S1 Simplification of the first-layer kernel

In this section, we get explicit control in spectral norm of the difference between the empirical (i.e. finite-size) NTK and the version in eqn. (22) that arises through the simplification of the first-layer kernel $K_1$ in eqn. (12). We will use the notation $A_i = (A_{i1}, \ldots, A_{in})$, where $A_{ij}$ is defined similarly. Recall from eqns. (8) and (9) that the empirical NTK is given by

$$K := \frac{X^TX}{n_0} \odot \frac{(F')^T \text{diag}(W_2)^2 F'}{n_1} + \frac{F^TF}{n_1} + \gamma I \tag{S1}$$

and from eqn. (22) the simplified kernel is given by

$$K_{\text{simpl}} := \zeta \frac{X^TX}{n_0} + (\eta' - \zeta)I + \frac{F^TF}{n_1} + \gamma I, \tag{S2}$$

Also, define

$$R := \zeta' \frac{11^T}{n_0}, \tag{S3}$$

where $\zeta' := [\mathbb{E}_{z \sim \mathcal{N}(0,1)} \sigma''(z)]^2$. In this section, we show for any $\varepsilon, \delta > 0$

$$\mathbb{P} \left\{ \|K - K_{\text{simpl}} - R\| > n_0^{2\varepsilon-1/4} \right\} < \delta \tag{S4}$$

for sufficiently large $n_0$.

Let $\mathbb{E}$ be expectation over $W_1$ and $W_2$ conditional on $X$. We note that with high-probability for any $\varepsilon > 0$ that $(X^TX/n_0)_{ab} = \delta_{ab} + O \left( n_0^{\varepsilon-1/2} \right)$ for all $a$ and $b$, and that $\|X^TX/n_0\| \leq n^\varepsilon$, since $X$ is i.i.d. Gaussian. The use of $O$ hides uniform constants.
Define
\[ \Delta_k := \frac{X^T X}{n_0} \odot \left( (W_2)_k^2 (F_{k_0}^l)^T F_{k_0}^l - \bar{M} \right) \quad \text{and} \quad \Delta := \frac{1}{n_1} \sum_{k=1}^{n_1} \Delta_k, \] (S5)
where \( \bar{M} := \mathbb{E}(F_{k_0}^l)^T F_{k_0}^l \) (which does not depend on \( k \)), so \( \mathbb{E}\Delta = 0 \). Then
\[ K - K_{\text{simp}} - R = \Delta + \left[ \frac{X^T X}{n_0} \odot (M - \bar{M}) - R \right] + \left( \frac{X^T X}{n_0} \odot M - K_{\text{simp}} \right), \] (S6)
where \( M := \zeta \mathbb{1}^T + (\eta' - \zeta) I \). Elementary arguments given in Sec. S1.2 show that, in operator norm, the two rightmost terms in eqn. (S6) are bounded by \( O(n_0^{3\varepsilon-1/2}) \). In Sec. S1.1, we bound \( \Delta \) by using the fact that, conditional on \( X \), \( \Delta \) is a sum of independent random matrices to apply the matrix Bernstein inequality (Tropp, 2015).

S1.1. Bounding \( \Delta \)

We start with a supremum bound on \( \|\Delta_k\| \). For any vector \( \mathbf{v} = \sum_k v_k \mathbf{e}_k \), we have
\[ \|\Delta_k \mathbf{v}\| \leq \sum_k |v_k| \|\Delta_k \mathbf{e}_k\| \leq n_1^{-1} \sup_{a,b} \left| (W_2)_a^2 F_{ka}^l F_{kb}^l - \bar{M} \cdot \|X^T X/n_0\| \sqrt{m} \right|, \] (S7)
by the Cauchy-Schwarz inequality. Note that by assumption on \( X \), eqn. (S7) is \( O \left( n_0^{2\varepsilon m^{1/2}/n_1} \right) = O(n_0^{3\varepsilon-1/2}) \).

Now we bound the variance term. Consider the \((a, b)\) entry of \( \mathbb{E}\Delta_k^2 \):
\[ \frac{1}{n_1^2} \sum_{t=1}^{m} \left( X^T X/n_0 \right)_{at} \left( X^T X/n_0 \right)_{bt} \mathbb{E} \left[ \left( (W_2)_a^2 F_{ka}^l F_{kb}^l - \bar{M}_{at} \right) \left( (W_2)_b^2 F_{kt}^l F_{kb}^l - \bar{M}_{tb} \right) \right] \]
\[ = \frac{1}{n_1^2} \sum_{t=1}^{m} \left( X^T X/n_0 \right)_{at} \left( X^T X/n_0 \right)_{bt} \left( 3\mathbb{E}[F_{ka}^l F_{kb}^l F_{ka}^l F_{kb}^l] - \mathbb{E}[F_{ka}^l F_{kb}^l] \mathbb{E}[F_{ka}^l F_{kb}^l] \right), \]
which we note is the same for all \( k \). We now calculate these 2- and 4-point expectations to leading order.

Since the entries of \( WX \) are multivariate Gaussian conditional on \( X \), we find
\[ \mathbb{E} F_{ka}^l F_{kb}^l = \mathbb{E} f'(Z_a) f'(Z_b), \] (S8)
where
\[ (Z_a, Z_b) \sim \mathcal{N} \left( 0, \frac{1}{n_0} \begin{pmatrix} X_{a, a} & X_{a, b} & X_{b, a} & X_{b, b} \end{pmatrix} \right) \]
\[ = \mathcal{N} \left( 0, \begin{pmatrix} 1 + O \left( n_0^{-1/2} \right) & O \left( n_0^{-1/2} \right) \\ O \left( n_0^{-1/2} \right) & 1 + O \left( n_0^{-1/2} \right) \end{pmatrix} \right). \] (S9)

Taylor expanding in the covariance term, one can show that, for all \( a \),
\[ \mathbb{E} f'(Z_a)^2 = \eta' + \mathbb{R} \left( \frac{X_{a, a} X_{a, a}}{n_0} - 1 \right) + O(n_0^{2\varepsilon-1}), \] (S11)
where \( \mathbb{R} := \mathbb{E} \left[ f''(Z)^2 + f'(Z)f'''(Z) \right] \), and for all \( a \neq b \),
\[ \mathbb{E} f'(Z_a) f'(Z_b) = \zeta + \frac{\xi''}{2} \left( \frac{X_{a, a} X_{a, a}}{n_0} + \frac{X_{b, b} X_{b, b}}{n_0} - 2 \right) + \zeta' \frac{X_{a, a} X_{b, b}}{n_0} + O(n_0^{2\varepsilon-1}), \] (S12)
where \( \xi := \mathbb{E} f'(Z) \) and \( \xi' := \mathbb{E} f''(Z) \). Using the same argument, we find
\[ \mathbb{E} (F_{ka}^l)^2 (F_{kb}^l)^2 = (\eta')^2 + O(n_0^{2\varepsilon-1/2}); \] (S13)
where $E$

Thus finally applying the matrix Bernstein inequality with $t = Cn_0^{4\varepsilon-1/4}$ for some sufficiently large constant $C$, we find for any $\delta > 0$

$$\mathbb{P}\left\{ \|\Delta\| > Cn_0^{4\varepsilon-1/4} \right\} < \delta$$

for sufficiently large $n_0$. Moreover, eqn. (S25) holds with $X$ random as it is independent of $W_1$ and $W_2$, and our assumptions on $X$ hold for any $\delta' > 0$ for sufficiently large $n_0$.

**S1.2. Bounding remaining terms**

Using eqns. (S11) and (S12), we have

$$\frac{X^T X}{n_0} \circ (\tilde{M} - M) - R = \tilde{R} \left( \frac{X^T X}{n_0} - I \right) \circ I + \xi \xi'' \frac{X^T X}{n_0} \circ (1 e^T + e 1^T) \circ (11^T - I)$$

$$+ \zeta X^T X \circ \left( \frac{X^T X}{n_0} \circ (11^T - I) - R + E, \right.$$}

where $E$’s diagonal entries are $O(n_0^{2\varepsilon-1})$ and off-diagonal entries are $O(n_0^{3\varepsilon-3/2})$. Taking the terms one by one, we first bound

$$\|\tilde{R} \left( \frac{X^T X}{n_0} - I \right) \circ I\| = \sup_{a} \left| \tilde{R} \left( \frac{X^T X}{n_0} - 1 \right) \right| = O(n_0^{\varepsilon-1/2})$$

