Vector norms

Vector norms

A norm is a way to measure the length or distance from $\mathbf{0}$ of a vector. It generalizes the absolute value. Properties:

1. $\|\mathbf{v}\| \geq 0$, with equality iff $\mathbf{v} = \mathbf{0}$;
2. $\|\mathbf{v}\alpha\| = \|\mathbf{v}\||\alpha|$ for $\alpha \in \mathbb{C}$;
3. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.

Many examples:

- $\|\mathbf{v}\|_\infty := \max_i |v_i|$;
- $\|\mathbf{v}\|_1 := \sum_i |v_i|$;
- ...

2-norm

Our preferred norm: 2-norm, or Euclidean norm

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_m^2} = \sqrt{\mathbf{v}^\top \mathbf{v}} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

(The subscript 2 is often omitted.)

Coincides with the concept of length in geometry.

Trivial but useful observation

Every vector is a multiple of a norm-1 vector:
$\mathbf{v} = \alpha \mathbf{u}$, where $\alpha = \|\mathbf{v}\|_2$ and $\mathbf{u} = \frac{1}{\alpha}\mathbf{v}$.

Orthogonal matrices

Why is the 2-norm nice? Because there are many matrices U that preserve it:

Orthogonal matrices

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

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

(Each one of these properties implies the others.)

Property

If U is orthogonal, then $\|U\mathbf{x}\|_2 = \|\mathbf{x}\|_2$, and more generally $(U\mathbf{x})^T(U\mathbf{y}) = \mathbf{x}^T\mathbf{y}$.

Geometric idea The transformation associated to U is a rotation or a mirror symmetry: lengths and angles are preserved.

Orthogonal vs. orthonormal

The columns $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ of an orthogonal matrix $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix}$ are orthonormal (why?):

$$\mathbf{u}_i^\top \mathbf{u}_j = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

. . . and so are its rows.

Sometimes, vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ such that $\mathbf{u}_i^\top \mathbf{u}_j = 0$ when $i \neq j$, without the second condition, are called orthogonal. This may be confusing.

Product of orthogonal matrices

The product of two orthogonal matrices is an orthogonal matrix.

Simple to verify if you remember a couple of facts from linear algebra:

'Shoe-sock property'

For any two matrices A, B:

- $(AB)^\top = B^\top A^\top$
- $(AB)^{-1} = B^{-1}A^{-1}$ (when A, B are square invertible)

Indeed,

$$(UV)^{-1} = V^{-1}U^{-1} = V^\top U^\top = (UV)^\top.$$

Orthogonal columns

We will often work with tall thin rectangular matrices with orthonormal columns:

$$U_0 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \in \mathbb{R}^{m \times n}. \qquad (m \geq n)$$

They are the first block of an orthogonal matrix: $U = \begin{bmatrix} U_0 & U_c \end{bmatrix}$.

Some exercises to get used to them:

1. Does $U_0^\top U_0 = I$? Does $U_0 U_0^\top = I$? (Which sizes would those I's need to be?)
2. Does $\|U_0\mathbf{v}\|_2 = \|\mathbf{v}\|_2$ hold for each $v \in \mathbb{R}^n$?
3. Does $\|\mathbf{w}^\top U_0\| = \|\mathbf{w}^\top\|$ hold for each $w \in \mathbb{R}^m$?

(Hint for proofs: consider blocks of $\begin{bmatrix} U_1 & U_2 \end{bmatrix}^\top \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$.)

Eigenvalues / vectors

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

Remember from linear algebra: almost all matrices A can be written as

$$A = V \Lambda V^{-1} = \begin{bmatrix} \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ \\ \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_m \end{bmatrix}$$
$$= \mathbf{v}_1 \lambda_1 \mathbf{w}_1^T + \mathbf{v}_2 \lambda_2 \mathbf{w}_2^T + \cdots + \mathbf{v}_m \lambda_m \mathbf{w}_m^T.$$

(Here, $\mathbf{w}_i^\top$ = rows of V^{-1}.)

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

`[V, D] = eig(A)` costs $O(m^3)$, in practice. We won't see algorithms here.

