SUPPLEMENTARY MATERIALS: Asymptotics of the Sketched Pseudoinverse[*]

Daniel LeJeune[†], Pratik Patil[‡], Hamid Javadi[§], Richard G. Baraniuk[¶], and Ryan J. Tibshirani[‡]

This document serves as a supplement to the paper "Asymptotics of the Sketched Pseudoinverse." The contents of this supplement are organized as follows. In section SM1, we collect some useful facts regarding Stieltjes transforms that are used in some of the proofs in later sections. In section SM2, we provide a detailed proof for Theorem 3.1. In section SM3, we provide proof for Theorem 4.1. In section SM4, we provide proofs of various properties regarding our main equivalences mentioned section 5 in the main paper. In section SM5, we give proofs for the application of our equivalence to the sketch-and-project method. Finally, in section SM6, we give proof for Theorem 7.2 which extends our results to free sketching.

SM1. Useful facts.

In this section, we jot down basic definitions and facts related Stieltjes transform that we will be using throughout the paper.

Let Q be a bounded nonnegative measure on $\mathbb{R}$. The Stieltjes transform of Q is defined at $z \in \mathbb{C}^+$ by

$$m_Q(z) = \int_{\mathbb{R}} \frac{1}{x - z} \, \mathrm{d}Q(x).$$

Fact SM1.1. *Let m be the Stieltjes transform of bounded measure Q on $\mathbb{R}_{\geq 0}$. Let $z \in \mathbb{C}^+$ with $\mathrm{Re}(z) < 0$. Then, $\mathrm{Im}(m(z)) \searrow 0$ as $\mathrm{Im}(z) \searrow 0$.*

Proof. Let $z = x + iy$ with $x < 0$ and $y > 0$. Since m is a Stieltjes transform of Q, we have

$$\mathrm{Im}(m(z)) = \mathrm{Im}\left(\int \frac{1}{r - z} \, \mathrm{d}Q(r)\right) = \mathrm{Im}\left(\int \frac{1}{r - (x + iy)} \, \mathrm{d}Q(r)\right) = \int \frac{y}{(r - x)^2 + y^2} \, \mathrm{d}Q(r).$$

Thus, we can bound

$$|\mathrm{Im}(m(z))| \leq \frac{y}{x^2} \int \mathrm{d}Q(r).$$

Since Q is a bounded measure, by letting $y \searrow 0$, one has $\mathrm{Im}(m(z)) \searrow 0$ as $\mathrm{Im}(z) \searrow 0$. ∎

---

[*]Supplementary material for SIMODS MS#M153026.
https://doi.org/10.1137/22M1530264

[†]Department of Statistics, Stanford University, Stanford, CA 94305 USA (daniel@dlej.net).
[‡]Department of Statistics, University of California, Berkeley, CA 94720 USA (pratikpatil@berkeley.edu, ryantibs@berkeley.edu).
[§]Google, Mountain View, CA 94043 USA (hamid71reza@gmail.com).
[¶]Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (richb@rice.edu).

We will be interested in the Stieltjes transforms of spectral measures. The spectral distribution of a symmetric matrix $\mathbf{A} \in \mathbb{C}^{p\times p}$ with eigenvalues $\lambda_1(\mathbf{A}), \dots, \lambda_p(\mathbf{A})$ is the probability distribution that places a point mass of $\frac{1}{p}$ at each eigenvalue

$$F_{\mathbf{A}}(\lambda) = \tfrac{1}{p}\sum_{i=1}^{p} \mathbb{1}\{\lambda_i \leq \lambda\}.$$

The matrices of interest for us will be the population covariance matrix $\mathbf{\Sigma} \in \mathbb{C}^{p\times p}$ and the sample covariance matrix $\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X}$ where $\mathbf{X} \in \mathbb{C}^{n\times p}$ is the random design matrix.

If the Stieltjes transform of spectrum of the sample covariance matrix $\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X}$ is

$$m(z) = \tfrac{1}{p}\mathrm{tr}[(\tfrac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_p)^{-1}], \tag{SM1.1}$$

then the so-called *companion* Stieltjes transform

$$v(z) = \tfrac{1}{n}\mathrm{tr}[(\tfrac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_n)^{-1}] \tag{SM1.2}$$

is the Stieltjes transform of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$ (and hence the prefix). The reason it is useful is that it is often easier to work with the companion Stieltjes transform than the Stieltjes transform. The following fact relates the companion Stieltjes transform to the Stieltjes transform.

Fact SM1.2. *The companion Stieltjes transform $v(z)$ can be expressed in terms of the Stieltjes transform $m(z)$ at $z \in \mathbb{C}^+$ as*

$$v(z) = \frac{p}{n}m(z) + \frac{1}{z}\left(\frac{p}{n} - 1\right).$$

Proof. Let $(\lambda_i)_{i=1}^r$ be the nonzero eigenvalues of $\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X}$ (which are also the nonzero eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$). Define $\Lambda(z) = \sum_{i=1}^r \frac{1}{\lambda_i - z}$. From (SM1.1) and (SM1.2), note that we can write

$$m(z) = \frac{\Lambda(z)}{p} - \frac{(p-r)}{pz}, \quad v(z) = \frac{\Lambda(z)}{n} - \frac{(n-r)}{nz}.$$

Combining these equations proves the claim. ∎

SM2. Proof of Theorem 3.1.

As a preliminary that we will need later, through a standard argument, we will first show that $\mathrm{Im}(c(z)) \nearrow 0$ as $\mathrm{Im}(z) \searrow 0$ in (3.2) for $z \in \mathbb{C}^+$ with $\mathrm{Re}(z) < 0$. To proceed, denote $\frac{1}{p}\mathrm{tr}\left[\mathbf{\Sigma}(c(z)\mathbf{\Sigma} - z\mathbf{I}_p)^{-1}\right]$ by $d(z)$. From the last part of Lemma 2.1, $d(z)$ is a Stieltjes transform of a certain positive measure on $\mathbb{R}_{\geq 0}$ with total mass $\frac{1}{p}\mathrm{tr}[\mathbf{\Sigma}]$. Since the operator norm of $\mathbf{\Sigma}$ is uniformly bounded in p, we have that $\frac{1}{p}\mathrm{tr}[\mathbf{\Sigma}]$ is bounded above by some constant independent of p. Combining this with Fact SM1.1, we have that $\mathrm{Im}(d(z)) \searrow 0$ as $\mathrm{Im}(z) \searrow 0$ for $z \in \mathbb{C}^+$ with $\mathrm{Re}(z) < 0$. Now manipulating (2.2), we can write

$$c(z) = \frac{1}{1 + \frac{p}{n}d(z)}.$$

Thus, we can conclude that $\mathrm{Im}(c(z)^{-1}) \searrow 0$ as $\mathrm{Im}(z) \searrow 0$ for $z \in \mathbb{C}^+$ with $\mathrm{Re}(z) < 0$. This in turn implies that $\mathrm{Im}(c(z)) \nearrow 0$ for $z \in \mathbb{C}^+$ with $\mathrm{Re}(z) < 0$.

We now begin the proof.

Proof. We start by considering $z \in \mathbb{C}^+$. To obtain (3.2), we multiply both sides of (2.1) by z:

$$\begin{aligned} z\left(\tfrac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_p\right)^{-1} &\simeq z\left(c(z)\boldsymbol{\Sigma} - z\mathbf{I}_p\right)^{-1} \\ &= \tfrac{z}{c(z)}\left(\boldsymbol{\Sigma} - \tfrac{z}{c(z)}\mathbf{I}_p\right)^{-1}. \end{aligned} \tag{SM2.1}$$

We will let $\zeta = \frac{z}{c(z)}$ shortly. First let $m(z) = \frac{1}{p}\mathrm{tr}\left[\left(c(z)\boldsymbol{\Sigma} - z\mathbf{I}_p\right)^{-1}\right]$. By an additional application of Lemma 2.1, $m(z)$ is asymptotically equal to $\frac{1}{p}\mathrm{tr}\left[\left(\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}\right)^{-1}\right]$, the Stieltjes transform of the spectrum of $\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X}$. Now note that we can write (2.2) in terms of $m(z)$ as

$$\frac{1}{c(z)} - 1 = \frac{p}{n}m(z).$$

We can manipulate the equation in the display above into the following form:

$$-\frac{c(z)}{z} = \frac{p}{n}m(z) + \frac{1}{z}\left(\frac{p}{n} - 1\right).$$

From the relationship between Stieltjes and the companion Stieltjes transforms in Fact SM1.2, this means that $-\frac{c(z)}{z}$ is asymptotically equal to $v(z) = \frac{1}{n}\mathrm{tr}[\left(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}\right)^{-1}]$, the companion Stieltjes transform of the spectrum of $\frac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X}$. Thus, letting $\zeta = \frac{z}{c(z)}$ in (SM2.1), we have that

$$z\left(\tfrac{1}{n}\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_p\right)^{-1} \simeq \zeta(\boldsymbol{\Sigma} - \zeta\mathbf{I}_p)^{-1},$$

and that asymptotically, $\zeta = -\frac{1}{v(z)}$ is the unique solution in $\mathbb{C}^+$ to (3.3) for $z \in \mathbb{C}^+$. Moreover, through analytic continuation, one can extend this relationship to the real line outside the support of the spectrum of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$ where by the similar argument as for $c(z)$ above, both $v(z)$ and ζ are real.

It remains to determine the interval for which the analytic continuation coincides with a unique solution to (3.3) for a given z. Let $z_0 \in \mathbb{R}$ denote the most negative zero of v. Then for all $z < z_0$, $\zeta \in \mathbb{R}$ is well-defined, asymptotically being a solution to

$$z - \zeta = -\zeta\tfrac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma} - \zeta\mathbf{I}_p\right)^{-1}\right], \tag{SM2.2}$$

which is an algebraic manipulation of (2.2). However, as we will now show, the solution to this equation is not in general unique, so we will show that the most negative solution for ζ is the correct analytic continuation of the corresponding solution in $\mathbb{C}^+$.

