APPENDIX

A. Proof of Proposition 1

According to the updating formula of the null cell $x_{ij} \in \mathcal{C} \cap M$ in Line 4 in Algorithm 1, we have

$$\begin{aligned}&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}\right)\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)} - \eta\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}},\ \forall x_{ij} \in M\right).\end{aligned}$$

Given $\eta \leq \epsilon / \left\|\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}}\right\|$, i.e., $\left\|-\eta\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}}\right\| \leq \epsilon$, referring to the first-order Taylor expansion [45], it follows

$$\begin{aligned}&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)} - \eta\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}},\ \forall x_{ij} \in M\right)\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)}, \forall x_{ij} \in M\right)\\ &+ \sum_{x_{ij}\in M}\left\|-\eta\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}}\right\| \cdot \frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}}\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)}, \forall x_{ij} \in M\right)\\ &- \eta\sum_{x_{ij}\in M}\left\|\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}}\right\|^2\\ \leq&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)}, \forall x_{ij} \in M\right)\\ =&\Theta\left(\mathcal{C}'^{(k)}, \mathcal{G}\right).\end{aligned}$$

B. Proof of Proposition 2

We first consider the left term $\Theta(\mathbf{C}_1', \mathcal{G} \mid \mathbf{C}_2')$ of Formula 4, which leads to

$$\begin{aligned}&\Theta(\mathbf{C}_1', \mathcal{G} \mid \mathbf{C}_2')\\ =&\sum_{g_p\in\mathcal{G}}\sum_{C_{ip}'\in\mathbf{C}_1'}\|x_{ip}' - g_p(C_{ip}' \setminus \{x_{ip}'\})\|^2\\ =&\Theta(\mathbf{C}_1', \mathcal{G}).\end{aligned}$$

According to Definition 6, we know $\mathbf{C}_1 \cap \mathbf{C}_2 \cap M = \emptyset$, which means that the missing cells in $\mathbf{C}_2$ do not appear in $\mathbf{C}_1$. Therefore, we also have

$$\begin{aligned}&\Theta(\mathbf{C}_1', \mathcal{G} \mid \mathbf{C}_2'')\\ =&\sum_{g_p\in\mathcal{G}}\sum_{C_{ip}'\in\mathbf{C}_1'}\|x_{ip}' - g_p(C_{ip}' \setminus \{x_{ip}'\})\|^2\\ =&\Theta(\mathbf{C}_1', \mathcal{G}).\end{aligned}$$

It completes the proof.

C. Proof of Proposition 3

Consider the initialization of $\mathcal{C}_m = \{\mathbf{C}_1, \dots, \mathbf{C}_u\}$ in Line 2 in Algorithm 2, which follows

$$\mathcal{C}' = \mathbf{C}_1' \cup \mathbf{C}_2' \cup \dots \cup \mathbf{C}_u' \cup \{\mathcal{C}' \setminus \mathcal{C}_m'\}.$$

According to Definition 3 for the imputation cost, we can obtain

$$\begin{aligned}\Theta\left(\mathcal{C}', \mathcal{G}\right) =&\Theta\left(\mathbf{C}_1', \mathcal{G}\right) + \Theta\left(\mathbf{C}_2', \mathcal{G}\right) + \dots\\ &+ \Theta\left(\mathbf{C}_u', \mathcal{G}\right) + \Theta\left(\mathcal{C}' \setminus \mathcal{C}_m', \mathcal{G}\right).\end{aligned}$$

Combing with Proposition 2, for any $\mathbf{C}_i, \mathbf{C}_j \in \mathcal{C}_m$, they always have

$$\Theta\left(\mathbf{C}_i', \mathcal{G} \mid \mathbf{C}_j'\right) = \Theta\left(\mathbf{C}_i', \mathcal{G} \mid \mathbf{C}_j''\right) = \Theta\left(\mathbf{C}_i', \mathcal{G}\right),$$

where $\mathbf{C}_j'$ and $\mathbf{C}_j''$ are two different fillings of $\mathbf{C}_j$.

Moreover, for any $\mathbf{C}_i, \mathbf{C}_j \in \mathcal{C}_m$ at the k-th round update in Algorithm 2, they always hold

$$\begin{aligned}&\Theta\left(\mathbf{C}_i'^{(k+1)}, \mathcal{G} \mid \mathbf{C}_j'^{(k+1)}\right)\\ =&\Theta\left(\mathbf{C}_i'^{(k+1)}, \mathcal{G} \mid \mathbf{C}_j'^{(k)}\right)\\ =&\Theta\left(\mathbf{C}_i'^{(k+1)}, \mathcal{G}\right)\\ =&\Theta\left(\mathbf{C}_i'^{(k+1)}, \mathcal{G} \mid x_{lq}'^{(k+1)} = x_{lq}'^{(k)} - \eta\frac{\partial\Theta(\mathbf{C}_i'^{(k)}, \mathcal{G})}{\partial x_{lq}'^{(k)}},\right.\\ &\left.\forall x_{lq} \in \mathbf{C}_i \cap M\right).\end{aligned}$$

Therefore, for the k-th round update in Algorithm 2, referring to Line 4 in Algorithm 1, it follows

$$\begin{aligned}&\Theta(\mathcal{C}'^{(k+1)}, \mathcal{G})\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)} - \eta\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}},\right.\\ &\left.\forall x_{ij} \in M\right)\\ =&\Theta\left(\mathbf{C}_1'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)} - \eta\frac{\partial\Theta(\mathbf{C}_1'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}},\right.\\ &\left.\forall x_{ij} \in \mathbf{C}_1 \cap M\right)\\ &+ \dots\\ &+ \Theta\left(\mathbf{C}_u'^{(k+1)}, \mathcal{G} \mid x_{ij}'^{(k+1)} = x_{ij}'^{(k)} - \eta\frac{\partial\Theta(\mathbf{C}_u'^{(k)}, \mathcal{G})}{\partial x_{ij}'^{(k)}},\right.\\ &\left.\forall x_{ij} \in \mathbf{C}_u \cap M\right)\\ &+ \Theta\left(\mathcal{C}' \setminus \mathcal{C}_m', \mathcal{G}\right).\end{aligned}$$

That is, PCDI Algorithm 2 returns the same result $\mathcal{C}'$ with CDI Algorithm 1 for fixed updates.

