The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization

IAEA International Atomic Energy Agency

**2044-8**

**Summer School and Advanced Workshop on Trends and Developments in Linear Algebra**

***22 June - 10 July, 2009***

# Tensor decompositions and tensors ranks: general concepts and algorithms

Eugene Tyrtyshnikov

*Russian Academy of Sciences*
*Moscow*
*Russia*

# TENSOR DECOMPOSITIONS AND TENSOR RANKS: GENERAL CONCEPTS AND ALGORITHMS

Eugene Tyrtyshnikov

Institute of Numerical Mathematics
Russian Academy of Sciences

# WHAT ARE TENSORS?

TENSOR = MULTI-INDEX ARRAY = MULTI-WAY ARRAY = MULTI-DIMENSIONAL MATRIX:

$$\boldsymbol{A} = [\boldsymbol{a}_{\boldsymbol{ij\ldots k}}]$$

$$\boldsymbol{i} \in \boldsymbol{I}, \quad \boldsymbol{j} \in \boldsymbol{J}, \quad \ldots, \quad \boldsymbol{k} \in \boldsymbol{K}$$

Number of different indices is *dimension*.

Indices are called also *modes*.

Cardinalities of index ranges $\boldsymbol{I}, \boldsymbol{J}, \ldots, \boldsymbol{K}$ are *mode sizes*.

In case of dimension $\boldsymbol{d}$ and mode sizes $\boldsymbol{n_1}, \boldsymbol{n_2}, \ldots, \boldsymbol{n_d}$,
$\boldsymbol{A}$ is a *tensor of size* $\boldsymbol{n_1} \times \boldsymbol{n_2} \times \ldots \times \boldsymbol{n_d}$.

Talking of tensors, tacitly assume that $\boldsymbol{d} \geq \boldsymbol{3}$.

# TENSORS AND MATRICES

Let $\boldsymbol{A} = [\boldsymbol{a}_{ijklm}]$.
Consider pairs of complementary *long indices*

$$(\boldsymbol{ij}) \quad \text{and} \quad (\boldsymbol{klm})$$

$$(\boldsymbol{kl}) \quad \text{and} \quad (\boldsymbol{ijm})$$

$$\ldots\ldots\ldots$$

Then $\boldsymbol{A}$ gives rise to several matrices:

$$\boldsymbol{B}_1 = [\boldsymbol{b}_{(ij),(klm)}],$$

$$\boldsymbol{B}_2 = [\boldsymbol{b}_{(kl),(ijm)}]$$

$$\ldots\ldots\ldots$$

with

$$\boldsymbol{b}_{(ij),(klm)} = \boldsymbol{b}_{(kl),(ijm)} = \ldots = \boldsymbol{a}_{ijklm}$$

# MODE UNFOLDING MATRICES

$$\boldsymbol{A}_1 = [\boldsymbol{a}_{i,(jklm)}]$$
$$\boldsymbol{A}_2 = [\boldsymbol{a}_{j,(iklm)}]$$
$$\boldsymbol{A}_3 = [\boldsymbol{a}_{k,(ijlm)}]$$
$$\boldsymbol{A}_4 = [\boldsymbol{a}_{l,(ijkm)}]$$
$$\boldsymbol{A}_5 = [\boldsymbol{a}_{m,(ijkl)}]$$

Columns of unfolding matrices are called *mode vectors*.

If $\boldsymbol{d} = \boldsymbol{3}$, typical names are *columns, rows, fibers*.

Ranks of unfolding matrices are called *mode ranks* or *Tucker ranks*.


L. R. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika*, V. 31, P. 279–311 (1966).

# TENSOR-BY-MATRIX MULTIPLICATIONS

Also called *mode contractions*.

Given a tensor $\boldsymbol{A} = [\boldsymbol{a}_{ijk}]$ and matrices

$$\boldsymbol{U} = [\boldsymbol{u}_{i'i}], \quad \boldsymbol{V} = [\boldsymbol{v}_{j'j}], \quad \boldsymbol{W} = [\boldsymbol{w}_{k'k}],$$

define new tensors

$$\boldsymbol{A}^{\boldsymbol{U}} = \boldsymbol{A} \times_1 \boldsymbol{U} = [\boldsymbol{a}^{\boldsymbol{U}}_{i'jk}]$$
$$\boldsymbol{A}^{\boldsymbol{V}} = \boldsymbol{A} \times_2 \boldsymbol{V} = [\boldsymbol{a}^{\boldsymbol{V}}_{ij'k}]$$
$$\boldsymbol{A}^{\boldsymbol{W}} = \boldsymbol{A} \times_3 \boldsymbol{W} = [\boldsymbol{a}^{\boldsymbol{W}}_{ijk'}]$$

as follows:

$$\boldsymbol{a}^{\boldsymbol{U}}_{i'jk} = \sum_i \boldsymbol{u}_{i'i}\, \boldsymbol{a}_{ijk} \quad \Leftrightarrow \quad \boldsymbol{A}^{\boldsymbol{U}}_1 = \boldsymbol{U}\boldsymbol{A}_1$$
$$\boldsymbol{a}^{\boldsymbol{V}}_{ij'k} = \sum_j \boldsymbol{v}_{j'j}\, \boldsymbol{a}_{ijk} \quad \Leftrightarrow \quad \boldsymbol{A}^{\boldsymbol{V}}_2 = \boldsymbol{V}\boldsymbol{A}_2$$
$$\boldsymbol{a}^{\boldsymbol{W}}_{ijk'} = \sum_k \boldsymbol{w}_{k'k}\, \boldsymbol{a}_{ijk} \quad \Leftrightarrow \quad \boldsymbol{A}^{\boldsymbol{W}}_3 = \boldsymbol{W}\boldsymbol{A}_3$$

## WHY CONTRACTIONS?

Let $\boldsymbol{A} = [\boldsymbol{a}_{ijk}]$ be $\boldsymbol{n} \times \boldsymbol{n} \times \boldsymbol{n}$ and mode ranks be equal to $\boldsymbol{r} \ll \boldsymbol{n}$. Consider $\boldsymbol{QR}$ decompositions of unfolding matrices

$$\boldsymbol{A}_1 = \boldsymbol{Q}_1\boldsymbol{R}_1, \quad \boldsymbol{A}_2 = \boldsymbol{Q}_2\boldsymbol{R}_2, \quad \boldsymbol{A}_3 = \boldsymbol{Q}_3\boldsymbol{R}_2$$

$\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3$ are orthogonal $\boldsymbol{n} \times \boldsymbol{r}$ matrices.

