A randomized block sampling approach to the canonical polyadic decomposition of large-scale tensors

Nico Vervliet

Joint work with Lieven De Lathauwer

SIAM AN17, July 13, 2017
Classification of hazardous gasses using e-noses

Classify 900 experiments containing 72 time series with 26,000 samples each.
Overview

Decomposing large-scale tensors

Randomized block sampling

Experimental results

Chemo-sensing application
Canonical polyadic decomposition

- Sum of $R$ rank-1 terms

\[ T = c_1 a_1 b_1 + \cdots + c_R a_R b_R \]

Mathematically, for a general $N$th order tensor

\[ T = \sum_{r=1}^{R} a_r \otimes \cdots \otimes a_r \]

where $A(1), A(2), \ldots, A(N)$ are the core tensors.
Canonical polyadic decomposition

Sum of $R$ rank-1 terms

\[ \mathcal{T} = \sum_{r=1}^{R} c_1 a_1^{(r)} b_1^{(r)} + \cdots + c_R a_R^{(r)} b_R^{(r)} \]

Mathematically, for a general $N$th order tensor $\mathcal{T}$

\[ \mathcal{T} = \sum_{r=1}^{R} a_r^{(1)} \otimes a_r^{(2)} \otimes \cdots \otimes a_r^{(N)} \]

\[ = \begin{bmatrix} A^{(1)}, A^{(2)}, \ldots, A^{(N)} \end{bmatrix} \]
Computing a CPD

- Optimization problem:

\[
\min_{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}} \frac{1}{2} \left\| \mathbf{T} - \left[ \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)} \right] \right\|_F^2
\]
Computing a CPD

- Optimization problem:

\[
\min_{A^{(1)}, A^{(2)}, \ldots, A^{(N)}} \frac{1}{2} \left\| \mathcal{T} - \begin{bmatrix} A^{(1)} & A^{(2)} & \ldots & A^{(N)} \end{bmatrix} \right\|_F^2
\]

- Algorithms
  - Alternating least squares
  - CPOPT [Acar et al. 2011a]
  - (Damped) Gauss–Newton [Phan et al. 2013]
  - (Inexact) nonlinear least squares [Sorber et al. 2013]
Curse of dimensionality

- Suppose $N$th order $\mathcal{T} \in \mathbb{C}^{I \times I \times \cdots \times I}$, then
- number of entries: $I^N$
- memory and time complexity: $O(I^N)$

Example [Vervliet et al. 2014]

Ninth-order tensor with $I = 100$ and rank $R = 5$:
- number of entries: $10^{18}$
- number of variables: $4500$
Curse of dimensionality

- Suppose $N$th order $\mathcal{T} \in \mathbb{C}^{I \times I \times \cdots \times I}$, then
  - number of entries: $I^N$
  - memory and time complexity: $O(I^N)$
  - number of variables: $NIR$
Curse of dimensionality

- Suppose $N$th order $\mathcal{T} \in \mathbb{C}^{I \times I \times \cdots \times I}$, then
- number of entries: $I^N$
- memory and time complexity: $O(I^N)$
- number of variables: $NIR$

Example [Vervliet et al. 2014]

Ninth-order tensor with $I = 100$ and rank $R = 5$:
- number of entries: $10^{18}$
- number of variables: 4500
How to handle large tensors?

- **Use incomplete tensors**
  Acar et al. 2011b; Vervliet et al. 2014; Vervliet et al. 2016a

- **Exploit sparsity**

- **Compress the tensor**
  Sidiropoulos et al. 2014; Oseledets and Tyrtyshnikov 2010; Vervliet et al. 2016b

- **Decompose subtensors and combine results**
  Papalexakis et al. 2012; Phan and Cichocki 2011

- **Parallel**
  Liavas and Sidiropoulos 2015 + many of the above
Overview

Decomposing large-scale tensors

Randomized block sampling

Experimental results

Chemo-sensing application
Randomized block sampling CPD: idea
Randomized block sampling CPD: idea
Randomized block sampling CPD: idea

Take sample
Randomized block sampling CPD: idea
Randomized block sampling CPD: idea