Consider the two sides of (SM2.2). The left-hand side is linear in ζ, and the right-hand side is concave for $\zeta < \lambda_{\min}^+(\boldsymbol{\Sigma})$. To see this, observe that

$$\begin{aligned} \frac{\partial^2}{\partial\zeta^2}\left(-\zeta\tfrac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma} - \zeta\mathbf{I}_p\right)^{-1}\right]\right) &= \frac{\partial}{\partial\zeta}\left(-\tfrac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma} - \zeta\mathbf{I}_p\right)^{-1}\right] - \zeta\tfrac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma} - \zeta\mathbf{I}_p\right)^{-2}\right]\right) \\ &= \frac{\partial}{\partial\zeta}\left(-\tfrac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma} - \zeta\mathbf{I}_p\right)^{-2}\right]\right) \\ &= -\tfrac{2}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma} - \zeta\mathbf{I}_p\right)^{-3}\right] < 0. \end{aligned}$$

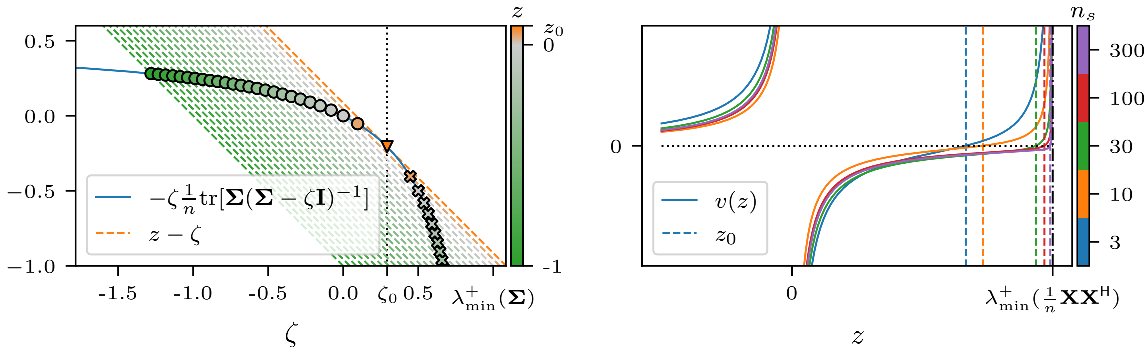

Figure SM1. ***Left:*** *Numerical illustration of the solutions to* (SM2.2) *for* $\boldsymbol{\Sigma} = \mathbf{I}$ *and* $\frac{p}{n} = \frac{1}{2}$. *The right-hand side of* (SM2.2) *is a fixed function of* ζ *(blue, solid), but the left-hand-side is a line with slope* -1 *shifted by* z *(orange to green, dashed). Solutions are the most negative intersections of the curves (circles), and not the most positive intersections (x's). The greatest possible value of* z *yielding an intersection,* $z = z_0$ *(triangle), gives* $\zeta = \zeta_0$ *(dotted). For this example, we know that* $z_0 = (1 - \sqrt{\frac{p}{n}})^2 \approx 0.0858$ *since the spectrum of* $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$ *follows the Marchenko–Pastur distribution.* ***Right:*** *Illustration of the convergence of* z_0 *to* $\lambda^+_{\min}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$. *For* $\boldsymbol{\Sigma} = \mathbf{I}$, $p = 500$, $n = 1000$, *we draw a random* $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$ *and compute its eigenvalues. To simulate increasing the dimensionality of the matrix while keeping* $\lambda^+_{\min}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$ *fixed, we then take a subsample of size* n_s *of the eigenvalues, comprised of* $\lambda^+_{\min}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$ *and* $n_s - 1$ *other eigenvalues chosen uniformly at random. We then plot* $v(z)$ *(solid) using this subsample. For any finite* n_s, z_0 *(dashed) will always lie between* 0 *and* $\lambda^+_{\min}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$, *but* z_0 *approaches* $\lambda^+_{\min}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$ *as* n_s *tends to infinity.*

A linear function and a concave function can intersect at zero, one, or two points. If at one point, this must occur at the unique point (z_1, ζ_1), $\zeta_1 < \lambda^+_{\min}(\boldsymbol{\Sigma})$ for which the derivatives of each side of (SM2.2) coincide, satisfying

$$1 = \tfrac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}^2 \left(\boldsymbol{\Sigma} - \zeta_1 \mathbf{I}_p\right)^{-2}\right].$$

This right-hand side of this equation sweeps the range $(0, \infty)$ for $\zeta_1 \in (-\infty, \lambda^+_{\min}(\boldsymbol{\Sigma}))$, so such a (z_1, ζ_1) always exists. Furthermore, since the solutions ζ are continuous as a function z, the analytic continuation of the complex solution to the reals of the map $z \mapsto \zeta$ with domain $(-\infty, z_1)$ must have image of either $(-\infty, \zeta_1)$ or $(\zeta_1, \lambda^+_{\min}(\boldsymbol{\Sigma}))$. The correct image must be $(-\infty, \zeta_1)$, which we illustrate in Figure SM1 (left).

To see why this must be the correct image, consider $z = x + i\varepsilon$ for a fixed $\varepsilon > 0$ with x very negative. Rewriting (SM2.2), we have the form of (3.3):

$$z = \zeta\left(1 - \tfrac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma} - \zeta \mathbf{I}_p\right)^{-1}\right]\right). \tag{SM2.3}$$

We begin by considering the behavior of the trace term. Let $\zeta = \chi + i\xi$, and suppose that $\chi < \frac{x}{2}$, which means that χ is also very negative. The trace is a sum of terms of the form

$$\frac{\sigma}{\sigma - \zeta} = \frac{\sigma(\sigma - \chi + i\xi)}{(\sigma - \chi)^2 + \xi^2}.$$

Let $g(\zeta)$ and $h(\zeta)$ denote the real and imaginary parts of $\frac{1}{n}\mathrm{tr}[\boldsymbol{\Sigma}\,(\boldsymbol{\Sigma}-\zeta\mathbf{I}_p)^{-1}]$. For x (and therefore χ) sufficiently negative, this gives us the simple bounds

$$|g(\zeta)| \le \frac{\frac{p}{n}\sigma_{\max}(\boldsymbol{\Sigma})}{-\chi} \le \frac{2\frac{p}{n}\sigma_{\max}(\boldsymbol{\Sigma})}{-x},$$
$$|h(\zeta)| \le \frac{\frac{p}{n}\sigma_{\max}(\boldsymbol{\Sigma})\xi}{\chi^2} \le \frac{4\frac{p}{n}\sigma_{\max}(\boldsymbol{\Sigma})\xi}{x^2}.$$

We therefore have by (SM2.3) that

$$\chi = \frac{x(1-g(\zeta))+\varepsilon h(\zeta)}{(1-g(\zeta))^2+h(\zeta)^2}, \quad \xi = \frac{\varepsilon(1-g(\zeta))-xh(\zeta)}{(1-g(\zeta))^2+h(\zeta)^2}.$$

By our bounds on g and h, we can conclude that for sufficiently negative x, there exists $a>0$ and $0<b<1$ such that $|\xi| \le a\varepsilon + b|\xi|$, implying that $|\xi| \le \frac{a\varepsilon}{1-b}$, and therefore $|\xi|$ is bounded. Since $|\xi|$ is bounded, $|h(\zeta)|$ has an upper bound of the form $\frac{1}{x^2}$, so for any $c\in(\frac{1}{2},1)$ and sufficiently negative x, we have the bound $\chi \le cx$. Therefore, we can confirm that our supposition that $\chi < \frac{x}{2}$ leads to the unique solution with $\xi>0$, since for any $c'\in(0,1)$ we similarly have $\xi > c'\varepsilon > 0$ for sufficiently negative x. One can similarly argue that for solutions with $\chi \nearrow \lambda_{\min}^{+}(\boldsymbol{\Sigma})$, it must be that $\xi<0$, which is the solution in the wrong half-plane. By continuity of $z\mapsto\zeta$, identifying these extreme cases is sufficient to identify the correct image. Therefore, for real-valued $z<z_1$, the correct ζ is the most negative solution, which is the unique $\zeta<\zeta_1$, and ζ is undefined for $z>z_1$.

Lastly, we argue that asymptotically, $z_0=z_1$. In the case $n<p$, this is straightforward, as the most negative zero of v must lie between the two most negative distinct eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$. This is because there is a pole at each distinct eigenvalue, so the entire range $(-\infty,\infty)$ (including crossing 0) is mapped to by v between each successive pair of distinct eigenvalues. When $n<p$, there is not a point mass at 0, so these two most negative eigenvalues must converge to the same value as the discrete eigenvalue distribution converges to a continuous distribution, and this value marks the beginning of the continuous support of the spectrum of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$, so $z_0 \to \lambda_{\min}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$. Moreover, ζ, being asymptotically equal to $\frac{1}{v}$, is undefined only on the support of the limiting spectrum and continuous elsewhere; therefore by the argument in the previous paragraph, the solution to (SM2.2) does not exist for $z>z_1$, and it must be that $\lambda_{\min}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}) \to z_1$.

