Using the equation which defines $\mathcal{R}_1$, we find that the variable r satisfies the following relationship at $\omega \to 0$

$$\alpha = \frac{1}{M}\sum_k \frac{\lambda_k r}{1+\lambda_k r} \tag{70}$$

After solving this implicit equation, we can find the limiting value of $i\omega\mathcal{H}_k(\omega)$ as

$$H_k^\infty = \lim_{\tau\to\infty} H_k(\tau) = \lim_{\omega\to 0} i\omega\mathcal{H}_k(\omega) = \frac{1}{1+\lambda_k r} \tag{71}$$

Next, we can work out the scaling of the correlation functions in the limit of low frequency. We define the following limiting quantities based on a scaling analysis performed on our correlation functions for small ω

$$\begin{aligned}
C_0^\infty &\equiv \lim_{t,s\to\infty} C_0(t,s) = \lim_{\omega,\omega'\to 0}(i\omega)(i\omega')\mathcal{C}_0(\omega,\omega') \\
C_1^\infty &\equiv \int_0^\infty\int_0^\infty dt' ds' C_1(t',s') = \lim_{\omega,\omega'\to 0}\mathcal{C}_1(\omega,\omega') \\
C_2^\infty &\equiv \int_0^\infty\int_0^\infty dt' ds' C_2(t',s') = \lim_{\omega,\omega'\to 0}\mathcal{C}_2(\omega,\omega') \\
C_3^\infty &\equiv \int_0^\infty\int_0^\infty dt' ds' C_3(t',s') = \lim_{\omega,\omega'\to 0}\mathcal{C}_3(\omega,\omega')
\end{aligned} \tag{72}$$

These limiting quantities satisfy the closed set of linear equations

$$\begin{aligned}
C_0^\infty &= \frac{1}{M}\sum_k \lambda_k (H_k^\infty)^2\left[(w_k^\star)^2 + \frac{1}{\nu}C_3^\infty + \frac{1}{\alpha}\lambda_k\left(1-\frac{\alpha}{\nu}\right)^2 C_1^\infty\right] \\
C_1^\infty &= \frac{r^2}{(1-\frac{\alpha}{\nu})^2}C_0^\infty \\
C_2^\infty &= \frac{1}{\alpha M}\sum_k \lambda_k (H_k^\infty)^2 C_1^\infty + \frac{r^2}{(1-\frac{\alpha}{\nu})^2}\frac{1}{M}\sum_k \lambda_k^2 (H_k^\infty)^2[(w_k^\star)^2 + \nu^{-1}C_3^\infty] \\
C_3^\infty &= \left(1-\frac{\alpha}{\nu}\right)^2 C_2^\infty
\end{aligned} \tag{73}$$

These equations can be solved for $\{C_0^\infty, C_1^\infty, C_2^\infty, C_3^\infty\}$. Simplifying the expressions to a two-variable system, we find

$$\begin{aligned}
C_0^\infty &= \frac{1}{M}\sum_k \lambda_k (H_k^\infty)^2\left[(w_k^\star)^2 + \frac{1}{\nu}\left(1-\frac{\alpha}{\nu}\right)^2 C_2^\infty + \frac{1}{\alpha}\lambda_k r^2 C_0^\infty\right] \\
C_2^\infty &= \frac{r^2}{\alpha(1-\alpha/\nu)^2 M}\sum_k \lambda_k (H_k^\infty)^2 C_0^\infty + \frac{r^2}{(1-\frac{\alpha}{\nu})^2}\frac{1}{M}\sum_k \lambda_k^2 (H_k^\infty)^2\left[(w_k^\star)^2 + \frac{1}{\nu}\left(1-\frac{\alpha}{\nu}\right)^2 C_2^\infty\right]
\end{aligned}$$

This expression recovers the ridgeless limit of the replica results of (Atanasov et al., 2023; Zavatone-Veth & Pehlevan, 2023) and the random matrix analysis of (Simon et al., 2023).

Under-parameterized Case $\nu < \alpha$: Following the same procedure, we note that for $\nu < \alpha$ that $\mathcal{R}_3 \sim \mathcal{O}(i\omega)$ and $\mathcal{R}_1 \sim 1 - \frac{\nu}{\alpha}$. We thus find the following equation for $r = \lim_{\omega\to\infty}(i\omega)^{-1}\mathcal{R}_1(\omega)\mathcal{R}_3(\omega)$.

