SUPPLEMENTARY MATERIALS: On the Exact Computation of Linear Frequency Principle Dynamics and Its Generalization*

Tao Luo†, Zheng Ma‡, Zhi-Qin John Xu§, and Yaoyu Zhang¶

SM1. Fourier transform table.

We list the results of one-dimensional Fourier transform in Table SM1 and high-dimensional Fourier transform in Table SM2 used in our proofs.

Table SM1

Fourier transform for 1-dimensional functions used in our proofs.

Function of x	Fourier transform with respect to x
$g(ax)$	$\frac{1}{\|a\|}\mathcal{F}[g](\frac{\xi}{a})$
$g(x-c)$	$\mathcal{F}[g](\xi)\mathrm{e}^{-2\pi\mathrm{i}c\xi}$
$x^k g(x)$	$(\frac{\mathrm{i}}{2\pi})^k \frac{\mathrm{d}^k}{\mathrm{d}\xi^k}\mathcal{F}[g](\xi)$
$g^{(k)}(x)$	$(2\pi\mathrm{i}\xi)^k\mathcal{F}[g](\xi)$
1	$\delta(\xi)$
x^k	$(\frac{\mathrm{i}}{2\pi})^k\delta^{(k)}(\xi)$
$\delta(x-x_0)$	$\mathrm{e}^{-2\pi\mathrm{i}x_0\xi}$
$H(x)$ (Heaviside)	$\frac{1}{\mathrm{i}2\pi\xi}+\frac{1}{2}\delta(\xi)$
$\mathrm{ReLU}(x)$	$-\frac{1}{4\pi^2\xi^2}+\frac{\mathrm{i}}{4\pi}\delta'(\xi)$
$\tanh(x)$	$-\mathrm{i}\pi\mathrm{csch}(\pi^2\xi)$
$\mathrm{Sigmoid}(x)$	$-\mathrm{i}\pi\mathrm{csch}(2\pi^2\xi)+\frac{1}{2}\delta(\xi)$
$\mathrm{sech}^2(x)$	$2\pi^2\xi\mathrm{csch}(\pi^2\xi)$
$x\mathrm{sech}^2(x)$	$\mathrm{i}\pi\left(1-\pi^2\xi\coth(\pi^2\xi)\right)\mathrm{csch}(\pi^2\xi)$

SM2. Proof.

Lemma. *(Lemma 2 in main text) Given any nonzero vector $\boldsymbol{w}\in\mathbb{R}^d$ with $\hat{\boldsymbol{w}}=\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$, we have*

$$\frac{1}{\|\boldsymbol{w}\|^d}\delta_{\hat{\boldsymbol{w}}}\left(\frac{\boldsymbol{x}}{\|\boldsymbol{w}\|}\right)=\delta_{\boldsymbol{w}}(\boldsymbol{x}). \tag{SM2.1}$$

*Supplementary material for SIMODS MS#M144440.
https://doi.org/10.1137/21M1444400

†Corresponding author. School of Mathematical Sciences, CMA-Shanghai, Institute of Natural Sciences, MOE-LSC, and Qing Yuan Research Institute, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China (luotao41@sjtu.edu.cn).

‡School of Mathematical Sciences, CMA-Shanghai, Institute of Natural Sciences, MOE-LSC, and Qing Yuan Research Institute, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China (zhengma@sjtu.edu.cn).

§School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSC, and Qing Yuan Research Institute, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China (xuzhiqin@sjtu.edu.cn).

¶School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSC, and Qing Yuan Research Institute, Shanghai Jiao Tong University, and Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai, 200240, People's Republic of China (zhyy.sjtu@sjtu.edu.cn).

Table SM2
Fourier transform for d-dimensional functions used in our proofs.

Function of $\boldsymbol{x}$	Fourier transform with respect to $\boldsymbol{x}$
$g(a\boldsymbol{x})$	$\frac{1}{\lvert a\rvert^d}\mathcal{F}[g](\frac{\boldsymbol{\xi}}{a})$
$\delta(\boldsymbol{x}-\boldsymbol{x}_0)$	$\mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{\xi}^\intercal\boldsymbol{x}_0}$
$g(\boldsymbol{\nu}^\intercal\boldsymbol{x})$ (unit vector $\boldsymbol{\nu}$)	$\delta_{\boldsymbol{\nu}}(\boldsymbol{\xi})\mathcal{F}[g](\boldsymbol{\xi}^\intercal\boldsymbol{\nu})$
$g(\boldsymbol{w}^\intercal\boldsymbol{x}+b)$	$\delta_{\boldsymbol{w}}(\boldsymbol{\xi})\mathcal{F}[g](\frac{\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}{\lVert\boldsymbol{w}\rVert})\mathrm{e}^{2\pi\mathrm{i}\frac{b}{\lVert\boldsymbol{w}\rVert}\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}$
$g(\boldsymbol{w}^\intercal\boldsymbol{x}+\lVert\boldsymbol{w}\rVert c)$	$\delta_{\boldsymbol{w}}(\boldsymbol{\xi})\mathcal{F}[g](\frac{\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}{\lVert\boldsymbol{w}\rVert})\mathrm{e}^{2\pi\mathrm{i}c\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}$
$\boldsymbol{x}g(\boldsymbol{x})$	$\frac{\mathrm{i}}{2\pi}\nabla\mathcal{F}[g](\boldsymbol{\xi})$
$\boldsymbol{x}^\perp g(\boldsymbol{w}^\intercal\boldsymbol{x}+\lVert\boldsymbol{w}\rVert c)$	$\frac{\mathrm{i}}{2\pi}\nabla_{\boldsymbol{\xi}^\perp}\left[\delta_{\boldsymbol{w}}(\boldsymbol{\xi})\mathcal{F}[g](\frac{\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}{\lVert\boldsymbol{w}\rVert})\mathrm{e}^{2\pi\mathrm{i}c\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}\right]$
$\boldsymbol{x}g(\boldsymbol{w}^\intercal\boldsymbol{x}+b)$	$\frac{\mathrm{i}}{2\pi}\nabla_{\boldsymbol{\xi}}\left[\delta_{\boldsymbol{w}}(\boldsymbol{\xi})\mathcal{F}[g](\frac{\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}{\lVert\boldsymbol{w}\rVert})\mathrm{e}^{2\pi\mathrm{i}b\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}/\lVert\boldsymbol{w}\rVert}\right]$

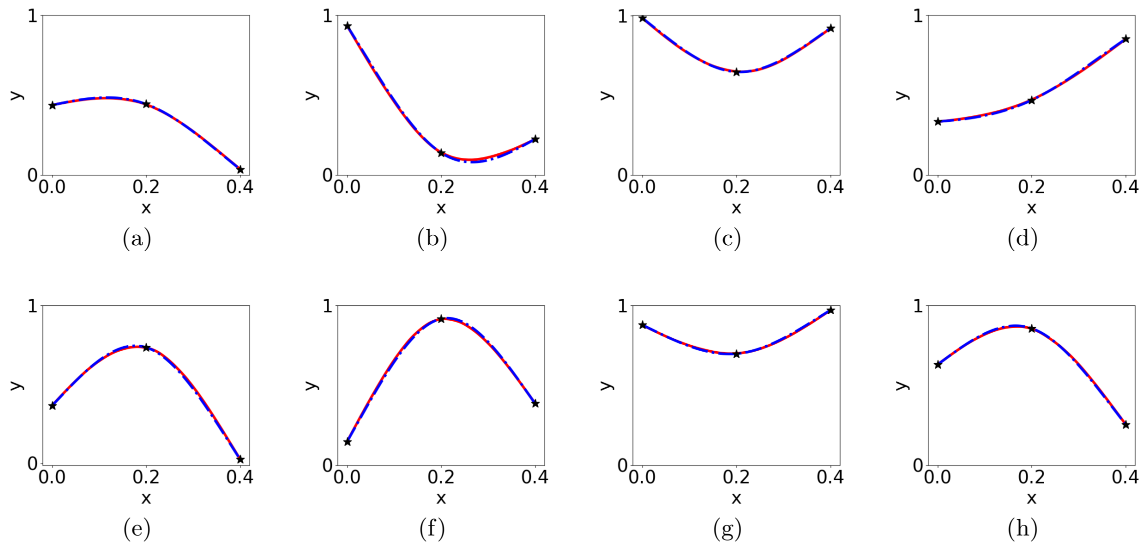

Figure SM1. *f_{NN} (red solid) vs. f_{LFP} (blue dashed dot) for a 1-d problem. The setting is the same as the case in Figure 1(a) in main text, except that the label for each data is randomly selected from $[0,1]$ and the uniform distribution half width U is randomly selected from $[3,6]$. Each subfigure is one trial.*

Proof. This is proved by changing of variables. In fact, for any $\phi\in\mathcal{S}(\mathbb{R}^d)$, we have

$$\begin{aligned}\left\langle\frac{1}{\lVert\boldsymbol{w}\rVert^d}\delta_{\hat{\boldsymbol{w}}}\left(\frac{\cdot}{\lVert\boldsymbol{w}\rVert}\right),\phi(\cdot)\right\rangle_{\mathcal{S}'(\mathbb{R}^d),\mathcal{S}(\mathbb{R}^d)} &= \langle\delta_{\hat{\boldsymbol{w}}}(\cdot),\phi(\lVert\boldsymbol{w}\rVert\cdot)\rangle_{\mathcal{S}'(\mathbb{R}^d),\mathcal{S}(\mathbb{R}^d)}\\ &= \int_{\mathbb{R}}\phi\left(\lVert\boldsymbol{w}\rVert y\hat{\boldsymbol{w}}\right)\mathrm{d}y\\ &= \int_{\mathbb{R}}\phi(y\boldsymbol{w})\,\mathrm{d}y\\ &= \langle\delta_{\boldsymbol{w}}(\cdot),\phi(\cdot)\rangle_{\mathcal{S}'(\mathbb{R}^d),\mathcal{S}(\mathbb{R}^d)}\,.\end{aligned}$$ ■

Lemma. *(Lemma 2 in main text) For any unit vector $\boldsymbol{\nu} \in \mathbb{R}^d$, any nonzero vector $\boldsymbol{w} \in \mathbb{R}^d$ with $\hat{\boldsymbol{w}} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$, and $g \in \mathcal{S}'(\mathbb{R})$ with $\mathcal{F}[g] \in C(\mathbb{R})$, we have, in the sense of distribution,*

$$\text{(a)} \quad \mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\boldsymbol{\nu}^\intercal \boldsymbol{x})](\boldsymbol{\xi}) = \delta_{\boldsymbol{\nu}}(\boldsymbol{\xi})\mathcal{F}[g](\boldsymbol{\xi}^\intercal \boldsymbol{\nu}), \tag{SM2.2}$$

$$\text{(b)} \quad \mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\boldsymbol{w}^\intercal \boldsymbol{x} + b)](\boldsymbol{\xi}) = \delta_{\boldsymbol{w}}(\boldsymbol{\xi})\mathcal{F}[g]\left(\frac{\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)\mathrm{e}^{2\pi \mathrm{i}\frac{b}{\|\boldsymbol{w}\|}\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}, \tag{SM2.3}$$

$$\text{(c)} \quad \mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[\boldsymbol{x} g(\boldsymbol{w}^\intercal \boldsymbol{x} + b)](\boldsymbol{\xi}) = \frac{\mathrm{i}}{2\pi}\nabla_{\boldsymbol{\xi}}\left[\delta_{\boldsymbol{w}}(\boldsymbol{\xi})\mathcal{F}[g]\left(\frac{\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)\mathrm{e}^{2\pi \mathrm{i}\frac{b}{\|\boldsymbol{w}\|}\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}\right]. \tag{SM2.4}$$

Proof. Let $\phi \in \mathcal{S}(\mathbb{R}^d)$ be any test function.

