Appendix: Proofs of the theorems

A.1 Proof of the generalization error bound

In this section, we prove Theorem 4.6, i.e., the generalization error bound. To this end, the covering number of the set of the values of the loss functions must be estimated. Although the data set is given as

$$(x_1^{(1)}, \dots, x_1^{(m_{\mathrm{step}})}, x_2^{(1)}, \dots, x_{m_{\mathrm{orbit}}}^{(m_{\mathrm{step}})}) \subset B^{m_{\mathrm{step}} \times m_{\mathrm{orbit}}},$$

from Assumption 4.1, these data are obtained by repeatedly applying the map ψ to the initial data $(x_1^{(1)}, \dots, x_{m_{\mathrm{orbit}}}^{(1)}) \subset B^{m_{\mathrm{orbit}}}$. This operation induces a map $\mathcal{O} : B^{m_{\mathrm{orbit}}} \to B^{m_{\mathrm{step}} \times m_{\mathrm{orbit}}}$. Using the Lipschitz constant of the map ψ, for each $(x_1^{(1)}, \dots, x_1^{(m_{\mathrm{step}})}, x_2^{(1)}, \dots, x_{m_{\mathrm{orbit}}}^{(m_{\mathrm{step}})}), (\tilde{x}_1^{(1)}, \dots, \tilde{x}_1^{(m_{\mathrm{step}})}, \tilde{x}_2^{(1)}, \dots, \tilde{x}_{m_{\mathrm{orbit}}}^{(m_{\mathrm{step}})})$, we have

$$\begin{aligned}
&\|(x_1^{(1)}, \dots, \dots, \dots, x_{m_{\mathrm{orbit}}}^{(m_{\mathrm{step}})}) - (\tilde{x}_1^{(1)}, \dots, \tilde{x}_{m_{\mathrm{orbit}}}^{(m_{\mathrm{step}})})\| \\
&= \|(x_1^{(1)}, \dots, \psi^{m_{\mathrm{step}}}(x_1^{(1)}), x_2^{(1)}, \dots, \psi^{(m_{\mathrm{step}})}(x_{m_{\mathrm{orbit}}}^{(1)})) - (\tilde{x}_1^{(1)}, \dots, \psi^{(m_{\mathrm{step}})}\tilde{x}_{m_{\mathrm{orbit}}}^{(1)}))\| \\
&= \|x_1^{(1)} - \tilde{x}_1^{(1)}\| + \dots + \|\psi^{m_{\mathrm{step}}}(x_1^{(1)}) - \psi^{m_{\mathrm{step}}}(\tilde{x}_1^{(1)})\| \\
&\quad + \dots + \|\psi^{(m_{\mathrm{step}})}(x_{m_{\mathrm{orbit}}}^{(1)}) - \psi^{(m_{\mathrm{step}})}(x_{m_{\mathrm{orbit}}}^{(1)})\| \\
&\le m_{\mathrm{step}} \max\{1, \rho_\psi^{m_{\mathrm{step}}}\}(\|x_1^{(1)} - \tilde{x}_1^{(1)}\| + \dots + \|x_{m_{\mathrm{orbit}}}^{(1)} - \tilde{x}_{m_{\mathrm{orbit}}}^{(1)}\|) \\
&= m_{\mathrm{step}} \max\{1, \rho_\psi^{m_{\mathrm{step}}}\}\|(x_1^{(1)} - \tilde{x}_1^{(1)}, \dots, x_{m_{\mathrm{orbit}}}^{(1)} - \tilde{x}_{m_{\mathrm{orbit}}}^{(1)})\|,
\end{aligned}$$

which shows the Lipschitz constant of $\mathcal{O}$ is bounded by $m_{\mathrm{step}} \max\{1, \rho_\psi^{m_{\mathrm{step}}}\}$. Hence, we obtain

$$N(r, \mathcal{O}(B^{m_{\mathrm{orbit}}})) \le N(\frac{r}{m_{\mathrm{step}} \max\{1, \rho_\psi^{m_{\mathrm{step}}}\}}, B^{m_{\mathrm{orbit}}}).$$

