Matrix Dyson equation for correlated linearizations and test error of random features regression

Hugo Latourelle-Vigeant[1] **and Elliot Paquette**[1]

[1] *Department of Mathematics and Statistics, McGill University, Montreal, Canada, e-mail:* hugo.latourelle-vigeant@mail.mcgill.ca; elliot.paquette@mcgill.ca

Abstract: This paper develops some theory of the matrix Dyson equation (MDE) for correlated linearizations and uses it to solve a problem on asymptotic deterministic equivalent for the test error in random features regression. The theory developed for the correlated MDE includes existence-uniqueness, spectral support bounds, and stability properties of the MDE. This theory is new for constructing deterministic equivalents for pseudoresolvents of a class of correlated linear pencils. In the application, this theory is used to give a deterministic equivalent of the test error in random features ridge regression, in a proportional scaling regime, wherein we have conditioned on both training and test datasets.

Keywords and phrases: anisotropic global law, empirical test error, Gaussian equivalence, linearization, matrix Dyson equation, pseudo-resolvent, random features.

1. Introduction

Contemporary artificial neural networks have found widespread applications in diverse domains. A notable trend in modern neural network design is the increasing size and complexity of these models. In practical applications, there is a prevalent use of highly overparametrized models. These overparametrized models exhibit exceptional capacity, showcasing the ability to perfectly fit the training data, even in scenarios where the labels are pure noise [ZBH$^+$21] and nonetheless generalize well. Despite the remarkable practical success of neural networks, a considerable gap exists between theoretical understanding and real-world performance in machine learning. Neural networks pose unique challenges for analysis due to two key factors. Firstly, the high dimensionality of these models often leads to behaviors that defy conventional statistical knowledge. Secondly, the presence of non-linear activation functions, which are known to enhance the expressive capacity of neural networks, further complicates analysis.

In recent years, random matrix theory has been used to provide valuable insights into the behavior of neural networks and other machine learning models. A notable line of research at the confluence of random matrix theory and neural networks combines the linear pencil method with operator-valued free probability to analyze the training and test errors of simple machine learning models [MP22, ALP19, AP20a, AP20b, TAP21b, TAP21a].

Beyond the linear model, the random features model introduced in [RR07] stands out as one of the simplest models with significant expressive power. In contrast to the linear model, the random features model incorporates a non-linear activation function and can be overparametrized. Despite its simplicity, the random features model provides a mathematically tractable framework that proves instrumental in studying phenomena observed in real-life machine learning models, such as multiple descent [MM22, AP20b] and implicit generalization [Cho22, JcS+20]. Extensive studies have delved into the training and test error of ridge regression with random features in high dimensions [dSB21, GLK+20, MM22, MMM22, HMRT22, MP22, TAP21a, TAP21b, ALP19]. Notably, the non-linear random features model is connected to a simpler linear Gaussian model through a universality phenomenon [GLK+20, GLR+22, MS22, HL23].

The concept of linearization, or linear pencil method, entails representing rational functions of random matrices as blocks of inverses of larger random matrices which depend linearly on its random matrix inputs. These linearizations possess simpler correlation structures, rendering them more amenable to certain types of analysis. Beyond their application in analyzing the training and test error of neural networks, linearizations have been extensively studied in the context of random matrix theory and free probability [HMS18, FOBS06, EKN20, HT05, And13, BMS17, HMV06]. This exploration naturally leads to the study of pseudo-resolvents, or generalized resolvents. Pseudo-resolvents are inherently more challenging to study than resolvents due to the absence of a spectral parameter spanning the entire diagonal. Nonetheless, one effective approach to study pseudo-resolvent involves analyzing a fixed-point equation known as the matrix Dyson equation for linearizations [And13, EKN20, FKN23].

1.1. Main contributions

In this work, we introduce an extension of the matrix Dyson equation (MDE) framework tailored specifically for linearizations (in particular those with correlated blocks – see Section 2.2 for further discussion on how this relates to existing MDE theory). Within this framework, we derive an anisotropic global law for a broad class of pseudo-resolvents with general correlation structures. Our approach provides a systematic method to construct a deterministic equivalent as the solution of a matrix fixed-point equation. By employing a linearization trick, our methodology becomes a versatile tool for finding deterministic equivalents for rational expressions involving random matrices. More generally, we further develop the MDE framework, offering tools that we believe can be applied to solve various other problems in this domain.

We then apply our framework to derive an asymptotically exact representation for the empirical test error of random features ridge regression. Specifically, we confirm [LLC18, Conjecture 1], with the additional assumption that the norm of the training and test random features matrices has bounded fourth moments. As a consequence, we establish a general Gaussian equivalence principle for the empirical test error of random feature ridge regression.

1.2. Notation

We adhere to the following conventions throughout our work. Lowercase letters (e.g., v) represent real or complex scalars or vectors, while uppercase letters (e.g., M) denote real or complex matrices. Calligraphic letters are utilized for denoting sets (e.g., $\mathscr{M}$) or operators on the space of complex matrices (e.g., $\mathcal{S}$). The symbol $\mathbb{H}$ refers to the upper half of the complex plane. We use $\mathbb{R}_{\geq 0}$ to represent the set of non-negative real numbers and $\mathbb{R}_{>0}$ to represent the set of positive real numbers.

When considering a matrix $M \in \mathbb{C}^{\ell\times\ell}$ or an operator on the space of $\ell \times \ell$ matrices, we often employ the blockwise representation:

$$M = \begin{bmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{bmatrix}.$$

Here, $M_{1,1} \in \mathbb{C}^{n\times n}$, $M_{1,2} \in \mathbb{C}^{n\times d}$, $M_{2,1} \in \mathbb{C}^{d\times n}$, and $M_{2,2} \in \mathbb{C}^{d\times d}$.

Given a complex matrix M, we denote its real transpose by M^T and the conjugate transpose by M^*. When dealing with matrix sub-blocks, $M^*_{i,j} = (M_{i,j})^*$ denotes the conjugate transpose of the (i, j) sub-block of M.

We use $\|\cdot\|$ to represent the standard Euclidean norm when applied to a vector and the operator norm when applied to either matrices or complex matrix-valued functions. Additionally, we employ $\|\cdot\|_F$ to denote the Frobenius norm and $\|\cdot\|_*$ for the nuclear norm. The trace of a matrix is denoted by tr.

1.3. Organization

The paper is structured as follows: we present the primary result concerning the asymptotic empirical test error of random feature ridge regression in the first section. This section also includes a concise discussion of the theorem and its implications. We then give in Section 2.2 a discussion of the related work to this main theorem, both from a method and application perspective. In Section 3, we introduce a mathematical framework designed to derive anisotropic global laws for general pseudo-resolvents. The main aspects of this framework are presented in this Section 3, with the proofs and more in-depth discussions deferred to the appendices. Finally, in Section 4 we apply our framework to establish and prove the main result. Additional supporting details are in Section B.

2. Main result

Consider a supervised training problem with a labeled dataset $\mathscr{D} = \{(x_j, y_j)\}_{j=1}^{n_{\text{train}}}$ with $x_j \in \mathbb{R}^{n_0}$ and $y_j \in \mathbb{R}$ for every $j \in \{1, 2, \ldots, n_{\text{train}}\}$. For conciseness, let $X \in \mathbb{R}^{n_{\text{train}}\times n_0}$ be the matrix with jth rows corresponding to x_j^T and y be the vectors of labels. We wish to learn a relation between the inputs x_j and the outputs y_j by minimizing the ℓ_2-regularized norm-squared loss

$$\min_{w\in\mathbb{R}^d} \|y - Aw\|^2 + \delta\|w\|^2 \tag{1}$$

where $A = n^{-\frac{1}{2}}\sigma(XW) \in \mathbb{R}^{n_{\text{train}}\times d}$ for some random matrix $W \in \mathbb{R}^{n_0\times d}$, some ridge parameter $\delta \in \mathbb{R}_{>0}$ and a λ_σ-Lipschitz activation function σ. Following the setup of [LLC18], we will assume that $W = \varphi(Z)$ for some $Z \in \mathbb{R}^{n_0\times d}$ with independent standard normal entries and φ a λ_φ-Lipschitz function. The Lipschitz constants λ_σ and λ_φ should be independent of the dimension of the problem in the sense that, as $n \to \infty$ with $n_{\text{train}} \propto n_{\text{test}} \propto n_0 \propto d$, $\limsup_{n\to\infty}(\lambda_\varphi \vee \lambda_\sigma) < \infty$. In other words, we are fitting a *random features model* using ridge regression.

The minimization problem in (1) admits the closed form solution $w = A^T(AA^T + \delta I_{n_{\text{train}}})^{-1}y$ which is called the ridge estimator. Given another labeled dataset $\widehat{\mathscr{D}} = \{(\widehat{x}_j, \widehat{y}_j)\}_{j=1}^{n_{\text{test}}}$, we can compute the empirical test error, or out-of-sample error, using the squared norm of the residuals

$$E_{\text{test}} := \|\widehat{y} - \widehat{A}w\|^2 = \|\widehat{y} - \widehat{A}A^T(AA^T + \delta I_{n_{\text{train}}})^{-1}y\|^2 \tag{2}$$

with $\widehat{A} = n^{-\frac{1}{2}}\sigma(\widehat{X}W) \in \mathbb{R}^{n_{\text{test}}\times d}$. We ask that $\limsup_{n\to\infty} \max\{\|X\|, \|\widehat{X}\|\} < \infty$ and similarly for the label vectors y and $\widehat{y}$.

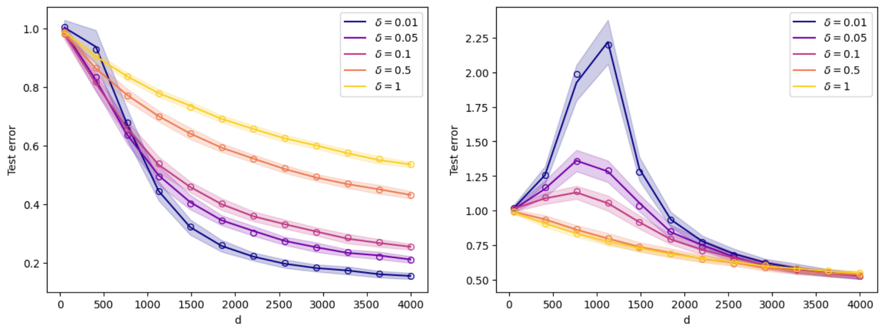

FIG 1. *E_{test} vs the deterministic approximation for various odd activation functions with different sizes of hidden layers d and ridge parameter δ. Left: Error function activation ($\sigma(x) = \operatorname{erf}(x)$); Right: Sign activation ($\sigma(x) = \operatorname{sign}(x)$).*

As it is often the case in random matrix theory, we expect that a law of large number will take effect and that the test error E_{test} will concentrate around a deterministic quantity depending on the first and second moments of A and $\widehat{A}$. Consequently, we will assume that $\{(a_j^T, \hat{a}_j^T)^T\}_{j=1}^d$, representing the columns of A and $\widehat{A}$, are i.i.d. random vectors with

$$\mathbb{E}[(a_1^T, \hat{a}_1^T)^T] = 0 \quad \text{and} \quad \mathbb{E}[(a_1^T, \hat{a}_1^T)^T(a_1^T, \hat{a}_1^T)] = \begin{bmatrix} K_{AA^T} & K_{A\widehat{A}^T} \\ K_{\widehat{A}A^T} & K_{\widehat{A}\widehat{A}^T} \end{bmatrix},$$

where we emphasize the expectation is only with respect to the randomness in W. Here, K_{AA^T}, $K_{A\widehat{A}^T}$, $K_{\widehat{A}A^T}$, and $K_{\widehat{A}\widehat{A}^T}$ encode the covariance between the entries of A and $\widehat{A}$. We obtain the following result which verifies [LLC18, Conjecture 1] under an additional boundedness assumption.

Theorem 2.1. *Suppose that $n_{\text{train}}, d, n_{\text{test}}, n_0 \propto n$ and $\limsup_{n\to\infty} \max\{\mathbb{E}[\|A\|^4], \mathbb{E}[\|\widehat{A}\|^4]\} < \infty$. Let α be the unique non-positive real number satisfying*

$$\alpha = -(1 + \operatorname{tr}(K_{AA^T}(\delta I_{n_{train}} - d\alpha K_{AA^T})^{-1}))^{-1} \in \mathbb{R}_{\leq 0}$$

and denote $M = (\delta I_{n_{\text{train}}} - d\alpha K_{AA^T})^{-1}$ as well as

$$\beta = \frac{\alpha^2 \operatorname{tr}(K_{\widehat{A}\widehat{A}^T} + d\alpha K_{\widehat{A}A^T} M(I_{n_{\text{train}}} + \delta M)K_{A\widehat{A}^T})}{1 - \|\sqrt{d}\alpha K_{AA^T}^{\frac{1}{2}} M K_{AA^T}^{\frac{1}{2}}\|_F^2} \in \mathbb{R}_{\geq 0}.$$

Then, $d\beta\|K_{AA^T}^{\frac{1}{2}} My\|^2 + \|d\alpha K_{\widehat{A}A^T} My + \widehat{y}\|^2 - E_{\text{test}} \xrightarrow[n\to\infty]{a.s.} 0.$

We prove Theorem 2.1 in Section 4.

2.1. Discussion

Let us briefly discuss some aspects regarding Theorem 2.1.

2.1.1. Behaviour of denominator

While not obvious at first, we show in Lemma 4.6 that $1 - \|\sqrt{d}\alpha K_{AA^T}^{1/2} M K_{AA^T}^{1/2}\|_F^2$ is positive and bounded away from 0 as $n \to \infty$ in the setting of Theorem 2.1. This implies that β, and therefore E_{test}, is well-behaved in the proportional limit.

2.1.2. Numerical result

Concerning numerical simulations, Theorem 2.1 indicates that to compute an asymptotic deterministic approximation for E_{test}, it suffices to solve a single scalar fixed-point equation. In particular, we demonstrate in Lemma 4.8 that the iterates $\{\alpha_k\}_{k\in\mathbb{N}_0}$ obtained by iterating

$$\alpha_{k+1} = -\left(1 + \operatorname{tr}\left(K_{AA^T}(\delta I_{n_{\text{train}}} - \alpha_k d K_{AA^T})^{-1}\right)\right)^{-1}$$

for every $k \in \mathbb{N}$ with arbitrary $\alpha_0 \in \mathbb{R}_{\leq 0}$ converge to α as $k \to \infty$. Moreover, when φ is the identity, the kernel matrices K_{AA^T}, $K_{A\widehat{A}^T}$, and $K_{\widehat{A}\widehat{A}^T}$ can be efficiently computed using [LLC18, Table 1]. We employ these techniques to generate Figure 1. The corresponding code is available online.

2.1.3. Relation to kernel regression

The term $\|d\alpha K_{\widehat{A}A^T} My + \widehat{y}\|^2 = \|\widehat{y} - dK_{\widehat{A}A^T}(dK_{AA^T} + (-\delta/\alpha)I_{n_{\text{train}}})^{-1}y\|^2$ in the asymptotic expression for the test error corresponds precisely to the norm squared test error of kernel ridge regression with ridge parameter $-\delta/\alpha \in \mathbb{R}_{>0}$. This is reminiscent of [JcS+20] and the generalization in [Cho22], and is related to the implicit regularization of the random features model. In fact, the proof of Theorem 2.1 recovers both of those results.

2.1.4. Boundedness assumption

The conditions $\limsup_{n\to\infty} \mathbb{E}[\|A\|^4] < \infty$ and $\limsup_{n\to\infty} \mathbb{E}[\|\widehat{A}\|^4] < \infty$ are satisfied in scenarios where the data matrices exhibit a form of approximate orthogonality, as discussed in [FW20], or when the data matrix consists of concentrated random vectors, as described in [LCM21, Assumption 2]. Importantly, these assumptions encompass cases where the data matrices are independent and comprised of i.i.d. standard normal entries, a setting that has been widely studied.

We also note that extensions of Theorem 2.1 to multiple layers (which we do not pursue), are reduced to establishing the norm-control on the kernels AA^T and $\widehat{A}\widehat{A}^T$.

2.2. Related work

In this section, we contextualize our results within the existing literature.

2.2.1. Random features

The random features model of [RR07] has garnered significant attention in scientific research and has proven to be a successful benchmark for studying the behavior of more intricate machine learning models. This model can be viewed as a two-layer neural network, where the first layer is frozen at random initialization, and only the second layer is trained.

The random features model has been extensively studied in various settings. The version with linear activation functions has garnered significant attention [WX20, MG21]. With non-linear activation functions, previous research has delved into the test error of random features trained using kernel regression under a student-teacher model, with data samples drawn from an isotropic distribution [AP20b]. Similarly, the test error of random features trained with ridge regression has been investigated under similar settings, employing both non-rigorous replica methods [GLK+20] and rigorous analyses [MM22, MMM22, ALP19]. These investigations have been extended to scenarios involving anisotropic data [HMRT22, MP22] and covariate shift [TAP21a, TAP21b]. Our study diverges from these works by focusing on the empirical test error of random features ridge regression, without assuming specific data models or distributions beyond some boundedness conditions.

Our main result regarding the test error of random features ridge regression shares similarities with the work of [LLC18], who established an asymptotically exact expression for the training error of random features ridge regression [LLC18]. The authors conjectured that our main result (stated in Theorem 2.1) holds without the additional conditions imposing bounded fourth moments for the norm of the random features matrices. The important paper [LCM21] resolves this conjecture in the special case of random Fourier features. Both [LCM21, LLC18] employ leave-one-out techniques and concentration of measure arguments in their approaches. Although we also utilize a leave-one-out argument to establish universality, our overall approach differs fundamentally, providing flexibility for addressing more complex scenarios where leave-one-out approaches may not be as straightforward to apply.

2.2.2. Conjugate kernel

The conjugate kernel, a concept inherently tied to the random features model, has been a subject of study in the realm of random matrix theory. Previous works, such as [PW17, BP21], derive a deterministic equivalent for the random feature model when dealing with isotropic data and weight matrices. An extension to nearly orthogonal data is given in [WZ23, FW20]. A similar outcome is established by [Cho22, LLC18], utilizing concentration of measure and leave-one-out methods, and in [PS21] using resolvent methods. Beyond bulk laws, [BP22] explore outlier eigenvalue of the conjugate kernel. Notably, the proof of our main result provides a deterministic equivalent for the conjugate kernel, akin to the findings of [Cho22, LLC18], employing a significantly different approach.

2.2.3. Gaussian equivalence

Theorem 2.1 establishes a Gaussian equivalence principle, indicating that every random features model trained with ridge regression, as described in the statement of Theorem 2.1, performs equivalently to a surrogate Gaussian model with a matching covariance structure. However, it is important to note, as mentioned in [LLC18], that the distribution of the input data can impact the performance of the random features model. This influence stems from the fact that, although there is Gaussian equivalence at the level of random feature matrices, the distribution of the input may influence the covariance matrices K_{AA^T}, $K_{A\widehat{A}^T}$, $K_{\widehat{A}A^T}$, and $K_{\widehat{A}\widehat{A}^T}$, which are directly linked to the performance of the random features model.