D. Proof of Proposition 4

Given

$$\eta \leq min\left\{\epsilon / \left\|\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial x_{ij}'^{(k)}}\right\|, \epsilon / \left\|\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}}\right\|\right\},$$

which indicates that

$$\left\|-\eta\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial x_{ij}'^{(k)}}\right\| \leq \epsilon$$

and

$$\left\|-\eta\frac{\partial\Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}}\right\| \leq \epsilon.$$

According to the proof of Proposition 1, for $\left\| -\eta \frac{\partial \Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial x_{ij}'^{(k)}} \right\| \leq \epsilon$, it leads to

$$\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k)}\right) \leq \Theta\left(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)}\right).$$

Following Line 6 in Algorithm 3, we have

$$\begin{aligned}&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k+1)}\right)\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k+1)} \mid \Phi_{\mathcal{G}'}^{(k+1)} = \Phi_{\mathcal{G}'}^{(k)} - \eta \frac{\partial \Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}}\right).\end{aligned}$$

Moreover, according to the first-order Taylor expansion, given $\left\| -\eta \frac{\partial \Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}} \right\| \leq \epsilon$, it further has

$$\begin{aligned}&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k+1)} \mid \Phi_{\mathcal{G}'}^{(k+1)} = \Phi_{\mathcal{G}'}^{(k)} - \eta \frac{\partial \Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}}\right)\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k+1)} \mid \Phi_{\mathcal{G}'}^{(k+1)} = \Phi_{\mathcal{G}'}^{(k)}\right)\\ &+ \left\| -\eta \frac{\partial \Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}} \right\| \cdot \frac{\partial \Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}}\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k+1)} \mid \Phi_{\mathcal{G}'}^{(k+1)} = \Phi_{\mathcal{G}'}^{(k)}\right)\\ &- \eta \left\| \frac{\partial \Theta(\mathcal{C}'^{(k)}, \mathcal{G}'^{(k)})}{\partial \Phi_{\mathcal{G}'}^{(k)}} \right\|^2\\ \leq&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k+1)} \mid \Phi_{\mathcal{G}'}^{(k+1)} = \Phi_{\mathcal{G}'}^{(k)}\right)\\ =&\Theta\left(\mathcal{C}'^{(k+1)}, \mathcal{G}'^{(k)}\right).\end{aligned}$$

E. Proof of Lemma 5

According to Formula 5, we know

$$\begin{aligned}&\mathbb{E}\|\Phi_{\mathcal{G}'}^{(\kappa)} - \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}\|^2\\ =& \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)} - \eta \sum_{t=0}^{\tau_\kappa - 1} \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa+t-\tau_{\kappa-\tau_\kappa+t})})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa+t-\tau_{\kappa-\tau_\kappa+t})}} \right.\\ &\left. - \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)} \right\|^2.\end{aligned}$$

Combining with Assumptions 1.2 and 1.3, it further leads to

$$\begin{aligned}&\mathbb{E}\|\Phi_{\mathcal{G}'}^{(\kappa)} - \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}\|^2\\ =&\eta^2 \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \sum_{t=0}^{\tau_\kappa - 1} \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa+t-\tau_{\kappa-\tau_\kappa+t})})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa+t-\tau_{\kappa-\tau_\kappa+t})}} \right\|^2\\ \overset{Assumption\ 1.3}{\leqslant}& \eta^2 \left\| \tau_\kappa V \right\|^2\\ =&\tau_k^2 \eta^2 V^2\\ \overset{Assumption\ 1.2}{\leqslant}& T^2 \eta^2 V^2.\end{aligned}$$

F. Proof of Proposition 6

We start from Formula 5, combining with Line 9 in Algorithm 4 and Proposition 1, it has

$$\begin{aligned}&\mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa+1)}\right)\\ =& \mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa+1)} \mid \Phi_{\mathcal{G}'}^{(\kappa+1)} = \Phi_{\mathcal{G}'}^{(\kappa)} - \eta \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}}\right).\end{aligned}$$

Then, according to Assumption 1.1, we have

$$\begin{aligned}&\mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa+1)}\right)\\ \overset{Assumption\ 1.1}{\leqslant}& \mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa)}\right)\\ &- \eta \mathbb{E}_{P_j \sim \mathbf{P}} \left\langle \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa)}}, \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\rangle\\ &+ \frac{L\eta^2}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\|^2\\ =&\mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa)}\right)\\ &+ \frac{\eta}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa)}} - \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\|^2\\ &- \frac{\eta}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa)}} \right\|^2\\ &- \frac{\eta}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\|^2\\ &+ \frac{L\eta^2}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\|^2.\end{aligned}$$

Combining with Assumption 1.3, we can obtain

$$\begin{aligned}&\mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa+1)}\right)\\ \leq&\mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa)}\right)\\ &+ \frac{\eta}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa)}} - \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\|^2\\ &- \frac{\eta}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa)}} \right\|^2\\ &+ \frac{L\eta^2}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\|^2\\ \overset{Assumption\ 1.3}{\leqslant}& \mathbb{E}_{P_j \sim \mathbf{P}} \Theta\left(\mathcal{C}_j', \mathcal{G}'^{(\kappa)}\right)\\ &+ \frac{\eta}{2} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa)}} - \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa-\tau_\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}} \right\|^2\\ &- \frac{\eta}{2} \mathbb{E}_{P_j \sim \mathbf{P}} \left\| \frac{\partial \Theta(\mathcal{C}_j', \mathcal{G}'^{(\kappa)})}{\partial \Phi_{\mathcal{G}'}^{(\kappa)}} \right\|^2 + \frac{L\eta^2 V^2}{2}.\end{aligned}$$

Moreover, according to Assumption 1.1, it has

$$\begin{aligned}
&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'_j,\mathcal{G}'^{(\kappa+1)}\right)\\
\overset{Assumption\ 1.1}{\leqslant}&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'_j,\mathcal{G}'^{(\kappa)}\right)\\
&+\frac{\eta L^2}{2}\mathbb{E}\|\Phi_{\mathcal{G}'}^{(\kappa)}-\Phi_{\mathcal{G}'}^{(\kappa-\tau_\kappa)}\|^2\\
&-\frac{\eta}{2}\mathbb{E}_{P_j\sim\mathbf{P}}\left\|\frac{\partial\Theta(\mathcal{C}'_j,\mathcal{G}'^{(\kappa)})}{\partial\Phi_{\mathcal{G}'}^{(\kappa)}}\right\|^2+\frac{L\eta^2V^2}{2}.
\end{aligned}$$

Referring to Lemma 5, it follows

$$\begin{aligned}
&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'_j,\mathcal{G}'^{(\kappa+1)}\right)\\
\overset{Lemma\ 5}{\leqslant}&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'_j,\mathcal{G}'^{(\kappa)}\right)\\
&-\frac{\eta}{2}\mathbb{E}_{P_j\sim\mathbf{P}}\left\|\frac{\partial\Theta(\mathcal{C}'_j,\mathcal{G}'^{(\kappa)})}{\partial\Phi_{\mathcal{G}'}^{(\kappa)}}\right\|^2\\
&+\frac{\eta L^2}{2}T^2\eta^2V^2+\frac{L\eta^2V^2}{2}.
\end{aligned}$$

