Tensors + Multilinear Algebra

Def: A real K-array or Kth order tensor of dimension $(M_1 \ldots M_K)$ is a set of $M_1 \times \ldots \times M_K$ points in $\mathbb{R}$ indexed as follows

$$Y = \{ y_{i_1 \ldots i_K} : i_1 \in 1..M_1, \ldots, i_K \in 1..M_K \} \subset \mathbb{R}^M$$

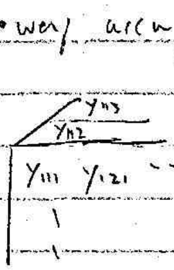
Def (alt). Let $I_1 = \{1..M_1\} \ldots I_K = \{1..M_K\}$. Y is a map from $I_1 \times \ldots \times I_K$ to $\mathbb{R}$

$I_1 \ldots I_K$ are sometimes called the modes of the array

Ex: vectors: $y = (y_1 \ldots y_{M_1}) \in \mathbb{R}^{M_1}$ is a 1-array

Ex: matrices $Y = \{ y_{ij} : i = 1..M_1, j = 1..M_2 \} \in \mathbb{R}^{M_1 \times M_2}$ is a 2-array

Ex: 3-way array $Y = \{ y_{ijk}, i = 1..M_1, j = 1..M_2, k = 1..M_3 \} \in \mathbb{R}^{M_1 \times M_2 \times M_3}$



Scalar Products + Norms

vector scalar product: $x, y \in \mathbb{R}^M$, $\langle x, y \rangle = \sum_{i=1}^{M} x_i y_i$

Matrix scalar product: $X, Y \in \mathbb{R}^{M_1 \times M_2}$ $\langle X, Y \rangle = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} x_{ij} y_{ij}$

$$= tr(X^T Y) = tr(Y X^T) = tr(Y^T X) = tr(X Y^T)$$

tensor scalar product: $X, Y \in \mathbb{R}^{M_1 \times M_2 \times M_3}$

$$\langle X, Y \rangle = \sum_i \sum_j \sum_k x_{ijk} y_{ijk}$$

$$= \langle vec(X), vec(Y) \rangle$$

Orthogonality: X, Y are orthogonal if $\langle X, Y \rangle = 0$

(F-)Norm: The (Frobenius) norm of a tensor X is

$$\|X\| = \langle X, X \rangle^{1/2}$$

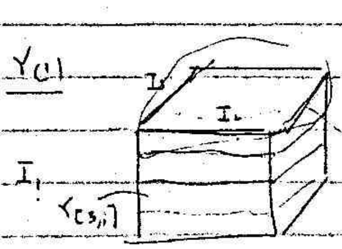
Matricizations, vectorizations

Let $Y \in \mathbb{R}^{M_1 \times M_2 \times M_3}$. Define matrices $Y_{(1)} \in \mathbb{R}^{M_1 \times (M_2 M_3)}$ "mode-1 matricization"

$Y_{(2)} \in \mathbb{R}^{M_2 \times (M_1 M_3)}$ " " 2 "

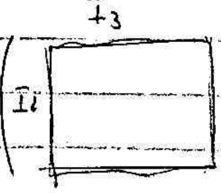
$Y_{(3)} \in \mathbb{R}^{M_3 \times (M_1 M_2)}$ " mode 3 matricization

$Y_{(1)}$



$$Y_{(1)} = \begin{bmatrix} \mathrm{vec}(Y_{[1,:]}) \longrightarrow \\ \mathrm{vec}(Y_{[2,:]}) \longrightarrow \\ \vdots \\ \mathrm{vec}(Y_{[M_1,:]}) \longrightarrow \end{bmatrix} \in \mathbb{R}^{M_1 \times (M_2 M_3)}$$

$$\mathrm{vec}(Y_{[k,:]}) = \mathrm{vec}\begin{bmatrix} y_{k11} & y_{k12} & \cdots & y_{k1M_3} \\ \vdots & & & \\ y_{kM_21} & & & y_{kM_2M_3} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} y_{k11} \\ y_{k21} \\ \vdots \\ y_{kM_21} \end{bmatrix} \\ \begin{bmatrix} y_{k12} \\ \vdots \\ y_{kM_22} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} y_{k1M_3} \\ \vdots \\ y_{kM_2M_3} \end{bmatrix} \end{bmatrix}$$