Some of the works on the random features model rely upon or prove a Gaussian equivalence principle, for instance [DL20]. This Gaussian equivalence postulate that every random feature model is equivalent to a surrogate Gaussian model with matching covariance. Gaussian equivalence theorems have been established in [GLK+20] using a non-rigorous replica method, rigorously in [GLR+22, HMRT22] for random features ridge regression. Then, [HL23] shows that a Gaussian equivalence principle holds more generally for random features under student teacher model with more general loss and regularization functions. Gaussian equivalence has also been established for deep random features [SCDL23]. Our Gaussian equivalence principle asserts that every random features model trained with ridge regression is equivalent—meaning it exhibits the same training and generalization error—to a Gaussian model with a matching covariance. This can be viewed as an extension of previous work to the empirical test error setting.

2.2.4. Matrix Dyson equation

The matrix Dyson equation, a self-consistent matrix equation, has proven to be a valuable tool for deriving local laws in various contexts. For an excellent introduction to this subject, we recommend [Erd19]. The vector version of the matrix Dyson equation has been employed to establish local laws for Generalized Wigner matrices [AEK19a, AKE17, AEK17]. Extending its applicability, the matrix Dyson equation has been used to investigate local laws for Hermitian matrices with correlations featuring fast decay, as well as those with slower correlation decay, particularly focusing on regular edges [AEKS20] and regions away from the support edges [EKS19]. A notable advantage of the matrix Dyson equation lies in the fact that its solution admits a Stieltjes representation. Leveraging this equation, detailed regularity properties of the self-consistent density of states have been explored [AEK18], and bounds on the spectrum of Kronecker random matrices have been established [AEKN19]. We explore a generalized version of the matrix Dyson equation, providing a framework for deriving deterministic equivalents for pseudo-resolvents.

2.2.5. Linearization

The concept of linearization gained prominence following the groundbreaking work of [HT05]. This work essentially demonstrated that to analyze a polynomial expression in matrices, it is sufficient to consider a linear polynomial with matrix coefficients [HT05]. A limitation of [HT05] is that their linearization trick does not preserve the self-adjointness of the polynomial expression. This issue was addressed by [And13], who proved that when the polynomial expression is self-adjoint, it is possible to choose the coefficients of the linearization in a way that maintains self-adjointness [And13]. This result was later extended to rational expressions [HMS18]. Notably, linearizations are highly non-unique, leading to the emergence of various other linearization techniques. For instance, [EKN20] introduced a minimal linearization. Importantly, the arguments supporting these linearization techniques are often constructive, enabling the explicit construction of suitable linearizations. The linearization trick has found application in free probability, allowing the study of polynomials of random matrices on both the global scale [And13, BMS17, HT05, HMS18, HMV06] and the local scale [EKN20, FKN23, And15]. The combination of the linearization trick, referred to as the pencil method in this context, along with operator-valued free probability has found successful applications in the study of simple neural networks [MP22, ALP19, AP20a, AP20b, TAP21b, TAP21a]. In this work, we develop a framework based on an extension of the matrix Dyson equation to study asymptotic properties of linearizations with a general correlation structure. This provides an alternative to the use of operator-valued free probability, and we believe that our framework could find multiple applications in machine learning.

2.2.6. Linearized matrix Dyson equation

The concept of linearization naturally gives rise to the study of pseudo-resolvents. Pseudo-resolvents are inherently more challenging to analyze than regular resolvents because there is no spectral parameter spanning the entire diagonal. Nonetheless, the matrix Dyson equation has been extended to study pseudo-resolvents. [And13] derived global laws for linearizations of polynomials in independent Wigner matrices using a matrix Dyson equation for linearization, which the author called the *Schwinger–Dyson equation* [And13]. This work was extended in [And15] to study the anticommutator on a local scale and more generally to study polynomials of independent matrices with independent centered entries and suitable normalization [EKN20, FKN23] under the name *Dyson equation for linearization (DEL)*.

Our approach differs significantly. While previous research has focused on linearizations with blocks of independent generalized Wigner matrices, we consider linearizations with arbitrary correlation structures. We are interested in studying pseudo-resolvents on a global scale, which, although less precise than the local scale, allows us to relax the assumptions of previous work. Additionally, global laws are sufficient to make assertions about machine learning models in many cases. Our approach also provides a novel perspective on studying the matrix Dyson equation. Notably, we analyze the Carathedory-Riffen-Finsler (CRF) pseudo-metric [Har03, Har79] to demonstrate that the matrix Dyson equation for linearizations is asymptotically stable under general assumptions.

3. Framework

We defer most of the proofs and discussion for this section to Section A. We focus on a class of real[1] self-adjoint linearizations denoted as

$$L = \begin{bmatrix} A & B^T \\ B & Q \end{bmatrix} \in \mathbb{R}^{\ell\times\ell}, \tag{3}$$

where $A \in \mathbb{R}^{n\times n}$ is a potentially random self-adjoint complex matrix, $Q \in \mathbb{R}^{d\times d}$ is a deterministic invertible self-adjoint matrix and $B \in \mathbb{R}^{d\times n}$ is a potentially random arbitrary matrix. Our primary interest lies in analyzing the behavior in high dimensions of the *pseudo-resolvent* $(L - z\Lambda)^{-1}$, where $\Lambda := \mathrm{BlockDiag}\{I_{n\times n}, 0_{d\times d}\}$ and $z \in \mathbb{H} := \{z \in \mathbb{C} : \Im[z] > 0\}$ represents the upper half of the complex plane.

Our framework relies on the *linearized matrix Dyson equation (MDE)*

$$(\mathbb{E}L - \mathcal{S}(M) - z\Lambda)\, M = I_\ell, \tag{4}$$

where the spectral parameter z is chosen from the upper half complex plane $\mathbb{H}$. Here, the *super-operator*

$$\mathcal{S} : M \in \mathbb{C}^{\ell\times\ell} \mapsto \mathbb{E}\begin{bmatrix} [(L-\mathbb{E}L)M(L-\mathbb{E}L)]_{1,1} & (A-\mathbb{E}A)M_{1,1}(B^T - \mathbb{E}B^T) \\ (B-\mathbb{E}B)M_{1,1}(A-\mathbb{E}A) & (B-\mathbb{E}B)M_{1,1}(B-\mathbb{E}B)^T \end{bmatrix} \in \mathbb{C}^{\ell\times\ell}. \tag{5}$$

is a positivity-preserving linear map which encodes the second moment of L. For conciseness, we let $s \in \mathbb{R}_{>0}$ such that $\|\mathcal{S}(W)\| \leq s\|W\|$ for every $W \in \mathbb{C}^{\ell\times\ell}$, but also $\|\mathcal{S}_{i,j}(W)\| \leq s\|W_{1,1}\|$ for all $(i,j) \in \{(1,2),(2,1),(2,2)\}$ and $W \in \mathbb{C}^{\ell\times\ell}$.

[1] Our framework can easily be extended to complex linearizations $L \in \mathbb{C}^{\ell\times\ell}$ with $\Im[L] \preceq 0$. Because we do not have any application for this generalization in mind, we do not pursue this direction for the sake of clarity.

In order to ensure the existence of a unique solution to the matrix Dyson equation, we need to restrict (4) to a suitable set. Consequently, we introduce the *admissible set*[2]

$$\mathscr{M} := \{f : \mathbb{H} \mapsto \mathscr{A} \text{ analytic}\}\,, \quad \mathscr{A} := \{W \in \mathbb{C}^{\ell\times\ell} : \Im[W] \succeq 0 \text{ and } \Im[W_{1,1}] \succ 0\}. \tag{6}$$

Our primary strategy for analyzing (4) involves initially establishing analogous results for a regularized version of the equation. This regularization typically simplifies the problem, enabling us to leverage existing knowledge. Subsequently, we demonstrate the feasibility of setting the regularization parameter to zero, effectively reverting to the original equation. Importantly, we ensure that the statements derived for the regularized variant remain valid in this limit, thereby providing valuable insights into the properties of (4). For this reason, we introduce the *regularized matrix Dyson equation (RMDE)*

$$(\mathbb{E}L - \mathcal{S}(M^{(\tau)}) - z\Lambda - i\tau I_\ell)M^{(\tau)} = I_\ell \tag{7}$$

for every $\tau > 0$. The corresponding admissible set is given by

$$\mathscr{M}_+ := \{f : \mathbb{H} \mapsto \mathscr{A}_+ \text{ analytic}\}\,, \quad \mathscr{A}_+ := \{W \in \mathbb{C}^{\ell\times\ell} : \Im[W] \succ 0\} \cap \mathscr{A}. \tag{8}$$

Our initial key result naturally revolves around establishing the existence and uniqueness of a solution for (4). In traditional matrix Dyson equation theory, wherein the spectral parameter spans the entire diagonal, the existence of a unique solution usually emerges comprehensively from [HFS07, Theorem 2.1]. Indeed, we leverage this theorem precisely to establish the existence and uniqueness of a solution to (7). However, due to the absence of a spectral parameter spanning the entire diagonal in our case, demonstrating the existence of a solution to (4) is not trivial and requires careful analysis. Nonetheless, by leveraging the suitable properties of the admissible set and the surrogate regularized matrix Dyson equation, we obtain the following existence and uniqueness result.

Theorem 3.1 (Existence and Uniqueness)**.** *There exists a unique analytic matrix-valued function $M \in \mathscr{M}$ such that $M(z)$ solves the MDE (4) for every $z \in \mathbb{H}$. Additionally, $\|M_{1,1}(z)\| \leq (\Im[z])^{-1}$ and*

$$M(z) = \begin{bmatrix} 0_{n\times n} & 0_{n\times d} \\ 0_{d\times n} & Q^{-1} \end{bmatrix} + \int_{\mathbb{R}} \frac{\Omega(\mathrm{d}\lambda)}{\lambda - z}$$

for all $z \in \mathbb{H}$, where Ω is a real Borel $\ell \times \ell$ positive semidefinite compactly supported [3] *measure satisfying*

$$\int_{\mathbb{R}} \Omega(\mathrm{d}\lambda) = \begin{bmatrix} I_n & -\mathbb{E}[B^T]Q^{-1} \\ -Q^{-1}\mathbb{E}[B] & Q^{-1}\mathbb{E}[BB^T]Q^{-1}. \end{bmatrix}.$$

For the rest this paper, we will utilize the notation M to represent the unique solution as ensured by Theorem 3.1, $M^{(\tau)}$ to represent the unique solution to the regularized linearized MDE (7), and we will omit the explicit mention of z when the context confines it to a fixed $z \in \mathbb{H}$.

---

[2]Alternatively, assuming that $A = \mathbb{E}A$ and $\mathbb{E}B = 0$, we can choose $\mathcal{S}(M) = \mathbb{E}[(L - \mathbb{E}L)M(L - \mathbb{E}L)]$ and consider the set

$$\mathscr{A} := \{W \in \mathbb{C}^{\ell\times\ell} : \Im[W] \succeq 0 \text{ and } \Im[W_{1,1}] \succ 0,\, W_{1,2} = W_{2,1} = 0\}.$$

Applying our framework under these assumptions requires no modifications, making it particularly suitable, for example, for the study of powers of Wigner matrices.

[3]The bound on the support is effective, and is given explicitly in Lemma A.5.

In the proof of Theorem 3.1, we define $M_{1,1}$ as the limit point of the normal family $\{M_{1,1}^{(\tau)} : \tau > 0\}$ as $\tau \to 0$. Decomposing (4) block-wise, we then observe that $M^{(\tau)}(z)$ converges to $M(z)$ in spectral norm for any fixed $z \in \mathbb{H}$ as τ approaches the origin from above. However, it will be considerably beneficial to quantify the extent to which $\|M(z) - M^{(\tau)}(z)\|$ varies with respect to $\tau \in \mathbb{R}_{>0}$ and $\ell \in \mathbb{N}$. Interestingly, when considering the analogous measure with respect to the pseudo-resolvent, it becomes apparent that $\|(L - z\Lambda)^{-1} - (L - z\Lambda - i\tau I_\ell)^{-1}\| \leq \tau \|(L - z\Lambda)^{-1}\|^2$, with the expectation that $\|(L - z\Lambda)^{-1}\|^2$ remains bounded as the dimension of the problem increases.

Assumption 1. For every $z \in \mathbb{H}$, there exists a function f and subsequence $\{\tau_k\} \subseteq \mathbb{R}_{>0}$ such that $\tau_k \to 0$, $f(\tau_k) \to 0$ and $\|M^{(\tau_k)}(z) - M(z)\| \leq f(\tau_k) + o_\ell(1)$ for all $k \in \mathbb{N}$ and every $\ell \in \mathbb{N}$ large enough.

It is noteworthy that Assumption 1 is fulfilled within the frameworks based on the matrix Dyson equation for linearization as detailed in [EKN20, And13, FKN23]. This satisfaction is explicitly indicated by [EKN20, Equation 4.11], [And13, Estimates 6.3.3.], and [FKN23, Equation A.25]. In general, the validity of Assumption 1 in these cases stems from the ability to construct a dimension-independent representation of the solution to the (R)MDE using tools from free probability. As asserted by [HT05, Lemma 5.4], such a representation exists whenever L takes the form $L = A_0 \otimes I_n + \sum_{j=1}^k A_i \otimes X_j$, where $\{A_j\}_{j=0}^k$ forms a collection of complex $d \times d$ self-adjoint matrices, and $\{X_j\}_{j=1}^k$ forms a collection of independent random matrices with $\{(X_j)_{a,a}\}_{a=1}^n \cup \{(\sqrt{2}\Re X_j)_{a,b}\}_{a<b} \cup \{(\sqrt{2}\Im X_j)_{a,b}\}_{a<b}$ being a collection of n^2 i.i.d. centered Gaussian random variables for every $j \in \{1, 2, \ldots, k\}$.

Furthermore, Assumption 1 is related to the *stability operator*. Following the notation in [AEKN19], the stability operator is defined as $\mathcal{L} : W \in \mathbb{C}^{\ell \times \ell} \mapsto W - M\mathcal{S}(W)M$. The concept of the stability operator is inherently connected to the analysis of the matrix Dyson equation [AEKN19, Erd19, AEK19b, FKN23]. The term stability operator is aptly chosen because, when it is both invertible and its inverse is bounded, it provides a means to establish the stability of the matrix Dyson equation through techniques like an implicit function theorem such as the one in [AEK19b, Lemma 4.10] as demonstrated in the work of [Erd19, EKN20]. The stability operator organically appears in the uniqueness argument, where its invertibility at infinity allows us to uniquely and recursively determine the power series expansion of the solution. The connection between the stability operator and Assumption 1 becomes apparent when we consider the derivative of $M^{(\tau)}(z)$ with respect to $i\tau$, which yields $\mathcal{L}(\partial_{i\tau} M(z)) = (M(z))^2$. Because $M(z)$ is bounded in operator norm, we can conclude that Assumption 1 is implied by the requirement of having an invertible stability operator with a bounded inverse.

We proceed to prove the asymptotic stability of the MDE. To this end, let $F(z) = \mathbb{E}(L - z\Lambda)^{-1} \in \mathscr{M}$ be the expected pseudo-resolvent. It is inconvenient to work directly with the expected pseudo-resolvent, and we will systematically prefer working with a regularized version of the same object. For each $\tau \in \mathbb{R}_{>0}$, we consider the expected regularized pseudo-resolvent $F^{(\tau)}(z) = \mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1} \in \mathscr{M}_+$ which satisfies

$$(\mathbb{E}L - \mathcal{S}(F^{(\tau)}(z)) - z\Lambda - i\tau I_\ell)F^{(\tau)}(z) = I_\ell + D^{(\tau)}, \tag{9}$$

where $D^{(\tau)}$ is a regularized perturbation term explicitly given by

$$D^{(\tau)} = \mathbb{E}\left[\left(\mathbb{E}L - L - \mathcal{S}(\mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1})\right)(L - z\Lambda - i\tau I_\ell)^{-1}\right]. \tag{10}$$

Essentially, we consider $F^{(\tau)}$ as a function that almost satisfies the MDE, up to an additive perturbation term $D^{(\tau)}$. By stability, we mean the property of the MDE that implies $F(z)$ is close — pointwise in spectral norm — to the solution $M(z)$ of (4) for every $z \in \mathbb{H}$ whenever the perturbation $D^{(\tau)}$ and the regularization parameter τ are small.

Since our primary objective is to investigate the behavior in the high-dimensional limit, it is essential for the super-operator $\mathcal{S}$, among other objects, to remain bounded as the problem dimension increases. We make the following assumption.

Assumption 2. Suppose that there exists $s \in \mathbb{R}_{>0}$ such that $\|\mathcal{S}(W)\| \leq s\|W\|$ for every $W \in \mathbb{C}^{\ell \times \ell}$ and $\limsup_{\ell\to\infty} s < \infty$. Furthermore, assume that $\limsup_{\ell\to\infty} \|\mathbb{E}L\| < \infty$ and $\limsup_{\ell\to\infty} \mathbb{E}\|(L - z\Lambda)^{-1}\|^2 < \infty$.

Using the fact that the mapping characterizing (7) is contractive with respect to the *Carathéodory-Riffen-Finsler pseudometric*, we establish the stability of the regularized MDE for every regularization parameter $\tau \in \mathbb{R}_{>0}$. Through a careful analysis, we further reveal that it is feasible to decrease the regularization parameter to zero at a specific rate while maintaining stability. This rate is contingent upon the rate at which the perturbation matrix vanishes. The resulting outcome is the following asymptotic stability of (4).

Theorem 3.2 (Asymptotic stability). *Suppose that* $\|D^{(\tau)}\| \xrightarrow{\ell\to\infty} 0$ *for every* $\tau \in \mathbb{R}_{>0}$. *Then, under Assumptions 1 and 2,* $\|M(z) - \mathbb{E}(L - z\Lambda)^{-1}\| \xrightarrow{\ell\to\infty} 0$ *for every* $z \in \mathbb{H}$.

Now that we have existence of a unique solution M to (4) as well as an asymptotic stability property, we want to show that $M(z)$ serves as a favorable asymptotic approximation for the pseudo-resolvent $(L - z\Lambda)^{-1}$. In view of Theorem 3.2, the focus shifts to proving that the perturbation matrix vanishes in norm as the problem dimension grows for every regularization parameter. There are various methods to establish this, depending on the assumptions about the linearization L. To apply our framework and prove Theorem 2.1, we naturally choose a route based on Gaussian concentration inequalities. This choice confines our theoretical considerations to linearizations characterized by Gaussian-concentrated entries.