For all processors, according to Definition 3 for the imputation cost, they have

$$\begin{aligned}
&\sum_{j=1}^{o}\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'_j,\mathcal{G}'^{(\kappa+1)}\right)\\
=&\mathbb{E}\Theta\left(\mathcal{C}',\mathcal{G}'^{(\kappa+1)}\right)\\
\leqslant&\sum_{j=1}^{o}\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'_j,\mathcal{G}'^{(\kappa)}\right)\\
&-\frac{\eta}{2}\sum_{j=1}^{o}\mathbb{E}_{P_j\sim\mathbf{P}}\left\|\frac{\partial\Theta(\mathcal{C}'_j,\mathcal{G}'^{(\kappa)})}{\partial\Phi_{\mathcal{G}'}^{(\kappa)}}\right\|^2\\
&+\frac{\eta L^2}{2}T^2\eta^2V^2o+\frac{L\eta^2V^2}{2}o\\
=&\mathbb{E}\Theta\left(\mathcal{C}',\mathcal{G}'^{(\kappa)}\right)\\
&-\frac{\eta}{2}\mathbb{E}\left\|\frac{\partial\Theta(\mathcal{C}',\mathcal{G}'^{(\kappa)})}{\partial\Phi_{\mathcal{G}'}^{(\kappa)}}\right\|^2\\
&+\frac{\eta L^2}{2}T^2\eta^2V^2o+\frac{L\eta^2V^2}{2}o.
\end{aligned}$$

Summing from $\kappa=0$ to $\kappa=K-1$, we can obtain

$$\begin{aligned}
&\mathbb{E}\Theta\left(\mathcal{C}',\mathcal{G}'^{(K)}\right)\\
\leqslant&\mathbb{E}\Theta\left(\mathcal{C}',\mathcal{G}'^{(0)}\right)-\frac{\eta}{2}\sum_{\kappa=0}^{K-1}\mathbb{E}\left\|\frac{\partial\Theta(\mathcal{C}',\mathcal{G}'^{(\kappa)})}{\partial\Phi_{\mathcal{G}'}^{(\kappa)}}\right\|^2\\
&+\frac{\eta^3L^2T^2V^2Ko}{2}+\frac{L\eta^2V^2Ko}{2}.
\end{aligned}$$

Considering the Definition 3 for the imputation cost, it leads to

$$\mathbb{E}\Theta\left(\mathcal{C}',\mathcal{G}'^{(K)}\right)\geqslant 0.$$

Finally, we can obtain the conclusion

$$\begin{aligned}
\sum_{\kappa=0}^{K-1}\mathbb{E}\left\|\frac{\partial\Theta(\mathcal{C}',\mathcal{G}'^{(\kappa)})}{\partial\Phi_{\mathcal{G}'}^{(\kappa)}}\right\|^2\leqslant&\frac{2\Theta\left(\mathcal{C}',\mathcal{G}'^{(0)}\right)}{\eta}\\
&+\eta oLKV^2(\eta LT^2+1).
\end{aligned}$$

G. Proof of Proposition 7

We start from Line 9 in Algorithm 4, it has

$$\begin{aligned}
&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(k_j+1)}_j,\mathcal{G}'\right)\\
=&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(k_j+1)}_j,\mathcal{G}'\mid\right.\\
&\left.x'^{(k_j+1)}_{ij}=x'^{(k_j)}_{ij}-\eta\frac{\partial\Theta(\mathcal{C}'^{(k_j)}_j,\mathcal{G}')}{\partial x'^{(k_j)}_{ij}},\forall x_{ij}\in\mathcal{C}_j\right).
\end{aligned}$$

Given $\eta\leq\epsilon/\left\|\frac{\partial\Theta(\mathcal{C}'^{(k_j)}_j,\mathcal{G}')}{\partial x'^{(k_j)}_{ij}}\right\|$, referring to the first-order Taylor expansion [45], it follows

$$\begin{aligned}
&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(k_j+1)}_j,\mathcal{G}'\right)\\
=&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(k_j)}_j,\mathcal{G}'\right)-\eta\sum_{x_{ij}\in\mathcal{C}_j}\left\|\frac{\partial\Theta(\mathcal{C}'^{(k_j)}_j,\mathcal{G}')}{\partial x'^{(k_j)}_{ij}}\right\|^2.
\end{aligned}$$

If the processor P_j updates the parameters κ_j times in total, summing from $k_j=0$ to $k_j=\kappa_j-1$, we can obtain

$$\begin{aligned}
&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(\kappa_j)}_j,\mathcal{G}'\right)\\
=&\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(0)}_j,\mathcal{G}'\right)-\eta\sum_{k_j=0}^{\kappa_j-1}\sum_{x_{ij}\in\mathcal{C}_j}\left\|\frac{\partial\Theta(\mathcal{C}'^{(k_j)}_j,\mathcal{G}')}{\partial x'^{(k_j)}_{ij}}\right\|^2.
\end{aligned}$$

Considering Definition 3 for the imputation cost, we know

$$\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(\kappa_j)}_j,\mathcal{G}'\right)\geqslant 0.$$

It leads to

$$\eta\sum_{k_j=0}^{\kappa_j-1}\sum_{x_{ij}\in\mathcal{C}_j}\left\|\frac{\partial\Theta(\mathcal{C}'^{(k_j)}_j,\mathcal{G}')}{\partial x'^{(k_j)}_{ij}}\right\|^2\leq\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(0)}_j,\mathcal{G}'\right).$$

For all processors, we can obtain the following conclusion according to Definition 3:

$$\begin{aligned}
&\sum_{j=1}^{o}\sum_{k_j=0}^{\kappa_j-1}\sum_{x_{ij}\in\mathcal{C}_j}\left\|\frac{\partial\Theta(\mathcal{C}'^{(k_j)}_j,\mathcal{G}')}{\partial x'^{(k_j)}_{ij}}\right\|^2\\
\leq&\frac{1}{\eta}\sum_{j=1}^{o}\mathbb{E}_{P_j\sim\mathbf{P}}\Theta\left(\mathcal{C}'^{(0)}_j,\mathcal{G}'\right)\\
=&\frac{\Theta\left(\mathcal{C}'^{(0)},\mathcal{G}'\right)}{\eta}.
\end{aligned}$$

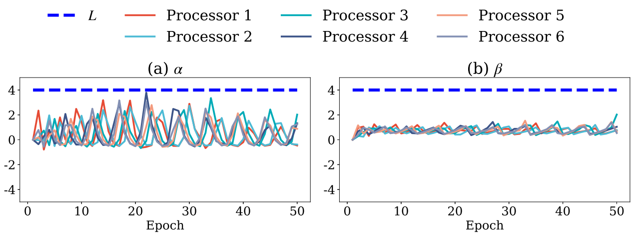

Fig. 14. L-smooth items over Energy dataset with 10% missing values

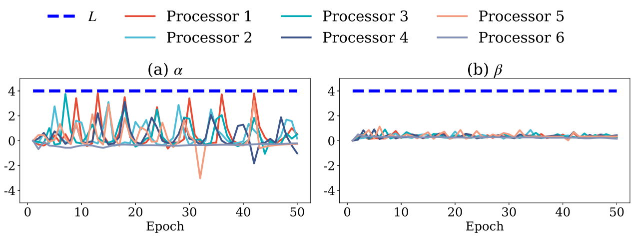

Fig. 15. L-smooth items over AirQuality dataset with 20% missing values

H. Empirical Evidence of L-smooth Assumption

The intuition of Assumption 1.1 is that the gradient of the imputation cost will not be too steep. In this section, we explore the empirical evidence for this assumption over multiple datasets.