To compute the loss functions $\mathcal{L}_1$ and $\mathcal{L}_2$, these extended data are first input into f_{NN} to form the set of the latent variables:

$$Z := \{(f_{\mathrm{NN}}(x), x) \mid x \in \mathcal{O}(B^{m_{\mathrm{orbit}}})\} \subset B^{m_{\mathrm{step}} \times m_{\mathrm{orbit}}} \times B^{m_{\mathrm{step}} \times m_{\mathrm{orbit}}}.$$

For each $x, \tilde{x} \in \mathcal{O}(B^{m_{\mathrm{orbit}}})$, we have

$$\|(f_{\mathrm{NN}}(x), x) - (f_{\mathrm{NN}}(\tilde{x}), \tilde{x})\| = \|f_{\mathrm{NN}}(x) - f_{\mathrm{NN}}(\tilde{x})\| + \|x - \tilde{x}\| \le (c_{\mathrm{enc}}\rho_{\mathrm{enc}} + 1)\|x - \tilde{x}\|.$$

This estimate shows that

$$N(r, Z) \le N\left(\frac{r}{(c_{\mathrm{enc}}\rho_{\mathrm{enc}} + 1)}, \mathcal{O}(B^{m_{\mathrm{orbit}}})\right).$$

Second, the set Z should be transformed to

$$\{(f_{\mathrm{NN}}(x), g_{\mathrm{NN}}(x), h_{\mathrm{NN}}(x), x) \mid x \in \mathcal{O}(B^{m_{\mathrm{orbit}}})\}.$$

Similarly to the estimation of the map from $\mathcal{O}(B^{m_{\mathrm{orbit}}})$ to Z, we get

$$N(r, \{(f_{\mathrm{NN}}(x), g_{\mathrm{NN}}(x), h_{\mathrm{NN}}(x), x) \mid x \in \mathcal{O}(B^{m_{\mathrm{orbit}}})\}) \le N(\frac{r}{c_{\mathrm{dec}}\rho_{\mathrm{dec}} + c_{\mathrm{symp}}\rho_{\mathrm{symp}} + 1}, Z).$$

Because we assume that the loss function $\mathcal{L}$ is $\rho_{\mathcal{L}}$-Lipschitz continuous, we have

$$\begin{aligned}
&N(r, \{\mathcal{L}(f_{\mathrm{NN}}(x), g_{\mathrm{NN}}(x), h_{\mathrm{NN}}(x), x) \mid x \in \mathcal{O}(B^{m_{\mathrm{orbit}}})\}) \\
&\le N(\frac{r}{\rho_{\mathcal{L}}(c_{\mathrm{dec}}\rho_{\mathrm{dec}} + c_{\mathrm{symp}}\rho_{\mathrm{symp}} + 1)}), Z).
\end{aligned}$$

Combining all of the above results yields the following inequality:

$$\begin{aligned}
&N(r, \{\mathcal{L}(f_{\mathrm{NN}}(x), g_{\mathrm{NN}}(x), h_{\mathrm{NN}}(x), x) \mid x \in \mathcal{O}(B^{m_{\mathrm{orbit}}})\}) \\
&\le N(\frac{r}{\rho_{\mathcal{L}}(c_{\mathrm{dec}}\rho_{\mathrm{dec}} + c_{\mathrm{symp}}\rho_{\mathrm{symp}} + 1)}), Z) \\
&\le N(r/\rho_{\mathcal{L}}(c_{\mathrm{dec}}\rho_{\mathrm{dec}} + c_{\mathrm{symp}}\rho_{\mathrm{symp}} + 1)(c_{\mathrm{dec}}\rho_{\mathrm{dec}} + c_{\mathrm{symp}}\rho_{\mathrm{symp}} + 1)), \mathcal{O}(B^{m_{\mathrm{orbit}}})) \\
&\le N(r/\rho_{\mathcal{L}}(c_{\mathrm{dec}}\rho_{\mathrm{dec}} + c_{\mathrm{symp}}\rho_{\mathrm{symp}} + 1)(c_{\mathrm{dec}}\rho_{\mathrm{dec}} + c_{\mathrm{symp}}\rho_{\mathrm{symp}} + 1)(m_{\mathrm{step}} \max\{1, \rho_\psi^{m_{\mathrm{step}}}\})), \\
&\qquad B^{m_{\mathrm{orbit}}}).
\end{aligned}$$

This shows Theorem 4.6.

