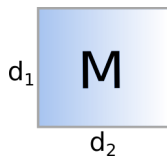


Higher-Order Low-Rank Regression

Guillaume Rabusseau, Hachem Kadri
McGill - RLLab / LIF - Qarma

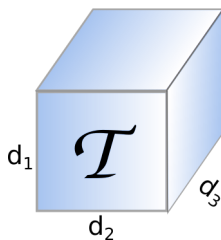
March 14, 2019

Tensors



$$\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}$$

$$\mathbf{M}_{ij} \in \mathbb{R} \text{ for } i \in [d_1], j \in [d_2]$$

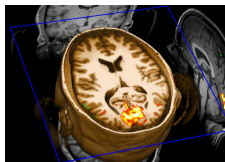
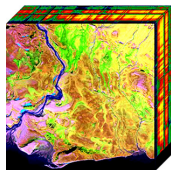


$$\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$$

$$(\mathcal{T}_{ijk}) \in \mathbb{R} \text{ for } i \in [d_1], j \in [d_2], k \in [d_3]$$

Tensors and Machine Learning

- (i) **Data** has a tensor structure: color image, video, multivariate time series...



- (ii) Tensor as **parameters** of a model: weighted tree automata, factorization machines...

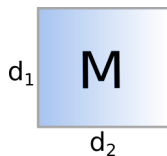
$$f(\mathbf{x}) = \sum_{i,j,k} \mathcal{W}_{i,j,k} \mathbf{x}_i \mathbf{x}_j \mathbf{x}_k$$

- (iii) Tensors as **tools**: tensor method of moments, system of polynomial equations, layer compression in neural networks...

Outline

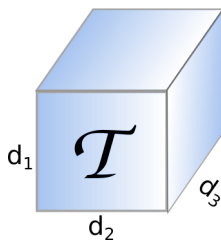
- 1 Introduction
 - Tensors
- 2 Low-Rank Regression with Tensor Responses

Tensors



$$\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}$$

$$\mathbf{M}_{ij} \in \mathbb{R} \text{ for } i \in [d_1], j \in [d_2]$$



$$\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$$

$$(\mathcal{T}_{ijk}) \in \mathbb{R} \text{ for } i \in [d_1], j \in [d_2], k \in [d_3]$$

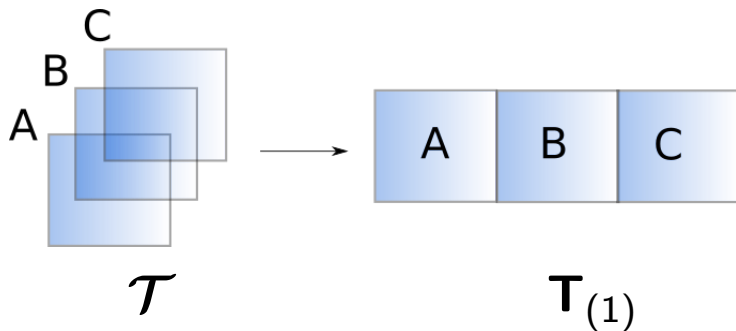
Tensors: Matricizations

- $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ can be reshaped into a matrix as

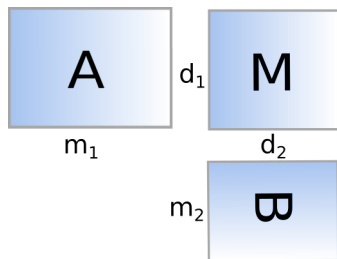
$$\mathbf{T}_{(1)} \in \mathbb{R}^{d_1 \times d_2 d_3}$$

$$\mathbf{T}_{(2)} \in \mathbb{R}^{d_2 \times d_1 d_3}$$

$$\mathbf{T}_{(3)} \in \mathbb{R}^{d_3 \times d_1 d_2}$$

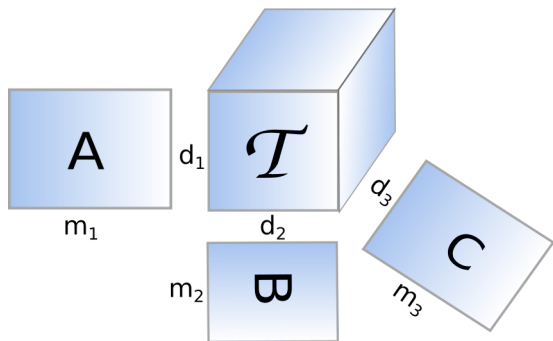


Tensors: Multiplication with Matrices



$$\mathbf{AMB}^T \in \mathbb{R}^{m_1 \times m_2}$$

Tensors: Multiplication with Matrices



$$\mathcal{T} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$
$$\in \mathbb{R}^{m_1 \times m_2 \times m_3}$$