Assumption 3. Suppose that $\gamma \in \mathbb{N}$, $g \sim \mathcal{N}(0, I_\gamma)$ and that there exists a map $\mathcal{C} : \mathbb{R}^\gamma \mapsto \mathbb{R}^{\ell\times\ell}$ such that $L \equiv L(g) = \mathcal{C}(g) + \mathbb{E}L$. Furthermore, assume that $\mathcal{C}$ is symmetric in the sense that $\mathcal{C}(x) = (\mathcal{C}(x))^T$ for every $x \in \mathbb{R}^\gamma$.

Under Assumption 3, we aim to decompose the perturbation matrix $D^{(\tau)}$ into terms that are amenable to analysis. To achieve this, define

$$\begin{aligned}\Delta(L, \tau; z) = &\ \mathbb{E}[(L - \mathbb{E}L)(L - z\Lambda - i\tau I_\ell)^{-1}] \\ &+ \mathbb{E}[(\tilde{L} - \mathbb{E}L)(L - z\Lambda - i\tau I_\ell)^{-1}(\tilde{L} - \mathbb{E}L)(L - z\Lambda - i\tau I_\ell)^{-1}] \end{aligned} \tag{11}$$

where $\tilde{L}$ is an i.i.d. copy of L,

$$\tilde{\mathcal{S}} : M \in \mathbb{C}^{\ell\times\ell} \mapsto \mathbb{E}\left[(L - \mathbb{E}L)M(L - \mathbb{E}L)\right] - \mathcal{S}(M) \in \mathbb{C}^{\ell\times\ell},^{4} \tag{12}$$

and consider the decomposition

$$D^{(\tau)} = \mathbb{E}\left[\mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1})(L - z\Lambda - i\tau I_\ell)^{-1}\right] - \mathcal{S}(F^{(\tau)})F^{(\tau)} \tag{13a}$$

$$+ \mathbb{E}\left[\tilde{\mathcal{S}}((L - z\Lambda - i\tau I_\ell)^{-1})(L - z\Lambda - i\tau I_\ell)^{-1}\right] \tag{13b}$$

$$- \Delta(L, \tau). \tag{13c}$$

[4] We may also remove any term in the upper-left block of $\mathbb{E}\left[(L - \mathbb{E}L)M(L - \mathbb{E}L)\right]$ from $\mathcal{S}$ and add them to $\tilde{\mathcal{S}}$ without changing any of our arguments.

The first perturbation term in (13a) arises from the use of the expected pseudo-resolvent in Theorem 3.2. To ensure that this perturbation term is asymptotically small, we require the super-operator $\mathcal{S}$ to be *averaging*. This implies that $\mathcal{S}((L-z\Lambda-i\tau I_\ell)^{-1})$ should exhibit a "law of large numbers" behavior and converge to a deterministic limit. While working directly with the pseudo-resolvent would eliminate this specific perturbation term from the expectation of $D^{(\tau)}$, such an approach would have its disadvantages. Utilizing the expected pseudo-resolvent, on the other hand, allows us to work with deterministic objects and leverage norm bounds. We derive a condition for $\mathcal{S}((L-z\Lambda-i\tau I_\ell)^{-1})$ to concentrate around its mean based on Gaussian concentration.

The second perturbation term, as expressed in (13b), arises from our specific definition of the super-operator and would not be present if we defined the super-operator as $\mathbb{E}[(L-\mathbb{E}L)M(L-\mathbb{E}L)]$. However, our chosen definition of the super-operator, coupled with the assumption $Q = \mathbb{E}Q$, ensures that the MDE can be determined by the upper-left $n \times n$ block. This distinction allows us to establish the existence of a solution to (4). Consequently, we view $\tilde{\mathcal{S}}$ as a correction term that be vanishing in ℓ.

Finally, (13c) posits that the matrix L should approximate a Gaussian distribution in the sense that it should asymptotically satisfy a matrix Stein lemma with a vanishing error. The quantity $\|\Delta(L,\tau)\|$ serves informally as a metric characterizing the distance between L and a matrix with Gaussian entries. Notably, the subsequent result demonstrate that $\Delta(L,\tau) = 0$ holds whenever L has Gaussian entries.

Lemma 3.1. *If $\tau \in \mathbb{R}_{>0}$, $z \in \mathbb{H}$ and Assumption 3 holds with a linear map $\mathcal{C}$, then $\Delta(L,\tau;z) = 0$.*

Proof. Let $j,k \in \{1,2,\dots,\ell\}$ be arbitrary. Consider $\mathcal{C}$ as a $\ell \times \ell \times \gamma$ tensor such that $[\mathcal{C}(g)]_{j,k} = \mathcal{C}_{j,k,\alpha} g_\alpha$. Here, we use Einstein's notation which means that we sum over every subscript appearing at least two times in a given expression. By Stein's lemma [Ste81, Lemma 1],

$$\begin{aligned}\mathbb{E}\left[(L-\mathbb{E}L)(L-z\Lambda-i\tau I_\ell)^{-1}\right]_{j,k} &= \mathbb{E}\left[\mathcal{C}_{j,m,\alpha} g_\alpha (L-z\Lambda-i\tau I_\ell)^{-1}_{m,k}\right] \\ &= \mathbb{E}\left[\mathcal{C}_{j,m,\alpha}\frac{\partial (L-z\Lambda-i\tau I_\ell)^{-1}_{m,k}}{\partial g_\alpha}\right]\end{aligned}$$

Let $e_\alpha \in \mathbb{R}^\gamma$ be the α-th canonical basis vector, $\delta \in \mathbb{R}_{>0}$ and $L_\delta = \mathcal{C}(g+\delta e_\alpha) + \mathbb{E}L$. Then,

$$\begin{aligned}(L_\delta - z\Lambda - i\tau I_\ell)^{-1}_{m,k} - (L-z\Lambda-i\tau I_\ell)^{-1}_{m,k} &= \left[(L_\delta - z\Lambda - i\tau I_\ell)^{-1}(L-L_\delta)(L-z\Lambda-i\tau I_\ell)^{-1}\right]_{m,k} \\ &= -\delta\left[(L_\delta - z\Lambda - i\tau I_\ell)^{-1}\mathcal{C}(e_\alpha)(L-z\Lambda-i\tau I_\ell)^{-1}\right]_{m,k}.\end{aligned}$$

Taking the limit of the quotient of this difference with δ as δ approaches 0, we get that

$$\frac{\partial (L-z\Lambda-i\tau I_\ell)^{-1}_{m,k}}{\partial g_\alpha} = -\left[(L-z\Lambda-i\tau I_\ell)^{-1}\mathcal{C}(e_\alpha)(L-z\Lambda-i\tau I_\ell)^{-1}\right]_{m,k}$$

and, consequently,

$$\mathbb{E}\left[(L-\mathbb{E}L)(L-z\Lambda-i\tau I_\ell)^{-1}\right]_{j,k} = -\mathbb{E}\left[\mathcal{C}_{j,m,\alpha}(L-z\Lambda-i\tau I_\ell)^{-1}_{m,a}\mathcal{C}_{a,b,\alpha}(L-z\Lambda-i\tau I_\ell)^{-1}_{b,k}\right].$$

Note that

$$\mathbb{E}\left[(L-\mathbb{E}L)^T N (L-\mathbb{E}L)\right]_{j,k} = \mathbb{E}\left[\mathcal{C}_{j,a,\alpha} g_\alpha N_{a,b} \mathcal{C}_{b,k,\beta} g_\beta\right] = \mathbb{E}\left[\mathcal{C}_{j,a,\alpha} N_{a,b}\mathcal{C}_{b,k,\alpha}\right]$$

for every $N \in \mathbb{R}^{\ell\times\ell}$ independent of L. The result follows. ☐

Bounding each term on the right-hand side of (13) individually, we may now control the perturbation matrix (10) in norm for every $\tau \in \mathbb{R}_{>0}$.

Theorem 3.3. *Let $\tau \in \mathbb{R}_{>0}$, $z \in \mathbb{H}$ and $D^{(\tau)}$ be the perturbation matrix in (10). Under Assumption 3, assume that the mapping $g \in (\mathbb{R}^{\gamma}, \|\cdot\|_2) \mapsto \mathcal{S}((L(g) - z\Lambda - i\tau I_\ell)^{-1}) \in (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_2)$ is λ-Lipschitz with respect to the operator norm. Then, there exists an absolute constant $c \in \mathbb{R}_{>0}$ such that*

$$\|D^{(\tau)}\| \leq c\tau^{-1}\sqrt{\ell}\lambda + \tau^{-2}\|\tilde{\mathcal{S}}\| + \|\Delta(L,\tau)\|.$$

As a direct outcome of Theorem 3.3, it follows that $\|D^{(\tau)}\|$ tends towards zero as the dimension ℓ approaches infinity under the conditions $\lim_{\ell\to\infty}\sqrt{\ell}\lambda = 0$, $\lim_{\ell\to\infty}\|\tilde{\mathcal{S}}\| = 0$ and $\lim_{\ell\to\infty}\|\Delta(L,\tau)\| = 0$ for every $\tau \in \mathbb{R}_{>0}$ small enough. In the proof of Theorem 2.1, we will upper bound the Lipschitz constant λ in Theorem 3.3 by $\lambda \leq \tau^{-2}\|\mathcal{S}\|_{F\to 2}\lambda_{\mathcal{C}}$ where $\|\mathcal{S}\|_{F\to 2}$ denote the operator norm of the map $\mathcal{S} : (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_F) \mapsto (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_2)$ and $\lambda_{\mathcal{C}}$ is the Lipschitz constant associated with the map $\mathcal{C} : (\mathbb{R}^{\gamma}, \|\cdot\|_2) \mapsto (\mathbb{R}^{\ell\times\ell}, \|\cdot\|_F)$. Then, $\lim_{\ell\to\infty}\sqrt{\ell}\lambda = 0$ will follow from $\|\mathcal{S}\|_{\|\cdot\|_F\to\|\cdot\|_2} \lesssim \ell^{-\frac{1}{2}}$ and $\lambda_{\mathcal{C}} \lesssim \ell^{-\frac{1}{2}}$.

By Lemma 3.1, it is trivial to control $\|\Delta(L,\tau)\|$ when L has Gaussian entries. Alternatively, an interpolation approach based on cumulant bounds in the spirit of [LP09, Proposition 3.1] appears to be a suitable avenue to extend Stein's lemma, and therefore Lemma 3.1 to more a larger class of distribution. In the proof of Theorem 2.1, we employ a leave-one-out strategy to demonstrate that $\|\Delta(L,\tau)\|$ is vanishing in ℓ for every $\tau \in \mathbb{R}_{>0}$.

The culmination of Theorem 3.2 and Theorem 3.3 along with these specified conditions signifies that $M(z)$ becomes a deterministic equivalent for the expected pseudo-resolvent $\mathbb{E}(L - z\Lambda)^{-1}$ across all $z \in \mathbb{H}$.

Corollary 3.1. *Let $z \in \mathbb{H}$ and λ be defined as in Theorem 3.3. Under Assumptions 1 to 3, suppose that $\lim_{\ell\to\infty}\sqrt{\ell}\lambda = \lim_{\ell\to\infty}\|\tilde{\mathcal{S}}\| = \lim_{\ell\to\infty}\|\Delta(L,\tau)\| = 0$ for every $\tau \in \mathbb{R}_{>0}$ small enough. Then, $\|\mathbb{E}(L - z\Lambda)^{-1} - M(z)\| \xrightarrow{\ell\to\infty} 0$.*

In certain scenarios, it is feasible to alleviate the reliance of Corollary 3.1 on Assumption 3. Notably, by employing a universality result such as the one presented in [BvH23, Lemma 6.11], one may directly argue that certain functionals of resolvent of random matrices do not depend on the distribution of the input.

The only remaining task is to establish that the expected pseudo-resolvent is itself a deterministic equivalent for the true pseudo-resolvent — a widely acknowledged fact that stems from a variety of methodologies. We present one such result, based on the assumptions used above.

Lemma 3.2. *Let $U \in \mathbb{C}^{\ell\times\ell}$ with $\|U\|_F \leq 1$ and assume that the map $g \in (\mathbb{R}^{\gamma}, \|\cdot\|_2) \mapsto (L(g) - z\Lambda)^{-1} \in (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_F)$ is λ-Lipschitz with $\lambda \asymp \ell^{-r}$ for some $r > 0$. Under Assumption 3, $\operatorname{tr}(U((L - z\Lambda)^{-1} - \mathbb{E}(L - z\Lambda)^{-1})) \xrightarrow[\ell\to\infty]{a.s.} 0$.*

Combining Corollary 3.1 and Lemma 3.2 through the utilization of Von Neumann's trace inequality, we derive the ensuing anisotropic law, presented here for the sake of comprehensiveness.

Corollary 3.2. *Under the settings of Corollary 3.1 and lemma 3.2, $\operatorname{tr}(U((L - z\Lambda)^{-1} - M(z))) \xrightarrow[\ell\to\infty]{a.s.} 0$ for every $U \in \mathbb{C}^{\ell\times\ell}$ with $\|U\|_* \leq 1$.*

4. Proof of main theorem

In this section, we provide some details concerning the proof of Theorem 2.1 using our framework. First, note that we may expand the test error in (2) as

$$E_{\text{test}} = \|\widehat{A}A^T \left(AA^T + \delta I_{n_{\text{train}}}\right)^{-1} y\|^2 - 2\widetilde{y}^T \widehat{A}A^T \left(AA^T + \delta I_{n_{\text{train}}}\right)^{-1} y + \|\widetilde{y}\|^2.$$

Each term in the above equations takes the form of a bilinear form, aligning well with the framework of deterministic equivalence. In view of this, our goal will be to use the matrix Dyson equation for linearization framework to determine deterministic equivalents for the matrix $\widehat{A}A^T \left(AA^T + \delta I_{n_{\text{train}}}\right)^{-1}$ and its square.

4.1. First deterministic equivalent

We start by considering $\widehat{A}A^T \left(AA^T + \delta I_{n_{\text{train}}}\right)^{-1}$. Let $\ell = n_{\text{train}} + d + 2n_{\text{test}}$ and consider the linearization

$$L = \begin{bmatrix} \delta I_{n_{\text{train}}} & A & 0_{n_{\text{train}}\times n_{\text{test}}} & 0_{n_{\text{train}}\times n_{\text{test}}} \\ A^T & -I_{d\times d} & 0_{d\times n_{\text{test}}} & \widehat{A}^T \\ 0_{n_{\text{test}}\times n_{\text{train}}} & 0_{n_{\text{test}}\times d} & 0_{n_{\text{test}}\times n_{\text{test}}} & -I_{n_{\text{test}}} \\ 0_{n_{\text{test}}\times n_{\text{train}}} & \widehat{A} & -I_{n_{\text{test}}} & 0_{n_{\text{test}}\times n_{\text{test}}} \end{bmatrix} \in \mathbb{R}^{\ell\times\ell}. \tag{14}$$

Taking $\Lambda := \text{BlockDiag}\{I_{n_{\text{train}}+d}, 0_{2n_{\text{test}}\times 2n_{\text{test}}}\}$, we use the Schur complement formula to express the pseudo-resolvent $(L - z\Lambda)^{-1}$ blow-wise as

$$(L - z\Lambda)^{-1} = \begin{bmatrix} R & (1+z)^{-1}RA & (1+z)^{-1}RA\widehat{A}^T & 0 \\ (1+z)^{-1}A^TR & \bar{R} & \bar{R}\widehat{A}^T & 0 \\ (1+z)^{-1}\widehat{A}A^TR & \widehat{A}\bar{R} & \widehat{A}\bar{R}\widehat{A}^T & -I_{n_{\text{test}}} \\ 0 & 0 & -I_{n_{\text{test}}} & 0 \end{bmatrix}.$$

Here $R := ((1+z)^{-1}AA^T + (\delta - z)I_{n_{\text{train}}})^{-1}$ represents a resolvent and $\bar{R} := -((1+z)I_d + (\delta - z)^{-1}A^TA)^{-1}$ is a co-resolvent. Indeed, the term $\lim_{z\to 0}(L - z\Lambda)^{-1}_{3,1} = \widehat{A}A^T \left(AA^T + \delta I_{n_{\text{train}}}\right)^{-1}$ holds the relevant expression. Therefore, it suffices to find a deterministic equivalent for the pseudo-resolvent $(L - z\Lambda)^{-1}$ and take the spectral parameter to zero.

The linearization in (14) yields the super-operator

$$\mathcal{S} : M : \mathbb{C}^{\ell\times\ell} \mapsto \begin{bmatrix} \text{tr}(M_{2,2})K_{AA^T} & 0 & 0 & \text{tr}(M_{2,2})K_{A\widehat{A}^T} \\ 0 & \rho(M)I_d & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \text{tr}(M_{2,2})K_{\widehat{A}A^T} & 0 & 0 & \text{tr}(M_{2,2})K_{\widehat{A}\widehat{A}^T} \end{bmatrix} \in \mathbb{C}^{\ell\times\ell}$$

where $\rho(M) := \text{tr}(K_{AA^T}M_{1,1} + K_{A\widehat{A}^T}M_{4,1} + K_{\widehat{A}A^T}M_{1,4} + K_{\widehat{A}\widehat{A}^T}M_{4,4})$. Then, $\mathcal{S}(M) = \mathbb{E}[(L - \mathbb{E}L)M(L - \mathbb{E}L)] - \tilde{\mathcal{S}}(M)$ holds with

$$\tilde{\mathcal{S}}(M) := \mathbb{E}\begin{bmatrix} 0 & K_{AA^T}M_{2,1}^T + K_{A\widehat{A}^T}M_{2,4}^T & 0 & 0 \\ M_{1,2}^TK_{AA^T} + M_{4,2}^TK_{\widehat{A}A^T} & 0 & 0 & M_{1,2}^TK_{A\widehat{A}^T} + M_{4,2}^TK_{\widehat{A}\widehat{A}^T} \\ 0 & 0 & 0 & 0 \\ 0 & K_{\widehat{A}A^T}M_{2,1}^T + K_{\widehat{A}\widehat{A}^T}M_{2,4}^T & 0 & 0 \end{bmatrix}.$$

By Theorem 3.1, there exists a unique solution $M \in \mathscr{M}$ such that $M(z)$ solves (4) for every $z \in \mathbb{H}$. Plugging-in the expression for the super-operator above and using the Schur complement formula, we find that

$$M(z) = \begin{bmatrix} ((\delta-z)I_{n_{\text{train}}} - \text{tr}(M_{2,2})K_{AA^T})^{-1} & 0 & -\text{tr}(M_{2,2})M_{1,1}K_{A\widehat{A}^T} & 0 \\ 0 & d^{-1}\,\text{tr}(M_{2,2})I_d & 0 & 0 \\ -\text{tr}(M_{2,2})K_{\widehat{A}A^T}M_{1,1} & 0 & \begin{matrix}(\text{tr}(M_{2,2}))^2 K_{\widehat{A}A^T}M_{1,1}K_{A\widehat{A}^T} \\ +\,\text{tr}(M_{2,2})K_{\widehat{A}\widehat{A}^T}\end{matrix} & -I_{n_{\text{test}}} \\ 0 & 0 & -I_{n_{\text{test}}} & 0 \end{bmatrix} \tag{15}$$

with $M_{2,2} = -(1 + z + \text{tr}(K_{AA^T}M_{1,1}))^{-1}I_d$.

