Linear Algebra I

Vectors and Matrices as Linear Operations on Vectors

Uwe Naumann

Informatik 12:
Software and Tools for Computational Engineering

RWTH Aachen University

Outline

Outline

Objective

- Introduction of essential concepts and terminology

Learning Outcomes

- You will understand
  - matrix products
  - vector norms and projections
  - matrices as linear operators on vectors.
- You will be able to
  - visualize concepts in $\boldsymbol{R}^2$
  - apply them to further topics in linear algebra.

Outline

For $A=(a_{i,j})\in \boldsymbol{R}^{m\times n}$, $\mathbf{v}=(v_j)\in \boldsymbol{R}^n$ we define the matrix × vector product

$$\mathbf{y}=A\cdot\mathbf{v}=(y_i)^{i=0,\ldots,m-1}=\left(\sum_{j=0,\ldots,n-1} a_{i,j}\cdot v_j\right)^{i=0,\ldots,m-1}\in \boldsymbol{R}^m$$

For $A=(a_{i,j})\in \boldsymbol{R}^{m\times n}$, $B=(b_{j,k})\in \boldsymbol{R}^{n\times p}$ we define the matrix × matrix product

$$C=A\cdot B=(c_{i,k})^{i=0,\ldots,m-1}_{k=0,\ldots,p-1}=\left(\sum_{j=0,\ldots,n-1} a_{i,j}\cdot b_{j,k}\right)^{i=0,\ldots,m-1}_{k=0,\ldots,p-1}\in \boldsymbol{R}^{m\times p}$$

Outline

Following the definition of matrix multiplication, two vector products can be defined.

For $\mathbf{a}, \mathbf{b} \in R^n$ the inner (scalar) vector product $c \in R$ is defined as

$$c = \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \cdot \mathbf{b} \quad \left(= \mathbf{b}^T \cdot \mathbf{a} = \langle \mathbf{b}, \mathbf{a} \rangle \right) \ .$$

For $\mathbf{a} \in R^m$, $\mathbf{b} \in R^n$ the outer (dyadic) vector product $C \in R^{m \times n}$ is defined as

$$C = \mathbf{a} \cdot \mathbf{b}^T \quad \left(= (\mathbf{b}^T \cdot \mathbf{a})^T \right) \ .$$

C has (both column and row) rank 1.

The magnitude of a vector $\mathbf{v} \in R^m$ is measured by its norms defined as

$$\|\mathbf{v}\|_k = \left(\sum_{i=0}^{m-1} |v_i|^k \right)^{\frac{1}{k}} \quad , \text{ e.g,}$$

1-norm: $$\|\mathbf{v}\|_1 = \sum_{i=0}^{m-1} |v_i|$$

2-norm: $$\|\mathbf{v}\|_2 = \sqrt{\sum_{i=0}^{m-1} |v_i|^2} = \sqrt{\sum_{i=0}^{m} v_i^2} = \sqrt{\mathbf{v}^T \cdot \mathbf{v}} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

∞ − norm: $$\|\mathbf{v}\|_\infty = \max_{0 \le i \le m-1} |v_i|$$

All vector norms $\| \cdot \| : \boldsymbol{R}^m \rightarrow \boldsymbol{R}$

1. are positive, i.e,

$$\forall \mathbf{v} \in \boldsymbol{R}^m : \|\mathbf{v}\| \geq 0; \quad \|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}_m$$

2. are homogeneous, i.e,

$$\forall \alpha \in \boldsymbol{R} : \|\alpha \cdot \mathbf{v}\| = |\alpha| \cdot \|\mathbf{v}\|$$

3. satisfy the triangle inequality (are subadditive), i.e,

$$\forall \mathbf{u}, \mathbf{v} \in \boldsymbol{R}^m : \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \ .$$

We assume $\|\mathbf{v}\| = \|\mathbf{v}\|_2$ (length of $\mathbf{v}$) unless stated otherwise.