A.2 Proof of the Hamiltonian interpolation

Let $\varepsilon_2 > 0$ be arbitrarily chosen. Suppose that the loss function for the training data satisfies $\mathcal{L} \leq \varepsilon_1$. Then, from Lemma 3.3, with probability at least $1-\delta$, it holds that

$$E[\mathcal{L}_2] \leq E[\mathcal{L}] \leq \varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}}(\mathcal{G}) + 3c\sqrt{\frac{2\ln\frac{4}{\delta}}{m_{\mathrm{orbit}}}}.$$

When this inequality holds, from Assumption 4.8, we have

$$\begin{aligned}\|h_{\mathrm{NN}} \circ \psi_{\Delta t}^{-1} - \mathrm{Id}\| &\leq c_{l_2} E[\mathcal{L}_2] \\ &\leq c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + 3c\sqrt{\frac{2\ln\frac{4}{\delta}}{m_{\mathrm{orbit}}}}).\end{aligned}$$

Thus, if we let $\hat{\delta}$ be

$$\hat{\delta} = 4\exp(-\frac{m_{\mathrm{orbit}}\varepsilon^2}{18c^2}),$$

at least probability $1-\hat{\delta}$, the following inequality holds:

$$\|h_{\mathrm{NN}} \circ \psi_{\Delta t}^{-1} - \mathrm{Id}\| \leq c_{l_2} E[\mathcal{L}_2] \leq c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + \varepsilon_2).$$

Then, from the above estimate, $h_{\mathrm{NN}} \circ \psi_{\Delta t}^{-1}$ is close to the identity. Also, this is a symplectic map because both h_{NN} and $\psi_{\Delta t}^{-1}$ are symplectic. Hence, from Theorem 3.1, there exists a Hamiltonian flow $\hat{h}_{\mathrm{NN}}$ that appoximates $h_{\mathrm{NN}} \circ \psi_{\Delta t}^{-1}$ within the error

$$O(c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + \varepsilon_2)\exp(-\frac{1}{c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + \varepsilon_2)})).$$

Because h_{NN} is written as

$$h_{\mathrm{NN}} = (h_{\mathrm{NN}} \circ \psi_{\Delta t}^{-1}) \circ \psi_{\Delta t} \simeq \hat{h}_{\mathrm{NN}} \circ \psi_{\Delta t},$$

h_{NN} is approximated by the composition of the two Hamiltonian flows $\hat{h}_{\mathrm{NN}}$ and $\psi_{\Delta t}$. The error analysis of the splitting method shows that there exists a Hamiltonian flow $\tilde{h}_{\mathrm{NN}}$ that approximates $\hat{h}_{\mathrm{NN}} \circ \psi_{\Delta t}$ within the error $O(\|\hat{h}_{\mathrm{NN}}\|\|\psi_{\Delta t}\|)$. This $\tilde{h}_n n$ approximates h_{NN}, and the approximation error is estimated by

$$\begin{aligned}&\|h_{\mathrm{NN}} - \tilde{h}_n n\| \leq \|h_{\mathrm{NN}} - \hat{h}_{\mathrm{NN}} \circ \psi_{\Delta t}\| + \|\hat{h}_{\mathrm{NN}} \circ \psi_{\Delta t} - \tilde{h}_n n\| \\ &= O(c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + \varepsilon_2)\exp(-\frac{1}{c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + \varepsilon_2)})) + O(\|\hat{h}_{\mathrm{NN}}\|\|\psi_{\Delta t}\|)\end{aligned}$$

Since $\|\hat{h}_{\mathrm{NN}}\|$ is $O(c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + \varepsilon_2))$, the approximation error is estimated by $O(c_{l_2}(\varepsilon_1 + 2\mathcal{R}_{m_{\mathrm{orbit}}} + \varepsilon_2))$. This completes the proof.