(a) By direct calculation, we have

$$\begin{aligned}
\langle \mathcal{F}_{\boldsymbol{x}\to\cdot}[g(\boldsymbol{\nu}^\intercal \boldsymbol{x})](\cdot), \phi(\cdot)\rangle_{\mathcal{S}'(\mathbb{R}^d),\mathcal{S}(\mathbb{R}^d)} &= \langle g(\boldsymbol{\nu}^\intercal \cdot), \mathcal{F}_{\boldsymbol{x}\to\cdot}[\phi(\boldsymbol{x})](\cdot)\rangle_{\mathcal{S}'(\mathbb{R}^d),\mathcal{S}(\mathbb{R}^d)} \\
&= \langle g(\cdot), \mathcal{F}_{y\to\cdot}[\phi(y\boldsymbol{\nu})](\cdot)\rangle_{\mathcal{S}'(\mathbb{R}),\mathcal{S}(\mathbb{R})} \\
&= \langle \mathcal{F}_{y\to\cdot}[g(y)](\cdot), \phi(\cdot\boldsymbol{\nu})\rangle_{\mathcal{S}'(\mathbb{R}),\mathcal{S}(\mathbb{R})} \\
&= \langle \mathcal{F}[g](\cdot\boldsymbol{\nu}^\intercal \boldsymbol{\nu}), \phi(\cdot\boldsymbol{\nu})\rangle_{\mathcal{S}'(\mathbb{R}),\mathcal{S}(\mathbb{R})} \\
&= \langle \delta_{\boldsymbol{\nu}}(\cdot)\mathcal{F}[g](\cdot^\intercal \boldsymbol{\nu}), \phi(\cdot)\rangle_{\mathcal{S}'(\mathbb{R}^d),\mathcal{S}(\mathbb{R}^d)}.
\end{aligned}$$

(b) By part (a), we have in the distributional sense

$$\mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\hat{\boldsymbol{w}}^\intercal \boldsymbol{x})](\boldsymbol{\xi}) = \delta_{\hat{\boldsymbol{w}}}(\boldsymbol{\xi})\mathcal{F}[g](\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}).$$

Note that

$$\mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\boldsymbol{x} - \boldsymbol{x}_0)](\boldsymbol{\xi}) = \mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g](\boldsymbol{\xi})\mathrm{e}^{-2\pi \mathrm{i}\boldsymbol{x}_0^\intercal \boldsymbol{\xi}},$$

then

$$\begin{aligned}
\mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\hat{\boldsymbol{w}}^\intercal \boldsymbol{x} + b)](\boldsymbol{\xi}) &= \mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\hat{\boldsymbol{w}}^\intercal(\boldsymbol{x} + b\hat{\boldsymbol{w}}))](\boldsymbol{\xi}) \\
&= \delta_{\hat{\boldsymbol{w}}}(\boldsymbol{\xi})\mathcal{F}[g](\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}})\mathrm{e}^{2\pi \mathrm{i} b\hat{\boldsymbol{w}}^\intercal \boldsymbol{\xi}}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\boldsymbol{w}^\intercal \boldsymbol{x} + b)](\boldsymbol{\xi}) &= \mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\hat{\boldsymbol{w}}^\intercal \|\boldsymbol{w}\|\boldsymbol{x} + b)](\boldsymbol{\xi}) \\
&= \frac{1}{\|\boldsymbol{w}\|^d}\mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[g(\hat{\boldsymbol{w}}^\intercal \boldsymbol{x} + b)]\left(\frac{\boldsymbol{\xi}}{\|\boldsymbol{w}\|}\right) \\
&= \frac{1}{\|\boldsymbol{w}\|^d}\delta_{\hat{\boldsymbol{w}}}\left(\frac{\boldsymbol{\xi}}{\|\boldsymbol{w}\|}\right)\mathcal{F}[g]\left(\frac{\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)\mathrm{e}^{2\pi \mathrm{i}\frac{b}{\|\boldsymbol{w}\|}\hat{\boldsymbol{w}}^\intercal \boldsymbol{\xi}} \\
&= \delta_{\boldsymbol{w}}(\boldsymbol{\xi})\mathcal{F}[g]\left(\frac{\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)\mathrm{e}^{2\pi \mathrm{i}\frac{b}{\|\boldsymbol{w}\|}\hat{\boldsymbol{w}}^\intercal \boldsymbol{\xi}}.
\end{aligned}$$

(c) This follows from part (b) and the fact that for any function $\tilde{g}(\boldsymbol{x})$ ∎

$$\mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[\boldsymbol{x}\tilde{g}(\boldsymbol{x})](\boldsymbol{\xi}) = \frac{\mathrm{i}}{2\pi}\nabla_{\boldsymbol{\xi}}\left[\mathcal{F}[g](\boldsymbol{\xi})\right].$$

Lemma. *(Lemma 3 in main text) The dynamics (4.9) has the following expression in the frequency domain for all* $\phi \in \mathcal{S}(\mathbb{R}^d)$

$$\langle \partial_t \mathcal{F}[u], \phi \rangle = -\langle \mathcal{L}[\mathcal{F}[u_\rho]], \phi \rangle, \tag{SM2.5}$$

where $\mathcal{L}[\cdot]$ *is called Linear F-Principle (LFP) operator is given by*

$$\mathcal{L}[\mathcal{F}[u_\rho]] = \int_{\mathbb{R}^d} \hat{K}(\boldsymbol{\xi}, \boldsymbol{\xi}') \mathcal{F}[u_\rho](\boldsymbol{\xi}') \, \mathrm{d}\boldsymbol{\xi}',$$

and

$$\hat{K}(\boldsymbol{\xi}, \boldsymbol{\xi}') := \mathbb{E}_{\boldsymbol{q}} \hat{K}_{\boldsymbol{q}}(\boldsymbol{\xi}, \boldsymbol{\xi}') := \mathbb{E}_{\boldsymbol{q}} \mathcal{F}_{\boldsymbol{x} \to \boldsymbol{\xi}}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}, \boldsymbol{q})] \cdot \overline{\mathcal{F}_{\boldsymbol{x}' \to \boldsymbol{\xi}'}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q})]}. \tag{SM2.6}$$

The expectation $\mathbb{E}_{\boldsymbol{q}}$ *is taken w.r.t. initial distribution of parameters.*

Proof. For any $\phi \in \mathcal{S}(\mathbb{R}^d)$. since $\partial_t u$ is in $\mathcal{S}'(\mathbb{R}^d)$ and locally integrable, we have

$$\begin{aligned}
\langle \partial_t \mathcal{F}[u], \phi \rangle &= \langle \partial_t u, \mathcal{F}[\phi] \rangle \\
&= \int_{\mathbb{R}^d} \partial_t u(\boldsymbol{x}, t) \int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}) \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{x} \cdot \boldsymbol{\xi}} \, \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{x} \\
&= -\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(\boldsymbol{x}, \boldsymbol{x}') u_\rho(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' \int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}) \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{x} \cdot \boldsymbol{\xi}} \, \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{x} \\
&= -\int_{\mathbb{R}^{3d}} K(\boldsymbol{x}, \boldsymbol{x}') u_\rho(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' \phi(\boldsymbol{\xi}) \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{x} \cdot \boldsymbol{\xi}} \, \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{x} \\
&= -\int_{\mathbb{R}^{3d}} \mathbb{E}_{\boldsymbol{q}} \nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}, \boldsymbol{q}) \cdot \nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q}) u_\rho(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' \phi(\boldsymbol{\xi}) \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{x} \cdot \boldsymbol{\xi}} \, \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{x} \\
&= -\mathbb{E}_{\boldsymbol{q}} \int_{\mathbb{R}^d} \nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q}) u_\rho(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' \cdot \int_{\mathbb{R}^{2d}} \nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}, \boldsymbol{q}) \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{x} \cdot \boldsymbol{\xi}} \phi(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{x} \\
&= -\mathbb{E}_{\boldsymbol{q}} \int_{\mathbb{R}^d} \nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q}) u_\rho(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' \cdot \langle \mathcal{F}_{\boldsymbol{x} \to \cdot}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}, \boldsymbol{q})](\cdot), \phi(\cdot) \rangle .
\end{aligned}$$

Since

$$\int_{\mathbb{R}^d} \nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q}) u_\rho(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' = \int_{\mathbb{R}^d} \overline{\mathcal{F}_{\boldsymbol{x}' \to \boldsymbol{\xi}'}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q})](\boldsymbol{\xi}')} \mathcal{F}_{\boldsymbol{x}' \to \boldsymbol{\xi}'}[u_\rho](\boldsymbol{\xi}') \, \mathrm{d}\boldsymbol{\xi}',$$

we have

$$\begin{aligned}
\langle \partial_t \mathcal{F}[u], \phi \rangle &= -\mathbb{E}_{\boldsymbol{q}} \int_{\mathbb{R}^d} \overline{\mathcal{F}_{\boldsymbol{x}' \to \boldsymbol{\xi}'}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q})](\boldsymbol{\xi}')} \mathcal{F}_{\boldsymbol{x}' \to \boldsymbol{\xi}'}[u_\rho](\boldsymbol{\xi}') \, \mathrm{d}\boldsymbol{\xi}' \cdot \langle \mathcal{F}_{\boldsymbol{x} \to \cdot}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}, \boldsymbol{q})](\cdot), \phi(\cdot) \rangle \\
&= -\mathbb{E}_{\boldsymbol{q}} \int_{\mathbb{R}^{2d}} \overline{\mathcal{F}_{\boldsymbol{x}' \to \boldsymbol{\xi}'}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}', \boldsymbol{q})](\boldsymbol{\xi}')} \cdot \mathcal{F}_{\boldsymbol{x} \to \boldsymbol{\xi}}[\nabla_{\boldsymbol{q}} \sigma^*(\boldsymbol{x}, \boldsymbol{q})](\boldsymbol{\xi}) \mathcal{F}_{\boldsymbol{x}' \to \boldsymbol{\xi}'}[u_\rho](\boldsymbol{\xi}') \, \mathrm{d}\boldsymbol{\xi}' \phi(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \\
&= -\int_{\mathbb{R}^{2d}} \hat{K}(\boldsymbol{\xi}, \boldsymbol{\xi}') \mathcal{F}[u_\rho](\boldsymbol{\xi}') \, \mathrm{d}\boldsymbol{\xi}' \phi(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \\
&= -\langle \mathcal{L}[\mathcal{F}[u_\rho]], \phi \rangle.
\end{aligned}$$

■

Theorem. *(Theorem 1 in main text) Suppose that Assumption 1 holds. If $\sigma_b \gg 1$, then the dynamics (4.9) has the following expression,*

$$\langle \partial_t \mathcal{F}[u], \phi\rangle = -\left\langle \mathcal{L}[\mathcal{F}[u_\rho]], \phi\right\rangle + O(\sigma_b^{-3}), \tag{SM2.7}$$

where $\phi \in \mathcal{S}(\mathbb{R}^d)$ is a test function and the LFP operator is given by

$$\begin{aligned}\mathcal{L}[\mathcal{F}[u_\rho]] = & \frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b\|\boldsymbol{\xi}\|^{d-1}}\mathbb{E}_{a,r}\left[\frac{1}{r}\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right)\cdot\mathcal{F}[\boldsymbol{g}_1]\left(\frac{-\|\boldsymbol{\xi}\|}{r}\right)\right]\mathcal{F}[u_\rho](\boldsymbol{\xi})\\ & -\frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b}\nabla\cdot\left(\mathbb{E}_{a,r}\left[\frac{1}{r\|\boldsymbol{\xi}\|^{d-1}}\mathcal{F}[g_2]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right)\mathcal{F}[g_2]\left(-\frac{\|\boldsymbol{\xi}\|}{r}\right)\right]\nabla\mathcal{F}[u_\rho](\boldsymbol{\xi})\right),\end{aligned} \tag{SM2.8}$$

where $\Gamma(\cdot)$ is the gamma function. The expectations are taken w.r.t. initial parameter distribution. Here $r = \|\boldsymbol{w}\|$ with the probability density $\rho_r(r) := \frac{2\pi^{d/2}}{\Gamma(d/2)}\rho_{\boldsymbol{w}}(r\boldsymbol{e}_1)r^{d-1}$, $\boldsymbol{e}_1 = (1,0,\cdots,0)^\intercal$.