To apply Theorem 3.3, it is necessary to demonstrate that $\|\Delta(L, \tau; z)\|$, as defined in (11), vanishes as $n \to \infty$ for every regularization parameter $\tau \in \mathbb{R}_{>0}$. To achieve this, we employ a leave-one-out method. While the ensuing argument involves detailed and intricate calculations, it is primarily a series of tedious steps. For brevity, we state the result here and defer the proof to Section B.

Lemma 4.1. *Fix $z \in \mathbb{H}$ and $\tau \in \mathbb{R}_{>0}$. Under the settings of Theorem 2.1,* $\lim_{\ell\to\infty} \|\Delta(L, \tau)\| = 0$ *for every $\tau \in \mathbb{R}_{>0}$.*

A key observation that greatly simplifies both the theoretical analysis of the MDE and enables us to derive an iterative procedure for computing its solution is the fact that we can treat the upper-left $n_{\text{train}} + d$ block of the MDE as a separate MDE. This insight allows us to effectively break down the problem and focus on a smaller sub-MDE. Let $L^{(\text{sub})}$ denote the upper-left $n_{\text{train}} + d$ block of L, define a new super-operator

$$\mathcal{S}^{(\text{sub})} : W \in \mathbb{C}^{n_{\text{train}}+d} \mapsto \begin{bmatrix} \text{tr}(W_{2,2})K_{AA^T} & 0 \\ 0 & \text{tr}(K_{AA^T}W_{1,1}) \end{bmatrix} \in \mathbb{C}^{n_{\text{train}}+d}$$

and a new sub-MDE mapping

$$\mathcal{F}^{(\text{sub})} : f \in \mathscr{M}_+^{(\text{sub})} \mapsto (\mathbb{E}L^{(\text{sub})} - \mathcal{S}^{(\text{sub})}(f(\cdot)) - (\cdot)I_{n_{\text{train}}+d})^{-1} \in \mathscr{M}_+^{(\text{sub})}.$$

Here, the set $\mathscr{M}_+^{(\text{sub})}$ refers to a subset of the usual set of $n_{\text{train}} + d \times n_{\text{train}} + d$ matrix-valued functions. Given that the sub-MDE has a spectral parameter spanning its diagonal, the iteration scheme $N_{k+1} = \mathcal{F}^{(\text{sub})}(N_k)$ converges to the unique solution of the sub-MDE $M^{(\text{sub})} = \mathcal{F}^{(\text{sub})}(M^{(\text{sub})})$ for any $N_0 \in \mathscr{M}_+^{(\text{sub})}$, as per general theory [HFS07, Theorem 2.1].

Considering that we want to find a deterministic equivalent for $(L - z\Lambda)^{-1}$ in a neighborhood of $z = 0$, the key lies in establishing control over $M(z)$ in the proximity of $z = 0$. This control is secured through the insights provided by the following lemma.

Lemma 4.2. *Let $z \in \mathbb{H}$ with $|z| < 1 \wedge \delta$ and $M \in \mathscr{M}$ be the unique solution to (15). Then, $\Re[M_{1,1}(z)] \succ 0$ and $\Re[M_{2,2}(z)] \prec 0$. Additionally, $\|M_{1,1}(z)\| \leq (\delta - \Re[z])^{-1}$ and $\|M_{2,2}(z)\| \leq (1 + \Re[z])^{-1}$.*

Proof. Let f be any $(n_{\text{train}} + d) \times (n_{\text{train}} + d)$ matrix-valued analytic function on $\mathbb{H}$ such that $\Im[f(z)] \succ 0$, $f_{1,2}(z) = N_{2,1}(z) = 0$ for every $z \in \mathbb{H}$. Further assume that $\Re[f_{1,1}(z)] \succeq 0$ and $\Re[f_{2,2}(z)] \preceq 0$ for every $z \in \mathbb{H}$ with $|z| < 1 \wedge \delta$. Using the resolvent identity,

$$\Re[\mathcal{F}^{(\text{sub})}(f)] = \mathcal{F}^{(\text{sub})}(f) \begin{bmatrix} \begin{matrix}(\delta-\Re[z])I_{n_{\text{train}}} \\ -\,\text{tr}(\Re[f_{2,2}])K_{AA^T}\end{matrix} & 0 \\ 0 & -(1+\Re[z]+\text{tr}(K_{AA^T}\Re[f_{1,1}])) \end{bmatrix} (\mathcal{F}^{(\text{sub})}(f))^* \tag{16}$$

where we omit the dependence of f on z. Thus, $\Re[\mathcal{F}_{1,1}^{(\text{sub})}(f)] \succ 0$ and $\Re[\mathcal{F}_{2,2}^{(\text{sub})}(f)] \prec 0$ for every $z \in \mathbb{H}$ with $|z| < 1 \wedge \delta$. Additionally, $\mathcal{F}_{1,2}^{(\text{sub})}(f) = \mathcal{F}_{2,1}^{(\text{sub})}(f) = 0$. Since the iterates $f_{k+1} = \mathcal{F}^{(\text{sub})}(f_k)$ converges to the unique solution to the sub-MDE, it must be the case that $\Re[M_{1,1}^{(\text{sub})}(z)] \succ 0$ and $\Re[M_{2,2}^{(\text{sub})}(z)] \prec 0$ for every $z \in \mathbb{H}$ with $|z| < 1 \wedge \delta$. In fact, by uniqueness of the solution to (4), that $\Re[M_{1,1}(z)] \succ 0$ and $\Re[M_{2,2}(z)] \prec 0$ for every $z \in \mathbb{H}$ with $|z| < 1 \wedge \delta$. Using the fact that $\mathcal{F}^{(\text{sub})}(M^{(\text{sub})}) = M^{(\text{sub})}$ and (16), we get

$$\Re[M_{1,1}] \succeq (\delta - \Re[z]) M_{1,1}(M_{1,1})^* \quad \text{and} \quad \Re[M_{2,2}] \preceq -(1 + \Re[z]) M_{2,2}(M_{2,2})^*$$

for every $z \in \mathbb{H}$ with $|z| \leq 1 \wedge \delta$. Since the spectral norm maintains the Loewner partial ordering and the spectral norm of the real and imaginary part of a complex matrix is upper-bounded by the spectral norm of the matrix itself, we derive $\|\Re[M_{1,1}]\| \geq (\delta - \Re[z])\|M_{1,1}(M_{1,1})^*\| = (\delta - \Re[z])\|M_{1,1}\|^2$ and $\|\Re[M_{2,2}]\| \geq (1 + \Re[z])\|M_{2,2}\|$. Rearranging yields the desired result. □

It will be useful later to not only have a deterministic equivalent for $(L - z\Lambda)^{-1}$, but also for $(L^{(\text{sub})} - zI_{n_{\text{train}}})^{-1}$.

Lemma 4.3. *Let $z \in \mathbb{H}$ with $|z| < \delta \wedge 1$ and $M \in \mathscr{M}$ be the unique solution to the sub-MDE. Under the settings of Theorem 2.1,* $\operatorname{tr}(U((L^{(\text{sub})} - zI_{n_{\text{train}}+d})^{-1} - M^{(\text{sub})}(z))) \xrightarrow[n\to\infty]{a.s.} 0$ *for every $U \in \mathbb{C}^{(\text{train}+d)\times \text{train}+d)}$ with $\|U\|_* \leq 1$.*

Proof. Let $D^{(\text{sub})} = \mathbb{E}[(\mathbb{E}L^{(\text{sub})} - L^{(\text{sub})} - \mathcal{S}^{(\text{sub})}(\mathbb{E}(L^{(\text{sub})} - zI_{n_{\text{train}}+d})^{-1}))(L^{(\text{sub})} - zI_{n_{\text{train}}+d})^{-1}]$ be the perturbation matrix, as defined in (10), associated with with linearization $L^{(\text{sub})}$. In particular, $(\mathbb{E}L^{(\text{sub})} - \mathcal{S}(\mathbb{E}(L^{(\text{sub})} - zI_{n_{\text{train}}+d})^{-1}) - zI_{\text{train}+d})\mathbb{E}(L^{(\text{sub})} - zI_{n_{\text{train}}+d})^{-1} = I_{\text{train}+d} + D^{(\text{sub})}$.

Since there is a spectral parameter spanning the entire diagonal in the sub-MDE, it follows from [AEKN19, Corollary 3.8] that Assumption 1 is satisfied for the sub-MDE. Because both $\|A\|$ and $\|\widehat{A}\|$ are centered and bounded in L^4, it is clear that $\limsup_{\ell\to\infty}(\|\mathcal{S}\| \vee \|\mathbb{E}L\|) < \infty$. It also follows from Hölder's inequality that $\limsup_{\ell\to\infty} \mathbb{E}\|(L - z\Lambda)^{-1}\|^2 < \infty$. In particular, Assumption 2 is satisfied.

Given that Assumption 3 is evidently satisfied, and we have demonstrated in Lemma 4.1 that $\|\Delta(L^{(\text{sub})}, \tau; z)\| \to 0$ as $n \to \infty$ for every $\tau \in \mathbb{R}_{>0}$, we only have to establish that the term involving the Lipschitz constant and the term involving the norm of $\tilde{\mathcal{S}}^{(\text{sub})}$ in Theorem 3.3 are asymptotically negligible.

We derive some useful norm bounds. Recall that $R = ((1+z)^{-1}AA^T + (\delta - z)I_{n_{\text{train}}})^{-1}$. For $|z| < 1 \wedge \delta$, $\Re[(1+z)^{-1}] \geq |1+z|^{-2}(1 - |z|) \geq (1 - |z|)/4 > 0$ and $\Re[\delta - z] \geq \delta - |z|$. Hence, $\Re[R] \geq (\delta - |z|)RR^*$ which implies that $\|R\| \leq (\delta - |z|)^{-1}$. A similar argument applied to $\bar{R} = -((1+z)I_d + (\delta - z)^{-1}A^TA)^{-1}$ gives $\|\bar{R}\| \leq (1 - |z|)^{-1}$. Furthermore, we know that $\|RA\|^2 = \|RAA^TR^*\| \leq \|RAA^T\|\|R\|$. By definition, $RAA^T = (1+z)I_{n_{\text{train}}} - (1+z)(\delta - z)R$. Thus, $\|RAA^T\| \leq 2(2+\delta)(\delta - |z|)^{-1}$ and $\|RA\| \leq \sqrt{2(2+\delta)}(\delta - |z|)^{-1}$.

Based on Assumption 3, write $L^{(\text{sub})} \equiv L^{(\text{sub})}(Z) = \mathcal{C}(Z) + \mathbb{E}L^{(\text{sub})}$ for $Z \in \mathbb{R}^{n_0 \times d}$ a matrix of i.i.d. standard normal entries and let λ be the Lipschitz constant associated to the function $Z \in (\mathbb{R}^{n_0\times d}, \|\cdot\|_F) \mapsto \mathcal{S}^{(\text{sub})}((L^{(\text{sub})}(Z) - zI_{n_{\text{train}}+d})^{-1}) \in (\mathbb{C}^{(n_{\text{train}}+d)\times(n_{\text{train}}+d)}, \|\cdot\|_2)$. As mentioned in the discussion following the statement of Theorem 3.3, $\lambda \leq (\Im[z])^{-2}\|\mathcal{S}^{(\text{sub})}\|_{F\mapsto 2}\lambda_{\mathcal{C}}$ where $\lambda_{\mathcal{C}}$ is the Lipschitz constant associated with map $\mathcal{C} : Z \in (\mathbb{R}^{n_0\times d}, \|\cdot\|_F) \mapsto \mathcal{C}(Z) \in (\mathbb{R}^{(n_{\text{train}}+d)\times(n_{\text{train}}+d)}, \|\cdot\|_F)$. For every $N \in \mathbb{C}^{(n_{\text{train}}+d)\times(n_{\text{train}}+d)}$, we can use Cauchy-Schwarz inequality to obtain

$$\|\mathcal{S}^{(\text{sub})}(N)\| \leq \|K_{AA^T}\||\operatorname{tr}(N_{2,2})| + |\operatorname{tr}(K_{AA^T}N_{1,1})| \leq (\sqrt{d} + \sqrt{n_{\text{train}}})\|K_{AA^T}\|\|N\|_F.$$

By Jensen's inequality, $\|K_{AA^T}\| = \|d^{-1}\mathbb{E}[AA^T]\| \le d^{-1}\mathbb{E}\|A\|^2$. In fact, by a similar argument, $\|K_{AA^T}\| \asymp \|K_{\widehat{A}A^T}\| \asymp \|K_{\widehat{A}A^T}\| \asymp \|K_{\widehat{A}\widehat{A}^T}\| \asymp n^{-1}$. Since we assumed that $\|A\|$ is bounded in L^4 and we are working in the proportional limit, $\|\mathcal{S}^{(\text{sub})}\|_{F\mapsto 2} \asymp n^{-1/2}$. Next, let $Z_1, Z_2 \in \mathbb{R}^{n_0\times d}$ and notice that $\|\mathcal{C}(Z_1) - \mathcal{C}(Z_2)\|_F \le n^{-1/2}\lambda_\sigma\lambda_\varphi\|X\|\|Z_1 - Z_2\|$. Since λ_σ, λ_φ and $\|X\|$ are all bounded by assumption, $\lambda_\sigma \asymp n^{-1/2}$ and $\lambda \lesssim (\Im[z])^{-1}n^{-1}$.

Finally, for every $N \in \mathbb{C}^{(n_{\text{train}}+d)\times(n_{\text{train}}+d)}$,

$$\tilde{\mathcal{S}}^{(\text{sub})}(N) = \mathbb{E}\begin{bmatrix} 0 & K_{AA^T}N_{2,1}^T + K_{A\widehat{A}^T}N_{2,4}^T \\ N_{1,2}^T K_{AA^T} + N_{4,2}^T K_{\widehat{A}A^T} & 0 \end{bmatrix}.$$

Since$\|K_{AA^T}\| \asymp \|K_{\widehat{A}A^T}\| \asymp \|K_{\widehat{A}A^T}\| \asymp \|K_{\widehat{A}\widehat{A}^T}\| \asymp n^{-1/2}$, $\|\tilde{\mathcal{S}}^{(\text{sub})}\| \asymp n^{-1/2}$.

Combining everything, the result follows from Corollary 3.2. □

The final prerequisite needed to establish that the solution to (15) acts as a deterministic equivalent for $(L - z\Lambda)^{-1}$, where L is defined in (14), for every $z \in \mathbb{H}$ within a neighborhood of the origin, is to confirm that Assumption 1 holds. We demonstrate this in the following lemma.

Lemma 4.4. *Suppose that $M \in \mathscr{M}$ is the unique solution to (15) and $M^{(\tau)}$ is the unique solution to the regularized version of the same equation. Then, both M and $M^{(\tau)}$ satisfy Assumption 1 for all $z \in \mathbb{H}$ with $|z| < \delta \wedge 1$.*

Proof. Fix $z \in \mathbb{H}$ with $|z| \le 1 \wedge \delta$, $\tau \in \mathbb{R}_{>0}$. Expanding the solution of the regularized MDE block-wise, we obtain

$$\begin{aligned}
M_{1,1}^{(\tau)} &= ((\delta - z - i\tau)I_{n_{\text{train}}} - \operatorname{tr}(M_{2,2}^{(\tau)})K_{AA^T} - (\operatorname{tr}(M_{2,2}^{(\tau)}))^2 K_{A\widehat{A}^T}M_{4,4}^{(\tau)}K_{\widehat{A}A^T})^{-1}\\
M_{2,2}^{(\tau)} &= -(1 + z + i\tau + \rho(M^{(\tau)}))^{-1}I_d\\
M_{3,3}^{(\tau)} &= (i\tau)^{-1}(i\tau I_{n_{\text{train}}}\operatorname{tr}(M_{2,2}^{(\tau)})K_{\widehat{A}\widehat{A}^T})M_{4,4}^{(\tau)} + (i\tau)^{-2}(\operatorname{tr}(M_{2,2}^{(\tau)}))^2 M_{4,4}^{(\tau)}K_{\widehat{A}A^T}M_{1,1}^{(\tau)}K_{A\widehat{A}^T}M_{4,4}^{(\tau)}\\
M_{3,4}^{(\tau)} &= -(i\tau)^{-1}M_{4,4}^{(\tau)}\\
M_{4,4}^{(\tau)} &= i\tau((1+\tau^2)I - i\tau\operatorname{tr}(M_{2,2}^{(\tau)})K_{\widehat{A}\widehat{A}^T})^{-1}\\
M_{1,3}^{(\tau)} &= (i\tau)^{-1}\operatorname{tr}(M_{2,2}^{(\tau)})M_{1,1}^{(\tau)}K_{A\widehat{A}^T}M_{4,4}^{(\tau)}\\
M_{1,4}^{(\tau)} &= -\operatorname{tr}(M_{2,2}^{(\tau)}(z))M_{1,1}^{(\tau)}(z)K_{A\widehat{A}^T}M_{4,4}^{(\tau)}(z).
\end{aligned}$$

It is established in Theorem 3.1 that $\|M_{1,1}\| \vee \|M_{2,2}\| \vee \|M_{1,1}^{(\tau)}\| \vee \|M_{2,2}^{(\tau)}\| \le (\Im[z])^{-1}$. Furthermore, by Cauchy–Schwarz, $|\operatorname{tr}(M_{2,2}^{(\tau)})| \le d\|M_{2,2}^{(\tau)}\|$. Considering that $\|K_{AA^T}\| \asymp \|K_{\widehat{A}A^T}\| \asymp \|K_{\widehat{A}A^T}\| \asymp \|K_{\widehat{A}\widehat{A}^T}\| \asymp n^{-1}$, there exists a constant $c_1 \in \mathbb{R}_{>0}$ such that $\|\operatorname{tr}(M_{2,2}^{(\tau)})K_{\widehat{A}\widehat{A}^T}\| \le c_1$ for every $\tau \in \mathbb{R}_{>0}$ and $n \in \mathbb{N}$. Therefore, taking $\tau \to 0$, we observe that $\|M_{4,4}^{(\tau)}\|$, $\|M_{1,4}^{(\tau)}\|$ and $\|M_{3,4}^{(\tau)} + I_{n_{\text{test}}}\|$ approach 0 uniformly in n.