Define the *Tucker core* tensor $\boldsymbol{G} = [\boldsymbol{g}_{\alpha\beta\gamma}]$ of *contracted* size $\boldsymbol{r} \times \boldsymbol{r} \times \boldsymbol{r}$:

$$\boldsymbol{G} = \boldsymbol{A} \times_1 \boldsymbol{Q}_1^\top \times_2 \boldsymbol{Q}_2^\top \times_3 \boldsymbol{Q}_3^\top \quad \text{i.e.} \quad \boldsymbol{g}_{\alpha\beta\gamma} = \sum_{i,j,k} \boldsymbol{a}_{ijk}\, \boldsymbol{q}^1_{i\alpha}\, \boldsymbol{q}^2_{j\beta}\, \boldsymbol{q}^3_{k\gamma}$$

## THEOREM

$$\boldsymbol{A} = \boldsymbol{G} \times_1 \boldsymbol{Q}_1 \times_2 \boldsymbol{Q}_2 \times_3 \boldsymbol{Q}_3 \quad \text{i.e.} \quad \boldsymbol{a}_{ijk} = \sum_{\alpha,\beta,\gamma} \boldsymbol{g}_{\alpha\beta\gamma}\, \boldsymbol{q}^1_{i\alpha}\, \boldsymbol{q}^2_{j\beta}\, \boldsymbol{q}^3_{k\gamma}$$

**IMPORTANT:** $\boldsymbol{A}$ is now represented in a *contracted form* with only $\boldsymbol{3nr + r^3 \ll n^3}$ parameters.

# TUCKER DECOMPOSITION

Regarded as *Tensor SVD* or *Higher Order SVD*:

$$\boldsymbol{A} = \boldsymbol{G} \times_1 \boldsymbol{Q}_1 \times_2 \boldsymbol{Q}_2 \times_3 \boldsymbol{Q}_3 \quad \text{i.e.} \quad \boldsymbol{a}_{ijk} = \sum_{\alpha,\beta,\gamma} \boldsymbol{g}_{\alpha\beta\gamma}\, \boldsymbol{q}^1_{i\alpha}\, \boldsymbol{q}^2_{j\beta}\, \boldsymbol{q}^3_{k\gamma}$$

Orthogonal matrices $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3$ are *Tucker factors* or *frame matrices*.

## THEOREM

Rows in each of unfolding matrices for the Tucker core can be made *orthogonal* and arranged in *length-decreasing order*.
Row lengths of unfoldings for $\boldsymbol{G}$ = singular values of unfoldings for $\boldsymbol{A}$.

PROOF is easy via SVD of unfolding matrices:

if $\boldsymbol{A}_1 = \boldsymbol{Q}_1 \boldsymbol{\Sigma}_1 \boldsymbol{V}_1$ then $(\boldsymbol{A} \times_1 \boldsymbol{Q}_1^\top)_1 = \boldsymbol{\Sigma}_1 \boldsymbol{V}_1$.

Same for other modes.

# TUCKER APPROXIMATIONS

$$a_{ijk} \approx \sum_{\alpha,\beta,\gamma} g_{\alpha\beta\gamma}\, q^1_{i\alpha}\, q^2_{j\beta}\, q^3_{k\gamma}$$

APPLICATIONS:

- Multi-way Principal Component Analysis
  (senior frame matrices are most informative).
- Tensor data compression
  (ignore small and get to reduced Tucker ranks $\ll$ mode sizes).
- New generation of numerical algorithms
  with all data in the Tucker format.
  Enjoy linear and even sublinear complexity in total size of data
  (could be petabytes).


I. Oseledets, D. Savostyanov, E. Tyrtyshnikov, *Linear algbera for tensor problems*, submitted to *Computing* (2008).

G. Beylkin, M. Mohlenkamp, Algorithms for numerical analysis in high dimensions, *SIAM J. Sci. Comput.*, 26 (6), pp. 2133-2159 (2005).

## CANONICAL DECOMPOSITION

$$\boldsymbol{a}_{ij\ldots k} = \sum_{t=1}^{\rho} \boldsymbol{u}_{it}\, \boldsymbol{v}_{jt}\ldots\, \boldsymbol{w}_{kt}$$

Minimal $\boldsymbol{\rho} = \textbf{tRank}$ is called *canonical rank* or *tensor rank* of $\boldsymbol{A}$.

## THEOREM

Let mode ranks be egual to $\boldsymbol{r}$. Then

$$\boldsymbol{r} \le \textbf{tRank}(\boldsymbol{A}) \le \boldsymbol{r}^2.$$

## CANONICAL APPROXIMATIONS

$$\boldsymbol{a}_{ij\ldots k} \approx \sum_{t=1}^{\rho} \boldsymbol{u}_{it}\, \boldsymbol{v}_{jt}\ldots\, \boldsymbol{w}_{kt}$$

play same compression role as Tucker.

**Could be better but not necessarily!**

## TENSOR RANKS IN COMPLEXITY THEORY

In the "row-by-column" rule for multiplication of $n \times n$ matrices we have $n^2$ multiplications. Can we reduce this number?

$$\begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$$

$$c_k = \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ijk}\, a_i\, b_j$$

Let $\rho =$ tensor rank of $h_{ijk}$ and canonical decomposition read

$$h_{ijk} = \sum_{t=1}^{\rho} u_{it}\, v_{jt}\, w_{kt} \quad \Rightarrow$$

$$c_k = \sum_{t=1}^{\rho} w_{kt} \left( \sum_{i=1}^{4} u_{it} a_i \right) \left( \sum_{j=1}^{n} v_{jt} b_j \right)$$

**Now we have $\rho$ multiplications!**

If $n = 2$ then $\rho = 7$ (Strassen, 1965).

By recursion $\Rightarrow$ only $O(n^{\log_2 7})$ multiplications for arbitrary $n$.

# TUCKER VS CANONICAL FOR MATRICES

$$\boldsymbol{a_{ij}} = \sum_{\alpha=1}^{r}\sum_{\beta=1}^{r} \boldsymbol{g_{\alpha\beta} q^1_{i\alpha} q^2_{j\beta}} \quad \Leftrightarrow \quad \boldsymbol{A} = \boldsymbol{Q_1 G Q_2^\top}$$

Tucker = a *pseudo-skeleton* decomposition of $\boldsymbol{A}$.

$$\boldsymbol{a_{ij}} = \sum_{t=1}^{\rho} \boldsymbol{u_{it} v_{jt}} \quad \Leftrightarrow \quad \boldsymbol{A} = \boldsymbol{U V^\top}$$

Canonical = a *skeleton* or *dyadic* decomposition of $\boldsymbol{A}$.

