Tensor-Based Dictionary Learning for Multidimensional Sparse Recovery

Florian Römer and Giovanni Del Galdo

Ilmenau University of Technology
Institute for Information Technology (InIT)

Contact Information

- Prof. Giovanni Del Galdo

Head of DVT Research Group, composed of
Digital Broadcasting Research Laboratory,
Ilmenau University of Technology
Wireless Distribution Systems / Digital Broadcasting Group,
Fraunhofer Institute for Integrated Circuits IIS

office: +49 3677 69-4280
mobile: +49 151 62 91 01 81
e-mail: giovanni.delgaldo@tu-ilmenau.de
giovanni.delgaldo@iis.fraunhofer.de

- Dr.-Ing. Florian Römer

Digital Broadcasting Research Laboratory,
Ilmenau University of Technology

office: +49 3677 69 4286
e-mail: florian.roemer@tu-ilmenau.de
florian.roemer@iis-extern.fraunhofer.de

Motivation

- Compressed Sensing
  - allows lossless acquisition of signals at a sampling rate below Nyquist
  - perfect knowledge of the dictionary is usually assumed
  - in practice, the dictionary may be unknown
    - partially (e.g., calibration, model mismatch, grid offset)
    - completely (e.g., "typical" structures in image coding)

  Dictionary learning
- State of the art
  - two of the most popular fundamental approaches: MOD [EAH00], K-SVD [AEB06]
  - many variations, e.g., SimCO, RLS-DA, discriminative K-SVD, …

[EAH00] K. Engan, S. O. Aase, and J. H. Husøy, "Multi-frame compression: Theory and design," *EURASIP Signal Process.*, vol. 80, no. 10, pp. 2121–2140, 2000.

[AEB06] M. Aharon, M. Elad, and A. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

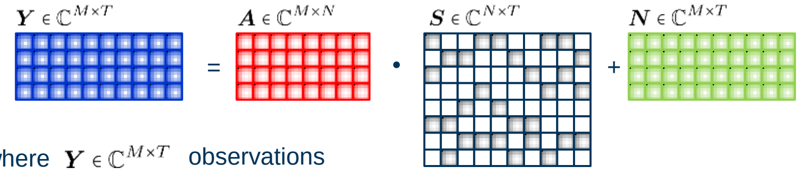
- Our contribution
  - tensors formulation for separable 2-D sparse recovery problems
  - extension of MOD and K-SVD, improved algorithms

Outline

- Motivation
- Data model
  - 2-D extension
- Dictionary learning algorithms
  - MOD → T-MOD
  - K-SVD → K-HOSVD
- Numerical Results
- Conclusions

Data model: 1-D sparse recovery problem

- 1-D noisy sparse recovery problem $\boldsymbol{Y} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N}$



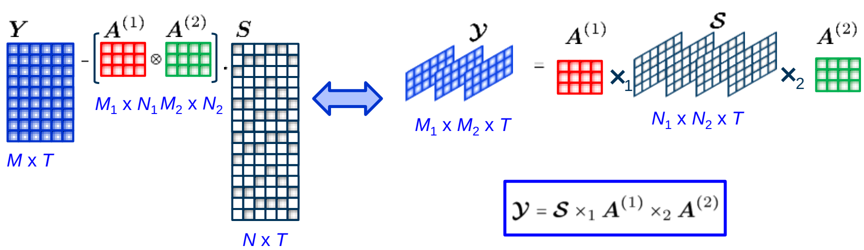
where $\boldsymbol{Y} \in \mathbb{C}^{M \times T}$ observations
$\boldsymbol{A} \in \mathbb{C}^{M \times N}$ overcomplete dictionary $M < N < T$
$\boldsymbol{S} \in \mathbb{C}^{N \times T}$ sparse coefficient matrix $\boldsymbol{S} = [\boldsymbol{s}_1, \quad \boldsymbol{s}_2, \quad \ldots, \boldsymbol{s}_T], \; \|\boldsymbol{s}_t\|_0 \leq K \; \forall t$
$\boldsymbol{N} \in \mathbb{C}^{M \times T}$ additive measurement noise

