Partial Trace Regression and Low-Rank Kraus Decomposition

Hachem Kadri[1], Stéphane Ayache[1], Riikka Huusari[2], Alain Rakotomamonjy[3,4], Liva Ralaivola[4]

[1]Aix-Marseille University, CNRS, LIS, Marseille, France
[2]Helsinki Institute for Information Technology, Aalto University, Espoo, Finland
[3]Université Rouen Normandie, LITIS, Rouen, France
[4]Criteo AI Lab, Paris

Trace Regression

$$y = \mathrm{tr}\left(B_*^\top X\right) + \epsilon$$

- Generalization of linear regression to **matrix input** X
  - $\rightarrow$ Spatio-temporal data, covariance descriptors, ...
- **Output** y is a **real** number

Trace Regression

$$y = \mathsf{tr}\left(B_*^\top X\right) + \epsilon$$

- Generalization of linear regression to **matrix input** X
  $\rightarrow$ Spatio-temporal data, covariance descriptors, ...
- **Output** y is a **real** number

Low-Rank Estimation [Koltchinskii et al., 2011]

$$\widehat{B} = \arg\min_{B} \sum_{i=1}^{\ell} \left(y_i - \mathsf{tr}(B^\top X_i)\right)^2 + \lambda \|B\|_1$$

Trace Regression

$$y = \mathsf{tr}\left(B_*^\top X\right) + \epsilon$$

- Generalization of linear regression to **matrix input** X
  $\rightarrow$ Spatio-temporal data, covariance descriptors, . . .
- **Output** y is a **real** number

PSD-contrained Estimation [Slawski et al., 2015]

$$\widehat{B} = \underset{B \in \mathbb{S}_p^+}{\arg\min} \sum_{i=1}^{\ell} \left(y_i - \mathsf{tr}(B^\top X_i)\right)^2$$

Trace Regression

$$y = \mathsf{tr}\left(B_*^\top X\right) + \epsilon$$

- Generalization of linear regression to **matrix input** X
  - $\rightarrow$ Spatio-temporal data, covariance descriptors, ...
- **Output** y is a **real** number
- Relevant to:
  - $\rightarrow$ Matrix completion
  - $\rightarrow$ Phase retrieval
  - $\rightarrow$ Quantum state tomography
  - $\rightarrow$ ...

Trace Regression

$$y = \mathsf{tr}\left(B_*^\top X\right) + \epsilon$$

- Generalization of linear regression to **matrix input** X
  $\rightarrow$ Spatio-temporal data, covariance descriptors, ...
- **Output** y is a **real** number
- Relevant to:
  $\rightarrow$ **Matrix completion**

$$\arg\min_B \|\mathcal{P}_\Omega(B^*) - \mathcal{P}_\Omega(B)\|^2 \ \text{s.t. } \operatorname{rank}(B) = r$$

Trace Regression

$$y = \mathsf{tr}\left(B_*^\top X\right) + \epsilon$$

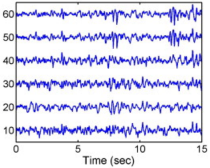
- Generalization of linear regression to **matrix input** X
  $\rightarrow$ Spatio-temporal data, covariance descriptors, ...
- **Output** y is a **real** number
- Relevant to:
  $\rightarrow$ **Matrix completion**

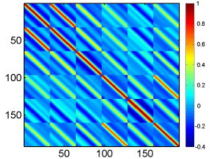
$$\arg\min_B \sum_{(i,j)\in\Omega} \left(\mathcal{P}_\Omega(B^*)_{ij} - \mathrm{tr}(B^\top E_{ij})\right)^2 \text{ s.t. } \mathrm{rank}(B) = r$$

Partial Trace Regression

Generalizes *Trace Regression* to the case when both **inputs** and **outputs** are **matrices**.

EEG data





Covariance matrix

Figure from [Williamson et al., 2012]

Domain adaptation

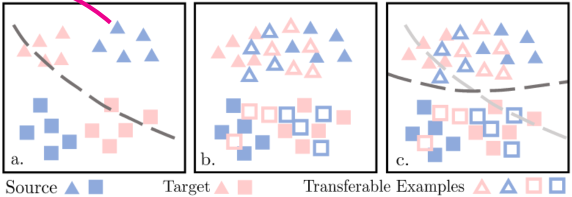


Figure from [Liu et al., 2019]

- Inspiration: **partial trace**, **CP maps** and **Kraus decomposition** in quantum computing

Notational Conventions

$\mathbb{M}_p := \mathbb{M}_p(\mathbb{R})$ the space of all $p \times p$ real **matrices**

