A PROOFS

A.1 Proof of Proposition 1

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

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

Given $\eta \leq \epsilon / \left\| \frac{\partial\Theta(C'^{(k)}, \mathcal{G})}{\partial x'^{(k)}_{ij}} \right\|$, i.e., $\left\| -\eta \frac{\partial\Theta(C'^{(k)}, \mathcal{G})}{\partial x'^{(k)}_{ij}} \right\| \leq \epsilon$, referring to the first-order Taylor expansion [40], it follows

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

A.2 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.

A.3 Proof of Proposition 3

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

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

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

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

Combing with Proposition 2, for any $\mathbf{C}_i, \mathbf{C}_j \in C_m$, 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 $\mathbf{C}_i, \mathbf{C}_j \in C_m$ at the k-th round update in Algorithm 2, they always hold

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

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

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

That is, PCIMD Algorithm 2 returns the same result C' with CIMD Algorithm 1 for fixed updates.

A.4 Proof of Proposition 4

Given

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

which indicates that

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

and

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

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

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

Following Line 6 in Algorithm 3, we have

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

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

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

A.5 Proof of Lemma 5

According to Formula 5, we know

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

Combining with Assumptions 1.2 and 1.3, it further leads to

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

A.6 Proof of Proposition 6

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

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

Then, according to Assumption 1.1, we have

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

Combining with Assumption 1.3, we can obtain

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

Moreover, according to Assumption 1.1, it has

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

Referring to Lemma 5, it follows

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

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

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

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

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

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

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

Finally, we can obtain the conclusion

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

A.7 Proof of Proposition 7

We start from Line 9 in Algorithm 4, it has

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

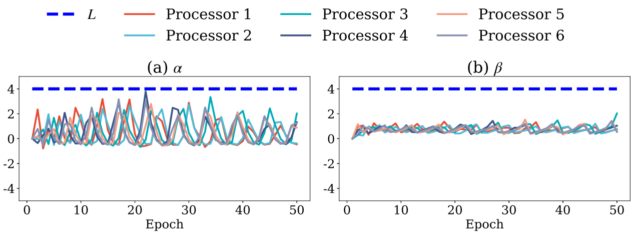

Figure 14: L-smooth items over Energy dataset with 10% missing values

Given $\eta \leq \epsilon / \left\| \frac{\partial \Theta(C_j'^{(k_j)}, \mathcal{G}')}{\partial x_{ij}'^{(k_j)}} \right\|$, referring to the first-order Taylor expansion [40], it follows

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

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

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

Considering Definition 3 for the imputation cost, we know

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

It leads to

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

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

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

A.8 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.