Take sample \rightarrow \text{Initialization} \rightarrow \text{Compute step} \rightarrow \text{Update}
Randomized block sampling CPD: algorithm

**input**: Data $\mathcal{T}$ and initial guess $\mathbf{A}^{(n)}$, $n = 1, ..., N$

**output**: $\mathbf{A}^{(n)}$, $n = 1, ..., N$ such that $\mathcal{T} \approx \left[ \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)} \right]$

**while** $k < K$ **and not converged** **do**

Create sample $\mathcal{T}_s$ and corresponding $\mathbf{A}_s^{(n)}$, $n = 1, \ldots, N$

Let $\bar{\mathbf{A}}_s^{(n)}$ be the result of 1 iteration in a restricted CPD algorithm on $\mathcal{T}_s$ with initial guess $\mathbf{A}_s^{(n)}$, $n = 1, \ldots, N$ and restriction $\Delta$

Update the *affected* variables $\mathbf{A}^{(n)}$ using $\bar{\mathbf{A}}_s^{(n)}$, $n = 1, \ldots, N$

$k \leftarrow k + 1$
input : Data $\mathcal{T}$ and initial guess $A^{(n)}$, $n = 1, \ldots, N$
output: $A^{(n)}$, $n = 1, \ldots, N$ such that $\mathcal{T} \approx [A^{(1)}, \ldots, A^{(N)}]$

while $k < K$ and not converged do

Create sample $\mathcal{T}_s$ and corresponding $A^{(n)}_s$, $n = 1, \ldots, N$

Let $\bar{A}^{(n)}_s$ be the result of 1 iteration in a restricted CPD
algorithm on $\mathcal{T}_s$ with initial guess $A^{(n)}_s$, $n = 1, \ldots, N$ and
restriction $\Delta$

Update the affected variables $A^{(n)}$ using $\bar{A}^{(n)}_s$, $n = 1, \ldots, N$

$k \leftarrow k + 1$
Ingredient 1: randomized block sampling

For a $6 \times 6$ tensor and block size $3 \times 2$:

$I_1 = \{3, 1, 2, 6, 5, 4\}$
$I_2 = \{1, 2, 4, 6, 3, 5\}$
Ingredient 1: randomized block sampling

For a $6 \times 6$ tensor and block size $3 \times 2$:

$\mathcal{I}_1 = \{3, 1, 2, 6, 5, 4\} \quad \mathcal{I}_1 = \{3, 1, 2, 6, 5, 4\}$

$\mathcal{I}_2 = \{1, 2, 4, 6, 3, 5\} \quad \mathcal{I}_2 = \{1, 2, 4, 6, 3, 5\}$
Ingredient 1: randomized block sampling

For a $6 \times 6$ tensor and block size $3 \times 2$:

$\mathcal{I}_1 = \{3, 1, 2, 6, 5, 4\}$

$\mathcal{I}_2 = \{1, 2, 4, 6, 3, 5\}$

$\mathcal{I}_1 = \{3, 1, 2, 6, 5, 4\}$

$\mathcal{I}_2 = \{1, 2, 4, 6, 3, 5\}$

$\mathcal{I}_1 = \{6, 1, 4, 2, 5, 3\}$

$\mathcal{I}_2 = \{1, 2, 4, 6, 3, 5\}$
Ingredient 2: restricted CPD algorithm

- **ALS variant**

\[
A_{k+1}^{(n)} = (1 - \alpha)A_k^{(n)} + \alpha T(n) \tilde{V}^{(n)}(\tilde{W}^{(n)})^{-1}
\]

Enforce restriction by \(\alpha = \Delta_k\).
Ingredient 2: restricted CPD algorithm

- ALS variant

\[ \mathbf{A}_{k+1}^{(n)} = (1 - \alpha) \mathbf{A}_k^{(n)} + \alpha \mathbf{T}_k \mathbf{V}_k^{(n)} (\mathbf{W}_k^{(n)})^{-1} \]

Enforce restriction by \( \alpha = \Delta_k \).