**The Neural Tangent Kernel in High Dimensions: Triple Descent and a Multi-Scale Theory of Generalization**

\[ \mathbb{E} (F_{ka}')^4 = C_4 + O(n_0^{2\varepsilon-1/2}); \] (S14)

for $l, a, b$ distinct,

\[ \mathbb{E} F_{ka}' F_{kb}' (F_{kl}')^2 = \zeta \eta' + O \left( n_0^{2\varepsilon-1/2} \right); \] (S15)

for $l \neq a$,

\[ \mathbb{E} F_{ka}' (F_{kl}')^3 = C_3 + O(n_0^{2\varepsilon-1/2}), \] (S16)

for some constants $C_3$ and $C_4$. Thus, we may write

\[ \sum_k \mathbb{E} \Delta_k^2 = \frac{1}{n_1} (X^T X/n_0)^2 \otimes M_2 + E, \] (S17)

where

\[ M_2 := (3\zeta \eta' - \zeta^2)11^T + 3\eta'(\eta' - \zeta)I \] (S18)

and

\[ E_{ab} := \frac{1}{n_1} \sum_l (X^T X/n_0)_{al} (X^T X/n_0)_{bl} \varepsilon_{abl} \] (S19)

for some $\varepsilon_{abl} = O(n_0^{2\varepsilon-1/2})$. We find $\| \frac{1}{n_1} (X^T X/n_0)^2 \otimes M_2 \| = O(n_0^3 / n_1)$ and

\[
\|E\| \leq \|E\|_F \\
= \left( \sum_{a,b} |E_{ab}|^2 \right)^{1/2} \\
= \left( \frac{1}{n_1^2} \sum_{a,b} \left( \sum_l (X^T X/n_0)^2_{al} (X^T X/n_0)^2_{bl} \varepsilon_{abl}^2 \right) \right)^{1/2} \\
= \sqrt{O(n_0^{6\varepsilon-1} n_1^{-2} m^2 (m n_0^{-2} + m n_0^{-1} + m n_0^{-2} + 1))} \\
= O(n_0^{3\varepsilon-1/2})
\]

using the Cauchy-Schwarz inequality and that assumption that all dimensions are on the same order.
Next, we bound
\[
\left\| \frac{\xi''}{2} \frac{XX}{n_0} \odot (e1^T + 1e^T) \odot (11^T - I) \right\| \leq O\left( n_0^{-1/2} \right).
\] (S28)

Eqn. (S28) can be demonstrated by taking the 4th power of the trace as in (El Karoui et al., 2010). This is expected, since the entries are mean zero and have variance order \( O\left( n_0^{-1} \right) \). Proving the spectral bound is a straightforward calculation using the independence of the entries of \( X \), but we avoid details here. The final term can also be bounded in this way, yielding,
\[
\left\| \xi' \frac{XX}{n_0} \odot \frac{XX}{n_0} \odot (11^T - I) - \mathcal{R} \right\| = O\left( n_0^{-1/2} \right).
\] (S29)

The inclusion of the matrix \( \mathcal{R} \) is necessary, due to the nonzero mean of the entries. See (El Karoui et al., 2010) for an example of this calculation.

Similarly using the assumptions on \( X \), we can bound the remaining diagonal matrix of eqn. (S6) as follows
\[
\left\| \left( \frac{XX}{n_0} \odot M - K_{\text{simp}} \right) \right\| = (\eta' - \zeta) \left\| \text{diag}(XX/n_0) - I \right\| = (\eta' - \zeta) \sup_a \left| \frac{1}{n_0} \sum_k X_{ka}^2 - 1 \right| = O\left( n_0^{-1/2} \right).
\] (S30)

Summing our bounds on \( \Delta \) and eqns. (S27)-(S30) completes the proof of eqn. (S4).

### S2. Gaussian equivalents

In this section we discuss the key arguments for existence of Gaussian equivalents and the linearizations of Sec. 4.2. As all the main elements of this argument have been established elsewhere, here we just provide the main intuitions and refer to prior work for the details.

Many of the statistics of random matrices are universal, that is, their limiting behavior as the matrix gets larger is insensitive to the detailed properties of their entries’ distributions. Considerable work has gone into demonstrating universality for an increasingly large class of random matrices and a growing number of detailed statistics. In our case, the test loss is a global measurement of several random matrices. This perspective gives some intuition for why we are able to replace many of the intractable terms in the expressions we analyze with tractable terms, which only need to match quite superficial properties of the distributions to ensure the limiting test loss is the same.

In Secs. S3 and S4, we use this replacement strategy in two distinct situations. The first is for terms of the form
\[
\text{tr}(AB) = \sum_{ij} A_{ij} B_{ji},
\] (S31)
for deterministic \( A \) and random \( B \). Under assumptions on \( A \) and \( B \), standard concentration inequalities can be used to describe the limiting behavior of sums like eqn. (S31). In our setting, one finds that this behavior only depends on the low-order moments of \( B \). By matching these low-order moments with Gaussian random variables, we can replace \( B \) with a Gaussian random matrix with the same limiting behavior. Note, often \( A \) is not actually deterministic, we are simply conditioning on it and only considering the randomness in \( B \). The approach is suitable for determining the average behavior of eqn. (S31) when we have control over the (weak) correlations in the entries of \( A \) and \( B \). Linearizing the matrices \( A \) and \( B \) in this setting is just a convenient bookkeeping device for performing these computations.

When one of the matrices in eqn. (S31) is inverted, the situation is more complex, and indeed this is the case for the kernel matrix \( K \) in expressions for the training and test loss. To apply the linear pencil algorithm, we have to replace the NTK in all expressions with a linearized version (see eqn. (22)), which is a rational expression of the i.i.d. Gaussian matrices, \( X, W_1, \text{etc.} \). In Sec. S1, we bounded the difference between the first-layer kernel and its linearization, thus removing the Hadamard product structure. It remains to linearize the second-layer kernel, i.e. linearize \( F \). This has been discussed in previous works, see (Mei & Montanari, 2019; Adlam et al., 2019; Péché et al., 2019; Benigni & Péché, 2019).
It should be expected that a linearized version of $F$ will lead to the same asymptotic statistics due to some very general results on the limiting behavior of expressions of the form,

$$\text{tr} \left( A \frac{1}{B - z I} \right),$$  \hspace{1cm} (S32)

where $A$ is symmetric and $z \in \mathbb{C}^+$. The resolvent $(B - z)^{-1}$ is intimately related to the spectral properties of $B$. Recently, isotropic results for quite general $A$ have been developed for matrices with correlated entries, which show that under certain assumptions the limiting behavior of eqn. (S32) depends only on the low-order moments of $B$. Specifically, the limiting behavior of eqn. (S32) is described by the matrix Dyson equation in many cases. For a summary of these results and related topics see e.g. (Erdos, 2019). While we do not explicitly show the correlation structure of $K$ meets the conditions known to suffice for the matrix Dyson equation, the assumptions in Sec. 2 imply that the correlations between entries of $K$ are weak, which is the essential ingredient.

Finding Gaussian equivalents for $A$ and $B$ in expressions like eqns. (S31) and (S32) is relatively simple in our case. We encounter terms for which the matrix $B$ depends on some other random matrix $C$ through a coordinate-wise nonlinear function $f(C)$. For such cases, Taylor expanding the function $f$ is the key tool to finding these equivalents (see e.g. (Adlam et al., 2019) for more details on this type of approach).

### S3. Exact asymptotics for the training loss

#### S3.1. Decomposition of terms

The model’s predictions on the training set, $\hat{y}(X)$, take a simple form,

$$\hat{y}(X) = N_0(X) + (Y - N_0(X))K^{-1}K(X,X)$$  \hspace{1cm} (S33)

$$= Y - \gamma(Y - N_0(X))K^{-1}. $$  \hspace{1cm} (S34)

The expected training loss can be written as,

$$E_{\text{train}} = \frac{1}{m} \mathbb{E} \text{tr} \left( (Y - \hat{y}(X))(Y - \hat{y}(X))^\top \right)$$  \hspace{1cm} (S35)

$$= \frac{\gamma^2}{m} \mathbb{E} \text{tr} \left( (Y - N_0(X))^\top (Y - N_0(X))K^{-2} \right)$$  \hspace{1cm} (S36)

$$= T_1 + \nu T_2 $$  \hspace{1cm} (S37)

where $\nu = 0$ with centering and $\nu = 1$ without it and,

$$T_1 = \frac{\gamma^2}{m} \mathbb{E} \text{tr}(Y^\top Y K^{-2})$$  \hspace{1cm} (S38)

$$T_2 = \frac{\gamma^2}{m} \mathbb{E} \text{tr}(N_0(X)^\top N_0(X)K^{-2}). $$  \hspace{1cm} (S39)

Note we can suppress the terms linear in $N_0$ since they vanish in expectation owing to the linear dependence on the mean-zero random variable $\omega$. Here $K = K(X,X) + \gamma I_m$ is the linearized NTK and is given by,

$$K = \sigma^2 W_2 \left[(\eta' - \zeta)I_m + \frac{\zeta X^\top X}{n_0} \right] + \frac{F^\top F}{n_1} + \gamma I_m $$  \hspace{1cm} (S40)

This substitution can be justified using the result of Sec. S1:

$$\frac{\gamma^2}{m} \mathbb{E} \text{tr} \left( Y^\top Y K^{-2}_{\text{simp}} \right) - \frac{\gamma^2}{m} \mathbb{E} \text{tr}(Y^\top Y K^{-2}) = \frac{\gamma^2}{m} \mathbb{E} \left( Y^\top \left( K^{-2}_{\text{simp}} - K^{-2} \right) Y \right)$$  \hspace{1cm} (S41)

$$\leq \frac{\gamma^2}{mn_0} \mathbb{E} |Y^\top 1|^2 + \frac{\gamma^2}{m} \mathbb{E} \|Y\|^2 \|R + K^{-2}_{\text{simp}} - K^{-2}\| = o(1). $$  \hspace{1cm} (S42)

Eqn. (S39) is similar.
we have,

\[ \mathbb{E}_{W_2} N_0(X)^T N_0(X) = \sigma_{W_2}^2 K - \sigma_{W_2}^2 \left[ \sigma_{W_2}^2 (\eta' - \zeta) + \gamma \right] I_m - \sigma_{W_2}^4 \frac{\zeta X^T X}{n_0}, \]  \hspace{1cm} (S43)

since \( \mathbb{E}_{W_2} N_0(X)^T N_0(X) = \sigma_{W_2}^2 / n_1 F^T F \).

