Tensors in Computations I

Lek-Heng Lim

University of Chicago

thanks

- Simons Foundation Award no. 663281 granted to the Institute of Mathematics of the Polish Academy of Sciences for 2021–2023
- Arieh for the opportunity to write book chapter
    - L.-H. Lim, "Tensors in computations," *Acta Numer.*, **30** (2021), pp. 555–764
- Arieh and Karolina for the opportunity to give this series of lectures

introduction

tensors

notion of tensors captures three great ideas:

- equivariance
- multilinearity
- separability

important alike in physics, mathematics, and computations

three definitions

- roughly correspond to three common definitions of a tensor
- chronologically
  ① a multi-indexed object that satisfies tensor transformation rules
  ② a multilinear map
  ③ an element of a tensor product of vector spaces
- all three definitions remain useful today
- our goals
  - introduce tensors through the lens of linear algebra
  - highlight their roles in computations

tensors in computations

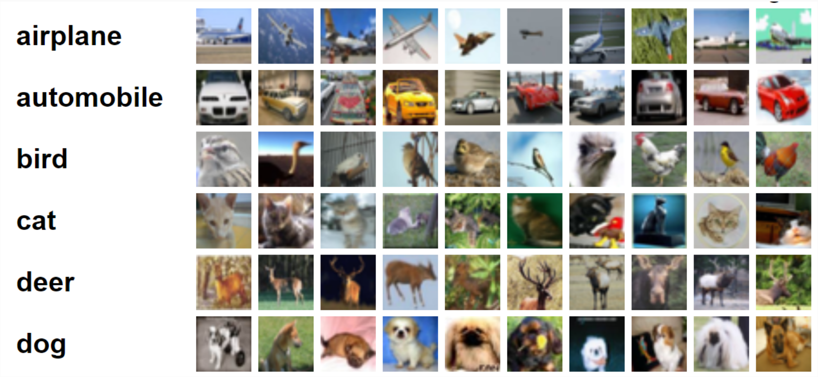
- decompositional approach to matrix computations
- interior point methods
- equivariant neural networks
- multidimensional Fourier, Laplace, Z, cosine transforms
- cryptographic multilinear maps
- tensor product bases, frames, kernels, multiresolution analyses
- fast integer and fast matrix multiplication algorithms
- fast multipole method
- separable ODEs, integral equations, Hamiltonians
- separation of variables in PDEs, integral, finite difference equations
- Grover quantum search
- Hartree–Fock approximation
- tensor networks and DMRG
- nonlinear separable convex optimization
- Smolyak's quadrature
- and more

motivation: equivariance

equivariance

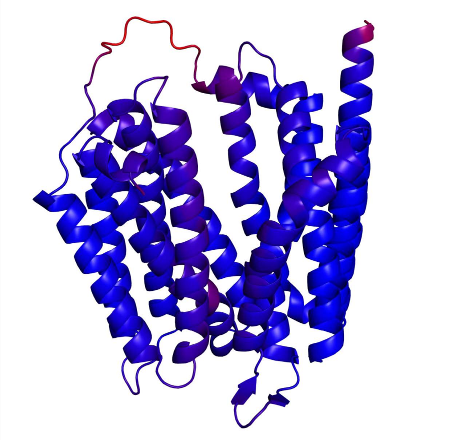
used to esoteric but not anymore

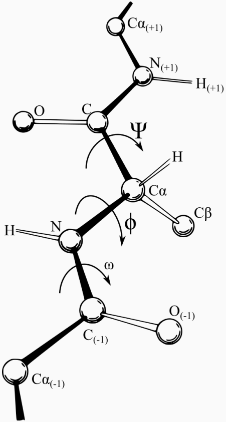
- CIFAR-10 computer vision dataset: best result obtained with equivariant neural network [Cohen–Welling, 2016]



more recently

- CASP14 protein folding competition: winning entry by Google DeepMind's AlphaFold 2 uses equivariant neural network [Jumper et al, 2020]





as old as tensors

- Woldemar Voigt, *Die fundamentalen physikalischen Eigenschaften der Krystalle in elementarer Darstellung*, Verlag Von Veit, Leipzig, 1898.

- "An abstract entity represented by an array of components that are functions of coordinates such that, under a transformation of cooordinates, the new components are related to the transformation and to the original components in a definite way."
- highlighted part = equivariance

tensors via transformation rules

what is a tensor?