- 2-D sparse recovery problem
  - if the manifold is separable (e.g., 2-D DOAE with a uniform rectangular array or joint DOA/DOD) and a separable 2-D sampling grid is chosen, then

$$\boldsymbol{A} = \boldsymbol{A}^{(1)} \otimes \boldsymbol{A}^{(2)} \qquad \begin{aligned} \boldsymbol{A}^{(1)} &= \left[\boldsymbol{a}_1^{(1)}, \ldots, \boldsymbol{a}_{N_1}^{(1)}\right] \in \mathbb{C}^{M_1 \times N_1} & M &= M_1 \cdot M_2 \\ \boldsymbol{A}^{(2)} &= \left[\boldsymbol{a}_1^{(2)}, \ldots, \boldsymbol{a}_{N_2}^{(2)}\right] \in \mathbb{C}^{M_2 \times N_2} & N &= N_1 \cdot N_2 \end{aligned}$$

Multilinear structure

- The 2-D model possesses an interesting multilinear structure



$$\boxed{\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)}}$$

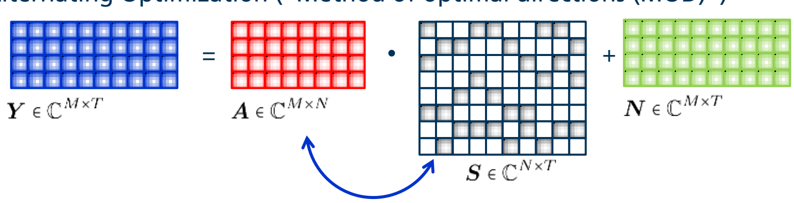
⇒ Tucker-2 decomposition with a sparse core tensor

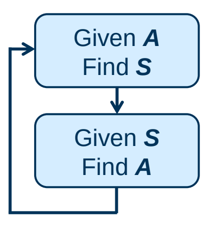
Outline

- Motivation
- Data model
  - 2-D extension
- Dictionary learning algorithms
  - MOD → T-MOD
  - K-SVD → K-HOSVD
- Numerical Results
- Conclusions

Dictionary Learning via MOD

- Alternating Optimization (“Method of optimal directions (MOD)”)





$$\min_{\boldsymbol{S}} \|\boldsymbol{Y} - \boldsymbol{A} \cdot \boldsymbol{S}\|_F^2 \text{ s.t. } \|\boldsymbol{s}_t\|_0 \le K, t = 1, 2, \ldots, T$$

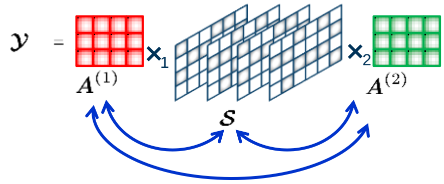
noisy sparse recovery problem

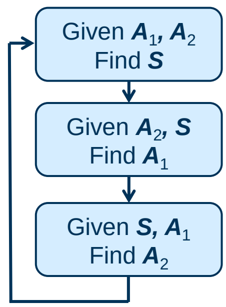
$$\min_{\boldsymbol{A}} \|\boldsymbol{Y} - \boldsymbol{A} \cdot \boldsymbol{S}\|_F^2 \text{ s.t. } \|\boldsymbol{a}_n\|_2 = 1, n = 1, 2, \ldots, N$$

linear least squares problem

MOD in 2-D: The Tensor-MOD procedure

- Alternating Optimization ("Method of optimal directions (MOD)")





$$\min_{S} \left\| Y - \left(A^{(1)} \otimes A^{(2)}\right) \cdot S \right\|_F^2 \ \text{s.t.}\ \|s_t\|_0 \le K, t = 1, 2, \ldots, T$$

noisy sparse recovery problem

$$\min_{A^{(1)}} \left\| \mathcal{Y} - \mathcal{S} \times_1 A^{(1)} \times_2 A^{(2)} \right\|_F^2 \ \text{s.t.}\ \left\|a_{n_1}^{(1)}\right\|_2 = 1, n_1 = 1, 2, \ldots, N_1$$

