e.g. if $T \in \mathbb{R}^{4\times 3\times 2}$

$$\begin{array}{c} \\ i \downarrow \end{array}
\begin{array}{c} j \rightarrow \\ \begin{matrix} a_{111} & a_{121} & a_{131} \\ a_{211} & a_{221} & a_{231} \\ a_{311} & a_{321} & a_{331} \\ a_{411} & a_{421} & a_{431} \end{matrix} \end{array}
\qquad \overset{k}{\nearrow} \qquad
\begin{matrix} a_{112} & a_{122} & a_{132} \\ a_{212} & a_{222} & a_{232} \\ a_{312} & a_{322} & a_{332} \\ a_{412} & a_{422} & a_{432} \end{matrix}$$

mode 1 will be $I \times JK$ matrix $= 4 \times (3\cdot 2)$

$$\begin{array}{c} \\ i \downarrow \end{array}
\begin{array}{c} \overset{j}{\longrightarrow} \qquad\qquad\qquad \overset{k}{\frown} \\ \left[\begin{array}{ccc:ccc} a_{111} & a_{121} & a_{131} & a_{112} & a_{122} & a_{132} \\ a_{211} & a_{221} & a_{231} & a_{212} & a_{222} & a_{232} \\ a_{311} & a_{321} & a_{331} & a_{312} & a_{322} & a_{332} \\ a_{411} & a_{421} & a_{431} & a_{412} & a_{422} & a_{432} \end{array}\right] \end{array}$$

mode 1 fibers

mode 2 will be $J \times IK$ matrix $= 3 \times (4\cdot 2)$

$$\begin{array}{c} \\ j \downarrow \end{array}
\begin{array}{c} \overset{i}{\longrightarrow} \qquad\qquad\qquad \overset{k}{\frown} \\ \left[\begin{array}{cccc:cccc} a_{111} & a_{211} & a_{311} & a_{411} & a_{112} & a_{212} & a_{312} & a_{412} \\ a_{121} & a_{221} & a_{321} & a_{421} & a_{122} & a_{222} & a_{322} & a_{422} \\ a_{131} & a_{231} & a_{331} & a_{431} & a_{132} & a_{232} & a_{332} & a_{432} \end{array}\right] \end{array}$$

mode 2 fibers

mode 3 will be $K \times IJ$ $= 2 \times (4\cdot 3)$

$$\begin{array}{c} \\ k \downarrow \end{array}
\begin{array}{c} \overset{i}{\longrightarrow} \qquad\qquad \overset{j}{\frown} \qquad\qquad \overset{j}{\frown} \\ \left[\begin{array}{cccc:cccc:cc} a_{111} & a_{211} & a_{311} & a_{411} & a_{121} & a_{221} & a_{321} & a_{421} & a_{131} & a_{231} \\ a_{112} & a_{212} & a_{312} & a_{412} & a_{122} & a_{222} & a_{322} & a_{422} & a_{132} & a_{232} \end{array}\right. \end{array}$$

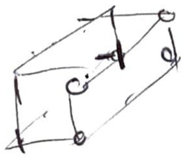
$$\begin{matrix} a_{331} & a_{431} \\ a_{332} & a_{432} \end{matrix}$$

mode 3 fibers

Outer products

- given vectors $\vec{u} \in \mathbb{R}^m$, $\vec{v} \in \mathbb{R}^n$, $\vec{w} \in \mathbb{R}^p$, the outer product $\vec{u} \circ \vec{v} \circ \vec{w}$ is the order 3 tensor whose ijk-th entry is $u_i v_j w_k$

e.g. $T = \begin{bmatrix}1\\1\end{bmatrix} \circ \begin{bmatrix}1\\0\end{bmatrix} \circ \begin{bmatrix}1\\-1\end{bmatrix}$

$=$  $\xrightarrow{\text{unfold}} T_1 = \left[\begin{array}{cc:cc} 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right]$

- analogue of uv^T in 2d
  - sometimes written $\vec{u} \circ \vec{v}$

- a matrix A is rank 1 exactly when it can be written $\vec{u}\vec{v}^T$ for some (nonzero) $\vec{u}, \vec{v}$

- we define a (order 3) tensor T to be rank 1 if it can be written $T = \vec{u} \circ \vec{v} \circ \vec{w}$ for some $\vec{u}, \vec{v}, \vec{w} \neq \vec{0}$.

- In such a T: all columns multiples of one another
  - " rows " " "
  - and more: all faces " " "

- We can express unfoldings of rank 1 tensors $\vec{u} \circ \vec{v} \circ \vec{w}$ in terms of $\otimes$

- Let's extract general formulae from an example.