For $n>p$, we apply similar reasoning; however, we must take care to consider the point mass of the spectrum at 0. This means that $z_0\in(0,\lambda_{\min}^{+}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}))$, because like before, the first zero must lie between the two most negative distinct eigenvalues, as we illustrate in Figure SM1 (right). However, asymptotically, it must be that $z_0 \nearrow z_1 = \lambda_{\min}^{+}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$. This is most easily seen by a contradiction argument. Suppose we have $z_0 < z_1-\varepsilon$ for some $\varepsilon>0$. Because $\frac{1}{v}$ has a pole at z_0, $-\zeta=\frac{1}{v}\nearrow\infty$ as $z\searrow z_0$. In particular, this means that $\frac{1}{v}$ is discontinuous at z_0, tending to ∞ from the right. Meanwhile, as argued above, $\lambda_{\min}^{+}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})\to z_1$, and we know that for $z<z_1$, $\zeta<\zeta_1\in(-\infty,\lambda_{\min}^{+}(\boldsymbol{\Sigma}))$. This is a contradiction, because on the one hand ζ is upper bounded by $\lambda_{\min}^{+}(\boldsymbol{\Sigma})$ for any $z\in(z_0,z_1)$, but on the other hand $\frac{1}{v}$ can be made arbitrarily large by taking $z\searrow z_0$. Therefore, we must have, asymptotically, that $z_0=z_1=\lambda_{\min}^{+}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}})$. For this reason, in the theorem statement, we denote $\zeta_0=\zeta_1$. ■

SM3. Proof of Theorem 4.1 for positive semidefinite $\mathbf{A}$.

Proof. We begin by proving the equivalence (4.3) and then show that the limit as $\lambda \to 0$ is well-behaved when we multiply by $\mathbf{A}^{1/2}$ to obtain (4.2).

Let $\mathbf{A}_\delta \triangleq \mathbf{A} + \delta \mathbf{I}_p$, $\mathbf{U} \triangleq \mathbf{S}(\mathbf{S}^\mathsf{H}\mathbf{A}\mathbf{S} + \lambda \mathbf{I}_q)^{-1}\mathbf{S}^\mathsf{H}$, and $\mathbf{V} \triangleq (\mathbf{A} + \mu \mathbf{I}_p)^{-1}$. By the Woodbury matrix identity, we have the following two identities:

$$\mathbf{S}(\mathbf{S}^\mathsf{H}\mathbf{A}_\delta\mathbf{S} + \lambda \mathbf{I}_q)^{-1}\mathbf{S}^\mathsf{H} = \mathbf{U} - \delta \mathbf{U}\left(\mathbf{I}_p + \delta \mathbf{U}\right)^{-1}\mathbf{U},$$
$$(\mathbf{A}_\delta + \lambda \mathbf{I}_p)^{-1} = \mathbf{V} - \delta \mathbf{V}\left(\mathbf{I}_p + \delta \mathbf{V}\right)^{-1}\mathbf{V}.$$

If either $\lambda \neq 0$ or $\limsup \frac{q}{p} < \liminf r(\mathbf{A})$, then we can conclude that (see, e.g., [SM1]) that $\|(\mathbf{S}^\mathsf{H}\mathbf{A}_\delta\mathbf{S} + \lambda \mathbf{I}_q)^{-1}\|_{\mathrm{op}}$ is almost surely uniformly bounded and that μ is bounded away from zero (see Remark 5.4). Thus, since $\|\mathbf{S}\|_{\mathrm{op}}$ is also almost surely bounded asymptotically, $\|\mathbf{U}\|_{\mathrm{op}}$ and $\|\mathbf{V}\|_{\mathrm{op}}$ are asymptotically bounded by constants $C_\mathbf{U}$ and $C_\mathbf{V}$, respectively. Therefore, for $\delta < \frac{1}{2}\min\{C_\mathbf{U}, C_\mathbf{V}\}$, we have the following bound on the trace functional difference:

$$\begin{aligned}\text{(SM3.1)}\quad & \limsup \left|\operatorname{tr}\left[\boldsymbol{\Theta}\left(\mathbf{S}(\mathbf{S}^\mathsf{H}\mathbf{A}_\delta\mathbf{S} + \lambda \mathbf{I}_q)^{-1}\mathbf{S}^\mathsf{H} - (\mathbf{A}_\delta + \lambda \mathbf{I}_p)^{-1}\right)\right] - \operatorname{tr}\left[\boldsymbol{\Theta}(\mathbf{U} - \mathbf{V})\right]\right| \\ & \quad \le \tfrac{\delta}{2}\|\boldsymbol{\Theta}\|_{\mathrm{tr}}(C_\mathbf{U}^2 + C_\mathbf{V}^2).\end{aligned}$$

Thus, as $\delta \searrow 0$, the trace functionals converge uniformly over p for $\boldsymbol{\Theta}$ with uniformly bounded trace norm. We can therefore apply the Moore–Osgood Theorem to interchange limits, such that almost surely

$$\begin{aligned}\lim_{p \nearrow \infty} \left|\operatorname{tr}\left[\boldsymbol{\Theta}(\mathbf{U} - \mathbf{V})\right]\right| &= \lim_{\delta \searrow 0} \lim_{p \nearrow \infty} \left|\operatorname{tr}\left[\boldsymbol{\Theta}\left(\mathbf{S}(\mathbf{S}^\mathsf{H}\mathbf{A}_\delta\mathbf{S} + \lambda \mathbf{I}_q)^{-1}\mathbf{S}^\mathsf{H} - (\mathbf{A}_\delta + \lambda \mathbf{I}_p)^{-1}\right)\right]\right| \\ &= 0.\end{aligned}$$

To prove the equivalence in (4.2), we can apply the equivalence in (4.3) proved above unless $\lambda = 0$ and $\limsup \frac{q}{p} \ge \liminf r(\mathbf{A})$. We need only consider $\limsup \lambda_0 < 0$, so it suffices to consider $\liminf \frac{q}{p} > \limsup r(\mathbf{A})$ (see Remark 5.1). The condition $\limsup \lambda_0 < 0$ implies that there exists $c_\lambda > 0$ such that $\lambda_{\min}^{+}(\mathbf{S}^\mathsf{H}\mathbf{A}\mathbf{S}) > c_\lambda$. Therefore, $\|\mathbf{A}^{1/2}\mathbf{S}(\mathbf{S}^\mathsf{H}\mathbf{A}\mathbf{S} + \lambda \mathbf{I}_q)^{-1}\|_{\mathrm{op}}$ is almost surely uniformly bounded in p for all $\lambda \in D_\lambda$, where $D_\lambda = \{z \in \mathbb{C} \colon |z| < \frac{c_\lambda}{2}\}$. We now need to bound $\|\mathbf{A}^{1/2}(\mathbf{A} + \mu \mathbf{I}_p)^{-1}\|_{\mathrm{op}}$. From the definition of μ_0 in (4.1), we observe that

$$\frac{p}{q}\frac{r(\mathbf{A})\lambda_{\max}(\mathbf{A})^2}{(\lambda_{\max}(\mathbf{A}) + \mu_0)^2} \le 1 \le \frac{p}{q}\frac{r(\mathbf{A})\lambda_{\min}^{+}(\mathbf{A})^2}{(\lambda_{\min}^{+}(\mathbf{A}) + \mu_0)^2},$$

from which we can conclude for the case that $\frac{q}{p} > r(\mathbf{A})$ and $\lambda_{\min}^{+}(\mathbf{A}) > 0$, we can bound

(SM3.2)
$$\left(\tfrac{pr(\mathbf{A})}{q} - 1\right)\lambda_{\max}(\mathbf{A}) < \left(\sqrt{\tfrac{pr(\mathbf{A})}{q}} - 1\right)\lambda_{\max}(\mathbf{A}) \le \mu_0 \le \left(\sqrt{\tfrac{pr(\mathbf{A})}{q}} - 1\right)\lambda_{\min}^{+}(\mathbf{A}) < 0.$$

Since $\liminf \frac{q}{p} > \limsup r(\mathbf{A})$ and $\liminf \lambda_{\min}^{+}(\mathbf{A}) > 0$, we therefore must have $\limsup \mu_0 < 0$. Define the set $D_\mu = \{z \in \mathbb{C} \colon |z| < \frac{-\limsup \mu_0}{2}\}$. Since $-\liminf \lambda_{\min}^{+}(\mathbf{A}) \le \mu_0$, for all $\mu \in D_\mu$, we must have the bound

$$\left\|\mathbf{A}^{1/2}\left(\mathbf{A}+\mu\mathbf{I}_p\right)^{-1}\right\|_{\mathrm{op}} \le \frac{2\left\|\mathbf{A}^{1/2}\right\|_{\mathrm{op}}}{-\limsup \mu_0}.$$

We also know from (4.4) that

$$|\lambda| = |\mu|\left|1-\tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}_p\right)^{-1}\right]\right|.$$

One can confirm that the second factor on the right-hand side is uniformly lower bounded away from 0 for $\mu \in D_\mu$ using the first bound in (SM3.2). Let $D_p = \{\lambda : \mu(\lambda) \in D_\mu\}$ be the inverse image of D_μ under the map $\lambda \mapsto \mu$ for each p. By the above arguments, the set $D = D_\lambda \cap \limsup D_p$ is an open set over which the functions

$$f_p(\lambda) = \left|\mathrm{tr}\left[\mathbf{\Theta}\left(\mathbf{A}^{1/2}\mathbf{S}\left(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S}+\lambda\mathbf{I}_q\right)^{-1}\mathbf{S}^{\mathsf{H}} - \mathbf{A}^{1/2}(\mathbf{A}+\mu\mathbf{I}_q)^{-1}\right)\right]\right|$$

converge uniformly as $\lambda \to 0$ over p. By Montel's theorem, these functions form a normal family. Since $f_p(\lambda) \searrow 0$ pointwise for $\lambda \neq 0$, this implies that $f_p(0) \searrow 0$. ■

SM4. Proofs in Section 5.

We collect the proofs of the various properties of the equivalences obtained in our paper.

SM4.1. Proof of Remark 5.1.