**Tensor (canonical) rank seems to be a true generalization of the matrix rank concept.**

**However, tensor rank for dimension $\geq$ 3 and matrix rank have *noticably different properties*.**

# KRONECKER PRODUCT REPRESENTATION

Tucker decomposition:

$$\boldsymbol{A} = \sum_{\alpha,\beta,\gamma} \boldsymbol{g}_{\alpha\beta\gamma}\, \boldsymbol{u}_{\alpha} \otimes \boldsymbol{v}_{\beta} \otimes \boldsymbol{w}_{\gamma}$$

Canonical decomposition:

$$\boldsymbol{A} = \sum_{t} \boldsymbol{u}_{t} \otimes \boldsymbol{v}_{t} \otimes \boldsymbol{w}_{t}$$

# KRUSKAL (ESSENTIAL) UNIQUENESS

Minimal canonical decomposition

$$\boldsymbol{A} = \sum_t \boldsymbol{u}_t \otimes \ \boldsymbol{v}_t \otimes \ \boldsymbol{w}_t$$

is said to be *essentially unique* if

$$\sum_t \boldsymbol{u}_{it} \otimes \ \boldsymbol{v}_{jt} \otimes \ \boldsymbol{w}_{kt} = \sum_t \bar{\boldsymbol{u}}_{it} \otimes \ \bar{\boldsymbol{v}}_{jt} \otimes \ \bar{\boldsymbol{w}}_{kt}$$

implies that, upon some reordering,

$$\boldsymbol{u}_t \mid\mid \bar{\boldsymbol{u}}_t, \qquad \boldsymbol{v}_t \mid\mid \bar{\boldsymbol{v}}_t, \qquad \boldsymbol{w}_t \mid\mid \bar{\boldsymbol{w}}_t,$$

$$||\boldsymbol{u}_t \otimes \boldsymbol{v}_t \otimes \boldsymbol{w}_t|| = ||\bar{\boldsymbol{u}}_t \otimes \bar{\boldsymbol{v}}_t \otimes \bar{\boldsymbol{w}}_t||.$$

Matrix skeleton (dyadic) decomposition is NOT ESSENTIALLY UNIQUE.
This becomes an obstacle in Principal Component Analysis,
e.g. in `separation of signals`.
Despite that, tensors possess ESSENTIAL UNIQUENESS
(under mild assumptions).

## INDEPENDENT COMPONENT RECONSTRUCTION

EXAMPLE (De Lathauwer) where the PCA fails.

Assume we need to separate two independent zero-mean signals

$$x_1(t) = \sqrt{2}\sin t, \qquad x_2(t) = \begin{cases} 1 & \text{if} \quad k\pi \le t < k\pi + \pi/2, \\ -1 & \text{if} \quad k\pi + \pi/2 \le t < k\pi + \pi, \end{cases}$$

defined on the interval $0 \le t \le 4\pi$ and mixed by a matrix

$$A = \begin{pmatrix} -1 & -3\sqrt{3} \\ 3\sqrt{3} & -5 \end{pmatrix}.$$

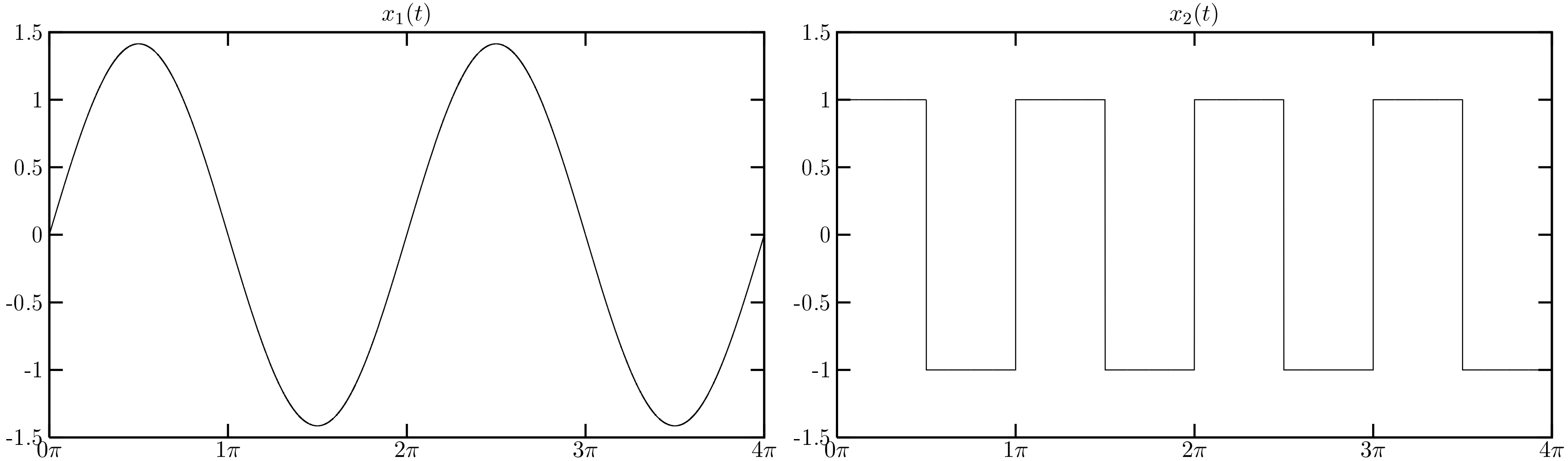


FIG. 1. Signals to be separated.

