Contents

A Proofs for Ridge Regression

In this section we provide the calculations associated to ridge regression. Section A.1 provides the proof of Proposition 1. Section A.2 provides the calculation for the oracle estimator presented in remark 1. Section A.3 provides some preliminary calculations related to random matrix theory. Section A.4 gives the proof of Theorem 1. Section A.5 provides the proof of Corollary 1. Section A.6 provides the calculations associated to the strong and weak features model.

A.1 Proof of Proposition 1

In the proof of this result, it is useful to indicate dependence on the true parameter $\beta^\star$ by denoting the risk $R_{\beta^\star}(\cdot) = \mathbf{E}_\epsilon[\|\Sigma^{1/2}(\cdot - \beta^\star)\|_2^2] + \sigma^2$. We also denote by $\mathcal{E}_{\beta^\star}(\beta) = R_{\beta^\star}(\beta) - R_{\beta^\star}(\beta^\star) = \|\Sigma^{1/2}(\beta - \beta^\star)\|_2^2$ the excess risk of $\beta \in \mathbb{R}^d$ when the true parameter is $\beta^\star \in \mathbb{R}^d$.

Lemma 1 *Let $V_1, \dots, V_k \subset \mathbb{R}^d$ (with $k \le d$) denote the eigenspaces of Σ. For $j = 1, \dots, k$, let $U_j \in O(V_j)$ be a linear isometry of V_j, and let $U \in \mathbb{R}^{d\times d}$ be the linear isometry defined by $Uv = U_j v$ for $v \in V_j$, $j = 1, \dots, k$. Then,*

$$\mathbf{E}_{X,\epsilon}[R_\beta(\widehat{\beta}_\lambda) - R_\beta(\beta)] = \mathbf{E}_{X,\epsilon}[R_{U\beta}(\widehat{\beta}_\lambda) - R_{U\beta}(U\beta)]\,.$$

Proof 1 (Proof of Lemma 1) *Denote $\beta' = U\beta$, as well as $X' = XU^{-1}$ and $x' = U^{-1}x$. Let $\widehat{\beta}'_\lambda$ the Ridge estimator computed on data (X', Y), namely*

$$\widehat{\beta}'_\lambda = (X'^\top X' + \lambda n I)^{-1} X'^\top Y = (UX^\top X U^{-1} + \lambda n I)^{-1} U X^\top Y = U\widehat{\beta}_\lambda\,.$$

Then, $y = \langle \beta, x\rangle + \sigma\epsilon = \langle \beta', x'\rangle + \sigma\epsilon$, hence the best linear predictor of y based on x' is β'. In addition, x' has distribution $\mathcal{N}(0, U^{-1}\Sigma U) = \mathcal{N}(0, \Sigma)$, where $U^{-1}\Sigma U = \Sigma$ comes from the fact that U is an isometry on the eigenspaces V_j of Σ. This implies that (X', ϵ) has the same distribution as (X, ϵ), and thus $\mathbf{E}_{\epsilon,X}[\mathcal{E}_{\beta'}(\widehat{\beta}'_\lambda)] = \mathbf{E}_{\epsilon,X}[\mathcal{E}_{\beta'}(\widehat{\beta}_\lambda)]$. On the other hand,

$$\mathcal{E}_{\beta'}(\widehat{\beta}'_\lambda) = \|\Sigma^{1/2}(\widehat{\beta}'_\lambda - \beta')\|_2^2 = \|\Sigma^{1/2}U(\widehat{\beta}_\lambda - \beta)\|_2^2 = \|\Sigma^{1/2}(\widehat{\beta}_\lambda - \beta)\|_2^2 = \mathcal{E}_\beta(\widehat{\beta}_\lambda)$$

(note that $\|\Sigma^{1/2}U \cdot\|_2^2 = \|U\Sigma^{1/2}\cdot\|_2^2 = \|\Sigma^{1/2}\cdot\|_2^2$ as U commutes with $\Sigma^{1/2}$ and is an isometry), so that $\mathbf{E}_{\epsilon,X}[\mathcal{E}_{\beta'}(\widehat{\beta}'_\lambda)] = \mathbf{E}_{\epsilon,X}[\mathcal{E}_\beta(\widehat{\beta}_\lambda)]$. This proves that $\mathbf{E}_{\epsilon,X}[\mathcal{E}_{\beta'}(\widehat{\beta}_\lambda)] = \mathbf{E}_{\epsilon,X}[\mathcal{E}_\beta(\widehat{\beta}_\lambda)]$.

We now turn to the proof of Proposition 1:

Proof 2 (Proof of Proposition 1) *Let $V_1, \dots, V_k$ denote the eigenspaces of Σ, with distinct eigenvalues $\tau'_1 > \cdots > \tau'_k$. Let $U_1, \dots, U_k$ be independent random isometries, where U_j is distributed according to the uniform (Haar) measure on the orthogonal group of V_j. Define U to be the random isometry acting as U_j on V_j, and let $\beta = U\beta^\star$ and Π its distribution.*

Note that U is of the form of Lemma 1, hence $\mathbf{E}_{X,\epsilon}[\mathcal{E}_{U\beta^\star}(\widehat{\beta}_\lambda)] = \mathbf{E}_{X,\epsilon}[\mathcal{E}_{\beta^\star}(\widehat{\beta}_\lambda)]$ and thus

$$\mathbf{E}_{\beta\sim\Pi}\mathbf{E}_{X,\epsilon}[\mathcal{E}_\beta(\widehat{\beta}_\lambda)] = \mathbf{E}_U\mathbf{E}_{X,\epsilon}[\mathcal{E}_{U\beta^\star}(\widehat{\beta}_\lambda)] = \mathbf{E}_{X,\epsilon}[\mathcal{E}_{\beta^\star}(\widehat{\beta}_\lambda)]\,. \tag{8}$$

Now, let $\beta'_j \in V_j$ be the orthogonal projection of $\beta^\star$ on V_j, so that $U\beta^\star = \sum_{j=1}^k U_j\beta'_j$. We have $\mathbf{E}[U\beta^\star] = 0$ since $\mathbf{E}[U_j] = 0$ for all j. In addition, the distribution of $U_j\beta^\star$ is invariant by rotation (since R_jU_j has the same distribution as U_j for any fixed rotation R_j), hence $\mathbf{E}[(U_j\beta'_j)(U_j\beta'_j)^\top] = t_j I_{V_j}$ (with I_{V_j} the identity on V_j), where letting $d_j = \dim(V_j)$,

$$d_j \cdot t_j = \mathrm{tr}\mathbf{E}[(U_j\beta'_j)(U_j\beta'_j)^\top] = \mathbf{E}[\|U_j\beta'_j\|_2^2] = \|\beta'_j\|_2^2\,, \tag{9}$$

hence $t_j = \|\beta'_j\|_2^2/d_j$. In addition, if $j \neq l$, by independence of U_j, U_l,

$$\mathbf{E}[(U_j\beta'_j)(U_l\beta'_l)^\top] = \mathbf{E}[U_j\beta'_j\beta'^\top_l U_l^\top] = \mathbf{E}[U_j]\beta'_j\beta'^\top_l\mathbf{E}[U_l]^\top = 0\,. \tag{10}$$

Hence, Π has covariance $\sum_{j=1}^k(\|\beta'_j\|^2/d_j)I_{V_j}$, which is precisely $\Phi(\Sigma)/d$ where Φ is defined as in (5). The proof is concluded by noting that the quantity $\mathbf{E}_{\epsilon,X}\mathcal{E}_\beta(\widehat{\beta}_\lambda)$ is quadratic in β, hence if Π' is another distribution on $\mathbb{R}^d$ with mean 0 and covariance $\Phi(\Sigma)/d$, then $\mathbf{E}_{\beta\sim\Pi'}\mathbf{E}_{\epsilon,X}\mathcal{E}_\beta(\widehat{\beta}_\lambda) = \mathbf{E}_{\beta\sim\Pi}\mathbf{E}_{\epsilon,X}\mathcal{E}_\beta(\widehat{\beta}_\lambda) = \mathbf{E}_{\epsilon,X}\mathcal{E}_{\beta^\star}(\widehat{\beta}_\lambda)$.

