Tensors

On an order-3 tensor $B$, for each of the modes $i \in [3] := \{1, 2, 3\}$:
- size of the $i$-th mode: $I_i$
- mode-$i$ fibers: fixing every index but the $i$-th, e.g., mode-1 fiber: $B_{:, :, :}$
- mode-$i$ unfolding: matrix $B_{:, :, :}$, whose columns are mode-$i$ fibers

decomp: CP, Tucker, tensor-train, ...

$$l_1 \cdot l_2 \cdot l_3 = \begin{bmatrix} A \cdot y_y \cdot y_z \end{bmatrix}$$

Figure 1: Tucker decomposition with multilinear rank $(r_1, r_2, r_3)$: $B = X_1 x_1^T + X_2 x_2^T + X_3 x_3^T$.

tensor completion

Given a partially observed $B_{\text{obs}} \in \mathbb{R}^{l_1 \times l_2 \times l_3}$, we have
- observation pattern $\Omega \in \{0, 1\}^{l_1 \times l_2 \times l_3}$, $\Omega_{i,j,k} = 0$ if $B_{i,j,k}$ is observed, and 0 otherwise.
- observation probability $P \in \mathbb{R}^{l_1 \times l_2 \times l_3}$: $P_{i,j,k} = P(\Omega_{i,j,k} = 1) = P(B_{i,j,k} \text{is observed})$

missingness types
- $(P_{\text{obs}})$ missing-completely-at-random (MCAR) uniform
- $(P_{\text{obs}})$ missing-not-at-random (MNAR) non-uniform

1-bit matrix completion

Given a binary matrix $Y \in \{0, 1\}^{m \times n}$, predict the parameter matrix $M \in \mathbb{R}^{m \times n}$.

Assumptions:
- $M$ is approximately low rank.
- $Y$ is a matrix function: $Y: R \rightarrow \{0, 1\}^m$, such that $P(Y_{i,j} = 1) = \sigma(M_{i,j})$ for $i \in [m]$.

Low rank surrogates for $M$: low nuclear norm, low max norm, ...

Our problem formulation: MNAR tensor completion

Input: MNAR data tensor $B_{\text{obs}} \in \mathbb{R}^{l_1 \times l_2 \times l_3}$.

Assumptions:
- true data tensor $B \in \mathbb{R}^{l_1 \times l_2 \times l_3}$ is approximately low multilinear rank
- noiseless observation: $(B_{\text{obs}})_{i,j,k} = B_{i,j,k}$ if $B_{i,j,k}$ is observed, and 0 otherwise.

- unknown parameter tensor $\hat{A} \in \mathbb{R}^{l_1 \times l_2 \times l_3}$ has the same rank structure as $B$.

1-bit observation: With the observation probability parameter $P \in \mathbb{R}^{l_1 \times l_2 \times l_3}$, $P(B_{\text{obs}})_{i,j,k} = \sigma(B_{i,j,k})$, in which $\sigma: R \rightarrow \{0, 1\}$ is a non-decreasing link function.

Algorithm Step 1: propensity recovery

Given a mask tensor $Q$, get a predicted propensity tensor $\hat{P}$.

algorithm hyperparameters
- Choice 1: convex proximal-proximal-gradient $\tau$ and $\gamma$
- Choice 2: nonconvex gradient descent target rank and step size

Choice 1: convex and provable
- get the square set and square unfolding [5] of $\Omega$.
- predict parameter $A$ by logistic loss minimization (by proximal-proximal-gradient [6])

$$A_{i,j,k} := \arg \min_{A_{i,j,k}} \sum_{i,j \in \Omega} - (A_{i,j,k} \log (\sigma(A_{i,j,k})) - (1 - A_{i,j,k}) \log (1 - \sigma(A_{i,j,k})))$$

where $\sigma_{i,j,k} = \{1 \in \mathbb{R}^{l_1 \times l_2 \times l_3} \mid \|\|F\|\| \leq \gamma\}$. $\hat{P}_{i,j,k} = \sigma(A_{i,j,k})$

Algorithm Step 2: tensor completion

Given $\hat{P}$ and MNAR observations $B_{\text{obs}}$, get $B$.

- Form an entrywise inverse propensity estimator for data tensor $B$ as $\hat{B} = \text{SO-HOSVD}(Q \hat{P})$.
- Do Tucker decomposition on $X(\hat{B})$.

Estimate $B$ by $\hat{B} = W(\hat{P}) x_1^T Q_1(\hat{P}) x_2^T Q_2(\hat{P})$.

Theoretical guarantees

Upper bound for propensity recovery error [1, 3]: Assumption that $\hat{P} = A(\hat{A})$. A given set $S \subset [X]$, together with the following assumptions:
- $A_{\hat{P}}$ has bounded nuclear norm: there exists a constant $\theta > 0$ such that $\|A_{\hat{P}}\| \leq \theta \|\|T\|\|_{\text{F}}$.
- Entries of $A$ have bounded absolute value: there exists a constant $\alpha > 0$ such that $|A_{ij,kl}| \leq \alpha$.

Suppose we run the convex propensity recovery algorithm with thresholds satisfying $\tau \geq \theta$ and $\gamma \geq \alpha$ to obtain an estimate $\hat{P}$ of $P$. With $L_i := \sup_{x,y} \|x \cdot y\| / \|x\|_2$, there exists a universal constant $C > 0$ such that $\|E\| \leq C \frac{\|P - \hat{P}\|}{\|P\|}$, with probability at least $1 - \frac{1}{L_i}$.

Optimality of the square unfolding for propensity recovery: Instate the same conditions as the previous lemma on propensity recovery error, and further assume that there exists a constant $c > 0$ such that $cr_i^2 \leq L_i$ for every $i \in [N]$. Then $S = \{S\}$ gives the tightest upper bound on the propensity estimation error $\|P - \hat{P}\|$ among all unfolding sets $S \subset [X]$. and MNAR tensor completion on synthetic data:

- real video tensor from $[4]$: $B \in \{0, 1\}^{29 \times 9 \times 255}$
- synthetic parameter tensor $A = (B - 128)/64$

Bibliography