Proof. For simplicity, we assume that $b \sim \mathcal{N}(0,\sigma_b^2)$, $\sigma_b \gg 1$ in this proof. It is straightforward to extend the proof to general distributions for b as long as it is zero-mean and with variance $\sigma_b \gg 1$.

1. Divide into two parts. Note that

$$\begin{pmatrix}\boldsymbol{g}_1(\boldsymbol{w}^\intercal\boldsymbol{x}+b)\\ \boldsymbol{x}g_2(\boldsymbol{w}^\intercal\boldsymbol{x}+b)\end{pmatrix} = \begin{pmatrix}\partial_a[a\sigma(\boldsymbol{w}^\intercal\boldsymbol{x}+b)]\\ \partial_b[a\sigma(\boldsymbol{w}^\intercal\boldsymbol{x}+b)]\\ \nabla_{\boldsymbol{w}}[a\sigma(\boldsymbol{w}^\intercal\boldsymbol{x}+b)]\end{pmatrix} = \nabla_{\boldsymbol{q}}\sigma^*(\boldsymbol{x},\boldsymbol{q}). \tag{SM2.9}$$

One can split the Fourier transformed kernel $\hat{K}$ into two parts, more precisely,

$$\hat{K} = \mathbb{E}_{\boldsymbol{q}}\hat{K}_{\boldsymbol{q}}, \quad \hat{K}_{\boldsymbol{q}} = \hat{K}_{a,b} + \hat{K}_{\boldsymbol{w}},$$

where

$$\begin{aligned}\hat{K}_{\boldsymbol{q}}(\boldsymbol{\xi},\boldsymbol{\xi}') &= \mathbb{E}_{\boldsymbol{q}}\mathcal{F}_{\boldsymbol{x}\to\boldsymbol{\xi}}[\nabla_{\boldsymbol{q}}\sigma^*(\boldsymbol{x},\boldsymbol{q})]\cdot\overline{\mathcal{F}_{\boldsymbol{x}'\to\boldsymbol{\xi}'}[\nabla_{\boldsymbol{q}}\sigma^*(\boldsymbol{x}',\boldsymbol{q})]},\\ \hat{K}_{a,b}(\boldsymbol{\xi},\boldsymbol{\xi}') &= \mathcal{F}\left[\boldsymbol{g}_1(\boldsymbol{w}^\intercal\boldsymbol{x}+b)\right]\cdot\overline{\mathcal{F}\left[\boldsymbol{g}_1(\boldsymbol{w}^\intercal\boldsymbol{x}'+b)\right]},\\ \hat{K}_{\boldsymbol{w}}(\boldsymbol{\xi},\boldsymbol{\xi}') &= \mathcal{F}\left[\boldsymbol{x}g_2(\boldsymbol{w}^\intercal\boldsymbol{x}+b)\right]\cdot\overline{\mathcal{F}\left[\boldsymbol{x}g_2(\boldsymbol{w}^\intercal\boldsymbol{x}'+b)\right]}.\end{aligned}$$

For any $\phi,\psi \in \mathcal{S}(\mathbb{R}^d)$, we have

$$\langle\hat{K}_{\boldsymbol{q}},\phi\otimes\psi\rangle := \langle\hat{K}_{\boldsymbol{q}},\phi\otimes\psi\rangle_{\mathcal{S}'(\mathbb{R}^{2d}),\mathcal{S}(\mathbb{R}^{2d})} = \int_{\mathbb{R}^{2d}}\hat{K}_{\boldsymbol{q}}(\boldsymbol{\xi},\boldsymbol{\xi}')\phi(\boldsymbol{\xi})\psi(\boldsymbol{\xi}')\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\xi}'. \tag{SM2.10}$$

The expressions for $\hat{K}_{a,b}$ and $\hat{K}_{\boldsymbol{w}}$ are similar.

2. Calculate $\hat{K}_{a,b}(\boldsymbol{\xi},\boldsymbol{\xi}')$. Since

$$\hat{K}_{a,b}(\boldsymbol{\xi},\boldsymbol{\xi}') = \delta_{\boldsymbol{w}}(\boldsymbol{\xi})\delta_{\boldsymbol{w}}(\boldsymbol{\xi}')\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\boldsymbol{\xi}'^\intercal\hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)}\mathrm{e}^{2\pi\mathrm{i}b(\boldsymbol{\xi}-\boldsymbol{\xi}')^\intercal\hat{\boldsymbol{w}}/\|\boldsymbol{w}\|},$$

we have

$$\begin{aligned}\langle \hat{K}_{a,b}, \phi\otimes\psi\rangle &= \int_{\mathbb{R}^{2d}} \delta_{\boldsymbol{w}}(\boldsymbol{\xi})\delta_{\boldsymbol{w}}(\boldsymbol{\xi}')\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\boldsymbol{\xi}^\intercal\hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\boldsymbol{\xi}'^\intercal\hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|}\right)}\mathrm{e}^{2\pi\mathrm{i}b(\boldsymbol{\xi}-\boldsymbol{\xi}')^\intercal\hat{\boldsymbol{w}}/\|\boldsymbol{w}\|}\phi(\boldsymbol{\xi})\psi(\boldsymbol{\xi}')\,\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{\xi}'\\ &= \int_{\mathbb{R}\times\mathbb{R}} \phi(\eta\boldsymbol{w})\psi(\eta'\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1](\eta')}\mathrm{e}^{2\pi\mathrm{i}b(\eta-\eta')}\,\mathrm{d}\eta\,\mathrm{d}\eta'.\end{aligned}$$

By assumption $b\sim\mathcal{N}(0,\sigma_b^2)$, i.e., $\rho_b(b)=\frac{1}{\sqrt{2\pi}\sigma_b}\mathrm{e}^{-\frac{b^2}{2\sigma_b^2}}$, then $\mathcal{F}[\rho_b](\eta)=\mathrm{e}^{-2\pi^2\sigma_b^2\eta^2}$.

$$\begin{aligned}\mathbb{E}_b\left(\mathrm{e}^{2\pi\mathrm{i}b(\eta-\eta')}\right) &= \int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}\sigma_b}\mathrm{e}^{-b^2/2\sigma_b^2}\mathrm{e}^{2\pi\mathrm{i}b(\eta-\eta')}\,\mathrm{d}b\\ &= \mathcal{F}[\rho_b]\left(-(\eta-\eta')\right)\\ &= \mathrm{e}^{-2\pi^2\sigma_b^2(\eta-\eta')^2}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}_b\left[\langle \hat{K}_{a,b}, \phi\otimes\psi\rangle\right] &= \int_{\mathbb{R}\times\mathbb{R}} \phi(\eta\boldsymbol{w})\psi(\eta'\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1](\eta)}\mathbb{E}_b\left[\mathrm{e}^{2\pi\mathrm{i}b(\eta-\eta')}\right]\mathrm{d}\eta\,\mathrm{d}\eta'\\ &= \int_{\mathbb{R}\times\mathbb{R}} \phi(\eta\boldsymbol{w})\psi(\eta'\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1](\eta)}\mathrm{e}^{-2\pi^2\sigma_b^2(\eta-\eta')^2}\,\mathrm{d}\eta\,\mathrm{d}\eta'.\end{aligned}$$

Applying the Laplace method, we have

$$\begin{aligned}\mathbb{E}_b\left[\langle \hat{K}_{a,b}, \phi\otimes\psi\rangle\right] &= \int_{\mathbb{R}} \phi(\eta\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\left[\int_{\mathbb{R}}\psi(\eta'\boldsymbol{w})\overline{\mathcal{F}[\boldsymbol{g}_1](\eta)}\mathrm{e}^{-2\pi^2\sigma_b^2(\eta-\eta')^2}\,\mathrm{d}\eta'\right]\mathrm{d}\eta\\ &= \int_{\mathbb{R}} \phi(\eta\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\left[\psi(\eta\boldsymbol{w})\overline{\mathcal{F}[\boldsymbol{g}_1](\eta)}\frac{1}{\sqrt{2\pi}\sigma_b}+O(\sigma_b^{-3})\right]\mathrm{d}\eta\\ &= \frac{1}{\sqrt{2\pi}\sigma_b}\int_{\mathbb{R}} \phi(\eta\boldsymbol{w})\psi(\eta\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1](\eta)}\,\mathrm{d}\eta+O(\sigma_b^{-3}).\end{aligned}$$

Next we consider the expectation with respect to $\boldsymbol{w}$. Up to error of order $O(\sigma_c^{-3})$, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{w},b}\left[\langle \hat{K}_{a,b}, \phi\otimes\psi\rangle\right] &= \mathbb{E}_{\boldsymbol{w}}\left[\frac{1}{\sqrt{2\pi}\sigma_b}\int_{\mathbb{R}} \phi(\eta\boldsymbol{w})\psi(\eta\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1](\eta)}\,\mathrm{d}\eta\right]\\ &= \int_{\mathbb{R}^{d+1}}\frac{1}{\sqrt{2\pi}\sigma_b}\phi(\eta\boldsymbol{w})\psi(\eta\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta)\cdot\overline{\mathcal{F}[\boldsymbol{g}_1](\eta)}\rho_{\boldsymbol{w}}(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}\,\mathrm{d}\eta.\end{aligned}$$