Next we recall the substitution (14),

\[ Y \rightarrow Y^{\text{lin}} = \frac{1}{\sqrt{n_0 n_1}} \omega \Omega X + \mathcal{E}, \]  \hspace{1cm} (S44)

which can be used to calculate the expectation over \( \omega \) and \( \Omega \) to leading order (i.e. with remainder terms \( o(1) \)) using the approach of eqn. (S31). Concretely,

\[
\frac{\gamma^2}{m} \mathbb{E}_{\omega, \Omega, \mathcal{E}} \text{tr}(Y^T Y K^{-2}) = \frac{\gamma^2}{m} \mathbb{E}_{\omega, \Omega, \mathcal{E}} \text{tr}(Y^{\text{lin}}^T Y^{\text{lin}} K^{-2}) + o(1) = \frac{\gamma^2}{m} \text{tr} \left[ \left( \frac{1}{n_0} X^T X + \sigma^2_{W_2} I_m \right) K^{-2} \right] + o(1). \]  \hspace{1cm} (S45)

Putting these pieces together, we can write for \( \tau_1 = \tau_1(\gamma) \) and \( \tau_2 = \tau_2(\gamma) \),

\[
T_1 = -\gamma^2 (\sigma^2_{W_2} \tau'_1 + \tau'_2)
\]

\[
T_2 = \sigma_{W_2}^2 \gamma^2 \left( \tau_1 + (\sigma^2_{W_2} (\eta' - \zeta) + \gamma) \tau'_1 + \sigma^2_{W_2} \zeta \tau'_2 \right),
\]

where,

\[
\tau_1 = \frac{1}{m} \text{tr}(K^{-1}), \quad \text{and} \quad \tau_2 = \frac{1}{m} \text{tr} \left( \frac{1}{n_0} X^T X K^{-1} \right).
\]

Self-consistent equations for \( \tau_1 \) and \( \tau_2 \) can be computed using the resolvent method, as was done in (Adlam et al., 2019) for the case of \( \sigma_{W_2} = 0 \). In order to pave the way for the analysis of the test error, we instead demonstrate how to compute these traces using operator-valued free probability.

**Remark S1.** In the remainder of this section, and in Sec. S4, we assume at times that \( \sigma \) is non-linear (so that \( \eta' > \zeta \) and \( \eta > \zeta \) and/or \( \gamma > 0 \) in order that certain denominator factors are non-zero). The linear and/or ridgeless cases can be obtained by limits of our general results, or through special cases of the pertinent intermediate formulas.

### S3.2. Linear pencils

To begin, we construct linear pencils for \( \tau_1 \) and \( \tau_2 \). Using the linearization eqn. (13), a straightforward block-matrix inversion confirms that

\[
\tau_1 = \mathbb{E} \text{tr} \left( [Q_T^{-1}]_{1,1} \right), \quad \text{and} \quad \mathbb{E} \tau_2 = \text{tr} \left( [Q_T^{-1}]_{2,4} \right),
\]

where,

\[
Q_T = \begin{pmatrix}
I_m \left( \gamma + \sigma^2_{W_2} (\eta' - \zeta) \right) & \frac{\zeta X^T}{n_0} \sigma^2_{W_2} & \sqrt{\eta - \zeta} \Theta_F & \sqrt{\eta - \zeta} \Theta_F \\
-\sqrt{\eta - \zeta} \Theta_F & I_{n_0} & 0 & 0 \\
0 & -\sqrt{\zeta W_1 \zeta} & I_{n_1} & 0 \\
0 & 0 & \sqrt{\zeta W_1 \zeta} & -\sqrt{\zeta W_1 \zeta} \phi \phi \end{pmatrix}.
\]

(S50)

The matrix \( Q_T \) is not self-adjoint, but a self-adjoint representation can be obtained from it by doubling the dimensionality. In particular, letting

\[
\bar{Q}_T = \begin{pmatrix}
0 & Q_T^T \\
Q_T & 0 
\end{pmatrix},
\]

we have,

\[
\tau_1 = \mathbb{E} \text{tr} \left( [\bar{Q}_T^{-1}]_{1,5} \right), \quad \text{and} \quad \mathbb{E} \text{tr} \left( [\bar{Q}_T^{-1}]_{2,8} \right).
\]

(S55)

Observe that \( \bar{Q}_T \) is a self-adjoint matrix whose blocks are either constants or proportional to one of \( \{ X, X^T, W_1, W_1^T, \Theta_F, \Theta_F^T \} \); let us denote the constant terms as \( Z \). As such, we can directly utilize the results of (Far et al., 2006; Mingo & Speicher, 2017) to compute the necessary traces.
S3.3. Operator-valued Stieltjes transform

The traces can be extracted from the operator-valued Stieltjes transform $G : M_d(\mathbb{C})^+ \to M_d(\mathbb{C})^+$, which is a solution of the equation,

$$ZG = I_d + \eta(G)G,$$  \hspace{1cm}  (S53)  

where $d$ is the number of blocks, $\eta : M_d(\mathbb{C}) \to M_d(\mathbb{C})$ defined by

$$[\eta(D)]_{ij} = \sum_{kl} \sigma(i, k; l, j)\alpha_k D_{kl},$$  \hspace{1cm}  (S54)  

where $\alpha_k$ is dimensionality of the $k$th block and $\sigma(i, k; l, j)$ denotes the covariance between the entries of the blocks $i,j$ block of $\bar{Q}$ and entries of the $k,l$ block of $Q$. Eqn. (S53) may admit many solutions, but there is a unique solution such that $\Im G \succ 0$ for $\Im Z \succ 0$.

The constants $Z$, the entries of $\sigma$, and therefore the equations (S54) are manifest by inspection of the block matrix representation for $\bar{Q}_T$. Although the matrix representation of the equations is too large to reproduce here, we can nevertheless extract the equations satisfied by each entry of $G$.

The equations satisfied by the operator-valued Stieltjes transform $G$ of $\bar{Q}_T$ induce the following structure on $G$,

$$G = \begin{pmatrix} \tau_1 & 0 & 0 & 0 \\ 0 & g_3 & 0 & \tau_2 \\ 0 & 0 & g_4 & 0 \\ 0 & g_6 & 0 & g_5 \end{pmatrix}$$  \hspace{1cm}  (S55)  

where,

$$G_{12} = \begin{pmatrix} \tau_1 & 0 & 0 & 0 \\ 0 & g_3 & 0 & \tau_2 \\ 0 & 0 & g_4 & 0 \\ 0 & g_6 & 0 & g_5 \end{pmatrix}$$  \hspace{1cm}  (S56)  

and the independent entry-wise component functions $g_i$, $\tau_1$ and $\tau_2$ satisfy the following system of polynomial equations,

$$0 = \sqrt{\zeta} g_6 \psi - \zeta g_3 g_4 \sqrt{n_0}$$  \hspace{1cm}  (S57)  

$$0 = \sqrt{\zeta} g_3 (r_2 - g_3 r_1)$$  \hspace{1cm}  (S58)  

$$0 = \sqrt{\zeta} g_5 (g_5 g_6 r_1) + \sqrt{n_0} \phi$$  \hspace{1cm}  (S59)  

$$0 = -\zeta g_4 g_5 g_6 (\zeta r_1 \sigma_{W_2}^2 + \phi)$$  \hspace{1cm}  (S60)  

$$0 = \sqrt{\zeta} g_3 \psi + \sqrt{n_0} (\phi - \zeta g_4 \tau_2)$$  \hspace{1cm}  (S61)  

$$0 = \phi - g_4 (\tau_1 \psi (\eta - \zeta) + \zeta \tau_2 \psi + \phi)$$  \hspace{1cm}  (S62)  

$$0 = -\zeta g_4 \tau_2 - g_3 (\zeta r_1 \sigma_{W_2}^2 + \phi) + \phi$$  \hspace{1cm}  (S63)  

$$0 = -\sqrt{\zeta} g_5 \psi - \sqrt{n_0} \tau_2 (\zeta r_1 \sigma_{W_2}^2 + \phi)$$  \hspace{1cm}  (S64)  

$$0 = \sqrt{n_0} (\phi - g_3 (\zeta r_1 \sigma_{W_2}^2 + \phi)) - \sqrt{\zeta} g_6 \tau_1 \psi$$  \hspace{1cm}  (S65)  

$$0 = \sqrt{n_0} (1 - \tau_1 (\gamma + g_4 (\eta - \zeta) + \sigma_{W_2}^2 (\eta' + \zeta (g_3 - 1))) - \sqrt{\zeta} g_6 \tau_1 \psi. \hspace{1cm}  (S66)$$