Proof. Recall from (5.2) that for $\alpha \in (0,\infty)$,

$$\lambda_0(\alpha) = \mu_0\left(1-\tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu_0(\alpha)\mathbf{I})^{-1}\right]\right). \tag{SM4.1}$$

From the statement of Remark 5.1, $\lim_{\alpha\nearrow\infty}\mu_0(\alpha) = -\lambda_{\min}^{+}(\mathbf{A})$. We will argue below that

$$\lim_{\alpha\nearrow\infty}\frac{\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu_0(\alpha)\mathbf{I})^{-1}\right]}{\alpha} = 0, \tag{SM4.2}$$

which combined with (SM4.1) provides the desired result.

Observe that the limit on the left-hand side of (SM4.2) is in the indeterminate ∞/∞ form because $\lim_{\alpha\nearrow\infty}\mu(\alpha) = -\lambda_{\min}^{+}(\mathbf{A})$ and thus $\lim_{\alpha\nearrow\infty}\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu(\alpha)\mathbf{I})^{-1}\right] = \infty$. To evaluate the limit, we will appeal to L'Hôpital's rule. The derivative of the denominator with respect to α is 1, while the derivative of the numerator with respect to α is

$$\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu_0(\alpha)\mathbf{I})^{-2}\right]\frac{\partial\mu_0(\alpha)}{\partial\alpha}. \tag{SM4.3}$$

Implicitly differentiating (5.1) with respect to α, we have

$$1 = \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+\mu_0(\alpha)\mathbf{I})^{-3}\right]\frac{\partial\mu_0(\alpha)}{\partial\alpha}. \tag{SM4.4}$$

Substituting for $\frac{\partial\mu_0(\alpha)}{\partial\alpha}$ from (SM4.4) into (SM4.3), we can write the derivative of the numerator as

$$\frac{\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu_0(\alpha)\mathbf{I})^{-2}\right]}{\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+\mu_0(\alpha)\mathbf{I})^{-3}\right]}.$$

As $\alpha \nearrow \infty$ and $\mu_0(\alpha) \searrow -\lambda_{\min}^{+}(\mathbf{A})$, the limit of the quantity in the display above becomes

$$\lim_{\alpha\nearrow\infty} \frac{\lambda_{\min}^{+}(\mathbf{A})}{(\lambda_{\min}^{+}(\mathbf{A}) + \mu_0(\alpha))^2} \cdot \frac{(\lambda_{\min}^{+}(\mathbf{A}) + \mu(\alpha))^3}{(\lambda_{\min}^{+}(\mathbf{A}))^2} = \lim_{\alpha\nearrow\infty} 1 + \frac{\mu_0(\alpha)}{\lambda_{\min}^{+}(\mathbf{A})} = 1 - 1 = 0.$$

Thus, we can conclude that (SM4.2) holds, and the statement then follows. The remaining claims follow by similar calculations. ∎

SM4.2. Proof of Remark 5.2.

Proof. We start by noting that

$$\begin{aligned}\lim_{x\searrow 0} \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A} + x\mathbf{I})^{-2}\right] &= \lim_{x\searrow 0} \tfrac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i^2(\mathbf{A})}{(\lambda_i(\mathbf{A}) + x)^2} \\ &= \lim_{x\searrow 0} \tfrac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i(\mathbf{A})}{\lambda_i(\mathbf{A}) + x} = \lim_{x\searrow 0} \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A} + x\mathbf{I})^{-1}\right] \\ &= \tfrac{1}{p} \sum_{i=1}^{p} \mathbb{1}\{\lambda_i(\mathbf{A}) > 0\} = r(\mathbf{A}).\end{aligned}$$

Now, write the first equation in (4.1) in terms of α as

$$\alpha = \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}^2\left(\mathbf{A} + \mu_0\mathbf{I}\right)^{-2}\right].$$

Thus, when $\alpha = r(\mathbf{A})$, we have $\mu_0 = 0$ as the solution to the first equation of (4.1). Because $\mu \mapsto \frac{1}{p}\mathrm{tr}[\mathbf{A}^2(\mathbf{A} + \mu\mathbf{I})^{-2}]$ is monotonically decreasing in μ, if $\alpha < r(\mathbf{A})$, we have $\mu_0 > 0$, while if $\alpha > r(\mathbf{A})$, we have $\mu_0 < 0$.

Next we argue about sign pattern of λ_0. When $\alpha > r(\mathbf{A})$, we have

$$\begin{aligned}\alpha = \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A} + \mu_0\mathbf{I})^{-2}\right] = \tfrac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i^2(\mathbf{A})}{(\lambda_i(\mathbf{A}) + \mu_0)^2} &\overset{(a)}{>} \tfrac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i(\mathbf{A})}{\lambda_i(\mathbf{A}) + \mu_0} \\ &= \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A} + \mu_0\mathbf{I})^{-1}\right],\end{aligned}$$

where the inequality (a) follows because $\mu_0 < 0$. From (4.1), it thus follows that $\lambda_0 < 0$. Similarly, when $\alpha < r(\mathbf{A})$, note that

$$\begin{aligned}\alpha = \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A} + \mu_0\mathbf{I})^{-2}\right] = \tfrac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i^2(\mathbf{A})}{(\lambda_i(\mathbf{A}) + \mu_0)^2} &\overset{(b)}{<} \tfrac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i(\mathbf{A})}{\lambda_i(\mathbf{A}) + \mu_0} \\ &= \tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A} + \mu_0\mathbf{I})^{-1}\right],\end{aligned}$$

where inequality (b) follows from the fact that

$$0 < \left(\frac{\lambda_{\min}^{+}(\mathbf{A})}{\lambda_{\min}^{+}(\mathbf{A}) + \mu_0}\right)^2 < \frac{\lambda_{\min}^{+}(\mathbf{A})}{\lambda_{\min}^{+}(\mathbf{A}) + \mu_0} < 1,$$

since $\mu_0 > 0$ in this case and $\lambda_{\min}^{+}(\mathbf{A}) > 0$. From (4.1), it thus again follows that $\lambda_0 < 0$. This completes the proof. ∎

SM4.3. Proof of Proposition 5.3.

Proof. The claims follow from simple derivative calculations. We split into two cases, one with respect to λ, and the other with respect to α.

SM4.3.1. Monotonicity with respect to λ. For a fixed α, implicitly differentiating the fixed-point equation (4.4) with respect to λ, we obtain

$$1 = \frac{\partial \mu}{\partial \lambda} - \left(\tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-1}\right] - \mu\tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2}\right]\right)\frac{\partial \mu}{\partial \lambda}. \tag{SM4.5}$$

Note the following algebraic simplification:

$$\begin{aligned}\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-1} - \mu\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2} &= \mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-1}\left(\mathbf{I}-\mu\left(\mathbf{A}+\mu\mathbf{I}\right)^{-1}\right)\\ &= \mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-1}\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-1} = \mathbf{A}^2\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2}.\end{aligned} \tag{SM4.6}$$

Substituting (SM4.6) into (SM4.5), we have

$$\frac{\partial \mu}{\partial \lambda} = \frac{1}{1-\frac{1}{q}\mathrm{tr}\left[\mathbf{A}^2\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2}\right]}. \tag{SM4.7}$$

Observe that $\mu \mapsto \frac{1}{q}\mathrm{tr}[\mathbf{A}^2(\mathbf{A}+\mu\mathbf{I})^{-2}]$ is monotonically decreasing function of μ over (μ_0,∞) and because $1 = \frac{1}{q}\mathrm{tr}[\mathbf{A}^2(\mathbf{A}+\mu_0\mathbf{I})^{-2}]$ from the first equation in (4.1), the denominator of (SM4.7) is positive over (μ_0,∞). Consequently, $\frac{\partial\mu}{\partial\lambda}$ is positive, and μ is a monotonically increasing function of λ. Finally, note that as $\lambda \searrow \lambda_0$, $\mu(\lambda) \searrow \mu_0$, and as $\lambda \nearrow \infty$, $\mu(\lambda) \nearrow \infty$. This completes the proof of the first part.

SM4.3.2. Monotonicity with respect to α. We begin by writing (4.4) in α as

$$\lambda = \mu\left(1-\tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right]\right). \tag{SM4.8}$$

For a fixed λ, implicitly differentiating (4.4) with respect to α, we have

$$0 = \frac{\partial \mu}{\partial \alpha} + \frac{\mu}{\alpha^2}\frac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right] - \frac{1}{\alpha}\left(\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right] - \mu\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-2}\right]\right)\frac{\partial \mu}{\partial \alpha}. \tag{SM4.9}$$

Solving for $\frac{\partial\mu}{\partial\alpha}$, we obtain

$$\frac{\partial \mu}{\partial \alpha} = \frac{-\frac{1}{\alpha^2}\mu\frac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right]}{1-\frac{1}{\alpha}\frac{1}{p}\left(\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}-\mu\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-2}\right]\right)}.$$

Similar to the part above, substituting the relation (SM4.6) into (SM4.9) and simplifying yields

$$\frac{\partial \mu}{\partial \alpha} = \frac{-\frac{1}{\alpha}\mu\frac{1}{q}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right]}{1-\frac{1}{q}\mathrm{tr}\left[\mathbf{A}^2\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2}\right]}. \tag{SM4.10}$$

Because the denominator of (SM4.10) is positive from (4.1) as argued above and $\mathrm{tr}[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}]$ is positive for $\mu \in (\mu_0, \infty)$, the sign of $\frac{\partial \mu}{\partial \alpha}$ is opposite the sign of μ. Because when $\lambda \geq 0$, $\mu \geq 0$ (from the first part of Remark 5.4), in this case, $\frac{\partial \mu}{\partial \alpha}$ is negative, and μ is monotonically decreasing in α. When $\lambda < 0$, for $\alpha \leq r(\mathbf{A})$, we have $\mu(\lambda) \geq 0$ (from the second part of Remark 5.4). Thus, over $(0, r(\mathbf{A}))$, μ is monotonically decreasing in α. On the other hand, for $\alpha > r(\mathbf{A})$, $\mu(\lambda) < 0$ (since $\mathrm{sign}(\mu(\lambda)) = \mathrm{sign}(\lambda)$ and $\lambda < 0$), and consequently, μ is monotonically increasing in α over $(r(\mathbf{A}), \infty)$.