Consequently, we may consider $\operatorname{BlockDiag}\{M_{1,1}^{(\tau)}, M_{2,2}^{(\tau)}\}$ as nearly satisfying the sub-MDE up to an additive perturbation matrix given block-wise by $D_{1,1}^{(\tau)} = (i\tau I_{n_{\text{train}}} + (\operatorname{tr}(M_{2,2}^{(\tau)}))^2 K_{A\widehat{A}^T}M_{4,4}^{(\tau)}K_{\widehat{A}A^T})M_{1,1}^{(\tau)}$, $D_{2,2}^{(\tau)} = (i\tau + \operatorname{tr}(K_{A\widehat{A}^T}M_{4,1}^{(\tau)} + K_{\widehat{A}A^T}M_{1,4}^{(\tau)} + K_{\widehat{A}\widehat{A}^T}M_{4,4}))M_{2,2}^{(\tau)}$ and $D_{1,2}^{(\tau)} = D_{2,1}^{(\tau)} = 0$. Indeed, $D^{(\text{sub})}$ is vanishing, in norm, as $\tau \to 0$ uniformly in n. Adapting the proof of Theorem 3.2, we leverage the stability property of the MDE to verify that $\|M_{1,1}^{(\tau)} - M_{1,1}\| \vee \|M_{2,2}^{(\tau)} - M_{2,2}\| \lesssim \tau$. Then, we can easily pass this convergence to $M_{1,3}^{(\tau)}$, $M_{3,3}^{(\tau)}$ using the block-wise decomposition, and the other blocks using the fact that the solution to the regularized matrix Dyson equation is Hermitian. □

Lemma 4.5. *Let $z \in \mathbb{H}$ with $|z| < \delta \wedge 1$ and $M \in \mathscr{M}$ be the unique solution to (15). Under the settings of Theorem 2.1,* $\operatorname{tr}(U((L-z)^{-1} - M(z))) \xrightarrow[n\to\infty]{a.s.} 0$ *for every $U \in \mathbb{C}^{\ell\times\ell}$ with $\|U\|_* \leq 1$.*

Proof. Utilizing a similar argument as in the proof of Lemma 4.3, we observe that Assumption 2 is satisfied, $\lim_{n\to\infty} \sqrt{n}\lambda = 0$, and $\limsup_{n\to\infty} \|\tilde{\mathcal{S}}\| = 0$, where λ is the Lipschitz constant defined in Theorem 3.3. In particular, since $\|\Delta(L, \tau; z)\| \to 0$ as $n \to \infty$ for every $\tau \in \mathbb{R}_{>0}$ by Lemma 4.1, it follows from Theorem 3.3 that $\|D^{(\tau)}\| \to 0$ as $n \to \infty$ for every $t \in \mathbb{R}_{>0}$. By Lemma 4.4, Assumption 1 holds. The application of Corollary 3.2 yields the desired result. □

Having established that the solution to (15) acts as a deterministic equivalent for the pseudo-resolvent $(L - z\Lambda)^{-1}$ linked to (14), we aim to retrieve the expression in (2) by taking the spectral parameter to 0. To accomplish this, we need further control over the MDE near the origin.

Lemma 4.6. *Let $z \in \mathbb{H}$ with $|z| < \delta \wedge 1$ and $M \in \mathscr{M}$ the unique solution to (15). Then, for every $\epsilon \in (0, 2^{-1}]$ with $(2(\delta - \Re[z])^{-2} d^2 \|K_{AA^T}\|^2 - 1 - \Re[z])\epsilon \leq 2^{-1}(1 + \Re[z])$,*

$$1 - d\|K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \|M_{2,2}(z)\|^2 \geq \epsilon.$$

Proof. Fix $z \in \mathbb{H}$ with $|z| < 1 \wedge \delta$ and write $M \equiv M(z)$. Let $\epsilon \in (0, 2^{-1}]$ such that

$$(2(\delta - \Re[z])^{-2} d^2 \|K_{AA^T}\|^2 - 1 - \Re[z])\epsilon \leq 2^{-1}(1 + \Re[z])$$

and assume, by contradiction, that $d\|K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \|M_{2,2}(z)\|^2 > 1 - \epsilon$. Using the definition of $M_{1,1}$ and $M_{2,2}$ and repeatedly applying (16),

$$\begin{aligned}
\operatorname{tr}(K_{AA^T}\Re[M_{1,1}]) &= (\delta - \Re[z]) \operatorname{tr}(K_{AA^T} M_{1,1} M_{1,1}^*) - \operatorname{tr}(\Re[M_{2,2}]) \operatorname{tr}(K_{AA^T} M_{1,1} K_{AA^T} M_{1,1}^*) \\
&= (\delta - \Re[z]) \operatorname{tr}(K_{AA^T} M_{1,1} M_{1,1}^*) \\
&\quad + d\|M_{2,2}\|^2 \|K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 (1 + \Re[z]) \\
&\quad + d\|M_{2,2}\|^2 \|K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \operatorname{tr}(K_{AA^T}\Re[M_{1,1}]) \\
&\geq (\delta - \Re[z]) \operatorname{tr}(K_{AA^T} M_{1,1} M_{1,1}^*) + (1 - \epsilon)(1 + \Re[z]) \\
&\quad + (1 - \epsilon) \operatorname{tr}(K_{AA^T}\Re[M_{1,1}]).
\end{aligned}$$

Solving for $\operatorname{tr}(K_{AA^T}\Re[M_{1,1}])$, we obtain that $\operatorname{tr}(K_{AA^T}\Re[M_{1,1}]) = t\epsilon^{-1}$ with

$$t := (\delta - \Re[z]) \operatorname{tr}(K_{AA^T} M_{1,1} M_{1,1}^*) + (1 - \epsilon)(1 + \Re[z]).$$

In particular, taking the real part of $M_{2,2}$, we have

$$-\Re[M_{2,2}] = \|M_{2,2}\|^2 (1 + \Re[z] + \operatorname{tr}(K_{AA^T}\Re[M_{1,1}])) I_d \succeq \|M_{2,2}\|^2 (1 + \Re[z] + t\epsilon^{-1}) I_d.$$

By taking the norm on both sides and leveraging the properties that the spectral norm preserves the Loewner partial ordering and that the spectral norm of the real and imaginary part of a complex matrix is bounded above by the spectral norm of the matrix itself, we obtain

$$\|M_{2,2}\|^2 (1 + \Re[z] + t\epsilon^{-1}) \leq \|\Re[M_{2,2}]\| \leq \|M_{2,2}\|.$$

Rearranging, this implies that $\|M_{2,2}\| \le (1+\Re[z]+t\epsilon^{-1})^{-1}$. By Lemma 4.2 and the definition of ϵ,

$$d\|K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \|M_{2,2}(z)\|^2 \le \frac{(\delta-\Re[z])^{-2} d^2 \|K_{AA^T}\|^2}{1+\Re[z]+t\epsilon^{-1}} \le 2^{-1}.$$

This is a contradiction. Thus, it must be the case that

$$1 - d\|K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \|M_{2,2}(z)\|^2 \ge \epsilon$$

for every $\epsilon \in (0, 2^{-1}]$ with $(2\delta^{-2}d^2\|X\|^4 - 1 - \Re[z])\epsilon \le 2^{-1}(1+\Re[z])$. □

The statement of Lemma 4.6 is intricate because the left-hand side of the inequality $(2\delta^{-2}d^2\|K_{AA^T}\|^2 - 1 - \Re[z])\epsilon \le 2^{-1}(1+\Re[z])$ may be negative. Nevertheless, the essence of Lemma 4.6 lies in the fact that the quantity $1 - d\|K_{AA^T}^{1/2} M_{1,1}(z) K_{AA^T}^{1/2}\|_F^2 \|M_{2,2}(z)\|^2$ can be consistently bounded away from 0 regardless of the dimension.

Now that we have some control on the solution of the MDE when the spectral parameter is close to the origin, we still need to continuously extend the function M to its boundary point 0. To do so, we analytically extend M by reflection to the lower complex plane $\{z \in \mathbb{H} : \Im[z] < 0\}$ through an open interval containing the origin.

Lemma 4.7. *The unique solution to (4) M can be extended analytically to the lower-half complex plane through the open interval $(-(1\wedge\delta), 1\wedge\delta)$.*

Proof. Using the definition of matrix imaginary part and the resolvent identity, we obtain the system of equations

$$\begin{cases} \Im[M_{1,1}] = M_{1,1}(\Im[z] + \operatorname{tr}(\Im[M_{2,2}])K_{AA^T})(M_{1,1})^* \\ \operatorname{tr}(\Im[M_{2,2}]) = d\|M_{2,2}\|^2(\Im[z] + \operatorname{tr}(K_{AA^T}\Im[M_{1,1}])). \end{cases}$$

Combing the two equalities, we get

$$d^{-1}\operatorname{tr}(\Im[M_{2,2}])\left(1 - \|\sqrt{d}K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \|M_{2,2}\|^2\right) = \|M_{2,2}\|^2 \Im[z](1 + \|\sqrt{d}K_{AA^T}^{\frac{1}{2}} M_{1,1}^{(1)}\|_F^2).$$

By Lemma 4.6, $1 - d\|K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \|M_{2,2}(z)\|^2 > 0$ uniformly on $\{z \in \mathbb{H} : |z| \le \epsilon\}$ for every $0 < \epsilon < 1\wedge\delta$. Using Lemma 4.2,

$$d^{-1}\operatorname{tr}(\Im[M_{2,2}]) \le \frac{\Im[z](1 + \|\sqrt{d}(\delta-\Re[z])^{-1}K_{AA^T}^{\frac{1}{2}}\|_F^2)}{1 - \|\sqrt{d}K_{AA^T}^{\frac{1}{2}} M_{1,1}(z) K_{AA^T}^{\frac{1}{2}}\|_F^2 \|M_{2,2}\|^2}.$$

Thus, we observe that $\Im[M_{2,2}(z)] \downarrow 0$ uniformly as $\Im[z] \downarrow 0$ on $(-\epsilon, \epsilon)$ and similarly for $\|\Im[M_{1,1}]\|$. Since $M_{1,1}$ and $M_{2,2}$ fully define the solution of the MDE, $\|\Im[M(z)]\|$ vanishes uniformly for $\Re[z] \in (-\epsilon, \epsilon)$ as $\Im[z] \downarrow 0$.

By Stieltjes inversion lemma, the positive semidefinite measure in Theorem 3.1 has no support in $(-\epsilon, \epsilon)$. We conclude with [GT97, Lemma 5.6]. □

We may now remove the spectral parameter in Lemma 4.5.

Corollary 4.1. *Let $M \in \mathscr{M}$ be the unique solution to (15). Under the settings of Theorem 2.1,* $\operatorname{tr}(U(L^{-1} - M(0))) \xrightarrow[n\to\infty]{a.s.} 0$ *for every $U \in \mathbb{C}^{\ell\times\ell}$ with $\|U\|_* \le 1$.*

(15) has a very nice property: it is fully defined by the scalar $\operatorname{tr}(M_{2,2})$. By fully defined, we mean that if we can compute $\operatorname{tr}(M_{2,2})$ we may explicitly construct the full solution of the MDE. Using the sub-MDE defined above, we get the following numerical result.

Lemma 4.8. *Suppose that $M(0)$ solves (15) when $z = 0$. Let*

$$T : x \in \mathbb{R}_{<0} \mapsto -\left(1 + \operatorname{tr}\left(K_{AA^T}(\delta I_{n_{\text{train}}} - dxK_{AA^T})^{-1}\right)\right)^{-1} \in \mathbb{R}_{<0}$$

and consider the iterates $\{\alpha_k\}_{k\in\mathbb{N}_0}$ obtained via $\alpha_{k+1} = T(\alpha_k)$ for every $k \in \mathbb{N}$ with arbitrary $\alpha_0 \in \mathbb{R}_{\leq 0}$. Then,

$$M(0) = \begin{bmatrix} (\delta I_{n_{\text{train}}} - d\alpha K_{AA^T})^{-1} & 0 & -d\alpha M_{1,1}(0)K_{A\widehat{A}^T} & 0 \\ 0 & \alpha I_d & 0 & 0 \\ -d\alpha K_{\widehat{A}A^T}M_{1,1}(0) & 0 & (d\alpha)^2 K_{\widehat{A}A^T}M_{1,1}(0)K_{A\widehat{A}^T} + d\alpha K_{\widehat{A}\widehat{A}^T} & -I_{n_{\text{test}}} \\ 0 & 0 & -I_{n_{\text{test}}} & 0 \end{bmatrix}$$

where $\alpha := d^{-1}\operatorname{tr}(M_{2,2}(0)) = \lim_{k\to\infty}\alpha_k$.

Proof. In order to use Earle-Hamilton fixed-point theorem [EH70], we consider the set $\mathscr{S}$ of complex matrices $N \in \mathbb{C}^{\ell\times\ell}$ with $N_{1,2} = N_{2,1} = 0$, $\Re[N_{1,1}] \succ 0$ and $\Re[N_{2,2}] \prec 0$. Slightly abusing notation, consider

$$\mathcal{F}^{(\text{sub})} : N \in \mathscr{S} \mapsto (\mathbb{E}L - \mathcal{S}^{(\text{sub})}(N))^{-1} \in \mathscr{S}$$

with $\mathscr{S} := \{N \in \mathbb{C}^{\ell\times\ell} : \Re[N_{1,1}] \succ 0, \Re[N_{2,2}] \prec 0, N_{1,2} = N_{2,1} = 0\}$. Using an argument analogous to the one in [HFS07], we get the existence of a unique matrix $N \in \mathscr{S}$ such that $\mathcal{F}^{(\text{sub})}(N) = N$. By uniqueness, $N = M_{1,1}(0)$. Additionally, by Earle-Hamilton fixed point theorem, the sequence $\{N_k : k \in \mathbb{N}_0\}$ with $N_{k+1} = \mathcal{F}^{(\text{sub})}(N_k)$ for every $k \in \mathbb{N}$ converges to $M_{1,1}(0)$ for every $N_0 \in \mathscr{S}$. Choosing $N_0 = \operatorname{BlockDiag}\{I_{n_{\text{train}}}, \alpha_0 I_d\}$ gives the result. ☐

4.2. Second deterministic equivalent

We now consider the squared matrix $(AA^T + \delta I_{n_{\text{train}}})^{-1}A\widehat{A}^T\widehat{A}A^T(AA^T + \delta I_{n_{\text{train}}})^{-1}$. Notice that

$$(L^{-2})_{1,1} = R^2 + RA^TAR + RA\widehat{A}^T\widehat{A}AR$$

and $(L^{(\text{sub})})^{-2}_{1,1} = R^2 + RA^TAR$ for $R := (AA^T + \delta I_{n_{\text{train}}})^{-1}$. Rearranging, we get that

$$(AA^T + \delta I_{n_{\text{train}}})^{-1}A\widehat{A}^T\widehat{A}A(AA^T + \delta I_{n_{\text{train}}})^{-1} = (L^{-2})_{1,1} - (L^{(\text{sub})})^{-2}_{1,1}. \tag{17}$$

Therefore, it suffices to find deterministic equivalents for $(L^{(\text{sub})})^{-2}$ and L^{-2} to obtain an anisotropic law for the random matrix $(AA^T + \delta I_{n_{\text{train}}})^{-1}A\widehat{A}^T\widehat{A}A(AA^T + \delta I_{n_{\text{train}}})^{-1}$.

Lemma 4.9. *Under the settings of Theorem 2.1, let $\alpha = d^{-1}\operatorname{tr}(M_{2,2}(0))$ as in Lemma 4.8, $R = (AA^T + \delta I_{n_{\text{train}}})^{-1}$ and define*

$$\beta = \frac{\alpha^2 \operatorname{tr}\left(K_{\widehat{A}\widehat{A}^T} + d\alpha K_{\widehat{A}A^T}M_{1,1}(0)(I_{n_{\text{train}}} + \delta M_{1,1}(0))K_{A\widehat{A}^T}\right)}{1 - \|\sqrt{d}\alpha K^{\frac{1}{2}}_{AA^T}M_{1,1}(0)K^{\frac{1}{2}}_{AA^T}\|_F^2} \in \mathbb{R}_{\geq 0}.$$

Then, $\operatorname{tr} U(RA\widehat{A}^T\widehat{A}A^TR - d\beta M_{1,1}(0)K_{AA^T}M_{1,1}(0) - M_{1,3}(0)M_{3,1}(0)) \xrightarrow[n\to\infty]{a.s.} 0$ for every $U \in \mathbb{C}^{n_{\text{train}}\times n_{\text{train}}}$ with $\|U\|_* \leq 1$.

Proof. First, we note that $i\tau \mapsto (L - i\tau)^{-1}$ is an analytic function with $\partial_{i\tau}(L - i\tau)^{-1} = (L - i\tau)^{-2}$. Overloading notation, let $M^{(\zeta)} \in \mathscr{M}_+$ be the unique solution to the MDE $(\mathbb{E}L - \mathcal{S}(M^{(\zeta)}) - \zeta I_\ell)M^{(\zeta)} = I_\ell$ where $\zeta \in \mathbb{H}$. By the proof of [EKN20, Theorem 2.14], the function $\zeta \mapsto M^{(\zeta)}(0)$ is analytic on $\mathbb{H}$. Adapting a general argument resembling to the one in [SCDL23, equation (174)], it follows from Cauchy's integral formula that

$$(L - i\tau)^{-2} - \partial_{i\tau} M^{(\tau)}(0) = \partial_{i\tau}\left((L - i\tau)^{-1} - M^{(\tau)}(0)\right) = \frac{1}{2\pi}\oint_\gamma \frac{(L - \zeta)^{-1} - M^{(\zeta)}(0)}{(\zeta - i\tau)^2}\mathrm{d}\zeta$$

where γ forms a counterclockwise circle of radius $\tau/2$ around $i\tau$. We know that $M^{(\zeta)}(0)$ is a deterministic equivalent for $(L - \zeta)^{-1}$ for every fixed $\zeta \in \mathbb{H}$. By the resolvent identity, $\zeta \mapsto (L - \zeta I_\ell)$ is $4/\theta^2$-Lipschitz on $\{z \in \mathbb{H} : \Im[z] \geq \tau/2\}$. Similarly, by the proof of [EKN20, Theorem 2.14], the function $\zeta \mapsto M^{(\zeta)}$ is $(2/\tau)^{12}$-Lipschitz on $\{z \in \mathbb{H} : \Im[z] \geq \tau/2\}$. Therefore, we obtain $\operatorname{tr}(U((L - i\tau)^{-2} - \partial_{i\tau} M^{(\tau)}(0))) \xrightarrow[n\to\infty]{a.s.} 0$ for every $\tau \in \mathbb{R}_{>0}$ and $U \in \mathbb{C}^{\ell\times\ell}$ with $\|U\|_* \leq 1$. Taking the derivative of (7), we obtain $\partial_{i\tau} M^{(\tau)}(0) = M^{(\tau)}(0)(\mathcal{S}(\partial_{i\tau} M^{(\tau)}(0)) + I_\ell)M^{(\tau)}(0)$ or, relating this equation to the stability operator, $\mathcal{L}^{(\tau)}(\partial_{i\tau} M^{(\tau)}(0)) = (M^{(\tau)}(0))^2$ with $\mathcal{L}^{(\tau)} : N \in \mathbb{C}^{\ell\times\ell} \mapsto N - M^{(\tau)}(0)\mathcal{S}(N)M^{(\tau)}(0)$.