A.2 Proof of Oracle Estimator (Remark 1)

Since the risk is quadratic, the average risk (integrated over the prior) of any estimator linear in Y only depends on the first two moments of the prior, hence one can assume that the prior is Gaussian (namely, $\mathcal{N}(0, r^2\Phi(\Sigma)/d)$) without loss of generality. In this case, a standard computation shows that the posterior is $\mathcal{N}(\widetilde{\beta}, [X^\top X + (\sigma^2 d/r^2)\Phi(\Sigma)^{-1}]^{-1})$, where $\widetilde{\beta}$ is the estimator defined in (4). Finally, since the risk is quadratic, the Bayes-optimal estimator is the posterior mean, which corresponds to $\widetilde{\beta}$.

A.3 Random Matrix Theory Preliminaries

We now introduce some useful properties of the Stieltjes transform as well as its companion transform. Firstly, we know the companion transform satisfies the Silverstein equation (Silverstein and Combettes, 1992; Silverstein and Choi, 1995)

$$-\frac{1}{v(z)} = z - \gamma \int \frac{\tau}{1+\tau v(z)} dH(\tau). \tag{11}$$

We then have for $z \in \mathcal{S} := \{u + iv : v \neq 0, \text{ or } v = 0, u > 0\}$, the companion transform $v(z)$ is the unique solution to the Silverstein equation with $v(z) \in \mathcal{S}$ such that the sign of the imaginary part is preserved $\text{sign}(\text{Im}(v(z))) - \text{sign}(\text{Im}(z))$. The above can then be differentiated with respect to z to obtain a formula for $v'(z)$ in terms of $v(z)$:

$$\frac{\partial v(z)}{\partial z} = \Big(\frac{1}{v(z)^2} - \gamma \int \frac{\tau^2}{(1+\tau v(z))^2} dH(\tau)\Big)^{-1}$$

Meanwhile from from the equality $\gamma(m(z) + 1/z) = v(z) + 1/z$ we note that we have the following equalities

$$\begin{aligned} 1 - \gamma(1 - \lambda m(-\lambda)) &= \lambda v(-\lambda) \\ 1 - \lambda m(-\lambda) &= \gamma^{-1}(1 - \lambda v(-\lambda)) \\ m(-\lambda) - \lambda m'(-\lambda) &= \gamma^{-1}(v(-\lambda) - \lambda v'(-\lambda)) \end{aligned} \tag{12}$$

which we will readily use to simplify/rewrite a number of the limiting functions.

A.4 Proof of Theorem 1

We begin with the decomposition into bias and variance terms following (Dobriban et al., 2018). The difference for the ridge parameter can be denoted

$$\widehat{\beta}_\lambda - \beta^\star = -\lambda\big(\frac{X^\top X}{n} + \lambda I\big)^{-1}\beta^\star + \sigma\big(\frac{X^\top X}{n} + \lambda I\big)^{-1}\frac{X^\top \epsilon}{n}$$

And thus taking expectation with respect to the noise in the observations ϵ

$$\begin{aligned} \mathbf{E}_\epsilon[R(\widehat{\beta}_\lambda)] - R(\beta^\star) &= \mathbf{E}_\epsilon[\|\Sigma^{1/2}(\widehat{\beta}_\lambda - \beta^\star)\|_2^2] \\ &= \mathbf{E}_\epsilon[\|\Sigma^{1/2}(\widehat{\beta}_\lambda - \mathbf{E}_\epsilon[\widehat{\beta}_\lambda])\|_2^2] + \|\Sigma^{1/2}(\mathbf{E}_\epsilon[\widehat{\beta}_\lambda] - \beta^\star)\|_2^2 \\ &= \sigma^2 \mathbf{E}_\epsilon\big[\big\|\Sigma^{1/2}\big(\frac{X^\top X}{n} + \lambda I\big)^{-1}\frac{X^\top \epsilon}{n}\big\|_2^2\big] + \lambda^2\|\Sigma^{1/2}\big(\frac{X^\top X}{n} - \lambda I\big)^{-1}\beta^\star\|_2^2 \\ &= \frac{\sigma^2}{n}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Sigma\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\frac{X^\top X}{n}\Big) \\ &\quad \lambda^2 \operatorname{Tr}\Big((\beta^\star)^\top\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Sigma\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\beta^\star\Big) \end{aligned}$$

Taking expectation with respect to $\mathbf{E}_{\beta^\star}$ we arrive at

$$\begin{aligned} \mathbf{E}_{\beta^\star}[\mathbf{E}_\epsilon[R(\widehat{\beta}_\lambda)] - R(\beta^\star)] &= \frac{\sigma^2}{n}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Sigma\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\frac{X^\top X}{n}\Big) \\ &\quad \frac{\lambda^2 r^2}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Sigma\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Phi(\Sigma)\Big) \\ &= \sigma^2\gamma\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Sigma\Big) - \lambda\sigma^2\gamma\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-2}\Sigma\Big) \\ &\quad + \frac{\lambda^2 r^2}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Sigma\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Phi(\Sigma)\Big) \end{aligned}$$

It is now a matter of showing the asymptotic almost sure convergence of the following three functionals

$$\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-1}\Sigma\Big),\quad \frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-2}\Sigma\Big)$$
$$\text{and } \frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-1}\Sigma\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-1}\Phi(\Sigma)\Big)$$

The limit of the first trace quantity comes directly from (Ledoit and Péché, 2011) meanwhile the limit of the second trace quantity is proven in (Dobriban et al., 2018). The third trace quantity depends upon the source condition Φ and computing its limit is one of the main technical contributions of this work. The limits for these objects is summarised within the following Lemma, the proof of which provides the key steps for computing the limit involving the source function.

Lemma 2 *Under the assumptions of Theorem 1 for any $\lambda > 0$ we have almost surely as $n, d \to \infty$ with $d/n = \gamma$*

$$\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-1}\Sigma\Big) \to \frac{1-\lambda m(-\lambda)}{1-\gamma(1-\lambda m(-\lambda))} \tag{13}$$
$$\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-2}\Sigma\Big) \to \frac{m(-\lambda)-\lambda m'(-\lambda)}{\big(1-\gamma(1-\lambda m(-\lambda))\big)^2} \tag{14}$$
$$\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-1}\Sigma\Big(\frac{X^\top X}{n}+\lambda I\Big)^{-1}\Phi(\Sigma)\Big) \to \frac{\Theta^\Phi(-\lambda)+\lambda\frac{\partial\Theta^\Phi(-\lambda)}{\partial\lambda}}{\big(1-\gamma(1-\lambda m(-\lambda))\big)^2} \tag{15}$$

The result is arrived at by plugging in the above limits and noting from the definition of the Companion Transform v that $1-\gamma(1-\lambda m(-\lambda)) = \lambda v(-\lambda)$, $1-\lambda m(-\lambda) = \gamma^{-1}(1-\lambda v(-\lambda))$ and, taking derivatives, $m(-\lambda)-\lambda m'(-\lambda) = \gamma^{-1}(v(-\lambda)-\lambda v'(-\lambda))$. The proof of Lemma 2, which is the key technical step in the proof of Theorem 1, is provided in Appendix B.

A.5 Proof of Corollary 1

In this section we provide the proof of Corollary 1. It will be broken into three parts associated to the three cases $\Phi(x) = x$, $\Phi(x) = 1$ and $\Phi(x) = 1/x$.