It is straightforward algebra to eliminate $g_3, g_4, g_5$ and $g_6$ from the above equations. A simple set of equations for $\tau_1$ and $\tau_2$ follows,

$$0 = \phi (\zeta r_2 \tau_1 + \psi (r_2 - \tau_1)) + \zeta r_1 \tau_2 \psi (\gamma (\eta - \zeta) + 1) + \zeta r_1 \tau_2 \sigma_{W_2}^2 (\zeta (r_2 - \tau_1) \psi + \tau_1 \psi \eta' + \phi)$$  \hspace{1cm}  (S67)  

$$0 = \zeta r_1^2 \tau_2 (\eta' - \eta) \sigma_{W_2}^2 + \zeta r_1 \tau_2 (\gamma (\eta - \zeta) + 1 - (r_2 - \tau_1) \phi (\zeta (r_2 - \tau_1) + \eta \tau_1)). \hspace{1cm}  (S68)$$

Although these equations admit multiple solutions, the general results of (Far et al., 2006; Mingo & Speicher, 2017) guarantee that the correct root is given by the unique solutions $\tau_1, \tau_2 : \mathbb{C}^+ \to \mathbb{C}^+$ which are analytic in the upper half-plane.

It will prove useful to obtain expressions for $\tau_1' (\gamma)$ and $\tau_2' (\gamma)$. By differentiating eqns. (S67) and (S68) with respect to $\gamma$, we find

$$\tau_1' = -\frac{\zeta^2 r_2^2 (\psi \tau_1^2 - \phi^2)}{\psi \tau_1^2 (\zeta^2 (r_2 + 1)^2 + \phi (\zeta r_2 + \eta) (\zeta (r_2 (2r_2 + 3) + \eta)) + \zeta^2 \phi^2 (r_2 + 1)^2 (\phi r_2^2 - 1))}$$  \hspace{1cm}  (S69)  

$$\tau_2' = \frac{\zeta^2 r_2^2 (\psi \tau_2^2 - \phi^2)}{\psi \tau_2^2 (\zeta^2 (r_2 + 1)^2 + \phi (\zeta r_2 + \eta) (\zeta (r_2 (2r_2 + 3) + \eta)) + \zeta^2 \phi^2 (r_2 + 1)^2 (\phi r_2^2 - 1))}$$
As described in Sec. 4.2, without loss of generality we can consider the case of a linear teacher, so that as in Sec. S3, we suppress the terms linear in $\omega$ where we have introduced some auxiliary variables to ease the presentation, 

$$\tilde{\tau}_1 = \sigma^2_{W_2} \zeta \tilde{\tau}_2 + \phi \tilde{\tau}_2 \quad \text{and} \quad \tilde{\tau}_2 = -1 + \tau_2/\tau_1.$$  

(S71)

### S4. Exact asymptotics for the test loss

#### S4.1. Decomposition of terms

As described in Sec. 4.3, the test loss can be written as,

$$E_{\text{test}} = E(y - \hat{y}(x))^2 = E_1 + E_2 + E_3$$  

(S72)

with

$$
E_1 = \mathbb{E} \text{tr}(y(x)y(x)^\top) + \mathbb{E} \text{tr}(N_0(x)N_0(x)^\top) \\
E_2 = -2\mathbb{E} \text{tr}(K_{x}^\top K^{-1}y(x)^\top) - 2\mathbb{E} \text{tr}(K_{x}^\top K^{-1}N_0(x)^\top N_0(x)) \\
E_3 = \mathbb{E} \text{tr}(K_{x}^\top K^{-1}y(x)^\top y(x)) + \mathbb{E} \text{tr}(F_{\text{test}} K^{-1}N_0(x)^\top N_0(x)K^{-1}K_{x}).
$$

(S73) (S74) (S75)

As in Sec. S3, we suppress the terms linear in $\omega$ as they vanish in expectation. The Neural Tangent Kernels $K = K(X, X) + \gamma I$ and $K_x = K(X, x)$ are given by,

$$K = \sigma^2_{W_2} [\eta' - \zeta] I_m + \frac{\zeta X^\top X}{n_0} + \frac{F^\top F}{n_1} + \gamma I_m, \quad \text{and} \quad K_x = \frac{\sigma^2_{W_2} \zeta}{n_0} X^\top x + \frac{1}{n_1} F^\top f,$$  

(S76)

where the substitution for the linearized NTK is justified as in Sec. S3 using the spectral norm bound of Sec. S1.

Using the cyclicity and linearity of the trace, the expectation over x requires the computation of

$$
\mathbb{E}_x K_x K_x^\top, \quad \mathbb{E}_x y(x) K_x^\top, \quad \mathbb{E}_x y(x)y(x)^\top, \quad \mathbb{E}_x N_0(x) K_x^\top, \quad \text{and} \quad \mathbb{E}_x N_0(x) N_0(x)^\top.
$$

(S77)

As described in Sec. 4.2, without loss of generality we can consider the case of a linear teacher, so that $\eta_t = \zeta_t = 1$ and (16) and (15) become

$$y \rightarrow y^{\text{lin}} = \frac{\sqrt{\tau}}{\sqrt{n_0\tau}} \omega \Omega x + \sqrt{\eta_t - \zeta_t} \frac{1}{\sqrt{n_t}} \omega \theta_y = \frac{1}{\sqrt{n_0 n_t}} \omega \Omega x \quad \text{and} \quad f \rightarrow f^{\text{lin}} = \frac{\sqrt{\tau}}{\sqrt{n_0}} W_1 x + \sqrt{\eta - \zeta} \theta f.$$  

(S78)

Using these substitutions, the expectations over x are now trivial and we readily find,

$$
\mathbb{E}_x K_x K_x^\top = \frac{\sigma^4_{W_2} \zeta^2}{n_0} X^\top X + \frac{\sigma^2_{W_2} \zeta}{n_0 n_1^{3/2}} (X^\top W_1^\top F + F^\top W_1 X) + \frac{1}{n_1^2} F^\top \left( \frac{\zeta}{n_0} W_1 W_1^\top + (\eta - \zeta) I_{n_1} \right) F
$$

(S79)

$$
\mathbb{E}_x y(x) K_x^\top = \frac{\sigma^2_{W_2} \zeta}{n_0 n_1^{3/2}} \omega \Omega X + \frac{\sqrt{\tau}}{n_0 n_1 \sqrt{n_t}} \omega \Omega W_1^\top F
$$

(S80)

$$
\mathbb{E}_x y(x)y(x)^\top = \frac{1}{n_0 n_t} \omega \Omega^\top \omega^\top
$$

(S81)

$$
\mathbb{E}_x N_0(x) K_x^\top = \frac{\sigma^2_{W_2} \zeta}{n_0 n_1^{3/2}} W_2 W_1 X + \frac{1}{n_1^{3/2}} W_2 (\frac{\zeta}{n_0} W_1 W_1^\top + (\eta - \zeta) I_{n_1}) F
$$

(S82)

$$
\mathbb{E}_x \text{tr}(N_0(x) N_0(x)^\top) = \sigma^2_{W_2} \eta.
$$

(S83)

One may interpret the substitutions in eqn. (S78) as a tool to calculate the expectations above to leading order as it leads to terms like eqn. (S31). Next we recall the substitution (S44),

$$Y \rightarrow Y^{\text{lin}} = \frac{1}{\sqrt{n_0 n_t}} \omega \Omega X + \mathcal{E}.$$  

(S84)
As above, we consider the leading order behavior with respect to the random variables $\omega$, $\Omega$, and $W_2$ using eqn. (S31) to find

$$E_{\omega,\Omega,E} \left[ Y^T Y \right] = \frac{1}{n_0} X^T X + \sigma_2^2 I_m \quad \text{(S85)}$$

$$E_{\omega,\Omega,E,W_2} \left[ Y^T E_\omega y(x) K_x^T \right] = \frac{\sigma_2^2 \zeta}{n_0^n} X^T X + \frac{\sqrt{\zeta}}{n_1^{3/2}} X^T W_1^T F \quad \text{(S86)}$$

$$E_{W_2} \left[ N_0(X)^T N_0(X) \right] = \frac{\sigma_2^2}{n_1} F^T F \quad \text{(S87)}$$

$$E_{W_2} \left[ N_0(X)^T E_\omega N_0(x) K_x^T \right] = \frac{\sigma_2^2 \zeta^{3/2}}{n_0^{3/2} n_1} F^T W_1 X + \frac{\sigma_2^2}{n_1^2} F^T \left( \frac{\zeta}{n_0} W_1 W_1^T + (\eta - \zeta) I_{n_1} \right) F. \quad \text{(S88)}$$