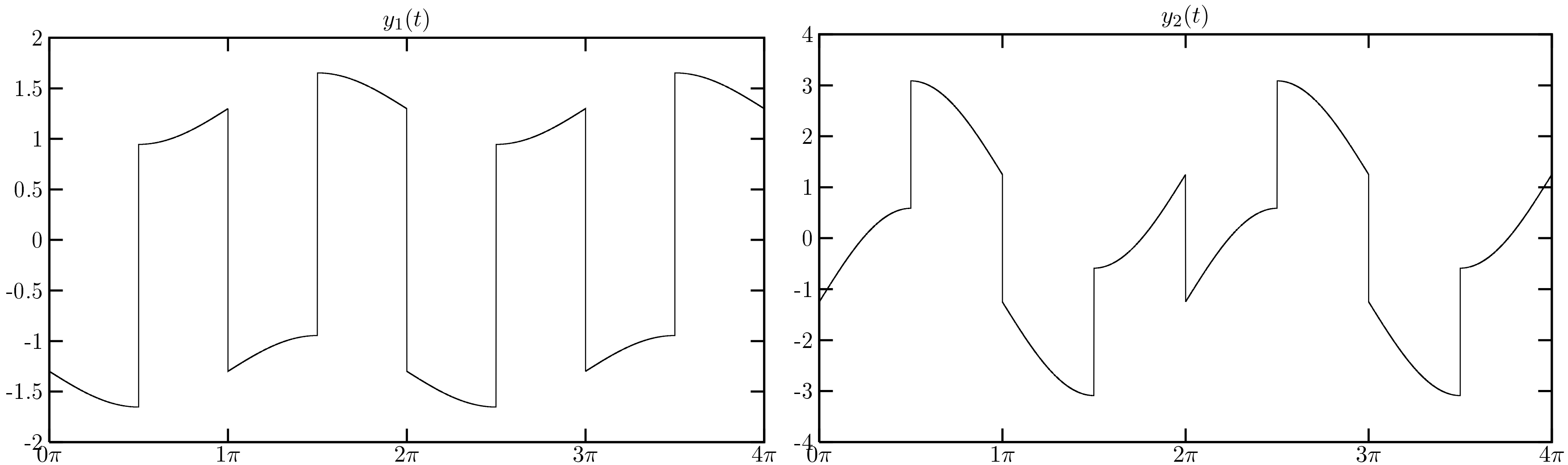


**Fig. 2.** Observations of linearly mixed signals.

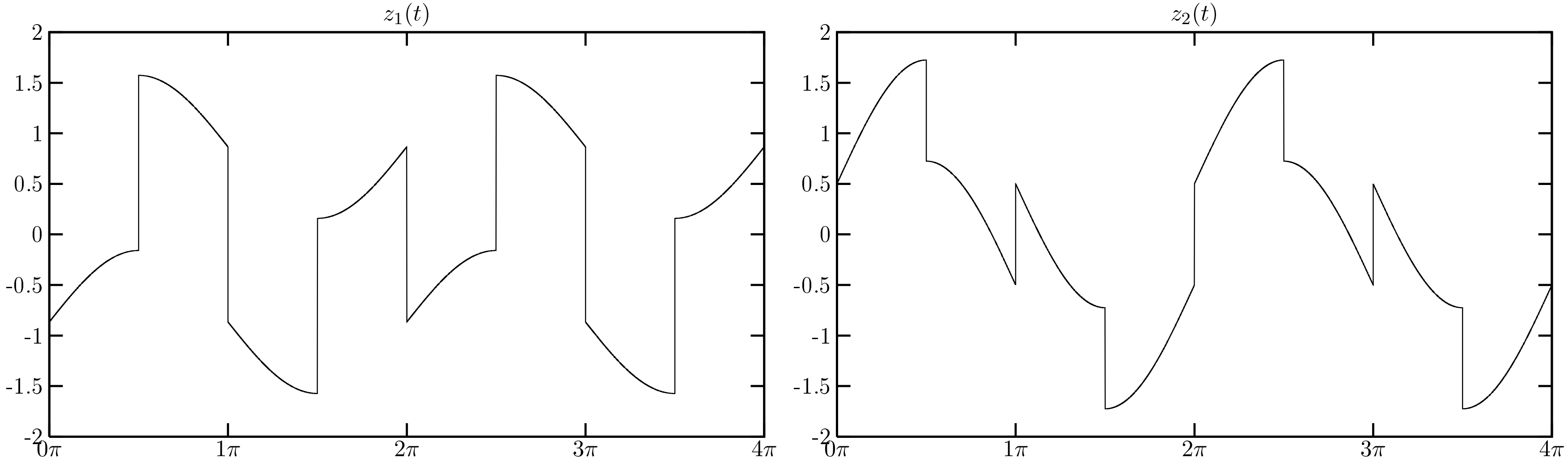


FIG. 3. Seperation results produced by the PCA.

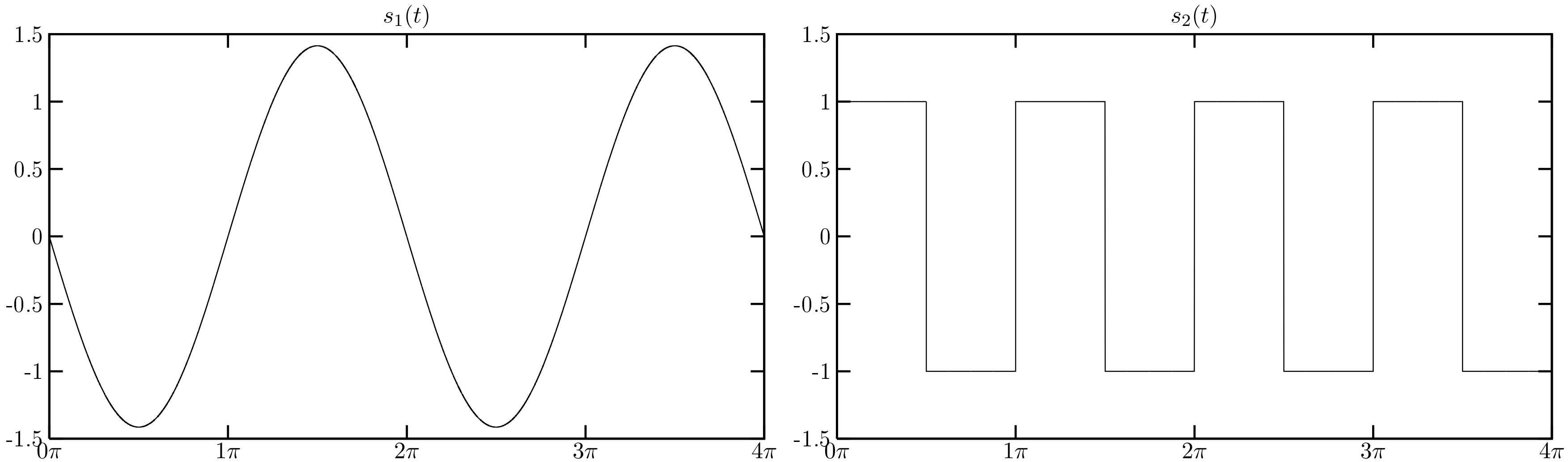


FIG. 4. Separation results of the tensor technique.

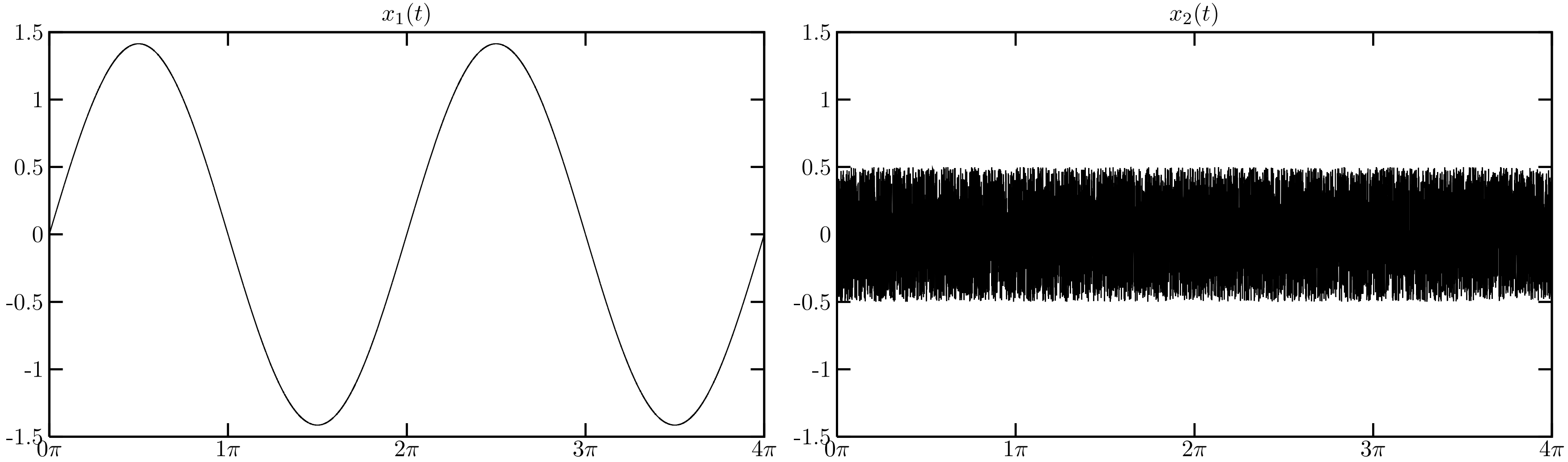


FIG. 1. Signals to be separated.