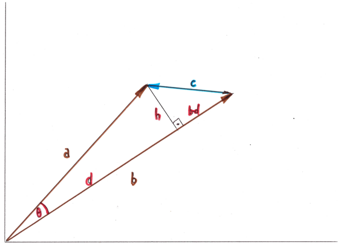
Operations on vectors are characterized by their effect on norms and relative positions, i.e. the angle spanned by two vectors in R^n. The following law of cosines turns out to be fundamental.

$$\|\mathbf{a}-\mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2 \cdot \|\mathbf{a}\| \cdot \|\mathbf{b}\| \cdot \cos(\theta)$$

For the picture on the right, Pythagoras yields $\|\mathbf{c}\|^2 = (\|\mathbf{b}\|-d)^2+h^2$ and $\|\mathbf{a}\|^2 = d^2 + h^2$ implying

$$\|\mathbf{c}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2 \cdot \|\mathbf{b}\| \cdot d$$

and with $d = \|\mathbf{a}\| \cdot \cos(\theta)$ proving the law.



A similar argument holds for h outside of the triangle spanned by $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$.

The angle θ spanned by two vectors $\mathbf{a}, \mathbf{b} \in R^m$ is characterized by

$$\cos(\theta) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} .$$

The above follows from the law of cosines with

$$\|\mathbf{a} - \mathbf{b}\|^2 = \langle \mathbf{a} - \mathbf{b}, \mathbf{a} - \mathbf{b} \rangle = \sum_{i=0}^{m-1} a_i^2 + b_i^2 - 2 \cdot a_i \cdot b_i$$

$$= \sum_{i=0}^{m-1} a_i^2 + \sum_{i=0}^{m-1} b_i^2 - 2 \cdot \sum_{i=0}^{m-1} a_i \cdot b_i = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2 \cdot \langle \mathbf{a}, \mathbf{b} \rangle$$

$$= \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2 \cdot \|\mathbf{a}\| \cdot \|\mathbf{b}\| \cdot \cos(\theta)$$

Software and Tools for Computational Engineering | RWTH AACHEN UNIVERSITY

From the law of cosines follows immediately, that the scalar projection of $\mathbf{a} \in R^m$ onto $\mathbf{b} \in R^m$ is given by

$$a_{\mathbf{b}} = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{b}\|} = \cos(\theta) \cdot \|\mathbf{a}\| \; .$$

The corresponding vector projection of is given by

$$\mathbf{a}_{\mathbf{b}} = a_{\mathbf{b}} \cdot \frac{\mathbf{b}}{\|\mathbf{b}\|}$$

(vector of length $a_{\mathbf{b}}$ pointing into same direction as $\mathbf{b}$).

The scalar projection of $\mathbf{a} \in R^m$ onto the i-th Cartesian basis vector $\mathbf{e}_i \in R^m$ is given by $a_{\mathbf{e}_i} = a_i$.

The corresponding vector projection is given by $\mathbf{a}_{\mathbf{e}_i} = a_i \cdot \mathbf{e}_i$.

Outline

Matrices $A \in \boldsymbol{R}^{m \times n}$ induce linear functions $\dot{F} : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^m$:

$$\dot{\mathbf{y}} = \dot{F}(\dot{\mathbf{v}}) \equiv A \cdot \dot{\mathbf{v}}$$

and $\bar{F} : \boldsymbol{R}^m \rightarrow \boldsymbol{R}^n$:

$$\bar{\mathbf{v}} = \bar{F}(\bar{\mathbf{y}}) \equiv A^T \cdot \bar{\mathbf{y}}$$

Linearity follows immediately from

$$\dot{F}(\dot{\mathbf{v}}_1 + \dot{\mathbf{v}}_2) = A \cdot (\dot{\mathbf{v}}_1 + \dot{\mathbf{v}}_2) = A \cdot \dot{\mathbf{v}}_1 + A \cdot \dot{\mathbf{v}}_2 = \dot{F}(\dot{\mathbf{v}}_1) + \dot{F}(\dot{\mathbf{v}}_2)$$

and

$$\dot{F}(\alpha \cdot \dot{\mathbf{v}}) = A \cdot \alpha \cdot \dot{\mathbf{v}} = \alpha \cdot A \cdot \dot{\mathbf{v}} = \alpha \cdot \dot{F}(\dot{\mathbf{v}})$$

(and similarly for $\bar{F}$).