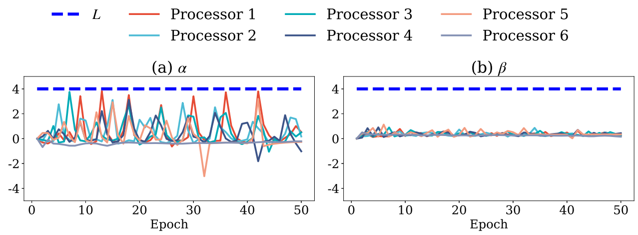

Figure 15: L-smooth items over AirQuality dataset with 20% missing values

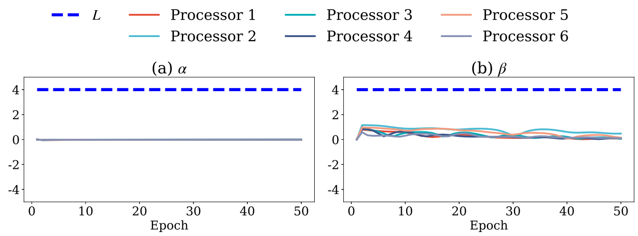

Figure 16: L-smooth items over Ethanol dataset with 10% missing values

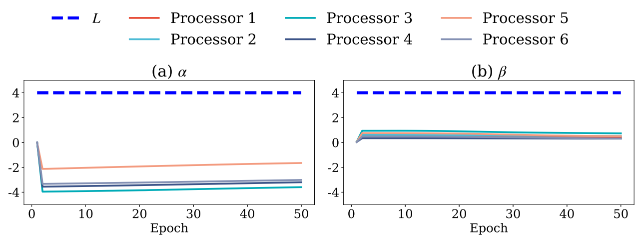

Figure 17: L-smooth items over MIMIC-III dataset with 21.84% missing values

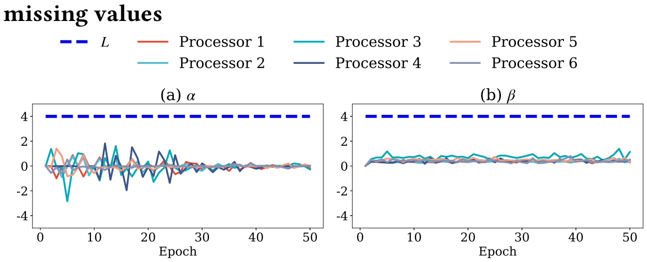

Figure 18: L-smooth items over GPS dataset with 42.98% missing values

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

$$\begin{aligned} \Theta(C_j', \mathcal{G}'^{(b)}) \leqslant &\Theta(C_j', \mathcal{G}'^{(a)}) + \frac{\partial \Theta(C_j', \mathcal{G}'^{(a)})}{\partial \Phi^{(a)}_{\mathcal{G}'}} (\Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'}) \\ &+ \frac{L}{2} \| \Phi^{(b)}_{\mathcal{G}'} - \Phi^{(a)}_{\mathcal{G}'} \|^2, \end{aligned}$$

and the second one is

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

We derive the following inequation from the first condition

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

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

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

Let $\alpha=\frac{2\left(\Theta(C'_j,\mathcal{G}'^{(b)})-\Theta(C'_j,\mathcal{G}'^{(a)})-\frac{\partial\Theta(C'_j,\mathcal{G}'^{(a)})}{\partial\Phi^{(a)}_{\mathcal{G}'}}(\Phi^{(b)}_{\mathcal{G}'}-\Phi^{(a)}_{\mathcal{G}'})\right)}{\|\Phi^{(b)}_{\mathcal{G}'}-\Phi^{(a)}_{\mathcal{G}'}\|^2}$, $\beta=\frac{\left\|\frac{\partial\Theta(C'_j,\mathcal{G}'^{(b)})}{\partial\Phi^{(b)}_{\mathcal{G}'}}-\frac{\partial\Theta(C'_j,\mathcal{G}'^{(a)})}{\partial\Phi^{(a)}_{\mathcal{G}'}}\right\|}{\|\Phi^{(b)}_{\mathcal{G}'}-\Phi^{(a)}_{\mathcal{G}'}\|}$. 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 PCIHMD algorithm with six processors trained for 50 epochs.

As shown in Figures 14-18, it is easy to find an L (e.g. 4) satisfying the two items of L-smooth for the training process of the PCIHMD 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.

A.9 Intratemporal and Intertemporal Patterns

To illustrate the relationships among attribute values, we show all the intertemporal and intertemporal patterns in each dataset next.

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

$$\begin{aligned}x_{i1}=&(4.27e-3)x_{i2}+(1.61e-3)x_{i3}+(5.34e-4)x_{i4}\\&-(8.10e-4)x_{i5}+(1.47e-3)x_{i6}+(1.13e-3)x_{i7}\\&+(1.45e-3)x_{i8}-(6.66e-4)x_{i9}+(6.20e-1)x_{i-1,1}\\&+(3.79e-1)x_{i+1,1}+(4.00e-4)\end{aligned}$$

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

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.

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

AirQuality has a schema consisting of 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 shown 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-1,1} + 0.14x_{i+1,1} - 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-1,1} + 0.13x_{i+1,1} + 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-1,1} + 0.07x_{i+1,1} - 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-1,1} + 0.10x_{i+1,1} + 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-1,1} + 0.18x_{i+1,1} + 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-1,1} + 0.29x_{i+1,1} + 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-1,1} + 0.38x_{i+1,1} - 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-1,1} + 0.08x_{i+1,1} + 0.03 \\
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-1,1} + 0.25x_{i+1,1} + 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-1,1} + 0.32x_{i+1,1} + 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-1,1} + 0.27x_{i+1,1} + 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-1,1} + 0.34x_{i+1,1}
\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 a full list of dependency models on MIMIC-III.

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

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.

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