Here we assume that $\rho_{\boldsymbol{w}}$ is radially symmetric so $\rho_{\boldsymbol{w}}(\boldsymbol{w})$ is a function of $r:=\|\boldsymbol{w}\|$ only. By using spherical coordinate system, we have

$$\begin{aligned}1 &= \int_{\mathbb{R}^d}\rho_{\boldsymbol{w}}(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}\\ &= \int_{\mathbb{R}^d}\rho_{\boldsymbol{w}}(\|\boldsymbol{w}\|\boldsymbol{e}_1)\,\mathrm{d}\boldsymbol{w}\\ &= \int_{\mathbb{R}^+}\int_{\mathbb{S}^{d-1}}\rho_{\boldsymbol{w}}(r\boldsymbol{e}_1)r^{d-1}\,\mathrm{d}\hat{\boldsymbol{w}}\,\mathrm{d}r\\ &= \int_{\mathbb{R}^+}\rho_r(r)\,\mathrm{d}r,\end{aligned}$$

where $\hat{\boldsymbol{w}} \in \mathbb{S}^{d-1}$ and we define

$$\rho_r(r) := \int_{\mathbb{S}^{d-1}} \rho_{\boldsymbol{w}}(r\boldsymbol{e}_1) r^{d-1} \, \mathrm{d}\hat{\boldsymbol{w}} = \frac{2\pi^{d/2}}{\Gamma(d/2)} \rho_{\boldsymbol{w}}(r\boldsymbol{e}_1) r^{d-1}, \tag{SM2.11}$$

where $\Gamma(\cdot)$ is the gamma function. Then we introduce the following change of variables,

$$\begin{cases} \boldsymbol{\zeta} = \eta \boldsymbol{w}, \\ r = \|\boldsymbol{w}\|, \end{cases}$$

whose the Jacobian determinant is

$$\det\left(\frac{\partial(\boldsymbol{\zeta}, r)}{\partial(\boldsymbol{w}, \eta)}\right) = \det \begin{bmatrix} \eta & 0 & \cdots & 0 & w_1 \\ 0 & \eta & \cdots & 0 & w_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \eta & w_d \\ w_1/r & w_2/r & \cdots & w_d/r & 0 \end{bmatrix} = -r\eta^{d-1} = -r\left(\frac{\|\boldsymbol{\zeta}\|}{r}\right)^{d-1}.$$

Thus

$$\begin{cases} \boldsymbol{w} = \dfrac{r\boldsymbol{\zeta}}{\|\boldsymbol{\zeta}\|} \\ \eta = \dfrac{\|\boldsymbol{\zeta}\|}{r}, \end{cases} \tag{SM2.12}$$

and its Jacobian determinant is

$$\det\left(\frac{\partial(\boldsymbol{w}, \eta)}{\partial(\boldsymbol{\zeta}, r)}\right) = -\frac{r^{d-1}}{r\|\boldsymbol{\zeta}\|^{d-1}}.$$

So one can obtain,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{w},b}\left[\langle \hat{K}_{a,b}, \phi \otimes \psi \rangle\right] &= \int_{\mathbb{R}^{d+1}} \frac{1}{\sqrt{2\pi}\sigma_b} \phi(\eta\boldsymbol{w})\psi(\eta\boldsymbol{w})\mathcal{F}[\boldsymbol{g}_1](\eta) \cdot \overline{\mathcal{F}[\boldsymbol{g}_1](\eta)} \rho_{\boldsymbol{w}}(r\boldsymbol{e}_1) \, \mathrm{d}\boldsymbol{w} \, \mathrm{d}\eta \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^+} \frac{1}{\sqrt{2\pi}\sigma_b} \phi(\boldsymbol{\zeta})\psi(\boldsymbol{\zeta})\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\zeta}\|}{r}\right) \cdot \overline{\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\zeta}\|}{r}\right)} \frac{r^{d-1}}{r\|\boldsymbol{\zeta}\|^{d-1}} \rho_{\boldsymbol{w}}(r\boldsymbol{e}_1) \, \mathrm{d}\boldsymbol{\zeta} \, \mathrm{d}r \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^+} \frac{1}{\sqrt{2\pi}\sigma_b} \phi(\boldsymbol{\zeta})\psi(\boldsymbol{\zeta})\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\zeta}\|}{r}\right) \cdot \overline{\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\zeta}\|}{r}\right)} \frac{1}{r\|\boldsymbol{\zeta}\|^{d-1}} \left[\frac{\Gamma(d/2)}{2\pi^{d/2}} \rho_r(r)\right] \mathrm{d}\boldsymbol{\zeta} \, \mathrm{d}r \\
&= \frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b} \int_{\mathbb{R}^d} \phi(\boldsymbol{\zeta}) \int_{\mathbb{R}^+} \left[\frac{1}{r\|\boldsymbol{\zeta}\|^{d-1}} \mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\zeta}\|}{r}\right) \cdot \overline{\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\zeta}\|}{r}\right)}\right] \psi(\boldsymbol{\zeta})\rho_r(r) \, \mathrm{d}r \, \mathrm{d}\boldsymbol{\zeta},
\end{aligned}$$

Therefore taking $\psi = \mathcal{F}[u_\rho]$, we have

$$\begin{aligned}
\mathcal{L}_{a,b}[\mathcal{F}[u_\rho]] &= \frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b\|\boldsymbol{\xi}\|^{d-1}} \mathbb{E}_{a,r}\left[\frac{1}{r}\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right) \cdot \overline{\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right)}\right] \mathcal{F}[u_\rho](\boldsymbol{\xi}) \\
&= \frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b\|\boldsymbol{\xi}\|^{d-1}} \mathbb{E}_{a,r}\left[\frac{1}{r}\mathcal{F}[\boldsymbol{g}_1]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right) \cdot \mathcal{F}[\boldsymbol{g}_1]\left(-\frac{\|\boldsymbol{\xi}\|}{r}\right)\right] \mathcal{F}[u_\rho](\boldsymbol{\xi}).
\end{aligned} \tag{SM2.13}$$

3. Calculate $\hat{K}_{\boldsymbol{w}}(\boldsymbol{\xi}, \boldsymbol{\xi}')$. Since

$$\hat{K}_{\boldsymbol{w}}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \frac{1}{4\pi^2} \nabla_{\boldsymbol{\xi}} \left[\delta_{\boldsymbol{w}}(\boldsymbol{\xi}) \mathcal{F}[g_2] \left(\frac{\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|} \right) \mathrm{e}^{2\pi \mathrm{i} b \boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}/\|\boldsymbol{w}\|} \right] \cdot \nabla_{\boldsymbol{\xi}'} \left[\delta_{\boldsymbol{w}}(\boldsymbol{\xi}') \overline{\mathcal{F}[g_2] \left(\frac{\boldsymbol{\xi}'^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|} \right)} \mathrm{e}^{-2\pi \mathrm{i} b \boldsymbol{\xi}'^\intercal \hat{\boldsymbol{w}}/\|\boldsymbol{w}\|} \right],$$

we have

$$\begin{aligned}
&\langle \hat{K}_{\boldsymbol{w}}, \phi \otimes \psi \rangle \\
&= \frac{1}{4\pi^2} \int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}) \nabla_{\boldsymbol{\xi}} \left[\delta_{\boldsymbol{w}}(\boldsymbol{\xi}) \mathcal{F}[g_2] \left(\frac{\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|} \right) \mathrm{e}^{2\pi \mathrm{i} b \boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}/\|\boldsymbol{w}\|} \right] \mathrm{d}\boldsymbol{\xi} \\
&\quad \cdot \int_{\mathbb{R}^d} \psi(\boldsymbol{\xi}') \nabla_{\boldsymbol{\xi}'} \left[\delta_{\boldsymbol{w}}(\boldsymbol{\xi}') \overline{\mathcal{F}[g_2] \left(\frac{\boldsymbol{\xi}'^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|} \right)} \mathrm{e}^{-2\pi \mathrm{i} b \boldsymbol{\xi}'^\intercal \hat{\boldsymbol{w}}/\|\boldsymbol{w}\|} \right] \mathrm{d}\boldsymbol{\xi}' \\
&= \frac{1}{4\pi^2} \int_{\mathbb{R}^d} \nabla_{\boldsymbol{\xi}} \phi(\boldsymbol{\xi}) \delta_{\boldsymbol{w}}(\boldsymbol{\xi}) \mathcal{F}[g_2] \left(\frac{\boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|} \right) \mathrm{e}^{2\pi \mathrm{i} b \boldsymbol{\xi}^\intercal \hat{\boldsymbol{w}}/\|\boldsymbol{w}\|} \, \mathrm{d}\boldsymbol{\xi} \\
&\quad \cdot \int_{\mathbb{R}^d} \nabla_{\boldsymbol{\xi}'} \psi(\boldsymbol{\xi}') \delta_{\boldsymbol{w}}(\boldsymbol{\xi}') \overline{\mathcal{F}[g_2] \left(\frac{\boldsymbol{\xi}'^\intercal \hat{\boldsymbol{w}}}{\|\boldsymbol{w}\|} \right)} \mathrm{e}^{-2\pi \mathrm{i} b \boldsymbol{\xi}'^\intercal \hat{\boldsymbol{w}}/\|\boldsymbol{w}\|} \, \mathrm{d}\boldsymbol{\xi}' \\
&= \int_{\mathbb{R}\times\mathbb{R}} \nabla\phi(\eta \boldsymbol{w}) \cdot \nabla\psi(\eta' \boldsymbol{w}) \mathcal{F}[g_2](\eta) \cdot \overline{\mathcal{F}[g_2](\eta')} \mathrm{e}^{2\pi \mathrm{i} b(\eta - \eta')} \, \mathrm{d}\eta \, \mathrm{d}\eta'.
\end{aligned}$$

By the same computation as for $\hat{K}_{a,b}(\boldsymbol{\xi}, \boldsymbol{\xi}')$, we can get

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{w},b} \left[\langle \hat{K}_{\boldsymbol{w}}, \phi \otimes \psi \rangle \right] \\
&= \frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b} \int_{\mathbb{R}^d} \nabla\phi(\boldsymbol{\zeta}) \cdot \int_{\mathbb{R}^+} \left[\frac{1}{r\|\boldsymbol{\zeta}\|^{d-1}} \mathcal{F}[g_2] \left(\frac{\|\boldsymbol{\zeta}\|}{r} \right) \cdot \overline{\mathcal{F}[g_2] \left(\frac{\|\boldsymbol{\zeta}\|}{r} \right)} \right] \nabla\psi(\boldsymbol{\zeta}) \rho_r(r) \, \mathrm{d}r \, \mathrm{d}\boldsymbol{\zeta} \\
&= \frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b} \int_{\mathbb{R}^d} \nabla\phi(\boldsymbol{\zeta}) \cdot \mathbb{E}_{a,r} \left[\frac{1}{r\|\boldsymbol{\zeta}\|^{d-1}} \mathcal{F}[g_2] \left(\frac{\|\boldsymbol{\zeta}\|}{r} \right) \cdot \overline{\mathcal{F}[g_2] \left(\frac{\|\boldsymbol{\zeta}\|}{r} \right)} \right] \nabla\psi(\boldsymbol{\zeta}) \, \mathrm{d}\boldsymbol{\zeta} \\
&= -\frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b} \int_{\mathbb{R}^d} \phi(\boldsymbol{\zeta}) \nabla \cdot \left\{ \mathbb{E}_{a,r} \left[\frac{1}{r\|\boldsymbol{\zeta}\|^{d-1}} \mathcal{F}[g_2] \left(\frac{\|\boldsymbol{\zeta}\|}{r} \right) \cdot \overline{\mathcal{F}[g_2] \left(\frac{\|\boldsymbol{\zeta}\|}{r} \right)} \right] \nabla\psi(\boldsymbol{\zeta}) \right\} \mathrm{d}\boldsymbol{\zeta}.
\end{aligned}$$

