Journal Watch:

IEEE Transanctions Signal Processing, August 1, 2017

Saurabh Khanna,
Signal Processing for Communication Lab, ECE, IISc

Online Categorical Subspace Learning for Sketching Big Data with Misses

Authors: Yannig Shen, Morteza Mardani and G. B. Giannakis

- Categorical PCA
  - low dimensional sketching of high dimensional categorical data
- Categorical Subspace Learning (CSL) scheme is proposed that learns the latent structure of the categorial data
- Three generative models are considered.
  - Probit
    - models data samples as quantized values of low-dimensional analog signal
  - Togit
    - models censored data
  - Logit
    - generalizes logistic regression to unsupervised case

Online Categorical Subspace Learning for Sketching Big Data with Misses

- Blind Probit Model:

$$\mathcal{F}_{\text{probit}}^{(J)}(x) := s_j \text{ if } x \in (\eta_j, \eta_{j+1}]$$
$$\text{for } j = 0, 1, \ldots, J-1$$

$$\begin{aligned} y_{i,t} &= \mathcal{F}_{\text{probit}}^{(J)}(x_{i,t} + v_{i,t}) \\ x_{i,t} &:= \mathbf{u}_i^\top \psi_t, \qquad i \in \Omega_t \end{aligned}$$

- Goal is to find $\{\psi_t\}_{t=1}^{T}$ and $\mathbf{U}$ given $\{y_{i,t}\}$
- **Example:**
  - The patient's survival, or, death is a binary response that depends upon several factors such as age, weight, gender, as well as the treatment dose and specifications.

Online Categorical Subspace Learning for Sketching Big Data with Misses

- Blind Tobit Model:

$$\mathcal{F}^{l}_{\text{tobit}}(x) := \begin{cases} \eta_u & x \geq \eta_u \\ \eta_l & x \leq \eta_l \\ x & x \in (\eta_l, \eta_u). \end{cases}$$

$$\begin{aligned} y_{i,t} &= \mathcal{F}_{\text{tobit}}(x_{i,t} + v_{i,t}) \\ x_{i,t} &:= \mathbf{u}_i^{\top} \psi_t, \qquad i \in \Omega_t. \end{aligned}$$

- **Example:**
  - If the patient dies naturally within the study period, one knows precisely the survival time. However, if the patient dies before or after the study, where no accurate data is collected, only an upper or a lower bound is available on the patient age.

Online Categorical Subspace Learning for Sketching Big Data with Misses

- Blind Logit Model:

$$\mathcal{F}_{\text{logit}}(x_{i,t}) := \Pr(y_{i,t} = s) = \frac{1}{1 + \exp((1-2s)x_{i,t})}, i \in \Omega_t.$$

$$\log \frac{\Pr(y_{i,t} = s_j)}{\Pr(y_{i,t} = s_0)} = \psi_t^\top \mathbf{u}_i^{(j)}, \quad j = 1, \ldots, J-1$$

$$\Pr(y_{i,t} = s_j) = \frac{\exp(\psi_t^\top \mathbf{u}_i^{(j)})}{1 + \sum_{k=1}^{J-1} \exp(\psi_t^\top \mathbf{u}_i^{(k)})}, j = 1, \ldots, J-1.$$

- **Example:**
  - The Logit model can predict the survival chance within a certain period of time.

Online Categorical Subspace Learning for Sketching Big Data with Misses

- Rank Regularized ML estimation of model parameters Ψ and **U**

$$\min_{\mathbf{X}=\mathbf{U}\boldsymbol{\Psi}} -\log \mathcal{L}\left(\{y_{i,\tau}, i \in \Omega_\tau\}_{\tau=1}^{T}; \mathbf{U}, \boldsymbol{\Psi}\right) + \lambda \|\mathbf{X}\|_*$$