- For vectors, we note $\mathcal{T} \bullet_n \mathbf{v} = \mathcal{T} \times_n \mathbf{v}^T$

Tucker Decomposition and Multilinear Rank

- Multilinear rank:

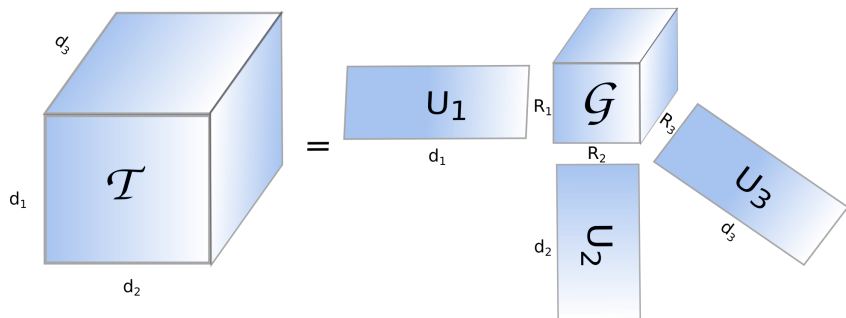
$$\text{rank}_{ml}(\mathcal{T}) = (R_1, R_2, R_3) \Leftrightarrow R_i = \text{rank}(\mathbf{T}_{(i)}) \text{ for } i = 1, 2, 3$$

Tucker Decomposition and Multilinear Rank

- Multilinear rank:

$$\text{rank}_{ml}(\mathcal{T}) = (R_1, R_2, R_3) \Leftrightarrow R_i = \text{rank}(\mathbf{T}_{(i)}) \text{ for } i = 1, 2, 3$$

- Tucker decomposition: $\mathcal{T} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ (*)
with $\mathbf{U}_i^\top \mathbf{U}_i = \mathbf{I}$ for all i .

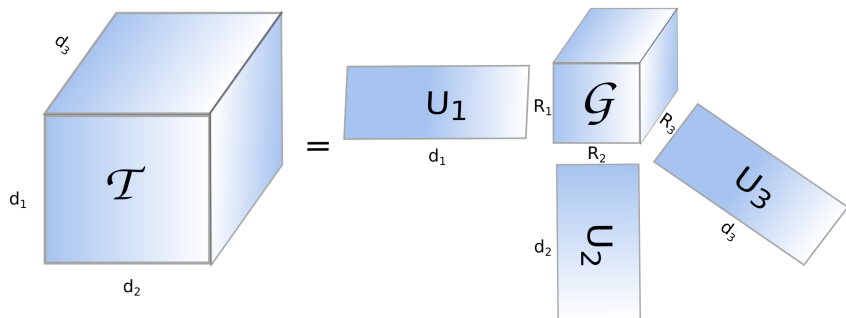


Tucker Decomposition and Multilinear Rank

- Multilinear rank:

$$\text{rank}_{ml}(\mathcal{T}) = (R_1, R_2, R_3) \Leftrightarrow R_i = \text{rank}(\mathbf{T}_{(i)}) \text{ for } i = 1, 2, 3$$

- Tucker decomposition: $\mathcal{T} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ (*)
with $\mathbf{U}_i^\top \mathbf{U}_i = \mathbf{I}$ for all i .



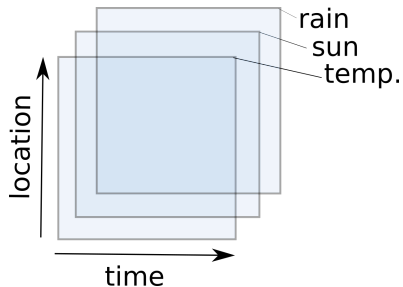
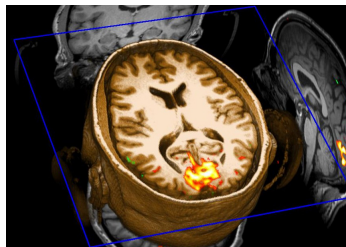
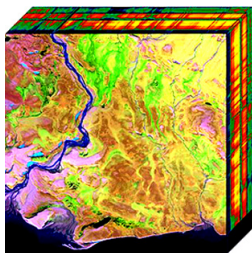
- Multilinear rank = the smallest (R_1, R_2, R_3) such that (*) holds.

