Multilinear algebra, tensors, and tensor decompositions

Melissa Zhang

MAT 167, UC Davis

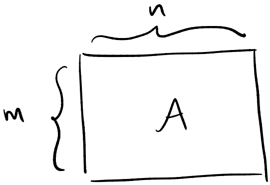
Lecture 28

Reminder

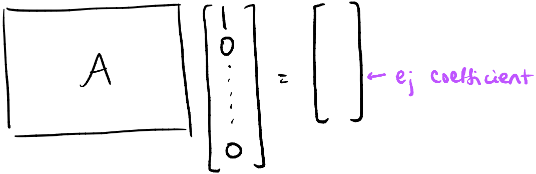
Pretty please fill out the Course Evaluation!
(Due tomorrow, June 8th)

A perspective on linear algebra

Matrix $A \in \mathbb{R}^{m \times n}$ = m-by-n array:



Entry a_{ij} of A = coefficient of e_i in Ae_j:

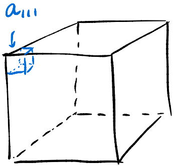


3D arrays?

$A_{ijk} = A[i][j][k]$

What if we had $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$? What would would this array mean?

$A_{ijk} = A[i][j][k]$



What would entry a_{ijk} represent?

Example: Handwritten digits

Suppose we have a training set with n images, each 16×16 pixels, manually classified into 10 classes

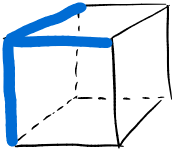
$$[0], [1], \ldots, [8], [9].$$

- Hence each image is a 16×16 matrix, and we have n of these.
- Previously, we handled this by reshaping the matrix into a long vector:
$$\mathbb{R}^{16\times 16} \to \mathbb{R}^{256}$$

Key point

Using tensors, we can keep the image as a 2D array.

Terminology



Suppose we have a tensor

$$\mathcal{A} \in \mathbb{R}^{l \times m \times n}.$$

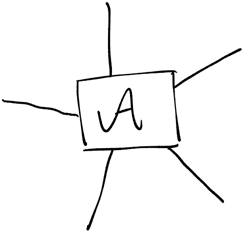
- The tensor $\mathcal{A}$ is sometimes called a *3-mode array*.
- The 3 "dimensions" of the array are called the *modes*.
- The *dimensions* [1] of $\mathcal{A}$ are l, m, n.
  - Just as you'd describe the *dimensions* of a cardboard box.

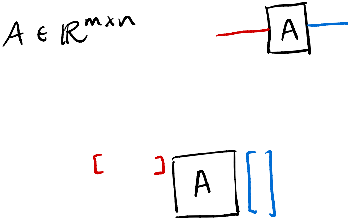
Note that a matrix is a *2-mode array* using this terminology.

[1] in the colloquial sense

d-mode tensors?

For any $d \in \mathbb{N}$, we can define a d-mode tensor. After all, a d-mode tensor is just a d-dimensional array:





d-mode tensors?

The use of tensors in data analysis applications was pioneered by researchers in psychometrics and chemometrics in the 1960s.

Examples

- $d = 2$ for our handwritten digit classification problem
- $d = 5$ for some facial recognition application

What is d for a color photo?

2 modes for pixels

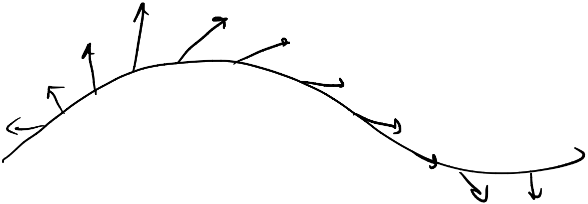
1 mode for rgb color values

$\Rightarrow d = 3$

Vector fields ⇝ tensor fields

In physics and mathematics, tensors show up all the time.

Vector field = 1-mode tensor field



Vector fields $\rightsquigarrow$ tensor fields

A taste of Riemannian geometry:

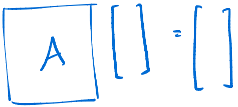
Example: Riemann curvature tensor field R

Let T_pS^2 denote the *tangent plane* to S^2 at a point p:
The Riemann curvature tensor R_p at point p is a 4-mode tensor:

$$R_p : T_pS^2 \times T_pS^2 \times T_pS^2 \to T_pS^2$$

Actually, this is $(3,1)$-tensor field because there are 3 input vectors and 1 output vector.

matrix:



1 input
1 output

(1+1 tensor)

$A \in \mathbb{R}^{m \times n} \Rightarrow 2$ tensor

Multilinear algebra

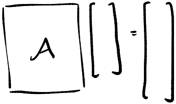
Back to 3-mode tensors:

$$\mathcal{A} \in \mathbb{R}^{l \times m \times n}$$

A priori, let's think of $\mathcal{A}$ as just a collection of coefficients.