$$\nu = \frac{1}{M}\sum_k \frac{\lambda_k r}{\lambda_k r + 1} \tag{74}$$

where as before $H_k^\infty = \frac{1}{1+\lambda_k r}$. The analogous scaling argument for small ω gives us the following set of well-defined limiting quantities

$$\begin{aligned}
C_0^\infty &\equiv \lim_{t,s\to\infty} C_0(t,s) = \lim_{\omega,\omega'\to 0}(i\omega)(i\omega')\mathcal{C}(\omega,\omega') \\
C_1^\infty &\equiv \lim_{t,s\to\infty} C_1(t,s) = \lim_{\omega,\omega'\to 0}(i\omega)(i\omega')\mathcal{C}_1(\omega,\omega') \\
C_2^\infty &\equiv \lim_{t,s\to\infty} C_2(t,s) = \lim_{\omega,\omega'\to 0}(i\omega)(i\omega')\mathcal{C}_2(\omega,\omega') \\
C_3^\infty &\equiv \int_0^\infty\int_0^\infty dtds\, C_3(t,s) = \lim_{\omega,\omega'\to 0}\mathcal{C}_3(\omega,\omega').
\end{aligned} \tag{75}$$

where these limiting correlation values satisfy

$$
\begin{aligned}
C_0^\infty &= \frac{1}{M}\sum_k \lambda_k (H_k^\infty)^2 \left[(w_k^\star)^2 + \frac{1}{\nu} C_3^\infty + \frac{1}{\alpha}\lambda_k \frac{r^2}{(1-\frac{\nu}{\alpha})^2} C_1^\infty \right] \\
C_1^\infty &= \left(1 - \frac{\nu}{\alpha}\right)^2 C_0^\infty \\
C_2^\infty &= \frac{1}{\alpha M}\sum_k \lambda_k (H_k^\infty)^2 C_1^\infty + \frac{1}{M}\left(1-\frac{\nu}{\alpha}\right)^2 \sum_k \lambda_k^2 (H_k^\infty)^2 \left[(w_k^\star)^2 + \frac{1}{\nu} C_3^\infty \right] \\
C_3^\infty &= \frac{r^2}{(1-\frac{\nu}{\alpha})^2} C_2^\infty .
\end{aligned}
\tag{76}
$$

This is again a closed linear system of equations for the variables $\{C_0^\infty, C_1^\infty, C_2^\infty, C_3^\infty\}$. In the next section, we recover the result for kernel regression where $\nu \to \infty$ and the learning curve for infinite data $\alpha \to \infty$ with respect to model size ν.

C.2. Learning Curves for Kernel Regression $\nu, t \to \infty$

In the $t \to \infty$ and $\nu \to \infty$ limit we recover the learning curve for kernel regression with eigenvalues λ_k. To match the notation of (Canatar et al., 2021), we define

$$
\lim_{\omega \to 0} (i\omega)^{-1} \mathcal{R}_1(\omega) \equiv \alpha \kappa^{-1} \tag{77}
$$

which generates the following self-consistent equation for κ

$$
1 = \frac{1}{M}\sum_k \frac{\lambda_k}{\lambda_k \alpha + \kappa}. \tag{78}
$$

Plugging this into the expression for the loss, we find

$$
\begin{aligned}
&(i\omega)(i\omega')\mathcal{C}_0(\omega,\omega') \sim \frac{1}{M}\sum_k \lambda_k \frac{\kappa^2}{(\kappa+\lambda_k\alpha)^2} [(w_k^\star)^2 + \alpha^{-1}\lambda_k \mathcal{C}_1(\omega,\omega')] \\
&\mathcal{C}_1(\omega,\omega') = (i\omega)(i\omega')\alpha^2\kappa^{-2}\mathcal{C}_0(\omega,\omega')
\end{aligned}
\tag{79}
$$

Letting $C_\infty \equiv \lim_{\omega,\omega'\to 0} (i\omega)(i\omega')\mathcal{C}(s,s')$, we have

$$
\begin{aligned}
C_\infty &= \frac{1}{M}\sum_k \lambda_k \frac{\kappa^2}{(\kappa+\lambda_k\alpha)^2}(w_k^\star)^2 + C_\infty \frac{\alpha}{M}\sum_k \frac{\lambda_k^2}{(\lambda_k\alpha+\kappa)^2} \\
&= \frac{1}{1-\gamma}\sum_k \lambda_k (w_k^\star)^2 \frac{\kappa^2}{(\kappa+\lambda_k\alpha)^2} \ , \quad \gamma = \frac{\alpha}{M}\sum_k \frac{\lambda_k^2}{(\lambda_k\alpha+\kappa)^2}
\end{aligned}
\tag{80}
$$

The variable κ decreases from $\left[\frac{1}{M}\sum_k \lambda_k, 0\right]$ as $\alpha \in [0,1]$. For $\alpha > 1$ we have $\kappa = 0$. The quantity $\frac{1}{1-\gamma}$ comes from overfitting due to variance from the randomly sampled dataset.

D. Early Time Dynamics (High-Frequency Range)

In this section, we explore the early time dynamical effects of this model. Similar to how the late time dynamical effects could be measured by examining the low frequency $\omega \ll 1$ part of the response and correlation functions, in this section, we analyze the high frequency components $\omega \gg 1$. We start by noting the following expansions valid near $\omega \to \infty$

$$
\begin{aligned}
\mathcal{R}_1(\omega) &\sim 1 - \frac{1}{\alpha(i\omega)}\left[\frac{1}{M}\sum_k \lambda_k\right] + \mathcal{O}(\omega^{-2}) \\
\mathcal{R}_3(\omega) &\sim 1 - \frac{1}{\nu(i\omega)}\left[\frac{1}{M}\sum_k \lambda_k\right] + \mathcal{O}(\omega^{-2})
\end{aligned}
\tag{81}
$$