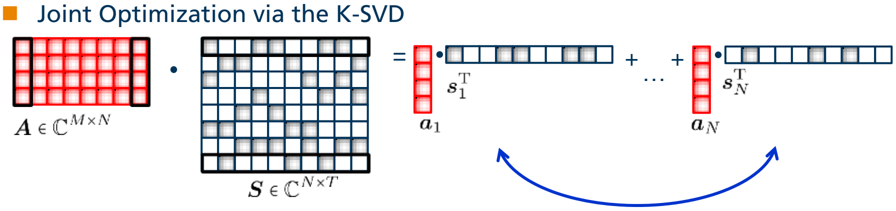
linear least squares problem

$$\min_{A^{(2)}} \left\| \mathcal{Y} - \mathcal{S} \times_1 A^{(1)} \times_2 A^{(2)} \right\|_F^2 \ \text{s.t.}\ \left\|a_{n_2}^{(2)}\right\|_2 = 1, n_2 = 1, 2, \ldots, N_2$$

linear least squares problem

Dictionary Learning via the K-SVD

- Joint Optimization via the K-SVD



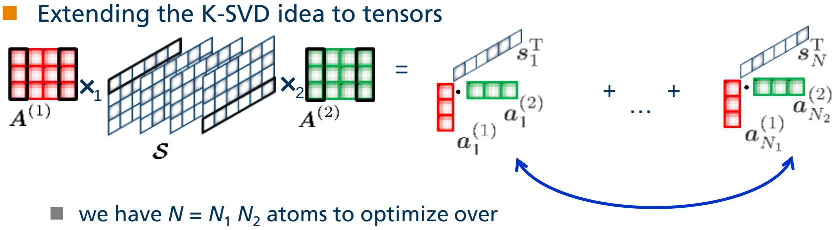
- instead of iterating between A and S, iterate between atoms and update the m-th column of A and the m-th row of S *jointly*

$$\min_{a_n, s_n^T} \|Y - A \cdot S\|_F^2 = \min_{a_n, s_n^T} \Bigg\| \underbrace{Y - \sum_{p=1, p\neq n}^{N} a_p \cdot s_p^T}_{Y_n} - a_n \cdot s_n^T \Bigg\|_F^2 = \min_{a_n, s_n^T} \|Y_n - a_n \cdot s_n^T\|_F^2$$

- rank-one approximation problem: closed-form solution given by truncated SVD of Y_n

K-SVD in 2-D: The K–“Higher-Order SVD” (K-HOSVD)

- Extending the K-SVD idea to tensors



  - we have $N = N_1\, N_2$ atoms to optimize over

$$\min_{\boldsymbol{a}_{n_1}^{(1)}, \boldsymbol{a}_{n_2}^{(2)}, \boldsymbol{s}_n^{\mathrm{T}}} \left\| \mathcal{Y} - \mathcal{S} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \right\|_{\mathrm{F}}^2 = \min_{\boldsymbol{a}_{n_1}^{(1)}, \boldsymbol{a}_{n_2}^{(2)}, \boldsymbol{s}_n^{\mathrm{T}}} \left\| \mathcal{Y}_{n_1, n_2} - \boldsymbol{a}_{n_1}^{(1)} \circ \boldsymbol{a}_{n_2}^{(2)} \circ \boldsymbol{s}_n \right\|_{\mathrm{F}}^2$$