Let us consider the two items in Assumption 1.1, the first one is

$$\Theta(\mathcal{C}'_j, \mathcal{G}'^{(b)}) \leqslant \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)}) + \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)})}{\partial \Phi^{(a)}_{\mathcal{G}'}} (\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}) + \frac{L}{2} \|\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}\|^2,$$

and the second one is

$$\left\| \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(b)})}{\partial \Phi^{(b)}_{\mathcal{G}'}} - \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)})}{\partial \Phi^{(a)}_{\mathcal{G}'}} \right\| \leqslant L \|\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}\|.$$

We derive the following inequation from the first condition

$$\left[\frac{2\left(\Theta(\mathcal{C}'_j, \mathcal{G}'^{(b)}) - \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)})\right)}{\|\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}\|^2} - \frac{2\left(\frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)})}{\partial \Phi^{(a)}_{\mathcal{G}'}} (\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'})\right)}{\|\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}\|^2} \right] \leqslant L.$$

Similarly, we can obtain the inequation from the second condition as follows,

$$\frac{\left\| \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(b)})}{\partial \Phi^{(b)}_{\mathcal{G}'}} - \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)})}{\partial \Phi^{(a)}_{\mathcal{G}'}} \right\|}{\|\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}\|} \leqslant L.$$

Let $\alpha = \frac{2\left(\Theta(\mathcal{C}'_j, \mathcal{G}'^{(b)}) - \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)}) - \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)})}{\partial \Phi^{(a)}_{\mathcal{G}'}} (\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'})\right)}{\|\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}\|^2}$, $\beta = \frac{\left\| \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(b)})}{\partial \Phi^{(b)}_{\mathcal{G}'}} - \frac{\partial \Theta(\mathcal{C}'_j, \mathcal{G}'^{(a)})}{\partial \Phi^{(a)}_{\mathcal{G}'}} \right\|}{\|\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}\|}$. To verify whether the two items of L-smooth are widely applicable to the imputation cost of various datasets, we present α and β of the PCDIH algorithm with six processors trained for 50 epochs.

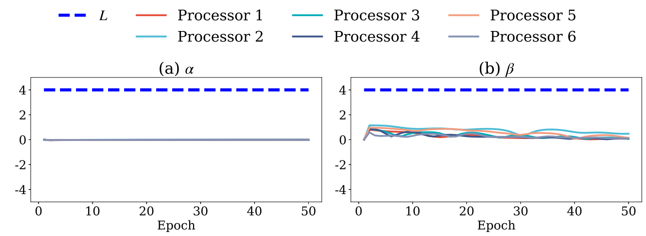

Fig. 16. L-smooth items over Ethanol dataset with 10% missing values

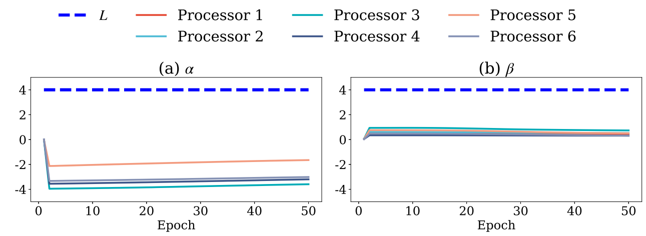

Fig. 17. L-smooth items over MIMIC-III dataset with 21.84% missing values

As shown in Figures 14-20, it is easy to find an L (e.g. 4 for Energy, AirQuality, Ethanol, MIMIC-III, GPS, 20 for IMU, 130 for Weer) satisfying the two items of L-smooth for the training process of the PCDIH model on various datasets. Note that since Ethanol and MIMIC-III datasets are preprocessed with Z-score normalization, as shown in Figure 16 and Figure 17, α and β values of these two datasets are smoother.

I. Memory Consumption

Table V presents the experimental results of the memory consumption for different methods over various datasets with different missing rates. According to the memory consumption in Table V and time costs in Table III, our methods have the reasonable scalability on large-scale Ethanol and Weer datasets, which demonstrate the contributions of our parallel imputation strategies.

J. Imputation Contexts

We report the number of maximal self-connected contexts over various datasets with different missing rates in Table VI. It can be noticed that the vast majority of scenarios have multiple maximal self-connected contexts, demonstrating the rationality of our parallel algorithms. In addition, there are two maximal self-connected contexts in the IMU dataset with real missing values, since the main source of these missing data is that the recording starts with all the inertial measurement unit sensors closed, leading to consecutive incomplete tuples. This condition will increase the difficulty of data imputation while also affecting the efficiency of the parallel imputation. However, our algorithms can still be more efficient than most baselines in Tables III and V, which demonstrates the efficiency of our algorithms, even for the real-world incomplete datasets containing few maximal self-connected contexts.

K. Imputation Rounds

For all the experiments, we repeat the procedure five times under different random seeds over various datasets with

