# Supplemental Materials:
# Quantum-data-driven dynamical transition in quantum learning

Bingzhi Zhang,[1,2] Junyu Liu,[3,4,5,6] Liang Jiang,[3] and Quntao Zhuang[2,1,*]

[1] *Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA*
[2] *Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089, USA*
[3] *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA*
[4] *Department of Computer Science, The University of Chicago, Chicago, IL 60637, USA*
[5] *Kadanoff Center for Theoretical Physics, The University of Chicago, Chicago, IL 60637, USA*
[6] *Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260, USA*

(Dated: May 10, 2025)

The supplemental materials contain details of derivations and additional figures.

## CONTENTS



* qzhuang@usc.edu

## Supplementary Note 1. PROOF OF LEMMA 2

In this section, we provide the proof of Lemma 2. **Proof.** Recall $f_{\alpha\beta}(t)$ defined in Eq. (13), for convenience we also define

$$g_\gamma(t) = \sqrt{K_{\gamma\gamma}(t)}, \tag{1}$$

such that $f_{\alpha\beta}(t) = \sum_\gamma g_\gamma(t)\epsilon_\gamma(t)\lambda_{\gamma\alpha\beta}$.

We can derive the time-derivative of $\angle_{\alpha\beta}$ as follows.

$$\partial_t \angle_{\alpha\beta} = \frac{(\partial_t K_{\alpha\beta})\sqrt{K_{\alpha\alpha}K_{\beta\beta}} - K_{\alpha\beta}\partial_t\sqrt{K_{\alpha\alpha}K_{\beta\beta}}}{K_{\alpha\alpha}K_{\beta\beta}} \tag{2}$$

$$= -\frac{\eta}{N}\frac{\sum_\gamma \epsilon_\gamma\sqrt{K_{\gamma\gamma}}\left(\lambda_{\gamma\beta\alpha}\sqrt{K_{\alpha\alpha}} + \lambda_{\gamma\alpha\beta}\sqrt{K_{\beta\beta}}\right)}{\sqrt{K_{\alpha\alpha}K_{\beta\beta}}} - \frac{K_{\alpha\beta}}{K_{\alpha\alpha}K_{\beta\beta}}\frac{(\partial_t K_{\alpha\alpha})K_{\beta\beta} + K_{\alpha\alpha}\partial_t K_{\beta\beta}}{2\sqrt{K_{\alpha\alpha}K_{\beta\beta}}} \tag{3}$$

$$= -\frac{\eta}{N}\frac{\sum_\gamma \epsilon_\gamma\sqrt{K_{\gamma\gamma}}\left(\lambda_{\gamma\beta\alpha}\sqrt{K_{\alpha\alpha}} + \lambda_{\gamma\alpha\beta}\sqrt{K_{\beta\beta}}\right)}{\sqrt{K_{\alpha\alpha}K_{\beta\beta}}} + \frac{2\eta}{N}\frac{K_{\alpha\beta}}{K_{\alpha\alpha}K_{\beta\beta}}\left(\frac{\sum_\gamma \epsilon_\gamma\sqrt{K_{\gamma\gamma}}\lambda_{\gamma\alpha\alpha}\sqrt{K_{\alpha\alpha}}K_{\beta\beta}}{2\sqrt{K_{\alpha\alpha}K_{\beta\beta}}} + \frac{K_{\alpha\alpha}\sum_\gamma \epsilon_\gamma\sqrt{K_{\gamma\gamma}}\lambda_{\gamma\beta\beta}\sqrt{K_{\beta\beta}}}{2\sqrt{K_{\alpha\alpha}K_{\beta\beta}}}\right) \tag{4}$$

$$= -\frac{\eta}{N}\sum_\gamma \frac{\epsilon_\gamma\sqrt{K_{\gamma\gamma}}}{\sqrt{K_{\alpha\alpha}K_{\beta\beta}}}\left[\lambda_{\gamma\beta\alpha}\sqrt{K_{\alpha\alpha}} + \lambda_{\gamma\alpha\beta}\sqrt{K_{\beta\beta}} - K_{\alpha\beta}\left(\frac{\lambda_{\gamma\alpha\alpha}}{\sqrt{K_{\alpha\alpha}}} + \frac{\lambda_{\gamma\beta\beta}}{\sqrt{K_{\beta\beta}}}\right)\right] \tag{5}$$

$$= -\frac{\eta}{N}\sum_\gamma \epsilon_\gamma\sqrt{K_{\gamma\gamma}}\left[\frac{\lambda_{\gamma\beta\alpha} - \angle_{\alpha\beta}\lambda_{\gamma\beta\beta}}{\sqrt{K_{\beta\beta}}} + \frac{\lambda_{\gamma\alpha\beta} - \angle_{\alpha\beta}\lambda_{\gamma\alpha\alpha}}{\sqrt{K_{\alpha\alpha}}}\right] \tag{6}$$

$$= -\frac{\eta}{N}\sum_\gamma \epsilon_\gamma g_\gamma\left[\left(\frac{\lambda_{\gamma\beta\alpha}}{g_\beta} + \frac{\lambda_{\gamma\alpha\beta}}{g_\alpha}\right) - \left(\frac{\lambda_{\gamma\beta\beta}}{g_\beta} + \frac{\lambda_{\gamma\alpha\alpha}}{g_\alpha}\right)\angle_{\alpha\beta}\right]. \tag{7}$$

Then Eq. (7) can be simplified as

$$\mathrm{d}_t\angle_{\alpha\beta}(t) = -\frac{\eta}{N}\left[\left(\frac{f_{\beta\alpha}(t)}{g_\beta(t)} + \frac{f_{\alpha\beta}(t)}{g_\alpha(t)}\right) - \left(\frac{f_{\beta\beta}(t)}{g_\beta(t)} + \frac{f_{\alpha\alpha}(t)}{g_\alpha(t)}\right)\angle_{\alpha\beta}(t)\right]. \tag{8}$$

Suppose

$$\mathcal{A}_{\alpha\beta} \equiv \lim_{t\to\infty}\frac{\left(\frac{f_{\beta\alpha}(t)}{g_\beta(t)} + \frac{f_{\alpha\beta}(t)}{g_\alpha(t)}\right)}{\left(\frac{f_{\beta\beta}(t)}{g_\beta(t)} + \frac{f_{\alpha\alpha}(t)}{g_\alpha(t)}\right)} = \text{const}, \tag{9}$$

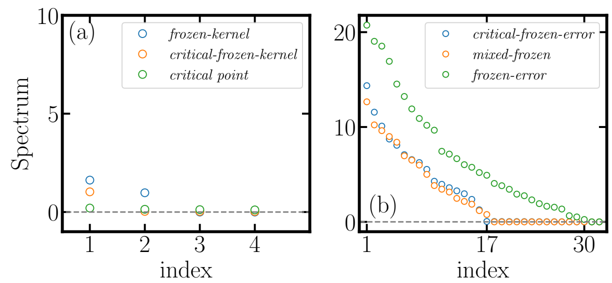
is a non-zero constant in $[-1, 1]$, at late time Eq. (8) can be simplified as

$$\mathrm{d}_t\angle_{\alpha\beta}(t) = -\frac{\eta}{N}\left(\frac{f_{\beta\beta}(t)}{g_\beta(t)} + \frac{f_{\alpha\alpha}(t)}{g_\alpha(t)}\right)\left[\mathcal{A}_{\alpha\beta} - \angle_{\alpha\beta}(t)\right]. \tag{10}$$

Therefore we obtain the fixed point

$$\angle_{\alpha\beta}(t) = \mathcal{A}_{\alpha\beta}. \tag{11}$$

■




Supplementary Figure 1. Spectrum of Hessian of loss function for different QNN training dynamics with two data. We plot the 4 and 32 largest eigenvalues in (a) and (b) separately. The setting is the same as in Fig. 2.

## Supplementary Note 2. HESSIAN SPECTRUM INTERPRETATION

In this section, we interpret the dynamical transition via the spectrum of Hessian of loss function in Eq. (2). To see this, we begin with the dynamical equation of variational parameters at the stable fixed point $\boldsymbol{\theta}^*$ as

$$\delta\boldsymbol{\theta} \simeq -\eta\mathbf{H}(\boldsymbol{\theta}^*)\left(\boldsymbol{\theta}-\boldsymbol{\theta}^*\right), \tag{12}$$

where $\mathbf{H}(\boldsymbol{\theta}^*)$ is the Hessian matrix of loss function with dimension $L\times L$ defined as

$$\mathbf{H}_{\ell_1\ell_2}(\boldsymbol{\theta}) = \frac{\partial^2\mathcal{L}}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}} = \sum_{\beta\in\Omega}\left(\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_2}} + \epsilon_\beta\frac{\partial^2\epsilon_\beta}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\right) \tag{13}$$

$$= \sum_{\beta\in S_E\setminus(S_E\cap S_K)}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_2}} + \sum_{\beta\in S_K\setminus(S_E\cap S_K)}\epsilon_\beta\frac{\partial^2\epsilon_\beta}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}. \tag{14}$$

In the above, the first equation comes from definition and the second equation adopts the definition of $S_E$ and $S_K$. We can regard Eq. (12) as an imaginary-time Schrödinger equation with $\mathbf{H}(\boldsymbol{\theta}^*)$ as an effective Hamiltonian. Therefore, it is natural to study the spectrum of $\mathbf{H}(\boldsymbol{\theta}^*)$. Clearly, we see the matrices in the first summation are only rank-1, while the others in general have rank much larger than one. For *frozen-kernel dynamics* with $S_E=\Omega$, the Hessian $\mathbf{H}_{\ell_1\ell_2}=\sum_{\beta\in S_E}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_2}}$ becomes sums of rank-1 matrices, resulting in a rank-$N$ matrix given an orthogonal input data set. Furthermore, one can see that the trace of Hessian is simply the trace of QNTK matrix $\mathrm{tr}(\mathbf{H}_{\ell_1\ell_2})=\sum_\beta K_{\beta\beta}$. When part of the data are targeted at the boundary leading to the *critical-frozen-kernel dynamics* with $S_K\subsetneq S_E=\Omega$, the rank of the Hamiltonian directly decreases to $N-|S_K|$. Specifically, at *critical point* with all data targeted at the boundary, all eigenvalues in the spectrum vanish at the fixed point. The above results are verified in Fig. 1(a). On the other hand, when there are data targeted beyond the accessible region, the Hessian of total error $\frac{\partial^2\epsilon_\beta}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}$ would significantly increase the number of positive eigenvalues in the spectrum. In fact, through numerical simulation (see Fig. 1(b)) we find that the number of positive eigenvalues in *mixed-frozen dynamics* is just $|S_E\setminus(S_E\cap S_K)|$ more than that for *critical-frozen-error dynamics*, and the *frozen-error dynamics* has many more positive eigenvalues compared to the others. Meanwhile, how the spectrum behaves with more data involved ~~still~~ remains unexplored as the rank may saturate to the number of parameters $L$. We leave that as an open question in future research.

## Supplementary Note 3. GAUGE INVARIANCE IN TRAINING DYNAMICS

In this section, we study the training dynamics under basis transformation. We begin with the MSE loss $\mathcal{L}=\frac{1}{2N}\sum_\alpha\epsilon_\alpha^2$. The inner product enables us to introduce an orthogonal matrix $S\in O(N)$, independent of both $\boldsymbol{\theta}$ and $t$, to transform the total error vector to

$$\epsilon_\alpha(\boldsymbol{\theta}) \to \sum_{\alpha'} S_{\alpha\alpha'}\epsilon_{\alpha'}(\boldsymbol{\theta}) \equiv \tilde{\epsilon}_\alpha(\boldsymbol{\theta}). \tag{15}$$

A direct result is that the MSE loss function is gauge invariant as

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}) = \frac{1}{2N}\sum_{\alpha}\tilde{\epsilon}^2_{\alpha}(\boldsymbol{\theta}) = \frac{1}{2N}\sum_{\alpha}\sum_{\alpha_1,\alpha_2} S_{\alpha\alpha_1}\epsilon_{\alpha_1}(\boldsymbol{\theta})S_{\alpha\alpha_2}\epsilon_{\alpha_2}(\boldsymbol{\theta}) \tag{16}$$

$$= \frac{1}{2N}\sum_{\alpha}\epsilon^2_{\alpha}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}), \tag{17}$$

where in the second line we apply $\sum_\alpha S_{\alpha\alpha_1}S_{\alpha\alpha_2} = \delta_{\alpha_1\alpha_2}$. Thus we can identify the orthogonal group as a global gauge symmetry since it is independent of $t$ and $\boldsymbol{\theta}$ as we state above. The gauge invariance can be concluded from its inner product structure. Following the definitions of QNTK and dQNTK in Eqs. (7), (10) of the main text, they are transformed as

$$K_{\alpha\beta}(\boldsymbol{\theta}) \to \sum_{\alpha',\beta'} S_{\alpha\alpha'}K_{\alpha'\beta'}(\boldsymbol{\theta})S_{\beta\beta'} \equiv \tilde{K}_{\alpha\beta}(\boldsymbol{\theta}), \tag{18}$$

$$\mu_{\gamma\alpha\beta}(\boldsymbol{\theta}) \to \sum_{\gamma',\alpha',\beta'} S_{\gamma\gamma'}S_{\alpha\alpha'}\mu_{\gamma'\alpha'\beta'}(\boldsymbol{\theta})S_{\beta\beta'} \equiv \tilde{\mu}_{\gamma\alpha\beta}(\boldsymbol{\theta}). \tag{19}$$

One can directly see that the QNTK and dQNTK do not exhibit the gauge invariance due to their outer product structure. However, one can easily check that $\mathrm{tr}(K) = \sum_\alpha K_{\alpha\alpha}$ is gauge invariant under the transformation.

For the dynamics, we begin with the gradient descent rule.

$$\delta\theta_\ell(t) = -\frac{\eta}{N}\sum_{\alpha}\tilde{\epsilon}_{\alpha}(\boldsymbol{\theta})\frac{\partial\tilde{\epsilon}_{\alpha}(\boldsymbol{\theta})}{\partial\theta_\ell} \tag{20}$$

$$= -\frac{\eta}{N}\sum_{\alpha}\sum_{\alpha_1,\alpha_2} S_{\alpha\alpha_1}\epsilon_{\alpha_1}(\boldsymbol{\theta})S_{\alpha\alpha_2}\frac{\partial\epsilon_{\alpha_2}(\boldsymbol{\theta})}{\partial\theta_\ell} \tag{21}$$

$$= -\frac{\eta}{N}\sum_{\alpha}\epsilon_{\alpha}(\boldsymbol{\theta})\frac{\partial\epsilon_{\alpha}(\boldsymbol{\theta})}{\partial\theta_\ell}, \tag{22}$$

which also preserves the gauge invariance.

For the first dynamical equation in Eqs. (9) of the main text, we have

$$\delta\tilde{\epsilon}_\alpha(t) + \frac{\eta}{N}\sum_{\beta}\tilde{K}_{\alpha\beta}(t)\tilde{\epsilon}_\beta(t) \tag{23}$$

$$= \sum_{\alpha'} S_{\alpha\alpha'}\delta\epsilon_{\alpha'}(t) + \frac{\eta}{N}\sum_{\beta,\alpha',\beta_1,\beta_2} S_{\alpha\alpha'}K_{\alpha'\beta_1}(t)S_{\beta\beta_1}S_{\beta\beta_2}\epsilon_{\beta_2}(t) \tag{24}$$

$$= \sum_{\alpha'} S_{\alpha\alpha'}\left(\delta\epsilon_{\alpha'}(t) + \frac{\eta}{N}\sum_{\beta_1} K_{\alpha'\beta_1}(t)\epsilon_{\beta_1}(t)\right) = 0. \tag{25}$$

Similarly, for the second one, we have

$$\delta\tilde{K}_{\alpha\beta}(t) + \frac{\eta}{N}\sum_{\gamma}\tilde{\epsilon}_\gamma(t)\left[\tilde{\mu}_{\gamma\alpha\beta}(t) + \tilde{\mu}_{\gamma\beta\alpha}(t)\right] \tag{26}$$

$$= \sum_{\alpha',\beta'} S_{\alpha\alpha'}\delta K_{\alpha'\beta'}(t)S_{\beta\beta'} + \frac{\eta}{N}\sum_{\substack{\gamma,\gamma_1,\gamma_2,\\ \alpha',\beta'}} S_{\gamma\gamma_1}\epsilon_{\gamma_1}(t)\left[S_{\gamma\gamma_2}S_{\alpha\alpha'}\mu_{\gamma_2\alpha'\beta'}(t)S_{\beta\beta'} + S_{\gamma\gamma_2}S_{\beta\beta'}\mu_{\gamma_2\beta'\alpha'}(t)S_{\alpha\alpha'}\right] \tag{27}$$

$$= \sum_{\alpha',\beta'} S_{\alpha\alpha'}\left[\delta K_{\alpha'\beta'}(t) + \frac{\eta}{N}\sum_{\gamma_1}\epsilon_{\gamma_1}\left(\mu_{\gamma_1\alpha'\beta'}(t) + \mu_{\gamma_1\beta'\alpha'}(t)\right)\right]S_{\beta\beta'} \tag{28}$$

$$= 0. \tag{29}$$

Therefore we can conclude that the dynamical equations in Eqs. (9) of the main text are gauge invariant under basis transformation from orthogonal group $O(N)$, which also suggests that $\tilde{\epsilon}_\alpha(t)\tilde{K}_{\alpha\alpha}(t) = 0, \forall\alpha$ are fixed points.

