Tensor Rank and Matrix Factorizations

Charles F. Van Loan
Department of Computer Science

Carla D. Moravitz Martin
Center for Applied Mathematics

Cornell University

Supported in part by the NSF contract CCR-9901988.

An Example and Some Questions

If

```
A = randn(2,2,2)
```

then

$$\begin{aligned}
\text{Prob}\,(\text{rank}(A) < 2) \;&=\; 0 \\
\text{Prob}\,(\text{rank}(A) = 2) \;&\approx\; .79 \\
\text{Prob}\,(\text{rank}(A) = 3) \;&\approx\; .21
\end{aligned}$$

If $\text{rank}(A) = 3$, how close is A to a rank 2 tensor?

Does this question make sense?

What Does "Rank(A) = 2" Mean?

We can find 2-by-2 matrices

$$X = (x_{ij}) = \left[\, x_1 \;\; x_2 \,\right] \qquad Y = (y_{ij}) = \left[\, y_1 \;\; y_2 \,\right] \qquad Z = (z_{ij}) = \left[\, z_1 \;\; z_2 \,\right]$$

so that

$$A = x_1 \circ y_1 \circ z_1 \;+\; x_2 \circ y_2 \circ z_2$$

$$\Longleftrightarrow$$

$$\mathrm{vec}(A) = z_1 \otimes y_1 \otimes x_1 \;+\; z_2 \otimes y_2 \otimes x_2$$

$$\Longleftrightarrow$$

$$a_{ijk} = x_{i1}y_{j1}z_{k1} + x_{i2}y_{j2}z_{k2} \qquad i = 1{:}2,\; j = 1{:}2,\; k = 1{:}2$$

$$\Longleftrightarrow$$

$$\begin{bmatrix} a_{111} & a_{121} \\ a_{211} & a_{221} \end{bmatrix} = z_{11} \cdot \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{21} \end{pmatrix}^T + z_{12} \cdot \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \begin{pmatrix} y_{12} \\ y_{22} \end{pmatrix}^T$$

$$\begin{bmatrix} a_{112} & a_{122} \\ a_{212} & a_{222} \end{bmatrix} = z_{21} \cdot \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{21} \end{pmatrix}^T + z_{22} \cdot \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \begin{pmatrix} y_{12} \\ y_{22} \end{pmatrix}^T$$

$$\Longleftrightarrow$$

$$A(:,:,1) = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} z_{11} & 0 \\ 0 & z_{12} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}^T$$

$$A(:,:,2) = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} z_{11} & 0 \\ 0 & z_{12} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}^T$$

Simultaneous Non-Orthogonal "SVDs" of the Slices

Thus, if we have the rank-2 expansion

$$A = X(:,1) \circ Y(:,1) \circ Z(:,1) \ + \ X(:,2) \circ Y(:,2) \circ Z(:,2)$$

then

$$A(:,:,1) \ = \ X \begin{bmatrix} z_{11} & 0 \\ 0 & z_{12} \end{bmatrix} Y^T \qquad A(:,:,2) \ = \ X \begin{bmatrix} z_{21} & 0 \\ 0 & z_{22} \end{bmatrix} Y^T$$

$$A(:,1,:) \ = \ X \begin{bmatrix} y_{11} & 0 \\ 0 & y_{12} \end{bmatrix} Z^T \qquad A(:,2,:) \ = \ X \begin{bmatrix} y_{21} & 0 \\ 0 & y_{22} \end{bmatrix} Z^T$$

$$A(1,:,:) \ = \ Y \begin{bmatrix} x_{11} & 0 \\ 0 & x_{12} \end{bmatrix} Z^T \qquad A(2,:,:) \ = \ Y \begin{bmatrix} x_{21} & 0 \\ 0 & x_{22} \end{bmatrix} Z^T$$

What Does "Rank(A) = 3" Mean?

We can find 2-by-3 matrices

$$X = (x_{ij}) = \left[\, x_1 \;\; x_2 \;\; x_3 \,\right] \qquad Y = (y_{ij}) = \left[\, y_1 \;\; y_2 \;\; y_3 \,\right] \qquad Z = (z_{ij}) = \left[\, z_1 \;\; z_2 \;\; z_3 \,\right]$$

so that

$$A = x_1 \circ y_1 \circ z_1 \; + \; x_2 \circ y_2 \circ z_2 \; + \; x_3 \circ y_3 \circ z_3$$

$$\Longleftrightarrow$$

$$\text{vec}(A) = z_1 \otimes y_1 \otimes x_1 \; + \; z_2 \otimes y_2 \otimes x_2 \; + \; z_3 \otimes y_3 \otimes x_3$$

$$\Longleftrightarrow$$

$$a_{ijk} = x_{i1}y_{j1}z_{k1} + x_{i2}y_{j2}z_{k2} + x_{i3}y_{j3}z_{k3} \qquad i = 1{:}2, \; j = 1{:}2, \; k = 1{:}2$$