Properties of matrices are defined in terms of their actions as linear operators on vectors.

A matrix $A \in \boldsymbol{R}^{n \times n}$ is symmetric if

$$A^T = A \, .$$

It follows that

$$\mathbf{y} = \dot{F}(\mathbf{v}) = A \cdot \mathbf{v} = A^T \cdot \mathbf{v} = \bar{F}(\mathbf{v}) \, .$$

For $A \in \boldsymbol{R}^{m \times n}$ both $A^T \cdot A$ and $A \cdot A^T$ are symmetric as

$$(A^T \cdot A)^T = A^T \cdot {A^T}^T = A^T \cdot A$$

$$(A \cdot A^T)^T = {A^T}^T \cdot A^T = A \cdot A^T \, .$$

Both matrices play fundamental roles in data analysis.

The inverse $A^{-1} \in \boldsymbol{R}^{n \times n}$ of an invertible (also: regular or non-singular) matrix $A \in \boldsymbol{R}^{n \times n}$ is defined by

$$A^{-1} \cdot A = I_n = A \cdot A^{-1} \; ,$$

where $I_n \in \boldsymbol{R}^{n \times n}$ denotes the identity in $\boldsymbol{R}^n$ mapping all vectors onto themselves.

A matrix is invertible if both its rows and its columns are linearly independent, i.e, no row / column can be written as a linear combination (weighted sum) of other rows / columns.

A matrix $A \in \boldsymbol{R}^{n \times n}$ is orthogonal if

$$A^{-1} = A^T \ .$$

$A \cdot \mathbf{v}$ amounts to a rotation of the coordinate system; i.e. angles θ between vectors $\mathbf{v}_1, \mathbf{v}_2 \in \boldsymbol{R}^n$ are preserved as

$$\begin{aligned}
\cos(\theta) &= \frac{\langle A \cdot \mathbf{v}_1, A \cdot \mathbf{v}_2 \rangle}{\|A \cdot \mathbf{v}_1\| \cdot \|A \cdot \mathbf{v}_2\|} = \frac{\mathbf{v}_1^T \cdot A^T \cdot A \cdot \mathbf{v}_2}{\sqrt{\langle A \cdot \mathbf{v}_1, A \cdot \mathbf{v}_1 \rangle} \cdot \sqrt{\langle A \cdot \mathbf{v}_2, A \cdot \mathbf{v}_2 \rangle}} \\
&= \frac{\mathbf{v}_1^T \cdot A^{-1} \cdot A \cdot \mathbf{v}_2}{\sqrt{\mathbf{v}_1^T \cdot A^T \cdot A \cdot \mathbf{v}_1} \cdot \sqrt{\mathbf{v}_2^T \cdot A^T \cdot A \cdot \mathbf{v}_2}} \\
&= \frac{\mathbf{v}_1^T \cdot \mathbf{v}_2}{\sqrt{\mathbf{v}_1^T \cdot A^{-1} \cdot A \cdot \mathbf{v}_1} \cdot \sqrt{\mathbf{v}_2^T \cdot A^{-1} \cdot A \cdot \mathbf{v}_2}} \\
&= \frac{\mathbf{v}_1^T \cdot \mathbf{v}_2}{\sqrt{\mathbf{v}_1^T \cdot \mathbf{v}_1} \cdot \sqrt{\mathbf{v}_2^T \cdot \mathbf{v}_2}} = \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \cdot \|\mathbf{v}_2\|}
\end{aligned}$$

and norms are preserved as, e.g,

$$\begin{aligned}\|A\cdot\mathbf{v}\| &= \sqrt{\langle A\cdot\mathbf{v}, A\cdot\mathbf{v}\rangle} = \sqrt{\mathbf{v}^T\cdot A^T\cdot A\cdot\mathbf{v}}\\ &= \sqrt{\mathbf{v}^T\cdot A^{-1}\cdot A\cdot\mathbf{v}} = \sqrt{\mathbf{v}^T\cdot\mathbf{v}} = \sqrt{\langle\mathbf{v},\mathbf{v}\rangle}\\ &= \|\mathbf{v}\|\end{aligned}$$

A is called orthonormal if $\|A\| = 1$.