$n = (n_1 - 1) \cdot N_2 + n_2$

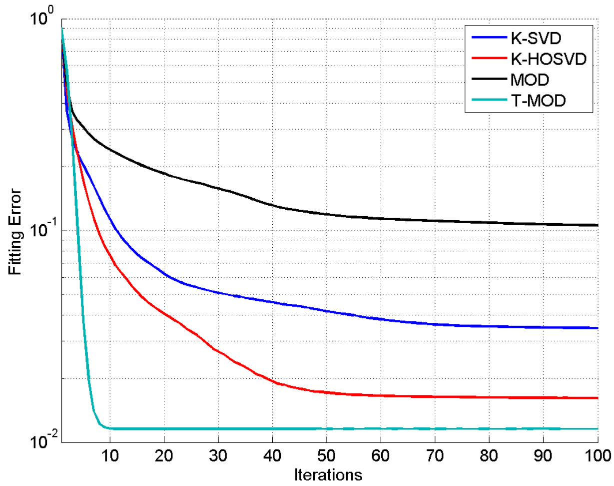
  - rank-one tensor approximation problem
    - Eckart-Young theorem does not generalize to tensors
      - LS-optimal: no closed-form solution, iterative scheme required [LMV00]
      - truncated Higher-Order SVD: closed-form, very close to LS-optimal

[LMV00] L. de Lathauwer, B. de Moor, and J. Vandewalle, “On the best rank-1 and rank-(r_1, r_2, …, r_n) approximation of higher-order tensors,” SIAM J. Matrix Anal. Appl., vol. 21, p. 1324-1342, 2000.

Outline

- Motivation
- Data model
  - 2-D extension
- Dictionary learning algorithms
  - MOD → T-MOD
  - K-SVD → K-HOSVD
- Numerical Results
- Conclusions

Random dictionary, known support

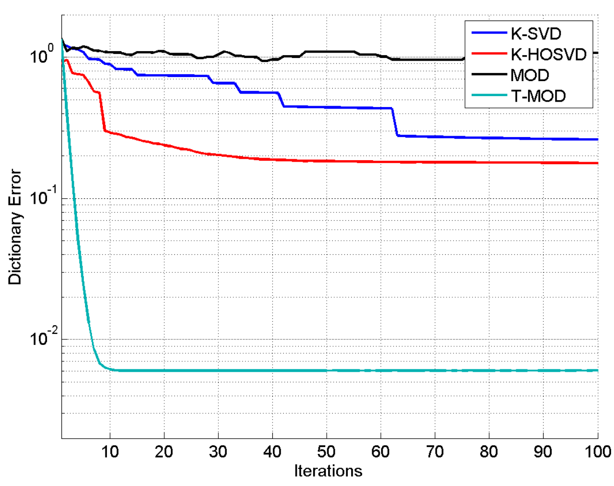


M = 4 x 4
N = 6 x 6
K = 3
T = 100
$P_N = 10^{-4}$
$\boldsymbol{N} \sim \text{i.i.d. } \mathcal{N}(0, P_N)$

$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)} \sim \text{i.i.d. } \mathcal{N}(0,1)$
$\hat{\boldsymbol{A}}_0^{(1)}, \hat{\boldsymbol{A}}_0^{(2)} \sim \text{i.i.d. } \mathcal{N}(0,1)$

$$\text{FE} = \frac{\|\hat{\boldsymbol{A}} \cdot \hat{\boldsymbol{S}} - \boldsymbol{A} \cdot \boldsymbol{S}\|_F}{\|\boldsymbol{A} \cdot \boldsymbol{S}\|_F}$$

