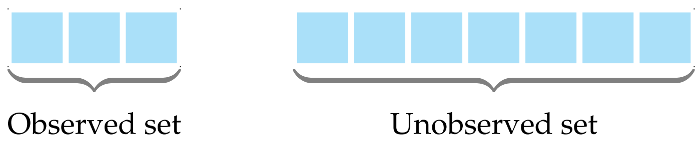
Model Settings in Imputation Experiments

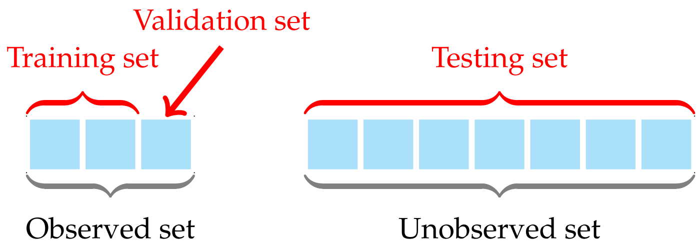
"Matrix and Tensor Model for Spatiotemporal Traffic Data Imputation and Forecasting"

Generally speaking, the selection of hyperparameters in the imputation methods is a little different from the forecasting methods. We do not have the validation set because:

- The imputation task separates the data into the observed set (i.e., training set) and the unobserved set (i.e., testing set), see below for an illustration.



- If one considers splitting the observed set into a training set and a validation set (see below), then it is possible to select hyperparameters in the imputation method as classical machine learning methods. However, the tuned hyperparameters on the training set would not match the observed set because the training set here is only a fraction of the observed set, causing model biases.



  One correct path would be: 1) Learning the model from the training set, 2) tuning hyperparameters on the validation set, and 3) using the training set to make the imputation on the testing set. However, the training set would be more sparse than the observed set, and therefore the imputation performance is not as good as learning the model from the observed set.

How to Select Hyperparameters in LATC and Baseline Imputation Methods?

We consider the following settings for our LATC model and the baseline models.

- LATC can be described as the following optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\mathcal{X}},\boldsymbol{Z}} \ \|\boldsymbol{\mathcal{X}}\|_{r,*}+\frac{\gamma}{2}\|\boldsymbol{Z}\|_{\boldsymbol{A},\mathcal{H}} \\ &\text{s.t.} \ \begin{cases} \boldsymbol{\mathcal{X}}=\mathcal{Q}(\boldsymbol{Z}) \\ \mathcal{P}_{\Omega}(\boldsymbol{Z})=\mathcal{P}_{\Omega}(\boldsymbol{Y}) \end{cases} \end{aligned} \tag{1}$$

  where we have some hyperparameters in the solution algorithms:

  - Hyperparameter λ: This value determines the algorithm's convergence, which is usually set as small values, e.g., $\lambda\in\{10^{-4},10^{-5}\}$. We can also set a very small value for λ, but it would cost a lot of iterations.

- Hyperparameter γ: This is the weight parameter of temporal autoregression. This value can be determined according to the importance of temporal modeling. If we want to reinforce temporal correlations of time series, we can use a relatively large value. In our experiments (on both Seattle and Portland traffic datasets), we set $\gamma \in \{\frac{\lambda}{10}, \frac{\lambda}{5}, \lambda, 5\lambda, 10\lambda\}$.
- Truncation parameter r: This value is important for the singular value thresholding, making the first r singular values fixed in the thresholding process. Generally speaking, this value can be set relatively small, but it is determined according to the sparsity of data. In our experiments (on both Seattle and Portland traffic datasets), we consider $r \in \{5, 10, 15, 20, 25\}$. For example, in Table 4.1, the best results for the nonrandom missing and the block-out missing are achieved as $r = 10$.
- Time lag set $\mathcal{H}$. Since tensorization $\mathcal{Q}(\cdot)$ in LATC can capture nonlocal time series trends, we only consider the local time lags. For example, we set $\mathcal{H} = \{1, 2, 3, 4, 5, 6\}$ (Seattle dataset) and $\mathcal{H} = \{1, 2, 3, 4\}$ (Portland dataset).

- As shown in Table 4.2, LRTC-TNN is an important baseline, which is also the special case of our LATC model without temporal autoregression (i.e., removing the weight parameter γ). The other hyperparameters of LRTC-TNN are set as the same as our LATC model. Thus, the comparison is fair and it is possible to highlight the importance of temporal autoregression.
- As shown in Table 4.2, LAMC is the matrix version of LATC. We also set the same hyperparameters in LAMC as our LATC model.
- As shown in Table 4.2, BTMF is a matrix factorization model whose hyperparameters only include the rank $R \in \mathbb{N}^{+}$ and the time lag set $\mathcal{H}$. We use the default time lag set (local time lags + nonlocal time lags), i.e., $\mathcal{H} = \{1, 2, 288\}$ (Seattle dataset) and $\mathcal{H} = \{1, 2, 96\}$ (Portland dataset). The rank R ($R \leq \min\{N, T\}$) is very important for the imputation task, we set some proper values (e.g., $R = 80$) that best performed in these imputation tasks.

How to Select Hyperparameters in LCR and Baseline Imputation Methods?

We consider the following settings for our LCR model and the baseline models.

- LCR can be solved by ADMM, and the augmented Lagrangian function is given by

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}) = \|\mathcal{C}(\boldsymbol{x})\|_* + \frac{\gamma}{2}\|\boldsymbol{\ell} \star \boldsymbol{x}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \langle \boldsymbol{w}, \boldsymbol{x} - \boldsymbol{z}\rangle + \frac{\eta}{2}\|\mathcal{P}_\Omega(\boldsymbol{z} - \boldsymbol{y})\|_2^2 \tag{2}$$

  where we have some hyperparameters:

  - Hyperparameter λ: This value determines the algorithm's convergence. Since we perform the thresholding process for ℓ_1-norm minimization in the frequency domain, we set λ according to the data length NT, the ratio could be a small value in our experiments, e.g, 10^{-5}. So we can see $\lambda = 10^{-5}NT$ used in our experiments.
  - Hyperparameter γ: This is the weight parameter of temporal regularization corresponding to the local trend modeling. As we want to highlight the importance of local trends, the value can be set as $\gamma = 10\lambda$.
  - Hyperparameter η: This value can be set as $\eta \to \infty$, corresponding to the constraint $\mathcal{P}_\Omega(\boldsymbol{z}) = \mathcal{P}_\Omega(\boldsymbol{y})$. In our model, we set η as a large default value, i.e., $\eta = 10^2\lambda$.

- Kernel size τ: If the given time series is very sparse, and adjacent data points around observations are mostly missing, then we should set a relatively large value for τ to borrow local information. In our experiment, we set $\tau = 1$ (or 2) for a relatively low missing rate and $\tau = 3$ for a high missing rate.

- Since the baseline models such as CTNNM and CircNNM are special cases of our LCR models without any temporal regularization, we can follow the same hyperparameters as LCR. By doing so, the imputation results can help highlight the importance of temporal regularization.

- For the remaining baseline models such as HaLRTC and LRTC-TNN, we follow similar settings as mentioned above.