- NLS variant

\[
\min_{\mathbf{p}_k} \frac{1}{2} \| \text{vec} (\mathcal{F}(\mathbf{x}_k)) - \mathbf{J}_k \mathbf{p}_k \|_2^2 \quad \text{s.t.} \quad \| \mathbf{p}_k \| \leq \Delta_k
\]

in which

\[ \mathcal{F} = \mathcal{T} - \left[ \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(N)} \right] \]
Ingredient 3: restriction

Use restriction of form

\[
\Delta_k = \begin{cases} 
\Delta_0 & \text{if } k < K_{\text{search}} \\
\hat{\Delta}_0 \cdot \alpha^{(k-K_{\text{search}})/Q} & \text{if } k \geq K_{\text{search}} 
\end{cases}
\]

Example (Selecting \(Q\))

For a 100 \times 100 \times 100 tensor and block size 25 \times 25 \times 25, \(Q = 4\).
Ingredient 3: restriction

Use restriction of form

\[ \Delta_k = \begin{cases} \Delta_0 & \text{if } k < K_{\text{search}} \\ \hat{\Delta}_0 \cdot \alpha^{(k-K_{\text{search}})/Q} & \text{if } k \geq K_{\text{search}} \end{cases} \]

Example (Selecting $Q$)

For a $100 \times 100 \times 100$ tensor and block size $25 \times 25 \times 25$, $Q = 4$
Ingredient 4: A stopping criterion

- Function evaluation $f_{val} = 0.5 \left\| T - \left[ A^{(1)}, \ldots, A^{(N)} \right] \right\|^2$

![Graph showing function evaluation $f_{val}$ and CPD Error over iterations](image)
Ingredient 4: A stopping criterion

- Function evaluation $f_{\text{val}} = 0.5 \| \mathcal{T} - [A^{(1)}, \ldots, A^{(N)}] \|^2$

- Step size
Intermezzo: Cramér–Rao bound

- Uncertainty of an estimate

\[\text{68\%} \]

\[-3\sigma \, -2\sigma \, -\sigma \, 0 \, \sigma \, 2\sigma \, 3\sigma\]
Intermezzo: Cramér–Rao bound

- Uncertainty of an estimate

\[ \text{CRB} \leq \sigma^2 \]
Intermezzo: Cramér–Rao bound

- Uncertainty of an estimate

\[ \text{CRB} \leq \sigma^2 \]

\[ \mathbf{C} = \tau^2 (\mathbf{J}^H \mathbf{J})^{-1} \]
Ingredient 4: Cramér–Rao bound based stopping criterion

- Experimental bound
  - Use estimates $A_k^{(n)}$
  - Use $f_{val}$ to estimate noise $\tau$

$$\text{Stopping criterion: } D_{CRB} = \left| \frac{A_k^{(n)}}{\sum_{n=1}^{N} I_n} \right| \leq \gamma$$
Ingredient 4: Cramér–Rao bound based stopping criterion

- Experimental bound
  - Use estimates $A_k^{(n)}$
  - Use $f_{\text{val}}$ to estimate noise $\tau$

- Stopping criterion:

$$D_{\text{CRB}} = \frac{1}{R \sum_n l_n} \sum_{n=1}^N \sum_{i=1}^{l_n} \sum_{r=1}^R \left| \frac{A_k^{(n)}(i, r) - A_{k-K_{\text{CRB}}}^{(n)}(i, r)}{\sqrt{C^{(n)}(i, r)}} \right| \leq \gamma$$
Ingredient 4: Cramér–Rao bound based stopping criterion

- Experimental bound
  - Use estimates $A_k^{(n)}$
  - Use $f_{\text{val}}$ to estimate noise $\tau$

- Stopping criterion:

$$D_{\text{CRB}} = \frac{1}{R \sum_n I_n} \sum_{n=1}^N \sum_{i=1}^{I_n} \sum_{r=1}^R \left| A_k^{(n)}(i, r) - A_{k - K_{\text{CRB}}}^{(n)}(i, r) \right| \sqrt{C^{(n)}(i, r)} \leq \gamma$$
Unrestricted phase vs restricted phase

- Unrestricted phase (1 + 2): converge to a neighborhood of an optimum
- Restricted phase (3): pull iterates towards optimum
Unrestricted phase vs restricted phase

- Unrestricted phase (1 + 2): converge to a neighborhood of an optimum
- Restricted phase (3): pull iterates towards optimum