Outline

- 1 Introduction
- 2 Low-Rank Regression with Tensor Responses
 - Preliminaries
 - Higher-Order Low-Rank Regression
 - Function with Low Multilinear Rank
 - Theoretical Guarantees
 - Experiments
 - Conclusion and Future Works

Introduction

- Data with tensor structure: EEG, hyperspectral images, videos, ...



Problem

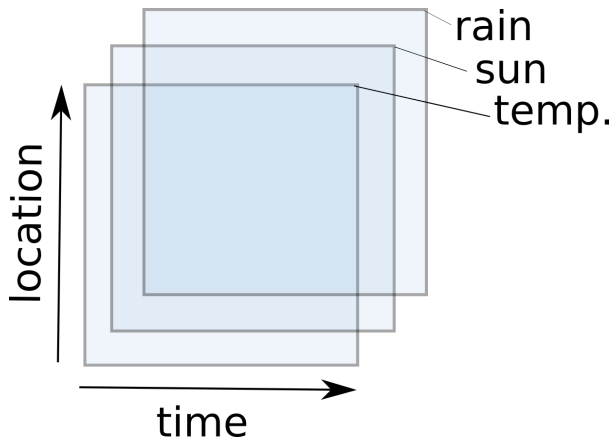
Learn $f : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_1 \times \dots \times d_p}$ from $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ where $\mathbf{y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.

Problem

Learn $f : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_1 \times \dots \times d_p}$ from $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ where $\mathbf{y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.

- Spatio-Temporal Forecasting (Bahadori et al., 2014)

$$f(\mathbf{x}) \in \mathbb{R}^{(\text{Times}) \times (\text{Locations}) \times (\text{Variables})}$$



Problem

Learn $f : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_1 \times \dots \times d_p}$ from $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ where $\mathbf{y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.

- Multilinear Multitask Learning (Romera-Paredes et al., 2013)

$$f(\mathbf{x}) \in \mathbb{R}^{(\text{Restaurant Critics}) \times (\text{Evaluation Criteria})}$$

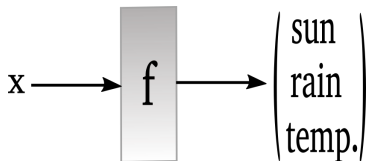
| Rest. 1 | Critic 1 | Critic 2 | Critic 3 |
|---------|----------|----------|----------|
| host | 5 | 3 | 6 |
| food | 7 | 8 | 6.5 |
| price | 5 | 6.5 | 4 |

| Rest. 2 | Critic 1 | Critic 2 | Critic 3 |
|---------|----------|----------|----------|
| host | 7 | 8 | 6 |
| food | 8.5 | 9 | 9 |
| price | 8 | 9.5 | 7 |

...

Multivariate Regression

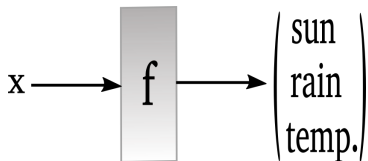
Learn $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ from samples $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ where $\mathbf{y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.



- Linear model: $f(\mathbf{x}) = \mathbf{W}^\top \mathbf{x}$ ($\mathbf{W} \in \mathbb{R}^{d \times p}$)

Multivariate Regression

Learn $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ from samples $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ where $\mathbf{y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.

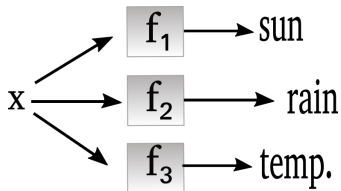


- Linear model: $f(\mathbf{x}) = \mathbf{W}^\top \mathbf{x}$ ($\mathbf{W} \in \mathbb{R}^{d \times p}$)
- Ordinary Least Squares

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W} \in \mathbb{R}^{d \times p}} \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|_F^2 \quad (\mathbf{X} \in \mathbb{R}^{N \times d}, \mathbf{Y} \in \mathbb{R}^{N \times p})$$

Multivariate Regression

Learn $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ from samples $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ where $\mathbf{y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.



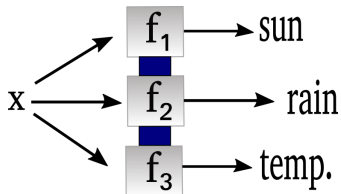
- Linear model: $f(\mathbf{x}) = \mathbf{W}^\top \mathbf{x}$ ($\mathbf{W} \in \mathbb{R}^{d \times p}$)
- Ordinary Least Squares

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W} \in \mathbb{R}^{d \times p}} \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|_F^2 \quad (\mathbf{X} \in \mathbb{R}^{N \times d}, \mathbf{Y} \in \mathbb{R}^{N \times p})$$

⇒ Equivalent to **perform p independent linear regressions!**
How can we capture linear dependencies in the output?