- Consider

$$T = \vec{u} \circ \vec{v} \circ w = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \circ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \circ \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{matrix} u_1 v_1 w_1 & u_1 v_2 w_1 & u_1 v_3 w_1 \\ u_2 v_1 w_1 & u_2 v_2 w_1 & u_2 v_3 w_1 \end{matrix} \quad \begin{matrix} u_1 v_1 w_2 & u_1 v_2 w_2 & u_1 v_3 w_2 \\ u_2 v_1 w_2 & u_2 v_2 w_2 & u_2 v_3 w_2 \end{matrix}$$

$$T_1 = \left[\begin{array}{ccc:ccc} u_1 v_1 w_1 & u_1 v_2 w_1 & u_1 v_3 w_1 & u_1 v_1 w_2 & u_1 v_2 w_2 & u_1 v_3 w_2 \\ u_2 v_1 w_1 & u_2 v_2 w_1 & u_2 v_3 w_1 & u_2 v_1 w_2 & u_2 v_2 w_2 & u_2 v_3 w_2 \end{array}\right]$$

observe: $w \otimes v = \begin{bmatrix} w_1 \vec{v} \\ w_2 \vec{v} \end{bmatrix} = \begin{bmatrix} w_1 v_1 \\ w_1 v_2 \\ w_1 v_3 \\ w_2 v_1 \\ w_2 v_2 \\ w_2 v_3 \end{bmatrix}$

so $(w \otimes v)^T = [w_1 v_1 \;\; w_1 v_2 \;\; w_1 v_3 \;\; w_2 v_1 \;\; w_2 v_2 \;\; w_2 v_3]$

so $u \otimes (w \otimes v)^T = \begin{bmatrix} u_1 (w \otimes v)^T \\ u_2 (w \otimes v)^T \end{bmatrix} =$

$= T_1$

thus is also $\vec{u}\,(\vec{w} \otimes \vec{v})^T$

! note

likewise can check:

$$T_2 = \vec{v} \otimes (\vec{w} \otimes \vec{u})^T = \vec{v}(\vec{w} \otimes \vec{u})^T$$

notice reversal of alphabetical order.

$$T_3 = \vec{w} \otimes (\vec{v} \otimes \vec{u})^T = \vec{w}(\vec{v} \otimes \vec{u})^T$$

CP Approximation to T

↳ tensor analogue of finding best rank k approx'n to a matrix A using SVD.

— Given $A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T$ of rank r, consider problem

Find X of rank k $(\leq r)$ s.t.

$$\|A - X\|_F \quad \text{is minimized}$$

— SVD says: sol'n very nice!

given by $X = A_k = \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_k \vec{u}_k \vec{v}_k^T$

$$= \sigma_1 \vec{u}_1 \circ \vec{v}_1 + \cdots + \sigma_k \vec{u}_k \circ \vec{v}_k$$

using $\circ$ notation

How to generalize to tensors T?

↳ first need to define rank(T) in general.

- Recall: SVD gives:

rank(A) = least # of rank 1 matrices needed to sum to A.

- We'll take this as our <u>definition</u> of rank for tensors!

i.e. define, given T, (order 3)

rank(T) = shortest length r of a sum of form

$$\sum_{i=1}^{r} \vec{u}_i \circ \vec{v}_i \circ \vec{w}_i$$

equalling T.

= least number of rank 1 tensors summing to T.

↳ now we can state an analogous "rank k approximation problem" for T

<u>Matrix</u>: minimize $\|A - X\|_F$

over all X's of rank k

<u>Tensor</u>: $\boxed{\begin{array}{l} \text{minimize } \|T - X\|_F \\ \text{over all } X\text{'s of the form} \\ X = \sum_{i=1}^{k} \vec{u}_i \circ \vec{v}_i \circ \vec{w}_i \end{array}}$ "CP approx'n problem" call this (*)

where k is fixed $\leq \text{rank}(T)$. (imagine $k \ll \text{rank}(T)$)

<u>note</u>: $\|T\|_F$ means exactly what you'd think for a tensor; if a_{ijk} is ijk-th entry then

$$\|T\|_F = \sqrt{\sum_{i,j,k} a_{ijk}^2}$$

Differences between matrix / tensor:

↳ no canonical "nice" sol'n to problem like in matrix case

↳ don't insist $\vec{u}$'s are either orthogonal or normal! likewise for $\vec{v}$'s, $\vec{w}$'s

↳ if we change k, $\vec{u}$'s, $\vec{v}$'s, $\vec{w}$'s will change in general - won't just add terms at the end.

<u>Main issue</u>: nothing so nice as the SVD exists for tensors in general.