Thus taking $\psi(\boldsymbol{\xi}) = \mathcal{F}[u_\rho](\boldsymbol{\xi})$, we have

(SM2.14)
$$\mathcal{L}_{\boldsymbol{w}}[\mathcal{F}[u_\rho](\boldsymbol{\xi})] = -\frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b} \nabla \cdot \left(\mathbb{E}_{a,r} \left[\frac{1}{r\|\boldsymbol{\xi}\|^{d-1}} \mathcal{F}[g_2] \left(\frac{\|\boldsymbol{\xi}\|}{r} \right) \mathcal{F}[g_2] \left(-\frac{\|\boldsymbol{\xi}\|}{r} \right) \right] \nabla \mathcal{F}[u_\rho](\boldsymbol{\xi}) \right).$$

Finally, one can plug (SM2.13) and (SM2.14) into (SM2.5) and obtain the dynamics (SM2.7).■

Corollary. *(Corollary 1 in main text) Suppose that Assumption 1 holds. If* $\sigma_b \gg 1$ *and* $\sigma = \mathrm{ReLU}$, *then the dynamics (4.9) has the following expression,*

$$\langle \partial_t \mathcal{F}[u], \phi \rangle = -\langle \mathcal{L}[\mathcal{F}[u_\rho]], \phi \rangle + O(\sigma_b^{-3}), \tag{SM2.15}$$

where $\phi \in \mathcal{S}(\mathbb{R}^d)$ is a test function and the LFP operator reads as

$$\begin{aligned}\mathcal{L}[\mathcal{F}[u_\rho]] = &\frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b}\mathbb{E}_{a,r}\left[\frac{r^3}{16\pi^4\|\boldsymbol{\xi}\|^{d+3}}+\frac{a^2 r}{4\pi^2\|\boldsymbol{\xi}\|^{d+1}}\right]\mathcal{F}[u_\rho](\boldsymbol{\xi})\\ &-\frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b}\nabla\cdot\left(\mathbb{E}_{a,r}\left[\frac{a^2 r}{4\pi^2\|\boldsymbol{\xi}\|^{d+1}}\right]\nabla\mathcal{F}[u_\rho](\boldsymbol{\xi})\right).\end{aligned} \tag{SM2.16}$$

The expectations are taken w.r.t. initial parameter distribution. Here $r = \|\boldsymbol{w}\|$ with the probability density $\rho_r(r) := \frac{2\pi^{d/2}}{\Gamma(d/2)}\rho_{\boldsymbol{w}}(r\boldsymbol{e}_1)r^{d-1}$, $\boldsymbol{e}_1 = (1,0,\cdots,0)^\intercal$.

Proof. Let

$$f_a(\boldsymbol{x}) := \nabla_a\left[a\mathrm{ReLU}(\boldsymbol{w}\cdot\boldsymbol{x}+b)\right] = \mathrm{ReLU}(\boldsymbol{w}\cdot\boldsymbol{x}+b), \tag{SM2.17}$$

$$g_a(z) := \mathrm{ReLU}(z), \tag{SM2.18}$$

$$f_b(\boldsymbol{x}) := \nabla_b\left[a\mathrm{ReLU}(\boldsymbol{w}\cdot\boldsymbol{x}+b)\right] = aH(\boldsymbol{w}\cdot\boldsymbol{x}+b), \tag{SM2.19}$$

$$g_b(z) := aH(z), \tag{SM2.20}$$

so $\boldsymbol{g}_1(z) = (g_a(z), g_b(z))^\intercal$ and $g_2(z) = g_b(z)$. Then

$$\mathcal{F}[g_a](\xi) = -\frac{1}{4\pi^2\xi^2}+\frac{\mathrm{i}}{4\pi}\delta'(\xi), \tag{SM2.21}$$

$$\mathcal{F}[g_b](\xi) = a\left[\frac{1}{\mathrm{i}2\pi\xi}+\frac{1}{2}\delta(\xi)\right], \tag{SM2.22}$$

By ignoring all $\delta(\xi)$ and $\delta'(\xi)$ related to only the trivial $\mathbf{0}$-frequency, we obtain

$$\frac{1}{r}\mathcal{F}[g_a]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right)\mathcal{F}[g_a]\left(\frac{-\|\boldsymbol{\xi}\|}{r}\right) = \frac{r^3}{16\pi^4\|\boldsymbol{\xi}\|^4}, \tag{SM2.23}$$

$$\frac{1}{r}\mathcal{F}[g_b]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right)\mathcal{F}[g_b]\left(\frac{-\|\boldsymbol{\xi}\|}{r}\right) = \frac{a^2 r}{4\pi^2\|\boldsymbol{\xi}\|^2}. \tag{SM2.24}$$

We then obtain (SM2.16) by plugging these into (SM2.8). ∎

Corollary. *(Corollary 2 in main text) Suppose that Assumption 1 holds. If $\sigma_b \gg 1$ and $\sigma = \tanh$, then the dynamics (4.9) has the following expression,*

$$\langle\partial_t\mathcal{F}[u],\phi\rangle = -\langle\mathcal{L}[\mathcal{F}[u_\rho]],\phi\rangle + O(\sigma_b^{-3}), \tag{SM2.25}$$

where $\phi \in \mathcal{S}(\mathbb{R}^d)$ is a test function and the LFP operator reads as

$$\begin{aligned}\mathcal{L}[\mathcal{F}[u_\rho]] = &\frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b\|\boldsymbol{\xi}\|^{d-1}}\mathbb{E}_{a,r}\left[\frac{\pi^2}{r}\mathrm{csch}^2\left(\frac{\pi^2\|\boldsymbol{\xi}\|}{r}\right)+\frac{4\pi^4 a^2\|\boldsymbol{\xi}\|^2}{r^3}\mathrm{csch}^2\left(\frac{\pi^2\|\boldsymbol{\xi}\|}{r}\right)\right]\mathcal{F}[u_\rho](\boldsymbol{\xi})\\ &-\frac{\Gamma(d/2)}{2\sqrt{2}\pi^{(d+1)/2}\sigma_b}\nabla\cdot\left(\mathbb{E}_{a,r}\left[\frac{4\pi^4 a^2}{r^3\|\boldsymbol{\xi}\|^{d-3}}\mathrm{csch}^2\left(\frac{\pi^2\|\boldsymbol{\xi}\|}{r}\right)\right]\nabla\mathcal{F}[u_\rho](\boldsymbol{\xi})\right).\end{aligned} \tag{SM2.26}$$

The expectations are taken w.r.t. initial parameter distribution. Here $r = \|\boldsymbol{w}\|$ with the probability density $\rho_r(r) := \frac{2\pi^{d/2}}{\Gamma(d/2)}\rho_{\boldsymbol{w}}(r\boldsymbol{e}_1)r^{d-1}$, $\boldsymbol{e}_1 = (1,0,\cdots,0)^\intercal$.

Proof. Let

$$f_a(\boldsymbol{x}) := \nabla_a \left[a \tanh(\boldsymbol{w}\cdot\boldsymbol{x}+b)\right] = \tanh(\boldsymbol{w}\cdot\boldsymbol{x}+b), \tag{SM2.27}$$
$$g_a(z) := \tanh(z), \tag{SM2.28}$$
$$f_b(\boldsymbol{x}) := \nabla_b \left[a \tanh(\boldsymbol{w}\cdot\boldsymbol{x}+b)\right] = a\mathrm{sech}^2(\boldsymbol{w}\cdot\boldsymbol{x}+b), \tag{SM2.29}$$
$$g_b(z) := a\mathrm{sech}^2(z), \tag{SM2.30}$$

so $\boldsymbol{g}_1(z) = (g_a(z), g_b(z))^\intercal$ and $g_2(z) = g_b(z)$. Then

$$\mathcal{F}[g_a](\xi) = -\mathrm{i}\pi\mathrm{csch}(\pi^2\xi), \tag{SM2.31}$$
$$\mathcal{F}[g_b](\xi) = 2\pi^2 a\xi\mathrm{csch}(\pi^2\xi). \tag{SM2.32}$$

By ignoring all $\delta(\xi)$ and $\delta'(\xi)$ related to only the trivial $\mathbf{0}$-frequency, we obtain

$$\frac{1}{r}\mathcal{F}[g_a]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right)\mathcal{F}[g_a]\left(\frac{-\|\boldsymbol{\xi}\|}{r}\right) = \frac{\pi^2}{r}\mathrm{csch}^2\left(\frac{\pi^2\|\boldsymbol{\xi}\|}{r}\right), \tag{SM2.33}$$
$$\frac{1}{r}\mathcal{F}[g_b]\left(\frac{\|\boldsymbol{\xi}\|}{r}\right)\mathcal{F}[g_b]\left(\frac{-\|\boldsymbol{\xi}\|}{r}\right) = \frac{4\pi^4 a^2\|\boldsymbol{\xi}\|^2}{r^3}\mathrm{csch}^2\left(\frac{\pi^2\|\boldsymbol{\xi}\|}{r}\right). \tag{SM2.34}$$

We then obtain (SM2.26) by plugging these into (4.17). ■

Lemma. *(Lemma 4 in main text) Suppose that H_1 and H_2 are two separable Hilbert spaces and $\mathcal{P}: H_1 \to H_2$ and $\mathcal{P}^*: H_2 \to H_1$ is the adjoint of $\mathcal{P}$. Then all eigenvalues of $\mathcal{P}^*\mathcal{P}$ and $\mathcal{P}\mathcal{P}^*$ are non-negative. Moreover, they have the same positive spectrum. If in particular, we assume that the operator $\mathcal{P}\mathcal{P}^*$ is surjective, then the operator $\mathcal{P}\mathcal{P}^*$ is invertible.*

Proof. We consider the eigenvalue problem $\mathcal{P}^*\mathcal{P}\phi_1 = \lambda\phi_1$. Taking inner product with ϕ_1, we have $\langle\phi_1, \mathcal{P}^*\mathcal{P}\phi_1\rangle_{H_1} = \lambda\|\phi_1\|^2_{H_1}$. Note that the left hand side is $\|\mathcal{P}\phi_1\|^2_{H_2}$ which is non-negative. Thus $\lambda \geq 0$. Similarly, the eigenvalues of $\mathcal{P}\mathcal{P}^*$ are also non-negative.

Now if $\mathcal{P}^*\mathcal{P}$ has a positive eigenvalue $\lambda > 0$, then $\mathcal{P}^*\mathcal{P}\phi_1 = \lambda\phi_1$ with non-zero vector $\phi_1 \in H_1$. It follows that $\mathcal{P}\mathcal{P}^*(\mathcal{P}\phi_1) = \lambda(\mathcal{P}\phi_1)$. It is sufficient to prove that $\mathcal{P}\phi_1$ is non-zero. Indeed, if $\mathcal{P}\phi_1 = 0$, then $\mathcal{P}^*\mathcal{P}\phi_1 = 0$ and $\lambda = 0$ which contradicts with our assumption. Therefore, any positive eigenvalue of $\mathcal{P}^*\mathcal{P}$ is an eigenvalue of $\mathcal{P}\mathcal{P}^*$. Similarly, any positive eigenvalue of $\mathcal{P}\mathcal{P}^*$ is an eigenvalue of $\mathcal{P}^*\mathcal{P}$.