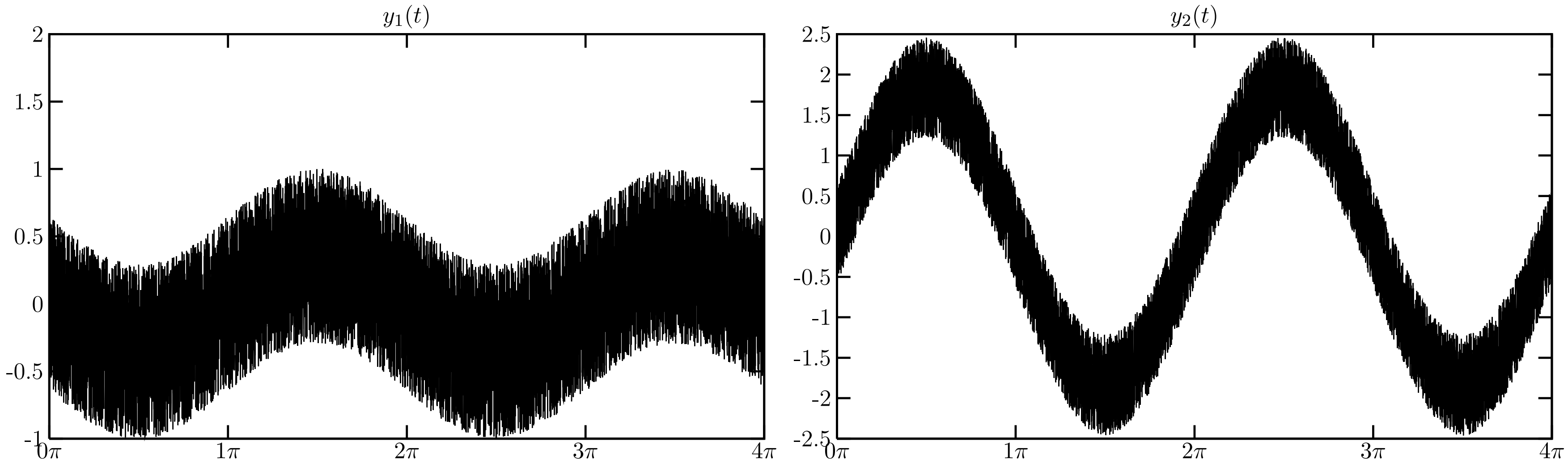


**Fig. 2.** Observations of linearly mixed signals.

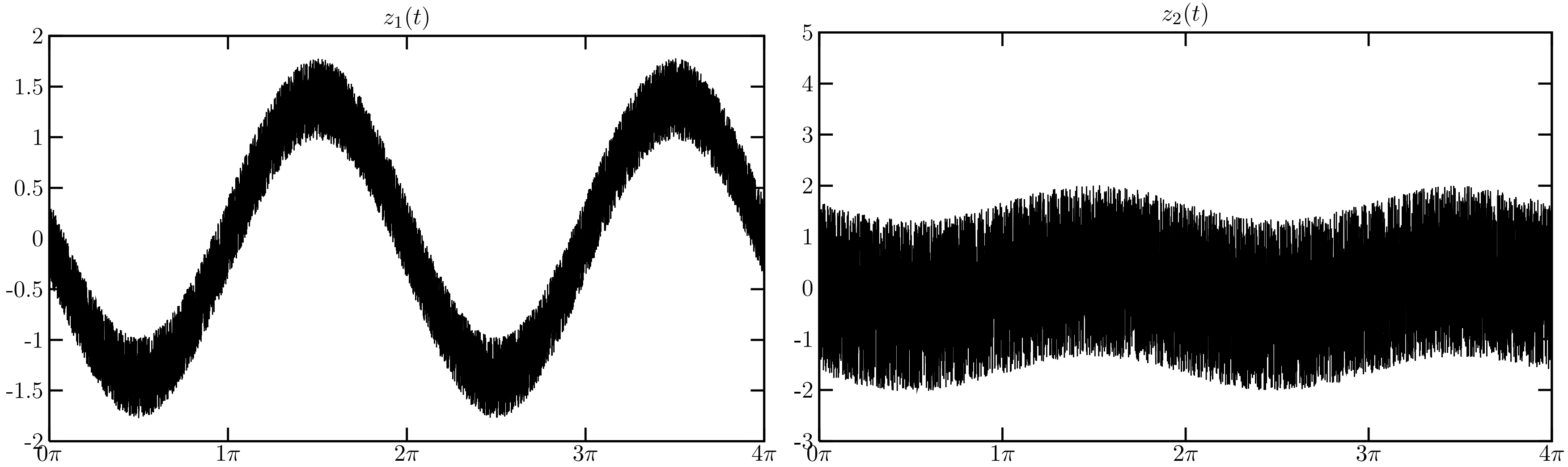


FIG. 3. Seperation results produced by the PCA.