Multivariate Regression

Learn $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ from samples $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$ where $\mathbf{y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.



- Linear model: $f(\mathbf{x}) = \mathbf{W}^\top \mathbf{x}$ ($\mathbf{W} \in \mathbb{R}^{d \times p}$)
- Ordinary Least Squares

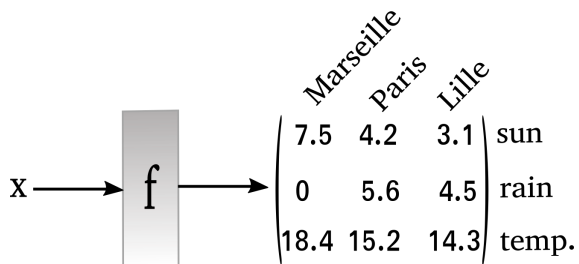
$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W} \in \mathbb{R}^{d \times p}} \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|_F^2 \quad (\mathbf{X} \in \mathbb{R}^{N \times d}, \mathbf{Y} \in \mathbb{R}^{N \times p})$$

- Reduced Rank Regression (Izenman, 1975)

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W} \in \mathbb{R}^{d \times p}} \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|_F^2 \quad \text{s.t. } \text{rank}(\mathbf{W}) \leq R$$

Tensor-valued Regression

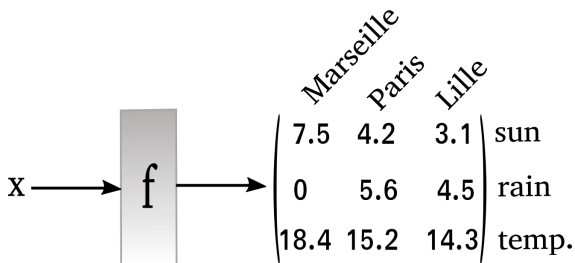
Learn $f : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_1 \times d_2}$ from $\{(\mathbf{x}^{(n)}, \mathcal{Y}^{(n)})\}_{n=1}^N$ where $\mathcal{Y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.



- Linear model: $f(\mathbf{x}) = \mathcal{W} \bullet_1 \mathbf{x}$ ($\mathcal{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}$)

Tensor-valued Regression

Learn $f : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_1 \times d_2}$ from $\{(\mathbf{x}^{(n)}, \mathcal{Y}^{(n)})\}_{n=1}^N$ where $\mathcal{Y}^{(n)} \simeq f(\mathbf{x}^{(n)})$.



- Linear model: $f(\mathbf{x}) = \mathcal{W} \bullet_1 \mathbf{x}$ ($\mathcal{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}$)
- **Low-Rank Regression for Tensor Structured Response**

$$\arg \min_{\mathcal{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}} \|\mathcal{W} \times_1 \mathbf{X} - \mathcal{Y}\|_F^2 \quad \text{s.t. } \text{rank}_{ml}(\mathcal{W}) \leq (R_0, R_1, R_2)$$

Regression Function with Low Multilinear Rank

If $f(\mathbf{x}) = \mathcal{W} \bullet_1 \mathbf{x}$ with $\text{rank}_{ml}(\mathcal{W}) \leq (R_0, R_1, \dots, R_p)$

Then $\mathcal{W} = \mathcal{G} \times_1 \mathbf{U}_0 \times_2 \cdots \times_{p+1} \mathbf{U}_p$

and $f(\mathbf{x}) = \mathcal{G} \times_1 \mathbf{x}^\top \mathbf{U}_0 \times_2 \mathbf{U}_1 \times_3 \cdots \times_{p+1} \mathbf{U}_p$

Regression Function with Low Multilinear Rank

If $f(\mathbf{x}) = \mathcal{W} \bullet_1 \mathbf{x}$ with $\text{rank}_{ml}(\mathcal{W}) \leq (R_0, R_1, \dots, R_p)$

Then $\mathcal{W} = \mathcal{G} \times_1 \mathbf{U}_0 \times_2 \cdots \times_{p+1} \mathbf{U}_p$

and $f(\mathbf{x}) = \mathcal{G} \times_1 \mathbf{x}^\top \mathbf{U}_0 \times_2 \mathbf{U}_1 \times_3 \cdots \times_{p+1} \mathbf{U}_p$

- $\text{rank}_{ml}(f(\mathbf{x})) \leq (R_1, \dots, R_p)$.