Next, suppose that $\mathcal{P}\mathcal{P}^*$ is surjective. We show that $\mathcal{P}\mathcal{P}^*\phi_2 = 0$ has only the trivial solution $\phi_2 = 0$. In fact, $\mathcal{P}\mathcal{P}^*\phi_2 = 0$ implies that $\|\mathcal{P}^*\phi_2\|^2_{H_1} = \langle\phi_2, \mathcal{P}\mathcal{P}^*\phi_2\rangle_{H_2} = 0$, i.e., $\mathcal{P}^*\phi_2 = 0$. Thanks to the surjectivity of $\mathcal{P}\mathcal{P}^*$, there exists a vector $\phi_3 \in H_2$ such that $\phi_2 = \mathcal{P}\mathcal{P}^*\phi_3$. Let $\phi_1 = \mathcal{P}^*\phi_3 \in H_1$. Hence $\phi_2 = \mathcal{P}\phi_1$ and $\mathcal{P}^*\mathcal{P}h_1 = 0$. Taking inner product with ϕ_1, we have $\|\mathcal{P}\phi_1\|^2_{H_2} = \langle\phi_1, \mathcal{P}^*\mathcal{P}\phi_1\rangle_{H_1} = 0$, i.e., $\phi_2 = \mathcal{P}\phi_1 = 0$. Therefore $\mathcal{P}\mathcal{P}^*$ is injective. This with the surjectivity assumption of $\mathcal{P}\mathcal{P}^*$ leads to that $\mathcal{P}\mathcal{P}^*$ is invertible. ■

Theorem. *(Theorem 2 in main text) Suppose that $\mathcal{P}\mathcal{P}^*$ is surjective. The above Problems (i) and (ii) are equivalent in the sense that $\phi_\infty = \phi_{\min}$. More precisely, we have*

$$\phi_\infty = h_{\min} = \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^{-1}(g - \mathcal{P}\phi_{\mathrm{ini}}) + \phi_{\mathrm{ini}}. \tag{SM2.35}$$

Proof. Let $\tilde{\phi} = \phi - \phi_{\text{ini}}$ and $\tilde{g} = g - \mathcal{P}\phi_{\text{ini}}$. Then it is sufficient to show the following problems (i') and (ii') are equivalent.

(i') The initial value problem

$$\begin{cases} \frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}t} = \mathcal{P}^*(\tilde{g} - \mathcal{P}\tilde{\phi}) \\ \tilde{\phi}(0) = 0. \end{cases}$$

(ii') The minimization problem

$$\min_{\tilde{\phi}} \|\tilde{\phi}\|_{H_1}^2, \quad \text{s.t.} \quad \mathcal{P}\tilde{\phi} = \tilde{g}.$$

We claim that $\tilde{\phi}_{\min} = \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g}$. Thanks to Lemma 4, $\mathcal{P}\mathcal{P}^*$ is invertible, and thus $\phi_{\min}$ is well-defined and satisfies that $\mathcal{P}\tilde{\phi} = \tilde{g}$. It remains to show that this solution is unique. In fact, for any $\tilde{\phi}$ satisfying $\mathcal{P}\tilde{\phi} = \tilde{g}$, we have

$$\begin{aligned} \langle \tilde{\phi} - \tilde{\phi}_{\min}, \tilde{\phi}_{\min} \rangle_{H_1} &= \langle \tilde{\phi} - \tilde{\phi}_{\min}, \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g} \rangle_{H_1} \\ &= \langle \mathcal{P}(\tilde{\phi} - \tilde{\phi}_{\min}), (\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g} \rangle_{H_2} \\ &= \langle \mathcal{P}\tilde{\phi}, (\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g} \rangle_{H_2} - \langle \mathcal{P}\tilde{\phi}_{\min}, (\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g} \rangle_{H_2} \\ &= 0. \end{aligned}$$

Therefore,

$$\|\tilde{\phi}\|_{H_1}^2 = \|\tilde{\phi}_{\min}\|_{H_1}^2 + \|\tilde{\phi} - \tilde{\phi}_{\min}\|_{H_1}^2 \geq \|\tilde{\phi}_{\min}\|_{H_1}^2.$$

The equality holds if and only if $\tilde{\phi} = \tilde{\phi}_{\min}$.

For problem (i'), from the theory of ordinary differential equations on Hilbert spaces, we have that its solution can be written as

$$\tilde{\phi}(t) = \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g} + \sum_{i \in I} c_i v_i \exp(-\lambda_i t),$$

where λ_i, $i \in \mathcal{I}$ are positive eigenvalues of $\mathcal{P}\mathcal{P}^*$, $\mathcal{I}$ is an index set with at most countable cardinality, and v_i, $i \in \mathcal{I}$ are eigenvectors in H_1. Thus $\tilde{\phi}_\infty = \tilde{\phi}_{\min} = \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g}$.

Finally, by back substitution, we have

$$\phi_\infty = \phi_{\min} = \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^{-1}\tilde{g} + \phi_0 = \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^{-1}(g - \mathcal{P}\phi_{\text{ini}}) + \phi_{\text{ini}}.$$

∎

Corollary. *(Corollary 4 in main text) Let H_1 and H_2 be two separable Hilbert spaces and $\Gamma : H_1 \to H_1$ be an injective operator. Define the Hilbert space $H_\Gamma := \text{Im}(\Gamma)$. Let $g \in H_2$ and $\mathcal{P} : H_\Gamma \to H_2$ be an operator such that $\mathcal{P}\mathcal{P}^* : H_2 \to H_2$ is surjective. Then $\Gamma^{-1} : H_\Gamma \to H_1$ exists and H_Γ is a Hilbert space with norm $\|\phi\|_{H_\Gamma} := \|\Gamma^{-1}\phi\|_{H_1}$. Moreover, the following two problems are equivalent in the sense that $\phi_\infty = \phi_{\min}$.*

(B1) *The initial value problem*

$$\begin{cases} \dfrac{\mathrm{d}\phi}{\mathrm{d}t} = \Gamma\Gamma^*\mathcal{P}^*(g - \mathcal{P}\phi) \\ \phi(0) = \phi_{\mathrm{ini}}. \end{cases}$$

(B2) *The minimization problem*

$$\min_{\phi-\phi_0\in H_\Gamma} \|\phi - \phi_{\mathrm{ini}}\|_{H_\Gamma},$$
$$s.t. \quad \mathcal{P}\phi = g.$$

Proof. The operator $\Gamma : H_1 \to H_\Gamma$ is bijective. Hence $\Gamma^{-1} : H_\Gamma \to H_1$ is well-defined and H_Γ with norm $\|\cdot\|_{H_\Gamma}$ is a Hilbert space. The equivalence result holds by applying Theorem 2 with proper replacements. More precisely, we replace ϕ by $\Gamma^{-1}\phi$ and $\mathcal{P}$ by $\mathcal{P}\Gamma$. ∎

Corollary. *(Corollary 5 in main text) Let $\gamma : \mathbb{R}^d \to \mathbb{R}^+$ be a positive function, h be a function in $L^2(\mathbb{R}^d)$ and $\phi = \mathcal{F}[h]$. The operator $\Gamma : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is defined by $[\Gamma\phi](\boldsymbol{\xi}) = \gamma(\boldsymbol{\xi})\phi(\boldsymbol{\xi})$, $\boldsymbol{\xi} \in \mathbb{R}^d$. Define the Hilbert space $H_\Gamma := \mathrm{Im}(\Gamma)$. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\intercal \in \mathbb{R}^{n\times d}$, $\boldsymbol{Y} = (y_1, \ldots, y_n)^\intercal \in \mathbb{R}^n$ and $\mathcal{P} : H_\Gamma \to \mathbb{R}^n$ be a surjective operator*

$$\mathcal{P} : \phi \mapsto \left(\int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}) \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{x}_1^\intercal\boldsymbol{\xi}} \,\mathrm{d}\boldsymbol{\xi}, \ldots, \int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}) \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{x}_n^\intercal\boldsymbol{\xi}} \,\mathrm{d}\boldsymbol{\xi} \right)^\intercal = (h(\boldsymbol{x}_1), \ldots, h(\boldsymbol{x}_n))^\intercal. \tag{SM2.36}$$

Then the following two problems are equivalent in the sense that $\phi_\infty = \phi_{\min}$.

(C1) *The initial value problem*

$$\begin{cases} \dfrac{\mathrm{d}\phi(\boldsymbol{\xi})}{\mathrm{d}t} = (\gamma(\boldsymbol{\xi}))^2 \displaystyle\sum_{i=1}^n \left(y_i \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{x}_i^\intercal\boldsymbol{\xi}} - \left[\phi * \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{x}_i^\intercal(\cdot)} \right](\boldsymbol{\xi}) \right) \\ \phi(0) = \phi_{\mathrm{ini}}. \end{cases}$$

(C2) *The minimization problem*

$$\min_{\phi-\phi_{\mathrm{ini}}\in H_\Gamma} \int_{\mathbb{R}^d} (\gamma(\boldsymbol{\xi}))^{-2} |\phi(\boldsymbol{\xi}) - \phi_{\mathrm{ini}}(\boldsymbol{\xi})|^2 \,\mathrm{d}\boldsymbol{\xi},$$
$$s.t. \quad h(\boldsymbol{x}_i) = y_i, \quad i = 1, \cdots, n.$$

Proof. Let $H_1 = L^2(\mathbb{R}^d)$, $H_2 = \mathbb{R}^n$, $g = \boldsymbol{Y}$. By definition, Γ is injective. Then by Corollary 4, we have that $\Gamma^{-1} : H_\Gamma \to L^2(\mathbb{R}^d)$ exists and H_Γ is a Hilbert space with norm $\|\phi\|_{H_\Gamma} := \|\Gamma^{-1}\phi\|_{L^2(\mathbb{R}^d)}$. Moreover, $\|\phi - \phi_{\mathrm{ini}}\|_{H_\Gamma}^2 = \int_{\mathbb{R}^d} (\gamma(\boldsymbol{\xi}))^{-2} |\phi(\boldsymbol{\xi}) - \phi_{\mathrm{ini}}(\boldsymbol{\xi})|^2 \,\mathrm{d}\boldsymbol{\xi}$. We note that $[\mathcal{P}^* Y](\boldsymbol{\xi}) = \sum_{i=1}^n y_i \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{x}_i^\intercal\boldsymbol{\xi}}$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$. Thus

$$\begin{aligned} [\mathcal{P}^*\mathcal{P}\phi](\boldsymbol{\xi}) &= \left[\mathcal{P}^* \left(\int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}') \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{x}_i^\intercal\boldsymbol{\xi}'} \,\mathrm{d}\boldsymbol{\xi}' \right)_{i=1}^n \right](\boldsymbol{\xi}) \\ &= \sum_{i=1}^n \int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}') \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{x}_i^\intercal\boldsymbol{\xi}'} \,\mathrm{d}\boldsymbol{\xi}' \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{x}_i^\intercal\boldsymbol{\xi}} \end{aligned}$$

$$= \sum_{i=1}^{n} \int_{\mathbb{R}^d} \phi(\boldsymbol{\xi}') \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal (\boldsymbol{\xi}-\boldsymbol{\xi}')} \,\mathrm{d}\boldsymbol{\xi}'$$

$$= \sum_{i=1}^{n} \left[\phi * \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal (\cdot)}\right](\boldsymbol{\xi}).$$