- for every complex question there is an answer that is clear, simple, and wrong

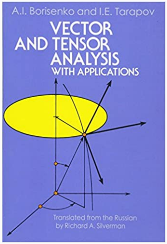
"a tensor is a multiway array"

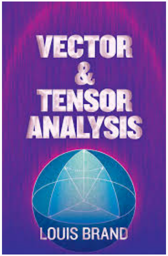
- unfortunately also widely believed — simple answer to complex question has its appeal
- indication that answer cannot be so simple: Einstein's letter to Sommerfeld, dated October 29, 1912
  - J. Earman, C. Glymour, "Lost in tensors: Einstein's struggles with covariance principles 1912–1916," *Stud. Hist. Phil. Sci.*, **9** (1978), no. 4, pp. 251–278

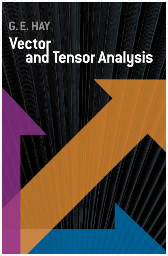
earliest definition

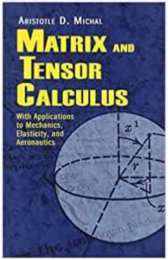
- trickiest among the three definitions
- Voigt's definition again:
    - "An abstract entity represented by an array of components that are functions of coordinates such that, under a transformation of cooordinates, the new components are related to the transformation and to the original components in a definite way"
- main issue: defines an entity by giving its change-of-bases formulas but without specifying the entity itself
- likely reason for notoriety of tensors as a difficult subject to master

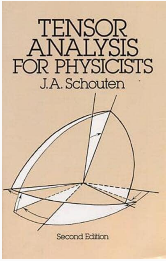
definition in Dover books c. 1950s

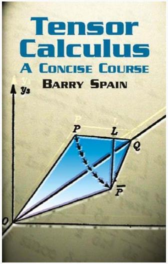


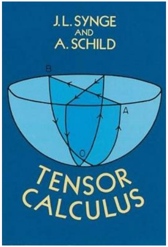


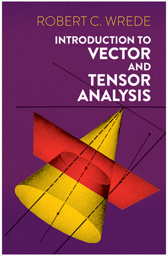












- "a multi-indexed object that satisfies certain transformation rules"

fortunately for us

- linear algebra as we know it today was a subject in its infancy when Einstein was trying to learn tensors
- vector space, linear map, dual space, basis, change-of-basis, matrix, matrix multiplication, etc, were all obscure notions back then
  - 1858: 3×3 matrix product (Cayley)
  - 1888: vector space and $n \times n$ matrix product (Peano)
  - 1898: tensor (Voigt)
- we enjoy the benefit of a hundred years of pedagogical progress
- next slides: look at tensor transformation rules in light of linear algebra and numerical linear algebra

eigen and singular values

- **eigenvalue and vectors:** $A \in \mathbb{C}^{n\times n}$, $A\mathbf{v} = \lambda\mathbf{v}$, for invertible $X \in \mathbb{C}^{n\times n}$,

$$(XAX^{-1})X\mathbf{v} = \lambda X\mathbf{v}$$

  - eigenvalue $\lambda' = \lambda$, eigenvector $\mathbf{v}' = X\mathbf{v}$, and $A' = XAX^{-1}$

- **singular values and vectors:** $A \in \mathbb{R}^{m\times n}$,

$$\begin{cases} A\mathbf{v} = \sigma\mathbf{u}, \\ A^\top\mathbf{u} = \sigma\mathbf{v} \end{cases}$$

  for orthogonal $X \in \mathbb{R}^{m\times m}$, $Y \in \mathbb{R}^{n\times n}$,

$$\begin{cases} (XAY^\top)Y\mathbf{v} = \sigma X\mathbf{u}, \\ (XAY^\top)^\top X\mathbf{u} = \sigma Y\mathbf{v} \end{cases}$$

  - singular value $\sigma' = \sigma$, left singular vector $\mathbf{u}' = X\mathbf{u}$, left singular vector $\mathbf{v}' = Y\mathbf{v}$, and $A' = XAY^\top$

matrix product and linear systems

- **matrix product:** $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times p}$, $C \in \mathbb{C}^{m\times p}$, $AB = C$, for any invertible X, Y, Z,

$$(XAY^{-1})(YBZ^{-1}) = XCZ^{-1}$$

  - $A' = XAY^{-1}$, $B' = YBZ^{-1}$, $C' = XCZ^{-1}$

- **linear system:** $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^{m}$, $A\mathbf{v} = \mathbf{b}$, for invertible X, Y,