Regression Function with Low Multilinear Rank

If $f(\mathbf{x}) = \mathcal{W} \bullet_1 \mathbf{x}$ with $\text{rank}_{ml}(\mathcal{W}) \leq (R_0, R_1, \dots, R_p)$

Then $\mathcal{W} = \mathcal{G} \times_1 \mathbf{U}_0 \times_2 \cdots \times_{p+1} \mathbf{U}_p$

and $f(\mathbf{x}) = \mathcal{G} \times_1 \mathbf{x}^\top \mathbf{U}_0 \times_2 \mathbf{U}_1 \times_3 \cdots \times_{p+1} \mathbf{U}_p$

- $\text{rank}_{ml}(f(\mathbf{x})) \leq (R_1, \dots, R_p)$.
- Low-dimensional mapping:
 - (i) project \mathbf{x} in \mathbb{R}^{R_0} as $\bar{\mathbf{x}} = \mathbf{U}_0^\top \mathbf{x}$,
 - (ii) perform a low-dimensional mapping $\bar{\mathcal{Y}} = \mathcal{G} \bullet_1 \bar{\mathbf{x}} \in \mathbb{R}^{R_1 \times \cdots \times R_p}$,
 - (iii) project back into the output space to get $\mathcal{Y} = \bar{\mathcal{Y}} \times_1 \mathbf{U}_1 \times_2 \cdots \times_p \mathbf{U}_p$.

Image Reconstruction from Noisy Measurements

- $\mathcal{W} \in \mathbb{R}^{3 \times 50 \times 50}$ is an RGB image.
- Data is generated by $\mathbf{Y} = \mathcal{W} \bullet_1 \mathbf{x} + \boldsymbol{\xi}$ where $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I})$ and $\boldsymbol{\xi}_{ij} \sim \mathcal{N}(0, 1)$.
- Training set of size 200.

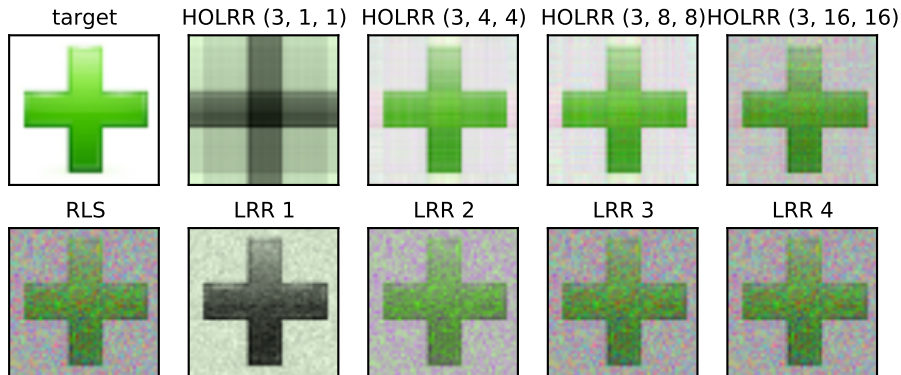


Image Reconstruction from Noisy Measurements

- $\mathcal{W} \in \mathbb{R}^{3 \times 44 \times 70}$ is an RGB image.

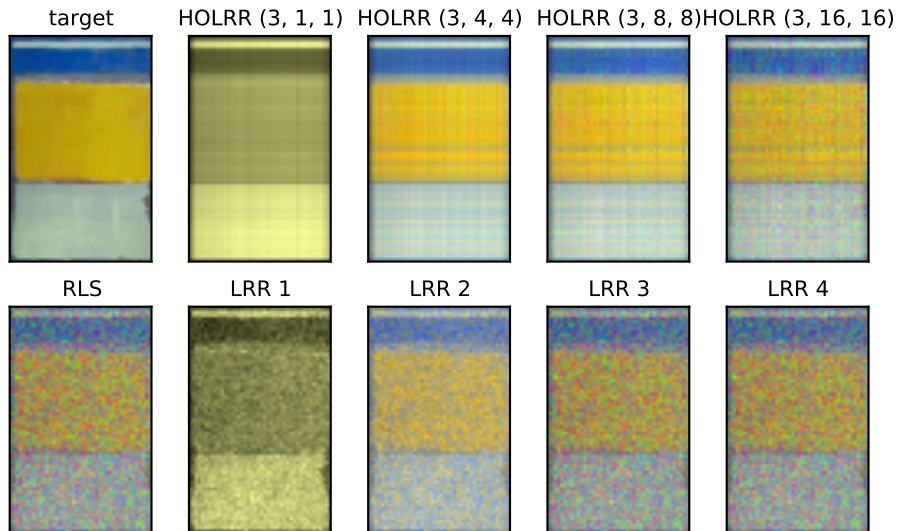
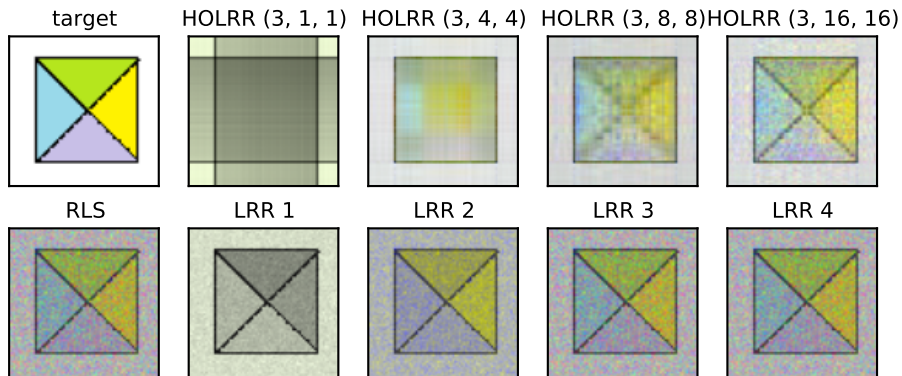


