., $I = VIV^{-1}$ for every invertible V.

Some matrices have only complex eigenvalues: $\begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}$.

Some matrices have fewer eigenvectors than we want: $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Neat result: for symmetric matrices, nothing goes wrong.

Spectral theorem

If $A = A^T$, then $A = U \Lambda U^{-1}$, where eigenvalues λ_i are all real, and we can take U orthogonal.

Quadratic forms

For a fixed symmetric matrix A, we can consider $f(x) = x^T Ax$.
Geometric idea: paraboloids. See more with prof. Frangioni.

Theorem

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

Proof: if $A = \Lambda$ diagonal,

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}^T \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_m x_m^2.$$

If we replace all λ_i with $\lambda_{\min}$ it gets smaller, and vice versa.

Otherwise, $x^T Ax = x^T(U\Lambda U^T)x = y^T \Lambda y$ for $y = U^T x$ with $\|y\| = \|x\|$.

Positive definiteness

Theorem

$\lambda_{\min}\|x\|^2 \leq x^T Ax \leq \lambda_{\max}\|x\|^2$, or $\lambda_{\min} \leq \frac{x^T Ax}{\|x\|^2} \leq \lambda_{\max}$

or, alternatively,

$$\lambda_{\min} \leq z^T Az \leq \lambda_{\max} \text{ for each } z \text{ with } \|z\| = 1.$$

In particular,

If $\lambda_i \geq 0$ for each eigenvalue of A, then $x^T Ax \geq 0$ for each vector x (A is called positive semidefinite).

If $\lambda_i > 0$ for each eigenvalue of A, then $x^T Ax > 0$ for each vector $x \neq 0$ (A is called positive definite).

Properties of $B^T B$

For any $B \in \mathbb{R}^{m \times n}$ (possibly rectangular), $B^T B$ is a valid product and gives a square, symmetric matrix (why?)

$B^T B$ is positive semidefinite: because $x^T B^T B x = \|Bx\|^2 \geq 0$.

The same properties hold also for BB^T (why?).

(Matlab examples)

Complex matrices

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

$\|x\|_2^2 = x^* x = \overline{x_1} x_1 + \overline{x_2} x_2 + \cdots + \overline{x_m} x_m$, which is always real ≥ 0.

$UU^* = I$: unitary matrix.

$A = A^*$: Hermitian matrix.

Exercises

1. Can an orthogonal matrix have an entry $U_{ij} > 1$? Why?
2. When is a diagonal matrix orthogonal?
3. When is an upper triangular matrix orthogonal? (Hint: use $U^T U = UU^T = I$.)
4. Is the inverse of an orthogonal matrix orthogonal?
5. What are eigenvalues and eigenvectors of a diagonal matrix?
6. Can you find vectors that attain each of the equality cases in $\lambda_{\min} \leq \frac{x^T A x}{\|x\|^2} \leq \lambda_{\max}$?