$$(XAY^{-1})(Y\mathbf{v}) = X\mathbf{b}$$

  - $A' = XAY^{-1}$, $\mathbf{b}' = X\mathbf{b}$, $\mathbf{v}' = Y\mathbf{v}$

ordinary and total least squares

- **ordinary least squares:** $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$,

$$\min_{v\in\mathbb{R}^n} \|A\mathbf{v} - \mathbf{b}\|^2 = \min_{v\in\mathbb{R}^n} \|(XAY^{-1})Y\mathbf{v} - X\mathbf{b}\|^2$$

  for orthogonal $X \in \mathbb{R}^{m\times m}$ and invertible $Y \in \mathbb{R}^{n\times n}$
  - $A' = XAY^{-1}$, $\mathbf{b}' = X\mathbf{b}$, $\mathbf{v}' = Y\mathbf{v}$, minimum value $\rho' = \rho$
- **total least squares:** $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, then

$$\begin{aligned} &\min \{\|E\|^2 + \|\mathbf{r}\|^2 : (A+E)\mathbf{v} = \mathbf{b} + \mathbf{r}\} \\ &= \min \{\|XEY^\intercal\|^2 + \|X\mathbf{r}\|^2 : (XAY^\intercal + XEY^\intercal)Y\mathbf{v} = X\mathbf{b} + X\mathbf{r}\} \end{aligned}$$

  for orthogonal $X \in \mathbb{R}^{m\times m}$ and orthogonal $Y \in \mathbb{R}^{n\times n}$
  - $A' = XAY^\intercal$, $E' = XEY^\intercal$, $\mathbf{b}' = X\mathbf{b}$, $\mathbf{r}' = X\mathbf{r}$, $\mathbf{v}' = Y\mathbf{v}$

rank, norm, determinant, intertia

- **rank, norm, determinant:** $A \in \mathbb{R}^{m \times n}$

  $$\text{rank}(XAY^{-1}) = \text{rank}(A), \quad \det(XAY^{-1}) = \det(A), \quad \|XAY^{-1}\| = \|A\|$$

  for X and Y invertible, special linear, or orthogonal, respectively
  - determinant identically zero whenever $m \neq n$
  - $\|\cdot\|$ either spectral, nuclear, or Frobenius norm
- **positive definiteness:** $A \in \mathbb{R}^{n \times n}$ positive definite iff

  $$XAX^\top \qquad \text{or} \qquad X^{-\top}AX^{-1}$$

  positive definite for any invertible $X \in \mathbb{R}^{n \times n}$

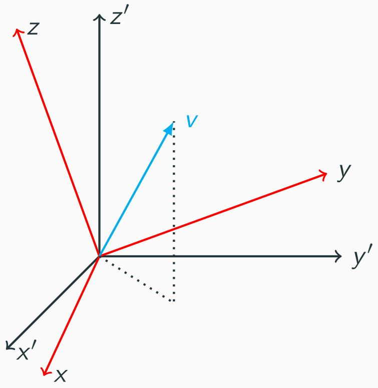
observation

- almost everything we study in linear algebra and numerical linear algebra satisfies tensor transformation rules
- different names, same thing:
  - equivalence of matrices: $A' = XAY^{-1}$
  - similarity of matrices: $A' = XAX^{-1}$
  - congruence of matrices: $A' = XAX^{\mathsf{T}}$
- almost everything we study in linear algebra and numerical linear algebra is about 0-, 1-, 2-tensors

0-, 1-, 2-tensor transformation rules

contravariant 1-tensor:	$\mathbf{a}' = X^{-1}\mathbf{a}$	$\mathbf{a}' = X\mathbf{a}$
covariant 1-tensor:	$\mathbf{a}' = X^{\mathsf{T}}\mathbf{a}$	$\mathbf{a}' = X^{-\mathsf{T}}\mathbf{a}$
covariant 2-tensor:	$A' = X^{\mathsf{T}}AX$	$A' = X^{-\mathsf{T}}AX^{-1}$
contravariant 2-tensor:	$A' = X^{-1}AX^{-\mathsf{T}}$	$A' = XAX^{\mathsf{T}}$
mixed 2-tensor:	$A' = X^{-1}AX$	$A' = XAX^{-1}$
contravariant 2-tensor:	$A' = X^{-1}AY^{-\mathsf{T}}$	$A' = XAY^{\mathsf{T}}$
covariant 2-tensor:	$A' = X^{\mathsf{T}}AY$	$A' = X^{-\mathsf{T}}AY^{-1}$
mixed 2-tensor:	$A' = X^{-1}AY$	$A' = XAY^{-1}$

