Canonical Polyadic Decomposition

Center for Computational and Data-Intensive Science and Engineering (CDISE)
Skolkowo Institute of Science and Technology (SKOLTECH), Moscow, Russia

February 8, 2022

Part I

CRLB

CRLB I

- **Squared angular error SAE**
  The angle α_1 between the true factor $\boldsymbol{a}_1$ and its estimate $\hat{\boldsymbol{a}}_1$ obtained through the CP decomposition is defined through its cosine

$$\cos\alpha_1 = \frac{\boldsymbol{a}_1^T \hat{\boldsymbol{a}}_1}{\|\boldsymbol{a}_1\| \, \|\hat{\boldsymbol{a}}_1\|} \,. \tag{1}$$

$$\mathrm{SAE}(\boldsymbol{a}_1, \hat{\boldsymbol{a}}_1) = \frac{\boldsymbol{a}_1^T \hat{\boldsymbol{a}}_1}{\|\boldsymbol{a}_1\| \, \|\hat{\boldsymbol{a}}_1\|}.$$

  In applications it is practical to characterize the error of the factor $\boldsymbol{a}_1$ in the decomposition by a scalar quantity.
  Cramér-Rao-induced bound (CRIB) for the squared angle serves a gauge of achievable accuracy of estimation/CP decomposition

CRLB I

- Let a vector parameter $\boldsymbol{\theta}$ containing all parameters of our model be arranged as

$$\boldsymbol{\theta} = [(\operatorname{vec} \mathbf{A}_1)^T, \ldots, (\operatorname{vec} \mathbf{A}_N)^T]^T$$

- The maximum likelihood solution for $\boldsymbol{\theta}$ consists in minimizing the least squares criterion

$$\mathcal{Q}(\boldsymbol{\theta}) = \|\mathcal{Y} - \hat{\mathcal{Y}}(\boldsymbol{\theta})\|_F^2$$

where $\|\cdot\|_F$ stands for the Frobenius norm.

CRLB II

- The Cramér-Rao lower bound for estimating $\boldsymbol{\theta}$, in general, for this estimation problem, is given as the inverse of the Fisher information matrix

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})$$

  where $\mathbf{J}(\boldsymbol{\theta})$ is the Jacobi matrix (matrix of the first-order derivatives) of $Q(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

- In other words, the Fisher information matrix is proportional to the approximate Hessian matrix of the criterion,

$$\mathbf{H}(\boldsymbol{\theta}) = \mathbf{J}^T(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})$$

Hessian

Theorem

The Hessian $\mathbf{H}$ *can be decomposed into low rank matrices under the form as*

$$\mathbf{H} = \mathbf{G} + \mathbf{Z}\,\mathbf{K}\,\mathbf{Z}^T \tag{2}$$

where $\mathbf{K} = [\mathbf{K}_{nm}]_{n,m=1}^N$ *contains submatrices* $\mathbf{K}_{nm}$ *given by*

$$\mathbf{K}_{nm} = (1 - \delta_{nm})\mathbf{P}_R \,\mathtt{diag}\,(\mathtt{vec}\,\Gamma_{nm}) \tag{3}$$

$\mathbf{P}_R$ *is the permutation matrix of dimension* $R^2 \times R^2$ *such that* $\mathtt{vec}\,\mathbf{X} = \mathbf{P}_R\,\mathtt{vec}\,(\mathbf{X}^T)$ *for any* $R \times R$ *matrix* $\mathbf{X}$*, and* δ_{nm} *is the Kronecker delta.*

$$\mathbf{G} \quad = \quad \mathtt{blkdiag}\,(\Gamma_{nn} \otimes \mathbf{I}_{I_n})_{n=1}^N \tag{4}$$

and

$$\mathbf{Z} = \mathtt{blkdiag}\,(\mathbf{I}_R \otimes \mathbf{A}_n)_{n=1}^N \tag{5}$$

CRLB($\boldsymbol{a}$) I

- The CRLB for the first column of $\mathbf{A}_1$, denoted simply as $\boldsymbol{a}_1$, is defined as σ^2 times the left-upper submatrix of $\mathbf{H}_E^{-1}$ of the size $I_1 \times I_1$,

$$\mathrm{CRLB}(\boldsymbol{a}_1) = \sigma^2\, [\mathbf{H}_E^{-1}]_{1:I_1,1:I_1} \ . \tag{6}$$

where the reduced Hessian, $\mathbf{H}_E$, is obtained from $\mathbf{H}$ by deleting $(N-1)R$ rows and corresponding columns, because the estimation of one element in the vectors $\boldsymbol{a}_r^{(n)}$, $r = 1, \ldots, R$, $n = 2, \ldots, N$ can be skipped

$$\mathbf{H}_E = \mathbf{E}\mathbf{H}\mathbf{E}^T \tag{7}$$

where

$$\mathbf{E} = \mathtt{blkdiag}\left(\mathbf{I}_{RI_1}, \mathbf{I}_R \otimes \mathbf{E}_2, \ldots, \mathbf{I}_R \otimes \mathbf{E}_N\right) \tag{8}$$

and $\mathbf{E}_n$ is an $(I_n - 1) \times I_n$ matrix of rank $I_n - 1$. For example, one can put $\mathbf{E}_n = [\mathbf{0}_{(I_n-1)\times 1}\ \mathbf{I}_{I_n-1}]$ for $n = 2, \ldots, N$.