TABLE V

IMPUTATION MEMORY CONSUMPTION (IN MIB) OF THE APPROACHES OVER VARIOUS DATASETS WITH DIFFERENT MISSING RATES

Approach	Energy					Ethanol					AirQuality				MIMIC-III			GPS	IMU	Weer
	10%	20%	30%	40%	80%	10%	20%	30%	40%	80%	20%	30%	40%	80%	30%	40%	80%	42.98%	0.46%	0.08%
BTMF	217.0	233.8	217.1	220.7	236.8	194.4	194.4	194.4	194.6	194.8	216.1	217.4	210.8	214.8	246.2	249.0	279.5	199.7	228.7	531.1
RecovDB	93.5	93.5	94.9	95.3	94.8	137.4	143.0	148.6	151.7	168.0	86.6	86.7	86.6	88.4	103.4	109.8	109.8	78.9	83.3	143.7
ORBITS	133.1	143.4	153.6	143.4	122.9	163.8	153.6	184.3	174.1	143.4	61.4	92.2	143.4	92.2	61.4	71.7	184.3	51.2	122.9	235.5
BayOTIDE	1130.0	1130.3	1130.4	1130.4	1096.0	1084.9	1084.7	1087.7	1080.8	981.2	715.4	715.9	716.0	703.2	3441.9	3416.0	3097.7	463.0	889.3	4382.7
BRITS	2813.4	2817.0	2818.2	2817.9	2818.9	2855.3	2859.2	2861.7	2864.6	2903.0	2813.5	2813.4	2815.0	2815.5	2830.1	2830.5	2835.3	2810.3	2811.6	3163.8
SAITS	2570.3	2575.0	2573.9	2573.7	2576.8	2570.2	2636.1	2643.6	2648.0	2648.0	2566.0	2567.8	2571.9	2570.2	2588.2	2591.8	2592.5	2562.0	2562.3	2568.5
GRIN	5312.5	5319.1	5321.2	5319.4	5320.4	5399.0	5399.9	5401.4	5400.4	5398.2	5321.3	5320.7	5321.2	5325.5	5334.8	5332.0	5336.7	5306.8	5309.4	5413.4
DAMR	1147.0	1168.4	1145.9	1147.4	1169.7	2295.9	2318.8	2306.7	2310.1	2295.1	874.4	875.3	868.2	868.8	1235.4	1238.5	1239.6	634.0	719.1	38863.1
E2GAN	3274.7	3272.6	3278.0	3278.7	3282.4	3317.7	3322.0	3324.0	3327.8	3363.0	3274.1	3267.8	3270.3	3275.7	3293.3	3296.6	3293.6	3264.2	3264.2	1248.9
MIWAE	2225.7	2227.4	2228.9	2227.4	2230.6	2316.2	2322.6	2326.9	2330.4	2347.0	2219.1	2221.8	2221.4	2224.0	2250.0	2250.3	2252.4	2217.9	2221.3	5413.4
CSDI	2411.2	2416.6	2418.0	2419.0	2420.3	2482.8	2487.1	2495.0	2499.3	2515.5	2411.9	2412.4	2416.3	2415.4	2553.9	2543.9	2536.3	2406.7	2409.5	2611.7
PriSTI	2562.0	2560.6	2561.8	2560.8	2564.2	2601.3	2602.5	2602.2	2599.5	2601.4	2557.2	2558.5	2558.1	2558.4	2568.1	2567.3	2568.7	2548.8	2553.9	2663.7
PCDI	393.9	402.6	413.9	442.6	517.7	317.1	320.4	326.6	329.0	399.1	387.5	523.4	523.3	702.1	857.1	994.2	1088.9	303.2	294.1	436.1
PCDIH	497.4	521.8	583.6	693.3	762.5	319.7	316.2	318.8	323.6	414.3	393.4	525.6	546.1	721.5	948.5	1003.2	1092.3	303.7	295.7	441.8

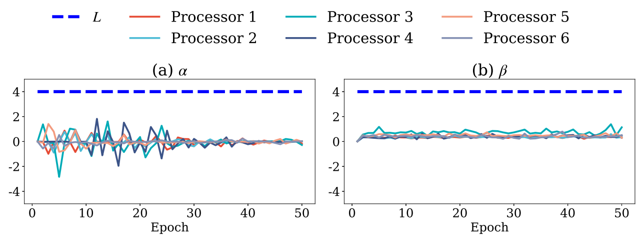

Fig. 18. L-smooth items over GPS dataset with 42.98% missing values

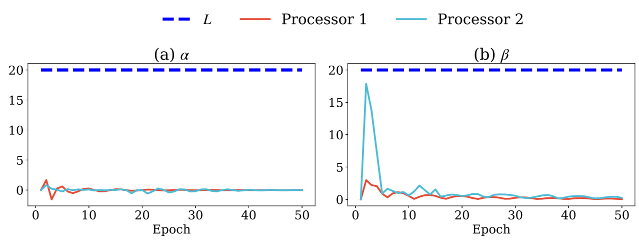

Fig. 19. L-smooth items over IMU dataset with 0.46% missing values

different missing rates, and we report the average results. Moreover, we add the new Table VII to detail the average number of imputation rounds under different settings for our methods. The imputation rounds are mainly influenced by several factors, such as the distribution of imputation contexts in different processors, temporal window size ω, dependency models $\mathcal{G}$ and the learning rate η. For instance, when using parallel algorithms, the rounds required for imputation on different processors may vary due to the differing difficulty levels in filling the incomplete imputation contexts, leading to variations in the number of rounds on different processors. Additionally, when the imputation temporal window size ω or the number of dependent models $\mathcal{G}$ increases, more intertemporal dependencies are considered, which may increase the number of rounds due to the rise in the task complexity. Similarly, when the learning rate η is too small, more imputation rounds are required to the convergence.

L. Intratemporal and Intertemporal Patterns

To illustrate the relationships among attribute values, we show all the intratemporal and intertemporal patterns in each dataset next. Specifically, to produce the dependency models of intratemporal and intertemporal patterns, for each attribute value, we extract the complete imputation contexts that can model the intertemporal dependency on this attribute over time and the intratemporal dependency between all the attribute values at the same time, following the process shown in Figure

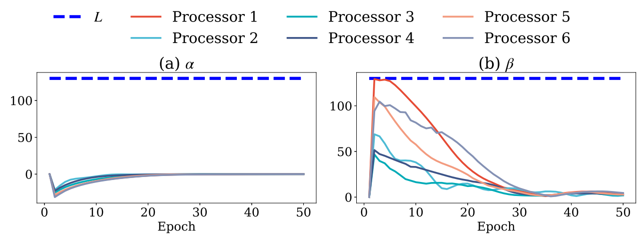

Fig. 20. L-smooth items over Weer dataset with 0.08% missing values

TABLE VI

NUMBER OF MAXIMAL SELF-CONNECTED CONTEXTS OVER VARIOUS DATASETS WITH DIFFERENT MISSING RATES

Dataset	Missing rate	Number of maximal self-connected contexts
Energy	10%	1495.80
	20%	995.60
	30%	431.60
	40%	138.60
	80%	1
Ethanol	10%	90.84
	20%	133.37
	30%	140.51
	40%	122.57
	80%	4.79
AirQuality	20%	26.80
	30%	7.20
	40%	4
	80%	1
MIMIC-III	30%	2552.4
	40%	1164.2
	80%	1
GPS	42.98%	110
IMU	0.46%	2
Weer	0.08%	195

3. We then train the dependency models using the gradient descent method over the complete imputation contexts for each attribute in turn. After training, the weights between different attributes at the same timestamp in the dependency model is associated with the intratemporal dependency and the intertemporal dependency is related to the weights between different timestamps for the same attribute, as defined in the previous works [63], [12].

As we can see below, for each dependency model, there typically exist different weights associated with different attribute values or the same attribute values with different timestamps, and the attribute values with larger weights contribute more to determine the predicted value.