simplest case: contravariant 1-tensor



- choose x-, y- and z-axes, $\mathbf{v}$ gets coordinates $\mathbf{a} \in \mathbb{R}^3$
- change axes to x'-, y'- and z'-axes with $X \in \mathrm{GL}(3)$
- nothing physical has changed, $\mathbf{v}$ still where it was
- coordinates must change in opposite way $\mathbf{a}' = X^{-1}\mathbf{a}$ to compensate

triply ambiguous

- transformation rules may mean different things

$$A' = XAY^{-1}, \quad A' = XAY^{\mathsf{T}}, \quad A' = XAX^{-1}, \quad A' = XAX^{\mathsf{T}}$$

  and more

- matrices in transformation rules may have different properties

$$X \in \mathrm{GL}(n),\ \mathrm{SL}(n),\ \mathrm{O}(n),$$
$$(X, Y) \in \mathrm{GL}(m) \times \mathrm{GL}(n),\ \mathrm{SL}(m) \times \mathrm{SL}(n),\ \mathrm{O}(m) \times \mathrm{O}(n),\ \mathrm{O}(m) \times \mathrm{GL}(n)$$

  and more

- alternative (but equivalent) forms just as common

$$A' = X^{-1}AY, \quad A' = X^{-1}AY^{-\mathsf{T}}, \quad A' = X^{-1}AX, \quad A' = X^{-1}AX^{-\mathsf{T}}$$

math perspective

- multi-indexed object $\lambda \in \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, etc, represents the tensor
- transformation rule $A' = XAY^{-1}$, $A' = XAY^{-1}$, $A' = XAX^{\mathsf{T}}$, etc, defines the tensor
- but the tensor has been left unspecified
- easily fixed with modern definitions ② and ③
- need a context in order to use definition ①
- is $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$ a tensor?
- it is a tensor if we are interested in, say, its eigenvalues and eigenvectors, in which case A transforms as a mixed 2-tensor

physics perspective

- remember definition ① came from physics — they don't ask
  - what is a tensor?

  but
  - is stress a tensor?
  - is deformation a tensor?
  - is electromagnetic field strength a tensor?
- unspecified quantity is placeholder for physical quantity like stress, deformation, etc
- it is a tensor if the multi-indexed object satisfies transformation rules under change-of-coordinates, i.e., definition ①
- makes perfect sense in a physics context
- is $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$ a tensor?
- it is a tensor if it represents, say, stress, in which case A transforms as a contravariant 2-tensor

higher order

multilinear matrix multiplication

- $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$
- $X \in \mathbb{R}^{m_1 \times n_1}, Y \in \mathbb{R}^{m_2 \times n_2}, \ldots, Z \in \mathbb{R}^{m_d \times n_d}$
- define

$$(X, Y, \ldots, Z) \cdot A = B$$

where $B \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ given by

$$b_{i_1 \cdots i_d} = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \cdots \sum_{j_d=1}^{n_d} x_{i_1 j_1} y_{i_2 j_2} \cdots z_{i_d j_d} a_{j_1 \cdots j_d}$$

- $d = 1$: reduces to $X\mathbf{a} = \mathbf{b}$ for $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$
- $d = 2$: reduces to

$$(X, Y) \cdot A = XAY^{\mathsf{T}}$$

higher-order transformation rules 1

- $X_1 \in \mathrm{GL}(n_1), X_2 \in \mathrm{GL}(n_2), \ldots, X_d \in \mathrm{GL}(n_d)$
- covariant d-tensor transformation rule:

$$A' = (X_1^{\mathsf{T}}, X_2^{\mathsf{T}}, \ldots, X_d^{\mathsf{T}}) \cdot A$$

- contravariant d-tensor transformation rule:

$$A' = (X_1^{-1}, X_2^{-1}, \ldots, X_d^{-1}) \cdot A$$

- mixed d-tensor transformation rule:

$$A' = (X_1^{-1}, \ldots, X_p^{-1}, X_{p+1}^{\mathsf{T}}, \ldots, X_d^{\mathsf{T}}) \cdot A$$

