Tensors in Data Analysis

Winter Semester 17/18
Block course

Dr. Pauli Miettinen

Tensors in Data Analysis

Winter Semester 2017/2018

9.–13. October

Exam: 20 October. Oral exam; time & place are announced later. Registration to HISPOS by 13 October. Exam material: these lecture notes & Kolda and Bader's "Tensor Decompositions and applications" (available from the course web page).

Tutorials: 4 problem sheets, 4 questions per sheet. To mark a problem solved, you must be present at the tutorial the whole time and be willing to present your solution. You must mark at least 8 problems to be allowed to take the final exam. There are no bonus points, but trying to solve all problems and attending all tutorials is strongly recommended.

Lectures: Lectures take place 10:15–11:45 and 12:30–14:00 every day. Tutorials follow at 14:15 trough 15:30 Tuesday to Friday. (On Tuesday, the tutorial session starts at 14:30.) Lecture notes are made available following the day's lectures.

Topics: A tentative list of topics is

- Monday: Tensor algebra and manipulations; products with matrices and vectors
- Tuesday: CP decomposition; algorithms; tensor rank
- Wednesday: Variations of the CP; applications thereof; Tucker3 decomposition
- Thusday: Variations of the Tucker; applications thereof; tensor train decomposition
- Friday: Applications; CORCONDIA

What is a Tensor?

Tensors are:

- Multi-linear mappings

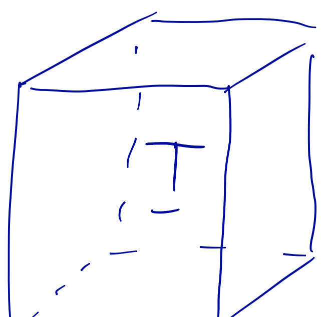
$$\underline{T}: \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \dots \times \mathbb{R}^{I_N} \to \mathbb{R}$$

- Multi-dimensional arrays

$$\underline{T} \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \dots \times \mathbb{R}^{I_N}$$

$$\underline{T} = (t_{i_1 i_2 \dots i_N})$$

- Generalisations of matrices
- N-ary relations
- Cubes



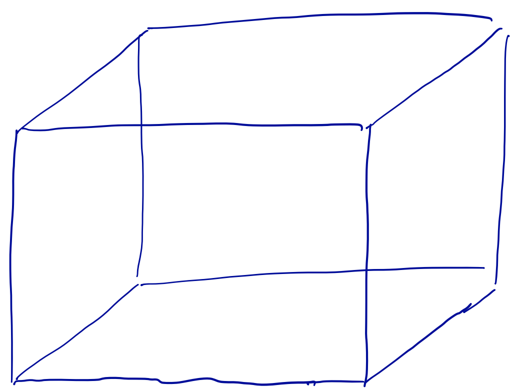
- Matrices and vectors are also tensors

Terminology

We say that a tensor is N-way array. A matrix is a 2-way array, a vector is a 1-way array. Other sources can use N-dimensional instead, but then a 3-dimensional vector is a 1-dimensional tensor. Yet others say rank-N, but we have another meaning for the word "rank".

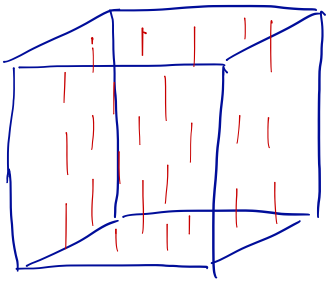
A 3-way tensor can be N-by-M-by-K dimensional (like vectors are n-dimensional).

3-way
5-by-7-by-3

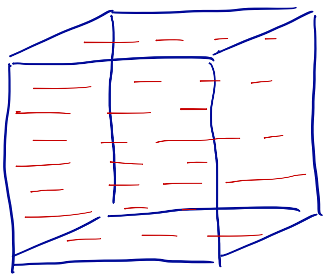


Modes and slices

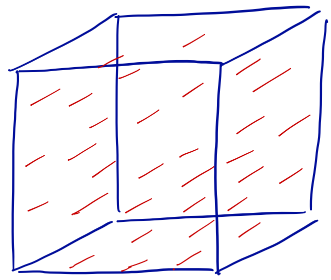
A 3-way tensor has three *modes*, *rows*, *columns*, and *tubes*. Vectors along any of the modes are called *fibres*. Matrices along any two modes are called *slices*.