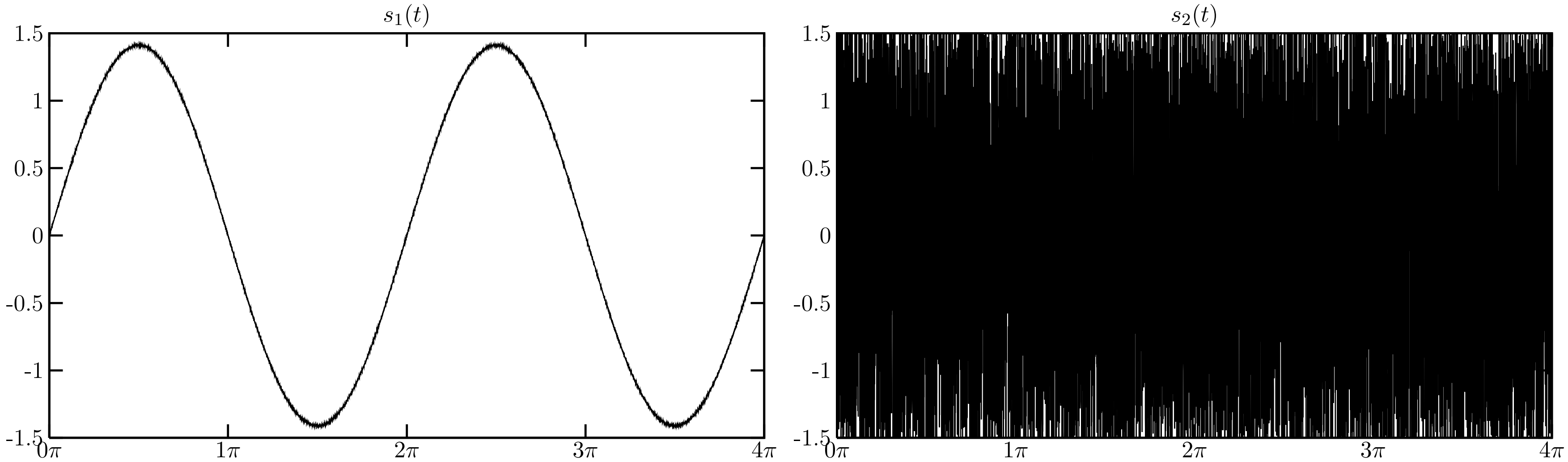


FIG. 4. Separation results of the tensor technique.

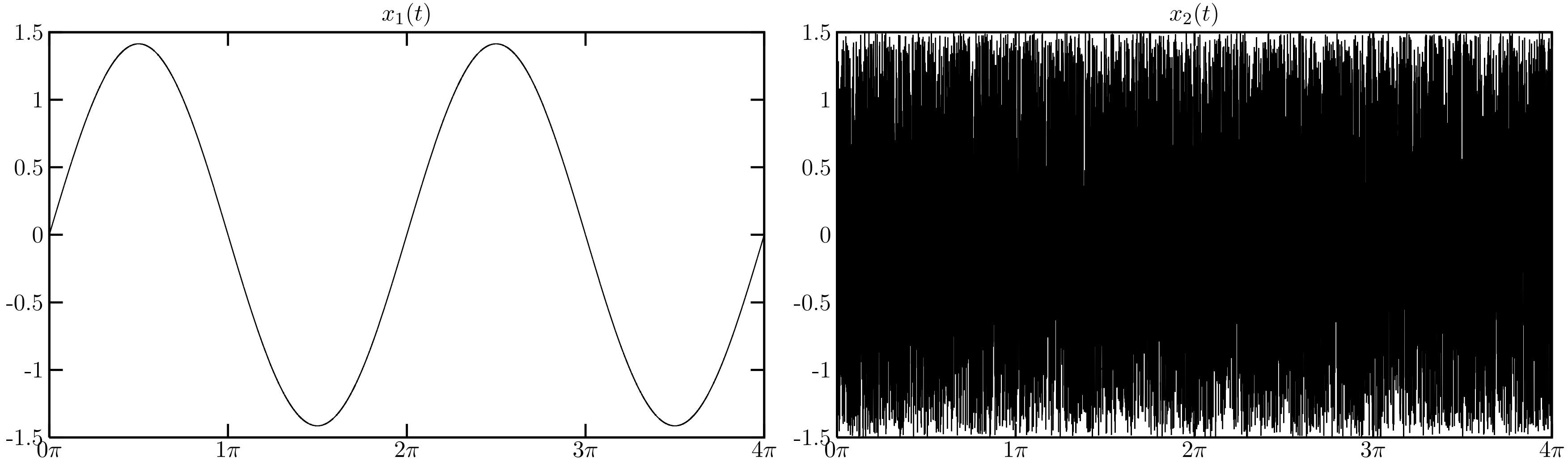


FIG. 1. Signals to be separated.