Finally, to obtain the limit of $\mu(\alpha)$ as $\alpha \searrow 0$, we write (SM4.8) as

$$\lambda\alpha = \mu\alpha - \mu\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right].$$

Now, for any $\lambda \in (\lambda_0, \infty)$, $\lim_{\alpha\searrow 0} \lambda\alpha = 0$. Thus, we have

$$\lim_{\alpha\searrow 0} \mu(\alpha) = \lim_{\alpha\searrow 0} f^{-1}(\alpha),$$

where $f(x) = \frac{1}{p}\mathrm{tr}[\mathbf{A}(\mathbf{A}+x\mathbf{I})^{-1}]$. Observe that function f is strictly decreasing over (μ_0, ∞), and $\lim_{x\nearrow\infty} f(x) = 0$. Hence, the function f^{-1} is strictly decreasing and $\lim_{\alpha\searrow 0} f^{-1}(\alpha) = \infty$. This provides us with the first limit. To obtain the limit of $\mu(\alpha)$ as $\alpha \nearrow \infty$, write from (4.4)

$$\mu = \lambda + \tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mu\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right].$$

Observe that $\frac{1}{p}\mathrm{tr}[\mu\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}]$ is bounded for $\mu \in (\mu_0, \infty)$. Thus, taking the limit $\alpha \nearrow \infty$, we conclude that $\lim_{\alpha\nearrow\infty} \mu(\alpha) = \lambda$. This finishes the second part, and completes the proof. ■

SM4.4. Proof of Remark 5.4.

Proof. We start by writing (4.4) in terms of α as

$$\lambda = \mu\left(1 - \tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right]\right).$$

For the subsequent argument, it will help to rearrange the terms in the equation in display above to arrive at the following equivalent equation:

$$1 - \frac{\lambda}{\mu} = \tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}\right]. \tag{SM4.11}$$

We consider two separate cases depending on $\lambda \geq 0$ and $\lambda < 0$.

Case $\lambda \geq 0$: Fix $\alpha > 0$. Observe that the left side of (SM4.11) is an increasing function of μ, and the right side of (SM4.11) is a decreasing function of μ. As μ varies from 0^+ to ∞, the right hand side decreases from $\frac{r(A)}{\alpha}$ to 0, while the left hand side increases from $-\infty$ to 1. Since $1 > 0$, there is a unique intersection for $\mu \geq 0$.

Case $\lambda < 0$: Fix $\alpha \leq r(\mathbf{A})$. For this subcase, from 5.2, $\mu_0 \geq 0$. Thus, there is a unique intersection for $\mu \geq 0$. Fix now $\alpha > r(\mathbf{A})$. For this subcase, the term in the parenthesis of (4.4) is positive. Thus, $\mathrm{sign}(\mu) = \mathrm{sign}(\lambda)$.

This completes all the three cases, and finishes the proof. ■

SM4.5. Proof of Proposition 5.5.

Proof. Recall that $\mu_0 > -\lambda_{\min}^{+}(\mathbf{A})$. For $x \in (\mu_0, \infty)$, observe that

$$\frac{\partial}{\partial x}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+x\mathbf{I})^{-1}\right] = -\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+x\mathbf{I})^{-2}\right] < 0,$$

$$\frac{\partial^2}{\partial x^2}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+x\mathbf{I})^{-1}\right] = 2\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+x\mathbf{I})^{-3}\right] > 0.$$

Thus, the function

$$x \mapsto \tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+x\mathbf{I})^{-1}\right]$$

is strictly decreasing and convex over (μ_0, ∞), and consequently the function

$$x \mapsto \tfrac{1}{q}\mathrm{tr}\left[x\mathbf{A}(\mathbf{A}+x\mathbf{I})^{-1}\right] = \tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}(\mathbf{I}-\mathbf{A}(\mathbf{A}+x\mathbf{I})^{-1})\right] = \tfrac{1}{q}\mathrm{tr}[\mathbf{A}] - \tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+x\mathbf{I})^{-1}\right]$$

is strictly increasing and concave over (μ_0, ∞). Hence, the function f (appearing in the right-hand side of (4.4) in μ) defined by

$$f(x) = x - x\tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+x\mathbf{I})^{-1}\right] = x\left(1 - \tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}(\mathbf{A}+x\mathbf{I})^{-1}\right]\right) \tag{SM4.12}$$

is strictly increasing and convex over (μ_0, ∞).

Now, observe from (4.4) that for a given λ, $\mu(\lambda) = f^{-1}(\lambda)$, where f is as defined in (SM4.12). Because inverse of a strictly increasing, continuous, and convex function is strictly increasing, continuous, and concave (see, e.g., Proposition 3 of [SM3]), we conclude that $\lambda \mapsto \mu(\lambda)$ where $\mu(\lambda)$ solves (4.4) is concave in λ over (λ_0, ∞). We remark that, more directly, we can also compute the second derivative of $\mu(\lambda)$ with respect to λ. From (SM4.7), we have

$$\frac{\partial \mu}{\partial \lambda} = \frac{1}{1 - \frac{1}{\alpha}\frac{1}{p}\mathrm{tr}\left[\mathbf{A}^2\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2}\right]}. \tag{SM4.13}$$

Taking partial derivative of (SM4.13) with respect to λ, we get

$$\frac{\partial^2 \mu}{\partial \lambda^2} = \frac{-2\frac{1}{\alpha}\frac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+\mu\mathbf{I})^{-3}\right]}{\left(1 - \frac{1}{\alpha}\frac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+\mu\mathbf{I})^{-2}\right]\right)^2}\frac{\partial \mu}{\partial \lambda} = \frac{-2\frac{1}{\alpha}\frac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+\mu\mathbf{I})^{-3}\right]}{\left(1 - \frac{1}{\alpha}\frac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+\mu\mathbf{I})^{-2}\right]\right)^3} < 0,$$

from which the concavity claim follows.

Using the concavity of μ in λ, we can write for $\lambda, \widetilde{\lambda} \in (\lambda_0, \infty)$,

$$\mu(\lambda) \le \mu(\widetilde{\lambda}) + \frac{\partial \mu}{\partial \lambda}\bigg|_{\lambda=\widetilde{\lambda}} (\lambda - \widetilde{\lambda}). \tag{SM4.14}$$

Now, from (4.4), for any $\widetilde{\lambda} \in (\lambda_0, \infty)$, we have

$$\mu(\widetilde{\lambda}) - \widetilde{\lambda} = \tfrac{1}{q}\mathrm{tr}\left[\mu(\widetilde{\lambda})\mathbf{A}(\mathbf{A}+\mu(\widetilde{\lambda})\mathbf{I})^{-1}\right] = \tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mu(\widetilde{\lambda})\mathbf{A}(\mathbf{A}+\mu(\widetilde{\lambda})\mathbf{I})^{-1}\right]. \tag{SM4.15}$$

Substituting in (SM4.15) in (SM4.14) yields

$$\mu(\lambda) \leq \frac{\partial \mu}{\partial \lambda}\Big|_{\lambda=\widetilde{\lambda}} \lambda + \tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mu(\widetilde{\lambda})\mathbf{A}(\mathbf{A}+\mu(\widetilde{\lambda})\mathbf{I})^{-1}\right].$$

From Proposition 5.3, $\lambda \mapsto \mu(\lambda)$ is monotonically increasing in λ and $\lim_{\lambda\nearrow\infty}\mu(\lambda)=\infty$. In addition, $\mu \mapsto \mathrm{tr}[\mathbf{A}^2(\mathbf{A}+\mu\mathbf{I})^{-2}]$ is monotonically decreasing in μ and $\lim_{\mu\nearrow\infty}\mathrm{tr}[\mathbf{A}^2(\mathbf{A}+\mu\mathbf{I})^{-2}] = 0$, while $\mu \mapsto \mathrm{tr}[\mu\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}]$ is monotonically increasing in μ, and $\lim_{\mu\nearrow\infty}\mathrm{tr}[\mu\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}] = \mathrm{tr}[\mathbf{A}]$. Thus, from (SM4.13), choosing $\widetilde{\lambda}$ large enough so that $\mu(\widetilde{\lambda})$ is large enough, for any $\epsilon > 0$, we can write

$$\frac{\partial \mu}{\partial \lambda}\Big|_{\lambda=\widetilde{\lambda}} = \frac{1}{1-\frac{1}{\alpha}\frac{1}{p}\mathrm{tr}\left[\mathbf{A}^2(\mathbf{A}+\mu(\widetilde{\lambda})\mathbf{I})^{-2}\right]} \leq 1+\epsilon, \tag{SM4.16}$$

$$\tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mu(\widetilde{\lambda})\mathbf{A}(\mathbf{A}+\mu(\widetilde{\lambda})\mathbf{I})^{-1}\right] \leq \tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}\left[\mathbf{A}\right]+\epsilon. \tag{SM4.17}$$

Combining (SM4.15)–(SM4.17), one then has

$$\mu(\lambda) \leq (1+\epsilon)\lambda + \tfrac{1}{\alpha}\tfrac{1}{p}\mathrm{tr}[\mathbf{A}] + \epsilon.$$

Since the inequality holds for any arbitrary ϵ, the desired upper bound on $\mu(\lambda)$ follows. For the lower bound, observe from (SM4.15) that for any $\lambda \in (\lambda_0, \infty)$

$$\mu(\lambda) = \lambda + \tfrac{1}{q}\mathrm{tr}\left[\mu(\lambda)\mathbf{A}(\mathbf{A}+\mu(\lambda)\mathbf{I})^{-1}\right].$$

From Remark 5.4, $\mu(\lambda) \geq 0$ either when $\lambda \geq 0$, or when $\alpha \leq r(\mathbf{A})$. In either of the cases, the term $\frac{1}{q}\mathrm{tr}[\mu(\lambda)\mathbf{A}(\mathbf{A}+\mu(\lambda)\mathbf{I})^{-1}]$ is positive, and thus $\mu(\lambda) \geq \lambda$. Finally, the limit as $\lambda \nearrow \infty$ follows simply by noting that $\mu(\lambda) \nearrow \infty$ and $\mathrm{tr}[\mu\mathbf{A}(\mathbf{A}+\mu\mathbf{I})^{-1}] \nearrow \mathrm{tr}[\mathbf{A}]$ as $\lambda \nearrow \infty$. This finishes the proof. ∎

SM4.6. Proof of Remark 5.6.