Using (13),

$$F \rightarrow F^{\text{lin}} = \frac{\sqrt{\zeta}}{\sqrt{n_0}} W_1 X + \frac{\sqrt{\eta - \zeta}}{\sqrt{n_1}} \Theta_F \quad \text{(S89)}$$

we can write,

$$\frac{\sqrt{\zeta}}{\sqrt{n_0}} F^T W_1 X + \frac{\sqrt{\eta}}{\sqrt{n_0}} X^T W_1^T F = F^T F + \frac{\zeta}{n_0} X^T W_1^T W_1 X - (\eta - \zeta) \Theta_F \Theta_F. \quad \text{(S90)}$$

Putting these pieces together, we have

$$E_1 = 1 + \nu \sigma_2^2 \eta \quad \text{(S91)}$$

$$E_2 = E_{21} + \nu E_{22} \quad \text{(S92)}$$

$$E_3 = E_{31} + E_{32} + \nu E_{33}, \quad \text{(S93)}$$

where $\nu = 0$ with centering and $\nu = 1$ without it,

$$E_{21} = - E \left( 2 \frac{\sigma_2^2 \zeta}{n_0^n} X K^{-1} X^T + \frac{\eta}{n_0 n_1} F K^{-1} F^T + \frac{\zeta}{n_0 n_1} W_1 X K^{-1} X^T W_1^T - \frac{\eta - \zeta}{n_0 n_1} \Theta_F K^{-1} \Theta_F \right) \quad \text{(S94)}$$

$$E_{22} = - 2 \frac{\sigma_2^2}{n_1} E \left( \frac{\sigma_2 \zeta^{3/2}}{n_0^{3/2}} K^{-1} F^T W_1 X + \frac{\zeta}{n_0 n_1} K^{-1} F^T W_1 W_1^T F + \frac{\eta - \zeta}{n_1} K^{-1} F^T F \right) \quad \text{(S95)}$$

$$E_{31} = \sigma_2^2 E \left( K^{-1} \Sigma_3 K^{-1} \right) \quad \text{(S96)}$$

$$E_{32} = \frac{1}{n_0} E \left( XK^{-1} \Sigma_3 K^{-1} X^T \right) \quad \text{(S97)}$$

$$E_{33} = \frac{\sigma_2^2}{n_1} E \left( F K^{-1} \Sigma_3 K^{-1} F^T \right), \quad \text{(S98)}$$

and,

$$\Sigma_3 = \frac{\sigma_2^2 \zeta^2}{n_0^n} X^T X + \left( \frac{\sigma_2 \zeta}{n_0 n_1} + \frac{\eta - \zeta}{n_1^2} \right) F^T F + \frac{\zeta}{n_0 n_1} F^T W_1 W_1^T F + \frac{\sigma_2^2 \zeta^2}{n_0^{3/2} n_1} X^T W_1^T W_1 X - \frac{\sigma_2 \zeta (\eta - \zeta)}{n_0 n_1} \Theta_F \Theta_F. \quad \text{(S99)}$$

### S4.2. Linear pencils

Repeated application of the Schur complement formula for block matrix inversion establishes the following representations for $E_{21}, E_{22}, E_{31}, E_{32}, E_{33}$.

#### S4.2.1. $E_{21}$

A linear pencil for $E_{21}$ follows from the representation,

$$E_{21} = E \left( U_{21}^T Q_{21}^{-1} V_{21} \right), \quad \text{(S100)}$$

where,

$$U_{21}^T = \left( 0 \quad 0 \quad 0 \quad 0 \quad -\frac{\eta - \zeta}{\eta} I_{n_1} \quad 0 \quad 0 \quad 0 \quad -\frac{I_{n_1}}{n_0} \quad 0 \quad 0 \right) \quad \text{(S101)}$$
\[ V_{21}^T = \begin{pmatrix} 0 & 0 & 0 & -\frac{\sqrt{\gamma_0 n_1} I_{n_0}}{\sqrt{\zeta}} & 0 & 0 & 0 & 0 & I_{n_1} & 0 & 0 & 0 \end{pmatrix} \] (S102)

and,

\[ Q_{21} = \begin{pmatrix} Q_{21}^{11} & 0 & 0 \\ 0 & Q_{21}^{22} & Q_{21}^{23} \\ 0 & 0 & Q_{21}^{33} \end{pmatrix} \] (S103)

with,

\[ Q_{21}^{11} = \begin{pmatrix} I_{m_1} (\gamma + \sigma_{W_2}^2 (\eta' - \zeta)) & \zeta X^T \sigma_{W_2}^2 & \sqrt{\eta - \zeta} \Theta_F & \sqrt{\gamma - \zeta} \Theta_F & \sqrt{\gamma X^T} \\ -X & I_{n_0} & 0 & 0 \\ -\sqrt{\eta - \zeta} \Theta_F & -\sqrt{\gamma X} & I_{n_1} & 0 \\ 0 & 0 & -W_1^T & I_{n_0} \end{pmatrix} \] (S104)

\[ Q_{21}^{22} = \begin{pmatrix} I_{m_1} (\gamma + \sigma_{W_2}^2 (\eta' - \zeta)) & 0 & \zeta X^T \sigma_{W_2}^2 & \sqrt{\eta - \zeta} \Theta_F & \sqrt{\gamma X^T} \\ -\Theta_F & I_{n_1} & -\sqrt{\gamma X} & 0 & 0 \\ -X & 0 & I_{n_0} & 0 & 0 \\ -\sqrt{\eta - \zeta} \Theta_F & 0 & -W_1^T & I_{n_1} & I_{n_0} \end{pmatrix} \] (S105)

\[ Q_{21}^{23} = \begin{pmatrix} -\Theta_F^T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\gamma X} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ I_{n_1} & 0 & 0 & 0 & 0 \end{pmatrix} \] (S106)

\[ Q_{21}^{33} = \begin{pmatrix} -\sqrt{\eta - \zeta} \Theta_F^T & I_{m_1} (\gamma + \sigma_{W_2}^2 (\eta' - \zeta)) & \sqrt{\eta - \zeta} \Theta_F & \sqrt{\gamma - \zeta} \Theta_F & \sqrt{\gamma X^T} \\ 0 & -\sqrt{\eta - \zeta} \Theta_F & I_{n_1} & -\sqrt{\gamma X} & 0 \\ 0 & 0 & -\sqrt{\eta - \zeta} \Theta_F & 0 & I_{n_0} \\ n_1 W_1^T & 0 & -W_1^T & I_{n_1} & I_{n_0} \end{pmatrix}. \] (S107)

S4.2.2. \( E_{22} \)

A linear pencil for \( E_{22} \) follows from the representation,

\[ E_{22} = \mathbb{E} \text{tr}(U_{22}^T Q_{22}^{-1} V_{22}^T), \] (S108)

where,

\[ U_{22}^T = \begin{pmatrix} 0 & 2\sqrt{\tau} I_{n_1} \sigma_{W_2}^2 \sqrt{n_0 (\eta - \zeta) + \zeta n_1 \sigma_{W_2}^2} & 0 & 0 \end{pmatrix} \] (S109)

\[ V_{22}^T = \begin{pmatrix} 0 & 0 & 0 & 0 & -n_1 I_{n_1} & 0 \end{pmatrix} \] (S110)

and,

\[ Q_{22} = \begin{pmatrix} I_{n_0} & 0 & -X & 0 & 0 & 0 \\ -W_1 & I_{n_1} & 0 & 0 & -\frac{\sqrt{\gamma X} \sigma_{W_2}^2}{\sqrt{n_0 n_1}} & 0 \\ -\frac{\sqrt{\gamma X^T} \sigma_{W_2}^2}{n_0} & 0 & I_{m_1} (\gamma + \sigma_{W_2}^2 (\eta' - \zeta)) & 0 & 0 & \frac{\sqrt{\eta - \zeta} \Theta_F}{n_1} & \frac{\sqrt{\gamma - \zeta} \Theta_F}{\sqrt{n_0 n_1}} \\ 0 & 0 & -\sqrt{\eta - \zeta} \Theta_F & I_{n_1} & 0 & w_2 \frac{1}{\sqrt{n_1 \sigma_{W_2}^2}} & 0 \\ 0 & -\frac{\sqrt{\gamma X} W_1}{\sqrt{n_0}} & 0 & -W_1^T & I_{n_0} & 0 & 0 \\ 0 & 0 & -\sqrt{\eta - \zeta} \Theta_F & 0 & 0 & I_{n_1} & 0 \\ 0 & 0 & 0 & 0 & -W_1^T & I_{n_0} \end{pmatrix}. \] (S111)
S4.2.3. $E_{31}$

A linear pencil for $E_{31}$ follows from the representation,

$$E_{31} = \mathbb{E}\text{tr}(U_{31}^T Q_{31}^{-1} V_{31}),$$

(S112)

where,

$$U_{31}^T = (m \sigma_x^2 I_m \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0) , \quad V_{31}^T = (0 \ 0 \ 0 \ 0 \ 0 \ I_m \ 0 \ 0)$$