In what follows, we omit the argument of $M^{(\tau)}$ and write $M^{(\tau)} \equiv M^{(\tau)}(0)$. Using easy but tedious computations, we decompose $\partial_{i\tau} M^{(\tau)}_{j,k} = C_{j,k} + D_{j,k}\operatorname{tr}(\partial_{i\tau} M^{(\tau)}_{2,2})$ for every $j, k \in \{(1,1),(1,4),(4,4)\}$ with

$$\begin{aligned}
C_{1,1} &:= M^{(\tau)}_{1,4}M^{(\tau)}_{4,1} + (M^{(\tau)}_{1,1})^2 + M^{(\tau)}_{1,3}M^{(\tau)}_{3,1},\\
D_{1,1} &:= M^{(\tau)}_{1,1}K_{AA^T}M^{(\tau)}_{1,1} + M^{(\tau)}_{1,1}A\widehat{A}^T M^{(\tau)}_{4,1} + M^{(\tau)}_{1,4}K_{\widehat{A}A^T}M^{(\tau)}_{1,1} + M^{(\tau)}_{1,4}K_{\widehat{A}\widehat{A}^T}M^{(\tau)}_{4,1},\\
C_{4,4} &:= M^{(\tau)}_{4,1}M^{(\tau)}_{1,4} + (M^{(\tau)}_{4,4})^2 + M^{(\tau)}_{4,3}M^{(\tau)}_{3,4},\\
D_{4,4} &:= M^{(\tau)}_{4,1}K_{AA^T}M^{(\tau)}_{1,4} + M^{(\tau)}_{4,1}K_{A\widehat{A}^T}M^{(\tau)}_{4,4} + M^{(\tau)}_{4,4}K_{\widehat{A}A^T}M^{(\tau)}_{1,4} + M^{(\tau)}_{4,4}K_{\widehat{A}\widehat{A}^T}M^{(\tau)}_{4,4},\\
C_{1,4} &:= M^{(\tau)}_{1,1}M^{(\tau)}_{1,4} + M^{(\tau)}_{1,4}M^{(\tau)}_{4,4} + M^{(\tau)}_{1,3}M^{(\tau)}_{3,4}, \text{ and}\\
D_{1,4} &:= M^{(\tau)}_{1,1}K_{AA^T}M^{(\tau)}_{1,4} + M^{(\tau)}_{1,1}K_{A\widehat{A}^T}M^{(\tau)}_{4,4} + M^{(\tau)}_{1,4}K_{\widehat{A}A^T}M^{(\tau)}_{1,4} + M^{(\tau)}_{1,4}K_{AA^T}M^{(\tau)}_{4,4}.
\end{aligned}$$

Taking the trace of the 2,2 block of $\partial_{i\tau} M^{(\tau)}(0)$, we get

$$\begin{aligned}
\operatorname{tr}(\partial_{i\tau} M^{(\tau)}_{2,2}) &= \operatorname{tr}((M^{(\tau)}_{2,2})^2)(\rho(\partial_{i\tau} M^{(\tau)}) + 1)\\
&= \operatorname{tr}((M^{(\tau)}_{2,2})^2)(\operatorname{tr}(K_{AA^T}C_{1,1} + K_{A\widehat{A}^T}C^T_{1,4} + K_{\widehat{A}A^T}C_{1,4} + K_{\widehat{A}\widehat{A}^T}C_{4,4}) + 1)\\
&\quad + \operatorname{tr}(\partial_{i\tau} M^{(\tau)}_{2,2})\operatorname{tr}((M^{(\tau)}_{2,2})^2)\operatorname{tr}(K_{AA^T}D_{1,1} + K_{A\widehat{A}^T}D^T_{1,4} + K_{\widehat{A}A^T}D_{1,4} + K_{\widehat{A}\widehat{A}^T}D_{4,4}).
\end{aligned}$$

By the proof of Lemma 4.5, we observe that there exists a function $f : \mathbb{R}_{>0} \mapsto \mathbb{R}_{\geq 0}$ with $\lim_{\tau\downarrow 0} f(\tau) = 0$ such that $\|D_{4,4}\| \leq f(\tau) + o_n(1)$, $\|D_{1,1} - M_{1,1}K_{AA^T}M_{1,1}\| \leq f(\tau) + o(1)$, $\|D_{1,4}\| = \|D^T_{4,1}\| \leq f(\tau) + o_n(1)$ and $\|M^{(\tau)}_{2,2} - M_{2,2}\| \leq f(\tau) + o_n(1)$. Since $\limsup_{n\to\infty} d\|K_{AA^T}\| = \limsup_{n\to\infty} d\|\mathbb{E}a_1a_1^T\| \leq \limsup_{n\to\infty} \mathbb{E}\|A\|^2 < \infty$, we have

$$\liminf_{n\to\infty} \sup\{\epsilon \in [0, 2^{-1}] : (2\delta^{-2}d^2\|K_{AA^T}\|^2 - 1 - \Re[z])\epsilon \leq 2^{-1}(1 + \Re[z])\} > 0$$

by Lemma 4.6 and, consequently,

$$|1 - \operatorname{tr}\left((M^{(\tau)}_{2,2})^2\right)\operatorname{tr}(K_{AA^T}D_{1,1} + K_{A\widehat{A}^T}D_{4,1} + K_{\widehat{A}A^T}D_{1,4} + K_{\widehat{A}\widehat{A}^T}D_{4,4})| \gg 0$$

for every $n \in \mathbb{N}$ large enough and $\tau \in \mathbb{R}_{>0}$ small enough. In particular, the limit $\lim_{\tau\downarrow 0} \partial_{i\tau} M^{(\tau)}$ exists and satisfies

$$\begin{aligned}\lim_{\tau\downarrow 0} d^{-1}\operatorname{tr}(\partial_{i\tau} M_{2,2}^{(\tau)}) &= \frac{\alpha^2(\operatorname{tr}(K_{AA^T}M_{1,1}^2 + K_{AA^T}M_{1,3}M_{3,1} - K_{A\widehat{A}^T}M_{3,1} - K_{\widehat{A}A^T}M_{1,3} + K_{\widehat{A}\widehat{A}^T}) + 1)}{1 - \|\sqrt{d}\alpha K_{AA^T}^{\frac{1}{2}} M_{1,1}(0) K_{AA^T}^{\frac{1}{2}}\|_F^2} \\ &= \beta + \frac{\alpha^2(1 + \operatorname{tr}(K_{AA^T}M_{1,1}^2))}{1 - \|\sqrt{d}\alpha K_{AA^T}^{\frac{1}{2}} M_{1,1}(0) K_{AA^T}^{\frac{1}{2}}\|_F^2}\end{aligned}$$

where we recall that $\alpha = d^{-1}\operatorname{tr}(M_{2,2})$ as defined in Lemma 4.8. Plugging this into the expression for $\partial_{i\tau} M_{1,1}^{(\tau)}$ and taking the limit as $\tau \downarrow 0$, we get that

$$d\beta M_{1,1}K_{AA^T}M_{1,1} + M_{1,3}M_{3,1} + M_{1,1}^2 + \frac{d\alpha^2(1 + \operatorname{tr}(K_{AA^T}M_{1,1}^2))}{1 - \|\sqrt{d}\alpha K_{AA^T}^{\frac{1}{2}} M_{1,1}(0) K_{AA^T}^{\frac{1}{2}}\|_F^2} M_{1,1}K_{AA^T}M_{1,1}$$

is an asymptotic deterministic equivalent for $(L^{-2})_{1,1}$.

Using a similar argument, we note that $\partial_z M^{(\mathrm{sub})}(0)$ is a deterministic equivalent for $(L^{(\mathrm{sub})})^{-2}$ and

$$\partial_z M_{1,1}^{(\mathrm{sub})} = M_{1,1}^2 + \frac{d\alpha^2(1 + \operatorname{tr}(K_{AA^T}M_{1,1}^2))}{1 - \|\sqrt{d}\alpha K_{AA^T}^{\frac{1}{2}} M_{1,1}(0) K_{AA^T}^{\frac{1}{2}}\|_F^2} M_{1,1}K_{AA^T}M_{1,1}.$$

We obtain the result by (17). □

5. Concluding remarks

In conclusion, our work extended the matrix Dyson equation framework to accommodate linearizations with general correlation structures. This extension allowed us to derive an asymptotically exact expression for the empirical test error of random features ridge regression.

Looking forward, our matrix Dyson equation framework holds promise for analyzing the largest eigenvalue of an anisotropic version of the conjugate kernel, thereby extending existing research in this direction [BP22]. While our result on the empirical test error has its limitations, we believe there are avenues for improvement. Firstly, the assumption of norm of the random features matrices being bounded in L^4 might be relaxed, allowing for more general scenarios, such as low-rank spikes, which could be explored with a more delicate analysis. Secondly, our empirical simulations, see Figure 1 for instance, suggest that the result might extend beyond Lipschitz activation functions, opening up possibilities for further exploration in this direction.

Acknowledgments

The authors would like to thank David Renfrew for helpful conversations relating to the Matrix Dyson equation. The author would also like to thank Konstantinos Christopher Tsiolis for his help at multiple stages of the project.

Funding

The first author was supported by the Canada CIFAR AI Chair Program (held at Mila by Courtney Paquette). The second author was supported by NSERC Discovery Grant RGPIN-2020-04974.

References

[AEK17] Oskari Ajanki, Laszlo Erdos, and Torben Krüger. Universality for general wigner-type matrices. *Probability Theory and Related Fields*, 169, 12 2017.

[AEK18] Johannes Alt, Laszlo Erdos, and Torben Krüger. The dyson equation with linear self-energy: spectral bands, edges and cusps. *Documenta Mathematica*, 25:1421–1539, 04 2018.

[AEK19a] Oskari Ajanki, László Erdős, and Torben Krüger. Quadratic vector equations on complex upper half-plane. *Memoirs of the American Mathematical Society*, 261(1261), 9 2019.

[AEK19b] Oskari H. Ajanki, László Erdős, and Torben Krüger. Stability of the Matrix Dyson Equation and Random Matrices with Correlations. *Probability Theory and Related Fields*, 173(1-2):293–373, 2 2019.

[AEKN19] Johannes Alt, László Erdős, Torben Krüger, and Yuriy Nemish. Location of the spectrum of Kronecker random matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55(2):661 – 696, 2019.

[AEKS20] Johannes Alt, László Erdős, Torben Krüger, and Dominik Schröder. Correlated random matrices: Band rigidity and edge universality. *The Annals of Probability*, 48(2), 3 2020.

[AKE17] Oskari Ajanki, Torben Krüger, and LÁszlÓ ERDŐS. Singularities of solutions to quadratic vector equations on the complex upper half-plane. *Communications on Pure and Applied Mathematics*, 70(9):1672–1705, 2017.

[ALP19] Ben Adlam, Jake Levinson, and Jeffrey Pennington. A random matrix perspective on mixtures of nonlinearities in high dimensions. In *International Conference on Artificial Intelligence and Statistics*, 2019.

[Alt18] Johannes Alt. *Dyson equation and eigenvalue statistics of random matrices*. Phd thesis, IST Austria, 07 2018.

[And13] Greg W. Anderson. Convergence of the largest singular value of a polynomial in independent wigner matrices. *The Annals of Probability*, 41(3B), 5 2013.

[And15] Greg W. Anderson. A local limit law for the empirical spectral distribution of the anticommutator of independent wigner matrices. *Annales de l'I.H.P. Probabilités et statistiques*, 51(3):809–841, 2015.

[AP20a] Ben Adlam and Jeffrey Pennington. The neural tangent kernel in high dimensions: Triple descent and a multi-scale theory of generalization. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20, 2020.

[AP20b] Ben Adlam and Jeffrey Pennington. Understanding double descent requires a fine-grained bias-variance decomposition. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[BMS17] Serban T. Belinschi, Tobias Mai, and Roland Speicher. Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2017(732):21–53, 2017.

[BP21] Lucas Benigni and Sandrine Péché. Eigenvalue distribution of some nonlinear models of random matrices. *Electronic Journal of Probability*, 26(none), 1 2021.

[BP22] Lucas Benigni and Sandrine Péché. Largest eigenvalues of the conjugate kernel of single-layered neural networks, 2022.

[BvH23] Tatiana Brailovskaya and Ramon van Handel. Universality and sharp matrix concentration inequalities, 2023.

[Cho22] Clément Chouard. Quantitative deterministic equivalent of sample covariance matrices with a general dependence structure, 2022.

[DL20] Oussama Dhifallah and Yue M. Lu. A precise performance analysis of learning with random features, 2020.

[dSB21] Stéphane d'Ascoli, Levent Sagun, and Giulio Biroli. Triple descent and the two kinds of overfitting: where and why do they appear? *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12):124002, 12 2021.

[EH70] Clifford J. Earle and Richard S Hamilton. A fixed point theorem for holomorphic mappings. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XVI, Berkeley, 1968)*, volume 16, pages 61–65, Rhode Island, 1970. AMS.

[EKN20] László Erdős, Torben Krüger, and Yuriy Nemish. Local laws for polynomials of wigner matrices. *Journal of Functional Analysis*, 278(12):108507, 2020.

[EKS19] László Erdős, Torben Krüger, and Dominik Schröder. Random matrices with slow correlation decay. *Forum of Mathematics, Sigma*, 7, 2019.

[Erd19] Laszlo Erdos. The matrix dyson equation and its applications for random matrices, 2019.

[FKN23] Jacob Fronk, Torben Krüger, and Yuriy Nemish. Norm convergence rate for multivariate quadratic polynomials of wigner matrices, 2023.

[FOBS06] Reza Rashidi Far, Tamer Oraby, Wlodzimierz Bryc, and Roland Speicher. Spectra of large block matrices, 2006.

[FW20] Zhou Fan and Zhichao Wang. Spectra of the conjugate kernel and neural tangent kernel for linear-width neural networks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[GLK+20] Federica Gerace, Bruno Loureiro, Florent Krzakala, Marc Mezard, and Lenka Zdeborova. Generalisation error in learning with random features and the hidden manifold model. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3452–3462. PMLR, 13–18 Jul 2020.

[GLR+22] Sebastian Goldt, Bruno Loureiro, Galen Reeves, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. The gaussian equivalence of generative models for learning with shallow neural networks. In *Mathematical and Scientific Machine Learning*, pages 426–471. PMLR, 2022.

[GT97] Fritz Gesztesy and Eduard Tsekanovskii. On matrix-valued herglotz functions. *Mathematische Nachrichten*, 218:61–138, 1997.

[Har79] Lawrence A. Harris. Schwarz-pick systems of pseudometrics for domains in normed linear spaces. In Jorge Albedo Barroso, editor, *Advances in Holomorphy*, volume 34 of *North-Holland Mathematics Studies*, pages 345–406. North-Holland, 1979.

[Har03] Lawrence Harris. Fixed point of holomorphic mappings for domains in banach spaces. *Abstract and Applied Analysis*, 2003, 2003.

[HFS07] J. William Helton, Reza Rashidi Far, and Roland Speicher. Operator-valued semicircular elements: Solving a quadratic matrix equation with positivity constraints. *International Mathematics Research Notices*, 2007.

[HL23] Hong Hu and Yue M. Lu. Universality laws for high-dimensional learning with random features. *IEEE Transactions on Information Theory*, 69(3):1932–1964, 2023.

[HMRT22] Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J. Tibshirani. Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, 50(2):949 – 986, 2022.

[HMS18] J. William Helton, Tobias Mai, and Roland Speicher. Applications of realizations (aka linearizations) to free probability. *Journal of Functional Analysis*, 274(1):1–79, 2018.

[HMV06] J. William Helton, Scott A. McCullough, and Victor Vinnikov. Noncommutative convexity arises from linear matrix inequalities. *Journal of Functional Analysis*, 240(1):105–191, 2006.

[HT05] Uffe Haagerup and Steen Thorbjørnsen. A new application of random matrices: $\mathrm{Ext}(c^*_{\mathrm{red}}(\mathbb{F}_2))$ is not a group. *Annals of Mathematics. Second Series*, 2, 2005.

[JcS+20] Arthur Jacot, Berfin Şimşek, Francesco Spadaro, Clément Hongler, and Franck Gabriel. Implicit regularization of random feature models. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

[LCM21] Zhenyu Liao, Romain Couillet, and Michael W Mahoney. A random matrix analysis of random fourier features: beyond the gaussian kernel, a precise phase transition, and the corresponding double descent. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12):124006, 12 2021.

[Led01] Michel Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Soc., 2001.

[LLC18] Cosme Louart, Zhenyu Liao, and Romain Couillet. A random matrix approach to neural networks. *The Annals of Applied Probability*, 28(2):1190 – 1248, 2018.

[LN17] Annemarie Luger and Mitja Nedic. A characterization of herglotz–nevanlinna functions in two variables via integral representations. *Arkiv för Matematik*, 55(1):199–216, 2017.

[LP09] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *The Annals of Probability*, 37(5), 9 2009.

[LS02] Tzon-Tzer Lu and Sheng-Hua Shiou. Inverses of 2 × 2 block matrices. *Computers & Mathematics with Applications*, 43(1):119–129, 2002.

[MG21] Gabriel Mel and Surya Ganguli. A theory of high dimensional regression with arbitrary correlations between input features and target functions: sample complexity, multiple descent curves and a hierarchy of phase transitions. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7578–7587. PMLR, 18–24 Jul 2021.

[MM22] Song Mei and Andrea Montanari. The generalization error of random features regression: Precise asymptotics and the double descent curve. *Communications on Pure and Applied Mathematics*, 75(4):667–766, 2022.

[MMM22] Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Generalization error of random feature and kernel methods: Hypercontractivity and kernel matrix concentration. *Applied and Computational Harmonic Analysis*, 59:3–84, 2022. Special Issue on Harmonic Analysis and Machine Learning.

[MP22] Gabriel Mel and Jeffrey Pennington. Anisotropic random feature regression in high dimensions. In *International Conference on Learning Representations*, 2022.

[MS22] Andrea Montanari and Basil N Saeed. Universality of empirical risk minimization. In *Conference on Learning Theory*, pages 4310–4312. PMLR, 2022.

[Pas05] L. A. Pastur. A simple approach to the global regime of gaussian ensembles of random matrices. *Ukrainian Mathematical Journal*, 57(6):936–966, 2005.

[PS21] Vanessa Piccolo and Dominik Schröder. Analysis of one-hidden-layer neural networks via the resolvent method. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

[PW17] Jeffrey Pennington and Pratik Worah. Nonlinear random matrix theory for deep learning. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[RR07] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

[SCDL23] Dominik Schröder, Hugo Cui, Daniil Dmitriev, and Bruno Loureiro. Deterministic equivalent and error universality of deep random features learning. In *International Conference on Machine Learning*, 2023.

