

**Assignment 4**

**Tensor decomposition for missing data completion**

Date assigned: 5 Nov. 2025

Date due: 19 Nov. 2025

---

The `traffic_speed.mat` data are collected from 323 sensors installed on a highway in 5 days. The traffic speed is averaged every 5 minutes; therefore, there are 288 data points per day. However, there are 20% missing data in `train.mat`, which need to be completed. Please apply tensor decomposition to analyze the data and answer the following questions:

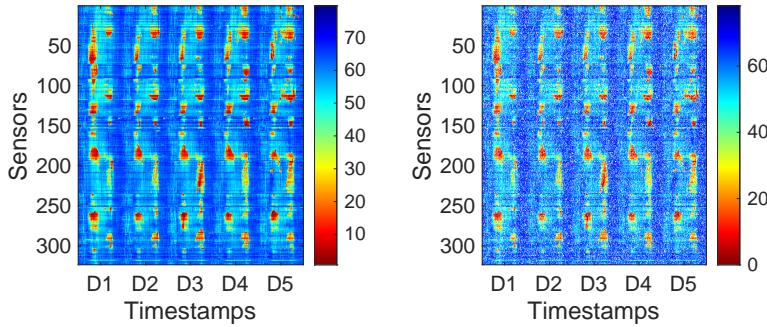


Figure 1: Left: full traffic speed data (ground truth); Right: training data with 20% missing data.

1. The current `train.mat` is a matrix, how do we organize it as a 3rd-order tensor to have a better missing data estimation? Explain the size of the tensor and the reason.
2. Apply CP decomposition to estimate the missing values when tensor rank  $r = 10, 20$  and  $50$ , respectively. Record the root-mean-squared-error (RMSE) of the estimated missing data points (test data) and plot the completed traffic speed data. Discuss your findings. (Note: the ground-truth is in `ground_truth.mat`.)
3. Apply Tucker decomposition to estimate the missing values under different tensor ranks, e.g.,  $r = [30, 20, 3]$ ,  $r = [40, 30, 2]$  or other combinations. Plot the completed traffic speed data and report the minimum RMSE you achieve and its corresponding tensor rank.
4. Discuss the difference between CP and Tucker decomposition. Explain which one is better.