Image Reconstruction from Noisy Measurements

- $\mathcal{W} \in \mathbb{R}^{3 \times 70 \times 70}$ is an RGB image.



Solving the Minimization Problem

Problem

$$\arg \min_{\mathcal{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}} \|\mathcal{W} \times_1 \mathbf{X} - \mathcal{Y}\|_F^2 \quad \text{s.t. } \text{rank}_{ml}(\mathcal{W}) \leq (R_0, R_1, R_2)$$

Solving the Minimization Problem

Problem

$$\arg \min_{\mathcal{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}} \|\mathcal{W} \times_1 \mathbf{X} - \mathcal{Y}\|_F^2 \quad \text{s.t. } \text{rank}_{ml}(\mathcal{W}) \leq (R_0, R_1, R_2)$$

Multilinear rank constraint is equivalent to:

$$\mathcal{W} = \mathcal{G} \times_1 \mathbf{U}_0 \times_2 \mathbf{U}_1 \times_3 \mathbf{U}_2$$

for $\mathcal{G} \in \mathbb{R}^{R_0 \times R_1 \times R_2}$, $\mathbf{U}_i \in \mathbb{R}^{d_i \times R_i}$, $\mathbf{U}_i^\top \mathbf{U}_i = \mathbf{I}$ for $0 \leq i \leq 2$

Solving the Minimization Problem

Problem

$$\arg \min_{\mathbf{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}} \|\mathbf{W} \times_1 \mathbf{X} - \mathbf{Y}\|_F^2 \quad \text{s.t. } \text{rank}_{ml}(\mathbf{W}) \leq (R_0, R_1, R_2)$$

is equivalent to:

Problem

$$\arg \min_{\mathbf{U}_0, \mathbf{U}_1, \mathbf{U}_2} \|\mathbf{Y} \times_1 \mathbf{\Pi}_0 \times_2 \mathbf{\Pi}_1 \times_3 \mathbf{\Pi}_2 - \mathbf{Y}\|_F^2 \quad \text{w.r.t. } \mathbf{U}_i \in \mathbb{R}^{d_i \times R_i}$$
$$\text{s.t. } \mathbf{U}_i^\top \mathbf{U}_i = \mathbf{I} \text{ for } 0 \leq i \leq 2,$$
$$\mathbf{\Pi}_0 = \mathbf{XU}_0 \left(\mathbf{U}_0^\top \mathbf{X}^\top \mathbf{XU}_0 \right)^{-1} \mathbf{U}_0^\top \mathbf{X}^\top, \quad \mathbf{\Pi}_i = \mathbf{U}_i \mathbf{U}_i^\top \text{ for } 1 \leq i \leq 2$$

Solving the Minimization Problem

Problem

$$\arg \min_{\mathbf{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}} \|\mathbf{W} \times_1 \mathbf{X} - \mathcal{Y}\|_F^2 \quad \text{s.t. } \text{rank}_{ml}(\mathbf{W}) \leq (R_0, R_1, R_2)$$

is equivalent to:

Problem

$$\arg \min_{\mathbf{U}_0, \mathbf{U}_1, \mathbf{U}_2} \|\mathcal{Y} \times_1 \mathbf{\Pi}_0 \times_2 \mathbf{\Pi}_1 \times_3 \mathbf{\Pi}_2 - \mathcal{Y}\|_F^2 \quad \text{w.r.t. } \mathbf{U}_i \in \mathbb{R}^{d_i \times R_i}$$
$$\text{s.t. } \mathbf{U}_i^\top \mathbf{U}_i = \mathbf{I} \text{ for } 0 \leq i \leq 2,$$
$$\mathbf{\Pi}_0 = \mathbf{XU}_0 \left(\mathbf{U}_0^\top \mathbf{X}^\top \mathbf{XU}_0 \right)^{-1} \mathbf{U}_0^\top \mathbf{X}^\top, \quad \mathbf{\Pi}_i = \mathbf{U}_i \mathbf{U}_i^\top \text{ for } 1 \leq i \leq 2$$

- Find 3 low-dimensional subspaces U_0, U_1, U_2 such that projecting \mathcal{Y} onto the spaces \mathbf{XU}_0, U_1, U_2 along the corresponding modes is close to \mathcal{Y} .
- NP-hard... Solve $\arg \min_{U_i} \|\mathcal{Y} \times_{i+1} \mathbf{\Pi}_i - \mathcal{Y}\|_F^2$ instead.

Higher-Order Low-Rank Regression

Algorithm (HOLRR)

Input: $\mathbf{X} \in \mathbb{R}^{N \times d_0}$, $\mathcal{Y} \in \mathbb{R}^{N \times d_1 \times \dots \times d_p}$, $\text{rank}(R_0, R_1, \dots, R_p)$.

1: $\mathbf{U}_0 \leftarrow$ top R_0 eigenvectors of $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}_{(1)} \mathbf{Y}_{(1)}^\top \mathbf{X}$

2: **for** $i = 1$ **to** p **do**

3: $\mathbf{U}_i \leftarrow$ top R_i eigenvectors of $\mathbf{Y}_{(i+1)} \mathbf{Y}_{(i+1)}^\top$

4: **end for**

5: $\mathbf{M} = (\mathbf{U}_0^\top \mathbf{X}^\top \mathbf{X} \mathbf{U}_0)^{-1} \mathbf{U}_0^\top \mathbf{X}^\top$

6: $\mathcal{G} \leftarrow \mathcal{Y} \times_1 \mathbf{M} \times_2 \mathbf{U}_1^\top \times_3 \dots \times_{p+1} \mathbf{U}_p^\top$

7: **return** $\mathcal{W} = \mathcal{G} \times_1 \mathbf{U}_0 \times_2 \dots \times_{p+1} \mathbf{U}_p$

Approximation Guarantees

Problem

$$(*) \quad \arg \min_{\mathcal{W} \in \mathbb{R}^{d_0 \times d_1 \times d_2}} \|\mathcal{W} \times_1 \mathbf{X} - \mathcal{Y}\|_F^2 \quad \text{s.t. } \text{rank}_{ml}(\mathcal{W}) \leq (R_0, R_1, R_2)$$

Theorem

Let \mathcal{W}^* be a solution of $(*)$ and let $\hat{\mathcal{W}}$ be the regression tensor returned by HOLRR. Then,

$$\|\hat{\mathcal{W}} \times_1 \mathbf{X} - \mathcal{Y}\|_F^2 \leq 3\|\mathcal{W}^* \times_1 \mathbf{X} - \mathcal{Y}\|_F^2.$$

Statistical Guarantees

- HOLRR is statistically consistent

Statistical Guarantees

- HOLRR is statistically consistent
- Generalization bound for the class of functions

$$\mathcal{F}_{ml} = \{\mathbf{x} \mapsto \mathcal{W} \bullet_1 \mathbf{x} : \text{rank}_{ml}(\mathcal{W}) = (R_0, R_1, R_2)\}.$$

Theorem

Let $\mathcal{L} : \mathbb{R}^{d_1 \times d_2} \rightarrow \mathbb{R}$ be a loss function bounded by M .

For all $h \in \mathcal{F}_{ml}$, for any $\delta > 0$, with probability at least $1 - \delta$:

$$R(h) \leq \hat{R}(h) + M \sqrt{\frac{2D \log \left(\frac{4e(p+2)d_0 d_1 d_2}{\max_{i \geq 0} d_i} \right) \log N}{N}} + M \sqrt{\frac{\log \left(\frac{1}{\delta} \right)}{2N}}$$

where $D = R_0 R_1 R_2 + \sum_{i=0}^2 R_i d_i$.