[Ste81] Charles M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics*, 9(6):1135 – 1151, 1981.

[Tao12] Terence Tao. *Topics in Random Matrix Theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

[TAP21a] Nilesh Tripuraneni, Ben Adlam, and Jeffrey Pennington. Covariate shift in high-dimensional random feature regression, 2021.

[TAP21b] Nilesh Tripuraneni, Ben Adlam, and Jeffrey Pennington. Overparameterization improves robustness to covariate shift in high dimensions. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[WX20] Denny Wu and Ji Xu. On the optimal weighted ℓ_2 regularization in overparameterized linear regression. In *Advances in Neural Information Processing Systems*, volume 33, pages 10112–10123. Curran Associates, Inc., 2020.

[WZ23] Zhichao Wang and Yizhe Zhu. Overparameterized random feature regression with nearly orthogonal data. In Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 8463–8493. PMLR, 25–27 Apr 2023.

[ZBH$^+$21] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM*, 64(3):107–115, feb 2021.

Appendix A: Theory for the linearized matrix Dyson equation

This section of the supplement is dedicated to establishing a framework, grounded in the matrix Dyson equation, for deriving anisotropic global laws for general pseudo-resolvents. We commence with a recapitulation of the settings, mirroring the presentation in the main text, supplemented by additional definitions crucial to our analysis. Subsequently, we introduce key properties of the matrix Dyson equation, laying the groundwork for proving the existence of a unique solution and demonstrating an asymptotic stability property.

A.1. Settings

We focus on a class of real[5] self-adjoint linearizations denoted as

$$L = \begin{bmatrix} A & B^T \\ B & Q \end{bmatrix} \in \mathbb{R}^{\ell \times \ell}, \tag{18}$$

where $A \in \mathbb{R}^{n\times n}$ is a potentially random self-adjoint complex matrix, $Q \in \mathbb{R}^{d\times d}$ is a deterministic invertible self-adjoint matrix and $B \in \mathbb{R}^{d\times n}$ is a potentially random arbitrary matrix. Our primary interest lies in analyzing the behavior in high dimensions of the *pseudo-resolvent* $(L - z\Lambda)^{-1}$, where $\Lambda := \mathrm{BlockDiag}\{I_{n\times n}, 0_{d\times d}\}$ and $z \in \mathbb{H} := \{z \in \mathbb{C} : \Im[z] > 0\}$ represents the upper half of the complex plane.

Our framework relies on the *linearized matrix Dyson equation (MDE)*

$$(\mathbb{E}L - \mathcal{S}(M) - z\Lambda)\, M = I_\ell, \tag{19}$$

where the spectral parameter z is chosen from the upper half complex plane $\mathbb{H}$. Here, the *super-operator*

$$\mathcal{S} : M \in \mathbb{C}^{\ell\times\ell} \mapsto \mathbb{E}\begin{bmatrix} [(L-\mathbb{E}L)M(L-\mathbb{E}L)]_{1,1} & (A-\mathbb{E}A)M_{1,1}(B^T-\mathbb{E}B^T) \\ (B-\mathbb{E}B)M_{1,1}(A-\mathbb{E}A) & (B-\mathbb{E}B)M_{1,1}(B-\mathbb{E}B)^T \end{bmatrix} \in \mathbb{C}^{\ell\times\ell}. \tag{20}$$

is a positivity-preserving linear map which encodes the second moment of L. For conciseness, we let $s \in \mathbb{R}_{>0}$ such that $\|\mathcal{S}(W)\| \le s\|W\|$ for every $W \in \mathbb{C}^{\ell\times\ell}$, but also $\|\mathcal{S}_{i,j}(W)\| \le s\|W_{1,1}\|$ for all $(i,j) \in \{(1,2),(2,1),(2,2)\}$ and $W \in \mathbb{C}^{\ell\times\ell}$. This condition is reminiscent of the upper-bound in the *flatness* condition commonly assumed in matrix Dyson equation literature [Erd19, AEKN19, Alt18]. Consequently, we will adopt this terminology to refer to this condition.

In order to ensure the existence of a unique solution to the matrix Dyson equation, we need to restrict (19) to a suitable set. Consequently, we introduce the *admissible set*[6]

$$\mathscr{M} := \{f : \mathbb{H} \mapsto \mathscr{A} \text{ analytic}\}, \quad \mathscr{A} := \{W \in \mathbb{C}^{\ell\times\ell} : \Im[W] \succeq 0 \text{ and } \Im[W_{1,1}] \succ 0\}. \tag{21}$$

Our primary strategy for analyzing (19) involves initially establishing analogous results for a regularized version of the equation. This regularization typically simplifies the problem,

[5] Our framework can easily be extended to complex linearizations $L \in \mathbb{C}^{\ell\times\ell}$ with $\Im[L] \preceq 0$. Because we do not have any application for this generalization in mind, we do not pursue this direction for the sake of clarity.

[6] Alternatively, assuming that $A = \mathbb{E}A$ and $\mathbb{E}B = 0$, we can choose $\mathcal{S}(M) = \mathbb{E}[(L-\mathbb{E}L)M(L-\mathbb{E}L)]$ and consider the set

$$\mathscr{A} := \{W \in \mathbb{C}^{\ell\times\ell} : \Im[W] \succeq 0 \text{ and } \Im[W_{1,1}] \succ 0,\ W_{1,2} = W_{2,1} = 0\}.$$

Applying our framework under these assumptions requires no modifications, making it particularly suitable, for example, for the study of powers of Wigner matrices.

enabling us to leverage existing knowledge. Subsequently, we demonstrate the feasibility of setting the regularization parameter to zero, effectively reverting to the original equation. Importantly, we ensure that the statements derived for the regularized variant remain valid in this limit, thereby providing valuable insights into the properties of (19). For this reason, we introduce the *regularized matrix Dyson equation (RMDE)*

$$(\mathbb{E}L - \mathcal{S}(M^{(\tau)}) - z\Lambda - i\tau I_\ell)M^{(\tau)} = I_\ell \tag{22}$$

for every $\tau > 0$. The corresponding admissible set is given by

$$\mathscr{M}_+ := \{f : \mathbb{H} \mapsto \mathscr{A}_+ \text{ analytic}\}, \quad \mathscr{A}_+ := \{W \in \mathbb{C}^{\ell\times\ell} : \Im[W] \succ 0\} \cap \mathscr{A}. \tag{23}$$

It will be convenient to sometimes view (19) as a fixed point equation, so we introduce the *MDE map*

$$\mathcal{F} : f \in \mathscr{M} \mapsto (\mathbb{E}L - \mathcal{S}(f(\cdot)) - (\cdot)\Lambda)^{-1} \in \mathscr{M}, \tag{24}$$

assuming its well-definedness, which we establish in Lemmas A.1 and A.2. With this definition, we can reexpress the MDE (19) as $M = \mathcal{F}(M)$. Whenever convenient, we will fix a spectral parameter $z \in \mathbb{H}$ and operate with $\mathcal{F}$ over $\mathscr{A}$. Similarly, the formulation of (22) becomes $M^{(\tau)} = \mathcal{F}^{(\tau)}(M^{(\tau)})$, where

$$\mathcal{F}^{(\tau)} : f \in \mathscr{M}_+ \mapsto (\mathbb{E}L - \mathcal{S}(f(\cdot)) - (\cdot)\Lambda - i\tau I_\ell)^{-1} \in \mathscr{M}_+ \tag{25}$$

is the *RMDE map*.

For every $\tau \in \mathbb{R}_{\geq 0}$, let

$$M_\star^{(\tau)} = (Q - i\tau I_d)^{-1} \quad \text{and} \quad M_\infty^{(\tau)} = \begin{bmatrix} 0_{n\times n} & 0_{n\times d} \\ 0_{d\times n} & M_\star^{(\tau)} \end{bmatrix} \tag{26}$$

such that $\Im[M_*^{(\tau)}] \succeq 0$ and denote $M_\star = M_\star^{(0)}$, $M_\infty = M_\infty^{(0)}$. We remind the reader that Q denotes the deterministic invertible lower-right $d \times d$ submatrix of the linearization L. We will see in Lemma A.4 that $M_\star^{(\tau)}$ corresponds precisely to the limit of $M^{(\tau)}(z)$ as $|z|$ diverges to infinity in the upper-half complex plane.

Since the RMDE (22) corresponds to the MDE (19) when $\tau = 0$, we will always write $M^{(0)} = M$, $\mathcal{F}^{(0)} = \mathcal{F}$, etc. Unless stated otherwise, we will assume that $z \in \mathbb{H}$ is fixed throughout the section. We will abuse notation and omit to write the dependence of M on z, using $M \equiv M(z)$ instead.

A.2. Main Properties

In this subsection, we will present and prove a series of properties of the (R)MDE. By doing so, we are essentially laying the groundwork for demonstrating the existence and uniqueness of a solution to the linearized MDE, as well as establishing its stability.

A.2.1. General properties

As mentioned earlier, the main challenge in our current framework arises from the fact that $(\mathbb{E}L - \mathcal{S}(M) - z\Lambda)^{-1}$ is *not* a resolvent. Consequently, the MDE does not directly inherit the desirable properties of the resolvent, such as simple bounds that only depend on the spectral parameter z. For instance, it is not immediately evident whether $\mathbb{E}L - \mathcal{S}(M) - z\Lambda$ is even invertible. The following lemma resolves this issue.

Lemma A.1. *Let $\tau \in \mathbb{R}_{\geq 0}$ and $W \in \mathscr{A}$. Then, $\mathbb{E}L - \mathcal{S}(W) - z\Lambda - i\tau I_\ell$ is invertible.*

Proof. Let $W \in \mathscr{A}$ be arbitrary. By definition of $\mathscr{A}$, $\Im[W] \succeq 0$. Since $\mathcal{S}$ is positivity-preserving, $\Im[\mathcal{S}(W)] \succeq 0$. If $\tau > 0$, it follows directly that $\Im[\mathbb{E}L - \mathcal{S}(W) - z\Lambda - i\tau I_\ell] \preceq -\tau$, which implies that $\mathbb{E}L - \mathcal{S}(W) - z\Lambda - i\tau I_\ell$ is non-singular.

Assume that $\tau = 0$ and let $v^* = (v_1^*, v_2^*)$ with $v_1 \in \mathbb{C}^n$ and $v_2 \in \mathbb{C}^d$ be a unitary vector in the kernel of $\mathbb{E}L - \mathcal{S}(W) - z\Lambda$. We will show that $v = 0$ and conclude that the kernel of $\mathbb{E}L - \mathcal{S}(W) - z\Lambda$ is trivial. Decomposing $\mathbb{E}L - \mathcal{S}(W) - z\Lambda$ into its real and imaginary parts, we have

$$0 = v^*(\mathbb{E}L - \mathcal{S}(W) - z\Lambda)v = v^*\Re[\mathbb{E}L - \mathcal{S}(W) - z\Lambda]v + iv^*\Im[\mathbb{E}L - \mathcal{S}(W) - z\Lambda]v.$$

Since both $\Re[\mathbb{E}L - \mathcal{S}(W) - z\Lambda]$ and $\Im[\mathbb{E}L - \mathcal{S}(W) - z\Lambda]$ are Hermitian, the quadratic forms are real and $v^*\Re[\mathbb{E}L - \mathcal{S}(W) - z\Lambda]v = v^*\Im[\mathbb{E}L - \mathcal{S}(W) - z\Lambda]v = 0$. By definition of the admissible set, $\Im[\mathcal{S}_{1,1}(W)] \succeq 0$ which implies that the imaginary part of the upper-left $n \times n$ block of $\mathbb{E}L - \mathcal{S}(W) - z\Lambda$ is negative definite. Consequently, it must be the case that $v_1 = 0$.

Returning to the equation $(\mathbb{E}L - \mathcal{S}(W) - z\Lambda)\, v = 0$, we have in particular that $(Q - \mathcal{S}_{2,2}(W))\, v_2 = 0$. Left-multiplying by v_2^* and decomposing the matrix $Q - \mathcal{S}_{2,2}(W)$ into its real and imaginary parts,

$$0 = v_2^*\Re[Q - \mathcal{S}_{2,2}(W)]v_2 + iv_2^*\Im[Q - \mathcal{S}_{2,2}(W)]v_2.$$

Again, since the real and imaginary parts of a matrix are Hermitian, the quadratic forms are real and $v_2^*\mathcal{S}_{2,2}(\Im[W])v_2 = -v_2^*\Im[Q - \mathcal{S}_{2,2}(W)]v_2 = 0$. By definition of $\mathcal{S}_{2,2}$, $v_2^*\mathcal{S}_{2,2}(\Im[W])v_2 = \mathbb{E}v_2^*(B - \mathbb{E}B)\Im[W_{1,1}](B - \mathbb{E}B)^T v_2$. Since $W \in \mathscr{A}$, $\Im[W_{1,1}] \succ 0$ and $(B - \mathbb{E}B)^T v_2 = 0$ almost surely. Going back to the equation $(Q - \mathcal{S}_{2,2}(W))\, v_2 = 0$, we obtain $Qv_2 = 0$. However, Q is non-singular so $v_2 = 0$. □

For every $W \in \mathscr{A}$, $\tau \in \mathbb{R}_{\geq 0}$ and $z \in \mathbb{H}$, the Schur complement of the lower-right $d \times d$ block of the matrix $\mathbb{E}L - \mathcal{S}(W) - z\Lambda - i\tau I_\ell$, $\mathbb{E}A - \mathcal{S}_{1,1}(W) - (z + i\tau)I_n$, has negative definite imaginary part. Hence, the Schur complement is non-singular. By [LS02, Theorem 2.1], this implies that the matrix $\mathbb{E}L - \mathcal{S}(W) - z\Lambda - i\tau I_\ell$ is invertible if and only if its lower-right $d \times d$ block is. Since we established non-singularity of the full matrix, we obtain the following corollary.

Corollary A.1. *Let $\tau \in \mathbb{R}_{\geq 0}$ and $W \in \mathscr{A}$. Then, the diagonal blocks of $(\mathbb{E}L - \mathcal{S}(W) - z\Lambda - i\tau I_\ell)^{-1}$ are invertible.*

This lemma provides insight into the invertibility of $\mathbb{E}L - \mathcal{S}(M) - z\Lambda$ and establishes the foundation for considering the MDE (22) as a fixed point equation $\mathcal{F}(M) = M$, along with its regularized counterpart. This perspective allows us to explore the existence and uniqueness of solutions by leveraging the extensive theory on fixed points. A first step in this direction is showing that $\mathcal{F}$ and $\mathcal{F}^{(\tau)}$ both map their respective domains to themselves. We adapt the argument from [HFS07].

Lemma A.2. *Let $\tau \in \mathbb{R}_{\geq 0}$, $z \in \mathbb{H}$ and $W \in \mathscr{A}$. Then,*

$$\Im[\mathcal{F}^{(\tau)}(W)] \succeq \tau\mathcal{F}^{(\tau)}(W)(\mathcal{F}^{(\tau)}(W))^*, \quad \Im[\mathcal{F}_{1,1}^{(\tau)}(W)] \succeq \Im[z]\mathcal{F}_{1,1}^{(\tau)}(W)(\mathcal{F}_{1,1}^{(\tau)}(W))^*$$

and $\|\mathcal{F}_{1,1}^{(\tau)}(W)\| \leq (\Im[z])^{-1}$. Furthermore, if $\tau > 0$, then $\|\mathcal{F}^{(\tau)}(W)\| \leq \tau^{-1}$.

Proof. By Lemma C.6,

$$\begin{aligned}\Im[\mathcal{F}^{(\tau)}(W)] &= \mathcal{F}^{(\tau)}(W)\Im[z\Lambda + i\tau I_\ell + \mathcal{S}(W) - \mathbb{E}L](\mathcal{F}^{(\tau)}(W))^* \\ &= \mathcal{F}^{(\tau)}(W)\Im[z\Lambda + i\tau I_\ell + \mathcal{S}(W)](\mathcal{F}^{(\tau)}(W))^*\end{aligned}$$

Since $\Im[W] \succeq 0$ and the super-operator is positivity-preserving, $\Im[z\Lambda + i\tau I_\ell + \mathcal{S}(W) - \mathbb{E}L] \succeq \Im[z\Lambda + i\tau I_\ell]$. In particular,

$$\Im[\mathcal{F}^{(\tau)}(W)] \succeq \mathcal{F}^{(\tau)}(W)\Im[z\Lambda + i\tau I_\ell](\mathcal{F}^{(\tau)}(W))^* \succeq \Im[z]\mathcal{F}^{(\tau)}(W)\Lambda(\mathcal{F}^{(\tau)}(W))^*.$$

Taking the upper-left $n \times n$ block in the equation above,

$$\Im[\mathcal{F}_{1,1}^{(\tau)}(W)] \succeq \Im[z]\mathcal{F}_{1,1}^{(\tau)}(W)(\mathcal{F}_{1,1}^{(\tau)}(W))^* \succ 0.$$

Since the (operator) norm preserve the Loewner ordering of positive semidefinite matrices, it follows from Lemma C.5 that

$$\|\mathcal{F}_{1,1}^{(\tau)}(W)\| \geq \|\Im[\mathcal{F}_{1,1}^{(\tau)}(W)]\| \geq \Im[z]\|\mathcal{F}_{1,1}^{(\tau)}(W)\|^2$$

Rearranging, we get the bound on $\|\mathcal{F}_{1,1}^{(\tau)}(W)\|$. We obtain the bound $\|\mathcal{F}^{(\tau)}(W)\| \leq \tau^{-1}$ using an analogous argument. ☐

The definitions of $\mathcal{F}$ and $\mathcal{F}^{(\tau)}$ in (24), (25) and the choice of sets in (21),(23) is supported by Lemmas A.1 and A.2. The lemma demonstrates that both maps map their respective domains to themselves.[7] This observation leads us to consider the possibility of proving the existence of a solution to the MDE by arguing that $\mathcal{F}$ is a contraction. However, it is important to note that Lemma A.2 also reveals a weaker control of the MDE in comparison to the RMDE. Specifically, we only have a norm bound on the upper-left $n \times n$ block of the map $\mathcal{F}$.

Nevertheless, we will leverage the favorable properties of the RMDE to establish the existence and uniqueness of a solution to (22). Subsequently, we will argue that the unique solution of the RMDE (22) converges to a solution of the MDE (19) as τ approaches zero. To accomplish this, a crucial step will be to lower bound the smallest singular value of $\Im[M_{1,1}^{(\tau)}]$ uniformly in τ, ensuring that $\lim_{\tau\downarrow 0}\Im[M_{1,1}^{(\tau)}] \succ 0$. This bound on the smallest eigenvalue of the imaginary part of $M_{1,1}^{(\tau)}$, combined with Lemma A.2, plays a vital role in controlling $\|M^{(\tau)}\|$ as $\tau \downarrow 0$.