We let $c = \frac{1}{M}\sum_k \lambda_k$. These can be plugged into the transfer function for mode k

$$\mathcal{H}_k(\omega) \sim \frac{1}{i\omega + \lambda_k - c(\alpha^{-1}+\nu^{-1})(i\omega)^{-1}} \sim \frac{1}{i\omega+\lambda_k}\left[1 + \frac{c\lambda_k(\alpha^{-1}+\nu^{-1})}{i\omega(i\omega+\lambda_k)}\right] + \mathcal{O}\left(\omega^{-2}\right) \tag{82}$$

Performing an inverse Fourier transform, we find the following early time asymptotics

$$\begin{aligned} H_k(t) &\sim e^{-\lambda_k t} + c\lambda_k(\alpha^{-1}+\nu^{-1}) \int \frac{d\omega}{2\pi} \frac{e^{i\omega t}}{i\omega(i\omega+\lambda_k)^2} \\ &= e^{-\lambda_k t} - c\lambda_k(\alpha^{-1}+\nu^{-1}) \frac{\partial}{\partial \lambda_k} \int \frac{d\omega}{2\pi}\frac{e^{i\omega t}}{i\omega(i\omega+\lambda_k)} \\ &= e^{-\lambda_k t} - c\lambda_k(\alpha^{-1}+\nu^{-1}) \frac{\partial}{\partial\lambda_k}\left[\frac{1}{\lambda_k} - \frac{1}{\lambda_k}e^{-\lambda_k t}\right] \\ &= e^{-\lambda_k t} + \frac{c(\alpha^{-1}+\nu^{-1})}{\lambda_k}\left[1 - e^{-\lambda_k t} - \lambda_k t\, e^{-\lambda_k t}\right] \end{aligned} \tag{83}$$

We see from this expression that the early time corrections always scale as $1/\alpha$ or $1/\nu$ and that these corrections build up over time. We also note that in this picture, $H_k(t)$ is minimized in the limit of large model and large data $\alpha, \nu \to \infty$ (limited data and limited model size strictly harm performance). A similar expansion can be performed for all of the correlation functions $\mathcal{C}(\omega, \omega')$ with $\omega, \omega' \gg 1$ which also give leading corrections which scale as $1/\alpha$ and $1/\nu$.

E. Buildup of Overfitting Effects

In this section, we derive a formula for the gap between test loss $\mathcal{L}(t)$ and train loss $\hat{\mathcal{L}}(t)$. We start from the following formula

$$v_1(t) = u_1(t) + \frac{1}{\alpha}\int_0^t ds R_{0,2}(t,s) v_1(s) \tag{84}$$

Moving the $v_1(t)$ term to the other side, and using the fact that $\langle u_1(t) u_1(s)\rangle = C_1(t,s)$, we find the following relationship between train and test loss

$$\begin{aligned} \mathcal{L}(t) = \langle u_1(t)u_1(t)\rangle &= \langle v_1(t) v_1(t)\rangle - \frac{2}{\alpha}\int_0^t dt' R_{0,2}(t,t')\left\langle v_1(t)v_1(t')\right\rangle \\ &+ \frac{1}{\alpha^2}\int_0^t dt' \int_0^t ds' R_{0,2}(t,t')R_{0,2}(t,s')\left\langle v_1(t')v_1(s')\right\rangle \\ &= \hat{\mathcal{L}}(t) - \frac{2}{\alpha}\int_0^t dt' R_{0,2}(t,t') C_1(t,t') + \frac{1}{\alpha^2}\int_0^t dt' \int_0^t ds' R_{0,2}(t,t')R_{0,2}(t,s')C_1(t',s'). \end{aligned} \tag{85}$$

To get a sense of these expressions at early and late timescales, we investigate the Fourier transforms at high $\omega \gg 1$ and low $\omega \ll 1$ frequencies respectively.

E.1. High Frequency Range / Early Time

The relationship between Fourier transforms at high frequencies $\omega \gg 1$ is

$$\mathcal{C}_0(\omega,\omega') = \frac{1}{\mathcal{R}_1(\omega)\mathcal{R}_1(\omega')}\,\mathcal{C}_1(\omega,\omega') \sim \mathcal{C}_1(\omega,\omega') + \frac{c}{\alpha(i\omega')}\mathcal{C}_1(\omega,\omega') + \frac{c}{\alpha(i\omega')}\mathcal{C}_1(\omega,\omega') + \mathcal{O}((i\omega)^{-2} + (i\omega')^{-2}) \tag{86}$$

where $c = \frac{1}{M}\sum_k \lambda_k$. Taking a Fourier transform back to real time gives us the following early time differential equation for the test-loss train loss gap

$$\partial_{ts}^2 [C_0(t,s) - C_1(t,s)] = \partial^2_{ts} C_1(t,s) + \frac{c}{\alpha}(\partial_t + \partial_s) C_1(t,s). \tag{87}$$

The above equation should hold for early times. We note that $C_0(t,t) - C_1(t,t) = \mathcal{L}(t) - \hat{\mathcal{L}}(t)$ exactly recovers the test-train gap.