Coefficients of what?

matrix



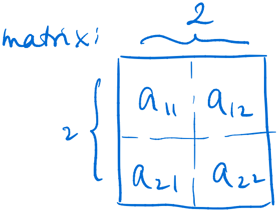
a_{ij} is coefficient of e_i in $A e_j$

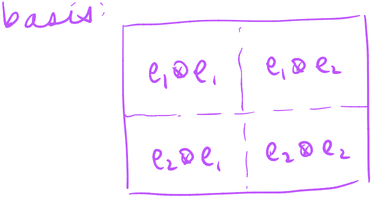
Relation to tensor products

Example

If $\mathbb{R}^2 = \langle e_1, e_2 \rangle$, then

$$\mathbb{R}^2 \otimes \mathbb{R}^2 = \langle e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2 \rangle$$





In general, if $\mathbb{R}^m = \langle e_1, e_2, \ldots, e_m \rangle$ and $\mathbb{R}^n = \langle e_1, e_2, \ldots, e_m \rangle$, then

$$\mathbb{R}^m \otimes_{\mathbb{R}} \mathbb{R}^n = \langle \{e_i \otimes e_j\}_{1 \le i \le m, 1 \le j \le n} \rangle.$$

Multilinear algebra

$$\mathcal{A} \in \mathbb{R}^{l \times m \times n}$$

Suppose we treat the first two dimensions as *input* and the last dimension as *output*. Then we think of $\mathcal{A}$ as a *multilinear transformation*:

$$\mathcal{A} : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$$

- input: 2 vectors $\vec{x} \in \mathbb{R}^l, \vec{y} \in \mathbb{R}^m$
- output: 1 vector $\vec{z} \in \mathbb{R}^n$

Why **multi**-linear?

Linear transformation $A : \mathbb{R} \to \mathbb{R}$:
For $\alpha \in \mathbb{R}$,

$$A(\alpha e_1) = \alpha A(e_1).$$

In general, $A(\sum c_i e_i) = \sum c_i A(e_i)$.

Multilinear transformation $\mathcal{A} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$:

$$\mathcal{A}(\alpha\, e_1, \beta\, e_1) = \alpha\beta\, \mathcal{A}(e_1, e_1)$$

In other words, the transformation $\mathcal{A}$ is determined by one piece of data: $\mathcal{A}(e_1, e_1) \in \mathbb{R}$.
In general, if $\mathcal{A}$ is multilinear, then it's really a transformation

$$\mathcal{A} : \mathbb{R}^l \otimes \mathbb{R}^m \to \mathbb{R}^n$$

and is determined by what it does to the tensors
$e_i \otimes e_j \in \mathbb{R}^l \otimes \mathbb{R}^m$, because all the coefficients float to the front.

Explicit example

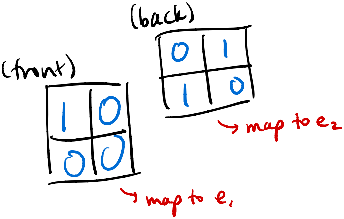
Consider the tensor

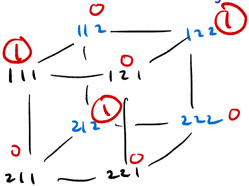
$$\mathcal{A} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$$

defined by

$$\mathcal{A}(e_1, e_1) = e_1 \qquad \mathcal{A}(e_1, e_2) = e_2$$
$$\mathcal{A}(e_2, e_1) = e_2 \qquad \mathcal{A}(e_2, e_2) = 0.$$

How can we write down the tensor $\mathcal{A}$ as a 3D array?