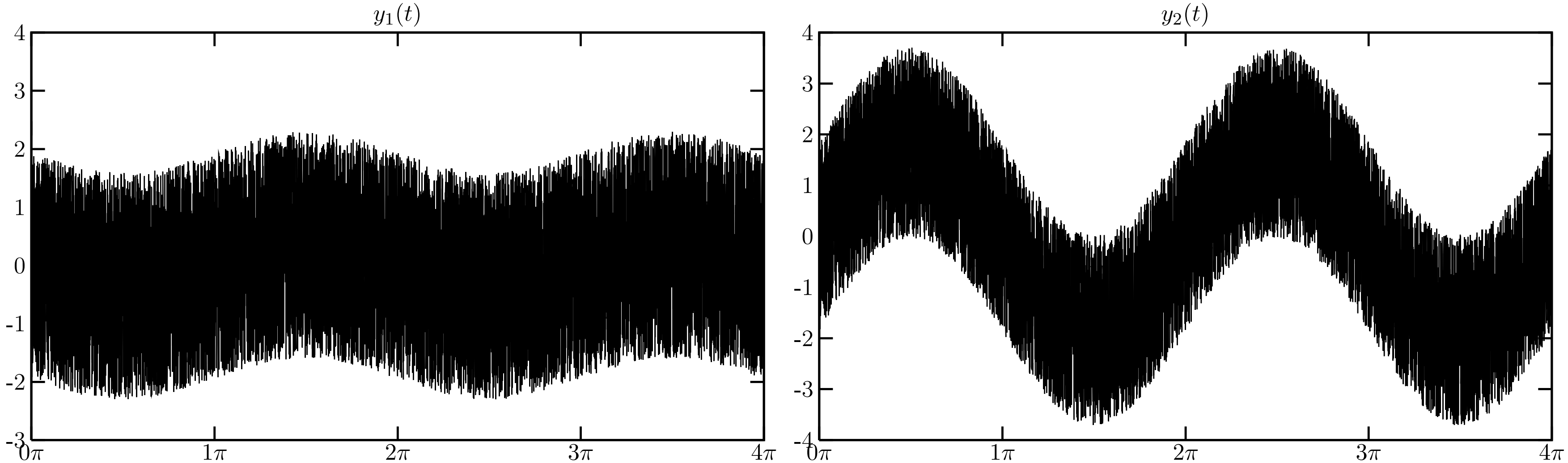


**Fig. 2.** Observations of linearly mixed signals.

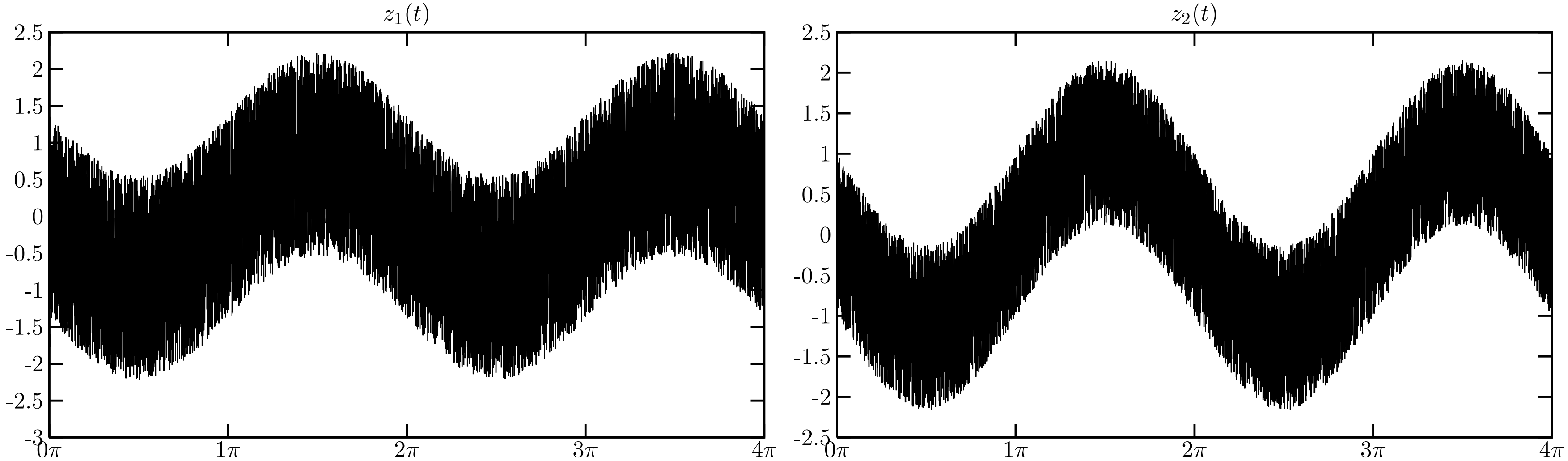


FIG. 3. Seperation results produced by the PCA.

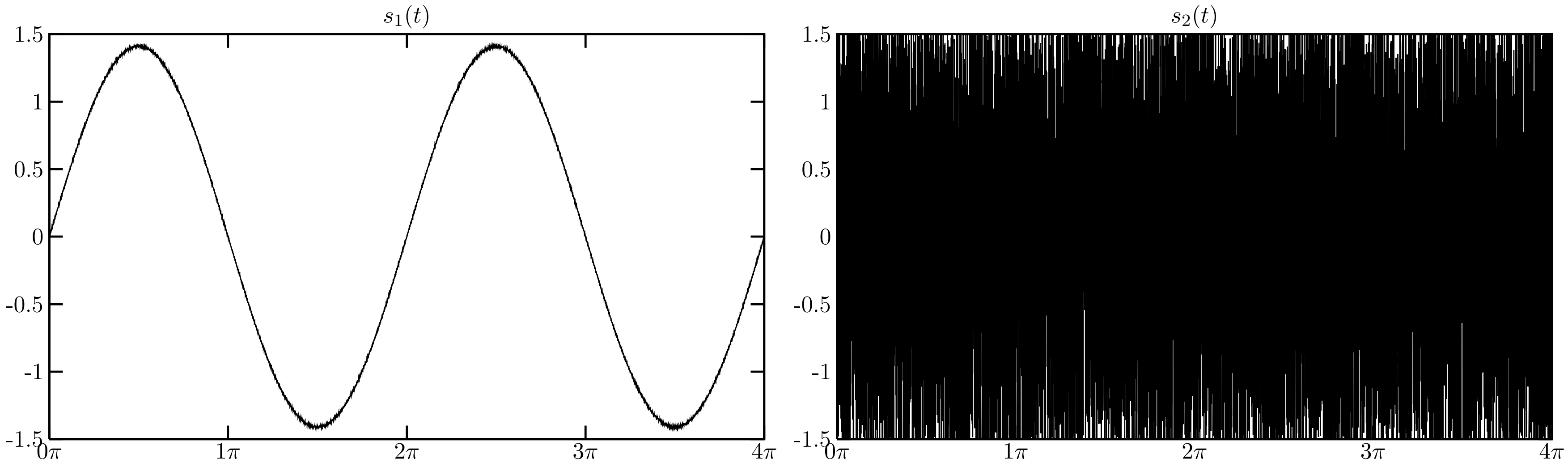


FIG. 4. Separation results of the tensor technique.

## KRUSKAL (ESSENTIAL) UNIQUENESS

Canonical decomposition

$$\boldsymbol{A} = \sum_{t=1}^{\rho} \boldsymbol{u}_t \otimes \boldsymbol{v}_t \otimes \boldsymbol{w}_t$$

is defined by matrices with $\boldsymbol{\rho}$ columns:

$$\boldsymbol{U} = [\boldsymbol{u}_{it}], \qquad \boldsymbol{V} = [\boldsymbol{v}_{jt}], \qquad \boldsymbol{W} = [\boldsymbol{w}_{kt}].$$

A matrix is said to have *Kruskal rank* $\boldsymbol{r}$ if
$\boldsymbol{r}$ is the maximal number s.t. any $\boldsymbol{r}$ columns are linearly independent.

## KRUSKAL THEOREM

Let the Kruskal ranks for $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ coincide with their ranks and

$$\mathbf{rank}\boldsymbol{U} + \mathbf{rank}\boldsymbol{V} + \mathbf{rank}\boldsymbol{W} \geq 2\boldsymbol{\rho} + 2.$$

Then this canonical decomposition is essentially unique.

J. B. Kruskal, Three-way arrays: rank and uniqueness for 3-way and n-way arrays, *Linear Algebra Appl.*, 18, pp. 95–138 (1977).

## SIMULTANEOUS DIAGONALIZATION

Tensor decomposition of an $\boldsymbol{n} \times \boldsymbol{n} \times \boldsymbol{n}$ tensor

$$\boldsymbol{a}_{ijk} = \sum_{t=1}^{\rho} \boldsymbol{u}_{it}\, \boldsymbol{v}_{jt}\, \boldsymbol{w}_{kt}$$

means simultaneous diagonalization of $\boldsymbol{n}$ slice matrices

$$\boldsymbol{M}_k = [\boldsymbol{a}_{ijk}] = \boldsymbol{U} \begin{bmatrix} \boldsymbol{w}_{k1} & & \\ & \ddots & \\ & & \boldsymbol{w}_{k\rho} \end{bmatrix} \boldsymbol{V}^\top$$

$\boldsymbol{U}$ and $\boldsymbol{V}$ are $\boldsymbol{n} \times \boldsymbol{\rho}$

## RELATED SIMULTANEOUS EIGENVALUE PROBLEM

$$\boldsymbol{M}_k \boldsymbol{x} = \boldsymbol{\lambda}_k \boldsymbol{y}$$

# $\mathbf{2 \times 2 \times 2}$ TENSORS

When tensor rank is equal to 2?
If so, we have simultaneously

$$\boldsymbol{M_1} = \boldsymbol{U W_1 V}$$
$$\boldsymbol{M_2} = \boldsymbol{U W_2 V}$$

If $\boldsymbol{M_2}$ is nonsingular, it follows that

$$\boldsymbol{M_1 M_2^{-1}} = \boldsymbol{U D U^{-1}}, \qquad \boldsymbol{D} = \boldsymbol{W_1 W_2^{-1}} \text{ is a diagonal matrix.}$$

## EXAMPLE

$$\boldsymbol{M_1} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \boldsymbol{M_2} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Tensor with slices $\boldsymbol{M_1}, \boldsymbol{M_2}$ must be of tensor rank $\geq 3$.