Random dictionary, known support

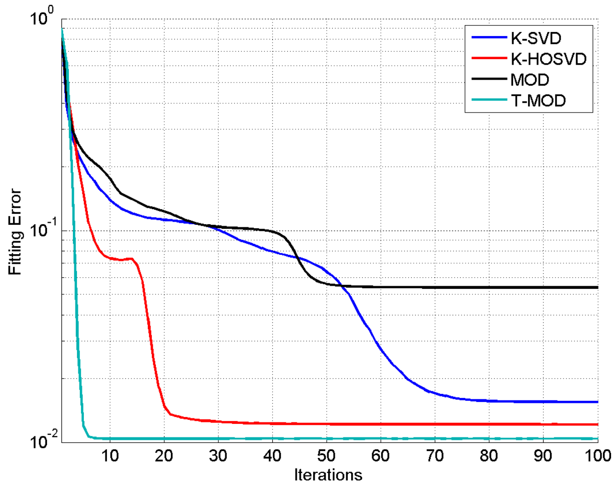


M = 4 x 4
N = 6 x 6
K = 3
T = 100
$P_N = 10^{-4}$
$\boldsymbol{N} \sim \text{i.i.d. } \mathcal{N}(0, P_N)$

$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)} \sim \text{i.i.d. } \mathcal{N}(0,1)$
$\hat{\boldsymbol{A}}_0^{(1)}, \hat{\boldsymbol{A}}_0^{(2)} \sim \text{i.i.d. } \mathcal{N}(0,1)$

$$\text{DE} = \frac{\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_F}{\|\boldsymbol{A}\|_F}$$

DCT dictionary, known support

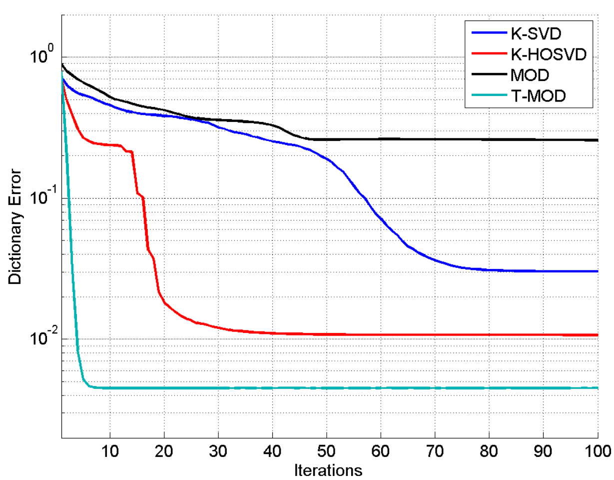


M = 4 x 4
N = 6 x 6
K = 3
T = 100
$P_N = 10^{-4}$
$\boldsymbol{N} \sim \text{i.i.d.}\ \mathcal{N}(0, P_{\text{N}})$

$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}: \text{DCT}$
$\hat{\boldsymbol{A}}_0^{(1)}, \hat{\boldsymbol{A}}_0^{(2)} \sim \text{i.i.d.}\ \mathcal{N}(0,1)$

$$\text{FE} = \frac{\|\hat{\boldsymbol{A}} \cdot \hat{\boldsymbol{S}} - \boldsymbol{A} \cdot \boldsymbol{S}\|_{\text{F}}}{\|\boldsymbol{A} \cdot \boldsymbol{S}\|_{\text{F}}}$$

DCT dictionary, known support

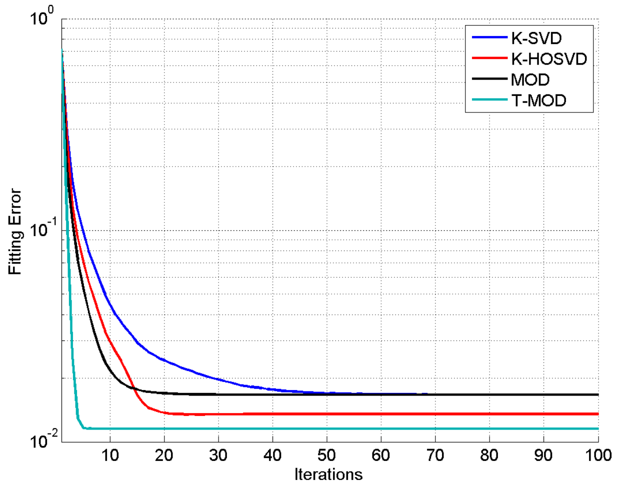


M = 4 x 4
N = 6 x 6
K = 3
T = 100
$P_N = 10^{-4}$
$\boldsymbol{N} \sim \text{i.i.d.}\ \mathcal{N}(0, P_{\text{N}})$