$= \mathrm{vec}\Big($  $\Big)$

$$Y_{(1)} = \begin{bmatrix} (y_{111} \cdots y_{1M_21}) & \cdots & (y_{11M_3} \cdots y_{1M_2M_3}) \\ (y_{211} \cdots y_{2M_21}) & \cdots & (y_{21M_3} \cdots y_{2M_2M_3}) \\ & \vdots & \end{bmatrix}$$

Example: (Kolda & Bader, 2009)

$$Y \in \mathbb{R}^{3\times 4\times 2} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}, \quad \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}$$

$$\rightarrow Y[1,,] = \begin{array}{cc} & I_3 \\ I_2 & \begin{bmatrix} 1 & 13 \\ 4 & 16 \\ 7 & 19 \\ 10 & 22 \end{bmatrix} \end{array}$$

$$\rightarrow Y_{(1)} = \begin{bmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & \cdots & \\ & & & & & & & \end{bmatrix}$$

Similarly:

$$Y_{(2)} = \begin{array}{cc} & m_1 \times m_3 \\ m_2 & \begin{bmatrix} \mathrm{vec}(Y[,1,]) & - \\ \mathrm{vec}(Y[,2,]) & - \\ \vdots & \\ \mathrm{vec}(Y[,m_2,]) & - \end{bmatrix} \end{array} \in \mathbb{R}^{m_2 \times (m_1 m_3)}$$

$$Y[,1,] = \begin{array}{cc} & m_3 \\ m_1 & \begin{bmatrix} 1 & 13 \\ 2 & 14 \\ 3 & 15 \end{bmatrix} \end{array} \quad Y[,2,] = \begin{bmatrix} 4 & 16 \\ 5 & 17 \\ 6 & 18 \end{bmatrix} \quad \cdots$$

$$Y_{(2)} = \begin{bmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 18 \\ \vdots & & & & & \end{bmatrix}$$

$$Y_{(3)} = \begin{array}{cc} & m_1 \times m_2 \\ m_3 & \begin{bmatrix} \mathrm{vec}(Y[,,1]) & - \\ \vdots & \\ \mathrm{vec}(Y[,,m_3]) & - \end{bmatrix} \end{array} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & \cdots & & & & & & & & \\ & & & & & & & & & & & \end{bmatrix}$$

* Matricizations vary across authors: see Kolda & Bader (2009) vs de Lathauwer (2000)

Vectorization: Define $\text{vec}(Y) = \text{vec}(Y_{(1)})$

"leftmost indices move fastest"

$$= \begin{pmatrix} y_{111} \\ y_{211} \\ \vdots \\ y_{m_1 11} \end{pmatrix} \begin{pmatrix} y_{1,2,1} \\ y_{2,2,1} \\ \vdots \\ y_{m_1 2 1} \end{pmatrix}$$

$(= c(Y)$ in R $)$

Multilinear operators + Array-Matrix Multiplication

Linear change-of-variables

Let $\underset{\sim}{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix}$, eg z_1 : math score
z_2 : science "
z_3 : read "
z_4 : writing "

For $B \in \mathbb{R}^{q \times p}$, let: $\underset{\sim}{x} = \underline{B}\, \underline{z} = \begin{bmatrix} \beta_1^T \text{—} \\ \vdots \\ \beta_q^T \text{—} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix} = \begin{bmatrix} \beta_1^T z \\ \vdots \\ \beta_q^T z \end{bmatrix}$

eg: $B = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 0 & 0 \end{bmatrix}$

$Bz = \begin{bmatrix} \text{overall avg} \\ \text{math/sci - read/write contrast} \\ \text{math - sci contrast} \end{bmatrix}$