The equivalence result then follows from Corollary 4. ■

Corollary. *(Corollary 6 in main text) Let* $\gamma : \mathbb{Z}^d \to \mathbb{R}^+$ *be a positive function defined on lattice* $\mathbb{Z}^d$ *and* $\phi = \mathcal{F}[h]$*. The operator* $\Gamma : \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ *is defined by* $[\Gamma\phi](\boldsymbol{k}) = \gamma(\boldsymbol{k})\phi(\boldsymbol{k})$, $\boldsymbol{k} \in \mathbb{Z}^d$*. Here* $\ell^2(\mathbb{Z}^d)$ *is set of square summable functions on the lattice* $\mathbb{Z}^d$*. Define the Hilbert space* $H_\Gamma := \mathrm{Im}(\Gamma)$*. Let* $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\intercal \in \mathbb{T}^{n\times d}$, $Y = (y_1, \ldots, y_n)^\intercal \in \mathbb{R}^n$ *and* $\mathcal{P} : H_\Gamma \to \mathbb{R}^n$ *be a surjective operator such as*

$$P : \phi \mapsto \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \phi(\boldsymbol{k}) \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{x}_1^\intercal \boldsymbol{k}}, \ldots, \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \phi(\boldsymbol{k}) \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{x}_n^\intercal \boldsymbol{k}}\right)^\intercal. \tag{SM2.37}$$

Then the following two problems are equivalent in the sense that $\phi_\infty = \phi_{\min}$*.*

(D1) *The initial value problem*

$$\begin{cases} \dfrac{\mathrm{d}\phi(\boldsymbol{k})}{\mathrm{d}t} = (\gamma(\boldsymbol{k}))^2 \displaystyle\sum_{i=1}^{n} \left(y_i \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal \boldsymbol{k}} - \left[\phi * \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal (\cdot)}\right](\boldsymbol{k})\right) \\ \phi(\boldsymbol{0}) = \phi_{\mathrm{ini}}. \end{cases}$$

(D2) *The minimization problem*

$$\min_{\phi-\phi_{\mathrm{ini}}\in H_\Gamma} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} (\gamma(\boldsymbol{k}))^{-2} |\phi(\boldsymbol{k}) - \phi_{\mathrm{ini}}(\boldsymbol{k})|^2,$$
$$\text{s.t.} \quad h(\boldsymbol{x}_i) = y_i, \quad i = 1, \cdots, n.$$

Proof. Let $H_1 = \ell^2(\mathbb{Z}^d)$, $H_2 = \mathbb{R}^n$, and $g = \boldsymbol{Y}$. By definition, Γ is injective. Then by Corollary 4, we have that $\Gamma^{-1} : H_\Gamma \to \ell^2(\mathbb{Z}^d)$ exists and H_Γ is a Hilbert space with norm $\|\phi\|_{H_\Gamma} := \|\Gamma^{-1}\phi\|_{\ell^2(\mathbb{Z}^d)}$. Moreover, $\|\phi - \phi_{\mathrm{ini}}\|_{H_\Gamma}^2 = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} (\gamma(\boldsymbol{k}))^{-2} |\phi(\boldsymbol{k}) - \phi_{\mathrm{ini}}(\boldsymbol{k})|^2$. We note that $[P^*\boldsymbol{Y}](\boldsymbol{k}) = \sum_{i=1}^{n} y_i \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal \boldsymbol{k}}$ for all $\boldsymbol{k} \in \mathbb{Z}^d$. Thus

$$[P^*P\phi](\boldsymbol{k}) = \left[P^* \left(\sum_{\boldsymbol{k}'\in\mathbb{Z}^d} \phi(\boldsymbol{k}') \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{x}_i^\intercal \boldsymbol{k}'}\right)_{i=1}^{n}\right](\boldsymbol{k})$$

$$= \sum_{i=1}^{n} \sum_{\boldsymbol{k}'\in\mathbb{Z}^d} \phi(\boldsymbol{k}') \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{x}_i^\intercal \boldsymbol{k}'} \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal \boldsymbol{k}}$$

$$= \sum_{i=1}^{n} \sum_{\boldsymbol{k}'\in\mathbb{Z}^d} \phi(\boldsymbol{k}') \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal (\boldsymbol{k}-\boldsymbol{k}')}$$

$$= \sum_{i=1}^{n} \left[\phi * \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{x}_i^\intercal (\cdot)}\right](\boldsymbol{k}).$$

The equivalence result then follows from Corollary 4. ∎

Lemma. *(Lemma 5 in main text)* (i) *For $\mathcal{H}_Q = \{h : \|h\|_\gamma \le Q\}$ with $\gamma : \mathbb{Z}^d \to \mathbb{R}^+$, we have*

$$\text{(SM2.38)}\qquad \mathrm{Rad}_S(\mathcal{H}_Q) \le \frac{1}{\sqrt{n}} Q \|\gamma\|_{\ell^2}.$$

(ii) *For $\mathcal{H}'_Q = \{h : \|h\|_\gamma \le Q, |\mathcal{F}[h](\mathbf{0})| \le c_0\}$ with $\gamma : \mathbb{Z}^{d*} \to \mathbb{R}^+$ and $\gamma^{-1}(\mathbf{0}) := 0$, we have*

$$\text{(SM2.39)}\qquad \mathrm{Rad}_S(\mathcal{H}'_Q) \le \frac{c_0}{\sqrt{n}} + \frac{1}{\sqrt{n}} Q \|\gamma\|_{\ell^2}.$$

Proof. We first prove (ii) since it is more involved. By the definition of the Rademacher complexity

$$\text{(SM2.40)}\qquad \mathrm{Rad}_S(\mathcal{H}'_Q) = \frac{1}{n}\mathbb{E}_{\boldsymbol{\tau}}\left[\sup_{h\in\mathcal{H}'_Q}\sum_{i=1}^n \tau_i h(\boldsymbol{x}_i)\right].$$

Let $\tau(\boldsymbol{x}) = \sum_{i=1}^n \tau_i \delta(\boldsymbol{x} - \boldsymbol{x}_i)$, where τ_i's are i.i.d. random variables with $\mathbb{P}(\tau_i = 1) = \mathbb{P}(\tau_i = -1) = \frac{1}{2}$. We have $\mathcal{F}[\tau](\boldsymbol{k}) = \int_\Omega \sum_{i=1}^n \tau_i \delta(\boldsymbol{x} - \boldsymbol{x}_i) \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k}^\intercal \boldsymbol{x}} \,\mathrm{d}\boldsymbol{x} = \sum_{i=1}^n \tau_i \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{k}^\intercal \boldsymbol{x}_i}$. Note that

$$\text{(SM2.41)}\qquad \sup_{h\in\mathcal{H}'_Q}\sum_{i=1}^n \tau_i h(\boldsymbol{x}_i) = \sup_{h\in\mathcal{H}'_Q}\sum_{i=1}^n \tau_i \bar{h}(\boldsymbol{x}_i) = \sup_{h\in\mathcal{H}'_Q}\sum_{i=1}^n \tau_i \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \overline{\mathcal{F}[h](\boldsymbol{k})} \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{k}^\intercal\boldsymbol{x}_i}$$

$$\text{(SM2.42)}\qquad = \sup_{h\in\mathcal{H}'_Q}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \mathcal{F}[\tau](\boldsymbol{k})\overline{\mathcal{F}[h](\boldsymbol{k})}.$$

By the Cauchy–Schwarz inequality,

$$\sup_{h\in\mathcal{H}'_Q}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \mathcal{F}[\tau](\boldsymbol{k})\overline{\mathcal{F}[h](\boldsymbol{k})}$$

$$\text{(SM2.43)}\qquad \le \sup_{h\in\mathcal{H}_Q}\left[\mathcal{F}[\tau](\mathbf{0})\overline{\mathcal{F}[h](\mathbf{0})} + \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}} (\gamma(\boldsymbol{k}))^2 |\mathcal{F}[\tau](\boldsymbol{k})|^2\right)^{1/2} \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}} (\gamma(\boldsymbol{k}))^{-2} |\overline{\mathcal{F}[h](\boldsymbol{k})}|^2\right)^{1/2}\right]$$

$$\text{(SM2.44)}\qquad \le c_0 |\mathcal{F}[\tau](\mathbf{0})| + Q\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}} (\gamma(\boldsymbol{k}))^2 |\mathcal{F}[\tau](\boldsymbol{k})|^2\right)^{1/2}.$$

Since $\mathbb{E}_{\boldsymbol{\tau}}|\mathcal{F}[\tau](\mathbf{0})| \le (\mathbb{E}_{\boldsymbol{\tau}}|\mathcal{F}[\tau](\mathbf{0})|^2)^{1/2} = \sqrt{n}$, $\mathbb{E}_{\boldsymbol{\tau}}|\mathcal{F}[\tau](\boldsymbol{k})|^2 = \mathbb{E}_{\boldsymbol{\tau}}\sum_{i,j=1}^n \tau_i\tau_j \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{k}^\intercal(\boldsymbol{x}_i - \boldsymbol{x}_j)} = n$, we obtain

$$\mathbb{E}_{\boldsymbol{\tau}}\left[\sup_{h\in\mathcal{H}'_Q}\sum_{i=1}^{n}\tau_i h(\boldsymbol{x}_i)\right] \le c_0\sqrt{n} + Q\mathbb{E}_{\boldsymbol{\tau}}\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}}(\gamma(\boldsymbol{k}))^2|\mathcal{F}[\tau](\boldsymbol{k})|^2\right)^{1/2} \tag{SM2.45}$$

$$\le c_0\sqrt{n} + Q\left(\mathbb{E}_{\boldsymbol{\tau}}\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}}(\gamma(\boldsymbol{k}))^2|\mathcal{F}[\tau](\boldsymbol{k})|^2\right)^{1/2} \tag{SM2.46}$$

$$= c_0\sqrt{n} + Q\sqrt{n}\|\gamma\|_{\ell^2}. \tag{SM2.47}$$

This leads to

$$\mathrm{Rad}_S(\mathcal{H}'_Q) \le \frac{c_0}{\sqrt{n}} + \frac{1}{\sqrt{n}}Q\|\gamma\|_{\ell^2}. \tag{SM2.48}$$

For (ii), the proof is similar to (i). We have

$$\mathbb{E}_{\boldsymbol{\tau}}\left[\sup_{h\in\mathcal{H}_Q}\sum_{\boldsymbol{k}\in\mathbb{Z}^{d}}\mathcal{F}[\tau](\boldsymbol{k})\overline{\mathcal{F}[h](\boldsymbol{k})}\right] \le Q\mathbb{E}_{\boldsymbol{\tau}}\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d}}(\gamma(\boldsymbol{k}))^2|\mathcal{F}[\tau](\boldsymbol{k})|^2\right)^{1/2} \le Q\sqrt{n}\|\gamma\|_{\ell^2}. \tag{SM2.49}$$

Therefore

$$\mathrm{Rad}_S(\mathcal{H}_Q) \le \frac{1}{\sqrt{n}}Q\|\gamma\|_{\ell^2}. \tag{SM2.50}$$