If A is orthogonal and symmetric then

$$A^{-1} = A^T = A$$

The magnitude of matrices $A \in \mathbf{R}^{m \times n}$ is typically measured in terms of norms derived from (induced by) the corresponding vector norms, e.g,

- 1-norm (maximum absolute column sum)

$$\|A\|_1 = \max_{j=0,\ldots,n-1} \sum_{i=0}^{m-1} |a_{i,j}|$$

- 2-norm (maximum singular value / eigenvalue of $A^T \cdot A$)

$$\|A\|_2 = \sqrt{\lambda_{\max}\left(A^T \cdot A\right)}$$

- ∞-norm (maximum absolute row sum)

$$\|A\|_\infty = \max_{i=0,\ldots,m-1} \sum_{j=0}^{n-1} |a_{i,j}|$$

Vector-induced matrix norms can be shown to be submultiplicative, i.e;

$$\forall A \in \mathbf{R}^{m \times n}, B \in \mathbf{R}^{n \times k} : \|A \cdot B\| \leq \|A\| \cdot \|B\| \ .$$

Vector-Induced Matrix Norms

Matrix Norms $\Rightarrow$ Vector Norms

For $n = 1$ the matrix $A = (a_{i,j}) \in \boldsymbol{R}^{m \times 1}$ becomes a vector $\mathbf{a} = (a_i) \in \boldsymbol{R}^m$. The matrix norms turn out to be equivalent to the corresponding vector norm, e.g,

- 1-norm

$$\|A\|_1 = \sum_{i=0}^{m-1} |a_{i,0}| = \sum_{i=0}^{m-1} |a_i| = \|\mathbf{a}\|_1$$

- 2-norm

$$\|A\|_2 = \sqrt{\lambda_{\max}\left(A^T \cdot A\right)} = \sqrt{\lambda_{\max}\left(\mathbf{a}^T \cdot \mathbf{a}\right)} = \sqrt{\mathbf{a}^T \cdot \mathbf{a}} = \|\mathbf{a}\|_2$$

- ∞-norm

$$\|A\|_\infty = \max_{i=0,\dots,m-1} |a_{i,0}| = \max_{i=0,\dots,m-1} |a_i| = \|\mathbf{a}\|_\infty$$

Relevant special cases of the impact of matrices $A \in \boldsymbol{R}^{n \times n}$ as linear operators acting on vectors $\mathbf{v} \in \boldsymbol{R}^n$ are characterized as pairs of eigenvalues $\lambda \in \boldsymbol{R}$ and -vectors $\mathbf{v} \in \boldsymbol{R}^n$ satisfying the equality

$$A \cdot \mathbf{v} = \lambda \cdot \mathbf{v} \,.$$

Eigenvectors of A are simply scaled by a the magnitude of the associated eigenvalue. Negative eigenvalues yield reversal of the orientation of the associated eigenvectors $\mathbf{v}$.

For symmetric matrices $A \in \boldsymbol{R}^{n \times n}$ $(A = A^T)$ one can show that all eigenvalues are non-complex $(\in \boldsymbol{R})$ and that the largest absolute eigenvalue $|\lambda_{\max}|$ quantifies the maximum stretching of any vector $\mathbf{v}$ under A, i.e,

$$\forall \mathbf{v} \in \boldsymbol{R}^n : \quad \|A \cdot \mathbf{v}\| \leq |\lambda_{\max}| \cdot \|\mathbf{v}\| \,.$$

For $B \in \boldsymbol{R}^{m \times n}$ eigenvalues and -vectors of $B^T \cdot B$ and $B \cdot B^T$ are referred to a singular values and vectors.

Outline

Summary

- Essential terminology and concepts
  - Vectors
  - Matrices as linear operations on vectors

Next Steps

- Visualize concepts in $\boldsymbol{R}^2$.
- Refer to literature for further details.
- Continue the course to find out more ...