- Variational form of nuclear norm

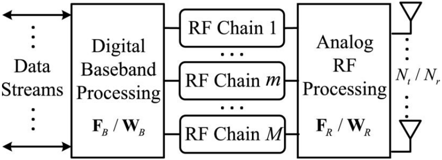
$$\|\mathbf{X}\|_* = \min_{\{\mathbf{U},\boldsymbol{\Psi}\}} \frac{1}{2}\left(\|\mathbf{U}\|_F^2 + \|\boldsymbol{\Psi}\|_F^2\right)$$
$$\text{s. to} \quad \mathbf{X} = \mathbf{U}\boldsymbol{\Psi}$$

- Equivalent Rank Regularized ML

$$\min_{\{\mathbf{U},\boldsymbol{\Psi}\}} -\log \mathcal{L}\left(\{y_{i,\tau}, i \in \Omega_\tau\}_{\tau=1}^{T}; \mathbf{U}, \boldsymbol{\Psi}\right) + \frac{\lambda}{2}\left(\|\mathbf{U}\|_F^2 + \|\boldsymbol{\Psi}\|_F^2\right).$$

Near-Optimal Hybrid Processing for Massive MIMO Systems via Matrix Decomposition

- Hybrid precoding architecture



- Baseband model before analog RF processing at RX

$$\mathbf{y} = \mathbf{H}\mathbf{F}_R\mathbf{F}_B\mathbf{s} + \mathbf{n}$$

- and after RX post processing

$$\tilde{\mathbf{y}} = \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{H}\mathbf{F}_R\mathbf{F}_B\mathbf{s} + \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{n}$$

- How to design the hybrid precoders $\mathbf{F}_B, \mathbf{F}_R$ and hybrid combiners $\mathbf{W}_B, \mathbf{W}_R$?

Near-Optimal Hybrid Processing for Massive MIMO Systems via Matrix Decomposition

- Conventional approach:

$$\begin{aligned} & \max \ R(\mathbf{F}_R, \mathbf{F}_B, \mathbf{W}_R, \mathbf{W}_B) \\ & \text{s.t. } ||\mathbf{F}_R\mathbf{F}_B||_F^2 = N_s, \\ & \quad \mathbf{F}_R \in \mathcal{F}_R, \ \mathbf{W}_R \in \mathcal{W}_R, \end{aligned}$$

- Proposed approach:
  - First learn the conventional precoder $\mathbf{F}^*$ and combiner $\mathbf{W}^*$ via unconstrained optimization
  - Then, use matrix decomposition to obtain hybrid precoders and combiners

$$\begin{aligned} & \min_{\mathbf{F}_R, \mathbf{F}_B} ||\mathbf{F}^\star - \mathbf{F}_R\mathbf{F}_B||_F \\ & \text{s.t. } ||\mathbf{F}_R\mathbf{F}_B||_F^2 = N_s, \\ & \quad \mathbf{F}_R \in \mathcal{F}_R. \end{aligned} \qquad \begin{aligned} & \min_{\mathbf{W}_R, \mathbf{W}_B} ||\mathbf{W}^\star - \mathbf{W}_R\mathbf{W}_B||_F \\ & \text{s.t. } \mathbf{W}_R \in \mathcal{W}_R. \end{aligned}$$

Near-Optimal Hybrid Processing for Massive MIMO Systems via Matrix Decomposition

- $\mathbf{F}^*$ found by maximizing the mutual information:

$$\mathbf{F}^* = \max_{\mathbf{F}} \log_2 \left(\left| \mathbf{I}_{N_s} + \frac{\gamma}{N_s} \mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H \right| \right).$$

- Once $\mathbf{F}_R, \mathbf{F}_B$ are found, one can find optimal $\mathbf{W}^*$ as

$$\begin{aligned}\mathbf{W}^\star = \mathbf{W}_{MMSE} &= \arg\min_{\mathbf{W}} \mathbb{E}\left[||\mathbf{s} - \mathbf{W}\mathbf{y}||_2 \right] \\ &= \frac{\sqrt{P}}{N_s} \left(\frac{P}{N_s} \mathbf{H}\mathbf{F}_R\mathbf{F}_B\mathbf{F}_B^H\mathbf{F}_R^H\mathbf{H}^H + \sigma^2 \mathbf{I}_{N_r} \right)^{-1} \mathbf{H}\mathbf{F}_R\mathbf{F}_B.\end{aligned}$$

