Model 1 will be $I \times J k$ matrix = $4 \times (7 \cdot 2)$

\[
\begin{bmatrix}
  a_{111} & a_{112} & a_{113} & a_{112} & a_{112} & a_{132} \\
  a_{211} & a_{221} & a_{222} & a_{222} & a_{222} & a_{232} \\
  a_{311} & a_{321} & a_{322} & a_{322} & a_{322} & a_{332} \\
  a_{411} & a_{421} & a_{422} & a_{422} & a_{422} & a_{432}
\end{bmatrix}
\]

Model 1 Gibbs

Model 2 will be $I \times J k$ matrix = $3 \times (4 \cdot 2)$

\[
\begin{bmatrix}
  a_{111} & a_{211} & a_{311} & a_{411} & a_{112} & a_{212} & a_{312} & a_{412} \\
  a_{121} & a_{221} & a_{321} & a_{421} & a_{122} & a_{222} & a_{322} & a_{422} \\
  a_{131} & a_{231} & a_{331} & a_{431} & a_{132} & a_{232} & a_{332} & a_{432}
\end{bmatrix}
\]

Mode 2 Gibbs

Model 3 will be $I \times J = 2 \times (4 \cdot 3)$

\[
\begin{bmatrix}
  a_{111} & a_{211} & a_{311} & a_{411} & a_{112} & a_{212} & a_{312} & a_{412} & a_{113} & a_{213} \\
  a_{122} & a_{222} & a_{322} & a_{422} & a_{123} & a_{223} & a_{323} & a_{423} & a_{132} & a_{232} \\
  a_{132} & a_{232} & a_{332} & a_{432} & a_{133} & a_{233} & a_{333} & a_{433}
\end{bmatrix}
\]

Mode 3 Gibbs
Outer products

- given vectors $\tilde{u} \in \mathbb{R}^m$, $\tilde{v} \in \mathbb{R}^n$, $\tilde{w} \in \mathbb{R}^p$
the outer product $\tilde{u} \circ \tilde{v} \circ \tilde{w}$ is the order 3 tensor whose $ijk$-th entry is $\tilde{u}_i \tilde{v}_j \tilde{w}_k$

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

$$= \begin{array}{c||c|c|c}
\text{unfold} \\
\hline
0 & 0 & 0 \\
\hline
1 & 0 & -1 \\
\hline
0 & 1 & 0 \\
\hline
0 & 0 & 0 \\
\end{array}$$

- analogue of $\tilde{u} \tilde{v}^T$ in 2D
  
  sometimes written $\tilde{u} \circ \tilde{v}$

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

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

- in such a $T$ all columns are multiples of one another and more, all rows are...
- We can express unfoldings of rank 1 tensors $u \circ v \circ w$ in terms of $\otimes$.

- Let's extract general formulae from an example.

Consider $u \circ v \circ w$

$$T = [u_1] \circ [v_1] \circ [w_1] = [u_1 v_1 w_1, u_1 v_2 w_1, u_1 v_3 w_1, u_2 v_1 w_1, u_2 v_2 w_1, u_2 v_3 w_1]$$

$$T_1 = [u_1 v_1 w_1, u_1 v_2 w_1, u_1 v_3 w_1, u_2 v_1 w_1, u_2 v_2 w_1, u_2 v_3 w_1]$$

Observe: $w \otimes v = [w_1, v] = [w_1 v_1, w_2 v_1, w_3 v_1, w_1 v_2, w_2 v_2, w_3 v_2, w_1 v_3, w_2 v_3]$.

So $(w \otimes v)^T = [w_1 v_1, w_1 v_2, w_1 v_3, w_2 v_1, w_2 v_2, w_2 v_3, w_3 v_1, w_3 v_2, w_3 v_3]$.

So $u \otimes (w \otimes v)^T = [u_1 (w \otimes v)^T] = [u_1, u_2 (w \otimes v)^T] = T_1$. Thus, $u \otimes (w \otimes v)^T = T_1$. 


likewise can check: $v (\tilde{u} \otimes \tilde{u})^T$
\[ T_2 = \tilde{v} \otimes (\tilde{u} \otimes \tilde{u})^T \] notice reversed order of alphabetical order.
\[ T_3 = \tilde{u} \otimes (\tilde{v} \otimes \tilde{u})^T = \tilde{v} (\tilde{v} \otimes \tilde{u})^T \]

---

CP Approximation to $T$

$\Psi$ tensor analogue of finding $\text{rank } k$ approximation to a matrix $A$ using SVD.

$A = \sigma_1 \tilde{u} \tilde{v} \tilde{w} + \cdots + \sigma_r \tilde{u} \tilde{v} \tilde{w}$

- Given a of rank $r$, consider problem

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

$\| A - x \|_F$ is minimized

- SVD says: sol'n very nice!

given by $x = A_k = \sigma_1 \tilde{u}_1 \tilde{v}_1 \tilde{w}_1 + \cdots + \sigma_k \tilde{u}_k \tilde{v}_k \tilde{w}_k$

$= \sigma_1 \tilde{u}_1 \tilde{v}_1 \tilde{w}_1 + \cdots + \sigma_k \tilde{u}_k \tilde{v}_k \tilde{w}_k$

\[ \text{wing notation} \]
How to generalize to tensor $T$?

to first need to define $\text{rank}(T)$ in general.