## Supplementary Note 4. DETAILED SOLUTIONS FOR THE CONVERGENCE DYNAMICS

In this section, we present the details on deriving the convergence solution perturbatively around the stable fixed point. For convenience, we re-print the dynamical equations of (17) in the main text here

$$\begin{cases} \partial_t \epsilon_\alpha(t) = -\frac{\eta}{N}\sum_\beta \angle_{\alpha\beta} g_\alpha(t) g_\beta(t) \epsilon_\beta(t); \\ \partial_t g_\alpha(t) = -\frac{\eta}{N}\sum_\beta \lambda_{\alpha\alpha\beta} g_\beta(t)\epsilon_\beta(t), \end{cases} \tag{30}$$

where we define $g_\alpha(t) \equiv \sqrt{K_{\alpha\alpha}(t)}$ to simplify the notation.

### Supplementary Note 4.1. Exponential convergence class

In this part, we study the exponential convergence class where $S_E \cap S_K = \emptyset$. The main idea for perturbatively solving the convergence dynamics towards a fixed point is to first focus on those quantities converging towards zero, and then apply the obtained solutions back to equations of the other equations.

#### *Supplementary Note 4.1.1. frozen-kernel dynamics*

For *frozen-kernel dynamics*, the fixed point is $\{(\epsilon_\alpha(\infty) = 0, K_{\alpha\alpha}(\infty) > 0)\}_{\alpha\in\Omega}$. The leading order of the first PDE in Eqs. (30) becomes

$$\partial_t \epsilon_\alpha(t) = -\frac{\eta}{N}\sum_{\beta\in\Omega} g_\alpha(\infty)\angle_{\alpha\beta} g_\beta(\infty)\epsilon_\beta(t) \tag{31}$$

$$= -\frac{\eta}{N}\sum_{\beta\in\Omega} K_{\alpha\beta}(\infty)\epsilon_\beta(t). \tag{32}$$

As $K_{\alpha\beta}(\infty)$ is symmetric and positive definite, we can diagonalize it as $K_{\alpha\beta} = \sum_{\alpha',\beta'} P_{\alpha\alpha'}\Lambda_{\alpha'\beta'}P^T_{\beta'\beta}$, where $\Lambda_{\alpha'\beta'}$ is a diagonal matrix consisting of eigenvalues $\{w_\alpha\}_{\alpha=1}^N$ of $K_{\alpha\beta}$. Thus, we can solve $\epsilon_\alpha$ as

$$\epsilon_\alpha(t) = \sum_{\beta\in\Omega} b_\beta P_{\alpha\beta} e^{-\eta w_\beta t/N}, \tag{33}$$

where $b_\beta$ are fitting parameters.

Plugging it into the second PDE in Eqs. (30), we have

$$\partial_t g_\alpha(t) = -\frac{\eta}{N}\sum_{\beta\in\Omega}\lambda_{\alpha\alpha\beta} g_\beta(\infty)\sum_{\gamma\in\Omega} b_\gamma P_{\beta\gamma} e^{-\eta w_\gamma t/N} \tag{34}$$

$$= -\frac{\eta}{N}\sum_{\gamma\in\Omega}\left(\sum_{\beta\in\Omega}\lambda_{\alpha\alpha\beta} g_\beta(\infty) P_{\beta\gamma}\right) b_\gamma e^{-\eta w_\gamma t/N}, \tag{35}$$

which can be solved as

$$g_\alpha(t) = g_\alpha(\infty) + \sum_{\gamma\in\Omega}\frac{b_\gamma}{w_\gamma}\left(\sum_{\beta\in\Omega}\lambda_{\alpha\alpha\beta} g_\beta(\infty) P_{\beta\gamma}\right) e^{-\eta v_\gamma t/N}, \tag{36}$$

In the asymptotic limit of $t \gg 1$, we can only keep track on the exponent with the smallest eigenvalue $w^* = \min\{w_\beta\}$, which determines the leading-order behavior, resulting in simpler solutions as

$$\begin{cases} \epsilon_\alpha(t) = b_{\gamma^*} P_{\alpha\gamma^*} e^{-\eta w^* t/N}; \\ g_\alpha(t) = g_\alpha(\infty) + \left(\sum_{\beta\in\Omega}\lambda_{\alpha\alpha\beta} g_\beta(\infty) P_{\beta\gamma^*}\right)\frac{b_{\gamma^*}}{w_{\gamma^*}} e^{-\eta w^* t/N}, \end{cases} \tag{37}$$

where $\gamma^* = \operatorname{argmin}_\beta w_\gamma$.

*Supplementary Note 4.1.2. frozen-error dynamics*

Inversely, for the *frozen-error dynamics*, the fixed point is $\{(\epsilon_\alpha(\infty) \neq 0, K_{\alpha\alpha}(\infty) = 0)\}_{\alpha\in\Omega}$, the second PDE in Eqs. (30) is reduced to

$$\partial_t g_\alpha(t) = -\frac{\eta}{N}\sum_{\beta\in\Omega}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty)g_\beta(t) = -\frac{\eta}{N}\sum_{\beta}F_{\alpha\beta}g_\beta(t), \tag{38}$$

where we define $F_{\alpha\beta} \equiv A_{\alpha\beta}\epsilon_\beta(\infty)$. Although $F_{\alpha\beta}$ is not symmetric in general, we can still perform diagonalization to obtain $F_{\alpha\beta} = \sum_{\alpha',\beta'} = P_{\alpha\alpha'}\Lambda_{\alpha'\beta'}P^{-1}_{\beta'\beta}$, where $\Lambda_{\alpha'\beta'} = w_{\alpha'\alpha'}\delta_{\alpha'\beta'}$ is the diagonal matrix of eigenvalues. Then $g_\alpha(t)$ can be solved as

$$g_\alpha(t) = \sum_{\beta\in\Omega} b_\beta P_{\alpha\beta}e^{-\eta w_\beta t/N}, \tag{39}$$

where $b_\beta$ are also free fitting parameters. One can then solve the dynamics of $\epsilon_\alpha(t)$ as

$$\partial_t\epsilon_\alpha(t) = -\frac{\eta}{N}\sum_{\beta\in\Omega}\angle_{\alpha\beta}\sum_{\gamma\in\Omega}b_\gamma P_{\alpha\gamma}e^{-\eta w_\gamma t/N}\sum_{\gamma'\in\Omega}b_{\gamma'}P_{\beta\gamma'}e^{-\eta w_{\gamma'}t/N}\epsilon_\beta(\infty) \tag{40}$$

$$= -\frac{\eta}{N}\sum_{\gamma,\gamma'}\left(\sum_{\beta}\angle_{\alpha\beta}\epsilon_\beta(\infty)P_{\beta\gamma'}\right)P_{\alpha\gamma}b_\gamma b_{\gamma'}e^{-\eta\left(w_\gamma+w_{\gamma'}\right)t/N}, \tag{41}$$

which leads to the solution as

$$\epsilon_\alpha(t) = \epsilon_\alpha(\infty) + \sum_{\gamma,\gamma'\in\Omega}\left(\sum_{\beta}\angle_{\alpha\beta}\epsilon_\beta(\infty)P_{\beta\gamma'}\right)\frac{P_{\alpha\gamma}b_\gamma b_{\gamma'}}{(w_\gamma+w_{\gamma'})}e^{-\eta(w_\gamma+w_{\gamma'})t/N}. \tag{42}$$

In the asymptotic limit, the leading-order solution is

$$\begin{cases} \epsilon_\alpha(t) = \epsilon_\alpha(\infty) + \left(\sum_\beta \angle_{\alpha\beta}\epsilon_\beta(\infty)P_{\beta\gamma^*}\right)\frac{P_{\alpha\gamma^*}b^2_{\gamma^*}}{2w^*}e^{-2\eta w^* t/N}; \\ g_\alpha(t) = b_{\gamma^*}P_{\alpha\gamma^*}e^{-\eta w^* t/N}, \end{cases} \tag{43}$$

where $\gamma^* = \operatorname{argmin}_\gamma w_\gamma$ and $w^* = w_{\gamma^*}$.

*Supplementary Note 4.1.3. mixed-frozen dynamics*

For the *mixed-frozen dynamics*, the fixed point is $\{(\epsilon_\alpha(\infty) = 0, K_{\alpha\alpha}(\infty) > 0)\}_{\alpha\in S_E} \cup \{(\epsilon_\alpha(\infty) \neq 0, K_{\alpha\alpha}(\infty) = 0)\}_{\alpha\in S_K}$. We first study the PDEs of $\{\epsilon_\alpha(t), \forall\alpha\in S_E\}$ and $\{g_\alpha(t), \forall\alpha\in S_K\}$, which can be reduced from Eqs. (30) as

$$\begin{cases} \partial_t\epsilon_\alpha(t) &= -\frac{\eta}{N}\left(\sum_{\beta\in S_E} g_\alpha(\infty)\angle_{\alpha\beta}g_\beta(\infty)\epsilon_\beta(t) + \sum_{\beta\in S_K} g_\alpha(\infty)\angle_{\alpha\beta}\epsilon_\beta(\infty)g_\beta(t)\right), \forall\alpha\in S_E; \\ \partial_t g_\alpha(t) &= -\frac{\eta}{N}\left(\sum_{\beta\in S_E}\lambda_{\alpha\alpha\beta}g_\beta(\infty)\epsilon_\beta(t) + \sum_{\beta\in S_K}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty)g_\beta(t)\right), \forall\alpha\in S_K. \end{cases} \tag{44}$$

Observing that the above linear PDEs can be reformed in a matrix form as

$$\partial_t\begin{pmatrix}[\epsilon_\alpha(t)]_{\alpha\in S_E}\\ [g_\alpha(t)]_{\alpha\in S_K}\end{pmatrix} = -\frac{\eta}{N}\begin{pmatrix}[K_{\alpha\beta}(\infty)]_{\alpha,\beta\in S_E} & [g_\alpha(\infty)\angle_{\alpha\beta}\epsilon_\beta(\infty)]_{\alpha\in S_E,\beta\in S_K}\\ [\lambda_{\alpha\alpha\beta}g_\beta(\infty)]_{\alpha\in S_K,\beta\in S_E} & [\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty)]_{\alpha,\beta\in S_K}\end{pmatrix}\begin{pmatrix}[\epsilon_\beta(t)]_{\beta\in S_E}\\ [g_\beta(t)]_{\beta\in S_K}\end{pmatrix}, \tag{45}$$

where '$[\cdot]_{\{\dots\}}$' indicate the vector or matrix form with indices constraints. Through the eigen-decomposition of the above matrix $P_{\alpha\alpha'}\Lambda_{\alpha'\beta'}P^{-1}_{\beta'\beta}$ with eigen-matrix $\Lambda_{\alpha'\beta'} = \mathrm{Diag}\{w_1,\cdots,w_N\}$, we obtain

$$\begin{cases} \epsilon_\alpha(t) = \sum_{\beta\in\Omega} b_\beta P_{\alpha\beta}e^{-\eta w_\beta t/N}, \forall\alpha\in S_E; \\ g_\alpha(t) = \sum_{\beta\in\Omega} b_\beta P_{\alpha\beta}e^{-\eta w_\beta t/N}, \forall\alpha\in S_K, \end{cases} \tag{46}$$

where $\{b_\beta\}_{\beta\in\Omega}$ are free fitting parameters. The PDE for $\{\epsilon_\alpha(t), \forall\alpha\in S_K\}$ becomes

$$\partial_t\epsilon_\alpha(t) = -\frac{\eta}{N}\sum_{\alpha'\in\Omega} b_{\alpha'}P_{\alpha\alpha'}e^{-\eta w_{\alpha'}t/N}\left(\sum_{\beta\in S_E}\angle_{\alpha\beta}g_\beta(\infty)\sum_{\beta'\in\Omega}b_{\beta'}P_{\beta\beta'}e^{-\eta w_{\beta'}t/N} + \sum_{\beta\in S_K}\angle_{\alpha\beta}\epsilon_\beta(\infty)\sum_{\beta'\in\Omega}b_{\beta'}P_{\beta\beta'}e^{-\eta w_{\beta'}t/N}\right) \tag{47}$$

$$= -\frac{\eta}{N}\sum_{\alpha',\beta'\in\Omega}\left(\sum_{\beta\in S_E}\angle_{\alpha\beta}g_\beta(\infty)P_{\beta\beta'} + \sum_{\beta\in S_K}\angle_{\alpha\beta}\epsilon_\beta(\infty)P_{\beta\beta'}\right)b_{\alpha'}b_{\beta'}P_{\alpha\alpha'}e^{-\eta\left(w_{\alpha'}+w_{\beta'}\right)t/N}, \tag{48}$$

leading to the solution

$$\epsilon_\alpha(t) = \epsilon_\alpha(\infty) + \sum_{\alpha',\beta'\in\Omega}\left(\sum_{\beta\in S_E}\angle_{\alpha\beta}g_\beta(\infty)P_{\beta\beta'} + \sum_{\beta\in S_K}\angle_{\alpha\beta}\epsilon_\beta(\infty)P_{\beta\beta'}\right)\frac{b_{\alpha'}b_{\beta'}P_{\alpha\alpha'}}{w_{\alpha'}+w_{\beta'}}e^{-\eta\left(w_{\alpha'}+w_{\beta'}\right)t/N}, \forall\alpha\in S_K. \tag{49}$$

Similarly, for $\{g_{\alpha\alpha}(t), \forall\alpha\in S_E\}$, we have

$$\partial_t g_\alpha(t) = -\frac{\eta}{N}\left(\sum_{\beta\in S_E}\lambda_{\alpha\alpha\beta}g_\beta(\infty)\sum_{\beta'\in\Omega}b_{\beta'}P_{\beta\beta'}e^{-\eta w_{\beta'}t/N} + \sum_{\beta\in S_K}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty)\sum_{\beta'\in\Omega}b_{\beta'}P_{\beta\beta'}e^{-\eta w_{\beta'}t/N}\right) \tag{50}$$

$$= -\frac{\eta}{N}\sum_{\beta'\in\Omega}\left(\sum_{\beta\in S_E}\lambda_{\alpha\alpha\beta}g_\beta(\infty)P_{\beta\beta'} + \sum_{\beta\in S_K}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty)P_{\beta\beta'}\right)b_{\beta'}e^{-\eta w_{\beta'}t/N}, \tag{51}$$

resulting in the solution

$$g_\alpha(t) = g_\alpha(\infty) + \sum_{\beta'\in\Omega}\left(\sum_{\beta\in S_E}\lambda_{\alpha\alpha\beta}g_\beta(\infty)P_{\beta\beta'} + \sum_{\beta\in S_K}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty)P_{\beta\beta'}\right)\frac{b_{\beta'}}{w_{\beta'}}e^{-\eta w_{\beta'}t/N}, \forall\alpha\in S_E. \tag{52}$$

In the asymptotic limit $t \gg 1$, we have the leading-order solution as

$$\begin{cases} \epsilon_\alpha(t) = b_{\gamma^*}P_{\alpha\gamma^*}e^{-\eta w^* t/N}, \forall\alpha\in S_E; \\ \epsilon_\alpha(t) = \epsilon_\alpha(\infty) + \left(\sum_{\beta\in S_E}\angle_{\alpha\beta}g_\beta(\infty)P_{\beta\gamma^*} + \sum_{\beta\in S_K}\angle_{\alpha\beta}\epsilon_\beta(\infty)P_{\beta\gamma^*}\right)\frac{b_{\gamma^*}^2 P_{\alpha\gamma^*}}{2w_{\gamma^*}}e^{-2\eta w^* t/N}, \forall\alpha\in S_K; \\ g_\alpha(t) = g_\alpha(\infty) + \left(\sum_{\beta\in S_E}\lambda_{\alpha\alpha\beta}g_\beta(\infty)P_{\beta\gamma^*} + \sum_{\beta\in S_K}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty)P_{\beta\gamma^*}\right)\frac{b_{\gamma^*}}{w_{\gamma^*}}e^{-\eta w^* t/N}, \forall\alpha\in S_E; \\ g_\alpha(t) = b_{\gamma^*}P_{\alpha\gamma^*}e^{-\eta w^* t/N}, \forall\alpha\in S_K, \end{cases} \tag{53}$$

where $\gamma^* = \operatorname{argmin}_\gamma w_\gamma$ and $w^* = w_{\gamma^*}$.

### Supplementary Note 4.2. Polynomial convergence class

In this section, we consider $S_E \cap S_K \neq \emptyset$, which corresponds to the polynomial convergence class.