- So How to answer our approx'n problem?

- First step: turn into a matrix problem, by unfolding our expression $(*)$ for $\mathcal{X}$.

- Let $U = [\vec{u}_1 | \cdots | \vec{u}_K]$

$V = [\vec{v}_1 | \cdots | \vec{v}_K]$

$W = [\vec{w}_1 | \cdots | \vec{w}_K]$

Then: $X_1 =$ mode 1 unfolding of $\mathcal{X}$

$$= \sum_{i=1}^{K} u_i \otimes (\vec{w}_i \otimes \vec{v}_i)^T = \sum_{i=1}^{K} \vec{u}_i (\vec{w}_i \otimes \vec{v}_i)^T$$

by our unfolding formula for $(\vec{u} \circ \vec{v} \circ \vec{w})_1$

We can condense this further using Khatri Rao product $\odot$

$=$

$$= [\vec{u}_1 | \cdots | \vec{u}_K] [\vec{w}_1 \otimes \vec{v}_1 | \cdots | \vec{w}_K \otimes \vec{v}_K]^T \quad \text{(why?)}$$

$$= U (W \odot V)^T$$

likewise can check

$$X_2 = V(W \odot U)^T$$

$$X_3 = W(V \odot U)^T$$

<u>Now</u>: we're trying to minimize

$$\|T - X\|_F$$

for our given tensor T.

This quantity is equal to

$$\|T_1 - X_1\|_F = \|T_2 - X_2\|_F = \|T_3 - X_3\|_F$$
$$= \|T - X\|_F$$

since all of these arrays have the same entries.

Hence if we can minimize any of

$$\|T_1 - X_1\|_F = \|T_1 - U(W \odot V)\|_F$$
$$\|T_2 - X_2\|_F = \|T_2 - V(W \odot U)\|_F$$
$$\|T_3 - X_3\|_F = \|T_3 - W(V \odot U)\|_F$$

we will have succeeded.

- These problems are hard!
- All three matrices U, V, W can potentially vary as we optimize

Idea: hold two of these matrices fixed (e.g. V, W) then solve for the third (U) using

$$\underset{U}{\text{minimize}} \; \| T_1 - U (W \odot V)^T \|_F$$

This has the general form:

$$\underset{X}{\text{minimize}} \; \| A - XB \|_F$$

where A, B are known.

We will see: this is a least squares problem

Lightning Round: least squares

- if we solve $A\vec{x} = \vec{b}$ for A square invertible, easy:

$$\vec{x} = A^{-1}\vec{b} \quad \text{is unique sol'n.}$$

- Can still "solve" $A\vec{x} = \vec{b}$ when A non-square/noninvertible in a way which is optimal — in some sense.

- "Solution" will be $\hat{x} = A^{+}\vec{b}$ where A^{+} is "pseudo inverse" of A — we get it from SVD.

- How: if $A = U\Sigma V^T$ then

$$A^{+} = V\Sigma^{+}U^T$$

where Σ^{+} is Σ^T but with diag entries σ_k replaced by $\frac{1}{\sigma_k}$ — unless they're 0!

- As always, two general situations for $A \in \mathbb{R}^{m \times n}$

<u>$m \geq n$</u>

"Tall-skinny" A

"overdetermined system" – i.e. $A\vec{x} = \vec{b}$ has no solutions (usually)

$$\underset{m \times n}{A} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}\ V^T, \qquad \Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ \hline & & 0 & & \end{bmatrix}$$

<u>then:</u>

$$\underset{n \times m}{A^+} = \underset{n\times n}{V}\ \underset{m\times n}{\Sigma^+}\ \underset{n\times n}{U^T}, \qquad \Sigma^+ = \left[\begin{array}{ccccc|c} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & 0 \\ & & & 0 & & \\ & & & & \ddots & \end{array}\right]$$

if A full rank (i.e. no zeroes in diag of Σ) then $A^+ A = I$, i.e. A^+ is a <u>left</u> <u>inverse</u> of A.

<u>$m \le n$</u>

"Short-Fat" A

"underdetermined" — i.e. $A\vec{x} = \vec{b}$ has ∞-many solutions (usually)

$$\underset{m \times n}{A} = \underset{m \times m}{U} \; \underset{m \times n}{\Sigma} \; \underset{n \times n}{V^T}, \qquad \Sigma = \left[\begin{array}{ccccc|c} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & 0 \\ & & & 0 & & \\ & & & & \ddots \; 0 & \end{array}\right]$$

$$\underset{n \times m}{A^+} = \underset{n \times n}{V} \; \underset{n \times m}{\Sigma^+} \; \underset{m \times m}{U^T}, \qquad \Sigma^+ = \left[\begin{array}{ccccc} \frac{1}{\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \; 0 \\ \hline & & 0 & & \end{array}\right]$$

if A full rank, then $AA^+ = I$, i.e.