$\mathbb{M}_p(\mathbb{M}_q)$ the space of $p \times p$ **block matrices** whose i, j entry is an element of $\mathbb{M}_q$

$\mathcal{L}(\mathbb{M}_p, \mathbb{M}_q)$ the space of **linear maps** from $\mathbb{M}_p$ to $\mathbb{M}_q$

From Trace to Partial Trace

- Trace

$$\operatorname{tr}\begin{pmatrix} \bullet & \bullet & \bullet & \cdots & \bullet \\ \bullet & \bullet & \bullet & \cdots & \bullet \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \bullet & \cdots & \bullet \end{pmatrix} = \bullet$$

From Trace to Partial Trace

- Trace of a block matrix

$$
\operatorname{tr}\left(\begin{array}{c|c|c}
\begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots\\ \bullet & \bullet & \cdots & \bullet\end{matrix} & \cdots & \begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots\\ \bullet & \bullet & \cdots & \bullet\end{matrix}\\
\hline
\vdots & \vdots & \vdots\\
\hline
\begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots\\ \bullet & \bullet & \cdots & \bullet\end{matrix} & \cdots & \begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots\\ \bullet & \bullet & \cdots & \bullet\end{matrix}
\end{array}\right) = \bullet
$$

From Trace to Partial Trace

- Partial-trace

$$\mathrm{tr}_m \left(\begin{array}{c|c|c} \begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \cdots & \bullet\end{matrix} & \cdots & \begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \cdots & \bullet\end{matrix} \\ \hline \vdots & \vdots & \vdots \\ \hline \begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \cdots & \bullet\end{matrix} & \cdots & \begin{matrix}\bullet & \bullet & \cdots & \bullet\\ \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \cdots & \bullet\end{matrix} \end{array}\right) = \begin{pmatrix} \bullet & \bullet & \cdots & \bullet \\ \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \cdots & \bullet \end{pmatrix}$$

The partial trace operation applied to $m \times m$-blocks of a $qm \times qm$ matrix gives a $q \times q$ matrix as an output.

Partial Trace Regression

$$Y = \mathrm{tr}_m\left(A_* X B_*^\top\right) + \epsilon$$

- **Matrix Input** $X \in \mathbb{M}_p$ and **Matrix Output** $Y \in \mathbb{M}_q$
- $A_*, B_* \in \mathbb{M}_{qm \times p}$ are the **unknown parameters** of the model
- We recover the **trace regression** model when $\mathbf{q = 1}$

Learning the Model Parameters

Our solution: **Kraus representation of completely positive maps**

Positive and Completely Positive Maps [Bhatia, 2009]

Positive maps

$\Phi \in \mathcal{L}(\mathbb{M}_p, \mathbb{M}_q)$ is **positive** if for all $M \in \mathbb{S}_p^+$, $\Phi(M) \in \mathbb{S}_q^+$

Positive and Completely Positive Maps [Bhatia, 2009]

m-Positive maps

$\Phi \in \mathcal{L}(\mathbb{M}_p, \mathbb{M}_q)$ is $\boldsymbol{m}$**-positive** if $\Phi_m : \mathbb{M}_m(\mathbb{M}_p) \to \mathbb{M}_m(\mathbb{M}_q)$ defined as

$$\Phi_m \begin{bmatrix} A_{11} & A_{12} & \dots \\ \vdots & \ddots & \\ A_{m1} & & A_{mm} \end{bmatrix} := \begin{bmatrix} \Phi(A_{11}) & \Phi(A_{12}) & \dots \\ \vdots & \ddots & \\ \Phi(A_{m1}) & & \Phi(A_{mm}) \end{bmatrix}$$

is positive.

Positive and Completely Positive Maps [Bhatia, 2009]

Completely positive maps

Φ is **completely positive** if it is m-positive for any $m \geq 1$

A Positive But Not Completely Positive Map

Example: the transpose map

- Define $\Phi : \mathbb{M}_2 \to \mathbb{M}_2$ by $\Phi(A) = A^\top$. Then $\Phi_1 \geq 0$ but $\Phi_2 \ngeq 0$.

$$\Phi_2\left(\left[\begin{array}{cc|cc}1&0&0&1\\0&0&0&0\\\hline 0&0&0&0\\1&0&0&1\end{array}\right]\right)=\left[\begin{array}{cc|cc}1&0&0&0\\0&0&1&0\\\hline 0&1&0&0\\0&0&0&1\end{array}\right]$$