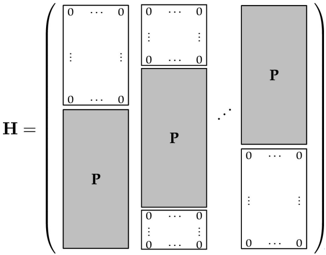
Tensor Decomposition With Several Block-Hankel Factors and Application in Blind System Identification

Authors: Frederic V Eeghem, M. Sorenson and L.D. Lathauwer

- Hankel matrix

$$\begin{bmatrix} a & b & c & d \\ b & c & d & e \\ c & d & e & f \\ d & e & f & g \end{bmatrix}$$

- Banded Block-Hankel structure



Tensor Decomp. With Several Blk-Hankel Factors and Application in Blind System Identification

Blind system identification problem

- Consider an LTI system with R inputs and M outputs, having $L+1$ coefficients per input-output channel.

$$y_m(n) = \sum_{r=1}^{R} \sum_{l=0}^{L} h_{mr}(l) s_r(n-l) + v(n)$$

- Goal is to determine $h_{mr}(n)$ based on the observed data sequence $y_m(n)$.

Tensor Decomp. With Several Blk-Hankel Factors and Application in Blind System Identification

Tensors

- Scalar is a rank-0 tensor (magnitude only)
- Vector is a tensor of rank-1 (magnitude and one direction)
- Matrix is a tensor of rank-2 (magnitude and two directions)
- Triad is a tensor of rank-3 (magnitude and three directions)

and so on...

Tensor Decomp. With Several Blk-Hankel Factors and Application in Blind System Identification

Canonical Polyadic Decomposition (CPD)

- Decomposition into a sum of rank-1 terms

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)}.$$

- The **rank of a tensor** $\mathcal{X}$ is equal to the minimal number of rank-1 tensors that yield X in a linear combination.
- The vectors $\left\{\mathbf{u}_r^{(n)}\right\}$ can be stacked as **factor matrices**

$$\mathbf{U}^{(n)} = \left[\begin{array}{ccc} \mathbf{u}_1^{(n)} & \cdots & \mathbf{u}_R^{(n)} \end{array} \right] \in \mathbb{C}^{I_n \times R},$$

- Compact representation of tensor $\mathcal{X}$:

$$\mathcal{X} = \left[\left[\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\right]\right].$$

Tensor Decomp. With Several Blk-Hankel Factors and Application in Blind System Identification

Uniqueness of Canonical Polyadic Decomposition

The PD of $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ $\left(\mathcal{X} = \left[\left[\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}\right]\right]\right)$ is unique if

$$\begin{cases} \mathbf{U}^{(3)} \text{ has full column rank R,} \\ \mathcal{C}_2\left(\mathbf{U}^{(1)}\right) \odot \mathcal{C}_2\left(\mathbf{U}^{(2)}\right) \text{ has full column rank,} \end{cases}$$

- $C_2(\mathbf{A})$ is defined as the k^{th} compound matrix of an $I \times R$ matrix $\mathbf{A}$, which is the $\binom{I}{k} \times \binom{R}{k}$ matrix containing the determinants of all $k \times k$ submatrices of $\mathbf{A}$.

Tensor Decomp. With Several Blk-Hankel Factors and Application in Blind System Identification

Blind system identification problem

$$y_m(n) = \sum_{r=1}^{R} \sum_{l=0}^{L} h_{mr}(l) s_r(n-l) + v(n)$$

- Generate pseudo-measurements (4$^{\text{th}}$ order cumulants)

$$\begin{aligned} & c_{m_1 m_2 m_3 m_4}(l_1, l_2, l_3) \\ & := \text{Cum}\left[y^*_{m_1}(n), y_{m_2}(n+l_1), y^*_{m_3}(n+l_2), y_{m_4}(n+l_3)\right] \\ & = \sum_{r=1}^{R} \gamma_r \sum_{l=0}^{L} h_{m_1 r}(l)^* h_{m_2 r}(l+l_1) h_{m_3 r}(l+l_2)^* h_{m_4 r}(l+l_3) \end{aligned}$$