Explicit example

$$\mathcal{A}(*,*,1) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathcal{A}(*,*,2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

These eight coefficients a_{ijk} $(i, j, k \in \{1, 2\})$ completely describe the tensor $\mathcal{A}$, and now we can compute what the multilinear transformation does to any **pair** of vectors $v, w \in \mathbb{R}^2$:

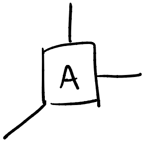
Example

Let $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 e_1 + v_2 e_2$ and $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = w_1 e_1 + w_1 e_2$. Then

$$\begin{aligned} \mathcal{A}(v, w) &= v_1 w_1 \mathcal{A}(e_1, e_1) + v_1 w_2 \mathcal{A}(e_1, e_2) \\ &\quad + v_2 w_1 \mathcal{A}(e_2, e_1) + v_2 w_2 \mathcal{A}(e_2, e_2) \\ &= v_1 w_1 e_1 + (v_1 w_2 + v_2 w_1) e_2 \\ &= \begin{bmatrix} v_1 w_1 \\ v_1 w_2 + v_2 w_1 \end{bmatrix}. \end{aligned}$$

Matrix decomposition $\rightsquigarrow$ tensor decomposition

Now back to viewing $\mathcal{A}$ just as a tensor (not a transformation):



Matrix SVD can be generalized to tensors in multiple ways. One such generalization is **higher order SVD** (HOSVD).

HOSVD

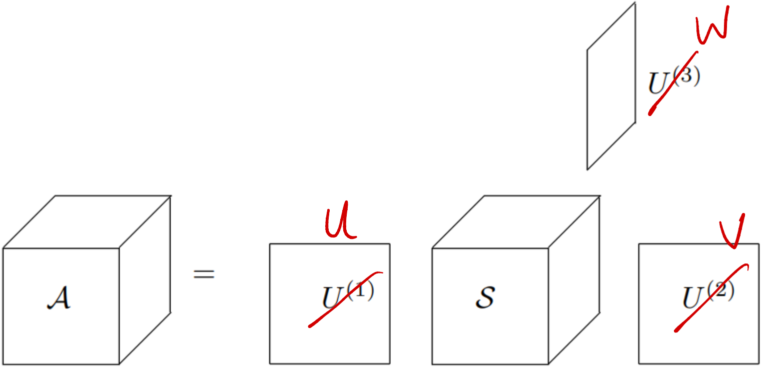


Figure 8.2. *Visualization of the HOSVD.*

HOSVD

$$\mathcal{A}_{ijk} = \sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} u_{ip} v_{jq} w_{kr} \mathcal{S}_{pqr}$$

The number (singular value) $\mathcal{S}_{pqr}$ reflects the variation by the combination of singular vectors u_p, v_q, and w_r.

Tensor decomposition

We've decomposed $\mathcal{A}$ into a sum of **rank-1 tensors** $u_p v_q w_r$!

We can now obtain a low-rank tensor approximation by choosing to keep only the components $u_p v_q w_r$ whose singular values $\mathcal{S}_{pqr}$ are significant or large.

HOSVD and handwritten digits

Example 8.4 in the text:

Given 131 handwritten "3" digits, where each image is a 16×16 matrix, compute the HOSVD of the $16 \times 16 \times 131$ tensor to get these singular values:

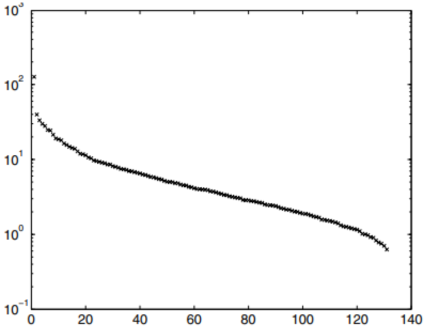


Figure 8.3. *The singular values in the digit (third) mode.*

HOSVD and handwritten digits

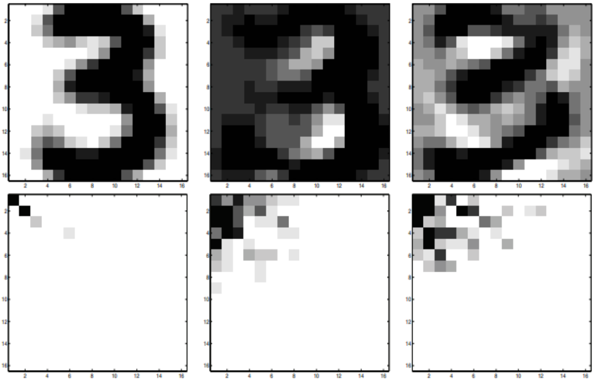


Figure 8.4. *The top row shows the three matrices A_1, A_2, and A_3, and the bottom row shows the three slices of the core tensor, $\mathcal{S}(:,:,1)$, $\mathcal{S}(:,:,2)$, and $\mathcal{S}(:,:,3)$ (absolute values of the components).*

These are the top three basis matrices for handwritten 3's.

Good luck on finals, and thanks for a great quarter!

Pretty please fill out the Course Evaluation!
(Due tomorrow, June 8th!)

In fact, you can fill it out right meow!