A.5.1 Case: $\Phi(x) = x$

The purpose of this section is to demonstrate, in the case $\Phi(x) = x$, how the functional $\Theta^\Phi(-\lambda)+\lambda\frac{\partial\Theta^\Phi(-\lambda)}{\partial\lambda}$ can be written in terms of the Stieltjes Transform $m(z)$. For this particular choice of Φ the asymptotics were calculated in (Chen et al., 2011), see also Lemma 7.9 in (Dobriban et al., 2018). We therefore repeat this calculation for completeness. Now, in this case we have

$$\Theta^\Phi(z) = \int \frac{\tau}{\tau(1-\gamma(1+zm(-\lambda)))-z}dH(\tau)$$

Following the steps are the start of the proof for Lemma 2.2 in (Ledoit and Péché, 2011), consider $1+zm(z)$

$$\begin{aligned}1+zm(z) &= \int 1+\frac{z}{\tau(1-\gamma(1+zm(z)))-z}dH(\tau)\\ &= \int \frac{\tau(1-\gamma(1+zm(z)))}{\tau(1-\gamma(1+zm(z)))-z}dH(\tau)\\ &= (1-\gamma(1+zm(z)))\Theta^\Phi(z)\end{aligned}$$

Solving for $\Theta^\Phi(z)$ we have

$$\Theta^\Phi(z) = \frac{1+zm(z)}{1-\gamma(1+zm(z))} = \frac{1}{\gamma}\Big(\frac{1}{1-\gamma(1+zm(z))}-1\Big)$$

Picking $z = -\lambda$ and differentiating with respect to λ we get

$$\frac{\partial \Theta^{\Phi}(-\lambda)}{\partial \lambda} = -\frac{m(-\lambda) - \lambda m'(-\lambda)}{(1-\gamma(1-\lambda m(-\lambda)))^2}$$

This leads to the final form

$$\begin{aligned}\frac{\Theta^{\Phi}(-\lambda) + \lambda \frac{\partial \Theta^{\Phi}(-\lambda)}{\partial \lambda}}{\big(1-\gamma(1-\lambda m(-\lambda))\big)^2} &= \frac{1-\lambda m(-\lambda)}{(1-\gamma(1-\lambda m(-\lambda)))^3} - \lambda \frac{m(-\lambda) - \lambda m'(-\lambda)}{(1-\gamma(1-\lambda m(-\lambda)))^4} \\ &= \frac{\gamma^{-1}(1-\lambda v(-\lambda))}{(\lambda v(-\lambda))^3} - \lambda \frac{\gamma^{-1}(v(-\lambda) - \lambda v'(-\lambda))}{(\lambda v(-\lambda))^4} \\ &= \frac{v'(-\lambda)}{\gamma \lambda^2 v(-\lambda)^4} - \frac{1}{\gamma(\lambda v(-\lambda))^2}\end{aligned}$$

where on the second equality we used (12). Multiplying through by λ^2 then yields the quantity presented.

A.5.2 Case: $\Phi(x) = 1$

The functional of interest in this case aligns with that calculated within (Dobriban et al., 2018), which we include below for completeness. In particular we have $\Theta^{\Phi}(-\lambda) = m(-\lambda)$ and as such we get

$$\Theta^{\Phi}(-\lambda) + \lambda \frac{\partial \Theta^{\Phi}(-\lambda)}{\partial \lambda} = m(-\lambda) - \lambda m'(-\lambda) = \gamma^{-1}(v(-\lambda) - \lambda v'(-\lambda))$$

where on the second equality we used (12). Dividing by $v(-\lambda)^2$ as well as adding the asymptotic variance we get, from Theorem 1, the limit as $n, d \to \infty$

$$\begin{aligned}\mathbf{E}_{\beta^\star}[\mathbf{E}_{\epsilon}[R(\widehat{\beta}_{\lambda})] - R(\beta^\star)] &\to \sigma^2 \frac{1-\lambda v(-\lambda)}{\lambda v(-\lambda)} - \lambda \sigma^2 \frac{v(-\lambda) - \lambda v'(-\lambda)}{(\lambda v(-\lambda))^2} + \frac{r^2}{\gamma} \frac{v(-\lambda) - \lambda v'(-\lambda)}{v(-\lambda)^2} \\ &= \sigma^2 \Big(\frac{v'(-\lambda)}{(v(-\lambda))^2} - 1 \Big) + \frac{r^2}{\gamma v(-\lambda)} - \frac{r^2 \lambda}{\gamma} \frac{v'(-\lambda)}{v(-\lambda)^2}\end{aligned}$$

A.5.3 Case: $\Phi(x) = 1/x$

The functional in the case $\Phi(x) = 1/x$ takes the form

$$\Theta^{\Phi}(z) = \int \frac{1}{\tau} \frac{1}{\tau(1-\gamma(1+zm(z))) - z} dH(\tau).$$

Observe that we have

$$\begin{aligned}\int \frac{1}{\tau} dH(\tau) + z\Theta^{\Phi}(z) &= \int \frac{1}{\tau} \Big(1 + \frac{z}{\tau(1-\gamma(1+zm(z))) - z} \Big) dH(\tau) \\ &= \int \frac{1}{\tau} \frac{\tau(1-\gamma(1+zm(z)))}{\tau(1-\gamma(1+zm(z))) - z} dH(\tau) \\ &= (1-\gamma(1+zm(z))) \int \frac{1}{\tau(1-\gamma(1+zm(z))) - z} dH(\tau) \\ &= (1-\gamma(1+zm(z)))m(z).\end{aligned}$$

Solving for $\Theta^{\Phi}(z)$ and plugging in the definition of the companion transform $v(z)$ we arrive at

$$\begin{aligned}\Theta^{\Phi}(z) &= \frac{1}{z}\big((1-\gamma(1+zm(z)))m(z) - \frac{1}{z}\int \frac{1}{\tau} dH(\tau)\big) \\ &= -v(z)\big(\frac{v(z)}{\gamma} + \frac{1}{z}(\frac{1}{\gamma} - 1)\big) - \frac{1}{z}\int \frac{1}{\tau} dH(\tau) \\ &= -\frac{v(z)^2}{\gamma} - \frac{v(z)}{z}(\frac{1}{\gamma} - 1) - \frac{1}{z}\int \frac{1}{\tau} dH(\tau).\end{aligned}$$

Fixing $z = -\lambda$ the quantity of interest then has the form

$$\Theta^{\Phi}(-\lambda) = -\frac{v(-\lambda)^2}{\gamma} + \frac{v(-\lambda)}{\lambda}\big(\frac{1}{\gamma} - 1\big) + \frac{1}{\lambda}\int \frac{1}{\tau} dH(\tau),$$

which when differentiated with respect to λ yields

$$\frac{\partial \Theta^{\Phi}(-\lambda)}{\partial \lambda} = \frac{2v(-\lambda)v'(-\lambda)}{\gamma} - \frac{1}{\lambda}\big(\frac{1}{\gamma} - 1\big)\big(\frac{v(-\lambda)}{\lambda} + v'(-\lambda)\big) - \frac{1}{\lambda^2}\int \frac{1}{\tau} dH(\tau).$$

Multiplying the above by λ and adding $\Theta^{\Phi}(-\lambda)$ brings us to

$$\Theta^{\Phi}(-\lambda) + \lambda \frac{\partial \Theta^{\Phi}(-\lambda)}{\partial \lambda} = 2\lambda \frac{v'(-\lambda)v(-\lambda)}{\gamma} - \frac{v(-\lambda)^2}{\gamma} - \big(\frac{1}{\gamma} - 1\big)v'(-\lambda).$$