(S113)

and, for $\beta = (n_0 (\zeta - \eta) - \zeta n_1 \sigma_W^2)$,

$$Q_{31} = \begin{pmatrix}
I_m (\gamma + \sigma_W^2 (\eta' - \zeta)) & \frac{\zeta X^T \sigma_W^2}{n_0} & \frac{\sqrt{\pi \pi \eta}}{n_1} & \frac{\zeta X^T \sigma_W^2}{n_0} & 0 & \frac{\sqrt{\pi \pi \eta} \beta}{n_1} & \frac{n_0 \pi}{n_1} & \frac{n_0 \pi}{n_1} & 0 \\
-X & I_{n_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\sqrt{\eta - \zeta} & I_{n_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -W_1^T & I_{n_0} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\frac{\zeta X W}{\sqrt{n_0}} & -\sqrt{\eta - \zeta} & I_{n_1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -W_1^T & I_{n_0} & 0 \\
\end{pmatrix}.
$$

(S114)

S4.2.4. $E_{32}$

A linear pencil for $E_{32}$ follows from the representation,

$$E_{32} = \mathbb{E}\text{tr}(U_{32}^T Q_{32}^{-1} V_{32}),$$

(S115)

where,

$$U_{32}^T = (0 \ I_{n_0} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0) , \quad V_{32}^T = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -\frac{\sqrt{n_0 n_1} I_{n_0}}{\sqrt{\zeta}})$$

(S116)

and, for $\beta = (n_0 (\zeta - \eta) - \zeta n_1 \sigma_W^2)$,

$$Q_{32} = \begin{pmatrix}
I_m (\gamma + \sigma_W^2 (\eta' - \zeta)) & \frac{\zeta X^T \sigma_W^2}{n_0} & \frac{\sqrt{\pi \pi \eta}}{n_1} & \frac{\zeta X^T \sigma_W^2}{n_0} & 0 & \frac{\sqrt{\pi \pi \eta} \beta}{n_1} & \frac{n_0 \pi}{n_1} & \frac{n_0 \pi}{n_1} & 0 \\
-X & I_{n_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\sqrt{\eta - \zeta} & I_{n_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -W_1^T & I_{n_0} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\frac{\zeta X W}{\sqrt{n_0}} & -\sqrt{\eta - \zeta} & I_{n_1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -W_1^T & I_{n_0} & 0 \\
\end{pmatrix}.
$$

(S117)

S4.2.5. $E_{33}$

A linear pencil for $E_{33}$ follows from the representation,

$$E_{33} = \mathbb{E}\text{tr}(U_{33}^T Q_{33}^{-1} V_{33}),$$

(S118)

where,

$$U_{33}^T = (0 \ I_{n_1} \sigma_W^2 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0) \quad (S119)$$

$$V_{33}^T = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -n_1 I_{n_1} \ 0) \quad (S120)$$
where $\alpha$ are the dimensions of the $k$-th block of $\tilde{Q}$ and entries of the $kl$ block of $\tilde{Q}$. Eqn. (S123) may admit many solutions, but there is a unique solution such that $\text{Im} G > 0$ for $\text{Im} Z > 0$.

The constants $Z$, the entries of $\sigma$, and therefore the equations (S124) are manifest by inspection of the block matrix representations for $\tilde{Q}$. Although the matrix representations are too large to reproduce here, we can nevertheless extract the equations satisfied by each entry of $G$, which we present in the subsequent sections.

S4.3.1. $E_{21}$

The equations satisfied by the operator-valued Stieltjes transform $G$ of $\tilde{Q}_{21}$ induce the following structure on $G$,

$$
G = \begin{pmatrix}
0 & \tilde{G}_{12} \\
\tilde{G}_{12}^T & 0
\end{pmatrix},
$$

(S125)
where,
\[
G_{12} = \begin{pmatrix}
  g_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & g_9 & 0 & g_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & g_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & g_{12} & 0 & g_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & g_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & g_1 & 0 & g_5 & 0 & g_4 & 0 & g_7 \\
 0 & 0 & 0 & 0 & 0 & 0 & g_9 & 0 & g_6 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{11} & 0 & g_{13} & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{12} & 0 & g_{10} & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_8 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_9 & g_6 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{12} \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix},
\]
\text{ (S126)}

and the independent entry-wise component functions \(g_i\) combine to produce the error \(E_{21}\) through the relation,
\[
E_{21} = \frac{g_4 (\eta - \zeta)}{n_0} + \frac{2 \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W}{\psi} - \frac{g_2}{n_0},
\]
\text{ (S127)}

and themselves satisfy the following system of polynomial equations,
\[
\begin{align*}
0 &= 1 - g_1 \\
0 &= \sqrt{\zeta} g_9 g_{11} \sqrt{n_0} - g_{12} \psi \\
0 &= \sqrt{\zeta} g_6 g_{11} \sqrt{n_0} - g_{10} \psi + \psi \\
0 &= g_7 (\eta - \zeta) + \sqrt{\zeta} g_6 g_{11} \sqrt{n_0} \\
0 &= g_9 (\eta - \zeta) - g_4 (\eta + \sigma_W (\eta - \zeta)) \\
0 &= -g_8 (g_9 g_6 - g_6 \sqrt{n_0} (\gamma + \sigma_W (\eta - \zeta))) \\
0 &= -\sqrt{\zeta} g_6 g_{12} - (g_{10} - 1) \sqrt{n_0} (\gamma + \sigma_W (\eta - \zeta)) \\
0 &= \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta + \gamma) + g_{8} (\sqrt{\zeta} g_{10} \psi + \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_{11} \sqrt{n_0} (\gamma + \sigma_W (\eta - \zeta)) + g_8 (\sqrt{\zeta} g_{11} \psi + \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
0 &= g_{11} (g_6 \psi (\eta - \zeta) + g_6 \psi (\eta - \zeta) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta))) + \phi (\gamma + \sigma_W (\eta - \zeta)) \\
0 &= g_{11} \sqrt{n_0} \sigma_W (\eta - \zeta) + g_6 \sqrt{n_0} \sigma_W (\eta - \zeta) - g_2 (\gamma + \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_6 \sqrt{n_0} (\phi (\eta - \zeta) + \sigma_W (\eta - \zeta)) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \psi) (\gamma + \sigma_W (\eta - \zeta)) \\
0 &= g_2 (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta) - \gamma) - g_6 (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) + \phi (\gamma + \sigma_W (\eta - \zeta)) \\
0 &= g_1 n_0 (g_6 \psi (\eta - \zeta) + \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) + g_2 (g_6 \psi (\eta - \zeta) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta))) \\
0 &= g_1 (g_6 \psi (\eta - \zeta) + \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) + g_2 (g_6 \psi (\eta - \zeta) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta))) \\
0 &= g_1 (g_6 \psi (\eta - \zeta) + \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) + g_2 (g_6 \psi (\eta - \zeta) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta))) \\
0 &= (g_6 - 1) \sqrt{n_0} (\psi (\eta - \zeta) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta))) + \phi (\gamma + \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_6 \sqrt{n_0} (\phi (\eta - \zeta) + \sigma_W (\eta - \zeta)) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_6 \sqrt{n_0} (\psi (\eta - \zeta) + \sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) + g_2 (g_6 \psi (\eta - \zeta) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta))) \\
0 &= g_1 g_6 \sqrt{n_0} (\phi (\eta - \zeta) + \sigma_W (\eta - \zeta)) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_6 \sqrt{n_0} (\phi (\eta - \zeta) + \sigma_W (\eta - \zeta)) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_6 \sqrt{n_0} (\phi (\eta - \zeta) + \sigma_W (\eta - \zeta)) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_6 \sqrt{n_0} (\phi (\eta - \zeta) + \sigma_W (\eta - \zeta)) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
0 &= g_1 g_6 \sqrt{n_0} (\phi (\eta - \zeta) + \sigma_W (\eta - \zeta)) - \phi (\sqrt{\zeta} g_6 \sqrt{n_0} \sigma_W (\eta - \zeta)) \\
\end{align*}
\]
\text{ (S128a)}

\text{ (S128b)}

\text{ (S128c)}

\text{ (S128d)}

\text{ (S128e)}

\text{ (S128f)}

\text{ (S128g)}

\text{ (S128h)}

\text{ (S128i)}

\text{ (S128j)}

\text{ (S128k)}

\text{ (S128l)}

\text{ (S128m)}

\text{ (S128n)}

\text{ (S128o)}

\text{ (S128p)}

\text{ (S128q)}

\text{ (S128r)}

\text{ (S128s)}

\text{ (S128t)}
After some straightforward algebra, one can eliminate all $g_i$ except for $g_6$ and $g_8$, which satisfy coupled polynomial equations. Those equations can be shown to be identical to eqn. (S48) by invoking the change of variables,

$$g_6 = -\sqrt{\zeta} \psi \tau_2, \quad \text{and} \quad g_8 = (\gamma + \sigma_\mathcal{W}_2^2 (\eta' - \zeta)) \tau_1. \quad (S129)$$

In terms of these variables, the error $E_{21}$ is given by,

$$E_{21} = 2(\tau_2/\tau_1 - 1). \quad (S130)$$

### S4.3.2. $E_{22}$

The equations satisfied by the operator-valued Stieltjes transform $G$ of $\tilde{Q}_{22}$ induce the following structure on $G$.