CRLB($\boldsymbol{a}$) II

- With this definition of $\mathbf{E}_n$, $\mathbf{H}_E$ is a Hessian for estimating the first factor matrix $\mathbf{A}_1$ and all other vectors $\boldsymbol{a}_r^{(n)}$, $r = 1, \ldots, R$, $n = 2, \ldots, N$ without their first elements

$$\mathbf{H}_E = \mathbf{G}_E + \mathbf{Z}_E \mathbf{K} \mathbf{Z}_E^T \tag{9}$$

where $\mathbf{G}_E = \mathbf{E}\mathbf{G}\mathbf{E}^T$ and $\mathbf{Z}_E = \mathbf{E}\mathbf{Z}$.

- Inverse of $\mathbf{H}_E$ can be written using a Woodbury matrix identity as

$$\mathbf{H}_E^{-1} = \mathbf{G}_E^{-1} - \mathbf{G}_E^{-1}\mathbf{Z}_E\mathbf{K}(\mathbf{I}_{NR^2} + \mathbf{Z}_E^T\mathbf{G}_E^{-1}\mathbf{Z}_E\mathbf{K})^{-1}\mathbf{Z}_E^T\mathbf{G}_E^{-1} \tag{10}$$

provided that the involved inverses exist.

CRLB($\boldsymbol{a}$) III

- **Theorem 2**: Let CRLB($\boldsymbol{a}_1$) be the Cramér-Rao bound on covariance matrix of unbiased estimators of $\boldsymbol{a}_1$. Then the Cramér–Rao–induced bound on the squared angular error between the true and estimated vector is

$$\mathrm{CRIB}(\boldsymbol{a}_1) = \frac{\mathrm{tr}[\Pi^{\perp}_{\mathbf{a}_1}\mathrm{CRLB}(\mathbf{a}_1)]}{||\mathbf{a}_1||^2} \tag{11}$$

where

$$\Pi^{\perp}_{\mathbf{a}_1} = \mathbf{I}_{I_1} - \mathbf{a}_1\mathbf{a}_1^T/||\mathbf{a}_1||^2 \tag{12}$$

is the projection operator to the orthogonal complement of $\mathbf{a}_1$ and tr(.) denotes trace of a matrix.

CRLB($\boldsymbol{a}$) IV

- **Theorem 3**: The CRIB($\boldsymbol{a}_1$) can be written in the form

$$\text{CRIB}(\boldsymbol{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left\{ (I_1 - 1) g_{11} - \operatorname{tr}\left[\mathbf{B}_0 \left((\boldsymbol{g}_{1,:}^T \boldsymbol{g}_{1,:}) \otimes \mathbf{X}_1 \right) \right] \right\} \quad (13)$$

where $\mathbf{B}_0$ is the submatrix of $\mathbf{B} = \mathbf{K}(\mathbf{I}_{NR^2} + \Psi \mathbf{K})^{-1}$, $\mathbf{B}_0 = \mathbf{B}_{1:R^2, 1:R^2}$,

$$\mathbf{X}_n = \mathbf{C}_n - \frac{1}{\mathbf{C}_{11}^{(n)}} \mathbf{C}_{:,1}^{(n)} \mathbf{C}_{:,1}^{(n)\,T} \quad (14)$$

for $n = 1, \ldots, N$, $\mathbf{C}_{11}^{(n)}$ and $\mathbf{C}_{:,1}^{(n)}$ denote the upper–right element and the first column of $\mathbf{C}_n$, respectively, and Ψ in the definition of $\mathbf{B}$ takes, for a special choice of matrices $\mathbf{E}_n$, the form

$$\Psi = \mathtt{blkdiag}\left(\Gamma_{11}^{-1} \otimes \mathbf{C}_1, \Gamma_{22}^{-1} \otimes \mathbf{X}_2, \ldots, \Gamma_{NN}^{-1} \otimes \mathbf{X}_N \right) \quad (15)$$

$$\Gamma_{nm} = \circledast_{k \neq n,m} \mathbf{C}_k \,, \qquad \mathbf{C}_k = \mathbf{A}_k^T \mathbf{A}_k \,. \quad (16)$$

Properties of the CRIB I

1. The CRIB in Theorems 3 and 4 depends on the factor matrices $\mathbf{A}_n$ only through the products $\mathbf{C}_n = \mathbf{A}_n^T \mathbf{A}_n$.
2. The CRIB is inversely proportional to the signal-to-noise ratio (SNR) of the factor of the interest (i.e. $\|\boldsymbol{a}_1\|^2/(\sigma^2 I_1)$) and independent of the SNR of the other factors, $\|\boldsymbol{a}_r\|^2/(\sigma^2 I_r)$, $r = 2, \ldots, R$.