Dividing the above by $v(-\lambda)^2$ and adding the limiting variance yields, from Theorem 1, the limit as $n, d \to \infty$

$$\begin{aligned}
&\mathbf{E}_{\beta^\star}[\mathbf{E}_{\epsilon}[R(\widehat{\beta}_\lambda)] - R(\beta^\star)] \\
&\to \sigma^2 \frac{1 - \lambda v(-\lambda)}{\lambda v(-\lambda)} - \lambda\sigma^2 \frac{v(-\lambda) - \lambda v'(-\lambda)}{(\lambda v(-\lambda))^2} + 2r^2\lambda \frac{v'(-\lambda)}{\gamma v(-\lambda)} - \frac{r^2}{\gamma} - r^2\big(\frac{1}{\gamma} - 1\big)\frac{v'(-\lambda)}{v(-\lambda)^2} \\
&= \sigma^2\big(\frac{v'(-\lambda)}{(v(-\lambda))^2} - 1\big) + 2r^2\lambda \frac{v'(-\lambda)}{\gamma v(-\lambda)} - \frac{r^2}{\gamma} + r^2\lambda \frac{\gamma - 1}{\gamma}\frac{v'(-\lambda)}{v(-\lambda)^2}
\end{aligned}$$

A.6 Strong and Weak Features Model

This section presents the calculations associated to the strong and weak features model. We begin giving the stationary point equation of the companion transform $v(t)$, after which we explicitly compute the limiting risk with the particular choice of $\Phi(x)$ in this case. Section A.6.1 there after gives explicit form for the companion transform in the ridgeless limit. Section A.6.2 gives the proof of Corollary 2 found within the main body of the manuscript.

We begin by recalling the limiting spectrum of the covariance Σ for the two Bulks Model is $dH(\tau) = \psi_1\delta_{\rho_1} + \psi_2\delta_{\rho_2}$. Recall we have $\psi_1 + \psi_2 = 1$ therefore we simply write $\psi_2 = 1 - \psi_1$. Using the Silverstein equations (11) the companion transform must satisfy

$$\frac{-1}{v(t)} = t - \gamma\Big(\frac{\psi_1\rho_1}{1 + \rho_1 v(t)} + \frac{(1 - \psi_1)\rho_2}{1 + \rho_2 v(t)}\Big), \tag{16}$$

meanwhile the derivative must satisfy

$$\begin{aligned}
&\frac{1}{(v(t))^2} = \frac{1}{v'(t)} + \gamma\Big(\frac{\psi_1\rho_1^2}{(1 + \rho_1 v(t))^2} + \frac{(1 - \psi_1)\rho_2^2}{(1 + \rho_2 v(t))^2}\Big) \\
&\implies v'(t) = \Big(\frac{1}{(v(t))^2} - \gamma\Big(\frac{\psi_1\rho_1^2}{(1 + \rho_1 v(t))^2} + \frac{(1 - \psi_1)\rho_2^2}{(1 + \rho_2 v(t))^2}\Big)\Big)^{-1}
\end{aligned} \tag{17}$$

as such given $v(t)$ we can compute the derivative. Rearranging (16) and denoting $v(t) = v$ the companion transform evaluated at t satisfies

$$\begin{aligned}
0 &= (1 + \rho_1 v)(1 + \rho_2 v) + tv(1 + \rho_1 v)(1 + \rho_2 v) - \gamma\psi_1\rho_1 v(1 + \rho_2 v) - \gamma(1 - \psi_1)\rho_2 v(1 + \rho_1 v) \\
&= t\rho_1\rho_2 v^3 + (t(\rho_1 + \rho_2) + (1 - \gamma)\rho_1\rho_2)v^2 + (t + \rho_1 + \rho_2 - \gamma\psi_1\rho_1 - \gamma(1 - \psi_1)\rho_2)v + 1
\end{aligned}$$

This cubic can then be solved computationally for different choices of t. In the case of the ridgeless limit $t \to 0$ in the overparameterised setting $\gamma > 1$, the above simplifies to a quadratic which can be solved, as shown in Section A.6.1.

Now, recall in the strong and weak features model the structure of the ground truth $\beta^\star$ is such that $\Phi(x) = \phi_1 \mathbb{1}_{x=\rho_1} + \phi_2 \mathbb{1}_{x=\rho_2}$. To compute the limiting risk, specifically the bias, we must then evaluate $\Theta^{\Phi}(-\lambda) + \lambda\frac{\partial \Theta^{\Phi}(-\lambda)}{\partial \lambda}$.

To this end, we have plugging $\Phi(x)$ into the definition of $\Theta^{\Phi}(z)$

$$\begin{aligned}\Theta^{\Phi}(z) &= \int \Phi(\tau)\frac{1}{\tau(1-\gamma(1+zm(z)))-z}dH(\tau)\\ &= \frac{\phi_1\psi_1}{\rho_1(1-\gamma(1+zm(z)))-z}+\frac{\phi_2(1-\psi_1)}{\rho_2(1-\gamma(1+zm(z)))-z}\\ &= \frac{\phi_1\psi_1}{-z(1+\rho_1 v(z))}+\frac{\phi_2(1-\psi_1)}{-z(1+\rho_2 v(z))}\end{aligned}$$

where on the last equality we used (12) to rewrite the above in terms of the companion transform. Plugging in the regularisation parameter $z=-\lambda$ we then get

$$\Theta^{\Phi}(-\lambda) = \frac{\phi_1\psi_1}{\lambda(1+\rho_1 v(-\lambda))}+\frac{\phi_1(1-\psi_1)}{\lambda(1+\rho_2 v(-\lambda))}.$$

To the end of computing $\frac{\partial\Theta^{\Phi}(-\lambda)}{\partial\lambda}$, we can differentiate the above to get

$$\frac{\partial\Theta^{\Phi}(-\lambda)}{\partial\lambda} = -\phi_1\psi_1\frac{1+\rho_1 v(-\lambda)-\lambda\rho_1 v'(-\lambda)}{\big(\lambda\rho_1 v(-\lambda)+\lambda\big)^2}-\phi_2(1-\psi_1)\frac{1+\rho_2 v(-\lambda)-\lambda\rho_2 v'(-\lambda)}{\big(\lambda\rho_2 v(-\lambda)+\lambda\big)^2},$$

which yields

$$\Theta^{\Phi}(-\lambda)+\lambda\frac{\partial\Theta^{\Phi}(-\lambda)}{\partial\lambda} = \phi_1\psi_1\frac{\rho_1 v'(-\lambda)}{(\rho_1 v(-\lambda)+1)^2}+\phi_2(1-\psi_1)\frac{\rho_2 v'(-\lambda)}{(\rho_2 v(-\lambda)+1)^2}$$

as required. The final form for the limiting risk is then

$$\begin{aligned}&\lim_{n,d\to\infty}\mathbf{E}_{\beta^\star}[\mathbf{E}_{\epsilon}[R(\widehat{\beta}_\lambda)]-R(\beta^\star)]\\ &= \sigma^2\frac{1-\lambda v(-\lambda)}{\lambda v(-\lambda)}-\lambda\sigma^2\frac{v(-\lambda)-\lambda v'(-\lambda)}{(\lambda v(-\lambda))^2}+r^2\sum_{i=1}^{2}\phi_i\psi_i\frac{\rho_i v'(-\lambda)}{(\rho_i v(-\lambda)+1)^2 v(-\lambda)^2}\\ &= -\sigma^2+\sigma^2\frac{v'(-\lambda)}{(v(-\lambda))^2}+r^2\sum_{i=1}^{2}\phi_i\psi_i\frac{\rho_i v'(-\lambda)}{(v(-\lambda)^2(\rho_i v(-\lambda)+1)^2}.\end{aligned}$$

A.6.1 Ridgeless Limit

To consider the Ridgeless limit $t\to 0$ of the companion transform $v(t)$, some care must be taken about which regime $\gamma<1$ or $\gamma>1$ we are in.

Underparameterised $\gamma<1$ Following the proof of Lemma 6.2 in (Dobriban et al., 2018) we have in the underparameterised case $\gamma<1$ the limit $\lim_{t\to 0_-} tv(t)=1-\gamma$.