- Recall: SVD gives:

$$\text{rank}(A) = \text{least \# of rank 1 matrices needed to sum to } A.$$ 

- We'll take this as our \textbf{definition} of rank for tensors!

i.e. define, given $T$, order $r$

$$\text{rank}(T) = \text{shortest length } r \text{ of a sum of form}$$

$$\sum_{(\ell)} \tilde{u}_\ell \tilde{v}_\ell \tilde{w}_\ell$$

\text{equalling } T.

= \text{least number of rank 1 tensors summing to } T.$$

So now we can state an analogous "rank $\leq$ approximation problem" for $T$. 

Matrix: \[
\text{minimize } \|A-X\|_F \\
\text{over all } X's \text{ of rank } K
\]

Tensor: \[
\text{minimize } \|T-X\|_F \\
\text{over all } X's \text{ of the form} \\
X = \sum_{i=1}^{K} \hat{u}_i \hat{v}_i \hat{w}_i \\
\text{call thy } (*)
\]

where \( K \) is fixed \( \leq \text{rank}(T) \). \((\text{imagine } K < \text{rank}(T))\)

Note: \( \|T\|_F \) mean exactly what you'd think for a tensor; \( \) if \( a_{ijk} \) is \( w-th \) entry then
\[
\|T\|_F = \sqrt{\sum_{ijk} a_{ijk}^2}
\]

---

Differences between matrix / tensor:
- no canonical "nice" sol'n to problem like in matrix case
- don't insist \( \hat{u}'s \) are either orthogonal or normal! likewise for \( \hat{v}'s, \hat{w}'s \)
- if we change \( K \), \( \hat{u}'s, \hat{v}'s, \hat{w}'s \) will change in general - so won't just add terms at the end.
Main issue: nothing so nice as the SVD exists for tensors in general.

- So how to answer our approximation problem?
- First step: turn into a matrix problem, by unfolding our expression (*) for $x$.

Let $U = [\hat{u}_1 \ldots \hat{u}_k]$

$V = [\hat{v}_1 \ldots \hat{v}_k]$

$W = [\hat{w}_1 \ldots \hat{w}_k]$

Then: $X_1 = \text{mode 1 unfolding of } x$

$$X_1 = \sum_{i=1}^{k} u_i \otimes (\hat{w}_i \otimes \hat{v}_i)^T = \sum_{i=1}^{k} u_i (\hat{w}_i \otimes \hat{v}_i)^T$$

by our unfolding formula for $(u \circ v \circ w)$.

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

$$= \quad \text{circled symbol}$$

$$= [\hat{u}_1 \ldots \hat{u}_k] [\hat{w}_1 \otimes \hat{v}_1 \ldots \hat{w}_k \otimes \hat{v}_k]^T \quad \text{(why?)}$$

$$= U (W \circ V)^T$$
Likewise, we can check

\[ x_2 = V (W \circ U) \]
\[ x_3 = U (V \circ U) \]

Now, we're trying to minimize

\[ \| T - x \|_F \]

For our given matrix \( T \).

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

\[ \| T_1 - x_1 \|_F = \| T_1 - U(W \circ V) \|_F \]
\[ \| T_2 - x_2 \|_F = \| T_2 - V(W \circ U) \|_F \]
\[ \| T_3 - x_3 \|_F = \| T_3 - W(V \circ U) \|_F \]

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

$$\text{minimize} \quad \|T - U(W \odot V)\|_F$$

This has the general form:

$$\text{minimize} \quad \|A - XB\|_F$$

where $A, B$ are known.

We will see: this is a least squares problem.
Lightning Round: Least squares

- If we solve \( A\hat{x} = b \) for a square matrix \( A \), easy:
  \[ \hat{x} = A^{-1}b \] is a unique solution.

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

- "Solution" will be \( \hat{x} = A^{+}b \) 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 \( \Sigma^{+} \) is \( \Sigma^{T} \) but with diagonal entries \( \sigma_{k} \) replaced by \( \frac{1}{\sigma_{k}} \) unless they're 0!