Energy collects the sensor readings with nine attributes, i.e., T1 (A_1), T2 (A_2), T3 (A_3), T4 (A_4), T5 (A_5), T6 (A_6), T7 (A_7), T8 (A_8) and T9 (A_9). Nine corresponding dependency

TABLE VII

AVERAGE IMPUTATION ROUNDS OF OUR METHODS OVER VARIOUS DATASETS WITH DIFFERENT MISSING RATES

Approach	Energy					Ethanol					AirQuality				MIMIC-III			GPS	IMU	Weer
	10%	20%	30%	40%	80%	10%	20%	30%	40%	80%	20%	30%	40%	80%	30%	40%	80%	42.98%	0.46%	0.08%
CDI	16.20	47.80	108.00	191.60	272.80	49.50	66.26	87.64	105.80	139.37	33.20	21.20	59.40	84.00	93.69	92.78	107.79	24.00	11.00	80.00
CDIH	18.00	124.20	350.80	581.60	615.40	50.00	68.30	84.72	109.78	122.24	32.60	33.00	51.00	73.80	95.78	96.66	108.33	30.00	6.00	83.00
PCDI	21.50	52.10	109.50	191.87	272.80	52.06	69.70	90.98	109.12	139.37	27.50	21.93	62.50	84.00	125.27	130.77	107.79	48.80	8.00	74.16
PCDIH	22.00	127.10	348.43	581.47	615.40	52.48	71.67	87.71	112.47	122.24	31.37	36.07	46.23	73.80	133.74	143.24	108.33	70.16	12.50	89.00

models (in the form of Formula 1) are listed below.

$$\begin{aligned}
x_{i1} =&(4.27e-3)x_{i2} + (1.61e-3)x_{i3} + (5.34e-4)x_{i4} \\
&- (8.10e-4)x_{i5} + (1.47e-3)x_{i6} + (1.13e-3)x_{i7} \\
&+ (1.45e-3)x_{i8} - (6.66e-4)x_{i9} + (6.20e-1)x_{i-1,1} \\
&+ (3.79e-1)x_{i+1,1} + (4.00e-4) \\
x_{i2} =&-(3.70e-3)x_{i1} + (1.61e-3)x_{i3} + (1.05e-3)x_{i4} \\
&+ (1.53e-4)x_{i5} - (2.37e-4)x_{i6} - (2.47e-4)x_{i7} \\
&+ (9.65e-4)x_{i8} + (2.99e-3)x_{i9} + (5.57e-1)x_{i-1,2} \\
&+ (4.49e-1)x_{i+1,2} + (1.50e-3) \\
x_{i3} =&-(2.62e-3)x_{i1} + (5.83e-4)x_{i2} - (1.00e-3)x_{i4} \\
&- (2.25e-3)x_{i5} - (9.82e-4)x_{i6} + (1.16e-5)x_{i7} \\
&- (8.67e-4)x_{i8} - (1.77e-3)x_{i9} + (5.95e-1)x_{i-1,3} \\
&+ (4.05e-1)x_{i+1,3} - (7.00e-4) \\
x_{i4} =&(-2.12e-3)x_{i1} + (2.20e-3)x_{i2} + (1.00e-3)x_{i3} \\
&+ (1.37e-3)x_{i5} + (1.36e-3)x_{i6} - (-5.30e-4)x_{i7} \\
&+ (1.37e-3)x_{i8} + (5.98e-4)x_{i9} + (5.97e-1)x_{i-1,4} \\
&+ (4.08e-1)x_{i+1,4} + (1.00e-3) \\
x_{i5} =&-(1.96e-3)x_{i1} - (7.71e-4)x_{i2} - (1.36e-3)x_{i3} \\
&- (9.06e-4)x_{i4} - (3.24e-4)x_{i6} - (4.67e-4)x_{i7} \\
&- (1.23e-3)x_{i8} - (2.89e-3)x_{i9} + (5.42e-1)x_{i-1,5} \\
&+ (4.60e-1)x_{i+1,5} + (8.00e-4) \\
x_{i6} =&-(1.97e-3)x_{i1} + (3.04e-3)x_{i2} + (2.55e-4)x_{i3} \\
&+ (1.03e-3)x_{i4} - (2.79e-4)x_{i5} - (1.90e-4)x_{i7} \\
&- (5.95e-4)x_{i8} + (5.48e-3)x_{i9} + (5.79e-1)x_{i-1,6} \\
&+ (4.23e-1)x_{i+1,6} + (1.80e-3) \\
x_{i7} =&-(1.65e-3)x_{i1} + (3.19e-4)x_{i2} + (3.19e-4)x_{i3} \\
&- (1.57e-4)x_{i4} - (1.36e-3)x_{i5} + (5.36e-4)x_{i6} \\
&- (3.92e-4)x_{i8} - (1.59e-3)x_{i9} + (6.15e-1)x_{i-1,7} \\
&+ (3.84e-1)x_{i+1,7} + (6.00e-4) \\
x_{i8} =&-(1.16e-3)x_{i1} + (1.99e-4)x_{i2} - (9.50e-4)x_{i3} \\
&- (5.73e-4)x_{i4} - (2.00e-3)x_{i5} + (9.76e-4)x_{i6} \\
&- (5.35e-4)x_{i7} - (2.67e-3)x_{i9} + (6.14e-1)x_{i-1,8} \\
&+ (3.85e-1)x_{i+1,8} + (1.30e-3) \\
x_{i9} =&(2.12e-3)x_{i1} + (5.83e-4)x_{i2} + (4.64e-4)x_{i3} \\
&+ (4.14e-4)x_{i4} - (3.64e-4)x_{i5} + (8.49e-4)x_{i6} \\
&+ (5.77e-4)x_{i7} + (1.53e-5)x_{i8} + (6.14e-1)x_{i-1,9} \\
&+ (3.90e-1)x_{i+1,9} + (9.00e-4)
\end{aligned}$$

Ethanol contains first measurement readings (A_1), second measurement readings (A_2) and third measurement readings (A_3), representing the raw spectral time series of water and ethanol solutions in whisky bottles. Please see the full list of dependency models on Ethanol below.

$$\begin{aligned}
x_{i1} =&0.24x_{i2} + 0.17x_{i3} - 0.02x_{i-1,1} + 0.62x_{i+1,1} - 0.01 \\
x_{i2} =&0.24x_{i1} - 0.05x_{i3} + 0.06x_{i-1,2} + 0.75x_{i+1,2} + 0.01 \\
x_{i3} =&0.10x_{i1} - 0.10x_{i2} + 0.22x_{i-1,3} + 0.77x_{i+1,3} - (1.36e-5)
\end{aligned}$$

AirQuality has a schema over CO (A_1), PT08.S1 (A_2), NMHC (A_3), C6H6 (A_4), PT08.S2 (A_5), NOx (A_6), PT08.S3 (A_7), NO2 (A_8), PT08.S4 (A_9), PT08.S5 (A_{10}), T (A_{11}), RH (A_{12}) and AH (A_{13}). The dependency models are below.