Overparameterised $\gamma>1$ Following the proof of Lemma 6.2 in (Dobriban et al., 2018) when $\gamma>1$ we have the limit $\lim_{t\to 0_-} v(t)=v(0)$. From dominated convergence theorem we can take the limit in the Silverstein equation (16) to arrive at the quadratic

$$\begin{aligned}0 &= (1+\rho_1 v)(1+\rho_2 v)-\gamma\psi_1\rho_1 v(1+\rho_2 v)-\gamma(1-\psi_1)\rho_2 v(1+\rho_1 v)\\ &= (1-\gamma)\rho_1\rho_2 v^2+(\rho_1+\rho_2-\gamma\psi_1\rho_1-\gamma(1-\psi_1)\rho_2)v+1\end{aligned}$$

Solving for v with the quadratic formula immediately gives

$$v(0) = \frac{-(\rho_1+\rho_2-\gamma\psi_1\rho_1-\gamma(1-\psi_1)\rho_2)-\sqrt{(\rho_1+\rho_2-\gamma\psi_1\rho_1-\gamma(1-\psi_1)\rho_2)^2-4(1-\gamma)\rho_1\rho_2}}{2(1-\gamma)\rho_1\rho_2}. \tag{18}$$

Recall from (Silverstein and Choi, 1995) we have that $v(z)\in\mathcal{S}$, as such we take the sign above which yields a non-negative quantity. Noting we we focus on the regime where $\gamma>1$, we see for the above to be non-negative we require the numerator to be negative, and thus, we take the negative sign.

A.6.2 Proof of Corollary 2

In this section we provide the proof of Corollary 2. The proof essentially requires computing the companion transform in this case and checking the sign of the asymptotic derivative at zero i.e. $R'_{\text{Asym}}(0)$. Let us begin by noting that when $\gamma = 2$ and $\psi_1 = \psi_2 = 1/2$ that the companion transform at zero is $v(0) = 1/\sqrt{\rho_1\rho_2}$. Let us now compute quantities related to both the first derivative $v'(0)$ and second derivative $v''(0)$. Using (17), we can, by dividing both sides by $v(t)^2$ and taking $t \to 0$, get

$$\frac{v'(0)}{v(0)^2} = \Big(1 - \frac{\rho_1 + \rho_2}{(\sqrt{\rho_1} + \sqrt{\rho_2})^2}\Big)^{-1} = \frac{1}{2}\Big(\sqrt{\frac{\rho_1}{\rho_2}} + \sqrt{\frac{\rho_2}{\rho_1}}\Big) + 1$$

Meanwhile, recall by differentiating both sides of the silverstein equations (11) in t we can get

$$\frac{v'(t)}{v(t)^2} = 1 + \gamma v'(t) \int \frac{\tau^2}{(1+\tau v(t))^2} dH(\tau).$$

Therefore, if we differentiate once more we get

$$\frac{v''(t)}{v(t)^2} - 2\frac{v'(t)^2}{(v(t))^3} = \gamma v''(t) \int \frac{\tau^2}{(1+\tau v(t))^2} dH(\tau) - 2\gamma v'(t)^2 \int \frac{\tau^3}{(1+\tau v(t))^3} dH(\tau),$$

and thus, multiplying through by $v(t)^3/v'(t)^2$ and rearranging we arrive at

$$\frac{v''(t)v(t)}{v'(t)^2}\Big[1 - \gamma \int \frac{\tau^2 v(z)^2}{(1+\tau v(z))^2} dH(\tau)\Big] = 2\Big[1 - \gamma \int \frac{\tau^3 v(z)^3}{(1+\tau v(z))^3} dH(\tau)\Big].$$

Furthermore, noting that $1 - \gamma \int \frac{\tau^2 v(z)^2}{(1+\tau v(z))^2} dH(\tau) = \frac{v'(t)}{v(t)^2}$ means we get the following equality for the second derivative

$$v''(t) = 2\Big[1 - \gamma \int \frac{\tau^3 v(t)^3}{(1+\tau v(t))^3} dH(\tau)\Big]\Big(\frac{v'(t)}{v(t)}\Big)^3.$$

Taking $t \to 0$ and plugging in the defintion of $v(0)$ yields the following, which will be required for the proof

$$v''(0) = 2\Big[1 - \frac{\rho_1^{3/2} + \rho_2^{3/2}}{(\sqrt{\rho_1} + \sqrt{\rho_2})^3}\Big]\Big(\frac{v'(0)}{v(0)}\Big)^3.$$

Now, let us compute the derivative of the asymmptotic risk $R_{\text{Asymm}}(\lambda)$ for the strong and weak features model. Bringing together the Bias and Variance terms, differentiating through by λ and dividing by σ^2 we get

$$\begin{aligned}\frac{1}{\sigma^2} R'_{\text{Asym}}(\lambda) = &\Big(2\frac{(v'(-\lambda))^2}{(v(-\lambda))^3} - \frac{v''(-\lambda)}{(v(-\lambda)^2)}\Big)\Big(1 + \frac{r^2}{\sigma^2}\sum_{i=1}^{2} \frac{\phi_i\psi_i\rho_i}{(\rho_i v(-\lambda)+1)^2}\Big) \\ &+ 2\Big(\frac{v'(-\lambda)}{v(-\lambda)}\Big)^2 \frac{r^2}{\sigma^2}\sum_{i=1}^{2} \frac{\phi_i\psi_i\rho_i^2}{(\rho_i v(-\lambda)+1)^3}\end{aligned}$$

Taking $\lambda \to 0$ and plugging in $\psi_1 = \psi_2 = 1/2$, $\psi_1\phi_1 + \psi_2\phi_2 = 1$, $\phi_1 + \phi_2 = 2$ as well as $v(0)$ into $\rho_i v(0)$ for $i = 1, 2$ yields

$$\frac{1}{\sigma^2} R'_{\text{Asym}}(0) = \Big(2\frac{v'(0)^2}{v(0)^3} - \frac{v''(0)}{v(0)^2)}\Big)\Big(1 + \frac{r^2}{\sigma^2}\frac{\rho_1\rho_2}{(\sqrt{\rho_1} + \sqrt{\rho_2})^2}\Big) + \Big(\frac{v'(0)}{v(0)}\Big)^2 \frac{r^2}{\sigma^2}(\rho_1\rho_2)^{3/2}\frac{\phi_1\sqrt{\rho_1} + \phi_2\sqrt{\rho_2}}{(\sqrt{\rho_1} + \sqrt{\rho_2})^3}.$$

Let us now plug in the second derivative $v''(0)$. In particular, note that we can write

$$\begin{aligned}2\frac{v'(0)^2}{v(0)^3} - \frac{v''(0)}{v(0)^2} &= 2\Big(\frac{v'(0)}{v(0)}\Big)^2 \frac{1}{v(0)}\Big(1 - \Big(1 - \frac{\rho_1^{3/2} + \rho_2^{3/2}}{(\sqrt{\rho_1} + \sqrt{\rho_2})^3}\Big)\frac{v'(0)}{v(0)^2}\Big) \\ &= -\Big(\frac{v'(0)}{v(0)}\Big)^2 \frac{1}{v(0)}\end{aligned}$$

where on the second equality we have used the equality for $\frac{v'(0)}{v(0)^2}$ from above to note that