$$G = \begin{pmatrix} 0 & G_{12} \\ G_{12}^\top & 0 \end{pmatrix}, \quad (S131)$$

where,

$$G_{12} = \begin{pmatrix} g_{11} & 0 & 0 & 0 & 0 & \cdots & g_7 \\ 0 & g_5 & g_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & g_8 & 0 & 0 & \cdots & 0 \\ g_3 & 0 & g_4 & 0 & g_6 & \cdots & 0 \\ 0 & 0 & 0 & 0 & g_{13} & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & g_{12} \end{pmatrix}, \quad (S132)$$

and the independent entry-wise component functions $g_i$ combine to produce the error $E_{22}$ through the relation,

$$E_{22} = \frac{2\sqrt{\zeta}g_6 \sigma_\mathcal{W}_2^2 (\psi (\eta - \zeta) + \zeta \sigma_\mathcal{W}_2^2)}{\sqrt{n_0 \psi}} + 2g_6 (\eta - \zeta) \sigma_\mathcal{W}_2^2, \quad (S133)$$

and themselves satisfy the following system of polynomial equations,

$$0 = \sqrt{\zeta} g_{11} g_{13} \sqrt{n_0} g_{14} \psi - g_{14} \psi \quad (S134a)$$
$$0 = \sqrt{\zeta} g_{12} g_{13} \sqrt{n_0} - g_{12} \psi + \psi \quad (S134b)$$
$$0 = g_{11} \psi (g_{12} \sqrt{n_0} - \sqrt{\zeta} g_{11} - g_{13} \sqrt{n_0} \sigma_\mathcal{W}_2^2) \quad (S134c)$$
$$0 = -g_1 \psi (\sqrt{\zeta} g_1 + g_3 \sqrt{n_0}) - g_1 \sqrt{n_0} \sigma_\mathcal{W}_2^2 \quad (S134d)$$
$$0 = g_1 \psi (g_5 \sqrt{n_0} - \sqrt{\zeta} g_2) - \sqrt{\zeta} g_1 \sigma_\mathcal{W}_2^2 \quad (S134e)$$
$$0 = g_1 \psi (\sqrt{\zeta} g_2 + g_4 \sqrt{n_0}) - \sqrt{\zeta} g_2 \sigma_\mathcal{W}_2^2 \quad (S134f)$$
$$0 = g_1 \psi (g_7 \sqrt{n_0} - \sqrt{\zeta} g_2) - (g_5 - 1) \sqrt{n_0} \sigma_\mathcal{W}_2^2 \quad (S134g)$$
$$0 = g_1 \psi (g_3 \sqrt{n_0} - \sqrt{\zeta} g_2) - (g_2 - 1) \sqrt{n_0} \sigma_\mathcal{W}_2^2 \quad (S134h)$$
$$0 = g_1 \psi (\sqrt{\zeta} g_5 + g_3 \sqrt{n_0}) - \sqrt{\zeta} (g_5 - 1) \sigma_\mathcal{W}_2^2 \quad (S134i)$$
$$0 = g_1 \psi (\sqrt{\zeta} g_5 + g_3 \sqrt{n_0}) - \sqrt{\zeta} (g_5 - 1) \sigma_\mathcal{W}_2^2 \quad (S134j)$$
$$0 = -g_1 \psi (\sqrt{\zeta} g_{10} g_{11} \psi - g_7 \sqrt{n_0} \phi (\gamma + \sigma_\mathcal{W}_2^2 (\eta' - \zeta))) \quad (S134k)$$
$$0 = -g_1 \psi (\sqrt{\zeta} g_{10} g_{12} \psi - g_7 \sqrt{n_0} \phi (\gamma + \sigma_\mathcal{W}_2^2 (\eta' - \zeta))) \quad (S134l)$$
$$0 = -g_1 \psi (\sqrt{\zeta} g_{10} g_{14} \psi - g_7 \sqrt{n_0} \phi (\gamma + \sigma_\mathcal{W}_2^2 (\eta' - \zeta))) \quad (S134m)$$
$$0 = g_1 (\sqrt{\zeta} (g_6 - \sqrt{\zeta} g_4 \sqrt{n_0} g_{13} + g_3 \sqrt{n_0}) - \sqrt{\zeta} (g_5 - 1) \sqrt{n_0} \sigma_\mathcal{W}_2^2) \quad (S134n)$$
$$0 = g_1 \psi (\sqrt{\zeta} g_6 + g_8 \sqrt{n_0}) - \sqrt{\zeta} g_7 \sigma_\mathcal{W}_2^2 + g_8 g_{13} \sqrt{n_0} \psi \quad (S134o)$$
$$0 = g_1 \psi (\sqrt{\zeta} g_6 + g_8 \sqrt{n_0}) - \sqrt{\zeta} g_7 \sigma_\mathcal{W}_2^2 + g_8 g_{13} \sqrt{n_0} \psi \quad (S134p)$$
$$0 = g_1 \psi (\sqrt{\zeta} g_6 + g_8 \sqrt{n_0}) - \sqrt{\zeta} g_7 \sigma_\mathcal{W}_2^2 + g_8 g_{13} \sqrt{n_0} \psi \quad (S134q)$$
$$0 = \sqrt{\zeta} g_{12} g_{13} \sqrt{n_0} \phi (\sigma_\mathcal{W}_2^2 (\zeta - \eta') - \gamma) + g_{14} \psi (\gamma \phi + \sigma_\mathcal{W}_2^2 (-\zeta \phi + \phi \eta' + g_{10})) \quad (S134r)$$
After some straightforward algebra, one can eliminate all $g_i$ except for $g_7$ and $g_{10}$, which satisfy coupled polynomial equations. Those equations can be shown to be identical to eqn. (S48) by invoking the change of variables, 

$$g_7 = \frac{-\sqrt{\psi}}{\sqrt{\tau_0}} \tau_2, \quad \text{and} \quad g_{10} = (\gamma + \sigma^2_W (\eta' - \zeta)) \tau_1. \quad (S135)$$

The error $E_{22}$ is then given by,

$$E_{22} = 2\zeta \left( \frac{\tau_2}{\tau_1} - 1 \right) + \frac{2\psi(\zeta(t_2 - t_1) + \eta_1)^2((t_2 - t_1)\phi + \zeta_1 t_2 \sigma^2_W)}{\zeta^2 \tau_2 \phi}. \quad (S136)$$

S4.3.3. $E_{31}$

The equations satisfied by the operator-valued Stieltjes transform $G$ of $\tilde{Q}_3$ induce the following structure on $G$,

$$G = \begin{pmatrix} 0 & G_{12} \\ G_{12}^T & 0 \end{pmatrix}, \quad (S137)$$

where,

$$G_{12} = \begin{pmatrix} g_5 & 0 & 0 & 0 & 0 & g_2 & 0 & 0 \\ 0 & g_6 & 0 & g_1 & g_3 & 0 & 0 & g_4 \\ 0 & 0 & g_8 & 0 & 0 & 0 & g_{12} & 0 \\ 0 & g_{11} & 0 & g_7 & g_{10} & 0 & 0 & g_9 \\ 0 & 0 & 0 & 0 & g_6 & 0 & 0 & g_1 \\ 0 & 0 & 0 & 0 & 0 & g_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_8 & 0 \\ 0 & 0 & 0 & 0 & 0 & g_{11} & 0 & g_7 \end{pmatrix}, \quad (S138)$$

and the independent entry-wise component functions $g_i$ give the error $E_{31}$ through the relation,

$$E_{31} = \frac{g_{21} \eta_0 \eta_2^2}{\phi(\gamma + \sigma^2_W (\eta' - \zeta))}, \quad (S139)$$

and themselves satisfy the following system of polynomial equations,

$$0 = \sqrt{\zeta} g_6 g_8 \sqrt{\eta_0} - g_{11} \psi \quad \text{(S140a)}$$

$$0 = \sqrt{\zeta} g_1 g_8 \sqrt{\eta_0} - g \psi + \psi \quad \text{(S140b)}$$
After some straightforward algebra, one can eliminate all $g_i$ except for $g_1$ and $g_5$, which satisfy coupled polynomial equations. Those equations can be shown to be identical to eqn. (S48) by invoking the change of variables,

$$ g_1 = -\frac{\sqrt{g_5}}{\sqrt{n_0}} \tau_2, \quad \text{and} \quad g_5 = (\gamma + \sigma^2_{W_2}(\eta' - \zeta)) \tau_1. $$