$$\iff$$

$$\begin{bmatrix} a_{111} & a_{121} \\ a_{211} & a_{221} \end{bmatrix} = z_{11} \cdot \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{21} \end{pmatrix}^T + z_{12} \cdot \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \begin{pmatrix} y_{12} \\ y_{22} \end{pmatrix}^T + z_{13} \cdot \begin{pmatrix} x_{13} \\ x_{23} \end{pmatrix} \begin{pmatrix} y_{13} \\ y_{23} \end{pmatrix}^T$$

$$\begin{bmatrix} a_{112} & a_{122} \\ a_{212} & a_{222} \end{bmatrix} = z_{21} \cdot \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{21} \end{pmatrix}^T + z_{22} \cdot \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \begin{pmatrix} y_{12} \\ y_{22} \end{pmatrix}^T + z_{23} \cdot \begin{pmatrix} x_{13} \\ x_{23} \end{pmatrix} \begin{pmatrix} y_{13} \\ y_{23} \end{pmatrix}^T$$

$$\iff$$

$$A(:,:,1) = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \begin{bmatrix} z_{11} & 0 & 0 \\ 0 & z_{12} & 0 \\ 0 & 0 & z_{13} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix}^T$$

$$A(:,:,2) = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \begin{bmatrix} z_{21} & 0 & 0 \\ 0 & z_{22} & 0 \\ 0 & 0 & z_{23} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix}^T$$

Simultaneous Non-Orthogonal "Partial" SVDs of the Slices

$$A = X(:,1)\circ Y(:,1)\circ Z(:,1) + X(:,2)\circ Y(:,2)\circ Z(:,2) + X(:,3)\circ Y(:,3)\circ Z(:,3)$$

$$\Longleftrightarrow$$

$$A(:,:,1) \;=\; X \begin{bmatrix} z_{11} & 0 & 0 \\ 0 & z_{12} & 0 \\ 0 & 0 & z_{13} \end{bmatrix} Y^T \qquad A(:,:,2) \;=\; X \begin{bmatrix} z_{21} & 0 & 0 \\ 0 & z_{22} & 0 \\ 0 & 0 & z_{23} \end{bmatrix} Y^T$$

$$A(:,1,:) \;=\; X \begin{bmatrix} y_{11} & 0 & 0 \\ 0 & y_{12} & 0 \\ 0 & 0 & y_{13} \end{bmatrix} Z^T \qquad A(:,2,:) \;=\; X \begin{bmatrix} y_{21} & 0 & 0 \\ 0 & y_{22} & 0 \\ 0 & 0 & y_{23} \end{bmatrix} Z^T$$

$$A(1,:,:) \;=\; Y \begin{bmatrix} x_{11} & 0 & 0 \\ 0 & x_{12} & 0 \\ 0 & 0 & x_{13} \end{bmatrix} Z^T \qquad A(2,:,:) \;=\; Y \begin{bmatrix} x_{21} & 0 & 0 \\ 0 & x_{22} & 0 \\ 0 & 0 & x_{23} \end{bmatrix} Z^T$$

Tensor Rank (General)

If $A = A(1{:}n_1, 1{:}n_2, 1{:}n_3)$ has rank r then there exist

$$X \in \mathbb{R}^{n_1 \times r} \qquad Y(:,j) \in \mathbb{R}^{n_2 \times r} \qquad Z(:,j) \in \mathbb{R}^{n_3 \times r}$$

so that

$$A = \sum_{j=1}^{r} X(:,j) \circ Y(:,j) \circ Z(:,j)$$

and r is minimal. I.e.,

$$\text{vec}(A) \;=\; \begin{bmatrix} \text{vec}(A(:\,,:\,,1)) \\ \text{vec}(A(:\,,:\,,2)) \\ \vdots \\ \text{vec}(A(:\,,:\,,n_3)) \end{bmatrix} \;=\; \sum_{j=1}^{r} Z(:,j) \otimes Y(:,j) \otimes X(:,j)$$

Simultaneous Non-Orthogonal "Partial SVDs"

$$A(:,:,j) \;=\; X \begin{bmatrix} z_{j1} & 0 & 0 & \cdots & 0 \\ 0 & z_{j2} & 0 & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & z_{jr} \end{bmatrix} Y^T \qquad j = 1{:}n_3$$

$$A(:,j,:) \;=\; X \begin{bmatrix} y_{j1} & 0 & 0 & \cdots & 0 \\ 0 & y_{j2} & 0 & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & y_{jr} \end{bmatrix} Z^T \qquad j = 1{:}n_2$$

$$A(j,:,:) \;=\; Y \begin{bmatrix} x_{j1} & 0 & 0 & \cdots & 0 \\ 0 & x_{j2} & 0 & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & x_{jr} \end{bmatrix} Z^T \qquad j = 1{:}n_1$$

Wishful: Simultaneous SVDs

It would be nice if X and Y were orthogonal:

$$A(:,:,j) \;=\; U \begin{bmatrix} z_{j1} & 0 & 0 & \cdots & 0 \\ 0 & z_{j2} & 0 & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & z_{jr} \end{bmatrix} V^T \qquad j = 1{:}n_3$$

How close can we get?