$$\begin{aligned}
x_{i1} =&0.16x_{i2} + 0.15x_{i3} + 0.13x_{i4} + 0.21x_{i5} + 0.23x_{i6} \\
&+ 0.22x_{i7} + 0.15x_{i8} - 0.07x_{i9} - 0.11x_{i10} - 0.21x_{i11} \\
&- 0.09x_{i12} + 0.19x_{i13} + 0.15x_{i-1,1} + 0.10x_{i+1,1} - 0.08 \\
x_{i2} =&0.30x_{i1} - 0.22x_{i3} + 0.11x_{i4} + 0.18x_{i5} + 0.02x_{i6} \\
&+ 0.11x_{i7} + 0.11x_{i8} + 0.55x_{i9} + 0.03x_{i10} - 0.14x_{i11} \\
&- 0.03x_{i12} + 0.06x_{i13} + 0.21x_{i-1,2} + 0.14x_{i+1,2} - 0.08 \\
x_{i3} =&1.24x_{i1} - 0.39x_{i2} + 1.97x_{i4} - 1.18x_{i5} - 0.80x_{i6} \\
&- 0.35x_{i7} - 0.13x_{i8} + 0.27x_{i9} - 0.08x_{i10} - 0.29x_{i11} \\
&- 0.09x_{i12} + 0.08x_{i13} + 0.20x_{i-1,3} + 0.13x_{i+1,3} + 0.42 \\
x_{i4} =&0.04x_{i1} - 0.03x_{i2} + 0.09x_{i3} + 0.46x_{i5} + 0.05x_{i6} \\
&+ 0.13x_{i7} - 0.03x_{i8} + 0.16x_{i9} + 0.05x_{i10} + 0.04x_{i11} \\
&+ 0.01x_{i12} - 0.08x_{i13} + 0.05x_{i-1,4} + 0.07x_{i+1,4} - 0.17 \\
x_{i5} =&0.17x_{i1} + 0.09x_{i2} - 0.06x_{i3} + 0.66x_{i4} - 0.11x_{i6} \\
&- 0.29x_{i7} - 0.04x_{i8} + 0.23x_{i9} - 0.01x_{i10} + 0.16x_{i11} \\
&+ 0.08x_{i12} - 0.26x_{i13} + 0.11x_{i-1,5} + 0.10x_{i+1,5} + 0.15 \\
x_{i6} =&0.14x_{i1} + 0.01x_{i2} - 0.04x_{i3} + 0.37x_{i4} - 0.28x_{i5} \\
&- 0.07x_{i7} - 0.03x_{i8} + 0.18x_{i9} - 0.01x_{i10} + 0.04x_{i11} \\
&+ 0.07x_{i12} - 0.18x_{i13} + 0.40x_{i-1,6} + 0.18x_{i+1,6} + 0.03 \\
x_{i7} =&0.28x_{i1} - 0.10x_{i2} - 0.07x_{i3} + 0.12x_{i4} - 0.48x_{i5} \\
&- 0.04x_{i6} - 0.06x_{i8} + 0.23x_{i9} - 0.02x_{i10} + 0.19x_{i11} \\
&+ 0.08x_{i12} - 0.27x_{i13} + 0.28x_{i-1,7} + 0.29x_{i+1,7} + 0.18 \\
x_{i8} =&0.18x_{i1} + 0.18x_{i2} + 0.04x_{i3} + 0.01x_{i4} + 0.04x_{i5} \\
&- 0.01x_{i6} + 0.10x_{i7} - 0.23x_{i9} + 0.10x_{i10} + 0.01x_{i11} \\
&- 0.01x_{i12} + 0.01x_{i13} + 0.41x_{i-1,8} + 0.38x_{i+1,8} - 0.01 \\
x_{i9} =&-0.05x_{i1} + 0.08x_{i2} + 0.03x_{i3} + 0.44x_{i4} + 0.15x_{i5} \\
&+ 0.13x_{i6} - 0.02x_{i7} - 0.18x_{i8} + 0.02x_{i10} + 0.25x_{i11} \\
&+ 0.1x_{i12} + 0.08x_{i13} + 0.03x_{i-1,9} + 0.08x_{i+1,9} + 0.03
\end{aligned}$$

$$
\begin{aligned}
x_{i10} =& -0.10x_{i1} + 0.19x_{i2} - 0.03x_{i3} + 0.53x_{i4} - 0.39x_{i5} \\
& - 0.06x_{i6} - 0.15x_{i7} + 0.05x_{i8} + 0.44x_{i9} - 0.41x_{i11} \\
& - 0.16x_{i12} + 0.09x_{i13} + 0.44x_{i-1,10} + 0.25x_{i+1,10} + 0.19 \\
x_{i11} =& 0.01x_{i1} - 0.01x_{i2} - 0.03x_{i3} - 0.05x_{i4} - 0.01x_{i5} \\
& - 0.09x_{i6} + 0.05x_{i7} + 0.07x_{i8} + 0.22x_{i9} - 0.05x_{i10} \\
& - 0.25x_{i12} + 0.29x_{i13} + 0.24x_{i-1,11} + 0.32x_{i+1,11} + 0.10 \\
x_{i12} =& -0.11x_{i1} - 0.08x_{i2} - 0.05x_{i3} + 0.03x_{i4} - 0.13x_{i5} \\
& + 0.10x_{i6} - 0.01x_{i7} + 0.05x_{i8} + 0.32x_{i9} - 0.03x_{i10} \\
& - 0.81x_{i11} + 0.56x_{i13} + 0.26x_{i-1,12} + 0.27x_{i+1,12} + 0.28 \\
x_{i13} =& 0.03x_{i1} + 0.07x_{i2} + 0.01x_{i3} + 0.8x_{i4} - 0.25x_{i5} \\
& - 0.02x_{i6} - 0.05x_{i7} - 0.03x_{i8} + 0.13x_{i9} + 0.03x_{i10} \\
& + 0.17x_{i11} + 0.01x_{i12} + 0.48x_{i-1,13} + 0.34x_{i+1,13}
\end{aligned}
$$

MIMIC-III records the daily data of patients, including Hours (A_1), Heart Rate (A_2), Mean blood pressure (A_3), Oxygen saturation (A_4), Respiratory rate (A_5) and Systolic blood pressure (A_6). Below is the list of dependency models.

$$
\begin{aligned}
x_{i1} =& 0.01x_{i2} - 0.02x_{i3} - 0.16x_{i4} + 0.01x_{i5} - 0.01x_{i6} \\
& + 0.47x_{i-1,1} + 0.54x_{i+1,1} + 0.10 \\
x_{i2} =& 0.02x_{i1} + 0.36x_{i3} + 0.05x_{i4} + 0.20x_{i5} - 0.27x_{i6} \\
& + 0.21x_{i-1,2} + 0.28x_{i+1,2} + 0.08 \\
x_{i3} =& -0.01x_{i1} + 0.17x_{i2} - 0.12x_{i4} - 0.04x_{i5} 0.83x_{i6} \\
& - 0.01x_{i-1,3} + 0.05x_{i+1,3} + 0.02 \\
x_{i4} =& 0.01x_{i1} + 0.07x_{i2} + 0.02x_{i3} - 0.04x_{i5} - 0.01x_{i6} \\
& + 0.15x_{i-1,4} + 0.35x_{i+1,4} + 0.26 \\
x_{i5} =& -0.02x_{i1} + 0.73x_{i2} - 0.44x_{i3} - 0.05x_{i4} + 0.31x_{i6} \\
& + 0.18x_{i-1,5} - 0.03x_{i+1,5} + 0.19 \\
x_{i6} =& 0.02x_{i1} - 0.16x_{i2} + 1.18x_{i3} + 0.18x_{i4} + 0.05x_{i5} \\
& + 0.01x_{i-1,6} - 0.04x_{i+1,6} - 0.08
\end{aligned}
$$