#### *Supplementary Note 4.2.1. Critical point*

When $S_E = S_K = \Omega$, it corresponds to the *critical point* with the fixed point $\{(\epsilon_\alpha(\infty) = 0, K_{\alpha\alpha}(\infty) = 0)\}_{\alpha\in\Omega}$. The PDEs for error and kernel are the same as in Eqs. (30), and to solve it, we take an ansatz solution

$$\begin{cases} \epsilon_\alpha(t) = c_\alpha^E/(c_0 + \eta t/N); \\ g_\alpha(t) = c_\alpha^G/\sqrt{c_0 + \eta t/N}, \end{cases} \tag{54}$$

with fitting parameters $\{c_\alpha^E, c_\alpha^G\}$.

*Supplementary Note 4.2.2. Critical-frozen-kernel dynamics*

When $S_K \subsetneq S_E = \Omega$, we have the fixed points $\{(\epsilon_\alpha(\infty) = 0, K_{\alpha\alpha}(\infty) = 0)\}_{\alpha \in S_K} \cup \{(\epsilon_\alpha(\infty) = 0, K_{\alpha\alpha}(\infty) > 0)\}_{\alpha \in S_E \setminus S_K}$. Initially, the interaction between different data is negligible, and we can expect that data from $S_K$ follows the dynamics of *critical point* while the one from $S_E \setminus S_K$ follows the dynamics of *frozen-kernel dynamics*, which suggests that the convergence of $\epsilon_\beta(t) g_\beta(t)$ from $S_K$, governed by Eqs. (30), is much faster compared to $S_E \setminus S_K$. Therefore, for the dynamics of error and kernel from $S_K$, we treat them as self-governed in a "free-field" theory as

$$\begin{cases} \partial_t \epsilon_\alpha(t) &= -\frac{\eta}{N} \sum_{\beta \in S_K} g_\alpha(t) \angle_{\alpha\beta} g_\beta(t) \epsilon_\beta(t), \forall \alpha \in S_K; \\ \partial_t g_\alpha(t) &= -\frac{\eta}{N} \sum_{\beta \in S_K} \lambda_{\alpha\alpha\beta} g_\beta(t) \epsilon_\beta(t), \forall \alpha \in S_K. \end{cases} \tag{55}$$

The solution of these "free-field" part can be described by Eqs. (54).

Plugging in the polynomial solutions, the PDE for error from $\alpha \in S_E \setminus S_K$ is

$$\partial_t \epsilon_\alpha(t) = -\frac{\eta}{N} g_\alpha(\infty) \left( \sum_{\beta \in S_E \setminus S_K} \angle_{\alpha\beta} g_\beta(\infty) \epsilon_\beta(t) + \sum_{\beta \in S_K} \frac{\angle_{\alpha\beta} c_\beta^E c_\beta^G}{(c_0 + \eta t/N)^{3/2}} \right). \tag{56}$$

At late time, when $\{g_\beta(\infty)\epsilon_\beta(t), \forall \beta \in S_E \setminus S_K\}$ is comparable to $(c_0 + \eta t/N)^{-3/2}$, the interactions cannot be neglected, and thus we take

$$\epsilon_\alpha(t) = \frac{b_\alpha}{(c_0 + \eta t/N)^{3/2}}, \forall \alpha \in S_E \setminus S_K, \tag{57}$$

with fitting parameters $b_\alpha$. Then $g_\alpha(t), \alpha \in S_E \setminus S_K$ can be obtained from

$$\partial_t g_\alpha(t) = -\frac{\eta}{N} \left( \sum_{\beta \in S_E \setminus S_K} \frac{\lambda_{\alpha\alpha\beta} g_\beta(\infty) b_\alpha}{(c_0 + \eta t/N)^{3/2}} + \sum_{\beta \in S_K} \frac{\lambda_{\alpha\alpha\beta} c_\beta^E c_\beta^G}{(c_0 + \eta t/N)^{3/2}} \right), \tag{58}$$

leading to

$$g_\alpha(t) = \left( \sum_{\beta \in S_E \setminus S_K} \lambda_{\alpha\alpha\beta} g_\beta(\infty) b_\alpha + \sum_{\beta \in S_K} \lambda_{\alpha\alpha\beta} c_\beta^E c_\beta^G \right) \frac{2}{\sqrt{c_0 + \eta t/N}} + g_\alpha(\infty), \forall \alpha \in S_E \setminus S_K. \tag{59}$$

To summarize, we have

$$\begin{cases} \epsilon_\alpha(t) = c_\alpha^E/(c_0 + \eta t/N), \forall \alpha \in S_K; \\ \epsilon_\alpha(t) = b_\alpha/(c_0 + \eta t/N)^{3/2}, \forall \alpha \in S_E \setminus S_K; \\ g_\alpha(t) = c_\alpha^G/\sqrt{c_0 + \eta t/N}, \forall \alpha \in S_K; \\ g_\alpha(t) = 2\left( \sum_{\beta \in S_E \setminus S_K} \lambda_{\alpha\alpha\beta} g_\beta(\infty) b_\alpha + \sum_{\beta \in S_K} \lambda_{\alpha\alpha\beta} c_\beta^E c_\beta^G \right)/\sqrt{c_0 + \eta t/N} + g_\alpha(\infty), \forall \alpha \in S_E \setminus S_K. \end{cases} \tag{60}$$

*Supplementary Note 4.2.3. Critical-frozen-error dynamics*

When $S_E \subsetneq S_K = \Omega$, the fixed point is described by: $\{(\epsilon_\alpha(\infty) = 0, K_{\alpha\alpha}(\infty) = 0)\}_{\alpha \in S_E} \cup \{(\epsilon_\alpha(\infty) \neq 0, K_{\alpha\alpha}(\infty) = 0)\}_{\alpha \in S_K \setminus S_E}$. Similar to the previous case, we apply the same method to solve the dynamics. For data from $S_E$, it is still described by Eq. (54), and for $g_\alpha, \forall \alpha \in S_K \setminus S_E$, the PDE for $g_\alpha(t)$ becomes

$$\partial_t g_\alpha(t) = -\frac{\eta}{N} \left( \sum_{\beta \in S_E} \frac{\lambda_{\alpha\alpha\beta} c_\beta^E c_\beta^G}{(c_0 + \eta t/N)^{3/2}} + \sum_{\beta \in S_K \setminus S_E} \lambda_{\alpha\alpha\beta} \epsilon_\beta(\infty) g_\beta(t) \right). \tag{61}$$

From the balance of r.h.s., we have

$$g_\alpha(t) = \frac{b_\alpha}{(c_0 + \eta t/N)^{3/2}}, \forall \alpha \in S_K \setminus S_E, \tag{62}$$

with free fitting parameters $b_\alpha$. One can then integrate over $t$ to find the dynamics for $\epsilon_\alpha(t), \forall\alpha \in S_K \setminus (S_E \cap S_K)$. Overall, we have

$$\begin{cases} \epsilon_\alpha(t) = c_\alpha^E/(c_0+\eta t/N), \forall\alpha \in S_E; \\ \epsilon_\alpha(t) = \frac{1}{2}\left[\sum_{\beta\in S_E} \angle_{\alpha\beta} b_\alpha c_\beta^E c_\beta^G + \sum_{\beta\in S_K\setminus S_E} \angle_{\alpha\beta} b_\alpha b_\beta \epsilon_\beta(\infty)\right]/\left(c_0+\eta t/N\right)^2 + \epsilon_\alpha(\infty), \forall\alpha\in S_K\setminus S_E; \\ g_\alpha(t) = c_\alpha^G/\sqrt{c_0+\eta t/N}, \forall\alpha\in S_E; \\ g_\alpha(t) = b_\alpha/(c_0+\eta t/N)^{3/2}, \forall\alpha\in S_K\setminus S_E. \end{cases} \tag{63}$$

### *Supplementary Note 4.2.4. Critical-mixed-frozen dynamics*

Finally, we extend our analyses to the case where the target values lie in all possible regions $\mathbb{R}$. With the same "free-field" approach, the data from $S_E \cap S_K$ can be described by Eq. (54). Then the dynamical equations for $\{\epsilon_\alpha, \forall\alpha\in S_E\setminus(S_E\cap S_K)\}$ and $\{g_\alpha, \forall S_K\setminus(S_E\cap S_K)\}$ become

$$\begin{cases} \partial_t\epsilon_\alpha(t) &= -\frac{\eta}{N} g_\alpha(\infty)\left(\sum_{\beta\in S_E\setminus(S_E\cap S_K)} \angle_{\alpha\beta} g_\beta(\infty)\epsilon_\beta(t) + \sum_{\beta\in S_E\cap S_K} \angle_{\alpha\beta} g_\beta(t)\epsilon_\beta(t) + \sum_{\beta\in S_K\setminus(S_E\cap S_K)} \angle_{\alpha\beta}\epsilon_\beta(\infty) g_\beta(t)\right); \\ \partial_t g_\alpha(t) &= -\frac{\eta}{N}\left(\sum_{\beta\in S_E\setminus(S_E\cap S_K)} \lambda_{\alpha\alpha\beta} g_\beta(\infty)\epsilon_\beta(t) + \sum_{\beta\in S_E\cap S_K}\lambda_{\alpha\alpha\beta} g_\beta(t)\epsilon_\beta(t) + \sum_{\beta\in S_K\setminus(S_E\cap S_K)}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty) g_\beta(\infty)\right). \end{cases} \tag{64}$$

As $\epsilon_\beta(t)g_\beta(t) = c_\beta^E c_\beta^G/(c_0+\eta t/N)^{3/2}, \forall\beta\in S_E\cap S_K$, we here take

$$\epsilon_\alpha(t) = \frac{b_\alpha^E}{(c_0+\eta t/N)^{3/2}}, \forall\alpha\in S_E\setminus(S_E\cap S_K), \tag{65}$$

$$g_\alpha(t) = \frac{b_\alpha^G}{(c_0+\eta t/N)^{3/2}}, \forall\alpha\in S_K\setminus(S_E\cap S_K), \tag{66}$$

with free fitting parameters $b_\alpha^E, b_\alpha^G$. By taking one additional step, one can find the solutions for the other errors and QNTKs. We summarize the solutions for errors and QNTKs as

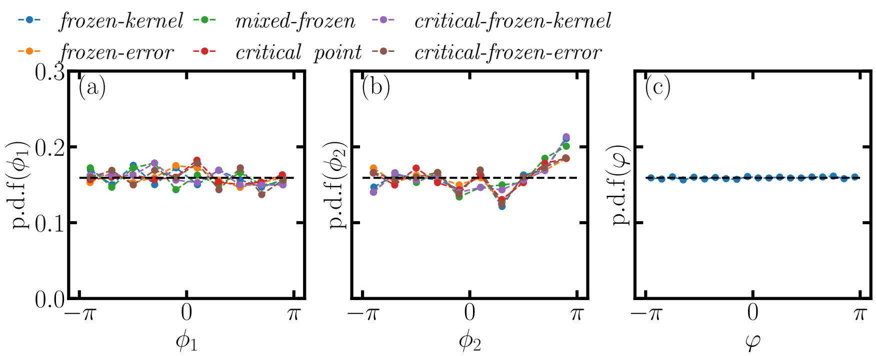
$$\begin{cases} \epsilon_\alpha(t) = c_\alpha^E/(c_0+\eta t/N), \forall\alpha\in S_E\cap S_K; \\ \epsilon_\alpha(t) = b_\alpha^E/(c_0+\eta t/N)^{3/2}, \forall\alpha\in S_E\setminus(S_E\cap S_K); \\ \epsilon_\alpha(t) = \frac{1}{2}\left(\sum_{\beta\in S_E\setminus(S_E\cap S_K)}\angle_{\alpha\beta} g_\beta(\infty) b_\beta^E + \sum_{\beta\in S_E\cap S_K}\angle_{\alpha\beta} c_\beta^E c_\beta^G + \sum_{\beta\in S_K\setminus(S_E\cap S_K)}\angle_{\alpha\beta}\epsilon_\beta(\infty) b_\beta^G\right) b_\alpha^G/\left(c_0+\eta t/N\right)^2 \\ \qquad +\epsilon_\alpha(\infty), \forall\alpha\in S_K\setminus(S_E\cap S_K); \\ g_\alpha(t) = 2\left(\sum_{\beta\in S_E\setminus(S_E\cap S_K)}\lambda_{\alpha\alpha\beta} g_\beta(\infty) b_\beta^E + \sum_{\beta\in S_E\cap S_K}\lambda_{\alpha\alpha\beta} c_\beta^E c_\beta^G + \sum_{\beta\in S_K\setminus(S_E\cap S_K)}\lambda_{\alpha\alpha\beta}\epsilon_\beta(\infty) b_\beta^G\right)/\sqrt{c_0+\eta t/N} \\ \qquad +g_\alpha(\infty), \alpha\in S_E\setminus(S_E\cap S_K); \\ g_\alpha(t) = c_\alpha^G/\sqrt{c_0+\eta t/N}, \forall\alpha\in S_E\cap S_K; \\ g_\alpha(t) = b_\alpha^G/(c_0+\eta t/N)^{3/2}, \forall\alpha\in S_K\setminus(S_E\cap S_K). \end{cases} \tag{67}$$

## Supplementary Note 5. RESTRICTED HAAR RANDOM ENSEMBLE

To provide an insight on the converged unitary in late time, we consider a multi-state preparation task with both input and target states $\{|\psi_\alpha\rangle\}$, $\{|\Phi_\alpha\rangle\}$ forming orthonormal sets, $\langle\psi_\alpha|\psi_\beta\rangle = \langle\Phi_\alpha|\Phi_\beta\rangle = \delta_{\alpha\beta}$. We can then formulate the ensemble of unitary (up to permutation) for the multi-state preparation task as

$$\mathcal{U}_{\rm RH} = \left\{U \middle| U = \begin{pmatrix} Q_N & \mathbf{0} \\ \mathbf{0} & V\end{pmatrix},\ Q_N = \operatorname{diag}\left(e^{i\phi_1},\dots,e^{i\phi_N}\right), \{\phi_\alpha\}_{\alpha=1}^N \sim \mathbb{U}[0,2\pi), V\in\mathcal{U}_{\rm Haar}(d-N)\right\}. \tag{68}$$

The unitary in the ensemble consists of two blocks, the first block $Q$ is a diagonal matrix of complex numbers with unity modulus and their corresponding angles are uniformly distributed within $[0, 2\pi)$ as there is no preferred distribution for the complex phases. The second block $V$ is sampled from Haar random unitaries with dimension $d-N$. Specifically, when $N \geq d-1$, $V$ reduces to a complex scalar $e^{i\phi}$ with $\phi$ uniformly distributed in $[0, 2\pi)$ as




Supplementary Figure 2. Distribution of complex angles. In (a), (b), we show the distribution of $\phi_1$ and $\phi_2$ from circuit unitaries at late time. We consider a $n = 2$ qubit multi-state preparation task, and the RPA consists of $L = 64$ parameters. In (c) we show the distribution of $\varphi$ generated from $d = 8$ haar random unitaries. The black dashed lines represent the p.d.f of uniform distribution $\mathbb{U}[-\pi, \pi)$.

well. The uniform distribution of $\phi_\alpha$ is verified in Fig. 2 (a) and (b) up to some fluctuations. Note that the ensemble $\mathcal{U}_{\rm RH}$ is a generalization of single-data restricted Haar ensemble discussed in Ref. [1].

To unveil ensemble properties of the restricted Haar ensemble, we focus on its frame potential [2], a quantity to represent the randomness of unitaries within the ensemble. Ahead of presenting the calculation details, we summarize the calculation results here. The $k$th frame potential of restricted Haar ensemble can be lower bounded by

$$\mathcal{F}^{(k)}_{\rm RH} \geq \begin{cases} \sum_{k_1={\rm even}}^{k} \sum_{k_2=0}^{k-k_1} \frac{k!}{((k_1/2)!)^2 k_2! (k-k_1-k_2)!} N^{k-k_1-k_2} \mathcal{F}^{(k_1/2+k_2)}_{\rm Haar}, & 1 \leq N < d-1 \\ \sum_{k_1={\rm even}}^{k} \frac{k!}{((k_1/2)!)^2 (k-k_1)!} d^{k-k_1}, & d-1 \leq N \leq d \end{cases}, \tag{69}$$

where the frame potential of Haar random unitaries is $\mathcal{F}^{(k)}_{\rm Haar} = k!$ [2]. Specifically for $k = 2$, the frame potential can be exactly solved as

$$\mathcal{F}^{(2)}_{\rm RH} = \begin{cases} 2N^2 + 3N + 2, & 1 \leq N < d-1 \\ 2d^2 - d, & d-1 \leq N \leq d \end{cases}. \tag{70}$$

In Fig. 3(a)-(b), we see that our lower bound (Eq. (69)) can characterize the leading order scaling of the exact $k$th frame potential for restricted Haar ensemble. Specifically, for $k = 2$, Eq. (70) (red line in Fig. 3 (b)) agrees with numerical results. In Fig. 3(a), the gap between $\mathcal{F}^{(k)}_{\rm RH}$ and $\mathcal{F}^{(k)}_{\rm Haar}$ enlarges with increasing $k$ for a fixed number of data $N$. On the other hand, in Fig. 3(b) for a specific order $k$ for example $k = 2$, the $\mathcal{F}^{(k)}_{\rm RH}$ increases with $N$ until convergence to a $d$-dependent constant, which is significantly different from the constant $\mathcal{F}^{(k)}_{\rm Haar} = k!$ of Haar ensemble. We can interpret the phenomena by the increasing number of constraints thus less degree of randomness of unitaries from $\mathcal{U}_{\rm RH}$ given more input data, leading to a larger frame potential.