CRIB for rank-2 tensors I

Theorem

$$\mathrm{CRIB}(\boldsymbol{a}_1) = \frac{l_1 - 1}{1 - h_1^2} + \frac{(1 - c_1^2)h_1^2}{1 - h_1^2} \frac{h_1^2 z(z+1) - y^2 - z}{(1 - c_1 y - h_1^2(z+1))^2 - h_1^2(y + c_1 z)^2}$$

where

$$h_n = \prod_{2 \le k \ne n}^{N} c_n \qquad \text{for} \quad n = 1, \ldots, N \tag{17}$$

$$y = -c_1 \sum_{n=2}^{N} \frac{h_n^2(1 - c_n^2)}{c_n^2 - h_n^2 c_1^2} \tag{18}$$

$$z = \sum_{n=2}^{N} \frac{1 - c_n^2}{c_n^2 - h_n^2 c_1^2}\,. \tag{19}$$

Example I

How to compute CRIB

Assume that $\sigma = 1$, then CRIB for estimation of the model $[\![\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N]\!]$

$$\text{crib} = \texttt{cribCP}(\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N\})$$

For other noise levels, CRIB is proportional to the noise level.

Example II

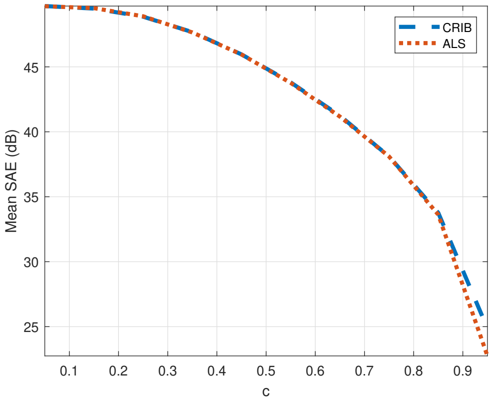

Figure: CRIB for estimation of $\mathbf{A}_n$ with identical collinear degree for all column vectors $\boldsymbol{a}_r^T \boldsymbol{a}_s = c$. Tensors are of size $30 \times 30 \times 30$ and rank-10.

Tensor decomposition through reshape I

- Assume that the tensor to-be decomposed is of dimension $N \geq 4$. The tensor can be reshaped to a lower dimensional tensor, which is easier to decompose, so that the first factor matrix remains unchanged. For example, consider $N = 4$.
- The tensor $\mathcal{Y}$ can be reshaped to a 3-way tensor

$$\mathcal{Y}_{res} = \sum_{r=1}^{R} \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ (\boldsymbol{a}_r^{(4)} \otimes \boldsymbol{a}_r^{(3)}) \,. \tag{20}$$

Both the original and the re-shaped tensors have the same number of elements ($I_1 I_2 I_3 I_4$) and the same noise added to them.

Tensor decomposition through reshape II

What is the accuracy of the factor matrix of the reshaped tensor compared to the original one?

The latter accuracy should be worse, because a decomposition of the reshaped tensor ignores structure of the third factor matrix. The question is, by how much worse.
If the difference were negligible, then it is advised to decompose the simpler tensor (of lower dimension).

-
- Let us examine tensors of rank 2. If the original tensor has correlations between columns of the factor matrices c_1, c_2, c_3 and c_4, the reshaped tensor has correlations c_1, c_2, and $c_3 c_4$, respectively.
- CRIB($\mathbf{a}_1$) of the reshaped tensor is independent of c_1, while CRIB of the original tensor is dependent on c_1, so there is a difference, in general.

Tensor decomposition through reshape III

- The difference will be smallest for $c_1 = 0$ (orthogonal factors) and largest for c_1 close to ± 1 (nearly or completely co-linear factors along the first dimension).
- The smallest difference between CRIB($\boldsymbol{a}_1$) for the reshaped tensor and for the original one is

$$\frac{c_2^2 + c_3^2 c_4^2 - 2c_2^2 c_3^2 c_4^2}{(1 - c_2^2)(1 - c_3^2 c_4^2)} - \frac{c_2^2 c_3^2 + c_2^2 c_4^2 + c_3^2 c_4^2 - 3c_2^2 c_3^2 c_4^2}{(1 - c_2^2 c_3^2 c_4^2)(2c_2^2 c_3^2 c_4^2 - c_2^2 c_3^2 - c_2^2 c_4^2 - c_3^2 c_4^2 + 1)}$$

  and the largest difference is

$$\frac{c_2^2 + c_3^2 c_4^2 - 2c_2^2 c_3^2 c_4^2}{(1 - c_2^2)(1 - c_3^2 c_4^2)} = \frac{c_2^2}{1 - c_2^2} + \frac{c_3^2 c_4^2}{1 - c_3^2 c_4^2} \, .$$

Tensor decomposition through reshape IV

- We can see that the difference may be large if the second or third factor matrix of the reshaped tensor has nearly co-linear columns ($c_2^2 \approx 1$ or $c_3^2 c_4^2 \approx 1$) .

References I