The error $E_{31}$ can then be written in terms of $\tau_1$ and its derivative $\tau'_1$ (S69),

$$ E_{31} = \sigma^2_{g}( - \tau'_1/\tau_1^2 - 1 ). $$

S4.3.4. $E_{32}$

The equations satisfied by the operator-valued Stieltjes transform $G$ of $Q_{32}$ induce the following structure on $G$,

$$ G = \begin{pmatrix} 0 & G_{12} \\ G_{12}^* & 0 \end{pmatrix}, $$

(S143)
where,

$$
G_{12} = \begin{pmatrix}
g_9 & 0 & 0 & 0 & 0 & 0 & g_6 & 0 & 0 \\
g_1 & g_3 & g_4 & g_7 & 0 & 0 & g_2 & 0 & 0 \\
g_0 & g_{10} & g_4 & g_{13} & 0 & 0 & g_5 & 0 & 0 \\
g_0 & 0 & g_{12} & 0 & 0 & 0 & g_{16} & 0 & 0 \\
g_0 & 0 & g_{15} & g_{11} & g_{14} & 0 & 0 & g_8 & 0 \\
g_0 & 0 & 0 & 0 & 0 & 0 & g_{10} & 0 & 0 \\
g_0 & 0 & 0 & 0 & 0 & 0 & g_4 & 0 & 0 \\
g_0 & 0 & 0 & 0 & 0 & 0 & g_{12} & 0 & 0 \\
g_0 & 0 & 0 & 0 & 0 & 0 & g_{15} & 0 & 0 \\
g_0 & 0 & 0 & 0 & 0 & 0 & g_{11} & 0 & 0 \\
\end{pmatrix},
$$

(S144)

and the independent entry-wise component functions $g_i$ give the error $E_{32}$ through the relation,

$$
E_{32} = -g_2n_0^{3/2}/(\sqrt[3]{\phi}),
$$

(S145)

and themselves satisfy the following system of polynomial equations,

$$
0 = \sqrt[3]{\phi}g_{10}g_{12}\sqrt{n_0} - g_{15}\phi 
$$

(S146a)

$$
0 = \sqrt[3]{\phi}g_{12}\sqrt{n_0} - g_{11}\phi + \phi 
$$

(S146b)

$$
0 = -\sqrt[3]{\phi}g_{10}\phi - g_{14}\sqrt{n_0}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146c)

$$
0 = -\sqrt[3]{\phi}g_{15}\phi - (g_{11} - 1)\sqrt{n_0}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146d)

$$
0 = -\sqrt[3]{\phi}g_{12}\phi - \sqrt[3]{\phi}g_{13}\phi - g_{4}\sqrt{n_0}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146e)

$$
0 = -\sqrt[3]{\phi}g_{14}\phi - \sqrt[3]{\phi}g_{15}\phi - g_{4}\sqrt{n_0}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146f)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146g)

$$
0 = g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) + g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146h)

$$
0 = g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) + g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146i)

$$
0 = (g_{10} - 1)\sqrt{n_0}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) + g_{12}(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146j)

$$
0 = -\sqrt[3]{\phi}(g_{11} + g_{12})\phi - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) + g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146k)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146l)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146m)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146n)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146o)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146p)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146q)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146r)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146s)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146t)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146u)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146v)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146w)

$$
0 = \sqrt[3]{\phi}g_{12}\phi g_{10}(\phi + \phi - \zeta) - g_{12}\phi(\gamma + \sigma_W^2(\eta' - \zeta)) 
$$

(S146x)
After some straightforward algebra, one can eliminate all $g_i$ except for $g_4$ and $g_9$, which satisfy coupled polynomial equations. Those equations can be shown to be identical to eqn. (S48) by invoking the change of variables,

$$g_4 = -\frac{\sqrt{\zeta g}}{\sqrt{\psi}} \tau_2, \quad \text{and} \quad g_9 = (\gamma + \sigma_{W_2}(\eta' - \zeta)) \tau_1.$$  

(S47)

In terms of $\tau_1$, $\tau_2$, and $\tau_2'$ (S70), the error $E_{32}$ is given by,

$$E_{32} = 1 - 2\tau_2/\tau_1 - \tau_2'/\tau_1.$$  

(S48)

S4.3.5. $E_{33}$

The equations satisfied by the operator-valued Stieltjes transform $G$ of $\tilde{Q}_{32}$ induce the following structure on $G$,

$$G = \begin{pmatrix} 0 & G_{12} \\ G_{12}^T & 0 \end{pmatrix},$$  

(S49)

where,

$$G_{12} = \begin{pmatrix} g_{13} & 0 & 0 & 0 & 0 & 0 & g_8 & 0 & 0 & 0 \\ 0 & g_1 & 0 & 0 & 0 & 0 & g_5 & 0 & 0 & 0 \\ 0 & 0 & g_1 & 0 & g_4 & 0 & g_6 & g_9 & 0 & 0 \\ 0 & 0 & 0 & g_{14} & 0 & g_6 & g_{17} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_{16} & 0 & 0 & 0 & g_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & g_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_{14} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_1 & 0 & g_5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{16} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{15} \end{pmatrix},$$  

(S50)

and the independent entry-wise component functions $g_i$ give the error $E_{32}$ through the relation,

$$E_{33} = -g_{32}n_0\sigma_{W_2}^2/\psi,$$  

(S51)

and themselves satisfy the following system of polynomial equations,

$$0 = \sqrt{\zeta g_{14}g_{16}}\sqrt{\mu_0} - g_{19}\psi$$  

(S52a)

$$0 = \sqrt{\zeta g_9g_{15}}\psi + \psi$$  

(S52b)

$$0 = -\sqrt{\zeta g_{13}g_{14}}\psi - g_9\sqrt{\mu_0}\phi(\gamma + \sigma_{W_2}(\eta' - \zeta))$$  

(S52c)

$$0 = -\sqrt{\zeta g_{13}g_{19}}\psi - (g_{15} - 1)\sqrt{\mu_0}\phi(\gamma + \sigma_{W_2}(\eta' - \zeta))$$  

(S52d)

$$0 = -\sqrt{\zeta g_{13}g_{19}}\psi - \sqrt{\zeta g_{13}g_{14}}\psi - g_e\sqrt{\mu_0}\phi(\gamma + \sigma_{W_2}(\eta' - \zeta))$$  

(S52e)

$$0 = -\sqrt{\zeta g_{13}g_{18}}\psi - \sqrt{\zeta g_{13}g_{19}}\psi - g_{10}\sqrt{\mu_0}\phi(\gamma + \sigma_{W_2}(\eta' - \zeta))$$  

(S52f)

$$0 = g_{13}g_{16}\psi(\zeta - \eta) - \phi(g_5 - \sqrt{\zeta g_{16}g_{16}}\sqrt{\mu_0})(\gamma + \sigma_{W_2}(\eta' - \zeta))$$  

(S52g)

$$0 = g_{13}g_{16}\psi(\zeta - \eta) - \phi(g_5 - \sqrt{\zeta g_{9}g_{16}}\sqrt{\mu_0})(\gamma + \sigma_{W_2}(\eta' - \zeta))$$  

(S52h)
The Neural Tangent Kernel in High Dimensions: Triple Descent and a Multi-Scale Theory of Generalization
After some straightforward algebra, one can eliminate all \( g_i \) except for \( g_6 \) and \( g_{13} \), which satisfy coupled polynomial equations. Those equations can be shown to be identical to eqn. (S48) by invoking the change of variables,

\[
g_6 = -\frac{\sqrt{\psi \phi}}{\sqrt{n_0}} \tau_2, \quad \text{and} \quad g_{13} = (\gamma + \sigma_{W_2}^2 (n' - \zeta)) \tau_1. \tag{S153}
\]

In terms of \( \tau_1, \tau_2 \), and their derivatives \( \tau'_1 \) (S69), \( \tau'_2 \) (S70), the error \( E_{33} \) is given by,

\[
E_{33} = \sigma_{W_2}^2 \left[ (\tau_1 + (\sigma_{W_2}^2 (n' - \zeta) + \gamma) \tau'_1 + \sigma_{W_2}^2 (n' - \zeta)) / \tau_1^2 - \eta \right] - E_{22}. \tag{S154}
\]

### S4.4 Total test error

Recall from eqns. (S72, S91-S93) that the total test error can be written as

\[
E_{\text{test}} = 1 + E_{21} + E_{31} + E_{32} + \nu (\eta \sigma_{W_2}^2 + E_{22} + E_{33}), \tag{S155}
\]

where \( \nu = 0 \) with centering and \( \nu = 1 \) without it. Combining the results from the previous subsections, we find

\[
E_{\text{test}} = 1 + 2(\tau_2/\tau_1 - 1) + \sigma_{W_2}^2 \left( \tau'_1 / \tau_1^2 - 1 \right) + 2 \tau_2 / \tau_1 + \tau'_2 / \tau_1^2
\]

\[
+ \nu \sigma_{W_2}^2 \left[ (\tau_1 + (\sigma_{W_2}^2 (n' - \zeta) + \gamma) \tau'_1 + \sigma_{W_2}^2 (n' - \zeta)) / \tau_1^2 \right]
\]

\[
= \tau'_2 / \tau_1 + \sigma_{W_2}^2 / \tau_1 + \nu \sigma_{W_2}^2 \left[ (\tau_1 + (\sigma_{W_2}^2 (n' - \zeta) + \gamma) \tau'_1 + \sigma_{W_2}^2 (n' - \zeta)) / \tau_1^2 \right] - \sigma_{W_2}^2
\]

\[
= (\gamma \tau_1)^{-2} E_{\text{train}} - \sigma_{W_2}^2,
\]

thereby establishing the result of the main theorem (27).