The detailed calculations of QNTK matrix and relative dQNTK averaged over restricted Haar ensemble can be Appendix Supplementary Note 7.

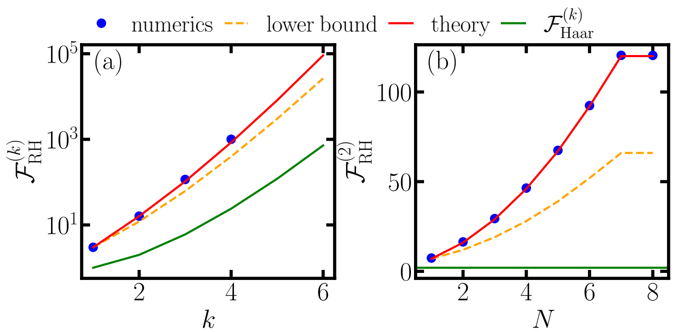
### Supplementary Note 5.1. Calculation details of frame potential

Following the definition, the $k$th frame potential of the restricted Haar ensemble unitaries becomes

$$\mathcal{F}^{(k)}_{\rm RH} = \frac{1}{|\mathcal{E}|^2} \sum_{U, U' \in \mathcal{U}_{\rm RH}} |\operatorname{tr}\left(U^\dagger U'\right)|^{2k} \tag{71}$$

$$= \frac{1}{|\mathcal{E}|^2} \sum_{V, V' \in \mathcal{U}_{\rm Haar}} |\operatorname{tr}\left(Q_N^\dagger Q'_N\right) + \operatorname{tr}\left(V^\dagger V'\right)|^{2k} \tag{72}$$

$$= \int_{\mathbb{U}[0,2\pi)} \prod_{\alpha=1}^{N} {\rm d}\phi_\alpha\, {\rm d}\phi'_\alpha \int_{\mathcal{U}_{\rm Haar}} {\rm d}V\, {\rm d}V' \left| \sum_{\alpha=1}^{N} e^{i(\phi'_\alpha - \phi_\alpha)} + \operatorname{tr}\left(V^\dagger V'\right) \right|^{2k} \tag{73}$$




Supplementary Figure 3. Frame potential of restricted Haar ensemble. In (a) the restricted Haar ensemble is in dimension of $d = 4$ with $N = 2$ data. In (b), the restricted Haar ensemble is in dimension $d = 8$ with various $N$. Blue dots are numerical results of an ensemble of $10^4$ unitaries sampled from $\mathcal{U}_{\rm RH}$. Red solid lines in (a) and (b) represent exact analytical results calculated from Eq. (75) and Eq. (70). The orange dashed lines represent the lower bound from Eq. (69). Green lines show the corresponding frame potential of haar random unitaries.

For convenience, we denote $z \equiv \operatorname{tr}\left(V^\dagger V'\right) = |z|e^{i\varphi}$, which is a complex scalar in general. As $V, V' \sim \mathcal{U}_{\rm Haar}$ without other limitations, we expect $\varphi \sim \mathbb{U}[0, 2\pi)$ (see example in Fig. 2 (c)), then we have

$$
\begin{aligned}
\mathcal{F}_{\rm RH}^{(k)} &= \int_{\mathbb{U}[0,2\pi)} \prod_{\alpha=1}^{N} \mathrm{d}\phi_\alpha \, \mathrm{d}\varphi \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}|z| \left[ N + 2\sum_{\alpha<\beta} \cos\left(\phi'_\alpha - \phi_\alpha - \phi'_\beta + \phi_\beta\right) + 2|z| \sum_\alpha \cos(\phi'_\alpha - \phi_\alpha - \varphi) + |z|^2 \right]^k && (74)\\
&= \int \prod_{i=1}^{\binom{N}{2}} \mathrm{d}x_i \prod_{j=1}^{N} \mathrm{d}y_j p_X(x_i) p_Y(y_j) \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}|z| \left[ N + 2\sum_{i=1}^{\binom{N}{2}} \cos(x_i) + 2|z| \sum_{j=1}^{N} \cos(y_j) + |z|^2 \right]^k && (75)\\
&\geq \int \mathrm{d}y_1 p_Y(y_1) \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}|z| \left[N + 2|z|\cos(y_1) + |z|^2\right]^k && (76)\\
&= \sum_{\substack{k_1,k_2=0\\k_1+k_2\leq k}} \int \mathrm{d}y_1 p_Y(y_1) \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}|z| \binom{k}{k_1, k_2} 2^{k_1} N^{k-k_1-k_2} \cos^{k_1}(y_1) |z|^{k_1+2k_2} && (77)\\
&= \sum_{\substack{k_1,k_2=0\\k_1+k_2\leq k}} \binom{k}{k_1, k_2} 2^{k_1} N^{k-k_1-k_2} \int \mathrm{d}y_1 p_Y(y_1) \cos^{k_1}(y_1) \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}|z| |z|^{k_1+2k_2} && (78)\\
&= \sum_{\substack{k_1,k_2=0\\k_1+k_2\leq k}} \binom{k}{k_1, k_2} 2^{k_1} N^{k-k_1-k_2} \mathbb{E}_{p_Y}\left[\cos^{k_1}(y_1)\right] \mathcal{F}_{\rm Haar}^{(k1/2+k_2)}, && (79)
\end{aligned}
$$

where in Eq. (75) we introduce the notation $x_i \equiv \phi_\alpha - \phi_\alpha - \phi'_\beta + \phi_\beta$ for $\alpha < \beta$ and $y_j \equiv \phi_\alpha - \phi_\alpha - \varphi$ for simplicity, and thus in total there are $\binom{N}{2}$ variables $x_i$ and $N$ variables $y_i$. As $\phi_\alpha, \varphi \sim \mathbb{U}[0, 2\pi)$, the distribution of $x_i$ and $y_j$ can be found to be

$$
p_X(x) = \begin{cases} \frac{(x+4\pi)^3}{96\pi^4}, & -4\pi \leq x \leq -2\pi \\ \frac{32\pi^3 - 12\pi x^2 - 3x^3}{96\pi^4}, & -2\pi \leq x \leq 0 \\ \frac{3x^3 - 12\pi x^2 + 32\pi^3}{96\pi^4}, & 0 \leq x \leq 2\pi \\ \frac{(4\pi - x)^3}{96\pi^4}, & 2\pi \leq x \leq 4\pi \end{cases}, \qquad p_Y(y) = \begin{cases} \frac{(y+4\pi)^2}{16\pi^3}, & -4\pi \leq y \leq -2\pi \\ \frac{2\pi^2 - 2\pi y - y^2}{8\pi^3}, & -2\pi \leq y \leq 0 \\ \frac{(y-2\pi)^2}{16\pi^3}, & 0 \leq y \leq 2\pi \end{cases} \qquad (80)
$$

The average $\mathbb{E}_{p_Y}[\cos^{k_1}(y_1)]$ can thus be evaluated as

$$\mathbb{E}_{p_Y}[\cos^{k_1}(y_1)] = \int_{-4\pi}^{2\pi} \mathrm{d}y_1 p_Y(y_1)\cos^{k_1}(y_1) \tag{81}$$

$$= \int_{-4\pi}^{2\pi} \mathrm{d}y_1 \frac{(y+4\pi)^2}{16\pi^3}\cos^{k_1}(y_1) + \int_{-2\pi}^{0} \mathrm{d}y_1 \frac{2\pi^2 - 2\pi y - y^2}{8\pi^3}\cos^{k_1}(y_1) + \int_{0}^{2\pi} \mathrm{d}y_1 \frac{(y-2\pi)^2}{16\pi^3}\cos^{k_1}(y_1) \tag{82}$$

$$= \int_{0}^{2\pi} \mathrm{d}y_1 \frac{y^2}{16\pi^3}\cos^{k_1}(y_1 - 4\pi) + \int_{0}^{2\pi} \mathrm{d}y_1 \frac{2\pi^2 - 2\pi(y-2\pi) - (y-2\pi)^2}{8\pi^3}\cos^{k_1}(y_1 - 2\pi) + \int_{0}^{2\pi} \mathrm{d}y_1 \frac{(y-2\pi)^2}{16\pi^3}\cos^{k_1}(y_1) \tag{83}$$

$$= \int_{0}^{2\pi} \mathrm{d}y_1 \frac{1}{2\pi}\cos^{k_1}(y_1) \tag{84}$$

$$= \frac{\left((-1)^{k_1}+1\right)^2 \Gamma\left(\frac{k_1+1}{2}\right)}{4\sqrt{\pi}\Gamma\left(\frac{k_1}{2}+1\right)}, \tag{85}$$

where in Eq. (83) we make the change of variables. Therefore, the frame potential can be reduced to

$$\mathcal{F}_{\mathrm{RH}}^{(k)} \geq \sum_{\substack{k_1,k_2=0\\k_1+k_2\leq k}} \binom{k}{k_1,k_2} 2^{k_1} N^{k-k_1-k_2} \mathbb{E}_{p_Y}\left[\cos^{k_1}(y_1)\right] \mathcal{F}_{\mathrm{Haar}}^{(k1/2+k_2)} \tag{86}$$

$$= \sum_{\substack{k_1,k_2=0\\k_1+k_2\leq k}} \binom{k}{k_1,k_2} 2^{k_1} N^{k-k_1-k_2} \frac{\left((-1)^{k_1}+1\right)^2 \Gamma\left(\frac{k_1+1}{2}\right)}{4\sqrt{\pi}\Gamma\left(\frac{k_1}{2}+1\right)} \mathcal{F}_{\mathrm{Haar}}^{(k1/2+k_2)} \tag{87}$$

$$= \sum_{k_1=\mathrm{even}}^{k} \sum_{k_2=0}^{k-k_1} \binom{k}{k_1,k_2} 2^{k_1} N^{k-k_1-k_2} \frac{\Gamma(k_1/2+1/2)}{\sqrt{\pi}\Gamma(k_1/2+1)} \mathcal{F}_{\mathrm{Haar}}^{(k1/2+k_2)} \tag{88}$$

$$= \sum_{k_1=\mathrm{even}}^{k} \sum_{k_2=0}^{k-k_1} \frac{k!}{k_1!k_2!(k-k_1-k_2)!} 2^{k_1} N^{k-k_1-k_2} \frac{2^{-k_1}\sqrt{\pi}\Gamma(k_1+1)}{\sqrt{\pi}\Gamma(k_1/2+1)^2} \mathcal{F}_{\mathrm{Haar}}^{(k1/2+k_2)} \tag{89}$$

$$= \sum_{k_1=\mathrm{even}}^{k} \sum_{k_2=0}^{k-k_1} \frac{k!}{((k_1/2)!)^2 k_2!(k-k_1-k_2)!} N^{k-k_1-k_2} \mathcal{F}_{\mathrm{Haar}}^{(k_1/2+k_2)}, \tag{90}$$

which holds for $N < d-1$. For a fixed $k$-th order, the leading order of the frame potential scales as $\mathcal{F}_{\mathrm{RH}}^{(k)} \sim N^k$. Specifically, for $k=2$, we can find the exact result from Eq. (75) as

$$\mathcal{F}_{\mathrm{RH}}^{(2)} = \int \prod_{i=1}^{\binom{N}{2}} \mathrm{d}x_i \prod_{j=1}^{N} \mathrm{d}y_j p_X(x_i) p_Y(y_j) \int_{\mathcal{U}_{\mathrm{Haar}}} \mathrm{d}|z| \left[ N + 2\sum_{i=1}^{\binom{N}{2}} \cos(x_i) + 2|z| \sum_{j=1}^{N} \cos(y_j) + |z|^2 \right]^2 \tag{91}$$

$$\begin{aligned} = \int \prod_{i=1}^{\binom{N}{2}} \mathrm{d}x_i \prod_{j=1}^{N} \mathrm{d}y_j p_X(x_i) p_Y(y_j) \int_{\mathcal{U}_{\mathrm{Haar}}} \mathrm{d}|z| \Bigg[ & N^2 + 4\sum_{i,i'=1}^{\binom{N}{2}} \cos(x_i)\cos(x_{i'}) + 4|z|^2 \sum_{j,j'=1}^{N} \cos(y_j)\cos(y_{j'}) + |z|^4 \\ & + 4N\sum_{i=1}^{\binom{N}{2}} \cos(x_i) + 4N|z|\sum_{j=1}^{N}\cos(y_j) + 2N|z|^2 + 8|z|\sum_{i,j}\cos(x_i)\cos(y_j) \\ & + 4|z|^2\sum_i \cos(x_i) + 2|z|^3 \sum_j \cos(y_j) \Bigg] \end{aligned} \tag{92}$$

$$= N^2 + 2\binom{N}{2} + 2N\mathcal{F}_{\mathrm{Haar}}^{(1)} + \mathcal{F}_{\mathrm{Haar}}^{(2)} + 2N\mathcal{F}_{\mathrm{Haar}}^{(1)} \tag{93}$$

$$= 2N^2 + 3N + 2, \tag{94}$$

where we utilize $\mathbb{E}_{p_X}[\cos(x)] = \mathbb{E}_{p_Y}[\cos(y)] = 0$ and $\mathbb{E}_{p_X}[\cos(x_i)\cos(x_{i'})] = \mathbb{E}_{p_Y}[\cos(y_i)\cos(y_{i'})] = \delta_{i,i'}/2$.

For $N \geq d-1$, the $k$th frame potential is reduced to

$$\mathcal{F}_{\mathrm{RH}}^{(k)} = \int_{\mathbb{U}[0,2\pi)} \prod_{\alpha=1}^{d} \mathrm{d}\phi_\alpha\, \mathrm{d}\phi'_\alpha \left| \sum_{\alpha=1}^{d} e^{i(\phi'_\alpha - \phi_\alpha)} \right|^{2k} \tag{95}$$

$$= \int_{\mathbb{U}[0,2\pi)} \prod_{\alpha=1}^{d} \mathrm{d}\phi_\alpha \left[ d + 2\sum_{\alpha<\beta} \cos\left(\phi'_\alpha - \phi_\alpha - \phi'_\beta + \phi_\beta\right) \right]^k \tag{96}$$

$$= \int \prod_{i=1}^{\binom{d}{2}} \mathrm{d}x_i p_X(x_i) \left[ d + 2\sum_{i=1}^{\binom{d}{2}} \cos(x_i) \right]^k \tag{97}$$

$$\geq \sum_{k_1} \binom{k}{k_1} 2^{k_1} \int \mathrm{d}x_1 p_X(x_1) \cos^{k_1}(x_1) d^{k-k_1} \tag{98}$$

$$= \sum_{k_1=\mathrm{even}}^{k} \binom{k}{k_1} 2^{k_1} d^{k-k_1} \frac{\Gamma(k_1/2+1/2)}{\sqrt{\pi}\Gamma(k_1/2+1)} \tag{99}$$

$$= \sum_{k_1=\mathrm{even}}^{k} \frac{k!}{((k_1/2)!)^2 (k-k_1)!} d^{k-k_1}. \tag{100}$$

Here we see that the $k$-th order frame potential leads to a constant only depending on the system dimension $d = 2^n$. For $k = 2$, we can also obtain the exact analytical result from Eq. 97 as

$$\mathcal{F}_{\mathrm{RH}}^{(2)} = \int \prod_{i=1}^{\binom{d}{2}} \mathrm{d}x_i p_X(x_i) \left[ d + 2\sum_{i=1}^{\binom{d}{2}} \cos(x_i) \right]^2 \tag{101}$$