- contravariant order p, covariant order $d - p$, or type $(p, d - p)$

higher-order transformation rules 2

- when $n_1 = n_2 = \cdots = n_d = n$, $X \in \mathrm{GL}(n)$
- covariant d-tensor transformation rule:

$$A' = (X^{\mathsf{T}}, X^{\mathsf{T}}, \dots, X^{\mathsf{T}}) \cdot A$$

- contravariant d-tensor transformation rule:

$$A' = (X^{-1}, X^{-1}, \dots, X^{-1}) \cdot A$$

- mixed d-tensor transformation rule:

$$A' = (X^{-1}, \dots, X^{-1}, X^{\mathsf{T}}, \dots, X^{\mathsf{T}}) \cdot A$$

- getting ahead of ourselves, with definition ②, difference is between multilinear

$$f : \mathbb{V}_1 \times \cdots \times \mathbb{V}_d \to \mathbb{R} \quad \text{and} \quad f : \mathbb{V} \times \cdots \times \mathbb{V} \to \mathbb{R}$$

change-of-cooordinates matrices

- $X_1, \dots, X_d$ or X may belong to:

$$\begin{aligned}
\mathrm{GL}(n) &= \{X \in \mathbb{R}^{n\times n} : \det(X) \neq 0\} \\
\mathrm{SL}(n) &= \{X \in \mathbb{R}^{n\times n} : \det(X) = 1\} \\
\mathrm{O}(n) &= \{X \in \mathbb{R}^{n\times n} : X^{\mathsf{T}} X = I\}, \\
\mathrm{SO}(n) &= \{X \in \mathbb{R}^{n\times n} : X^{\mathsf{T}} X = I,\ \det(X) = 1\} \\
\mathrm{U}(n) &= \{X \in \mathbb{C}^{n\times n} : X^* X = I\} \\
\mathrm{SU}(n) &= \{X \in \mathbb{C}^{n\times n} : X^* X = I,\ \det(X) = 1\} \\
\mathrm{O}(p,q) &= \{X \in \mathbb{R}^{n\times n} : X^{\mathsf{T}} I_{p,q} X = I_{p,q}\} \\
\mathrm{SO}(p,q) &= \{X \in \mathbb{R}^{n\times n} : X^{\mathsf{T}} I_{p,q} X = I_{p,q},\ \det(X) = 1\} \\
\mathrm{Sp}(2n,\mathbb{R}) &= \{X \in \mathbb{R}^{2n\times 2n} : X^{\mathsf{T}} J X = J\} \\
\mathrm{Sp}(2n) &= \{X \in \mathbb{C}^{2n\times 2n} : X^{\mathsf{T}} J X = J,\ X^* X = I\}
\end{aligned}$$

- $I := I_n$ is $n \times n$ identity, $I_{p,q} := \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix} \in \mathbb{R}^{n\times n}$, $J := \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \in \mathbb{R}^{2n\times 2n}$

change-of-cooordinates matrices

- again getting ahead of ourselves with definitions ② or ③,
  - if vector spaces involve have no extra structure, then $\mathrm{GL}(n)$
  - if inner product spaces, then $\mathrm{O}(n)$
  - if equipped with yet other structures, then whatever group that preserves those structures
- e.g., $\mathbb{R}^4$ equipped with Euclidean inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_0 y_0 + x_1 y_1 + x_2 y_2 + x_3 y_3$$

  want $X \in \mathrm{O}(4)$ or $\mathrm{SO}(4)$
- e.g., $\mathbb{R}^4$ equipped with Lorentzian scalar product,

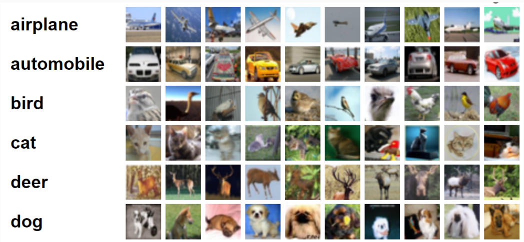
$$\langle \mathbf{x}, \mathbf{y} \rangle = x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3,$$

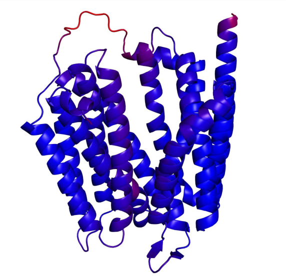
  want $X \in \mathrm{O}(1,3)$ or $\mathrm{SO}(1,3)$
- called Cartesian tensors or Lorentzian tensors respectively

transformation rule is key

why important (in machine learning)

- tensor transformation rules in modern parlance: equivariance
- we mentioned earlier equivariant neural networks





why important (in physics)