GPS consists of trajectory data with Sat-Lon (A_1), Sat-Lat (A_2), Map-Lon (A_3), Map-Lat (A_4), Altitude (A_5), Speed (A_6), hAccuracy (A_7) and vAccuracy (A_8) attributes. Similarly, eight dependency models are also listed below.

$$
\begin{aligned}
x_{i1} =& 0.05x_{i2} - 0.01x_{i3} - 0.05x_{i4} + 0.01x_{i5} - 0.01x_{i6} \\
& - 0.01x_{i7} - 0.01x_{i8} + 0.04x_{i-1,1} + 0.06x_{i+1,1} - 0.01 \\
x_{i2} =& -0.01x_{i1} - 0.01x_{i3} - 0.03x_{i4} - 0.01x_{i5} - 0.01x_{i6} \\
& 0.01x_{i7} + 0.01x_{i8} + 0.39x_{i-1,2} + 0.58x_{i+1,2} + 0.01 \\
x_{i3} =& 0.07x_{i1} + 0.14x_{i2} - 0.14x_{i4} - 0.14x_{i5} + 0.01x_{i6} \\
& - 0.01x_{i7} + 0.01x_{i8} + 0.39x_{i-1,3} + 0.58x_{i+1,3} - 0.04 \\
x_{i4} =& 0.01x_{i1} + 0.22x_{i2} - 0.06x_{i3} + 0.01x_{i5} - 0.01x_{i6} \\
& + 0.01x_{i7} + 0.01x_{i8} + 0.30x_{i-1,4} + 0.49x_{i+1,4} - 0.06 \\
x_{i5} =& 0.03x_{i1} + 0.03x_{i2} - 0.03x_{i3} - 0.02x_{i4} - 0.01x_{i6} \\
& - 0.01x_{i7} - 0.01x_{i8} + 0.53x_{i-1,5} + 0.48x_{i+1,5} - 0.01 \\
x_{i6} =& -0.03x_{i1} + 0.26x_{i2} - 0.06x_{i3} - 0.22x_{i4} - 0.13x_{i5} \\
& - 0.09x_{i7} - 0.03x_{i8} + 0.43x_{i-1,6} + 0.47x_{i+1,6} + 0.11 \\
x_{i7} =& -0.07x_{i1} + 0.05x_{i2} + 0.01x_{i3} - 0.04x_{i4} - 0.01x_{i5} \\
& - 0.02x_{i6} + 0.01x_{i8} + 0.47x_{i-1,7} + 0.46x_{i+1,7} + 0.07 \\
x_{i8} =& 0.02x_{i1} + 0.05x_{i2} - 0.01x_{i3} - 0.05x_{i4} - 0.01x_{i5} \\
& - 0.01x_{i6} + 0.01x_{i7} + 0.50x_{i-1,8} + 0.50x_{i+1,8} - 0.01
\end{aligned}
$$

IMU consists of the observations from an inertial measurement unit with ax_raw (A_1), ay_raw (A_2), az_raw (A_3), wx_raw (A_4), wy_raw (A_5), wz_raw (A_6). Below is a full list of dependency models over IMU dataset. Specially, we can observe that there is no strong intratemporal dependency between the different attributes, as they record data in different axes of the inertial measurement units. Therefore, it can be seen that our methods can self-adaptively learn small weights for other attributes (with parameters for $x_{i1}, \ldots, x_{i6}$) and mainly focus on intratemporal dependencies (with parameters for $x_{i-1,j}$, $x_{i+1,j}$) of the dependency model g_j.

$$
\begin{aligned}
x_{i1} =& -0.001x_{i2} + 0.001x_{i3} - 0.001x_{i4} - 0.003x_{i5} - 0.002x_{i6} \\
& + 0.415x_{i-1,1} + 0.585x_{i+1,1} + 0.003 \\
x_{i2} =& -0.001x_{i1} + 0.001x_{i3} + 0.002x_{i4} + 0.001x_{i5} + 0.001x_{i6} \\
& + 0.417x_{i-1,2} + 0.583x_{i+1,2} - 0.002 \\
x_{i3} =& 0.001x_{i1} + 0.001x_{i2} + 0.001x_{i4} - 0.001x_{i5} - 0.001x_{i6} \\
& + 0.479x_{i-1,3} + 0.521x_{i+1,3} - 0.001 \\
x_{i4} =& 0.001x_{i1} - 0.001x_{i2} - 0.001x_{i3} + 0.01x_{i5} + 0.01x_{i6} \\
& + 0.461x_{i-1,4} + 0.541x_{i+1,4} - 0.001 \\
x_{i5} =& 0.001x_{i1} - 0.001x_{i2} + 0.001x_{i3} - 0.001x_{i4} + 0.001x_{i6} \\
& + 0.417x_{i-1,5} + 0.584x_{i+1,5} - 0.001 \\
x_{i6} =& 0.001x_{i1} - 0.001x_{i2} - 0.001x_{i3} - 0.001x_{i4} - 0.001x_{i5} \\
& + 0.423x_{i-1,6} + 0.578x_{i+1,6} - 0.001
\end{aligned}
$$

Weer records weather data including Wind direction averaged over last 10 minutes of the past hour / Mean wind direction during the 10 minutes period preceding the time of observation (A_1), Hourly mean wind speed (A_2), Wind speed average over the last 10 minutes of the past hour / Mean wind direction during the 10 minutes period preceding the time of observation (A_3), Highest wind gust over the past hour / Maximum wind gust during the hourly division (A_4), Temperature (A_5), Please see the full list of dependency models below.

$$
\begin{aligned}
x_{i1} =& -0.10x_{i2} - 0.16x_{i3} + 0.22x_{i4} + 0.01x_{i5} \\
& + 0.32x_{i-1,1} + 0.32x_{i+1,1} + 0.07 \\
x_{i2} =& -0.01x_{i1} + 0.34x_{i3} + 0.27x_{i4} - 0.01x_{i5} \\
& + 0.25x_{i-1,2} + 0.12x_{i+1,2} - 0.01 \\
x_{i3} =& -0.01x_{i1} + 0.75x_{i2} + 0.13x_{i4} + 0.01x_{i5} \\
& - 0.05x_{i-1,3} + 0.19x_{i+1,3} - 0.01 \\
x_{i4} =& 0.01x_{i1} + 0.48x_{i2} + 0.06x_{i3} + 0.01x_{i5} \\
& + 0.25x_{i-1,4} + 0.22x_{i+1,4} - 0.01 \\
x_{i5} =& 0.01x_{i1} - 0.02x_{i2} + 0.03x_{i3} - 0.01x_{i4} \\
& + 0.5x_{i-1,5} + 0.5x_{i+1,5} - 0.01
\end{aligned}
$$