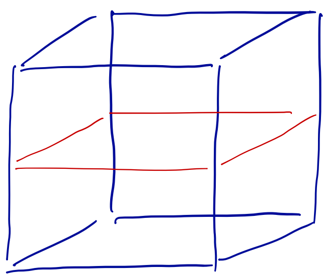
Mode-1 (column) fibres



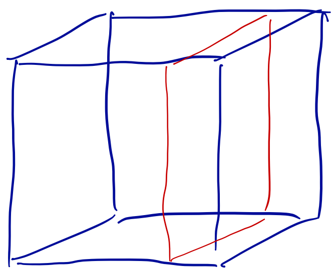
Mode-2 (row) fibres



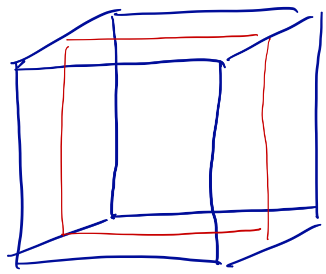
Mode-3 (tube) fibres



Horizontal slice



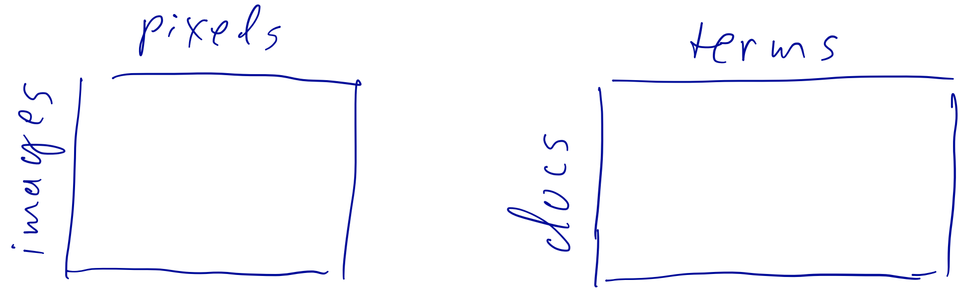
Lateral slice



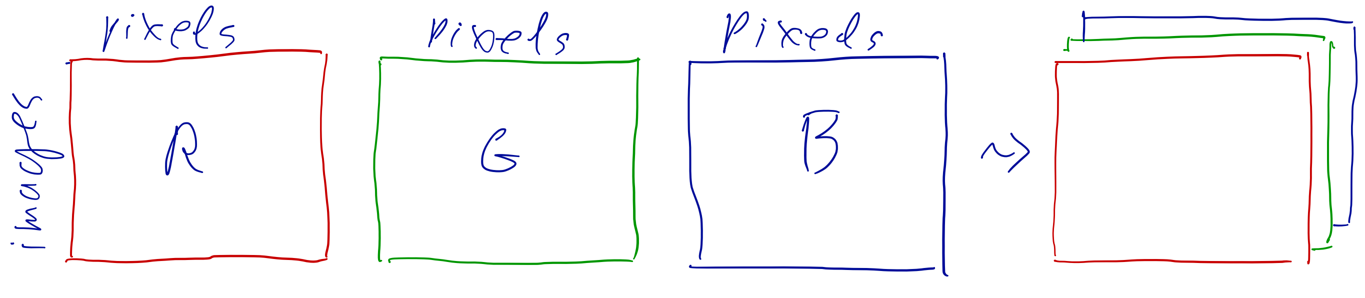
Frontal slice

Why Tensors?

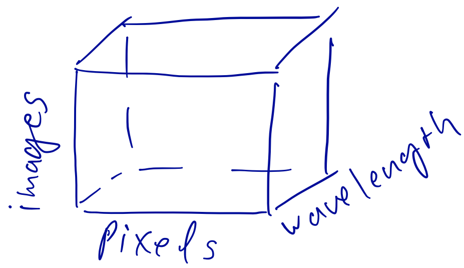
Matrices are a common way to store or interpret data in data analysis. One way to interpret matrices is to think them as binary relations or functions. Think, for example, values of pixels in images or frequency of terms in documents.



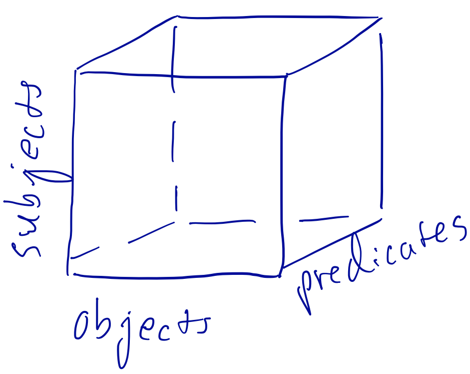
Tensors generalize this relation from binary to multi-ary. For example, if we want to store the values of different colours in images, we need three matrices, for red, green, and blue.



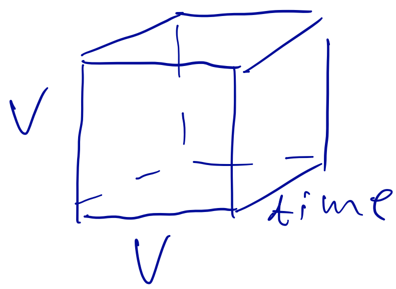
In hyperspectral imaging, the tensor would have more than just three colours.



Tensors can also be thought to contain many binary relations between the same entities. In <u>Resource Definition Framework</u> (RDF), all data is stored in subject-predicate-object (or (s,p,o)) triples. These triples can be seen as a 3-way binary tensor where each frontal slice has one predicate (i.e. relation between subjects and objects).



Another use of tensors is to store matrices (or tensors) over different discrete time stamps. For example, a series of adjacency matrices can be stored in a vertices-by-vertices-by-days tensor.



This is perhaps the most common way to "generalize" matrix-form data to tensors. Care must be taken, however, as most tensor techniques are invariant over permutations of the indices, that is, they don't take into account that time has a strict ordering.