An Embedding Idea

Suppose $A \in \mathbb{R}^{n\times n\times n}$ and $\text{rank}(A) = r > n$, then

$$A(:,:,j) \;=\; X \begin{bmatrix} z_{j1} & 0 & 0 & \cdots & 0 \\ 0 & z_{j2} & 0 & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & z_{jr} \end{bmatrix} Y^T \qquad j = 1{:}n.$$

If $U \in \mathbb{R}^{r\times r}$ is orthogonal

$$XU^T = \left[\; R \;\; 0 \;\right]$$

such that $R \in \mathbb{R}^{n\times n}$ is upper triangular and $V \in \mathbb{R}^{r\times r}$ is orthogonal so that

$$VY^T = \begin{bmatrix} S \\ 0 \end{bmatrix}$$

with $S \in \mathbb{R}^{n\times n}$ is upper triangular, then

$$R^{-1}A(:,:,j)S^{-1} \;=\; \left[\, I_n \;\; 0 \,\right] U \begin{bmatrix} z_{j1} & 0 & 0 & \cdots & 0 \\ 0 & z_{j2} & 0 & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & z_{jr} \end{bmatrix} V^T \begin{bmatrix} I_n \\ 0 \end{bmatrix}$$

This says that the act of finding a rank-r decomposition of A is (more or less) equivalent to finding $A_1, \ldots, A_n \in \mathbb{R}^{r\times r}$ so that

- The slices of A are the (1,1) blocks of the A_j

$$A_j(1{:}n, 1{:}n) = A(:,:,j) \qquad j = 1{:}n$$

- There exist upper triangular $\tilde{R}, \tilde{S} \in \mathbb{R}^{r\times r}$ so that the matrices

$$\tilde{R}^{-1} A_j \tilde{S}^{-1} \qquad j = 1{:}n$$

  have the same left and right singular vectors.

Embedding for Commutivity

We want to embed the slices into larger matrices and we'd like a friendly simultaneous diagonalization of those large matrices.

Think eigenvalues for a while.

Ignore Jordan block issues for a while.

If two matrices commute then we can find an eigenvector matrix that diagonalizes them both.

Given $A_1, A_2, A_3 \in \mathbb{R}^{n \times n}$, these matrices commute amongst themselves:

$$\tilde{A}_1 = \begin{bmatrix} A_1 & A_2 & A_3 \\ A_3 & A_1 & A_2 \\ A_2 & A_3 & A_1 \end{bmatrix} \qquad \tilde{A}_2 = \begin{bmatrix} A_2 & A_3 & A_1 \\ A_1 & A_2 & A_3 \\ A_3 & A_1 & A_2 \end{bmatrix} \qquad \tilde{A}_3 = \begin{bmatrix} A_3 & A_1 & A_2 \\ A_2 & A_3 & A_1 \\ A_1 & A_2 & A_3 \end{bmatrix}$$

A Question

Given $A_1, \ldots, A_N \in \mathbb{R}^{m \times n}$, can we find $\tilde{A}_1, \ldots, \tilde{A}_N \in \mathbb{R}^{\tilde{m} \times \tilde{n}}$ so that

- The SVDs of the $\tilde{A}_j$ involve the same U and V.
- $\tilde{A}_j(1{:}m, 1{:}n) = A_j, \quad j = 1{:}N.$

Back to the 2-by-2-by-2

If `A = randn(2,2,2)` then

$$\begin{aligned} \text{Prob}\,(\text{rank}(A) < 2) &= 0 \\ \text{Prob}\,(\text{rank}(A) = 2) &\approx .79 \\ \text{Prob}\,(\text{rank}(A) = 3) &\approx .21 \end{aligned}$$

There is a connection to a generalized eigenvalue problem.

The probability that the pencil

$$\begin{bmatrix} a_{111} & a_{121} \\ a_{211} & a_{221} \end{bmatrix} - \lambda \begin{bmatrix} a_{112} & a_{122} \\ a_{212} & a_{222} \end{bmatrix}$$

has complex eigenvalues is .21.

How near is such a pencil to a pencil with real eigenvalues?

The n-by-n-by-2

Using the generalized Schur decomposition, the probability that

```
A = randn(n,n,2)
```

has rank k is positive for

$$k = n,\ n+1, \ldots,\ \text{floor}(3n/2).$$

Conclusions

Not at all Schur about using nonorthogonal tensor decompositions!