Proof. We begin by rewriting (4.6) using (4.3):

$$\mu' = \frac{\frac{\mu^3}{q}\mathrm{tr}\left[\boldsymbol{\Psi}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2}\right]}{\mu\left(1-\frac{1}{q}\mathrm{tr}\left[\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}_p\right)^{-1}\right]\right)+\frac{\mu^2}{q}\mathrm{tr}\left[\mathbf{A}\left(\mathbf{A}+\mu\mathbf{I}\right)^{-2}\right]}.$$

After dividing both the numerator and denominator by μ, we note that the denominator has a form which has already been simplified in subsection SM4.3, and immediately obtain the factorization in terms of $\frac{\partial \mu}{\partial \lambda}$. ∎

SM5. Proofs in Section 6.

This section collects proofs for various results in section 6.

SM5.1. Proof of Equation (6.1).

Proof. First, we can write $\mathbf{b} = \mathbf{L}\mathbf{x}_*$. Next, we can subtract $\mathbf{x}_*$:

$$\mathbf{x}_t - \mathbf{x}_* = (\mathbf{I}_n - \mathbf{L}^{\mathsf{H}}\mathbf{S}_t(\mathbf{S}_t^{\mathsf{H}}\mathbf{L}\mathbf{L}^{\mathsf{H}}\mathbf{S}_t)^{\dagger}\mathbf{S}_t^{\mathsf{H}}\mathbf{L})(\mathbf{x}_{t-1} - \mathbf{x}_*).$$

Because $(\mathbf{I}_n - \mathbf{L}^{\mathsf{H}}\mathbf{S}_t(\mathbf{S}_t^{\mathsf{H}}\mathbf{L}\mathbf{L}^{\mathsf{H}}\mathbf{S}_t)^{\dagger}\mathbf{S}_t^{\mathsf{H}}\mathbf{L})$ is a projection matrix and therefore idempotent,

$$\begin{aligned}\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 &= (\mathbf{x}_{t-1} - \mathbf{x}_*)^{\mathsf{H}}(\mathbf{I}_n - \mathbf{L}^{\mathsf{H}}\mathbf{S}_t(\mathbf{S}_t^{\mathsf{H}}\mathbf{L}\mathbf{L}^{\mathsf{H}}\mathbf{S}_t)^{\dagger}\mathbf{S}_t^{\mathsf{H}}\mathbf{L})(\mathbf{x}_{t-1} - \mathbf{x}_*)\\ &\simeq (\mathbf{x}_{t-1} - \mathbf{x}_*)^{\mathsf{H}}(\mathbf{I}_n - \mathbf{L}^{\mathsf{H}}\big(\mathbf{L}\mathbf{L}^{\mathsf{H}} + \mu\mathbf{I}_p\big)^{-1}\mathbf{L})(\mathbf{x}_{t-1} - \mathbf{x}_*)\\ &\le \rho\|\mathbf{x}_{t-1} - \mathbf{x}_*\|_2^2,\end{aligned}$$

where the asymptotic equivalence is the result of applying Theorem 4.1 with $\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathsf{H}}$, and $\rho = \lambda_{\max}(\mathbf{I}_n - \mathbf{L}^{\mathsf{H}}\big(\mathbf{L}\mathbf{L}^{\mathsf{H}} + \mu\mathbf{I}_p\big)^{-1}\mathbf{L})$. Thus the stated convergence bound holds almost surely for any t. ∎

SM5.2. Proof of Remark 6.1.

Proof. Since $\lambda = 0$, we know that $\alpha \mapsto \mu$ is an invertible mapping from $(0, r(\mathbf{L}))$ onto $\mu \in (0, \infty)$ by Proposition 5.3 and Remark 5.4, while for $\alpha \ge r(\mathbf{L})$, $\mu = 0$ and therefore a solution is reached in $t_\varepsilon = 1$ steps. Thus, it remains only to consider $\mu \in (0, \infty)$. Generalizing to galactic inversion algorithms of complexity $O(m^{1+\delta}p)$, we can write the relative computation factor in terms of μ as

$$\alpha^{1+\delta}t_\varepsilon = \left(\tfrac{1}{n}\mathrm{tr}\left[\mathbf{L}\mathbf{L}^{\mathsf{H}}\big(\mathbf{L}\mathbf{L}^{\mathsf{H}} + \mu\mathbf{I}_n\big)^{-1}\right]\right)^{1+\delta}\left\lceil\frac{\log\left(\frac{a_+\|\mathbf{x}_0 - \mathbf{x}_*\|_2^2}{\varepsilon}\right)}{\log(1 + \frac{a_-}{\mu})}\right\rceil,$$

where $a_+ \triangleq \lambda_{\max}(\mathbf{L}\mathbf{L}^{\mathsf{H}})$ and $a_- \triangleq \lambda_{\min}^{+}(\mathbf{L}\mathbf{L}^{\mathsf{H}})$. For any fixed μ (and equivalently any fixed $\alpha < r(\mathbf{L})$), as $\varepsilon \searrow 0$, we clearly have $t_\varepsilon \nearrow \infty$. For fixed ε, the limiting behavior as $\mu \nearrow \infty$ (equivalently $\alpha \searrow 0$) is determined by the ratio

$$\frac{\left(\frac{1}{n}\mathrm{tr}\left[\mathbf{L}\mathbf{L}^{\mathsf{H}}\big(\mathbf{L}\mathbf{L}^{\mathsf{H}} + \mu\mathbf{I}_n\big)^{-1}\right]\right)^{1+\delta}}{\log(1 + \frac{a_-}{\mu})} = \frac{\left(\frac{1}{n}\mathrm{tr}\left[\mathbf{L}\mathbf{L}^{\mathsf{H}}\big(\mathbf{L}\mathbf{L}^{\mathsf{H}} + \mu\mathbf{I}_n\big)^{-1}\right]\right)^{1+\delta}}{\frac{a_-}{\mu} + o(\frac{1}{\mu})} \searrow 0. \quad ∎$$

SM6. Proofs for free sketching.

We first establish some notation and useful lemmas. We next provide the proof details for Theorem 7.2 and then provide minor derivation details for orthogonal sketching in Corollary 7.3.

With some abuse of notation, we will let $\mathbf{A}$ denote both the finite $p \times p$ matrix as well as the limiting element in the free probability space (which can be understood for example as being a bounded linear operator on a Hilbert space). We note that all notions that we need, in particular logarithms of determinants, are well defined in this limit as well, as long as they are appropriately normalized. For this reason, we define normalized versions $\overline{\log\det}(\mathbf{A}) \triangleq \frac{1}{p}\log\det(\mathbf{A})$ and $\overline{\mathrm{tr}}[\mathbf{A}] \triangleq \frac{1}{p}\mathrm{tr}[\mathbf{A}]$ which extend nicely to the limit.

We will also use the following straightforward result from differential calculus allowing us to draw conclusions about first derivatives from second derivatives.

Lemma SM6.1 (Controlling derivatives). *Let $g\colon \mathcal{T} \times \mathcal{Z} \subseteq \mathbb{C}^2 \to \mathbb{C}$ be holomorphic. Then, for each $t \in \mathcal{T}$, if $\inf_{z \in \mathcal{Z}} \left|\frac{\partial g(t,z)}{\partial t}\right| = 0$ and $\frac{\partial^2 g(t,z)}{\partial t \partial z} = 0$ for all $z \in \mathcal{Z}$, then $\frac{\partial g(t,z)}{\partial t} = 0$ for all $z \in \mathcal{Z}$.*

Proof. By the fundamental theorem of calculus, for some $z_0 \in \mathcal{Z}$,

$$\begin{aligned}\frac{\partial g(t,z)}{\partial t} &= \frac{\partial}{\partial t}\left(\int_{z_0}^{z} \frac{\partial g(t,u)}{\partial u} du + g(t,z_0)\right)\\ &= \int_{z_0}^{z} \frac{\partial^2 g(t,u)}{\partial t \partial u} du + \frac{\partial g(t,z_0)}{\partial t}\\ &= 0,\end{aligned}$$

where the final equality follows by our hypotheses since z_0 is arbitrary. ■

We lastly introduce a series of invertible transformations from free probability [SM5]:

$$G_{\mathbf{A}}(z) = \overline{\operatorname{tr}}\big[\big(z\mathbf{I} - \mathbf{A}\big)^{-1}\big] \longleftrightarrow M_{\mathbf{A}}(z) = \frac{1}{z} G_{\mathbf{A}}\left(\frac{1}{z}\right) - 1 \longleftrightarrow \mathcal{S}_{\mathbf{A}}(z) = \frac{1+z}{z} M_{\mathbf{A}}^{\langle -1 \rangle}(z),$$

which are the Cauchy transform (negative of the Stieltjes transform), moment generating series $M_{\mathbf{A}}(z) = \sum_{k=1}^{\infty} \overline{\operatorname{tr}}[\mathbf{A}^k] z^k$, and S-transform of $\mathbf{A}$, respectively. Here $M_{\mathbf{A}}^{\langle -1 \rangle}$ denotes inverse under composition of $M_{\mathbf{A}}$. We also recall the property of free products that $\mathcal{S}_{\mathbf{AB}}(z) = \mathcal{S}_{\mathbf{A}}(z)\mathcal{S}_{\mathbf{B}}(z)$, or equivalently $M_{\mathbf{AB}}^{\langle -1 \rangle}(z) = \frac{1+z}{z} M_{\mathbf{A}}^{\langle -1 \rangle}(z) M_{\mathbf{B}}^{\langle -1 \rangle}(z) = \mathcal{S}_{\mathbf{A}}(z) M_{\mathbf{B}}^{\langle -1 \rangle}(z)$.