$$\begin{aligned}
&1 - \Big(1 - \frac{\rho_1^{3/2} + \rho_2^{3/2}}{(\sqrt{\rho_1} + \sqrt{\rho_2})^3}\Big)\frac{v'(0)}{v(0)^2} \\
&= \frac{\rho_1^{3/2} + \rho_2^{3/2}}{(\sqrt{\rho_1} + \sqrt{\rho_2})^3} - \frac{1}{2}\sqrt{\frac{\rho_1}{\rho_2}} - \frac{1}{2}\sqrt{\frac{\rho_2}{\rho_1}} + \frac{\rho_1^{3/2} + \rho_2^{3/2}}{(\sqrt{\rho_1} + \sqrt{\rho_2})^3}(\sqrt{\frac{\rho_1}{\rho_2}} + \sqrt{\frac{\rho_2}{\rho_1}})\frac{1}{2} \\
&= \frac{(\rho_1^{3/2} + \rho_2^{3/2})(2\sqrt{\rho_1\rho_2} + \rho_1 + \rho_2) - (\rho_1 + \rho_2)(\sqrt{\rho_1} + \sqrt{\rho_2})^3}{2\sqrt{\rho_1\rho_2}(\sqrt{\rho_1} + \sqrt{\rho_2})^3} \\
&= \frac{(\rho_1^{3/2} + \rho_2^{3/2})(2\sqrt{\rho_1\rho_2} + \rho_1 + \rho_2) - (\rho_1 + \rho_2)(\sqrt{\rho_1} + \sqrt{\rho_2})^3}{2\sqrt{\rho_1\rho_2}(\sqrt{\rho_1} + \sqrt{\rho_2})^3} \\
&= \frac{(\rho_1^{3/2} + \rho_2^{3/2}) - (\rho_1 + \rho_2)(\sqrt{\rho_1} + \sqrt{\rho_2})}{2\sqrt{\rho_1\rho_2}(\sqrt{\rho_1} + \sqrt{\rho_2})} \\
&= -\frac{1}{2}.
\end{aligned}$$

Returning to the derivative of the asymptotic risk $R'_{\text{Asym}}(0)$, factoring out $\Big(\frac{v'(0)}{v(0)}\Big)^2$ and plugging in the definition of $v(0)$ gives

$$\frac{1}{\sigma^2}R'_{\text{Asym}}(0) = \Big(\frac{v'(0)}{v(0)}\Big)^2\sqrt{\rho_1\rho_2}\Big[-1 + \frac{r^2}{\sigma^2}\frac{\rho_1\rho_2}{(\sqrt{\rho_1} + \sqrt{\rho_2})^2}\Big(\frac{\phi_1\rho_1 + \phi_2\rho_2}{\sqrt{\rho_1} + \sqrt{\rho_2}} - 1\Big)\Big]$$

It is then clear that the sign of $R'_{\text{Asym}}(0)$ is governed by the quantity in the square brackets. This then yields the result.

B Proof of Lemma 2

In this section we provide the proof for Lemma 2. We recall that the limits (13) and (14) have been computed previously. In particular, Lemma 2.2 of (Ledoit and Péché, 2011) (the roles of d, n are swapped in their work, and thus, one must swap γ with $1/\gamma$) shows

$$\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-1}\Sigma\Big) \to \gamma^{-1}\Big(\frac{1}{1 - \gamma(1 - \lambda m(-\lambda))} - 1\Big)$$

Meanwhile Lemma 7.4 of (Dobriban et al., 2018) shows

$$\frac{1}{d}\operatorname{Tr}\Big(\Big(\frac{X^\top X}{n} + \lambda I\Big)^{-2}\Sigma\Big) \to \frac{m(-\lambda) - \lambda m'(-\lambda)}{(1 - \gamma(1 - \lambda m(-\lambda)))^2}$$

This leaves us to show the limit (15), for which we build upon the techniques (Ledoit and Péché, 2011) as well as (Chen et al., 2011).

We begin with the decomposition. Recall since the covariates are multivariate Gaussians, they can be rewritten as $X = Z\Sigma^{1/2}$ where $Z \in \mathbb{R}^{n\times d}$ is a matrix of independent standard normal Gaussian random variables. For $i = 1, \ldots, n$ the associated row in X is then denoted $X_i = Z_i\Sigma^{1/2}$. As such $X^\top X = \sum_{i=1}^n X_i^\top X_i = \sum_{i=1}^n \Sigma^{1/2}Z_i^\top Z_i\Sigma^{1/2}$. Let us then define $R_i(z) = \big(\frac{X^\top X}{n} - \frac{X_i^\top X_i}{n} - zI\big)^{-1}$. Using the Sherman-Morrison formula we then get

$$R(z) = (\frac{X^\top X}{n} - zI)^{-1} = R_i(z) - \frac{1}{n}\frac{R_i(z)\Sigma^{1/2}Z_i^\top Z_i\Sigma^{1/2}R_i(z)}{1 + \frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top} \tag{19}$$

Moreover we have

$$\frac{1}{n}\sum_{i=1}^n \Sigma^{1/2}Z_i^\top Z_i\Sigma^{1/2}R(z) = \frac{X^\top X}{n}R(z) = (\frac{X^\top X}{n} - zI)R(z) + zR(z) = I + zR(z)$$

Multiplying the above on the left by $\Phi(\Sigma)R(z)$, taking the trace and dividing by d yields

$$\begin{aligned}\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(z)\right)+z\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(z)^2\right)&=\frac{1}{d}\sum_{i=1}^{n}\frac{1}{n}Z_i\Sigma^{1/2}R(z)\Phi(\Sigma)R(z)\Sigma^{1/2}Z_i^\top\\&=\frac{1}{d}\sum_{i=1}^{n}\frac{1}{n}\frac{Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top}{\left(1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top\right)^2}\end{aligned}$$

where for $i=1,\dots,n$ we have plugged in (19) twice into for $R(z)$ to get

$$\begin{aligned}&Z_i\Sigma^{1/2}R(z)\Phi(\Sigma)R(z)\Sigma^{1/2}Z_i^\top\\&=Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R(z)\Sigma^{1/2}Z_i^\top-\frac{1}{n}\frac{Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R(z)\Sigma^{1/2}Z_i^\top}{1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top}\\&=\frac{Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R(z)\Sigma^{1/2}Z_i^\top}{1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top}\\&=\frac{1}{1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top}\\&\quad\times\left[Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top-\frac{1}{n}\frac{Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top}{1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top}\right]\\&=\frac{Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top}{\left(1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top\right)^2}.\end{aligned}$$

Choosing $z=-\lambda$ we then have that

$$\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(-\lambda)\right)-\lambda\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(-\lambda)^2\right)=\frac{1}{d}\sum_{i=1}^{n}\frac{\frac{1}{n}\operatorname{Tr}\left(\Sigma^{1/2}R(-\lambda)\Phi(\Sigma)R(-\lambda)\Sigma^{1/2}\right)}{\left(1+\frac{1}{n}\operatorname{Tr}(\Sigma R(-\lambda))\right)^2}+\delta \tag{20}$$

where the error term $\delta=\delta_1+\delta_2+\delta_3+\delta_4$ such that

$$\begin{aligned}\delta_1&=\frac{1}{d}\sum_{i=1}^{n}\frac{\frac{1}{n}\operatorname{Tr}\left(\Sigma^{1/2}R_i(-\lambda)\Phi(\Sigma)R_i(-\lambda)\Sigma^{1/2}\right)-\frac{1}{n}\operatorname{Tr}\left(\Sigma^{1/2}R(-\lambda)\Phi(\Sigma)R(-\lambda)\Sigma^{1/2}\right)}{\left(1+\frac{1}{n}\operatorname{Tr}(\Sigma R(-\lambda))\right)^2}\\\delta_2&=\frac{1}{d}\sum_{i=1}^{n}\frac{1}{n}\operatorname{Tr}\left(\Sigma^{1/2}R_i(-\lambda)\Phi(\Sigma)R_i(-\lambda)\Sigma^{1/2}\right)\left(\frac{1}{\left(1+\frac{1}{n}\operatorname{Tr}(\Sigma R_i(-\lambda))\right)^2}-\frac{1}{\left(1+\frac{1}{n}\operatorname{Tr}(\Sigma R(-\lambda))\right)^2}\right)\\\delta_3&=\frac{1}{d}\sum_{i=1}^{n}\frac{1}{n}\operatorname{Tr}\left(\Sigma^{1/2}R_i(-\lambda)\Phi(\Sigma)R_i(-\lambda)\Sigma^{1/2}\right)\\&\qquad\times\left(\frac{1}{\left(1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(-\lambda)\Sigma^{1/2}Z_i^\top\right)^2}-\frac{1}{\left(1+\frac{1}{n}\operatorname{Tr}(\Sigma R_i(-\lambda))\right)^2}\right)\\\delta_4&=\frac{1}{d}\sum_{i=1}^{n}\frac{\frac{1}{n}Z_i\Sigma^{1/2}R_i(-\lambda)\Phi(\Sigma)R_i(-\lambda)\Sigma^{1/2}Z_i^\top-\frac{1}{n}\operatorname{Tr}\left(\Sigma^{1/2}R_i(-\lambda)\Phi(\Sigma)R_i(-\lambda)\Sigma^{1/2}\right)}{\left(1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(z)\Sigma^{1/2}Z_i^\top\right)^2}\end{aligned}$$