A Completely Positive Map

Example

- Let $V \in \mathbb{M}_{q\times p}$. Define $\Phi : \mathbb{M}_p \to \mathbb{M}_q$ by $\Phi(A) = VAV^\top$. Then Φ is **completely positive**.

$$\Phi_2\left(\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]\right) = \left[\begin{array}{c|c} VA_{11}V^\top & VA_{12}V^\top \\ \hline VA_{21}V^\top & VA_{22}V^\top \end{array}\right]$$

$$= (I_2 \otimes V)\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right](I_2 \otimes V^\top)$$

Stinespring Representation

Stinespring's Theorem 1955

Let $\Phi \in \mathcal{L}(\mathbb{M}_p, \mathbb{M}_q)$. Φ writes as $\Phi(X) = \mathrm{tr}_m\left(AXA^\top\right)$ for some $A \in \mathbb{M}_{qm \times p}$ *if and only if* Φ is **completely positive**.

- Partial trace regression $\leftrightarrow$ Learning a completely positive map
- Partial trace version of the PSD-contrained trace regression
- Efficient optimization via **Kraus decomposition**

Kraus Representation

Choi's Theorem 1975, Kraus Decomposition 1971

Let $\Phi \in \mathcal{L}(\mathbb{M}_p, \mathbb{M}_q)$ be a completely positive linear map. Then there exist $A_j \in \mathbb{M}_{q\times p}$, $1 \leq j \leq r$, with $r \leq pq$ such that

$$\forall X \in \mathbb{M}_p, \quad \Phi(X) = \sum_{j=1}^{r} A_j X A_j^\top.$$

- Learning a completely positive map $\leftrightarrow$ Finding a Kraus decomposition
- Small values of r correspond to **low-rank** Kraus representation

Back to Partial Trace Regression

Low-Rank Kraus Estimation

$$\underset{A_j \in \mathbb{M}_{q\times p}}{\arg\min} \sum_{i=1}^{l} \ell\Big(Y_i, \sum_{j=1}^{r} A_j X_i A_j^\top\Big)$$

Back to Partial Trace Regression

Generalization Bound

$$\mathcal{F} = \{\Phi : \mathbb{M}_p \to \mathbb{M}_q : \Phi \text{ is completely positive and its Kraus rank is equal to } r\}$$

Under some assumptions on ℓ, for any $\delta > 0$, with probability at least $1 - \delta$, the following holds for all $h \in \mathcal{F}$,

$$R(h) \leq \hat{R}(h) + \gamma \sqrt{\frac{pqr \log(\frac{8epq}{r}) \log(\frac{l}{pqr})}{l}} + \gamma \sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{2l}}$$

Back to Matrix Completion

(Block) PSD Matrix Completion

Let $\Phi : \mathbb{M}_p \to \mathbb{M}_q$ be a linear mapping. Then the following conditions are equivalent:

1. Φ is **completely positive**.
2. The **block matrix** $\mathbf{M} \in \mathbb{M}_p(\mathbb{M}_q)$ defined by

$$\mathbf{M}_{ij} = \Phi(E_{ij}),\ 1 \leq i, j \leq p,$$

   is **positive**, where E_{ij} are the matrix units of $\mathbb{M}_p$.

Experiments

- (PSD) Matrix-to-Matrix Regression

- (Block) PSD Matrix Completion

(PSD) Matrix-to-Matrix Regression

- Simulated data: $\mathbf{10000}$ samples, $\mathbf{20 \times 20 \rightarrow 10 \times 10}$

Vectorized	Tensorized	Trace	Partial Trace

Model	**MSE**
Multivariate Regression	0.058 ± 0.0134
Reduced-Rank Regression	0.245 ± 0.1023
HOPLS	1.602 ± 0.0011
Tucker-NN	0.595 ± 0.0252
TensorTrain-NN	$\mathbf{0.001 \pm 0.0009}$
Trace Regression	0.028 ± 0.0144
Partial Trace Regression	$\mathbf{0.001 \pm 0.0008}$

(PSD) Matrix-to-Matrix Regression

- Simulated data: $\mathbf{100}$ samples, $\mathbf{20 \times 20 \rightarrow 10 \times 10}$

Model	**MSE**	**Param.**
TensorTrain-NN	0.662 ± 0.364	4,000
Partial Trace Regression	$\mathbf{0.007 \pm 0.013}$	$\mathbf{1,000}$