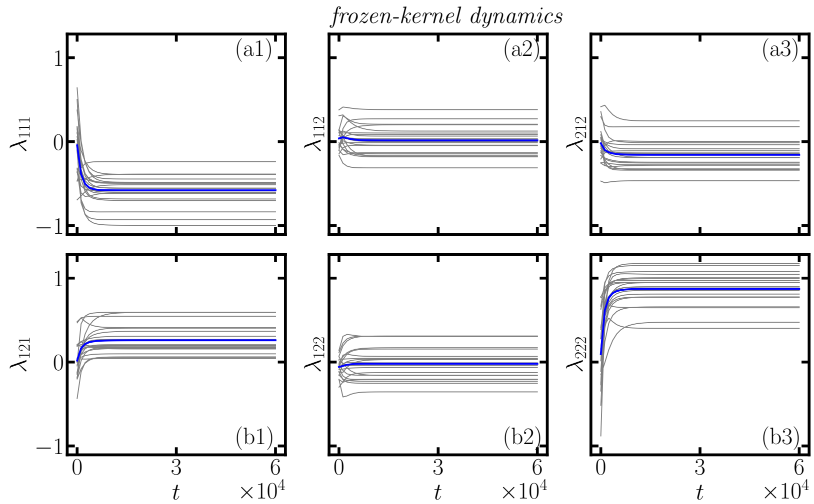
$$= \int \prod_{i=1}^{\binom{d}{2}} \mathrm{d}x_i p_X(x_i) \left[ d^2 + 4\sum_{i,i'=1}^{\binom{d}{2}} \cos(x_i)\cos(x_{i'}) + 4d\sum_{i=1}^{\binom{d}{2}} \cos(x_i) \right] \tag{102}$$

$$= 2d^2 - d. \tag{103}$$

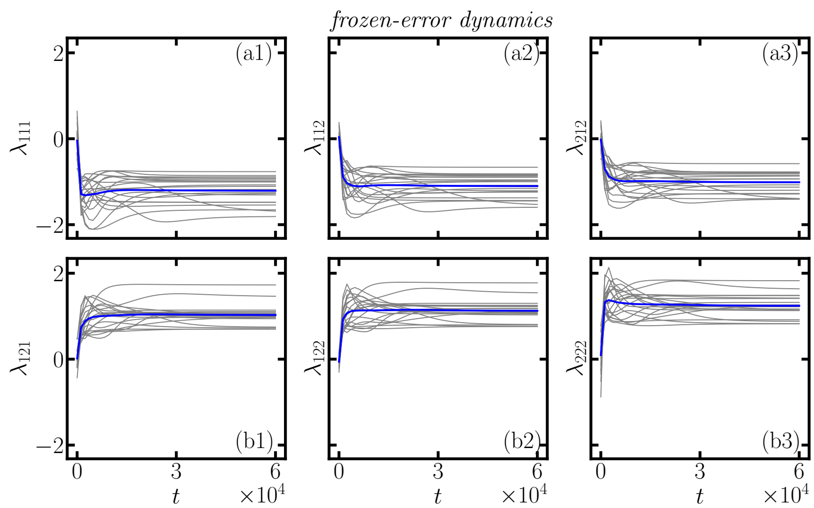
## Supplementary Note 6. ADDITIONAL NUMERICAL RESULTS

In the main text, we develop the coupled dynamical equations Eqs. (17) relying on an assumption that the relative dQNTK $\lambda_{\gamma\alpha\beta}(t) = \mu_{\gamma\alpha\beta}(t)/\sqrt{K_{\gamma\gamma}(t)K_{\beta\beta}(t)}$ converges to a constant at late time, and provide numerical results based on a generalized norm. In the following, we show the additional numerical evidence to support it for each type of dynamics. From the definition of $\lambda_{\gamma\alpha\beta}$, we see that $\lambda_{\gamma\alpha\beta} = \lambda_{\beta\alpha\gamma}$, and thus in the following we only present the independent elements. In Fig. 4, 5 and 6, we show the convergence of $\lambda_{\gamma\alpha\beta}$ for *frozen-kernel dynamics*, *frozen-error dynamics* and *mixed-frozen dynamics* in the exponential convergence class. In Fig. 7, 8, 9, 10, we plot its convergence for *critical point*, *critical-frozen-kernel dynamics*, *critical-frozen-error dynamics* and *critical-mixed-frozen dynamics* in the polynomial convergence class. In both convergence classes of dynamics, we see that every element of the relative dQNTK $\lambda_{\gamma\alpha\beta}$ converges to a constant in late time of training.

In Fig. 11 and Fig. 12, we show the convergence of geometric quantity $\angle_{\alpha\beta}(t)$ towards a constant for dynamics in exponential and polynomial convergence class, which supports Lemma. 2 in the main text. Indeed, the converged constant for each dynamics lies within the range $[-1, 1]$, indicating the geometric interpretation discussed in the main text.




Supplementary Figure 4. Dynamics of $\lambda_{\gamma\alpha\beta}$ for *frozen-kernel dynamics* in Fig. 5 (a1)-(c1) of the main text. Grey lines represent $\lambda_{\gamma\alpha\beta}$ of each random sample, and the blue lines represent the corresponding average.




Supplementary Figure 5. Dynamics of $\lambda_{\gamma\alpha\beta}$ for *frozen-error dynamics* in Fig. 5 (a2)-(c2) of the main text. Grey lines represent $\lambda_{\gamma\alpha\beta}$ of each random sample, and the blue lines represent the corresponding average.

## Supplementary Note 7. ADDITIONAL CALCULATIONS ON ENSEMBLE AVERAGE RESULTS

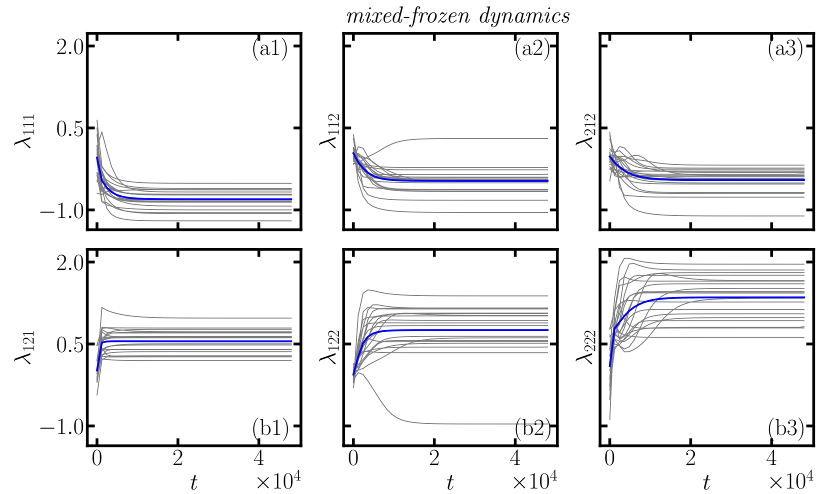
In this section, we present calculations for ensemble average of QNTK and dQNTK. As they are defined in terms of first and second-order derivatives, we first show the expression for gradients. From parameter-shift rule, the derivative of $\epsilon_\alpha = \langle\psi_\alpha|U^\dagger O_\alpha U|\psi_\alpha\rangle$ with $O_\alpha = |\Phi_\alpha\rangle\langle\Phi_\alpha|$ is

$$\frac{\partial\epsilon_\alpha}{\partial\theta_\ell} = \frac{i}{2}\langle\psi_\alpha|U_{\ell^-}^\dagger\left[X_\ell, O_{\alpha;\ell^+}\right]U_{\ell^-}|\psi_\alpha\rangle\,, \tag{104}$$

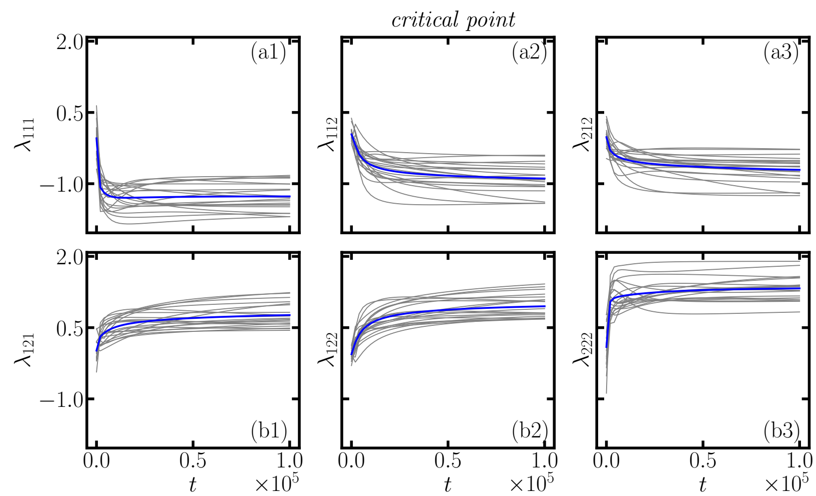
where we define the notation

$$U_{\ell^-} = \prod_{k=1}^{\ell-1} W_k V_k(\theta_k), U_{\ell^+} = \prod_{k=\ell}^{L} W_k V_k(\theta_k), \tag{105}$$

and $O_{\alpha;\ell^+} = U_{\ell^+}^\dagger O_\alpha U_{\ell^+}$. Thus the unitary for whole circuit becomes $U = U_{\ell^+}U_{\ell^-}$. One can show that Eq. (104) is equivalent to the parameter-shift rule [3, 4], and the expression here is also utilized in previous related works [1, 5].




Supplementary Figure 6. Dynamics of $\lambda_{\gamma\alpha\beta}$ for *mixed-frozen dynamics* in Fig. 5 (a3)-(c3) of the main text. Grey lines represent $\lambda_{\gamma\alpha\beta}$ of each random sample, and the blue lines represent the corresponding average.




Supplementary Figure 7. Dynamics of $\lambda_{\gamma\alpha\beta}$ for *critical point* in Fig. 6 (a1)-(c1) of the main text. Grey lines represent $\lambda_{\gamma\alpha\beta}$ of each random sample, and the blue lines represent the corresponding average.

The second order gradient assuming $\ell_1 < \ell_2$ and $\ell_1 = \ell_2 = \ell$ can be written in a similar way as

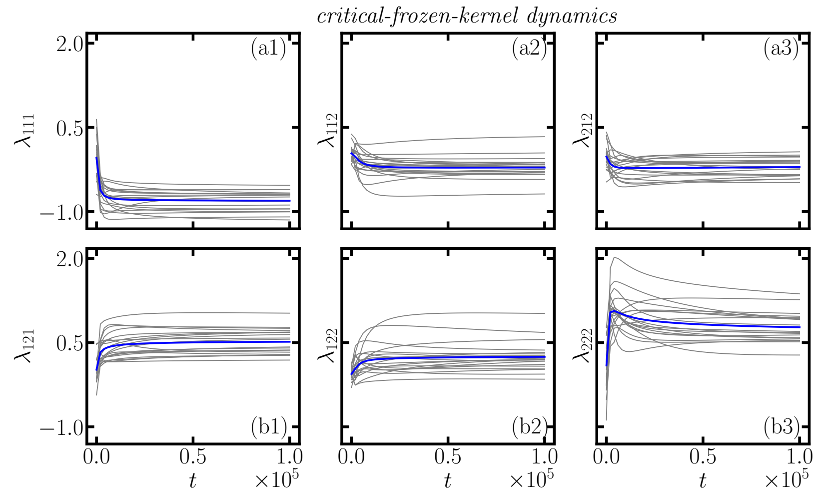
$$\frac{\partial^2 \epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}} = -\frac{1}{4}\langle\psi_\alpha| U^\dagger_{\ell_1^-}[X_{\ell_1}, U^\dagger_{\ell_1\to\ell_2}[X_{\ell_2}, U^\dagger_{\ell_2^+} O_\alpha U_{\ell_2^+}] U_{\ell_1\to\ell_2}] U_{\ell_1^-}|\psi_\alpha\rangle = -\frac{1}{4}\langle\psi_\alpha| U^\dagger_{\ell_1^-}[X_{\ell_1}, U^\dagger_{\ell_1\to\ell_2}[X_{\ell_2}, O_{\alpha;\ell_2^+}] U_{\ell_1\to\ell_2}] U_{\ell_1^-}|\psi_\alpha\rangle \tag{106}$$

$$\frac{\partial^2 \epsilon_\alpha}{\partial\theta_\ell^2} = -\frac{1}{4}\langle\psi_\alpha| U^\dagger_{\ell^-}[X_\ell, [X_\ell, O_{\alpha;\ell^+}]] U_{\ell^-}|\psi_\alpha\rangle\,, \tag{107}$$

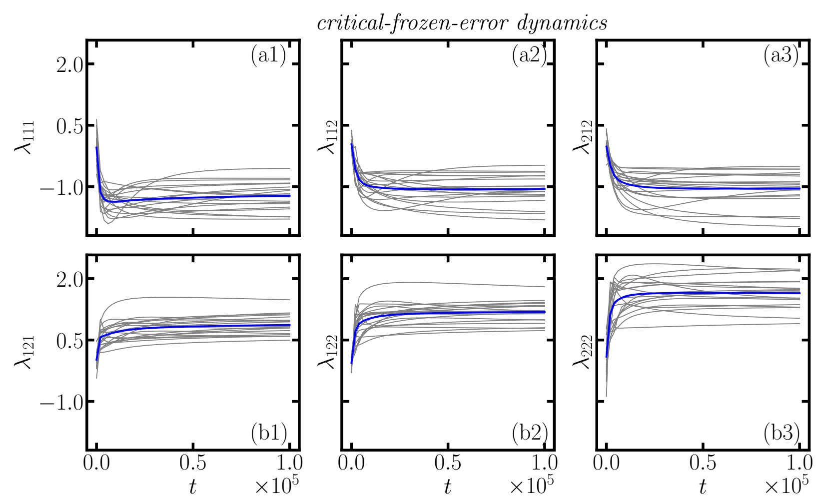
where

$$U_{\ell_1\to\ell_2} = \prod_{k=\ell_1}^{\ell_2-1} W_k V_k(\theta_k). \tag{108}$$

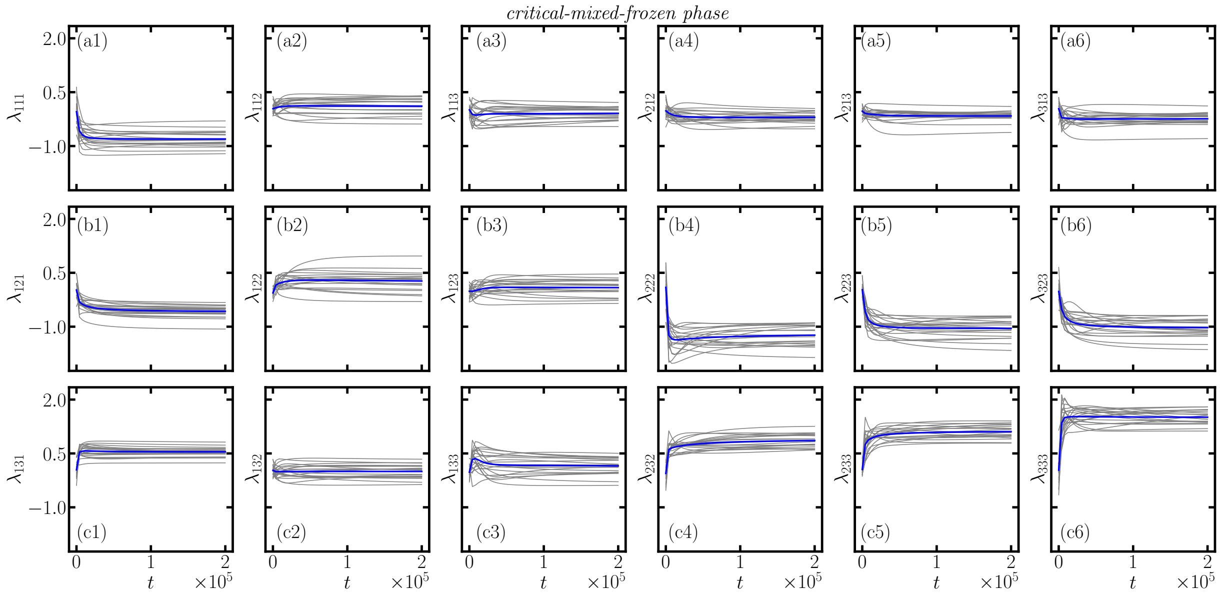
The ensemble average over Haar random unitaries are performed via symbolic calculation tools `RTNI` [6].



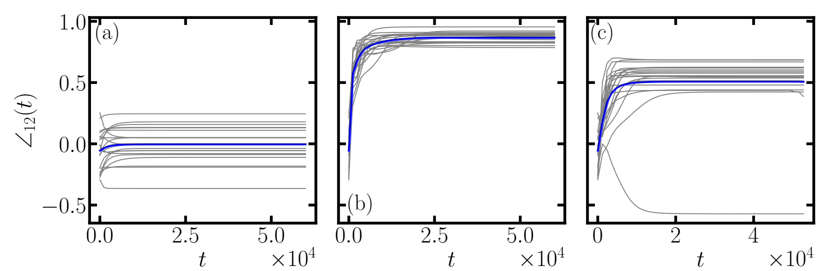

Supplementary Figure 8. Dynamics of $\lambda_{\gamma\alpha\beta}$ for *critical-frozen-kernel dynamics* in Fig. 6 (a2)-(c2) of the main text. Grey lines represent $\lambda_{\gamma\alpha\beta}$ of each random sample, and the blue lines represent the corresponding average.



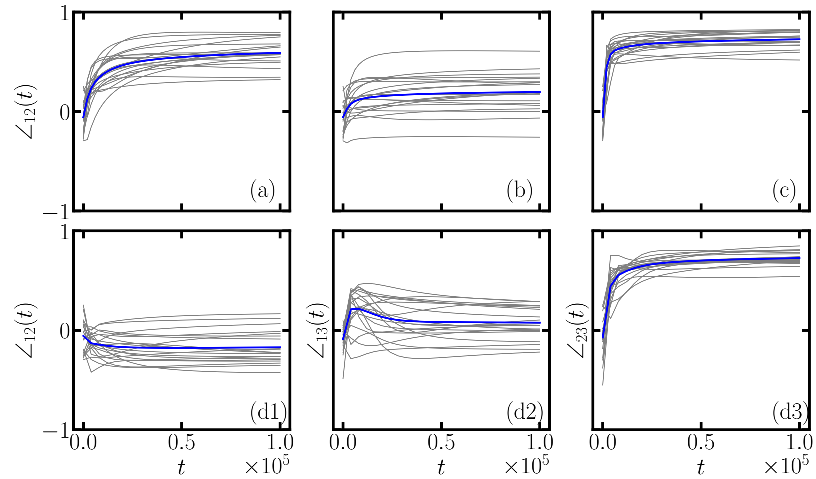

Supplementary Figure 9. Dynamics of $\lambda_{\gamma\alpha\beta}$ for *critical-frozen-kernel dynamics* in Fig. 6 (a3)-(c3) of the main text. Grey lines represent $\lambda_{\gamma\alpha\beta}$ of each random sample, and the blue lines represent the corresponding average.