Matrices: $\underset{n\times p}{Z}\underset{p\times q}{B^T}$ $\begin{bmatrix} z_1^T & — \\ \vdots & \\ z_n^T & — \end{bmatrix}\begin{bmatrix} | & & | \\ B_1 & \cdots & B_q \\ | & & | \end{bmatrix} = \begin{bmatrix} (Bz_1)^T & — \\ \vdots & \\ (Bz_n)^T & — \end{bmatrix}$

Example: $\underset{\sim}{z}_i$ = scores at time point i, Z = scores across all time points

$\underset{\sim}{x}_i = B\underset{\sim}{z}_i$ = composite scores at i, $X = ZB^T$ = comp. scores across time.

Linearity:

$g_B(Z) = ZB^T$, $g_B : Z \to X$ is a linear map

$$g_B(aZ_1 + bZ_2) = (aZ_1 + bZ_2)B^T$$
$$= a\,g_B(Z_1) + b\,g_B(Z_2)$$

also, $g_Z(B) = ZB^T$ is linear: $g_Z(aB_1 + bB_2) = a g_Z(B_1) + b g_Z(B_2)$

Bilinear maps: for $A \in \mathbb{R}^{m\times n}$, let $g_{AB}(Z) = AZB^T$

$$g_{AB}(Z) : \mathbb{R}^{n\times p} \to \mathbb{R}^{m\times q}$$

Example: z_{ij} = score on jth exam at time i, $i = 1 \ldots 4$

$\underset{\sim}{z}_i$ = scores at time i

$\underset{\sim}{x}_i = B\underset{\sim}{z}_i$ = composite scores at time i

Now let $\begin{pmatrix} y_{1j} \\ y_{2j} \end{pmatrix} = A\begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \end{pmatrix} = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & -1/2 & -1/2 \end{pmatrix}\begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \end{pmatrix} = \begin{pmatrix} \text{avg of comp. score } j \\ \text{contrast for composite score } j \end{pmatrix}$

$Y = AX$

$\quad = AZB^T$, a bilinear change of variables

The map $g_{AB}: Z \to AZB^T$

1) linear in Z
2) not linear in (A,B)
3) bilinear in (A,B)

1) $A(cZ_1 + dZ_2)B^T = cAZ_1B^T + dAZ_2B^T$

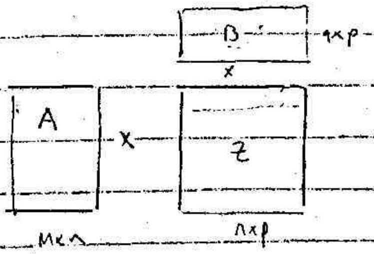
2) $(cA_1 + dA_2)Z(cB_1 + dB_2)^T \neq cA_1ZB_1^T + dA_2ZB_2^T$

3) $AZ(cB_1 + dB_2)^T = cAZB_1^T + dAZB_2^T$

$(cA_1 + dA_2)ZB^T = cA_1ZB^T + dA_2ZB^T$

Array-Matrix Multiplication

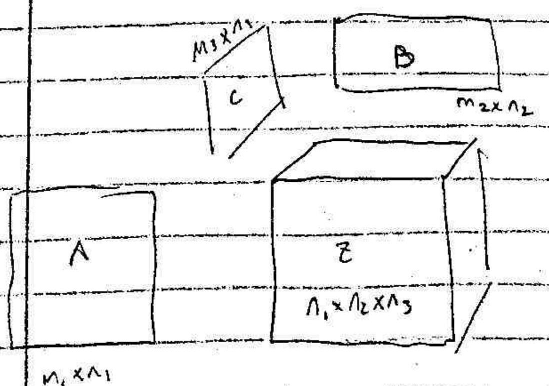
$AZB^T =$



A transforms the rows of Z

rows of AZ are LCs of rows of Z

B transforms columns of Z



How to define this?