∎

Lemma. *(Lemma 6 in main text) Suppose that the real-valued target function $f\in\mathcal{F}_\gamma(\Omega)$ and that the training dataset $\{(\boldsymbol{x}_i,y_i)\}_{i=1}^n$ satisfies $y_i=f(\boldsymbol{x}_i)$, $i=1,\cdots,n$. If $\gamma:\mathbb{Z}^d\to\mathbb{R}^+$, then there exists a unique solution h_n to the regularized model*

$$\min_{h-h_{\mathrm{ini}}\in\mathcal{F}_\gamma(\Omega)}\|h-h_{\mathrm{ini}}\|_\gamma,\quad s.t.\quad h(\boldsymbol{x}_i)=y_i,\quad i=1,\cdots,n. \tag{SM2.51}$$

Moreover, we have

$$\|h_n-h_{\mathrm{ini}}\|_\gamma \le \|f-h_{\mathrm{ini}}\|_\gamma. \tag{SM2.52}$$

Proof. By the definition of the FP-norm, we have $\|h_n-h_{\mathrm{ini}}\|_\gamma = \|\mathcal{F}[h]_n-\mathcal{F}[h]_{\mathrm{ini}}\|_{H_\Gamma}$. According to Corollary 6, the minimizer of problem (SM2.51) exists, i.e., h_n exists. Since the target function $f(x)$ satisfies the constraints $f(x_i)=y_i$, $i=1,\cdots,n$, we have $\|h_n-h_{\mathrm{ini}}\|_\gamma\le\|f-h_{\mathrm{ini}}\|_\gamma$. ∎

Lemma. *(Lemma 7 in main text) Suppose that the real-valued target function $f\in\mathcal{F}_\gamma(\Omega)$ and the training dataset $\{(\boldsymbol{x}_i,y_i)\}_{i=1}^n$ satisfies $y_i=f(\boldsymbol{x}_i)$, $i=1,\cdots,n$. If $\gamma:\mathbb{Z}^{d*}\to\mathbb{R}^+$ with $\gamma^{-1}(\mathbf{0}):=0$, then there exists a solution h_n to the regularized model*

$$\min_{h-h_{\mathrm{ini}}\in\mathcal{F}_\gamma(\Omega)}\|h-h_{\mathrm{ini}}\|_\gamma,\quad s.t.\quad h(\boldsymbol{x}_i)=y_i,\quad i=1,\cdots,n. \tag{SM2.53}$$

Moreover, we have

$$|\mathcal{F}[h_n-h_{\mathrm{ini}}](\mathbf{0})| \le \|f-h_{\mathrm{ini}}\|_\infty + \|f-h_{\mathrm{ini}}\|_\gamma\|\gamma\|_{\ell^2}. \tag{SM2.54}$$

Proof. Let $f' = f - h_{\text{ini}}$. Since $h_n(\boldsymbol{x}_i) - f(\boldsymbol{x}_i) = 0$ for $i = 1, \cdots, n$, we have $h_n(\boldsymbol{x}_i) - f'(\boldsymbol{x}_i) - h_{\text{ini}}(\boldsymbol{x}_i) = 0$. Therefore

$$|\mathcal{F}[h_n - h_{\text{ini}}](\mathbf{0})| = \left| f'(\boldsymbol{x}_i) - \sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}} \mathcal{F}[h_n - h_{\text{ini}}](\boldsymbol{k}) \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}^\intercal \boldsymbol{x}_i} \right| \tag{SM2.55}$$

$$\leq \|f'\|_\infty + \sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}} |\mathcal{F}[h_n - h_{\text{ini}}](\boldsymbol{k})| \tag{SM2.56}$$

$$\leq \|f'\|_\infty + \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}} (\gamma(\boldsymbol{k}))^2 \right)^{\frac{1}{2}} \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}} (\gamma(\boldsymbol{k}))^{-2} |\mathcal{F}[h_n - h_{\text{ini}}](\boldsymbol{k})|^2 \right)^{\frac{1}{2}} \tag{SM2.57}$$

$$\leq \|f'\|_\infty + \|h_n - h_{\text{ini}}\|_\gamma \|\gamma\|_{\ell^2} \tag{SM2.58}$$

$$\leq \|f'\|_\infty + \|f'\|_\gamma \|\gamma\|_{\ell^2}. \tag{SM2.59}$$

We remark that the last step is due to the same reason as Lemma 6. ■

Theorem. *(Theorem 3 in main text) Suppose that the real-valued target function $f \in \mathcal{F}_\gamma(\Omega)$, the training dataset $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ satisfies $y_i = f(\boldsymbol{x}_i)$, $i = 1, \cdots, n$, and h_n is the solution of the regularized model*

$$\min_{h - h_{\text{ini}} \in \mathcal{F}_\gamma(\Omega)} \|h - h_{\text{ini}}\|_\gamma, \quad s.t. \quad h(\boldsymbol{x}_i) = y_i, \quad i = 1, \cdots, n. \tag{SM2.60}$$

Then we have

(i) *given $\gamma : \mathbb{Z}^d \to \mathbb{R}^+$, for any $\delta \in (0,1)$, with probability at least $1 - \delta$ over the random training sample, the population risk has the bound*

$$R_{\mathcal{D}}(h_n) \leq \|f - h_{\text{ini}}\|_\gamma \|\gamma\|_{\ell^2} \left(\frac{2}{\sqrt{n}} + 4\sqrt{\frac{2\log(4/\delta)}{n}} \right). \tag{SM2.61}$$

(ii) *given $\gamma : \mathbb{Z}^{d*} \to \mathbb{R}^+$ with $\gamma(\mathbf{0})^{-1} := 0$, for any $\delta \in (0,1)$, with probability at least $1 - \delta$ over the random training sample, the population risk has the bound*

$$R_{\mathcal{D}}(h_n) \leq (\|f - h_{\text{ini}}\|_\infty + 2\|f - h_{\text{ini}}\|_\gamma \|\gamma\|_{\ell^2}) \left(\frac{2}{\sqrt{n}} + 4\sqrt{\frac{2\log(4/\delta)}{n}} \right). \tag{SM2.62}$$

Proof. Let $f' = f - h_{\text{ini}}$ and $Q = \|f'\|_\gamma$.

(i) Given $\gamma : \mathbb{Z}^d \to \mathbb{R}^+$, we set $\mathcal{H}_Q = \{h : \|h - h_{\text{ini}}\|_\gamma \leq Q\}$. According to Lemma 6, the solution of problem (SM2.60) $h_n \in \mathcal{H}_Q$. By the relation between generalization gap and Rademacher complexity [SM1, SM2],

$$|R_{\mathcal{D}}(h_n) - L_S(h_n)| \leq 2\mathrm{Rad}_S(\mathcal{H}_Q) + 2 \sup_{h,h'\in\mathcal{H}_Q} \|h - h'\|_\infty \sqrt{\frac{2\log(4/\delta)}{n}}. \tag{SM2.63}$$

One of the component can be bounded as follows

$$\sup_{h,h'\in\mathcal{H}_Q}\|h-h'\|_\infty \le \sup_{h\in\mathcal{H}_Q} 2\|h-h_{\mathrm{ini}}\|_\infty \tag{SM2.64}$$

$$\le \sup_{h\in\mathcal{H}_Q} 2\max_{\boldsymbol{x}}\left|\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\mathcal{F}[h-h_{\mathrm{ini}}](\boldsymbol{k})\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}^\intercal\boldsymbol{x}}\right| \tag{SM2.65}$$

$$\le \sup_{h\in\mathcal{H}_Q} 2\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\mathcal{F}[h-h_{\mathrm{ini}}](\boldsymbol{k})| \tag{SM2.66}$$

$$\le 2\sup_{h\in\mathcal{H}_Q}\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^d}(\gamma(\boldsymbol{k}))^2\right)^{\frac{1}{2}}\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^d}(\gamma(\boldsymbol{k}))^{-2}|\mathcal{F}[h-h_{\mathrm{ini}}](\boldsymbol{k})|^2\right)^{\frac{1}{2}} \tag{SM2.67}$$

$$\le 2Q\|\gamma\|_{\ell^2}. \tag{SM2.68}$$

By Lemma 5,

$$\mathrm{Rad}_S(\mathcal{H}_Q)\le\frac{1}{\sqrt{n}}Q\|\gamma\|_{\ell^2}. \tag{SM2.69}$$

By optimization problem (SM2.60), $L_S(h_n)\le L_S(f')=0$. Therefore we obtain

$$R_{\mathcal{D}}(h)\le\frac{2}{\sqrt{n}}\|f'\|_\gamma\|\gamma\|_{\ell^2}+4\|f'\|_\gamma\|\gamma\|_{\ell^2}\sqrt{\frac{2\log(4/\delta)}{n}}. \tag{SM2.70}$$

(ii) Given $\gamma:\mathbb{Z}^{d*}\to\mathbb{R}^+$ with $\gamma(\mathbf{0})^{-1}:=0$, set $c_0=\|f'\|_\infty+\|f'\|_\gamma\|\gamma\|_{\ell^2}$. By Lemma 5, 6, and 7, define $\mathcal{H}'_Q=\{h:\|h-h_{\mathrm{ini}}\|_\gamma\le Q,|\mathcal{F}[h-h_{\mathrm{ini}}](\mathbf{0})|\le c_0\}$, we obtain

$$\mathrm{Rad}_S(\mathcal{H}'_Q)\le\frac{1}{\sqrt{n}}\|f'\|_\infty+\frac{2}{\sqrt{n}}\|f'\|_\gamma\|\gamma\|_{\ell^2}. \tag{SM2.71}$$

Also

$$\sup_{h,h'\in\mathcal{H}'_Q}\|h-h'\|_\infty\le\sup_{h\in\mathcal{H}'_Q}2\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\mathcal{F}[h-h_{\mathrm{ini}}](\boldsymbol{k})| \tag{SM2.72}$$

$$\le 2\sup_{h\in\mathcal{H}'_Q}\left[|\mathcal{F}[h-h_{\mathrm{ini}}](\mathbf{0})|+\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}}(\gamma(\boldsymbol{k}))^2\right)^{\frac{1}{2}}\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^{d*}}(\gamma(\boldsymbol{k}))^{-2}|\mathcal{F}[h-h_{\mathrm{ini}}](\boldsymbol{k})|^2\right)^{\frac{1}{2}}\right] \tag{SM2.73}$$

$$\le 2\|f'\|_\infty+4\|f'\|_\gamma\|\gamma\|_{\ell^2}. \tag{SM2.74}$$

Then

$$R_{\mathcal{D}}(h_n)\le\frac{2}{\sqrt{n}}\|f'\|_\infty+\frac{4}{\sqrt{n}}\|f'\|_\gamma\|\gamma\|_{\ell^2}+\left(4\|f'\|_\infty+8\|f'\|_\gamma\|\gamma\|_{\ell^2}\right)\sqrt{\frac{2\log(4/\delta)}{n}}. \tag{SM2.75}$$ ■

REFERENCES

[1] P. L. Bartlett and S. Mendelson, *Rademacher and gaussian complexities: Risk bounds and structural results*, Journal of Machine Learning Research, 3 (2002), pp. 463–482.

[2] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*, Cambridge university press, 2014.