$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}: \text{DCT}$
$\hat{\boldsymbol{A}}_0^{(1)}, \hat{\boldsymbol{A}}_0^{(2)} \sim \text{i.i.d.}\ \mathcal{N}(0,1)$

$$\text{DE} = \frac{\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_{\text{F}}}{\|\boldsymbol{A}\|_{\text{F}}}$$

DCT dictionary, known support



M = 4 x 4
N = 6 x 6
K = 3
T = 100
$P_N = 10^{-4}$
$\boldsymbol{N} \sim \text{i.i.d. } \mathcal{N}(0, P_\mathrm{N})$

$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)} : \mathrm{DCT}$

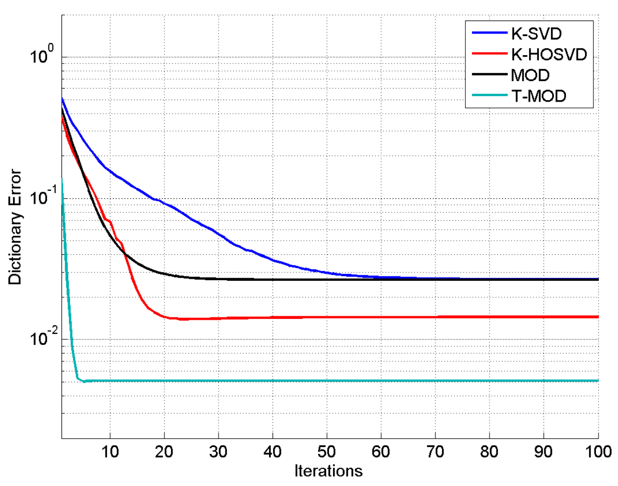
$\hat{\boldsymbol{A}}_0^{(r)} = \boldsymbol{A}_0^{(r)} + \boldsymbol{E}^{(r)}, \ r = 1, 2$

$\boldsymbol{E}^{(r)} \sim \text{i.i.d. } \mathcal{N}(0, P_\mathrm{E})$

$P_E = 10^{-1}$

$$\mathrm{FE} = \frac{\|\hat{\boldsymbol{A}} \cdot \hat{\boldsymbol{S}} - \boldsymbol{A} \cdot \boldsymbol{S}\|_\mathrm{F}}{\|\boldsymbol{A} \cdot \boldsymbol{S}\|_\mathrm{F}}$$

DCT dictionary, known support



M = 4 x 4
N = 6 x 6
K = 3
T = 100
$P_N = 10^{-4}$
$\boldsymbol{N} \sim \text{i.i.d. } \mathcal{N}(0, P_\mathrm{N})$

$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)} : \mathrm{DCT}$

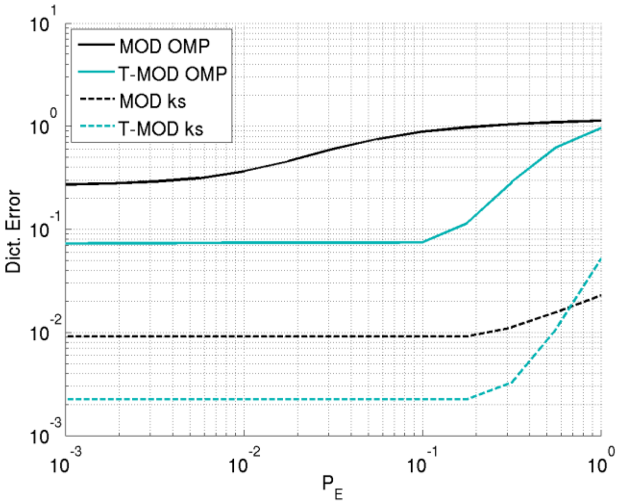
$\hat{\boldsymbol{A}}_0^{(r)} = \boldsymbol{A}_0^{(r)} + \boldsymbol{E}^{(r)}, \ r = 1, 2$