Define: $Z \times_1 A = \text{array}\left(\text{vec}\left(A\, Z_{(1)} \right), \ \text{dim}=(m_1, n_2, n_3) \right)$

"first mode multiplication

$$Z \times_2 B = \text{aperm}\left(\text{array}\left(\text{vec}(B\, Z_{(2)}), \ \text{dim}=(m_2, n_1, n_3) \right), \ c(2,1,3) \right)$$

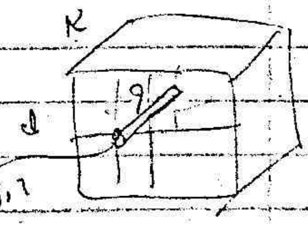
$$Z \times_3 = \text{aperm}\left(\text{array}\left(\text{vec}(C Z_{(3)}), \ \text{dim}=(m_3, n_1, n_2) \right), \ c(3,1,2) \right)$$

generally, $Z \times_k A$ is "$\underline{k^{th}}$-mode multiplication"

Formally: $\left(Z \times_k A \right) \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_{k-1} \times n_k \times m_{k+1} \cdots m_K}$

$$\left(Z \times_k A \right)_{i_1 \cdots i_{k-1} j\, i_{k+1} \cdots i_K} = \sum_{i_k=1}^{n_k} z_{i_1 \cdots i_K}\, a_{j i_k}$$

idea: for $K=3$,



let $\tilde{z}_{ij} = Z_{[i,j,]} \in \mathbb{R}^{m_n}$

Change of variables along this mode $\tilde{x}_{ij} = \underset{m_3 \times n_3}{A} \begin{pmatrix} z_{ij1} \\ \vdots \\ z_{ijn_3} \end{pmatrix}$

Properties of array-matrix multiplication

① $Y = Z \times_k A \iff Y_{(k)} = A Z_{(k)}$ (relation to matricization)

② a) $(Y \times_n A) \times_j B = (Y \times_j B) \times_n A$

b) $Y \times_n A \times_n B = Y \times_n (BA)$

(Pf: $(Y \times_n A)_{(n)} = A Y_{(n)}$

$((Y \times_n A) \times_n B)_{(n)} = B (Y \times_n A)_{(n)}$
$= B A Y_{(n)}$
$\iff Y \times_n A \times_n B = Y \times_n (BA)$)

Multilinear transforms

let $Z \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $A_k \in \mathbb{R}^{m_k \times n_k}$, $k = 1, 2, 3$

define $Z \times \{A_1, A_2, A_3\} := Z \times_1 A_1 \times_2 A_2 \times_3 A_3$

Most useful Property

$Y = Z \times \{A_1, A_2, A_3\} \iff Y_{(1)} = A_1 Z_{(1)} (A_3^T \otimes A_2^T)$

$Y_{(2)} = A_2 Z_{(2)} (A_3^T \otimes A_1^T)$

$Y_{(3)} = A_3 Z_{(3)} (A_2^T \otimes A_1^T)$

Compare to matrices

$Y = A_1 Z A_2^T \iff Y = Z \times \{A_1, A_2\}$

$\iff Y_{(1)} = A_1 Z_{(1)} A_2^T$

$Y_{(2)} = A_2 Z_{(2)} A_1^T$ ($Y_{(1)} = Y = A_1 Z A_2^T$, $Y_{(2)} = Y^T = A_2 Z^T A_1$)

Vectorization:

$$\mathrm{vec}(\mathcal{Z} \times \{A_1 \dots A_n\}) = (A_n \otimes A_{n-1} \otimes \cdots \otimes A_1)\, \underset{\sim}{z} \qquad (\underset{\sim}{z} = \mathrm{vec}(\mathcal{Z}))$$

Proof:

$$\mathrm{vec}(\ \text{"}\ ) = \mathrm{vec}\Big((\qquad)_{(1)} \Big)$$

$$= \mathrm{vec}\Big(A_1 Z_{(1)} (A_n^T \otimes \cdots \otimes A_2^T) \Big)$$

$$= (A_n \otimes \cdots \otimes A_1)\, \mathrm{vec}(Z_{(1)})$$

$$= (A_n \otimes \cdots \otimes A_1)\, \underset{\sim}{z}$$