A^+ is the <u>right inverse</u> for A.

In either case, define

$$\hat{x} = A^+ \vec{b}$$

<u>Fact</u> ① in underdetermined case, assuming sol'n to $A\vec{x} = \vec{b}$ exists (typically ∞-many) then $\hat{x}$ is a sol'n and $\|\hat{x}\|$ is minimal among all sol'ns.

② in overdetermined case, assuming no sol'n exists (typical), $\hat{x}$ is "close as possible" to a sol'n, i.e. $\|A\hat{x} - b\|$ is as small as possible among all $\vec{x} \in \mathbb{R}^n$.

<u>Ex</u>: Consider $A = U\Sigma V^T$ below:

$$\underbrace{\frac{1}{\sqrt{6}}\begin{bmatrix} \sqrt{3} & 1 & -\sqrt{2} \\ 0 & 2 & \sqrt{2} \\ \sqrt{3} & -1 & \sqrt{2} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{V^T} = \underbrace{\frac{1}{\sqrt{12}}\begin{bmatrix} 3\sqrt{3}+1 & -3\sqrt{3}+1 \\ 2 & 2 \\ 3\sqrt{3}-1 & -3\sqrt{3}-1 \end{bmatrix}}_{A}$$

<u>then</u>: $$A^+ = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/1 & 0 \end{bmatrix} \frac{1}{\sqrt{6}}\begin{bmatrix} \sqrt{3} & 0 & \sqrt{3} \\ 1 & 2 & -1 \\ -\sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix}$$

$$= \frac{1}{\sqrt{12}} \begin{bmatrix} 1/3 & 1 & 0 \\ -1/3 & 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & \sqrt{3} \\ 1 & 2 & -1 \\ -\sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix}$$

$$= \frac{1}{\sqrt{12}} \begin{bmatrix} \sqrt{3}/3+1 & 2 & \sqrt{3}/3-1 \\ -\sqrt{3}/3+1 & 2 & -\sqrt{3}/3-1 \end{bmatrix}$$

least squares "sol'n" to eq.

$$A\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

given by $\hat{x} = A^+ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$$= \frac{1}{\sqrt{12}} \begin{bmatrix} \sqrt{3}/3+1 & 2 & \sqrt{3}/3-1 \\ -\sqrt{3}/3+1 & 2 & -\sqrt{3}/3-1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \frac{1}{\sqrt{12}} \begin{bmatrix} \frac{2\sqrt{3}}{3}+2 \\ -\frac{2\sqrt{3}}{3}+2 \end{bmatrix}$$

- Now consider a different problem, given $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times k}$, find $X \in \mathbb{R}^{n \times k}$ so that

$$\|AX - B\|_F \quad \text{is minimized}$$

$$(\text{equiv: } \|AX - B\|_F^2 \quad \text{is minimized})$$

- This is a least squares problem in disguise!

To see: write $X = [\vec{x}_1 | \cdots | \vec{x}_k]$

so $AX = [A\vec{x}_1 | \cdots | A\vec{x}_k]$

So if $B = [\vec{b}_1 | \cdots | \vec{b}_k]$

then $\|AX - B\|_F^2$

$$= \|[A\vec{x}_1 - b_1 | \cdots | A\vec{x}_k - \vec{b}_k]\|_F^2$$

$$= \|A\vec{x}_1 - \vec{b}_1\|_2^2 + \cdots + \|A\vec{x}_k - \vec{b}_k\|_2^2$$

think of defn of $\|\cdot\|_F$

Now observe: since (F-norm)^2 of $AX - B$ is just sum of (2-norm)^2 of its columns if we find $\vec{x}_i$'s minimizing

$$\|A\vec{x}_1 - \vec{b}_1\|^2$$
$$\vdots$$
$$\|A\vec{x}_k - \vec{b}_k\|^2$$

i.e. solve a bunch of least squares problems

then letting $X = [\vec{x}_1 | \cdots | \vec{x}_k]$ minimizes $\|AX - B\|_F$!

Why: if there were a better matrix $X' = [\vec{x}_1' | \cdots | \vec{x}_k']$ then $\|A\vec{x}_i' - \vec{b}_i\|_2$ would have to be smaller for at least one i – contradiction than $\|A\vec{x}_i - b_i\|$

Summary: minimizing $\|AX - B\|_F$ given A, B is just a bunch of LS problems (ie column by column)