$\boldsymbol{E}^{(r)} \sim \text{i.i.d. } \mathcal{N}(0, P_\mathrm{E})$

$P_E = 10^{-1}$

$$\mathrm{DE} = \frac{\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_\mathrm{F}}{\|\boldsymbol{A}\|_\mathrm{F}}$$

Random dictionary, estimated support



M = 4 x 4
N = 6 x 6
K = 3
T = 300
$P_N = 10^{-4}$

$\boldsymbol{N} \sim \text{i.i.d. } \mathcal{N}(0, P_\text{N})$

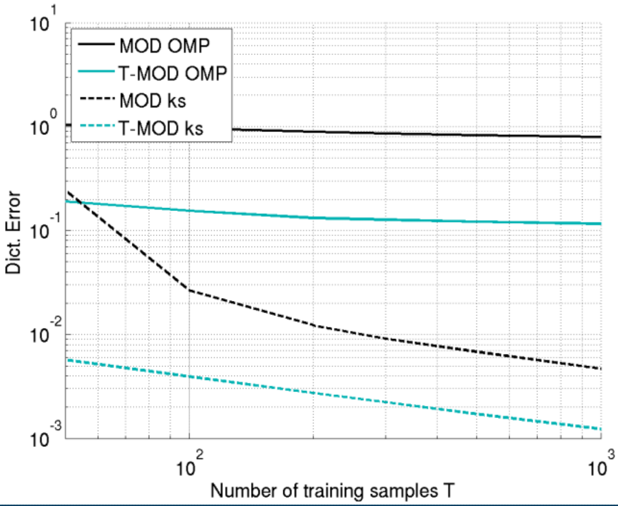
$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)} \sim \text{i.i.d. } \mathcal{N}(0,1)$

$\hat{\boldsymbol{A}}_0^{(r)} = \boldsymbol{A}_0^{(r)} + \boldsymbol{E}^{(r)}, \ r = 1,2$

$\boldsymbol{E}^{(r)} \sim \text{i.i.d. } \mathcal{N}(0, P_\text{E})$

$$\text{DE} = \frac{\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_\text{F}}{\|\boldsymbol{A}\|_\text{F}}$$

Random dictionary, estimated support



M = 4 x 4
N = 6 x 6
K = 3
P_E = 0.1
$P_N = 10^{-4}$

$\boldsymbol{N} \sim \text{i.i.d. } \mathcal{N}(0, P_\text{N})$

$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)} \sim \text{i.i.d. } \mathcal{N}(0,1)$

$\hat{\boldsymbol{A}}_0^{(r)} = \boldsymbol{A}_0^{(r)} + \boldsymbol{E}^{(r)}, \ r = 1,2$

$\boldsymbol{E}^{(r)} \sim \text{i.i.d. } \mathcal{N}(0, P_\text{E})$

$$\text{DE} = \frac{\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_\text{F}}{\|\boldsymbol{A}\|_\text{F}}$$

Conclusions

- 2-D sparse recovery problems on a separable manifold
  - Dictionary has a Kronecker structure
  - observation model can be expressed in tensor form
    - Tucker-2 with sparse core

- dictionary learning algorithms
  - tensor structure can be efficiently exploited
    - improved estimation accuracy
  - Demonstrated using two prominent examples
    - Method of Optimal Direction (MOD) → Tensor-MOD
    - K-SVD → K-Higher Order SVD
  - improved accuracy shown numerically