SM6.1. Proof of Theorem 7.2.

Proof. We begin with the simpler case where $\mathbf{\Theta}$ in the equivalence definition is such that $p\mathbf{\Theta}$ has uniformly bounded operator norm. For this proof, we will simply write $\mathbf{\Theta}$ instead of $p\mathbf{\Theta}$ to be compatible with the normalized trace. First, we can decompose $\mathbf{\Theta}$ into real and imaginary parts $\mathbf{\Theta} = \mathbf{\Theta}_{\mathrm{Re}} + i\mathbf{\Theta}_{\mathrm{Im}}$, so without loss of generality we can assume $\mathbf{\Theta}$ is real. Similarly, we note that $\overline{\operatorname{tr}}[\mathbf{\Theta}\mathbf{B}] = \overline{\operatorname{tr}}[\frac{1}{2}(\mathbf{\Theta} + \mathbf{\Theta}^{\mathsf{H}})\mathbf{B}]$ for any self-adjoint matrix $\mathbf{B} \in \mathbb{C}^{p \times p}$, so we can assume $\mathbf{\Theta}$ is symmetric and therefore diagonalizable without loss of generality. We let $\widetilde{\mathbf{S}} = (\mathbf{S}\mathbf{S}^{\mathsf{H}})^{1/2}$ and note that we can now work entirely in dimension p instead of both dimensions p and q:

$$\mathbf{S}\big(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S} - z\mathbf{I}_q\big)^{-1}\mathbf{S}^{\mathsf{H}} = \widetilde{\mathbf{S}}\big(\widetilde{\mathbf{S}}\mathbf{A}\widetilde{\mathbf{S}} - z\mathbf{I}_p\big)^{-1}\widetilde{\mathbf{S}}.$$

Consider now the limit where $(\mathbf{\Theta}, \mathbf{A}, \widetilde{\mathbf{S}})$ have converged spectrally with $\widetilde{\mathbf{S}}$ free from $\mathbf{\Theta}$ and $\mathbf{A}$. We need only show that for some $\zeta \in \mathbb{C}^+$,

$$\overline{\operatorname{tr}}[\mathbf{\Theta}\widetilde{\mathbf{S}}\big(\widetilde{\mathbf{S}}\mathbf{A}\widetilde{\mathbf{S}} - z\mathbf{I}\big)^{-1}\widetilde{\mathbf{S}}] = \overline{\operatorname{tr}}[\mathbf{\Theta}\big(\mathbf{A} - \zeta\mathbf{I}\big)^{-1}].$$

We now define parameterized operators $\mathbf{B}_{t,\zeta} = \mathbf{A} + t\mathbf{\Theta} - \zeta\mathbf{I}$ and $\mathbf{B}_{t,z}^{\widetilde{\mathbf{S}}} = \widetilde{\mathbf{S}}(\mathbf{A} + t\mathbf{\Theta})\widetilde{\mathbf{S}} - z\mathbf{I}$. By Jacobi's formula, we have the following two equalities

$$\begin{aligned}\overline{\operatorname{tr}}[\mathbf{\Theta}\big(\mathbf{A} - \zeta\mathbf{I}\big)^{-1}] &= \frac{\partial \overline{\log\det}(\mathbf{B}_{t,\zeta})}{\partial t}\Big|_{t=0},\\ \overline{\operatorname{tr}}[\mathbf{\Theta}\widetilde{\mathbf{S}}\big(\widetilde{\mathbf{S}}\mathbf{A}\widetilde{\mathbf{S}} - z\mathbf{I}\big)^{-1}\widetilde{\mathbf{S}}] &= \frac{\partial \overline{\log\det}(\mathbf{B}_{t,z}^{\widetilde{\mathbf{S}}})}{\partial t}\Big|_{t=0}.\end{aligned}$$

Suppose that $z \mapsto \zeta$ is a holomorphic map. Then another way of stating our condition to be proven is that for $t = 0$ and all $z \in \mathbb{C}^+$, we must have $\frac{\partial g(t,z)}{\partial t} = 0$, where

$$g(t,z) = \overline{\log\det}(\mathbf{B}_{t,\zeta}) - \overline{\log\det}(\mathbf{B}^{\widetilde{\mathbf{S}}}_{t,z}).$$

By Lemma SM6.1, it is sufficient to show that $\mathrm{Im}(\zeta) \nearrow \infty$ as $z \to i\infty$ (implying the condition $\inf_{z\in\mathcal{Z}} |\frac{\partial g(t,z)}{\partial t}| = 0$) and that $\frac{\partial^2 g(t,z)}{\partial t \partial z} = 0$ for all $z \in \mathbb{C}^+$.

We therefore seek a choice of $z \mapsto \zeta$ that satisfies these conditions. In particular, we need only to show that the last condition holds, and the rest will follow. The main idea is that we can control the derivative of g in t, which has a dependence on $\mathbf{\Theta}$, in terms of the derivative of g in z, which does not. For succinctness in the subsequent arguments, we will use the following notation for derivatives: for a function $f_t\colon \mathbb{C} \to \mathbb{C}$, we denote $\dot{f}_t(z) = \frac{\partial f_t(z)}{\partial t}$ and $f_t'(z) = \frac{\partial f_t(z)}{\partial z}$. That is, $\dot{f}_t$ is the derivative with respect to its index t, and f_t' is the derivative with respect to its argument (typically z). Although we omit the argument z of ζ, we let $\zeta' = \frac{\partial \zeta}{\partial z}$.

Define $\mathbf{A}_t \triangleq \mathbf{A} + t\mathbf{\Theta}$. Appealing again to Jacobi's formula, we have two further equalities:

$$\frac{\partial \overline{\log\det}(\mathbf{B}_{t,\zeta})}{\partial z} = -\overline{\mathrm{tr}}\big[\mathbf{B}_{t,\zeta}^{-1}\big]\zeta' = G_{\mathbf{A}_t}(\zeta)\zeta',$$
$$\frac{\partial \overline{\log\det}(\mathbf{B}^{\widetilde{\mathbf{S}}}_{t,z})}{\partial z} = -\overline{\mathrm{tr}}\big[\mathbf{B}^{\widetilde{\mathbf{S}}\,-1}_{t,\zeta}\big] = G_{\mathbf{A}_t\widetilde{\mathbf{S}}^2}(z).$$

The last equality follows because $\widetilde{\mathbf{S}}\mathbf{A}_t\widetilde{\mathbf{S}}$ has the same spectrum as $\mathbf{A}_t\widetilde{\mathbf{S}}^2$ (to see this, note that they have the same moments due to the cyclic invariance of the tracial state $\overline{\mathrm{tr}}$). We therefore need ζ such that at $t = 0$, for all $z \in \mathbb{C}^+$,

$$\dot{G}_{\mathbf{A}_t\widetilde{\mathbf{S}}^2}(z) = \dot{G}_{\mathbf{A}_t}(\zeta)\zeta'.$$

Equivalently, in terms of the moment generating series, we need

$$\frac{\dot{M}_{\mathbf{A}_t\widetilde{\mathbf{S}}^2}(\frac{1}{z})}{z} = \frac{\dot{M}_{\mathbf{A}_t}(\frac{1}{\zeta})\zeta'}{\zeta}. \tag{SM6.1}$$

This is finally the condition that we will show.

Now, from the property of free products, we know that for $m \in \mathbb{C}^-$,

$$M^{\langle -1\rangle}_{\mathbf{A}_t\widetilde{\mathbf{S}}^2}(m) = \mathcal{S}_{\widetilde{\mathbf{S}}^2}(m) M^{\langle -1\rangle}_{\mathbf{A}_t}(m).$$

Choose now $m = M_{\mathbf{A}_t\widetilde{\mathbf{S}}^2}(\frac{1}{z})$, which gives us

$$z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m) = \frac{1}{M^{\langle -1\rangle}_{\mathbf{A}_t}(m)} \iff m = M_{\mathbf{A}_t}\left(\frac{1}{z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)}\right). \tag{SM6.2}$$

Matching the forms of (SM6.1) and (SM6.2), we can form a guess of $\zeta = z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)$, which we can also prove is the correct choice. To do so, we note that m is parameterized by both t and z. We first implicitly differentiate with respect to t:

$$\dot{m} = \dot{M}_{\mathbf{A}_t}\left(\frac{1}{z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)}\right) - M'_{\mathbf{A}_t}\left(\frac{1}{z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)}\right)\frac{\mathcal{S}'_{\widetilde{\mathbf{S}}^2}(m)\dot{m}}{z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)^2},$$

which after plugging in $\zeta = z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)$ gives us

$$\dot{m} = \frac{\dot{M}_{\mathbf{A}_t}\left(\frac{1}{\zeta}\right)}{1 + M'_{\mathbf{A}_t}\left(\frac{1}{\zeta}\right)\frac{z\mathcal{S}'_{\widetilde{\mathbf{S}}^2}(m)}{\zeta^2}}.$$