- special relativity is essentially the observation that the laws of physics are invariant under Lorentz transformations in $\mathrm{O}(1,3)$ [Einstein, 1920]
- transformation rules under $\mathrm{O}(1,3)$-analogue of Givens rotations:

$$\begin{bmatrix} \cosh\theta & -\sinh\theta & 0 & 0 \\ -\sinh\theta & \cosh\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} \cosh\theta & 0 & -\sinh\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sinh\theta & 0 & \cosh\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} \cosh\theta & 0 & 0 & -\sinh\theta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\theta & 0 & 0 & \cosh\theta \end{bmatrix}$$

  enough to derive most standard results of special relativity
- "Geometric Principle: The laws of physics must all be expressible as geometric (coordinate independent and reference frame independent) relationships between geometric objects (scalars, vectors, tensors, ...) that represent physical entitities." [Thorne, 1973]

why important (in mathematics)

- deriving higher-order tensorial analogues not a matter of just adding more indices to

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad \sum_{j=1}^{n} a_{ij} x_j = \lambda x_i, \quad \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}$$

- need to satisfy tensor transformation rules
- e.g., $A \in \mathbb{R}^{2\times 2\times 2}$ has hyperdeterminant

$$\begin{aligned}\det(A) = {} & a_{000}^2 a_{111}^2 + a_{001}^2 a_{110}^2 + a_{010}^2 a_{101}^2 + a_{011}^2 a_{100}^2 \\ & - 2(a_{000}a_{001}a_{110}a_{111} + a_{000}a_{010}a_{101}a_{111} + a_{000}a_{011}a_{100}a_{111} \\ & + a_{001}a_{010}a_{101}a_{110} + a_{001}a_{011}a_{110}a_{100} + a_{010}a_{011}a_{101}a_{100}) \\ & + 4(a_{000}a_{011}a_{101}a_{110} + a_{001}a_{010}a_{100}a_{111}),\end{aligned}$$

- preserved by transformation $A' = (X, Y, Z) \cdot A$ for $X, Y, Z \in \mathrm{SL}(2)$
- just as determinant preserved by $A' = XAY^{\mathsf{T}}$ for $X, Y \in \mathrm{SL}(n)$

tensor multiplication?

- Hadamard product:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \circ \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$$

- seems a lot more obvious than standard matrix product

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

- matrix product satisfies transformation rule for mixed 2-tensors $(XAY^{-1})(YBZ^{-1}) = X(AB)Z^{-1}$, i.e., defined on tensors
- Hadamard product undefined on tensors — depends on coordinates
- product on $\mathbb{R}^{m\times n\times p}$ or $\mathbb{R}^{n\times n\times n}$ that satisfies 3-tensor transformation rules does not exist

identity tensor?

- identity matrix $I \in \mathbb{R}^{3\times 3}$

$$I = \sum_{i=1}^{3} \mathbf{e}_i \otimes \mathbf{e}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

  with $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathbb{R}^3$ standard basis vectors
- $(Q, Q) \cdot I = QIQ^\top = I$ for $Q \in \mathrm{O}(3)$, unique up to scalar multiples
- I is a Cartesian 2-tensor
- analogue in $\mathbb{R}^{3\times 3\times 3}$ is not

$$A = \sum_{i=1}^{3} \mathbf{e}_i \otimes \mathbf{e}_i \otimes \mathbf{e}_i \in \mathbb{R}^{3\times 3\times 3}$$

  because $(Q, Q, Q) \cdot A \neq A$

identity tensor?

- analogue is

$$J = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \varepsilon_{ijk} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \in \mathbb{R}^{3\times 3\times 3}$$

  where ε_{ijk} is the Levi-Civita symbol

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) = (1,2,3),(2,3,1),(3,1,2), \\ -1 & \text{if } (i,j,k) = (1,3,2),(2,1,3),(3,2,1), \\ 0 & \text{if } i = j, j = k, k = i \end{cases}$$

- $(Q,Q,Q) \cdot J = J$ for $Q \in \mathrm{O}(3)$, unique up to scalar multiples
- J is a Cartesian 3-tensor

why important (in computations)

two simple properties:

- **group:** change-of-coordinates matrices may be multiplied/inverted:
  - if X, Y orthogonal or invertible, so is XY
  - if X orthogonal or invertible, so is X^{-1}
- **group action:** transformation rules may be composed:
  - if $\mathbf{a}' = X^{-\top}\mathbf{a}$ and $\mathbf{a}'' = Y^{-\top}\mathbf{a}'$, then $\mathbf{a}'' = (YX)^{-\top}\mathbf{a}$
  - if $A' = XAX^{-1}$ and $A'' = YA'Y^{-1}$, then $A'' = (YX)A(YX)^{-1}$

plus one more fact about the change-of-coordinate matrices (next slides)

why important (in computations)