- As always, two general situations for \( \text{AeR}^{m} \)
\[
\text{if } A \text{ full rank (i.e. no zeroes in diag of } \Sigma) \text{ then } A^+ A = I, \text{ i.e. } A^+ U = C_{\text{left inverse of } A}.
\]
\[ m \leq n \]

"Short-Fat" \( A \)

"underdetermined" — i.e. \( A \tilde{x} = \mathbf{b} \) has many solutions (usually)

\[
\begin{bmatrix}
  A \\
  \vdots \\
  \end{bmatrix}
= \begin{bmatrix}
  u \\
  \vdots \\
  \end{bmatrix} \begin{bmatrix}
  \Sigma \\
  \vdots \\
  \end{bmatrix} \begin{bmatrix}
  v^T \\
  \vdots \\
  \end{bmatrix}
\]

\[
\begin{bmatrix}
  A^+ \\
  \vdots \\
  \end{bmatrix}
= \begin{bmatrix}
  V \\
  \vdots \\
  \end{bmatrix} \begin{bmatrix}
  \Sigma^+ \\
  \vdots \\
  \end{bmatrix} \begin{bmatrix}
  u^T \\
  \vdots \\
  \end{bmatrix}
\]

If \( A \) full rank, then \( AA^+ = \mathbf{I} \), i.e. \( A^+ \) the right inverse for \( A \).
In either case, define

\[ \tilde{E} = A^+ \tilde{b} \]

\( \tilde{E} \) in underdetermined case, assuming

\( \tilde{b} \) in

then \( \tilde{E} \) is a solution and \( \| \tilde{E} \|_1 \) is minimized among all solutions.

\( \tilde{E} \) in overdetermined case, assuming no solution exists (typical), \( \tilde{E} \) is "close as possible" to a solution, i.e., \( \| A \tilde{E} - \tilde{b} \|_1 \) is as small as possible among all \( \tilde{E} \in \mathbb{R}^n \).

**Example:** Consider \( A = USV^T \) below:

\[
\frac{1}{\sqrt{6}} \begin{bmatrix}
\sqrt{3} & 1 & -\sqrt{2} \\
0 & 2 & \sqrt{2} \\
\sqrt{3} & -1 & \sqrt{2}
\end{bmatrix}
\begin{bmatrix}
3 & 0 \\
0 & 1 \\
0 & 0
\end{bmatrix}
\frac{1}{\sqrt{2}} \begin{bmatrix}
1 & -1 \\
1 & 1
\end{bmatrix}
\frac{1}{\sqrt{12}} \begin{bmatrix}
3\sqrt{7}+1 & -3\sqrt{7}+1 \\
2 & 2 \\
3\sqrt{7}+1 & -3\sqrt{7}+1
\end{bmatrix}
\]

Then:

\[
A^+ = \frac{1}{\sqrt{12}} \begin{bmatrix}
-1 & 1 \\
1 & 1 \\
1 & -1
\end{bmatrix}
\begin{bmatrix}
\sqrt{3} & 0 & 0 \\
0 & 3 & 0 \\
0 & 0 & \sqrt{3}
\end{bmatrix}
\frac{1}{\sqrt{12}} \begin{bmatrix}
\sqrt{3} & 0 & \sqrt{3} \\
1 & 2 & -1 \\
-\sqrt{3} & \sqrt{3} & \sqrt{2}
\end{bmatrix}
\]
\[
\begin{align*}
\hat{\mathbf{x}} &= \frac{1}{\sqrt{12}} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 1 \\ 2 - 1 \end{bmatrix} \\
&= \frac{1}{\sqrt{12}} \begin{bmatrix} \frac{1}{\sqrt{3}} + 1 & 2 & \frac{1}{\sqrt{3}} - 1 \\ -\frac{1}{\sqrt{3}} + 1 & 2 & -\frac{1}{\sqrt{3}} - 1 \end{bmatrix} \\
&= \frac{1}{\sqrt{12}} \begin{bmatrix} \frac{2\sqrt{3}}{3} + 2 \\ -\frac{2\sqrt{3}}{3} + 2 \end{bmatrix}
\end{align*}
\]

Least square solution to \( \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} \).
Now consider a different problem, given \( A \in \mathbb{R}^{m \times n} \), \( B \in \mathbb{R}^{m \times k} \), and \( x \in \mathbb{R}^{n \times k} \), so that 

\[
\| A x - B \|_F \text{ is minimized}
\]

(equiv: \( \| A x - B \|_2 \) is minimized)

This is a least squares problem in disguise!

To see: write \( x = [x_1, \ldots, x_k] \)

so \( Ax = [Ax_1, \ldots, Ax_k] \)

so if \( B = [b_1, \ldots, b_k] \)

then 

\[
\| A x - B \|_F^2 = \| Ax_1 - b_1 \|_2^2 + \ldots + \| Ax_k - b_k \|_2^2
\]

which is the definition of \( \| \cdot \|_F \).
Now observe: Since \((F-\text{norm})^2\) of \(Ax-b\) is just sum of \((z-\text{norm})^2\) of its columns if we find \(x_i\)'s minimizing
\[
\|Ax_i - b\|^2
\]
\[
\|Ax_k - b_k\|^2
\]
\[\text{1.e. solve a bunch of least squares problems}\]
then writing \(x = [x_1 \ldots x_k]\) minimizes \(\|Ax-b\|_F^2\!\)

why: if there were a better matrix \(x' = [x_1 \ldots x_k]\) then \(\|Ax' - b\|^2\) would have to be smaller for at least one \(i\) contradicts
\[\text{then } \|Ax_i - b\|^2\]

Summary: minimizing \(\|Ax-b\|_F^2\) given \(A, B\) is just a bunch of LS problems (ge column by column)