Statistical Guarantees

- HOLRR is statistically consistent
- Generalization bound for the class of functions

$$\mathcal{F}_{ml} = \{\mathbf{x} \mapsto \mathcal{W} \bullet_1 \mathbf{x} : \text{rank}_{ml}(\mathcal{W}) = (R_0, R_1, R_2)\}.$$

Theorem

Let $\mathcal{L} : \mathbb{R}^{d_1 \times d_2} \rightarrow \mathbb{R}$ be a loss function bounded by M .

For all $h \in \mathcal{F}_{ml}$, for any $\delta > 0$, with probability at least $1 - \delta$:

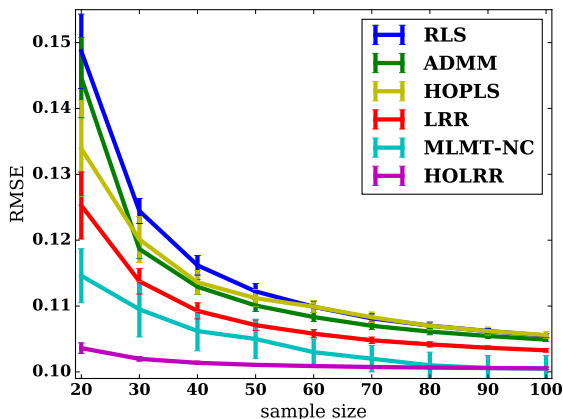
$$R(h) \leq \hat{R}(h) + M \sqrt{\frac{2D \log \left(\frac{4e(p+2)d_0 d_1 d_2}{\max_{i \geq 0} d_i} \right) \log N}{N}} + M \sqrt{\frac{\log \left(\frac{1}{\delta} \right)}{2N}}$$

where $D = R_0 R_1 R_2 + \sum_{i=0}^2 R_i d_i$.

- ▶ Without low-rank constraint: $\mathcal{O}(\sqrt{d_0 d_1 d_2})$

Experiments on Synthetic Data

RMSE w.r.t. training data set size.



Data is generated with $\mathcal{Y} = \mathcal{W} \bullet_1 \mathbf{x} + \boldsymbol{\xi}$ where $\mathcal{W} \in \mathbb{R}^{10 \times 10 \times 10 \times 10}$, $\text{rank}_{ml}(\mathcal{W}) = (6, 4, 4, 8)$, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\xi}_{i_1, i_2, i_3} \sim \mathcal{N}(0, 0.1)$.

Experiments on Real Data

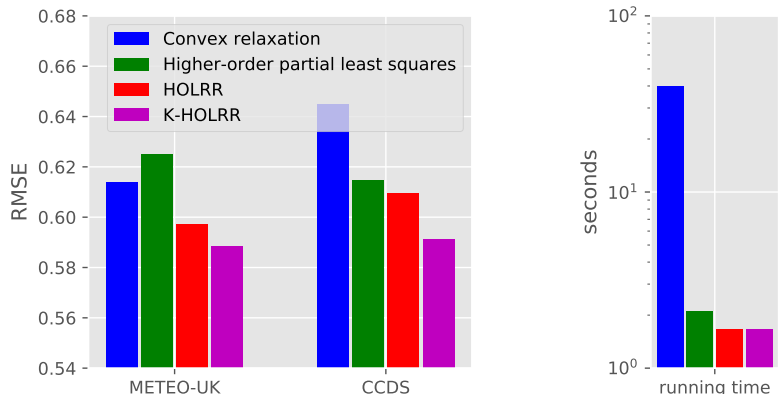


Figure: Task: predict meteorological variables in different locations from their values in the preceding 3 time steps (average over 10 runs). Output is of size 17×125 for CCDS and $5 \times 16 \times 5$ for METEO-UK.

Conclusion and Future Works

- Multilinear extension of low/reduced-rank regression
- Kernel extension \rightarrow nonlinear setting
- Fast (could be even faster), efficient, theoretical guarantees...

Conclusion and Future Works

- Multilinear extension of low/reduced-rank regression
- Kernel extension \rightarrow nonlinear setting
- Fast (could be even faster), efficient, theoretical guarantees...

- Theoretical comparative analysis of HOLRR vs. RLS/LRR (sample complexity analysis, minimax lower bound)
- Study the regularization properties of various notions of tensor rank
- HOLRR is limited to the squared error loss and does not take the tensor structure of the input into account.

- Bahadori, M. T., Yu, Q. R., and Liu, Y. (2014). Fast multivariate spatio-temporal analysis via low rank tensor learning. In *NIPS*.
- Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5(2):248–264.
- Romera-Paredes, B., Aung, M. H., Bianchi-Berthouze, N., and Pontil, M. (2013). Multilinear multitask learning. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1444–1452.