Supplementary Figure 10. Dynamics of $\lambda_{\gamma\alpha\beta}$ for *critical-frozen-kernel dynamics* in Fig. 7 of the main text. Grey lines represent $\lambda_{\gamma\alpha\beta}$ of each random sample, and the blue lines represent the corresponding average.




Supplementary Figure 11. Dynamics of $\angle_{12}(t)$ for exponential convergence class. From left to right we show $\angle_{12}(t)$ for *frozen-kernel dynamics*, *frozen-error dynamics* and *mixed-frozen dynamics*. Grey lines represent $\angle_{12}$ of each random sample, and the blue lines represent the corresponding average. The settings follow Fig. 5 of the main text.




Supplementary Figure 12. Dynamics of $\angle_{\alpha\beta}(t)$ for polynomial convergence class. We show $\angle_{12}(t)$ for (a) *critical point*, (b) *critical-frozen-kernel dynamics* and (c) *critical-frozen-error dynamics*. We plot $\angle_{12}(t), \angle_{13}(t), \angle_{23}(t)$ in (d1)-(d3) for *critical-mixed-frozen dynamics*. Grey lines represent $\angle_{12}$ of each random sample, and the blue lines represent the corresponding average. The settings of top and bottom panels follow Fig. 6 and Fig. 7 of the main text separately.

### Supplementary Note 7.1. Average QNTK under restricted Haar ensemble

For the QNTK $K_{\alpha\beta} = \sum_\ell \frac{\partial\epsilon_\alpha}{\partial\theta_\ell}\frac{\partial\epsilon_\beta}{\partial\theta_\ell}$, the restricted Haar ensemble average of product of derivatives become

$$\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial\epsilon_\alpha}{\partial\theta_\ell}\frac{\partial\epsilon_\beta}{\partial\theta_\ell}\right] = -\frac{1}{4}\int {\rm d}U_{\ell-}\,{\rm d}U_{\ell+}\,{\rm tr}\left(P_{\beta\alpha}U_{\ell-}^\dagger\left[X_\ell, O_{\alpha;\ell+}\right]U_{\ell-}P_{\alpha\beta}U_{\ell-}^\dagger\left[X_\ell, O_{\beta;\ell+}\right]U_{\ell-}\right) \tag{109}$$

$$= -\frac{1}{4}\int_{\mathcal{U}_{\rm RH}}{\rm d}U\int_{\mathcal{U}_{\rm Haar}}{\rm d}U_{\ell-}\Big[{\rm tr}\left(P_{\beta\alpha}U_{\ell-}^\dagger X_\ell U_{\ell-}O_{\alpha;U}P_{\alpha\beta}U_{\ell-}^\dagger X_\ell U_{\ell-}O_{\beta;U}\right) + {\rm tr}\left(P_{\beta\alpha}O_{\alpha;U}U_{\ell-}^\dagger X_\ell U_{\ell-}P_{\alpha\beta}O_{\beta;U}U_{\ell-}^\dagger X_\ell U_{\ell-}\right)$$
$$- {\rm tr}\left(P_{\beta\alpha}U_{\ell-}^\dagger X_\ell U_{\ell-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U_{\ell-}^\dagger X_\ell U_{\ell-}\right) - {\rm tr}\left(P_{\beta\alpha}O_{\alpha;U}U_{\ell-}^\dagger X_\ell U_{\ell-}P_{\alpha\beta}U_{\ell-}^\dagger X_\ell U_{\ell-}O_{\beta;U}\right)\Big] \tag{110}$$

$$= -\frac{1}{4}\int_{\mathcal{U}_{\rm RH}}{\rm d}U\Big[\frac{d\,{\rm tr}(O_{\alpha;U}P_{\alpha\beta})\,{\rm tr}(O_{\beta;U}P_{\beta\alpha}) - {\rm tr}(P_{\alpha\beta}O_{\beta;U}P_{\beta\alpha}O_{\alpha;U})}{d^2-1} + \frac{d\,{\rm tr}(P_{\alpha\beta}O_{\beta;U})\,{\rm tr}(P_{\beta\alpha}O_{\alpha;U}) - {\rm tr}(P_{\beta\alpha}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U})}{d^2-1}$$
$$-\frac{d\,{\rm tr}(P_{\beta\alpha})\,{\rm tr}(O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}) - {\rm tr}(P_{\beta\alpha}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U})}{d^2-1} - \frac{d\,{\rm tr}(P_{\alpha\beta})\,{\rm tr}(O_{\beta;U}P_{\beta\alpha}O_{\alpha;U}) - {\rm tr}(O_{\beta;U}P_{\beta\alpha}O_{\alpha;U}P_{\alpha\beta})}{d^2-1}\Big] \tag{111}$$

$$= -\frac{d}{4}\int_{\mathcal{U}_{\rm RH}}{\rm d}U\frac{{\rm tr}(O_{\alpha;U}P_{\alpha\beta})\,{\rm tr}(O_{\beta;U}P_{\beta\alpha}) + {\rm tr}(P_{\alpha\beta}O_{\beta;U})\,{\rm tr}(P_{\beta\alpha}O_{\alpha;U}) - {\rm tr}(P_{\beta\alpha})\,{\rm tr}(O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}) - {\rm tr}(P_{\alpha\beta})\,{\rm tr}(O_{\beta;U}P_{\beta\alpha}O_{\alpha;U})}{d^2-1} \tag{112}$$

$$= -\frac{d}{4}\int_{\mathcal{U}_{\rm RH}}{\rm d}U\frac{{\rm tr}(O_{\alpha;U}P_{\alpha\beta})\,{\rm tr}(O_{\beta;U}P_{\beta\alpha}) + {\rm tr}(P_{\alpha\beta}O_{\beta;U})\,{\rm tr}(P_{\beta\alpha}O_{\alpha;U}) - \langle\psi_\alpha|\psi_\beta\rangle\,{\rm tr}(O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}) - \langle\psi_\beta|\psi_\alpha\rangle\,{\rm tr}(O_{\beta;U}P_{\beta\alpha}O_{\alpha;U})}{d^2-1} \tag{113}$$

$$= -\frac{d}{4(d^2-1)}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[T_{\alpha\beta}^*T_{\alpha\alpha}T_{\beta\alpha}^*T_{\beta\beta} + T_{\beta\beta}^*T_{\beta\alpha}T_{\alpha\alpha}^*T_{\alpha\beta} - \langle\psi_\alpha|\psi_\beta\rangle\langle\Phi_\beta|\Phi_\alpha\rangle T_{\alpha\alpha}T_{\beta\beta}^* - \langle\psi_\beta|\psi_\alpha\rangle\langle\Phi_\alpha|\Phi_\beta\rangle T_{\beta\beta}T_{\alpha\alpha}^*\right] \tag{114}$$

$$= -\frac{d}{4(d^2-1)}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[T_{\alpha\beta}^*T_{\alpha\alpha}T_{\beta\alpha}^*T_{\beta\beta} - \delta_{\alpha\beta}T_{\alpha\alpha}T_{\beta\beta}^* + c.c.\right], \tag{115}$$

where $P_{\alpha\beta} = |\alpha\rangle\langle\beta|$ is an operator introduced for the convenience of unitary integral calculation utilizing `RTNI` and $T_{\alpha\beta} \equiv \langle\Phi_\alpha|U|\psi_\beta\rangle$. Here *c.c.* stands for complex conjugate.

For $\alpha = \beta$, we have

$$\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\right] = -\frac{d}{2(d^2-1)}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[|T_{\alpha\alpha}|^4 - |T_{\alpha\alpha}|^2\right] = \frac{d}{2(d^2-1)}o_\alpha(1-o_\alpha), \tag{116}$$

where we utilize $|T_{\alpha\alpha}|^2 = |\langle\Phi_\alpha|U|\psi_\alpha\rangle|^2 = o_\alpha$. On the other hand, for $\alpha \neq \beta$, it becomes

$$\begin{aligned}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\frac{\partial \epsilon_\beta}{\partial \theta_\ell}\right] &= -\frac{d}{4(d^2-1)}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[T^*_{\alpha\beta}T_{\alpha\alpha}T^*_{\beta\alpha}T_{\beta\beta} + c.c.\right] &(117)\\ &= -\frac{d}{4(d^2-1)}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[|T_{\alpha\beta}|e^{-i\phi_\beta}|T_{\alpha\alpha}|e^{i\phi_\alpha}|T_{\beta\alpha}|e^{-i\phi_\alpha}|T_{\beta\beta}|e^{i\phi_\beta} + c.c.\right] &(118)\\ &= -\frac{d}{2(d^2-1)}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[|T_{\alpha\beta}||T_{\alpha\alpha}||T_{\beta\alpha}||T_{\beta\beta}|\right], &(119)\end{aligned}$$

where in the second line, we utilize the definition of restricted Haar ensemble in Eq. (68). We see that the off-diagonal terms require extra information.

The average QNTK under restricted Haar ensemble becomes

$$\begin{aligned}\overline{K_{\alpha\alpha}(\infty)} &= L\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\right] = \frac{Ld}{2(d^2-1)}o_\alpha(1-o_\alpha) \simeq \frac{L}{2d}o_\alpha(1-o_\alpha), &(120)\\ \overline{K_{\alpha\beta}(\infty)} &= L\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\frac{\partial \epsilon_\beta}{\partial \theta_\ell}\right] = -\frac{Ld}{2(d^2-1)}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[|T_{\alpha\beta}||T_{\alpha\alpha}||T_{\beta\alpha}||T_{\beta\beta}|\right] \simeq -\frac{L}{2d}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[|T_{\alpha\beta}||T_{\alpha\alpha}||T_{\beta\alpha}||T_{\beta\beta}|\right], &(121)\end{aligned}$$

where we approximate them with $d \gg 1$ at the end.

### Supplementary Note 7.2. Average relative dQNTK under restricted Haar ensemble

Ahead of presenting the calculation details of relative QNTK, we summarize the results here.

$$\overline{\lambda_{\alpha\alpha\alpha}(\infty)} = \frac{\overline{\mu_{\alpha\alpha\alpha}(\infty)}}{\overline{K_{\alpha\alpha}(\infty)}} \simeq -\frac{1}{4d}\left[2(do_\alpha - 2) + L(2o_\alpha - 1)\right]. \tag{122}$$

In this section, we evaluate the relative dQNTK $\overline{\lambda_{\gamma\alpha\beta}(\infty)} = \overline{\mu_{\gamma\alpha\beta}(\infty)}/\sqrt{\overline{K_{\alpha\alpha}(\infty)K_{\beta\beta}(\infty)}}$. We first calculate $\overline{\mu_{\gamma\alpha\beta}(\infty)}$. Recall that $\mu_{\gamma\alpha\beta} = \sum_{\ell,\ell'}\frac{\partial\epsilon_\gamma}{\partial\theta_\ell}\frac{\partial^2\epsilon_\alpha}{\partial\theta_\ell\partial\theta_{\ell'}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell'}} = \sum_\ell \frac{\partial\epsilon_\gamma}{\partial\theta_\ell}\frac{\partial^2\epsilon_\alpha}{\partial\theta_\ell^2}\frac{\partial\epsilon_\beta}{\partial\theta_\ell} + \sum_{\ell\neq\ell'}\frac{\partial\epsilon_\gamma}{\partial\theta_\ell}\frac{\partial^2\epsilon_\alpha}{\partial\theta_\ell\partial\theta_{\ell'}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell'}}$, we then calculate the ensemble average of the two terms separately. As only $\lambda_{\alpha\alpha\beta}$ is utilized in the dynamical equations (see Eq. (30)), then we only consider ensemble average of $\mu_{\alpha\alpha\beta}$ in the following.

#### *Supplementary Note 7.2.1.* $\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial\epsilon_\alpha}{\partial\theta_\ell}\frac{\partial^2\epsilon_\alpha}{\partial\theta_\ell^2}\frac{\partial\epsilon_\beta}{\partial\theta_\ell}\right]$ *under restricted Haar ensemble*

We can expand it following the parameter-shift rule as

$$\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial\epsilon_\alpha}{\partial\theta_\ell}\frac{\partial^2\epsilon_\alpha}{\partial\theta_\ell^2}\frac{\partial\epsilon_\beta}{\partial\theta_\ell}\right] = \frac{1}{16}\int {\rm d}U_{\ell-}\,{\rm d}U_{\ell+}\,{\rm tr}\Big(P_{\alpha\alpha}U^\dagger_{\ell-}[X_\ell,[X_\ell,O_{\alpha;\ell+}]]U_{\ell-}P_{\alpha\beta}U^\dagger_{\ell-}\left[X_\ell,O_{\beta;\ell+}\right]U_{\ell-}P_{\beta\alpha}U^\dagger_{\ell-}\left[X_\ell,O_{\alpha;\ell+}\right]U_{\ell-}\Big) \tag{123}$$

$$\begin{aligned}= \frac{2}{16}\int_{\mathcal{U}_{\rm RH}}{\rm d}U\int_{\mathcal{U}_{\rm Haar}}{\rm d}U_{\ell-}\Big[&{\rm tr}\Big(P_{\alpha\alpha}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\beta;U}P_{\beta\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}\Big) + {\rm tr}\Big(P_{\alpha\alpha}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}\Big)\\ &- {\rm tr}\Big(P_{\alpha\alpha}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\beta;U}P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}\Big) - {\rm tr}\Big(P_{\alpha\alpha}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\beta\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}\Big)\\ &+ {\rm tr}\Big(P_{\alpha\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\alpha\beta}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\beta;U}P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}\Big)\\ &+ {\rm tr}\Big(P_{\alpha\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\beta\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}\Big)\\ &- {\rm tr}\Big(P_{\alpha\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\alpha\beta}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\beta;U}P_{\beta\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}\Big)\\ &- {\rm tr}\Big(P_{\alpha\alpha}U^\dagger_{\ell-}X_\ell U_{\ell-}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell-}X_\ell U_{\ell-}P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell-}X_\ell U_{\ell-}\Big)\Big].\end{aligned} \tag{124}$$

The first term is

$$\begin{aligned}
I_1 \equiv & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \int_{\mathcal{U}_{\mathrm{Haar}}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} O_{\beta;U} P_{\beta\alpha} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} O_{\alpha;U}\Big) \\
= & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \frac{d \operatorname{tr}(O_{\beta;U} P_{\beta\alpha}) \operatorname{tr}(O_{\alpha;U} P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta}) - \operatorname{tr}(O_{\alpha;U} P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U} P_{\beta\alpha})}{d^2-1} && (125) \\
= & \frac{1}{d^2-1} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ d T_{\beta\beta} T_{\beta\alpha}^* T_{\alpha\alpha} T_{\alpha\alpha}^* T_{\alpha\alpha} T_{\alpha\beta}^* - T_{\alpha\alpha} T_{\alpha\alpha}^* T_{\alpha\alpha} T_{\beta\beta}^* T_{\beta\beta} T_{\alpha\alpha}^* \right] && (126) \\
= & \frac{1}{d^2-1} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ d T_{\beta\beta} T_{\beta\alpha}^* T_{\alpha\alpha} |T_{\alpha\alpha}|^2 T_{\alpha\beta}^* - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 \right]. && (127)
\end{aligned}$$

The second term is

$$\begin{aligned}
I_2 \equiv & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \int_{\mathcal{U}_{\mathrm{Haar}}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} P_{\beta\alpha} O_{\alpha;U} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-}\Big) \\
= & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \frac{d \operatorname{tr}(P_{\beta\alpha} O_{\alpha;U}) \operatorname{tr}(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U}) - \operatorname{tr}(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U} P_{\beta\alpha} O_{\alpha;U})}{d^2-1} && (128) \\
= & \frac{1}{d^2-1} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ d T_{\alpha\beta} T_{\alpha\alpha}^* |T_{\alpha\alpha}|^2 T_{\beta\beta}^* T_{\beta\alpha} - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 \right] = I_1^*. && (129)
\end{aligned}$$

The third term is

$$\begin{aligned}
I_3 \equiv & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \int_{\mathcal{U}_{\mathrm{Haar}}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} O_{\beta;U} P_{\beta\alpha} O_{\alpha;U} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-}\Big) \\
= & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \frac{d \operatorname{tr}(O_{\beta;U} P_{\beta\alpha} O_{\alpha;U}) \operatorname{tr}(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta}) - \operatorname{tr}(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U} P_{\beta\alpha} O_{\alpha;U})}{d^2-1} && (130) \\
= & \frac{1}{d^2-1} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ d T_{\beta\beta} T_{\alpha\alpha}^* \langle \Phi_\alpha | \Phi_\beta \rangle T_{\alpha\alpha}^* T_{\alpha\alpha} \langle \psi_\beta | \psi_\alpha \rangle - |T_{\alpha\alpha}|^2 |T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2 \right] && (131) \\
= & \frac{1}{d^2-1} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ d \delta_{\alpha\beta} T_{\beta\beta} T_{\alpha\alpha}^* |T_{\alpha\alpha}|^2 - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 \right]. && (132)
\end{aligned}$$