- recall Givens rotation, Householder reflector, Gauss transform:

$$G = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos\theta & \cdots & -\sin\theta & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \sin\theta & \cdots & \cos\theta & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathrm{SO}(n),$$

$$H = I - \frac{2\mathbf{v}\mathbf{v}^{\mathsf{T}}}{\mathbf{v}^{\mathsf{T}}\mathbf{v}} \in \mathrm{O}(n), \qquad M = I - \mathbf{v}\mathbf{e}_i^{\mathsf{T}} \in \mathrm{GL}(n)$$

- $\mathbf{a}' = G\mathbf{a}$ rotation of $\mathbf{a}$ in (i, j)-plane by an angle θ
- $\mathbf{a}' = H\mathbf{a}$ reflection of $\mathbf{a}$ in the hyperplane with normal $\mathbf{v}/\|\mathbf{v}\|$
- for judiciously chosen $\mathbf{v}$, $\mathbf{a}' = M\mathbf{a} \in \operatorname{span}\{\mathbf{e}_{i+1}, \dots, \mathbf{e}_n\}$, i.e., has $(i+1)$th through nth coordinates zero

why important (in computations)

- facts about change-of-coordinate matrices in transformation rules
  - any $X \in \mathrm{SO}(n)$ is a product of Givens rotations
  - any $X \in \mathrm{O}(n)$ is a product of Householder reflectors
  - any $X \in \mathrm{GL}(n)$ is a product of elementary matrices
  - any unit lower triangular $X \in \mathrm{GL}(n)$ is a product of Gauss transforms
- in group theoretic lingo:
  - Givens roations generate $\mathrm{SO}(n)$
  - Householder reflectors generate $\mathrm{O}(n)$
  - elementary matrices generate $\mathrm{GL}(n)$
  - Gauss transforms generate lower unitriangular subgroup of $\mathrm{GL}(n)$

why important (in computations)

- algorithms in numerical linear algebra implicitly based on these:
  - apply a sequence of tensor transformation rules

$$
\begin{aligned}
&A \to X_1 A \to X_2(X_1 A) \to \cdots \to B \\
&A \to X_1^{-\mathsf{T}} A \to X_2^{-\mathsf{T}}(X_1^{-\mathsf{T}} A) \to \cdots \to B \\
&A \to X_1 A X_1^{\mathsf{T}} \to X_2(X_1 A X_1^{\mathsf{T}}) X_2^{\mathsf{T}} \to \cdots \to B \\
&A \to X_1 A X_1^{-1} \to X_2(X_1 A X_1^{-1}) X_2^{-1} \to \cdots \to B \\
&A \to X_1 A Y_1^{-1} \to X_2(X_1 A Y_1^{-1}) Y_2^{-1} \to \cdots \to B
\end{aligned}
$$

  - required X obtained as either $X_m X_{m-1} \dots X_1$ or its limit as $m \to \infty$
- caveat: in numerical linear algebra, we tend to view these transformation rules as giving matrix decompositions

examples

example: full-rank least squares

- tensor transformation rules for ordinary least squares: mixed 2-tensor $A' = XAY^{-1}$ with change-of-coordinates $(X, Y) \in \mathrm{O}(m) \times \mathrm{GL}(n)$
- method of solution essentially obtains

$$X = Q \in \mathrm{O}(m), \qquad Y = R^{-1} \in \mathrm{GL}(n)$$

  by applying a sequence of tensor transformation rules
- suppose $\mathrm{rank}(A) = n$, with sequence of tensor transformation rules

$$A \to Q_1^{\mathsf{T}} A \to Q_2^{\mathsf{T}}(Q_1^{\mathsf{T}} A) \to \cdots \to Q^{\mathsf{T}} A = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

  given by Householder QR algorithm, get

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

- practically Voigt's definition: transform problem into form where solution of transformed problem is related to original solution in a definite way

example: full-rank least squares

- minimum value is invariant Cartesian 0-tensor

$$\min \|A\mathbf{v} - \mathbf{b}\|^2 = \min \|Q^\intercal(A\mathbf{v} - \mathbf{b})\|^2 = \min \left\| \begin{bmatrix} R \\ 0 \end{bmatrix} \mathbf{v} - Q^\intercal \mathbf{b} \right\|^2$$

$$= \min \left\| \begin{bmatrix} R \\ 0 \end{bmatrix} \mathbf{v} - \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} \right\|^2 = \min \|R\mathbf{v} - \mathbf{c}\|^2 + \|\mathbf{d}\|^2 = \|\mathbf{d}\|^2$$

where

$$Q^\intercal \mathbf{b} = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}$$