As shown in section B.1 the error terms $|\delta_1|,|\delta_2|,|\delta_3|,|\delta_4|\to 0$ almost surely as $n,d\to\infty$. It is now a matter of computing the limits of the remaining terms. As discussed previously the limit of $\frac{1}{d}\operatorname{Tr}(\Sigma R(-\lambda))$ is known from (Ledoit and Péché, 2011). From the same work it is also known that

$$\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(-\lambda)\right)\to\Theta^{\Phi}(-\lambda). \tag{21}$$

That leaves us to compute the limit of $\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(-\lambda)^2\right)$. If we are to write $f_d(\lambda)=\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(-\lambda)\right)$ then note the derivative with respect to λ is $f_d'(\lambda)=-\frac{1}{d}\operatorname{Tr}\left(\Phi(\Sigma)R(-\lambda)^2\right)$. We wish to now study the limit of the $f_d'(\lambda)$ through the limit of $f_d(\lambda)$. To do so we will follow the steps in (Dobriban et al., 2018), which will require some definitions and the following theorem.

Let D be a domain, i.e. a connected open set of $\mathbb{C}$. A function $f:D\to\mathbb{C}$ is called analytic on D if it is differentiable as a function of the complex variable z on D. The following key theorem, sometimes known as Vitali's Theorem, ensures that the derivatives of converging analytic functions also converge.

Theorem 2 (Lemma 2.14 in (Bai and Silverstein, 2010)) *Let $f_1, f_2, \ldots$ be analytic on the domain D, satisfying $|f_n(z)| \leq M$ for every n and z in D. Suppose that there is an analytic function f on D such that $f_n(z) \to f(z)$ for all $z \in D$. Then it also holds that $f_n'(z) \to f'(z)$ for all $z \in D$*

Now we have from (Ledoit and Péché, 2011)

$$f_d(\lambda) \to \int \Phi(\tau) \frac{1}{\tau(1-\gamma(1-\lambda m(-\lambda))) + \lambda} dH(\tau)$$

for all $\lambda \in \mathcal{S} := \{u + iv : v \neq 0, \text{ or } v = 0, u > 0\}$. Checking the conditions of Theorem 2 we have that $f_d(\lambda)$ is an analytic function of λ on $\mathcal{S}$ and is bounded $|f_d(\lambda)| \leq \frac{\|\Phi(\Sigma)\|_2}{\lambda}$. To apply Theorem 2 it suffices to show that the limit $\Theta^{\Phi}(-\lambda)$ is analytical. To this end we invoke Morera's theorem which states if

$$\oint_{\gamma} \Theta^{\Phi}(-\lambda) d\lambda = 0$$

for any closed curve γ in the region $\mathcal{S}$ then $\Theta^{\Phi}(-\lambda)$ is analytic. We see this is the case by applying Fubini's Theorem as follows

$$\begin{aligned}\oint_{\gamma} \Theta^{\Phi}(-\lambda) d\lambda &= \oint_{\gamma} \int \Phi(\tau) \frac{1}{\tau(1-\gamma(1-\lambda m(-\lambda))) + \lambda} dH(\tau) d\lambda \\ &= \int \Phi(\tau) \underbrace{\oint_{\gamma} \frac{1}{\tau(1-\gamma(1-\lambda m(-\lambda))) + \lambda} d\lambda}_{=0} dH(\tau) = 0\end{aligned}$$

and noting that the inner integral is zero from Cauchy Theorem as $\frac{1}{\tau(1-\gamma(1-\lambda m(-\lambda)))+\lambda}$ is an analytical function of λ in $\mathcal{S}$ for any $\tau \in [h_1, h_2]$. By Theorem 2 we have that

$$-\frac{1}{d} \operatorname{Tr}\left(\Phi(\Sigma) R(-\lambda)^2\right) = f_d'(-\lambda) \to \frac{\partial \Theta^{\Phi}(-\lambda)}{\partial \lambda}. \tag{22}$$

The final limit (15) is arrived at by considering the limit as $d, n \to \infty$ of (20). Specifically, with the fact that $\delta \to 0$, bringing together (21), (22) and (13). Noting that (13) is applied to the square of $1 + \frac{1}{n} \operatorname{Tr}\left(\Sigma R(-\lambda)\right) = 1 + \gamma \frac{1}{d} \operatorname{Tr}\left(\Sigma R(-\lambda)\right) \to \frac{1}{1-\gamma(1-\lambda m(-\lambda))}$.

B.1 Showing $\delta \to 0$

To analyse these quantities we introduce the following concentration inequality from Lemma A.2 of (Paul, 2007) with $\delta = 1/3$.

Lemma 3 *Suppose y is $d-$dimensional Gaussian random vector $y \sim \mathcal{N}(0, I)$ and $C \in \mathbb{R}^{d \times d}$ is a symmetric matrix such that $\|C\| \leq L$. Then for all $0 < t < L$,*

$$\mathbf{P}\big(\frac{1}{d}|yCy^{\top} - \operatorname{Tr}(C)| > t\big) \leq 2 \exp\Big\{ - \frac{pt^2}{6L^2} \Big\}.$$

Furthermore, we will use the fact that the maximal eigenvalues are upper bounded

$$\|R(-\lambda)\|_2 \leq \frac{1}{\lambda} \quad \text{and} \quad \max_{1 \leq i \leq n} \|R_i(-\lambda)\|_2 \leq \frac{1}{\lambda}$$

We proceed to show that each of the error $\delta_1, \delta_2, \delta_3, \delta_4$ converge to zero almost surely.

Begin with δ_1. For $i = 1, \ldots, n$ by adding and subtracting $\operatorname{Tr}(\Sigma^{1/2} R(z) \Phi(\Sigma) R_i(z) \Sigma^{1/2})$ we can decompose

$$\begin{aligned}&\operatorname{Tr}\left(\Sigma^{1/2} R_i(-\lambda) \Phi(\Sigma) R_i(-\lambda) \Sigma^{1/2}\right) - \operatorname{Tr}\left(\Sigma^{1/2} R(z) \Phi(\Sigma) R(z) \Sigma^{1/2}\right) \\ &= \operatorname{Tr}\left((R_i(-\lambda) - R(-\lambda)) \Phi(\Sigma) R_i(-\lambda) \Sigma\right) + \operatorname{Tr}\left(\Sigma R(-\lambda) \Phi(\Sigma) (R_i(-\lambda) - R(-\lambda))\right)\end{aligned}$$