The forth term is

$$\begin{aligned}
I_4 \equiv & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \int_{\mathcal{U}_{\mathrm{Haar}}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} P_{\beta\alpha} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} O_{\alpha;U}\Big) \\
= & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \frac{d \operatorname{tr}(P_{\beta\alpha}) \operatorname{tr}(O_{\alpha;U} P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U}) - \operatorname{tr}(O_{\alpha;U} P_{\alpha\alpha} O_{\alpha;U} P_{\alpha\beta} O_{\beta;U} P_{\beta\alpha})}{d^2-1} && (133) \\
= & \frac{1}{d^2-1} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ d \langle \psi_\alpha | \psi_\beta \rangle T_{\alpha\alpha} T_{\alpha\alpha}^* T_{\alpha\alpha} T_{\beta\beta}^* \langle \Phi_\beta | \Phi_\alpha \rangle - |T_{\alpha\alpha}|^2 |T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2 \right] && (134) \\
= & \frac{1}{d^2-1} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ d \delta_{\alpha\beta} T_{\alpha\alpha} |T_{\alpha\alpha}|^2 T_{\beta\beta}^* - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 \right] = I_3^*. && (135)
\end{aligned}$$

The fifth term is

$$\begin{aligned}
I_5 \equiv & \int_{\mathcal{U}_{\mathrm{RH}}} \mathrm{d}U \int_{\mathcal{U}_{\mathrm{Haar}}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} O_{\alpha;U} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} P_{\alpha\beta} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-} O_{\beta;U} P_{\beta\alpha} O_{\alpha;U} U_{\ell^-}^{\dagger} X_\ell U_{\ell^-}\Big) \\
= & \delta_{\alpha\beta} \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ \frac{2(d+2) T_{\beta\beta} T_{\alpha\alpha}^* - 2 T_{\beta\beta} |T_{\alpha\alpha}|^2 \left( T_{\alpha\beta}^* + T_{\beta\alpha}^* + T_{\alpha\alpha}^* \right)}{d^3 + 3d^2 - d - 3} \right] \\
& + \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}} \left[ \frac{(d+2) T_{\alpha\alpha} T_{\beta\beta} |T_{\alpha\alpha}|^2 T_{\alpha\beta}^* T_{\beta\alpha}^* - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 - 2 |T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2}{d^3 + 3d^2 - d - 3} \right]. && (136)
\end{aligned}$$

The sixth term is

$$\begin{aligned} I_6 &\equiv \int_{\mathcal{U}_{\rm RH}} \mathrm{d}U \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} U_{\ell^-}^\dagger X_\ell U_{\ell^-} O_{\alpha;U} U_{\ell^-}^\dagger X_\ell U_{\ell^-} P_{\alpha\beta} O_{\beta;U} U_{\ell^-}^\dagger X_\ell U_{\ell^-} P_{\beta\alpha} U_{\ell^-}^\dagger X_\ell U_{\ell^-} O_{\alpha;U}\Big) \\ &= \delta_{\alpha\beta} \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{2(d+2) T_{\alpha\alpha} T_{\beta\beta}^* - 2|T_{\alpha\alpha}|^2 T_{\beta\beta}^* (T_{\alpha\alpha} + T_{\alpha\beta} + T_{\beta\alpha})}{d^3 + 3d^2 - d - 3} \right] \\ &\quad + \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{(d+2) T_{\alpha\beta} T_{\beta\alpha} |T_{\alpha\alpha}|^2 T_{\alpha\alpha}^* T_{\beta\beta}^* - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 - 2|T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2}{d^3 + 3d^2 - d - 3} \right] = I_5^*. \end{aligned} \tag{137}$$

The seventh term is

$$\begin{aligned} I_7 &\equiv \int_{\mathcal{U}_{\rm RH}} \mathrm{d}U \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} U_{\ell^-}^\dagger X_\ell U_{\ell^-} O_{\alpha;U} U_{\ell^-}^\dagger X_\ell U_{\ell^-} P_{\alpha\beta} U_{\ell^-}^\dagger X_\ell U_{\ell^-} O_{\beta;U} P_{\beta\alpha} U_{\ell^-}^\dagger X_\ell U_{\ell^-} O_{\alpha;U}\Big) \\ &= \delta_{\alpha\beta} \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{T_{\beta\beta} |T_{\alpha\alpha}|^2 \Big((d+2) T_{\alpha\alpha}^* - 2T_{\alpha\beta}^* - 2T_{\beta\alpha}^*\Big)}{d^3 + 3d^2 - d - 3} \right] \\ &\quad + \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{2(d+2) T_{\alpha\alpha} T_{\beta\beta} T_{\alpha\beta}^* T_{\beta\alpha}^* - 2T_{\alpha\alpha} T_{\beta\beta} |T_{\alpha\alpha}|^2 T_{\alpha\beta}^* T_{\beta\alpha}^* - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 - 2|T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2}{d^3 + 3d^2 - d - 3} \right]. \end{aligned} \tag{138}$$

The eighth (last) term is

$$\begin{aligned} I_8 &\equiv \int_{\mathcal{U}_{\rm RH}} \mathrm{d}U \int_{\mathcal{U}_{\rm Haar}} \mathrm{d}U_{\ell^-} \operatorname{tr}\Big(P_{\alpha\alpha} U_{\ell^-}^\dagger X_\ell U_{\ell^-} O_{\alpha;U} U_{\ell^-}^\dagger X_\ell U_{\ell^-} P_{\alpha\beta} O_{\beta;U} U_{\ell^-}^\dagger X_\ell U_{\ell^-} P_{\beta\alpha} O_{\alpha;U} U_{\ell^-}^\dagger X_\ell U_{\ell^-}\Big) \\ &= \delta_{\alpha\beta} \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{T_{\beta\beta}^* |T_{\alpha\alpha}|^2 \left((d+2) T_{\alpha\alpha} - 2T_{\alpha\beta} - 2T_{\beta\alpha}\right)}{d^3 + 3d^2 - d - 3} \right] \\ &\quad + \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{2(d+2) T_{\alpha\beta} T_{\beta\alpha} T_{\alpha\alpha}^* T_{\beta\beta}^* - 2T_{\alpha\beta} T_{\beta\alpha} |T_{\alpha\alpha}|^2 T_{\alpha\alpha}^* T_{\beta\beta}^* - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 - 2|T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2}{d^3 + 3d^2 - d - 3} \right] = I_7^*. \end{aligned} \tag{139}$$

Then we have

$$\mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{\partial \epsilon_\alpha}{\partial \theta_\ell} \frac{\partial^2 \epsilon_\alpha}{\partial \theta_\ell^2} \frac{\partial \epsilon_\beta}{\partial \theta_\ell} \right] = \frac{2}{16} \left( I_1 + I_2 - I_3 - I_4 + I_5 + I_6 - I_7 - I_8 \right)$$

$$= \frac{1}{8} \left( I_1 - I_3 + I_5 - I_7 + c.c. \right) \tag{140}$$

$$\begin{aligned} &= \frac{1}{8} \Bigg( \frac{1}{d^2 - 1} \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ d T_{\beta\beta} T_{\beta\alpha}^* T_{\alpha\alpha} |T_{\alpha\alpha}|^2 T_{\alpha\beta}^* - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 \right] - \frac{1}{d^2 - 1} \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ d \delta_{\alpha\beta} T_{\beta\beta} T_{\alpha\alpha}^* |T_{\alpha\alpha}|^2 - |T_{\alpha\alpha}|^4 |T_{\beta\beta}|^2 \right] \\ &+ \delta_{\alpha\beta} \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{2(d+2) T_{\beta\beta} T_{\alpha\alpha}^* - 2T_{\beta\beta} |T_{\alpha\alpha}|^2 \left(T_{\alpha\beta}^* + T_{\beta\alpha}^* + T_{\alpha\alpha}^*\right)}{d^3 + 3d^2 - d - 3} \right] + \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{(d+2) T_{\alpha\alpha} T_{\beta\beta} |T_{\alpha\alpha}|^2 T_{\alpha\beta}^* T_{\beta\alpha}^* - |T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2 \left(|T_{\alpha\alpha}|^2 + 2\right)}{d^3 + 3d^2 - d - 3} \right] \\ &- \delta_{\alpha\beta} \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{T_{\beta\beta} |T_{\alpha\alpha}|^2 \left((d+2) T_{\alpha\alpha}^* - 2T_{\alpha\beta}^* - 2T_{\beta\alpha}^*\right)}{d^3 + 3d^2 - d - 3} \right] - \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{2T_{\alpha\alpha} T_{\beta\beta} T_{\alpha\beta}^* T_{\beta\alpha}^* \left(d + 2 - |T_{\alpha\alpha}|^2\right) - |T_{\alpha\alpha}|^2 |T_{\beta\beta}|^2 \left(|T_{\alpha\alpha}|^2 + 2\right)}{d^3 + 3d^2 - d - 3} \right] + c.c. \Bigg) \end{aligned} \tag{141}$$

$$= \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{1}{8(d^2 - 1)} \left( \frac{(d+2)^2}{d+3} |T_{\alpha\alpha}|^2 - \frac{2(d+2)}{d+3} \right) \left( T_{\alpha\alpha} T_{\beta\beta} T_{\alpha\beta}^* T_{\beta\alpha}^* - \delta_{\alpha\beta} T_{\alpha\alpha} T_{\beta\beta}^* \right) + c.c. \right]. \tag{142}$$

For $\alpha = \beta$, it is reduced to

$$\mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{\partial \epsilon_\alpha}{\partial \theta_\ell} \frac{\partial^2 \epsilon_\alpha}{\partial \theta_\ell^2} \frac{\partial \epsilon_\alpha}{\partial \theta_\ell} \right] = \mathbb{E}_{\mathcal{U}_{\rm RH}} \left[ \frac{1}{4(d^2 - 1)} \left( \frac{(d+2)^2}{d+3} |T_{\alpha\alpha}|^2 - \frac{2(d+2)}{d+3} \right) \left( |T_{\alpha\alpha}|^4 - |T_{\alpha\alpha}|^2 \right) \right] \tag{143}$$

$$= \frac{(d+2)(o_\alpha - 1) o_\alpha ((d+2) o_\alpha - 2)}{4(d^2 - 1)(d+3)}, \tag{144}$$

where we denote $o_\alpha = \epsilon_\alpha(\infty) + y_\alpha$ for simplicity. On the other hand, for $\alpha \neq \beta$, it is reduced to

$$\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\frac{\partial^2 \epsilon_\beta}{\partial \theta_\ell^2}\frac{\partial \epsilon_\alpha}{\partial \theta_\ell}\right] = \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{1}{8(d^2-1)}\left(\frac{(d+2)^2}{d+3}|T_{\alpha\alpha}|^2 - \frac{2(d+2)}{d+3}\right)T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha} + c.c.\right] \tag{145}$$

$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{1}{4(d^2-1)}\left(\frac{(d+2)^2}{d+3}|T_{\alpha\alpha}|^2 - \frac{2(d+2)}{d+3}\right)|T_{\alpha\alpha}||T_{\beta\beta}||T_{\alpha\beta}|T_{\beta\alpha}|\right]. \tag{146}$$

*Supplementary Note 7.2.2.* $\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}[\frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_1}}\frac{\partial^2 \epsilon_\alpha}{\partial \theta_{\ell_1}\partial \theta_{\ell_2}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_2}}]$ *under restricted Haar ensemble*

The other part $\sum_{l_1\neq l_2}\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}[\frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_1}}\frac{\partial^2 \epsilon_\alpha}{\partial \theta_{\ell_1}\partial \theta_{\ell_2}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_2}}] = \sum_{\ell_1<\ell_2}\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\partial^2 \epsilon_\alpha}{\partial \theta_{\ell_1}\partial \theta_{\ell_2}}\left(\frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_1}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_2}} + \frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_2}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_1}}\right)\right]$, and specifically for $\alpha = \beta$, it can be simplified to $2\sum_{\ell_1<\ell_2}\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}[\frac{\partial^2 \epsilon_\alpha}{\partial \theta_{\ell_1}\partial \theta_{\ell_2}}\frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_1}}\frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_2}}]$. $\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\partial^2 \epsilon_\alpha}{\partial \theta_{\ell_1}\partial \theta_{\ell_2}}\left(\frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_1}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_2}} + \frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_2}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_1}}\right)\right]$ becomes

$$\begin{aligned}
&\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\partial^2 \epsilon_\alpha}{\partial \theta_{\ell_1}\partial \theta_{\ell_2}}\left(\frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_1}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_2}} + \frac{\partial \epsilon_\alpha}{\partial \theta_{\ell_2}}\frac{\partial \epsilon_\beta}{\partial \theta_{\ell_1}}\right)\right] \\
&= \frac{1}{16}\int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}\left[X_{\ell_1}, U^\dagger_{\ell_1\to\ell_2}\left[X_{\ell_2}, O_{\alpha;\ell_2^+}\right]U_{\ell_1\to\ell_2}\right]U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}\left[X_{\ell_1}, O_{\alpha;\ell_1^+}\right]U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_2^-}\left[X_{\ell_2}, O_{\beta;\ell_2^+}\right]U_{\ell_2^-}\Big) \\
&+ \frac{1}{16}\int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}\left[X_{\ell_1}, U^\dagger_{\ell_1\to\ell_2}\left[X_{\ell_2}, O_{\alpha;\ell_2^+}\right]U_{\ell_1\to\ell_2}\right]U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_2^-}\left[X_{\ell_2}, O_{\alpha;\ell_2^+}\right]U_{\ell_2^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}\left[X_{\ell_1}, O_{\beta;\ell_1^+}\right]U_{\ell_1^-}\Big).
\end{aligned} \tag{147}$$

We calculate the above two separately. The first one becomes

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial^2\epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\frac{\partial\epsilon_\alpha}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_2}}\right]\\
&=\frac{1}{16}\int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}\left[X_{\ell_1},U^\dagger_{\ell_1\to\ell_2}\left[X_{\ell_2},O_{\alpha;\ell_2^+}\right]U_{\ell_1\to\ell_2}\right]U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}\left[X_{\ell_1},O_{\alpha;\ell_1^+}\right]U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_2^-}\left[X_{\ell_2},O_{\beta;\ell_2^+}\right]U_{\ell_2^-}\Big)\\
&=\frac{1}{16}\int_{\mathcal{U}_{\rm RH}}\mathrm{d}U_{\rm RH}\int_{\mathcal{U}_{\rm Haar}}\mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\Big[\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&\quad+\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&\quad+\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&\quad+\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\Big].
\end{aligned}
\tag{148}
$$

The first term is

$$
\begin{aligned}
I_1&\equiv\int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&=\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{dT_{\alpha\alpha}T_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\beta}T^*_{\beta\alpha}-|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right].
\end{aligned}
\tag{149}
$$

The second term is

$$
\begin{aligned}
I_2&\equiv\int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)\\
&=\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\delta_{\alpha\beta}\frac{d|T_{\alpha\alpha}|^2T^*_{\beta\beta}\left(dT_{\alpha\beta}-T_{\alpha\alpha}\right)}{\left(d^2-1\right)^2}+\frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)}{\left(d^2-1\right)^2}\right].
\end{aligned}
\tag{150}
$$

The third term is

$$
\begin{aligned}
I_3&\equiv\int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)\\
&=\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{|T_{\alpha\alpha}|^2\left(|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)-dT_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}\right)}{\left(d^2-1\right)^2}+\delta_{\alpha\beta}\frac{d^2T_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\beta}}{\left(d^2-1\right)^2}\right].
\end{aligned}
\tag{151}
$$

The forth term is

$$I_4 \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\alpha\alpha}|T_{\alpha\alpha}|^2T^*_{\beta\beta}}{(d-1)(d+1)^2} - \frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right]. \tag{152}$$

The fifth term is

$$I_5 \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}|T_{\alpha\alpha}|^2\left(dT^*_{\alpha\beta}-T^*_{\alpha\alpha}\right)}{\left(d^2-1\right)^2} + \frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)}{\left(d^2-1\right)^2}\right] = I_2^*. \tag{153}$$

The sixth term is

$$I_6 \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{dT_{\alpha\beta}T_{\beta\alpha}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}T^*_{\beta\beta} - |T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_1^*. \tag{154}$$

The seventh term is

$$I_7 \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}}{(d-1)(d+1)^2} - \frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_4^*. \tag{155}$$

The eighth term is

$$I_8 \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{|T_{\alpha\alpha}|^2\left(|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)-dT_{\alpha\beta}T_{\beta\alpha}T^*_{\alpha\alpha}T^*_{\beta\beta}\right)}{\left(d^2-1\right)^2} + \delta_{\alpha\beta}\frac{d^2T_{\alpha\beta}|T_{\alpha\alpha}|^2T^*_{\beta\beta}}{\left(d^2-1\right)^2}\right] = I_3^*. \tag{156}$$