- The pseudo-measurements can be rearranged as a fourth-order tensor

$$\mathcal{T} = \sum_{r=1}^{R} \gamma_r \sum_{l=0}^{L} \mathbf{p}_r^{(l)*} \circ \mathbf{h}_r^{(l)} \circ \mathbf{h}_r^{(l)*} \circ \mathbf{h}_r^{(l)}$$

Tensor Decomp. With Several Blk-Hankel Factors and Application in Blind System Identification

- The pseudo-measurements (cumulants) rearranged as a fourth-order tensor

$$\mathcal{T} = \sum_{r=1}^{R} \gamma_r \sum_{l=0}^{L} \mathbf{p}_r^{(l)*} \circ \mathbf{h}_r^{(l)} \circ \mathbf{h}_r^{(l)*} \circ \mathbf{h}_r^{(l)}$$

where $p_r^{(l)}$ is the r^{th} column of $\mathbf{P}^{(l)}$, defined as

$$\mathbf{P}^{(l)} = \begin{bmatrix} h_{11}(l) & \cdots & h_{1R}(l) \\ \vdots & \ddots & \vdots \\ h_{M1}(l) & \cdots & h_{MR}(l) \end{bmatrix} \in \mathbb{C}^{M \times R},$$

- Compact representation of pseudo-measurements $\mathcal{T}$:

$$\mathcal{T} = [[\mathbf{G}^*, \mathbf{H}, \mathbf{H}^*, \mathbf{H}]],$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{P}^{(0)}\Gamma & \cdots & \mathbf{P}^{(L)}\Gamma \end{bmatrix} \in \mathbb{C}^{M \times R(L+1)}, \qquad \Gamma = \operatorname{diag}(\gamma)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}^{(0)} & \cdots & \mathbf{H}^{(L)} \end{bmatrix} \in \mathbb{C}^{M(2L+1) \times R(L+1)} \quad \text{(block Hankel)}$$

Tensor Decomp. With Several Blk-Hankel Factors and Application in Blind System Identification

- Compact representation of pseudo-measurements $\mathcal{T}$:

$$\mathcal{T} = [[\mathbf{G}^*, \mathbf{H}, \mathbf{H}^*, \mathbf{H}]] ,$$

$$\mathbf{G} = \left[\begin{array}{ccc} \mathbf{P}^{(0)}\Gamma & \cdots & \mathbf{P}^{(L)}\Gamma \end{array} \right] \in \mathbb{C}^{M \times R(L+1)}, \qquad \Gamma = \mathrm{diag}(\gamma)$$

$$\mathbf{H} = \left[\begin{array}{ccc} \mathbf{H}^{(0)} & \cdots & \mathbf{H}^{(L)} \end{array} \right] \in \mathbb{C}^{M(2L+1) \times R(L+1)} \quad \text{(block Hankel)}$$

- By unfolding $\mathcal{T}$ and rewriting in a matrix form, we get

$$\begin{aligned} \mathbf{T} &= \left[\mathbf{t}^{(1,1)}, \ldots, \mathbf{t}^{(1,M(2L+1))}, \mathbf{t}^{(2,1)}, \ldots, \mathbf{t}^{(M,M(2L+1))} \right] \\ &= (\mathbf{H}^* \odot \mathbf{H}) (\mathbf{G} \odot \mathbf{H}^*)^{\mathsf{T}}, \end{aligned}$$

- When is the above decomposition unique?

Other Interesting Papers:

- Low-Rank Phase Retrieval
- Robust Distributed Estimation by Networked Agents
- Parameter Estimation in Sensor Networks Under Probabilistic Censoring
- Matrix Product State for Higher-Order Tensor Compression and Classification
- Blind Multichannel Deconvolution and Convolutive Extensions of Canonical Polyadic and Block Term Decompositions