Next, noting that $\zeta' = \mathcal{S}_{\widetilde{\mathbf{S}}^2}(m) + z\mathcal{S}'_{\widetilde{\mathbf{S}}^2}(m)m'$, we differentiate (SM6.2) with respect to z:

$$m' = -M'_{\mathbf{A}_t}\left(\frac{1}{\zeta}\right)\frac{\zeta'}{\zeta^2} \implies \zeta' = \frac{\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)}{1 + M'_{\mathbf{A}_t}\left(\frac{1}{\zeta}\right)\frac{z\mathcal{S}'_{\widetilde{\mathbf{S}}^2}(m)}{\zeta^2}}.$$

We can deduce from the previous two equations and the fact that $\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m) = \frac{\zeta}{z}$ that

$$\dot{m} = \dot{M}_{\mathbf{A}_t}\left(\frac{1}{\zeta}\right)\frac{z\zeta'}{\zeta},$$

which is equivalent to (SM6.1), which we needed to show. Therefore, specializing to $t = 0$, we have that $\zeta = z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(M_{\mathbf{A}\widetilde{\mathbf{S}}^2}(\frac{1}{z}))$ makes the the second derivative condition of Lemma SM6.1 satisfied. Additionally, we have that $\operatorname{Im}(\zeta) \nearrow \infty$ as $z \to i\infty$: note that $M_{\mathbf{A}\widetilde{\mathbf{S}}^2}(\frac{1}{z}) = \overline{\operatorname{tr}}(\mathbf{A}\widetilde{\mathbf{S}}^2)\frac{1}{z} + o(\frac{1}{z})$ and similarly $\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m) = \frac{1}{\overline{\operatorname{tr}}(\widetilde{\mathbf{S}}^2)} + o(m)$, such that $\zeta = z(\frac{1}{\overline{\operatorname{tr}}(\widetilde{\mathbf{S}}^2)} + o(\frac{1}{z}))$.

To obtain the equation for ζ in terms of $\mathcal{S}_{\widetilde{\mathbf{S}}^2}$ and $M_{\mathbf{A}}$, combine $\zeta = z\mathcal{S}_{\widetilde{\mathbf{S}}^2}(m)$ and (SM6.2). To obtain the equation for ζ in terms of $\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S}$ and z, use the fact that $m = M_{\widetilde{\mathbf{S}}\mathbf{A}\widetilde{\mathbf{S}}}(\frac{1}{z})$.

Trace norm bounded $\mathbf{\Theta}$. For more general trace norm bounded $\mathbf{\Theta}$, such as rank one vector outer products, $p\mathbf{\Theta}$ does not have bounded operator norm and so the previous argument cannot be applied. However, with a stronger notion of freeness, called first-order or infinitesimal freeness [SM6], this extension is also possible. Following [SM6, SM2], the key condition is to require sufficiently fast convergence of $G_{\mathbf{A}\widetilde{\mathbf{S}}^2}(z)$ in p. Concretely, let $\widetilde{G}_{\mathbf{A}\widetilde{\mathbf{S}}^2}$ be the Cauchy transform of the multiplicative free convolution of the spectra of $\mathbf{A}$ and $\widetilde{\mathbf{S}}^2$—that is, what the Cauchy transform of $\mathbf{A}\widetilde{\mathbf{S}}^2$ would be if $\mathbf{A}$ and $\widetilde{\mathbf{S}}^2$ were free, which is not possible in finite dimensions. Then we need almost sure convergence in the sense that

$$G_{\mathbf{A}\widetilde{\mathbf{S}}^2}(z) = \widetilde{G}_{\mathbf{A}\widetilde{\mathbf{S}}^2}(z) + \epsilon(p) \quad \text{where} \quad \epsilon(p) = o(\tfrac{1}{p}).$$

Fortunately, this rate is known to hold in the almost sure sense when $\widetilde{\mathbf{S}}$ is a unitarily invariant ensemble [SM2, Theorem 3.5], so this assumption is satisfiable.

We apply the same approach as in the previous case when $p\mathbf{\Theta}$ had bounded operator norm. Even though $p\mathbf{\Theta}$ now does not converge to a limiting bounded operator, the first-order terms like $\dot{G}_{\mathbf{A}}(z)$ remain well-defined due to the bounded trace norm. We note that a trace norm bounded perturbation does not change the limiting spectral distribution, which means that $\mathbf{A}_t$ and $\mathbf{A}$ asymptotically have the same spectrum and thus the same result of multiplicative convolution with $\widetilde{\mathbf{S}}^2$. However, given some $t(p)$, we have the Taylor expansion

$$G_{\mathbf{A}_{t(p)}\widetilde{\mathbf{S}}^2}(z) = \widetilde{G}_{\mathbf{A}\widetilde{\mathbf{S}}^2}(z) + t(p)\dot{\widetilde{G}}_{\mathbf{A}\widetilde{\mathbf{S}}^2}(z) + O(t(p)^2 + \epsilon(p)).$$

Meanwhile, also taking the Taylor expansion of $G_{\mathbf{A}_{t(p)}}(\zeta)$,

$$G_{\mathbf{A}_{t(p)}}(\zeta)\zeta' = G_{\mathbf{A}}(\zeta)\zeta' + t(p)\dot{G}_{\mathbf{A}}(\zeta)\zeta' + O(t(p)^2).$$

Therefore, choosing $t(p) = \sqrt{\frac{1}{p}\epsilon(p)}$ and taking the derivative of these two expansions, we can finally say that

$$\begin{aligned}
&\operatorname{tr}\left[\mathbf{\Theta}\left(\left(\mathbf{A} - \zeta\mathbf{I}_p\right)^{-1}\zeta' - \widetilde{\mathbf{S}}\left(\widetilde{\mathbf{S}}\mathbf{A}\widetilde{\mathbf{S}} - z\mathbf{I}_p\right)^{-1}\widetilde{\mathbf{S}}\right)\right] \\
&\quad = \dot{G}_{\mathbf{A}}(\zeta)\zeta' - \dot{\widetilde{G}}_{\mathbf{A}\widetilde{\mathbf{S}}^2}(z) + O\left(t(p) + \tfrac{\epsilon(p)}{t(p)}\right) \\
&\quad = O\left(t(p) + \tfrac{\epsilon(p)}{t(p)}\right) \\
&\quad \xrightarrow{\text{a.s.}} 0,
\end{aligned}$$

where the final equality follows by choosing ζ as in the bounded operator norm case. Then by similar application of Lemma SM6.1 as before, we obtain the desired equivalence. ∎

SM6.2. Proof details for orthogonal sketching. To obtain the S-transform for the normalized orthogonal sketch, we first note that $\mathbf{Q}\mathbf{Q}^{\mathsf{H}}$ has q eigenvalues of $\frac{1}{\alpha}$ and $p-q$ eigenvalues of 0. Therefore, it has

$$M_{\mathbf{Q}\mathbf{Q}^{\mathsf{H}}}(z) = \overline{\operatorname{tr}}\left[\mathbf{Q}\mathbf{Q}^{\mathsf{H}}\left(\tfrac{1}{z}\mathbf{I} - \mathbf{Q}\mathbf{Q}^{\mathsf{H}}\right)^{-1}\right] = \frac{\alpha z}{\alpha - z},$$

which has inverse $M^{\langle -1\rangle}_{\mathbf{Q}\mathbf{Q}^{\mathsf{H}}}(w) = \frac{\alpha w}{\alpha + w}$ and therefore $S_{\mathbf{Q}\mathbf{Q}^{\mathsf{H}}}(w) = \frac{\alpha(1+w)}{\alpha+w}$.

To obtain the fixed point equation, we first solve $\gamma = \lambda S_{\mathbf{Q}\mathbf{Q}^{\mathsf{H}}}(w)$ for w:

$$w = \frac{\alpha(\lambda - \gamma)}{\gamma - \alpha\lambda}.$$

Then, we plug in $w = -\overline{\operatorname{tr}}[\mathbf{A}\left(\mathbf{A} + \gamma\mathbf{I}_p\right)^{-1}] = \gamma\overline{\operatorname{tr}}[\left(\mathbf{A} + \gamma\mathbf{I}_p\right)^{-1}] - 1$:

$$\gamma\overline{\operatorname{tr}}[\left(\mathbf{A} + \gamma\mathbf{I}_p\right)^{-1}] = \frac{\alpha(\lambda - \gamma)}{\gamma - \alpha\lambda} + 1 = \frac{\gamma(1 - \alpha)}{\gamma - \alpha\lambda}.$$

The stated relation follows directly.

REFERENCES

[SM1] Z. D. Bai and J. W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab., 26 (1998), pp. 316–345, https://doi.org/10.1214/aop/1022855421.

[SM2] G. Cébron, A. Dahlqvist, and F. Gabriel, *Freeness of type B and conditional freeness for random matrices*, preprint, arXiv:2205.01926, 2022.

[SM3] J.-B. Hiriart-Urruty and J.-E. Martínez-Legaz, *New formulas for the Legendre–Fenchel transform*, J. Math. Anal. Appl., 288 (2003), pp. 544–555, https://doi.org/10.1016/j.jmaa.2003.09.012.

[SM4] D. LeJeune, *Ridge regularization by randomization in linear ensembles*, PhD thesis, Rice University, 2022.

[SM5] J. Mingo and R. Speicher, *Free Probability and Random Matrices*, Fields Inst. Monogr., Springer New York, 2017, https://doi.org/10.1007/978-1-4939-6942-5.

[SM6] D. Shlyakhtenko, *Free probability of type-B and asymptotics of finite-rank perturbations of random matrices*, Indiana Univ. Math. J., 67 (2018), pp. 971–991, https://doi.org/10.1512/iumj.2018.67.7294.