Using (19) and letting $A = \Phi(\Sigma)R_i(-\lambda)\Sigma$ we then get

$$\begin{aligned}\frac{1}{n}|\operatorname{Tr}\big((R_i(-\lambda)-R(-\lambda))A\big)| &= \Big|\frac{1}{n^2}\frac{Z_i\Sigma^{1/2}R_i(-\lambda)AR_i(-\lambda)\Sigma^{1/2}Z_i^\top}{1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(-\lambda)\Sigma^{1/2}Z_i^\top}\Big| \\ &\leq \frac{\|A\|_2}{n}\Big|\frac{1}{n}\frac{Z_i\Sigma^{1/2}R_i(-\lambda)R_i(-\lambda)\Sigma^{1/2}Z_i^\top}{1+\frac{1}{n}Z_i\Sigma^{1/2}R_i(-\lambda)\Sigma^{1/2}Z_i^\top}\Big| \\ &\leq \frac{\|A\|_2}{n}\sup_x\Big|\frac{xR_i(-\lambda)^2x^\top}{1+xR_i(-\lambda)x^\top}\Big| \\ &\leq \frac{\|A\|_2}{n}\sup_x\Big|\frac{xR_i(-\lambda)^2x^\top}{xR_i(-\lambda)x^\top}\Big| \\ &\leq \frac{\|A\|_2}{n}\|R_i(-\lambda)\|_2 \\ &\leq \frac{\|A\|_2}{\lambda n} \\ &\leq \frac{\|\Phi(\Sigma)\|_2\|\Sigma\|_2}{\lambda^2 n}\end{aligned} \tag{23}$$

An identical calculation with $A = \Phi(\Sigma)R(-\lambda)\Sigma$ yields the same bound. This then yields with the lower bound $(1+\operatorname{Tr}(\Sigma^{1/2}R(-\lambda)\Sigma^{1/2})) \geq 1$

$$|\delta_1| \leq 2\frac{n}{d}\frac{\|\Phi(\Sigma)\|_2\|\Sigma\|_2}{\lambda^2 n}$$

and as such δ_1 goes to zero as $n, d \to \infty$ so that $d/n \to \gamma$.

Now consider the term δ_2. Note that for two positive numbers $a, b \geq 0$ we have

$$\begin{aligned}\frac{1}{(1+a)^2}-\frac{1}{(1+b)^2} &= \frac{(1+b)^2-(1+a)^2}{(1+a)^2(1+b)^2} \\ &= \frac{b^2+2b-a^2-2a}{(1+a)^2(1+b)^2} \\ &= \frac{b(b-a)+a(b-a)+2(b-a)}{(1+a)^2(1+b)^2} \\ &= (b-a)\frac{(b+1)+(a+1)}{(1+a)^2(1+b)^2} \\ &= (b-a)\big(\frac{1}{(1+a)^2(1+b)}+\frac{1}{(1+a)(1+b)^2}\big)\end{aligned}$$

and as such $|(1+a)^{-2}-(1+b)^{-2}| \leq 2|b-a|$. Using this with $a = \frac{1}{n}\operatorname{Tr}(\Sigma^{1/2}R_i(-\lambda)\Sigma^{1/2})$ and $b = \frac{1}{n}\operatorname{Tr}(\Sigma^{1/2}R(-\lambda)\Sigma^{1/2})$ whom are both non-negative, allows us to upper bound

$$\begin{aligned}\Big|\frac{1}{\big(1+\frac{1}{n}\operatorname{Tr}(\Sigma R_i(-\lambda))\big)^2}-\frac{1}{\big(1+\frac{1}{n}\operatorname{Tr}(\Sigma R(-\lambda))\big)^2}\Big| &\leq 2\frac{1}{n}\big|\operatorname{Tr}(\Sigma R_i(-\lambda))-\operatorname{Tr}(\Sigma R(-\lambda))\big| \\ &\leq \frac{2\|\Sigma\|}{\lambda n}\end{aligned}$$

where for the final inequality we used the argument (23) with $A = \Sigma$. Now, since the eigenvalues in the following trace are non-negative we can upper bound

$$\begin{aligned}\frac{1}{d}\big|\operatorname{Tr}\big(\Sigma^{1/2}R_i(-\lambda)\Phi(\Sigma)R_i(-\lambda)\Sigma^{1/2}\big)\big| &\leq \|\Sigma^{1/2}R_i(-\lambda)\Phi(\Sigma)R_i(-\lambda)\Sigma^{1/2}\|_2 \\ &\leq \|\Sigma^{1/2}\|_2^2\|\Phi(\Sigma)\|_2\|R_i(-\lambda)\|_2^2 \\ &\leq \frac{\|\Sigma^{1/2}\|_2^2\|\Phi(\Sigma)\|_2}{\lambda^2} \\ &= \frac{\|\Sigma\|_2\|\Phi(\Sigma)\|_2}{\lambda^2}\end{aligned} \tag{24}$$

Combining these two facts yields the upper bound

$$|\delta_2| \leq \frac{2\|\Sigma\|_2^2\|\Phi(\Sigma)\|_2}{\lambda^3 n}$$

which goes to zero as $n \to \infty$.

We now proceed to bound δ_3 and δ_4. With the bound on the trace (24) as well as using the bound $|(1+a)^{-2} - (1+b)^{-2}| \leq 2|b-a|$ we arrive at the bound for δ_3

$$\begin{aligned}|\delta_3| &\leq 2\frac{\|\Sigma\|_2\|\Phi(\Sigma)\|_2}{\lambda^2} \\ &\quad \times \max_{1\leq i\leq n}\Big|Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top - \operatorname{Tr}\left(\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}\right)\Big|.\end{aligned}$$

Meanwhile using that $1 + \frac{1}{n}Z_i\Sigma^{1/2}R_i(-\lambda)\Sigma^{1/2}Z_i^\top \geq 1$ we arrive at the bound for δ_4

$$|\delta_4| \leq \max_{1\leq i\leq n}\Big|Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top - \operatorname{Tr}\left(\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}\right)\Big|$$

We now show that $\max_{1\leq i\leq n}\Big|Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top - \operatorname{Tr}\left(\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}\right)\Big|$ converges to zero almost surely. Observe since we have the upper bound on the largest eigenvalue we have using Lemma 3 as well as union bound for $1 \leq i \leq n$ we have for $0 < t < \frac{\|\Sigma\|_2\|\Phi(\Sigma)\|_2}{\lambda^2}$

$$\begin{aligned}&\mathbf{P}\Big(\max_{1\leq i\leq n}\frac{1}{d}\Big|Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top - \operatorname{Tr}\left(\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}\right)\Big| \geq t\Big) \\ &\leq 2\exp\Big\{-\frac{dt^2\lambda^4}{6\|\Sigma\|_2^2\|\Phi(\Sigma)\|^2} + \log(n)\Big\} \qquad (25)\end{aligned}$$

Let $V_{n,d} := \max_{1\leq i\leq n}\frac{1}{d}\Big|Z_i\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}Z_i^\top - \operatorname{Tr}\left(\Sigma^{1/2}R_i(z)\Phi(\Sigma)R_i(z)\Sigma^{1/2}\right)\Big|$ and, for any $t > 0$, let $E_{n,d}(t)$ denote the event $\{V_{n,d} \geq t\}$ where $d = d_n$. Then, if $d = d_n$ satisfies $d_n/n \to \infty$, $\mathbf{P}(E_{n,d}) \leq 2n\exp\Big\{-\frac{dt^2\lambda^4}{6\|\Sigma\|_2^2\|\Phi(\Sigma)\|^2}\Big\} \leq 2n\exp\Big\{-\frac{\gamma nt^2\lambda^4}{12\|\Sigma\|_2^2\|\Phi(\Sigma)\|^2}\Big\}$ where the last inequality for n large enough that $d/n \geq \gamma/2$. Hence,

$$\sum_{n=1}^{\infty}\mathbf{P}(E_{n,d_n}(t)) < +\infty$$

so that, by the Borel-Cantelli lemma, almost surely, $V_{n,d} \geq t$ only holds for a finite number of values of n. This implies that, almost surely, $\limsup_{n\to\infty} V_{n,d} \leq t$. Note that this is true for every $t > 0$; letting $t = 1/k$ and taking a union bound over $k \geq 1$ shows that $\limsup_{n\to\infty} V_{n,d} = 0$ almost surely, *i.e.* $V_{n,d} \to 0$ almost surely.