Assumptions:
- CPD of rank $R$ exists
- SNR is high enough
- Most block dimensions $> R$
Unrestricted phase vs restricted phase

Unrestricted phase $(1 + 2)$: converge to a neighborhood of an optimum
Restricted phase $(3)$: pull iterates towards optimum

Assumptions

- CPD of rank $R$ exists
- SNR is high enough
- Most block dimensions $> R$
Overview

Decomposing large-scale tensors

Randomized block sampling

Experimental results

Chemo-sensing application
Experiment overview

- Experiments
  - Comparison ALS vs NLS (see paper)
  - Influence of block size
  - Influence of step size (see paper)
Experiment overview

- Experiments
  - Comparison ALS vs NLS (see paper)
  - Influence of block size
  - Influence of step size (see paper)

- Performance
  - 50 Monte Carlo experiments
  - CPD error

\[
\max_n \left| \frac{A_0^{(n)} - A_{\text{res}}^{(n)}}{A_0^{(n)}} \right|
\]
Experiment overview

- Experiments
  - Comparison ALS vs NLS (see paper)
  - Influence of block size
  - Influence of step size (see paper)

- Performance
  - 50 Monte Carlo experiments
  - CPD error
    \[ \max_n \left\| A_0^{(n)} - A_{\text{res}}^{(n)} \right\| / \left\| A_0^{(n)} \right\| \]

- \texttt{cpd\_rbs} in Tensorlab 3.0 [Vervliet et al. 2016c]
Influence of block size: setup

\[(4 \times 4 \times 2) \cdot \nu \quad \mathcal{U}(0, 1)\]

\[800 \times 800 \times 400\]

\[R = 20\]

No noise
Influence of block size on computation time

\[ (4 \times 4 \times 2) \cdot \nu \]

\[ R = 20, \mathcal{U}(0, 1) \]

800 \times 800 \times 400

No noise
Influence of block size on data accesses

800 × 800 × 400

(4 × 4 × 2) · ν

$R = 20, \mathcal{U}(0, 1)$

No noise
Influence of block size on accuracy

\[ E_{\text{CPD}} \]

\[ \nu \]

\[ R = 20, \mathcal{U}(0, 1) \]

\[ 800 \times 800 \times 400 \]

\[ (4 \times 4 \times 2) \cdot \nu \]

\[ 20 \text{ dB} \]
Influence of block size on accuracy

\[ 800 \times 800 \times 400 \quad (4 \times 4 \times 2) \cdot \nu \quad R = 20, \mathcal{U}(0, 1) \quad 20 \text{ dB} \]
Overview

Decomposing large-scale tensors

Randomized block sampling

Experimental results

Chemo-sensing application
Classify hazardous gasses

Does the sample contain CO, acetaldehyde or ammonia?

Strategy: classify using coefficients of spatiotemporal patterns.

\[
26000 \times 72 \times 900 \quad 100 \times 36 \times 100 \quad R = 5 \quad \text{Unknown}
\]
Classify hazardous gasses: results

- Resulting factor matrices

<table>
<thead>
<tr>
<th></th>
<th>Iterations</th>
<th>Time (s)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No restriction</td>
<td>3000</td>
<td>60</td>
<td>5.0</td>
</tr>
<tr>
<td>Restriction</td>
<td>9000</td>
<td>170</td>
<td>0.3–0.8</td>
</tr>
</tbody>
</table>

![Time, Sensor, Experiment graphs]
Classify hazardous gasses: results

- Resulting factor matrices

- Performance after clustering

| No restriction | 3000 | 60   | 5.0
|----------------|------|------|-----|
| Restriction    | 9000 | 170  | 0.3–0.8
Conclusion

- The randomized block sampling CPD algorithm enables the decomposition of larger tensors, using fewer data points and less memory.
- Block size controls accuracy, data accesses and time.
- Step size restriction improves accuracy.
- Cramér–Rao bound based stopping criterion combines noise and step information.
More details:

A randomized block sampling approach to the canonical polyadic decomposition of large-scale tensors

Nico Vervliet

Joint work with Lieven De Lathauwer

SIAM AN17, July 13, 2017
References 1


References II


References IV


References V