- solution of transformed problem $R\mathbf{v} = \mathbf{c}$ equals original solution, and may be obtained through back substitution, i.e., a sequence

$$\mathbf{c} \to Y_1^{-1}\mathbf{c} \to Y_2^{-1}(Y_1^{-1}\mathbf{c}) \to \cdots \to R^{-1}\mathbf{c} = \mathbf{v}$$

where Y_i's are Gauss transforms

example: Krylov subspaces

- $A \in \mathbb{R}^{n \times n}$ with all eigenvalues distinct and nonzero, arbitrary $\mathbf{b} \in \mathbb{R}^n$
- change-of-coordinates matrix K whose columns are

$$\mathbf{b}, A b, A^2\mathbf{b}, \ldots, A^{n-1}\mathbf{b}$$

  is invertible, i.e., $K \in \mathrm{GL}(n)$
- transformation rule gives

$$A = K \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix} K^{-1}$$

- seemingly trivial but when combined with other techniques, give powerful iterative methods for linear systems, least squares, eigenvalue problems, or evaluating various matrix functions

example: Krylov subspaces

- why not use more obvious

$$A = X \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} X^{-1}$$

  with change-of-coordinates matrix $X \in \mathrm{GL}(n)$ given by eigenvectors?
- much more difficult to compute than K
- one way is in fact to implicitly exploit relation between K and X:

$$\begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{n-1} \\ 1 & \lambda_3 & \lambda_3^2 & \ldots & \lambda_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_m & \lambda_m^2 & \ldots & \lambda_m^{n-1} \end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix} \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{n-1} \\ 1 & \lambda_3 & \lambda_3^2 & \ldots & \lambda_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_m & \lambda_m^2 & \ldots & \lambda_m^{n-1} \end{bmatrix}^{-1}$$

example: Newton method

- equality-constrained optimization

$$\begin{array}{ll} \text{minimize} & f(\mathbf{v}) \\ \text{subject to} & A\mathbf{v} = \mathbf{b} \end{array}$$

- strongly convex $f \in C^2(\Omega)$

$$\beta I \preceq \nabla^2 f(\mathbf{v}) \preceq \gamma I$$

- Newton step $\Delta\mathbf{v} \in \mathbb{R}^n$ defined by

$$\begin{bmatrix} \nabla^2 f(\mathbf{v}) & A^\intercal \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{v} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{v}) \\ 0 \end{bmatrix}$$

- Newton decrement $\lambda(v) \in \mathbb{R}$ defined by

$$\lambda(\mathbf{v})^2 := \nabla f(\mathbf{v})^\intercal \nabla^2 f(\mathbf{v})^{-1} \nabla f(\mathbf{v})$$

example: Newton method

- linear change of coordinates $X\mathbf{v}' = \mathbf{v}$ with $X \in \mathrm{GL}(n)$
- write $g(\mathbf{v}') = f(X\mathbf{v})$, then

coordinates	contravariant 1-tensor	$\mathbf{v}' = X^{-1}\mathbf{v}$
gradient	covariant 1-tensor	$\nabla g(\mathbf{v}') = X^{\mathsf{T}}\nabla f(X\mathbf{v})$
Hessian	covariant 2-tensor	$\nabla^2 g(\mathbf{v}') = X^{\mathsf{T}}\nabla^2 f(X\mathbf{v})X$
Newton step	contravariant 1-tensor	$\Delta\mathbf{v}' = X^{-1}\Delta\mathbf{v}$
Newton iterate	contravariant 1-tensor	$\mathbf{v}_k' = X^{-1}\mathbf{v}_k$
Newton decrement	invariant 0-tensor	$\lambda(\mathbf{v}_k') = \lambda(\mathbf{v}_k)$

- Newton method is tensorial, steepest descent is not

example: Newton method

- condition number of $X^\top \nabla^2 f(X\mathbf{v})X$ can be scaled to any desired value in $[1, \infty)$ with appropriate $X \in \mathsf{GL}(n)$
- Newton step independent of the condition number of $\nabla^2 f(\mathbf{v})$
- manifests as insensitivity to condition number in finite precision
- in practice Newton method gives solutions of high accuracy for $\kappa \approx 10^{10}$ when steepest descent already fails at $\kappa \approx 20$