What do eigenvalues tell us

Behavior under repeated application of a matrix:

$$A^k \mathbf{x} = (V\Lambda V^{-1})(V\Lambda V^{-1})\dots(V\Lambda V^{-1}) = V\Lambda^k V^{-1}\mathbf{x}$$
$$= V \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_m^k \end{bmatrix} V^{-1}\mathbf{x}$$

More generally:

Theorem

For any polynomial $p(x) = p_0 + p_1 x + p_2 x^2 + \cdots + p_d x^2$,

$$p(A) = p_0 I + p_1 A + \cdots + p_d A^d = V \begin{bmatrix} p(\lambda_1) & & & \\ & p(\lambda_2) & & \\ & & \ddots & \\ & & & p(\lambda_m) \end{bmatrix} V^{-1}.$$

What can go wrong

Eigenvalues are well-defined for each matrix (up to reordering).

Eigenvectors are not:

- If $\mathbf{v}$ is an eigenvector, any multiple $\alpha\mathbf{v}$ is, too.
- If $\mathbf{v}, \mathbf{w}$ are two eigenvectors with the same (repeated) eigenvalue λ, then any linear combination $\alpha\mathbf{v} + \beta\mathbf{w}$ is, too.
- Extreme case: any vector is an eigenvector of the identity matrix: $I = VIV^{-1}$ for any invertible V.

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

Some do not have enough eigenvectors to form a basis: $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Neat result: for symmetric matrices, nothing goes wrong.

Spectral theorem

If $A = A^T$, we can always find U, Λ s.t. $A = U\Lambda U^{-1}$. Moreover, eigenvalues λ_i are all real, and we can choose U orthogonal.

Quadratic forms

For a fixed symmetric matrix $Q = Q^T$, consider $f(\mathbf{x}) = \mathbf{x}^\top Q\mathbf{x}$.
Geometric idea: paraboloids. See more with prof. Frangioni.

Theorem

Let Q be a symmetric matrix with minimum and maximum eigenvalue $\lambda_{\min}, \lambda_{\max}$. For any vector $\mathbf{x} \in \mathbb{R}^n$, we have

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

Proof: First an easy case: $Q = \Lambda$ 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.

General case: $\mathbf{x}^T Q\,\mathbf{x} = \mathbf{x}^T(U\Lambda U^T)\mathbf{x} = \mathbf{c}^T\Lambda\,\mathbf{c}$ for $\mathbf{c} = U^T\mathbf{x}$ with $\|\mathbf{c}\| = \|\mathbf{x}\|$.

Positive definiteness

Note: given $\mathbf{x}$, $\mathbf{c} = U^T\mathbf{x} = U^{-1}\mathbf{x}$ is the vector of its coordinates in the basis of eigenvectors U.

Theorem

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

or, alternatively,

$$\lambda_{\min} \leq \mathbf{u}^T Q\mathbf{u} \leq \lambda_{\max} \text{ for each } \mathbf{u} \text{ with } \|\mathbf{u}\| = 1.$$

In particular,

If $\lambda_i \geq 0$ for each eigenvalue of Q, then $\mathbf{x}^T Q\mathbf{x} \geq 0$ for each vector $\mathbf{x}$ (Q is called positive semidefinite, $Q \succeq 0$).

If $\lambda_i > 0$ for each eigenvalue of Q, then $\mathbf{x}^T Q\mathbf{x} > 0$ for each vector $\mathbf{x} \neq \mathbf{0}$ (Q is called positive definite, $Q \succ 0$).

Moreover, these are 'if and only if'. (Why? When does equality hold in the theorem?)

Properties of $A^T A$

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

$A^T A$ is positive semidefinite: because $\mathbf{x}^T A^T A\mathbf{x} = \|A\mathbf{x}\|^2 \geq 0$.

The same properties hold also for AA^T. (Why?)

(Matlab examples)

Complex matrices

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

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

Some terminology changes:

$UU^* = I$: unitary matrix.

$Q = Q^*$: Hermitian matrix (capital letter, after Charles Hermite).

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: consider the diagonal entries in $U^T U = UU^T = I$.)
4. Is the inverse of an orthogonal matrix orthogonal?
5. Assume λ is an eigenvalue of an orthogonal matrix; show that $|\lambda| = 1$.
6. What are eigenvalues and eigenvectors of a diagonal matrix?
7. Can you find vectors that attain the two equality cases in $\lambda_{\text{min}} \leq \frac{\mathbf{x}^T Q \mathbf{x}}{\|\mathbf{x}\|^2} \leq \lambda_{\text{max}}$?

Reference: Trefethen-Bau book, chapter 2