The ninth term is

$$I_9 \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}T^*_{\alpha\alpha}\left(d-|T_{\alpha\alpha}|^2\right)}{\left(d^2-1\right)^2} - \frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(d-|T_{\alpha\alpha}|^2\right)}{\left(d^2-1\right)^2}\right]. \tag{157}$$

The tenth term is

$$I_{10} \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{dT_{\alpha\beta}T_{\beta\alpha}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}T^*_{\beta\beta} - |T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_1^*. \tag{158}$$

The eleventh term is

$$I_{11} \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1\to\ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}}{(d-1)(d+1)^2} - \frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_4^*. \tag{159}$$

The twelfth term is

$$I_{12} \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1 \nrightarrow \ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}\Big)$$

$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\left(d-|T_{\alpha\alpha}|^2\right)\left(dT_{\alpha\beta}T_{\beta\alpha}T^*_{\alpha\alpha}T^*_{\beta\beta}-|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\right)}{\left(d^2-1\right)^2}\right]. \tag{160}$$

The thirteenth term is

$$I_{13} \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1 \nrightarrow \ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)$$

$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{dT_{\alpha\alpha}T_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\beta}T^*_{\beta\alpha}-|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_1. \tag{161}$$

The fourteenth term is

$$I_{14} \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1 \nrightarrow \ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}\Big)$$

$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\alpha\alpha}T^*_{\beta\beta}\left(d-|T_{\alpha\alpha}|^2\right)}{\left(d^2-1\right)^2}-\frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(d-|T_{\alpha\alpha}|^2\right)}{\left(d^2-1\right)^2}\right] = I_9^*. \tag{162}$$

The fifteenth term is

$$I_{15} \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1 \nrightarrow \ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\beta;U}\Big)$$

$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\left(d-|T_{\alpha\alpha}|^2\right)\left(dT_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}-|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\right)}{\left(d^2-1\right)^2}\right] = I_{12}^*. \tag{163}$$

The sixteenth term is

$$I_{16} \equiv \int \mathrm{d}U_{\ell_1^-}\,\mathrm{d}U_{\ell_1 \nrightarrow \ell_2}\,\mathrm{d}U_{\ell_2^+}\,\mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}\Big)$$

$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\alpha\alpha}|T_{\alpha\alpha}|^2T^*_{\beta\beta}}{(d-1)(d+1)^2}-\frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_4. \tag{164}$$

Finally we have

$$\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\partial^2\epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\frac{\partial\epsilon_\alpha}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_2}}\right] = \frac{1}{16}\sum_{i=1}^{4}\left(I_{4i+1}+I_{4i+2}-I_{4i+3}-I_{4i+4}\right)$$

$$= \frac{1}{16}\left(2I_1+I_2-I_3-2I_4+I_9-I_{12}+c.c.\right) \tag{165}$$

$$+\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{d^2(|T_{\alpha\alpha}|^2-1)T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}}{(d^2-1)^2}+c.c.\right]+\delta_{\alpha\beta}\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{d^2T_{\alpha\alpha}T^*_{\beta\beta}(1-|T_{\alpha\alpha}|^2)}{(d^2-1)^2}+c.c.\right]\Bigg) \tag{166}$$

$$= \frac{1}{16}\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{d^2(2|T_{\alpha\alpha}|^2-1)\left(T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}-\delta_{\alpha\beta}T_{\alpha\alpha}T^*_{\beta\beta}\right)}{(d^2-1)^2}+c.c.\right]. \tag{167}$$

Specifically, for $\alpha=\beta$, we have

$$\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\partial^2\epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\frac{\partial\epsilon_\alpha}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_\alpha}{\partial\theta_{\ell_2}}\right] = \frac{d^2o_\alpha(o_\alpha-1)(2o_\alpha-1)}{8(d^2-1)^2}. \tag{168}$$

On the other hand, for $\alpha\neq\beta$, we have

$$\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\partial^2\epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\frac{\partial\epsilon_\alpha}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_2}}\right] = \frac{1}{16}\mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{d^2(2|T_{\alpha\alpha}|^2-1)T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}}{(d^2-1)^2}+c.c.\right] \tag{169}$$

$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{d^2(2|T_{\alpha\alpha}|^2-1)|T_{\alpha\alpha}||T_{\beta\beta}||T_{\alpha\beta}||T_{\beta\alpha}|}{8(d^2-1)^2}\right]. \tag{170}$$

Similarly, we next work on

$$\begin{aligned}
&\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial^2\epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\frac{\partial\epsilon_\alpha}{\partial\theta_{\ell_2}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_1}}\right]\\
&=\frac{1}{16}\int {\rm d}U_{\ell_1^-}\,{\rm d}U_{\ell_1\to\ell_2}\,{\rm d}U_{\ell_2^+}\,{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}\left[X_{\ell_1},U^\dagger_{\ell_1\to\ell_2}\left[X_{\ell_2},O_{\alpha;\ell_2^+}\right]U_{\ell_1\to\ell_2}\right]U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_2^-}\left[X_{\ell_2},O_{\alpha;\ell_2^+}\right]U_{\ell_2^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}\left[X_{\ell_1},O_{\beta;\ell_1^+}\right]U_{\ell_1^-}\Big)\\
&=\frac{1}{16}\int_{\mathcal{U}_{\rm RH}}{\rm d}U_{\rm RH}\int_{\mathcal{U}_{\rm Haar}}{\rm d}U_{\ell_1^-}\,{\rm d}U_{\ell_1\to\ell_2}\Big[{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&\quad+{\rm tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+{\rm tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&\quad+{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&\quad+{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad+{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&\quad-{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big).
\end{aligned}\tag{171}$$

The first is

$$\begin{aligned}
I_1&\equiv{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&=\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{dT_{\alpha\alpha}T_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\beta}T^*_{\beta\alpha}-|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right].
\end{aligned}\tag{172}$$

The second is

$$\begin{aligned}
I_2&\equiv{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)\\
&=\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\delta_{\alpha\beta}\frac{d|T_{\alpha\alpha}|^2T^*_{\beta\beta}(dT_{\beta\alpha}-T_{\alpha\alpha})}{(d^2-1)^2}+\frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)}{(d^2-1)^2}\right].
\end{aligned}\tag{173}$$

The third is

$$\begin{aligned}
I_3&\equiv{\rm tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)\\
&=\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{|T_{\alpha\alpha}|^2\left(|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)-dT_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}\right)}{(d^2-1)^2}+\delta_{\alpha\beta}\frac{d^2T_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\beta\alpha}}{(d^2-1)^2}\right].
\end{aligned}\tag{174}$$

The forth is

$$I_4 \equiv \mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\alpha\alpha}|T_{\alpha\alpha}|^2T^*_{\beta\beta}}{(d-1)(d+1)^2} - \frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right]. \tag{175}$$

The fifth is

$$I_5 \equiv \mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}|T_{\alpha\alpha}|^2\left(dT^*_{\beta\alpha}-T^*_{\alpha\alpha}\right)}{(d^2-1)^2} + \frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)}{(d^2-1)^2}\right] = I_2^*. \tag{176}$$

The sixth is

$$I_6 \equiv \mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{dT_{\alpha\beta}T_{\beta\alpha}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}T^*_{\beta\beta} - |T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_1^*. \tag{177}$$

The seventh is

$$I_7 \equiv \mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}}{(d-1)(d+1)^2} - \frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_4^*. \tag{178}$$

The eighth is

$$I_8 \equiv \mathrm{tr}\Big(P_{\beta\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{|T_{\alpha\alpha}|^2\left(|T_{\beta\beta}|^2\left(|T_{\alpha\alpha}|^2-d\right)-dT_{\alpha\beta}T_{\beta\alpha}T^*_{\alpha\alpha}T^*_{\beta\beta}\right)}{(d^2-1)^2} + \delta_{\alpha\beta}\frac{d^2T_{\beta\alpha}|T_{\alpha\alpha}|^2T^*_{\beta\beta}}{(d^2-1)^2}\right] = I_3^*. \tag{179}$$

The ninth is

$$I_9 \equiv \mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{dT_{\alpha\alpha}T_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\beta}T^*_{\beta\alpha} - |T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_1. \tag{180}$$

The tenth is

$$I_{10} \equiv \mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\delta_{\alpha\beta}\frac{dT_{\alpha\alpha}T^*_{\beta\beta}\left(d-|T_{\alpha\alpha}|^2\right)}{(d^2-1)^2} - \frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(d-|T_{\alpha\alpha}|^2\right)}{(d^2-1)^2}\right]. \tag{181}$$

The eleventh is

$$I_{11} \equiv \mathrm{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\nrightarrow\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{\left(d-|T_{\alpha\alpha}|^2\right)\left(dT_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}-|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\right)}{(d^2-1)^2}\right]. \tag{182}$$

The twelfth is

$$I_{12} \equiv \operatorname{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\delta_{\alpha\beta}\frac{dT_{\alpha\alpha}|T_{\alpha\alpha}|^2T^*_{\beta\beta}}{(d-1)(d+1)^2} - \frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I_4. \tag{183}$$

The thirteenth is

$$I_{13} \equiv \operatorname{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}T^*_{\alpha\alpha}\left(d-|T_{\alpha\alpha}|^2\right)}{\left(d^2-1\right)^2} - \frac{|T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\left(d-|T_{\alpha\alpha}|^2\right)}{\left(d^2-1\right)^2}\right] = I^*_{10}. \tag{184}$$

The fourteenth is

$$I_{14} \equiv \operatorname{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{dT_{\alpha\beta}T_{\beta\alpha}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}T^*_{\beta\beta} - |T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I^*_1. \tag{185}$$

The fifteenth is

$$I_{15} \equiv \operatorname{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}P_{\alpha\beta}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}O_{\beta;U}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\delta_{\alpha\beta}\frac{dT_{\beta\beta}|T_{\alpha\alpha}|^2T^*_{\alpha\alpha}}{(d-1)(d+1)^2} - \frac{|T_{\alpha\alpha}|^4|T_{\beta\beta}|^2}{(d-1)(d+1)^2}\right] = I^*_4. \tag{186}$$

The sixteenth (last) is

$$I_{16} \equiv \operatorname{tr}\Big(P_{\beta\alpha}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}P_{\alpha\alpha}O_{\alpha;U}U^\dagger_{\ell_1^-}X_{\ell_2,\ell_1\to\ell_2}U_{\ell_1^-}P_{\alpha\beta}O_{\beta;U}U^\dagger_{\ell_1^-}X_{\ell_1}U_{\ell_1^-}\Big)$$
$$= \mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\left(d-|T_{\alpha\alpha}|^2\right)\left(dT_{\alpha\beta}T_{\beta\alpha}T^*_{\alpha\alpha}T^*_{\beta\beta} - |T_{\alpha\alpha}|^2|T_{\beta\beta}|^2\right)}{\left(d^2-1\right)^2}\right] = I^*_{11}. \tag{187}$$

Therefore, we have

$$\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial^2\epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\frac{\partial\epsilon_\alpha}{\partial\theta_{\ell_2}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_1}}\right] = \frac{1}{16}\sum_{i=1}^4\left(I_{4i+1}+I_{4i+2}-I_{4i+3}-I_{4i+4}\right)$$
$$= \frac{1}{16}\left(2I_1+I_2-I_3-2I_4+I_{10}-I_{11}+c.c.\right) \tag{188}$$
$$= \frac{1}{16}\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{d^2(2|T_{\alpha\alpha}|^2-1)\left(T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha} - \delta_{\alpha\beta}T_{\alpha\alpha}T^*_{\beta\beta}\right)}{\left(d^2-1\right)^2} + c.c.\right], \tag{189}$$

which matches Eq. (167).

*Supplementary Note 7.2.3. Summary*

Combining Eq. (142), Eq. (167) and (189), we finally have

$$\overline{\mu_{\alpha\alpha\beta}(\infty)} = L\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial\epsilon_\gamma}{\partial\theta_\ell}\frac{\partial^2\epsilon_\alpha}{\partial\theta_\ell^2}\right] + L(L-1)\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{\partial\epsilon_\gamma}{\partial\theta_{\ell_1}}\frac{\partial^2\epsilon_\alpha}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}}\frac{\partial\epsilon_\beta}{\partial\theta_{\ell_2}}\right] \tag{190}$$
$$= L\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{d+2}{8(d^2-1)(d+3)}\left((d+2)|T_{\alpha\alpha}|^2-2\right)\left(T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha} - \delta_{\alpha\beta}T_{\alpha\alpha}T^*_{\beta\beta}\right) + c.c.\right]$$
$$+ L(L-1)\mathbb{E}_{\mathcal{U}_{\rm RH}}\left[\frac{d^2(2|T_{\alpha\alpha}|^2-1)\left(T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha} - \delta_{\alpha\beta}T_{\alpha\alpha}T^*_{\beta\beta}\right)}{16\left(d^2-1\right)^2} + c.c.\right]. \tag{191}$$

For $\alpha = \beta$, it can be simplified to

$$\begin{aligned}
&\overline{\mu_{\alpha\alpha\alpha}(\infty)} \\
&= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{L(d+2)}{4(d^2-1)(d+3)}\left((d+2)|T_{\alpha\alpha}|^2-2\right)\left(|T_{\alpha\alpha}|^4-|T_{\alpha\alpha}|^2\right)+L(L-1)\frac{d^2(2|T_{\alpha\alpha}|^2-1)\left(|T_{\alpha\alpha}|^4-|T_{\alpha\alpha}|^2\right)}{8\left(d^2-1\right)^2}\right] \quad (192)\\
&= \frac{Lo_\alpha(o_\alpha-1)}{8(d^2-1)}\left[\frac{2(d+2)}{d+3}\left((d+2)o_\alpha-2\right)+\frac{(L-1)d^2(2o_\alpha-1)}{d^2-1}\right], \quad (193)
\end{aligned}$$

and for $\alpha \neq \beta$, it becomes

$$\begin{aligned}
&\overline{\mu_{\alpha\alpha\beta}(\infty)} \\
&= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{L(d+2)}{8(d^2-1)(d+3)}\left((d+2)|T_{\alpha\alpha}|^2-2\right)T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}+\frac{L(L-1)d^2(2|T_{\alpha\alpha}|^2-1)T_{\alpha\alpha}T_{\beta\beta}T^*_{\alpha\beta}T^*_{\beta\alpha}}{16\left(d^2-1\right)^2}+c.c.\right] \quad (194)\\
&= \mathbb{E}_{\mathcal{U}_{\mathrm{RH}}}\left[\frac{L(d+2)}{4(d^2-1)(d+3)}\left((d+2)|T_{\alpha\alpha}|^2-2\right)|T_{\alpha\alpha}||T_{\beta\beta}||T_{\alpha\beta}||T_{\beta\alpha}|+\frac{L(L-1)d^2(2|T_{\alpha\alpha}|^2-1)|T_{\alpha\alpha}||T_{\beta\beta}||T_{\alpha\beta}||T_{\beta\alpha}|}{8\left(d^2-1\right)^2}\right] \quad (195)
\end{aligned}$$

With one more step, we can obtain the ensemble average relative dQNTK as

$$\overline{\lambda_{\alpha\alpha\alpha}(\infty)} = \frac{\overline{\mu_{\alpha\alpha\alpha}(\infty)}}{\overline{K_{\alpha\alpha}(\infty)}} = -\frac{1}{4d}\left[\frac{2(d+2)}{d+3}\left((d+2)o_\alpha-2\right)+\frac{(L-1)d^2(2o_\alpha-1)}{d^2-1}\right] \simeq -\frac{1}{4d}\left[2(do_\alpha-2)+L(2o_\alpha-1)\right], \quad (196)$$

$$(197)$$

where we approximate it with $L, d \gg 1$ at the end. The off-diagonal part $\overline{\lambda_{\alpha\alpha\beta}(\infty)} = \frac{\overline{\mu_{\alpha\alpha\beta}(\infty)}}{\sqrt{\overline{K_{\alpha\alpha}(\infty)}\,\overline{K_{\beta\beta}(\infty)}}}$ can be found from Eq. (195) and (120).


[1] B. Zhang, J. Liu, X.-C. Wu, L. Jiang, and Q. Zhuang, Dynamical phase transition in quantum neural networks with large depth, arXiv:2311.18144 (2023).
[2] D. A. Roberts and B. Yoshida, Chaos and complexity by design, Journal of High Energy Physics **2017**, 1 (2017).
[3] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Physical Review A **98**, 032309 (2018).
[4] G. E. Crooks, Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition, arXiv preprint arXiv:1905.13311 (2019).
[5] J. Liu, K. Najafi, K. Sharma, F. Tacchino, L. Jiang, and A. Mezzacapo, Analytic theory for the dynamics of wide quantum neural networks, Phys. Rev. Lett. **130**, 150601 (2023).
[6] M. Fukuda, R. König, and I. Nechita, Rtni—a symbolic integrator for haar-random tensor networks, J. Phys. A: Math. Theor. **52**, 425303 (2019).