$\rightarrow$ **Partial Trace Regression preserves PSD structure**

(PSD) Matrix-to-Matrix Regression

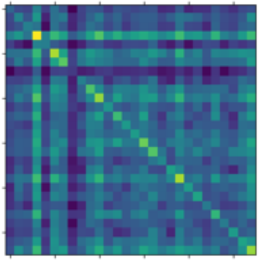
- Real BCI data: Mapping **covariance matrices** for domain adaptation

No adaptation	Optimal Transport (Full adaptation)	Partial Trace (Out-of-sample adaptation)

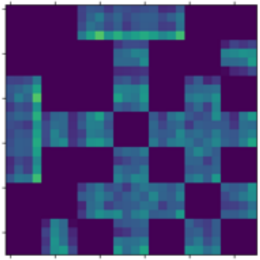
Subject	NoAdapt	FullAdapt	OoSAdapt
1	73.66	73.66	72.24
3	65.93	72.89	68.86
7	53.42	64.62	59.20
8	73.80	75.27	73.06
9	73.10	75.00	76.89

(Block) PSD Matrix Completion

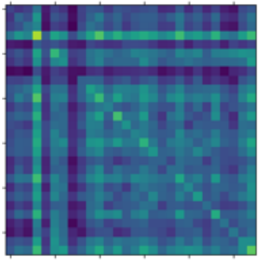
- Simulated data: Missing **blocks**, 28×28 $(p = 7, q = 4)$



Original



With missing values



Reconstructed

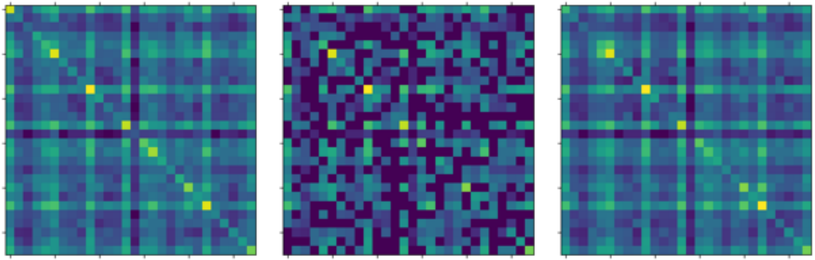
(Block) PSD Matrix Completion

- Simulated data: Missing **blocks**, $\mathbf{28 \times 28}$ $\boldsymbol{(p = 7, q = 4)}$

Model	**MSE**
TensorTrain-NN	3.942 ± 1.463
Trace Regression	2.996 ± 1.078
Partial Trace Regression	$\mathbf{0.572 \pm 0.019}$

(Block) PSD Matrix Completion

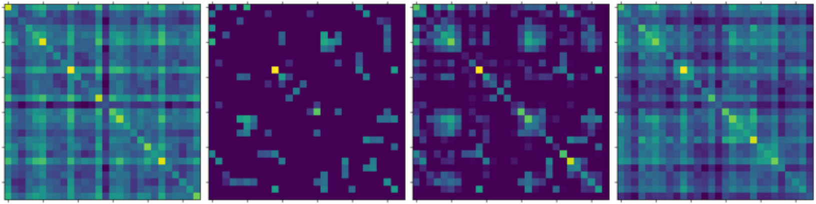
- Simulated data: Missing **entries**, 28×28 $(p = 7, q = 4)$



Original | With missing values | Reconstructed

(Block) PSD Matrix Completion

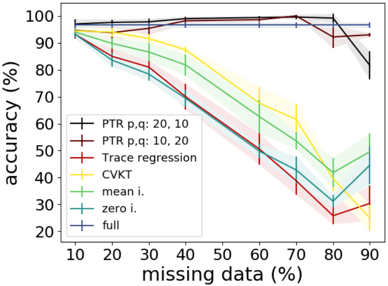
- Simulated data: Missing **entries**, 28×28 $(p = 7, q = 4)$



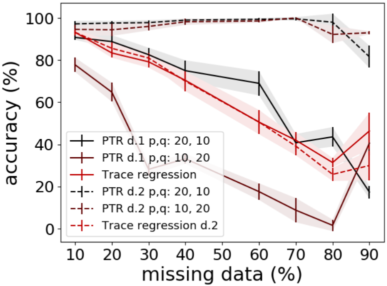
Original | With missing values | Reconstructed depth-1 | Reconstructed depth-2

(Block) PSD Matrix Completion

- Multiple features digits dataset: **Kernel matrix** completion in a multi-view setting



SVM accuracy



depth-1 Vs depth-2

Conclusion

- New model – Partial trace regression
- Novel concepts in ML – Completely postive maps and low-rank Kraus decomposition
- Theoretically well founded
- Promising performance

A bridge between Machine Learning and Quantum Computing