## COROLLARY

Tensor rank for a tensor of size $\mathbf{2 \times 2 \times 2}$ can be greater than $\mathbf{2}$.

It cannot exceed 4, but can it be greater than 3?

## PRESERVATION OF TENSOR RANK

### LEMMA

Tensor rank is invariant under mode contractions by nonsingular matrices.

### COROLLARY 1.

Tensor rank calculation for general $\mathbf{2} \times \mathbf{2} \times \mathbf{2}$ tensor reduces to a particular case of tensor with slices

$$\boldsymbol{M}_1 = \begin{bmatrix} * & \mathbf{0} \\ \mathbf{0} & * \end{bmatrix}, \qquad \boldsymbol{M}_2 = \begin{bmatrix} * & * \\ * & * \end{bmatrix}.$$

### COROLLARY 2.

Maximum of tensor ranks for $\mathbf{2} \times \mathbf{2} \times \mathbf{2}$ tensors is equal to 3.

# TENSOR RANK DEPENDS ON THE SUBFIELD

It does not happen for matrices!
However, for tensor over $\mathbb{R}$ tensor ranks over $\mathbb{R}$ and $\mathbb{C}$ may differ.

PROOF.
Consider $\mathbf{2} \times \mathbf{2} \times \mathbf{2}$ tensor with slices

$$\boldsymbol{M_1} = \begin{bmatrix} \mathbf{0} & \mathbf{-1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}, \qquad \boldsymbol{M_2} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}.$$

Matrix $\boldsymbol{M_1 M_2^{-1}}$ has eigenvalues $\pm\sqrt{\mathbf{-1}}$.

Hence, it cannot be diagonalized by a real similarity $\Rightarrow$
tensor rank over $\mathbb{R}$ is equal to 3.

But tensor rank over $\mathbb{C}$ is 2.

## RANK INSTABILITY

- Matrix rank can be made larger by arbitrarily small perturbation, but cannot be made smaller.
  The same for Tucker ranks.
- Tensor (canonical) rank ***may decrease*** by an arbitrary small perturbation, at least for some tensors.

**EXAMPLE** (could be 3D Laplacian)

$$\boldsymbol{a}\otimes\boldsymbol{a}\otimes\boldsymbol{b}+\boldsymbol{a}\otimes\boldsymbol{b}\otimes\boldsymbol{a}+\boldsymbol{b}\otimes\boldsymbol{a}\otimes\boldsymbol{a}=$$

$$\boldsymbol{a}\otimes(\boldsymbol{a}+\varepsilon\boldsymbol{b})\otimes(\boldsymbol{b}+\varepsilon^{-1}\boldsymbol{a})+(\boldsymbol{b}-\varepsilon^{-1}\boldsymbol{a})\times\boldsymbol{a}\otimes\boldsymbol{a}$$
$$+\varepsilon\boldsymbol{a}\otimes\boldsymbol{b}\otimes\boldsymbol{b}$$

Notice large numbers in a lower-rank tensor.

**DIFFICULTY**
Best approximation to a given tensor by tensors of a prescribed tensor rank ***may not exist***.

# BEST TENSOR APPROXIMATIONS

## THEOREM 1.

For a tensor of canonical rank $\boldsymbol{\rho}$,
best approximations of rank $\mathbf{1}$ and rank $\boldsymbol{\rho}$ always exist.

*Is it possible to produce an example of tensor with non-existence of best approximations of any rank strictly in between of $\mathbf{1}$ and $\boldsymbol{\rho}$?*

## THEOREM 2.

Best approximations of a prescribed tensor rank and a predetermined upper bound on moduli of the factor entries always exist.

## THEOREM 3.

Best approximations of a prescribed tensor rank with nonnegativity constraint for all entries of factors always exist.

## GENERIC RANKS

A minimal finite set $\mathcal{R}(n_1, ..., n_d) = \{r_s\}$ of positive integers s.t. tensors with tensor ranks from this set are dense in the set of all tensor of size

$$n_1 \times ... \times n_d$$

is said to consist of ***generic ranks*** for $n_1 \times ... \times n_d$ tensors.

Real $\mathbf{2 \times 2 \times 2}$ tensors has generic ranks 2 and 3.

$\mathbf{2}$ in $\sim \mathbf{79\%}$ and $\mathbf{3}$ in $\sim \mathbf{21\%}$ cases.

The set of complex $\mathbf{2 \times 2 \times 2}$ tensors has generic rank 2.

**THEOREM (HYPOTHESIS?)**
For complex tensors there is a single value of generic rank (depending on size).

**HYPOTHESIS (THEOREM?)**
For real tensors there could be onle two possible generic ranks (depending on size).

# ALTERNATING LEAST SQUARES

R. A. Harshman,
Foundations of the Parafac procedure: models and conditions for an explanatory multimodal factor analysis, *UCLA Working Papers in Phonetics*, 16: 1–84 (1970).

`Given` $\boldsymbol{A}$`, find an optimal canonical decomposition`
`with factor matrices` $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$
`with prescribed number of columns.`

ALS reads

- Freeze $\boldsymbol{V}, \boldsymbol{W}$ and substitute $\boldsymbol{U}$ with the best LS fit.
- Freeze $\boldsymbol{U}, \boldsymbol{W}$ and do the same with $\boldsymbol{V}$.
- Freeze $\boldsymbol{U}, \boldsymbol{V}$ and do the same with $\boldsymbol{W}$.
- Repeat until convergence.

Convergence theory?

## ONE STEP OF ALS

With $\boldsymbol{U}$ and $\boldsymbol{V}$ frozen, find $\boldsymbol{W}$ from the LS problem

$$\sum_{t=1}^{\rho} u_{it}\, v_{jt}\, w_{kt} \overset{\text{LS}}{=} a_{ijk}.$$

In the matrix form, find vectors of size $\rho$

$$x_k = \begin{bmatrix} w_{k1} \\ \cdots \\ w_{k\rho} \end{bmatrix}$$

solving

$$U \square V \overset{\text{LS}}{=} b_k \equiv \begin{bmatrix} a_{11k} \\ a_{12k} \\ \cdots \\ a_{nnk} \end{bmatrix}$$

$$U \square V = [u_1 \otimes v_1,\ \ldots,\ u_\rho \otimes v_\rho] \qquad \text{(Khartri-Rao product)}$$