In what follows, we will take advantage of the block structure of the (R)MDE. Using the Schur complement formula, we decompose (19) as

$$M_{1,1} = (\mathbb{E}A - (\mathbb{E}B^T - \mathcal{S}_{1,2}(M))C^{-1}(\mathbb{E}B - \mathcal{S}_{2,1}(M)) - \mathcal{S}_{1,1}(M) - zI_n)^{-1}, \tag{27a}$$

$$M_{2,2} = ((\mathbb{E}B - \mathcal{S}_{2,1}(M))\left(zI_n + \mathcal{S}_{1,1}(M) - \mathbb{E}A\right)^{-1}(\mathbb{E}B^T - \mathcal{S}_{1,2}(M)) - \mathcal{S}_{2,2}(M) + Q)^{-1}, \tag{27b}$$

$$M_{1,2} = -\mathcal{F}_{1,1}(M)(\mathbb{E}B^T - \mathcal{S}_{1,2}(M))C^{-1} \text{ and} \tag{27c}$$

$$M_{2,1} = -C^{-1}(\mathbb{E}B - \mathcal{S}_{2,1}(M))\mathcal{F}_{1,1}(M) \tag{27d}$$

where $C = Q - \mathcal{S}_{2,2}(M)$. It may sometimes be practical to work with the equivalent form

$$M_{2,2} = C^{-1} + C^{-1}(\mathbb{E}B - \mathcal{S}_{2,1}(M))\mathcal{F}_{1,1}(M)(\mathbb{E}B^T - \mathcal{S}_{1,2}(M))C^{-1}. \tag{27e}$$

We may decompose (22) similarly. In this case, we will write $C^{(\tau)} = Q - \mathcal{S}_{2,2}(M^{(\tau)}) - i\tau I_d$.

[7] To be more precise, we also need $\mathcal{F}$ and $\mathcal{F}^{(\tau)}$ to be holomorphic functions. However, this is clear from definition.

A.2.2. Large spectral parameter

Considering that Lemma A.2 ensures that any solution M to (19) must satisfy $M_{1,1}(z) \leq (\Im[z])^{-1}$ for every $z \in \mathbb{H}$, it is natural to consider the behavior of $M(z)$ for $z \in \mathbb{H}$ with $\Im[z]$ in a neighborhood of infinity. The following lemma, which is reminiscent of a condition for the Nevanlinna representation theorem, demonstrates that $\Im[M(z)]$ converges to fixed deterministic quantity as $\Im[z]$ increases.

Lemma A.3. *Fix $\tau \in \mathbb{R}_{\geq 0}$ and assume that $M \in \mathscr{M}$ such that, for all $z \in \mathbb{H}$, $M(z)$ solves the RMDE (22). Then, $\|M(z) - M_\infty\| \to 0$ as $\Im[z] \to \infty$.*

Proof. We proceed block-wise. By Lemma A.2, it is clear that $\|\mathcal{F}_{1,1}^{(\tau)}(M(z))\| \to 0$ as $\Im[z] \to \infty$. By Lemma C.6 and the fact that the super-operator is positivity-preserving,

$$\Im[(\mathbb{E}A - zI_n - \mathcal{S}_{1,1}(M(z)))^{-1}] \succeq \Im[z]\,(zI_n + \mathcal{S}_{1,1}(M(z)) - \mathbb{E}A)^{-1}\,(zI_n + \mathcal{S}_{1,1}(M(z)) - \mathbb{E}A)^{-*}$$

which implies that $\|(zI_n + \mathcal{S}_{1,1}(M(z)) - \mathbb{E}A)^{-1}\| \leq (\Im[z])^{-1} \to 0$ as $\Im[z] \to \infty$. By flatness and definition of the superoperator, $\max\{\|\mathcal{S}_{1,2}(M)\|, \|\mathcal{S}_{2,1}(M)\|, \|\mathcal{S}_{2,2}(M)\|\} \leq s\|M_{1,1}\| \to 0$ as $\Im[z] \to \infty$. Hence,

$$\|(\mathbb{E}B - \mathcal{S}_{2,1}(M))\,(zI_n + \mathcal{S}_{1,1}(M) - \mathbb{E}A)^{-1}\,(\mathbb{E}B^T - \mathcal{S}_{1,2}(M)) - \mathcal{S}_{2,2}(M)\| \xrightarrow{\Im[z]\uparrow\infty} 0$$

and, using the block-wise decomposition of $M_{2,2}$ in (27b) as well as the definition of $M_*^{(\tau)}$ in (26), $\|M_{2,2}(z) - M_*^{(\tau)}\| \to 0$ as $\Im[z] \to \infty$.

Finally, using the decomposition in (27c), (27d) and the fact that $||(\mathcal{S}_{2,2}(M(z)) + i\tau I_d - Q)^{-1}||$ is bounded as $\Im[z]$ increases, it follows that both $||M_{1,2}(z)||$ and $||M_{2,1}(z)||$ vanish as $\Im[z]$ approaches infinity. □

The purpose of Lemma A.3 is to play an intermediate role in establishing the existence and uniqueness of solutions. However, our ultimate goal is to characterize the underlying real spectral measure that is encoded in the pseudo-resolvent.

While the pseudo-resolvent and the (R)MDE exhibit favorable behavior when the spectral parameter is far away from the spectrum, it is important to note that the spectral information is primarily contained in the poles located at the eigenvalues of the underlying matrix. Therefore, in order to extract meaningful information about the spectrum of the pseudo-resolvent, we need to bring the spectral parameter close to the real line.

To accomplish this, the next lemma constructs a loose bound on the norm of any solution to (22), which holds for every spectral parameter large enough in norm regardless of the magnitude of its imaginary part. This bound allows us to explore the behavior of the solution for large spectral values without being restricted by its imaginary part.

Lemma A.4. *Fix $\tau \in \mathbb{R}_{\geq 0}$ and assume that $M \in \mathscr{M}$ such that, for all $z \in \mathbb{H}$, $M(z)$ solves the RMDE (22). Then, there exists some constant $c \in \mathbb{R}_{>0}$ such that $\|M(z) - M_\infty^{(\tau)}\| \leq c(|z| - \kappa)^{-1}$ for all $z \in \{z \in \mathbb{H} : |z| > \kappa + c\kappa^{-1}\}$ with $\kappa := 2\|(Q - i\tau I_d)^{-1}\|(||\mathbb{E}B|| + (2\|(Q - i\tau I_d)^{-1}\|)^{-1})^2 + ||\mathbb{E}A|| + (2\|(Q - i\tau I_d)^{-1}\|)^{-1} + s||M_\infty^{(\tau)}||$.*

Proof. Fix $z \in \mathbb{H}$ with $|z| > \kappa$ and let $M \equiv M(z)$. For clarity, we denote $a = ||\mathbb{E}A||$, $b = ||\mathbb{E}B||$, $m = ||M_\infty^{(\tau)}||$ and $q = \|(Q - i\tau I_d)^{-1}\|$. We will show that $\|M(z) - M_\infty^{(\tau)}\| \notin (c(|z| - \kappa)^{-1}, \kappa]$ for every $z \in \mathbb{H}$ with $|z| > \kappa + c\kappa^{-1}$. By Lemma A.3, $\|M(z) - M_\infty^{(\tau)}\|$ is in a neighborhood of the origin for every $z \in \mathbb{H}$ with $\Im[z]$ large enough. Since $z \mapsto \|M(z) - M_\infty^{(\tau)}\|$ is a continuous function on $\{z \in \mathbb{H} : |z| > \kappa + c\kappa^{-1}\}$, this will imply that $\|M(\{z \in \mathbb{H} : |z| > \kappa + c\kappa^{-1}\}) - M_\infty^{(\tau)}\| \subseteq [0, c(|z| - \kappa)^{-1}]$. Suppose that $||M - M_\infty^{(\tau)}|| \leq (2sq)^{-1}$ such that $\|M\| \leq \|M - M_\infty^{(\tau)}\| + \|M_\infty^{(\tau)}\| \leq (2sq)^{-1} + m$. Again, we consider the blocks separately using (27).

By definiftion of $\mathcal{S}_{2,2}$, $\|\mathcal{S}_{2,2}(M)\| \leq s\|M_{1,1}\| \leq (2q)^{-1}$. It follows from Lemma C.2 that

$$\| \left(\mathcal{S}_{2,2}(M) + i\tau I_d - Q\right)^{-1} \| \leq \|(Q - i\tau I_d)^{-1} \left(\mathcal{S}_{2,2}(M)(Q - i\tau I_d)^{-1} - I_d\right)^{-1} \| \leq 2q.$$

Furthermore, by subadditivity of the spectral norm,

$$\begin{aligned} &\|(\mathbb{E}B^T + \mathcal{S}_{1,2}(M)) \left(\mathcal{S}_{2,2}(M) + i\tau I_d - Q\right)^{-1} (\mathbb{E}B + \mathcal{S}_{2,1}(M)) + \mathbb{E}A - \mathcal{S}_{1,1}(M)\| \\ &\qquad \leq 2q(b + (2q)^{-1})^2 + a + (2q)^{-1} + sm = \kappa < |z|. \end{aligned}$$

By (27a) and Lemma C.2, $||M_{1,1}|| \leq (|z| - \kappa)^{-1}$. We now turn our attention to $M_{2,2}$. Using Lemma C.1 and equation (27b),

$$\begin{aligned} M_{2,2} - M_\star^{(\tau)} &= M_{2,2}\mathcal{S}_{2,2}(M)M_\star^{(\tau)} \\ &\quad - M_{2,2}(\mathbb{E}B + \mathcal{S}_{2,1}(M)) \left((z + i\tau)I_n + \mathcal{S}_{1,1}(M) - \mathbb{E}A\right)^{-1} (\mathbb{E}B^T + \mathcal{S}_{1,2}(M))M_\star^{(\tau)} \end{aligned}$$

By Lemma C.2, $\|((z + i\tau)I_n + \mathcal{S}_{1,1}(M) - \mathbb{E}A)^{-1}\| \leq (|z| - (2q)^{-1} - sm - a)^{-1}$. Thus,

$$\begin{aligned} \|M_{2,2} - M_\star^{(\tau)}\| &\leq m((2sq)^{-1} + m)s\|M_{1,1}\| \\ &\quad + m((2sq)^{-1} + m)(b + (2q)^{-1})^2 \left(|z| - (2q)^{-1} - sm - a\right)^{-1}. \end{aligned}$$

Plugging the bound for $\|M_{1,1}\|$ derived above and simplifying,

$$\|M_{2,2} - M_\star^{(\tau)}\| \leq m(s + (b + (2q)^{-1})^2)((2sq)^{-1} + m)(|z| - \kappa)^{-1}.$$

It only remains to treat $||M_{1,2}||$ and $||M_{2,1}||$, which we directly bound by

$$\max\{\|M_{1,2}\|, \|M_{2,1}\|\} \leq 2(b + (2q)^{-1})q\|M_{1,1}\| \leq 2(b + (2q)^{-1})q(|z| - \kappa)^{-1}$$

using (27c),(27d).

To summarize, we showed that for every $z \in \mathbb{H}$ with $|z| > sm + (2q)^{-1} + a + 2q(b + (2q)^{-1})^2$, $||M(z) - M_\infty^{(\tau)}|| \leq (2sq)^{-1}$ implies that

$$\|M(z) - M_\infty^{(\tau)}\| \leq \|M_{1,1}(z)\| + \|M_{1,2}\| + \|M_{2,1}\| + \|M_{2,2} - M_\star^{(\tau)}\| \leq c(|z| - \kappa)^{-1}$$

with $c := 1 + m(s + (b + (2q)^{-1})^2)((2sq)^{-1} + m) + 4(b + (2q)^{-1})q$. Choosing $|z| > \kappa + c\kappa^{-1}$ completes the proof. □

The proof of Lemma A.4 provides important bounds which are not explicitly stated in the statement of the lemma. We continue our treatment of the RMDE by upper-bounding the imaginary part of any solution to the RMDE when the spectral parameter is far away from the support.

Lemma A.5. *Fix $\tau \in \mathbb{R}_{\geq 0}$ and assume that $M \in \mathscr{M}$ such that, for all $z \in \mathbb{H}$, $M(z)$ solves the RMDE (22). Let κ and c be defined as in Lemma A.4. Then, there exists $\kappa_+ \geq \kappa + c\kappa^{-1}$ and constant $c_* \in \mathbb{R}_{>0}$ such that $\|\Im[M_{1,1}(z)]\| \leq c_*(|z| - \kappa)^{-2}(\tau + \Im[z])$ for every $z \in \{z \in \mathbb{H} : |z| \geq \kappa_+\}$. In particular, if $\tau = 0$, then $\|\Im[M(z)]\|$ converges uniformly to 0 as $\Im[z] \downarrow 0$ on $\{z \in \mathbb{H} : \Re[z] \geq \kappa_+\}$*

Proof. Let $z \in \mathbb{H}$ with $|z| > \kappa + c\kappa^{-1}$ and denote $M = M(z)$. Furthermore, let $m \equiv m(z) = c(|z|-\kappa)^{-1}$ be the bound in Lemma A.4. By Lemma C.6,(25), $\Im[M] = M\left(\Im[z]\Lambda + \tau I_\ell + \mathcal{S}(\Im[M])\right)M^*$. Decomposing block-wise and recalling that $M_{i,j}^* = (M_{i,j})^*$ denotes the conjugate transpose of the (i,j) sub-block, we get

$$\begin{aligned}\Im[M_{1,1}] = M_{1,1}\left((\Im[z]+\tau)I_n + \mathcal{S}_{1,1}(\Im[M])\right)M_{1,1}^* + M_{1,1}\mathcal{S}_{1,2}(\Im[M])M_{1,2}^* \\ + M_{1,2}\mathcal{S}_{2,1}(\Im[M])M_{1,1}^* + M_{1,2}\left(\tau I_n + \mathcal{S}_{2,2}(\Im[M])\right)M_{1,2}^*,\end{aligned}$$

$$\begin{aligned}\Im[M_{2,1}] = M_{2,1}\left((\Im[z]+\tau)I_n + \mathcal{S}_{1,1}(\Im[M])\right)M_{1,1}^* + M_{2,1}\mathcal{S}_{1,2}(\Im[M])M_{1,2}^* \\ + M_{2,2}\mathcal{S}_{2,1}(\Im[M])M_{1,1}^* + M_{2,2}\left(\tau I_n + \mathcal{S}_{2,2}(\Im[M])\right)M_{2,2}^*,\end{aligned}$$

$$\begin{aligned}\Im[M_{1,2}] = (\Im[z]+\tau)M_{1,1}M_{2,1}^\star + M_{1,1}\mathcal{S}_{1,1}(\Im[M])M_{2,1}^* + M_{1,1}\mathcal{S}_{1,2}(\Im[M])M_{2,2}^* \\ + M_{1,2}\mathcal{S}_{2,1}(\Im[M])M_{2,1}^* + \tau M_{1,2}M_{2,1}^* + M_{1,2}\mathcal{S}_{2,2}(\Im[M])M_{2,2}^*\end{aligned}$$

and

$$\begin{aligned}\Im[M_{2,2}] = (\Im[z]+\tau)M_{2,1}M_{2,1}^\star + M_{2,1}\mathcal{S}_{1,1}(\Im[M])M_{2,1}^* + M_{2,1}\mathcal{S}_{1,2}(\Im[M])M_{2,2}^* \\ + M_{2,2}\mathcal{S}_{2,1}(\Im[M])M_{2,1}^* + \tau M_{2,2}M_{2,1}^* + M_{2,2}\mathcal{S}_{2,2}(\Im[M])M_{2,2}^*.\end{aligned}$$

By Lemma A.4, $\|M_{1,1}\| \vee \|M_{1,2}\| \vee \|M_{2,1}\| \le m$ and $\|M\| \le \|M - M_\infty\| + \|M_\infty\| \le m + \|M_\infty\|$. For simplicity, we let $m_\infty = \|M_\infty\|$. If we take the norm of the expansion above, we may bound every term $\|M_{1,1}\|$, $\|M_{1,2}\|$ and $\|M_{2,1}\|$ by m and $\|M_{2,2}\|$ by $m + m_\infty$. Let $x = \|\mathcal{S}_{1,2}(\Im[M])\| \vee \|\mathcal{S}_{2,1}(\Im[M])\| \vee \|\mathcal{S}_{2,2}(\Im[M])\|$. Grouping the terms in the expansion for $\Im[M_{1,2}]$, $\Im[M_{2,1}]$ and $\Im[M_{2,2}]$ and bounding, in norm, by the worst case bound, we get

$$\begin{aligned}x \le m^2\Im[z] + 2(m+m_\infty)^2\tau + m^2\|\mathcal{S}_{1,1}(\Im[M])\| \\ + (m+m_\infty)^2\left(\|\mathcal{S}_{1,2}(\Im[M])\| + \|\mathcal{S}_{2,1}(\Im[M])\| + \|\mathcal{S}_{2,2}(\Im[M])\|\right).\end{aligned}$$

By flatness of the super-operator and the definition of $\mathcal{S}$,

$$\|\mathcal{S}_{1,2}(\Im[M])\| + \|\mathcal{S}_{2,1}(\Im[M])\| + \|\mathcal{S}_{2,2}(\Im[M])\| \le 3s\|\Im[M_{1,1}]\|$$

and $\|\mathcal{S}_{1,1}(\Im[M])\| \le s\|\Im[M_{1,1}]\| + 3sx$. Plugging this back into the inequality above and rearranging,

$$(1-3sm^2)x \le m^2\Im[z] + 2(m+m_\infty)^2\tau + 4s(m+m_\infty)^2\|\Im[M_{1,1}]\|.$$

By Lemma A.4, $m \to 0$ as $|z| \to \infty$. Hence, letting $\kappa_+ \ge \kappa + c\kappa^{-1}$ such that $3sm^2 < 1$ for all $|z| \ge \kappa_+$,

$$x \le \underbrace{\left(\frac{m^2}{1-3sm^2}\right)}_{c_z}\Im[z] + \underbrace{\left(\frac{2(m+m_\infty)^2}{1-3sm^2}\right)}_{c_\tau}\tau + \underbrace{\left(\frac{4s(m+m_\infty)^2}{1-3sm^2}\right)}_{c_M}\|\Im[M_{1,1}]\|.$$

Taking the norm of $\Im[M_{1,1}]$ using similar bounds,

$$\|\Im[M_{1,1}]\| \le m^2\Im[z] + 2m^2\tau + 6sm^2x + sm^2\|\Im[M_{1,1}]\|.$$

Using the inequality for x and grouping terms,

$$\|\Im[M_{1,1}]\| \leq (6c_z s + 1)m^2\Im[z] + (6c_\tau s + 2)m^2\tau + (6c_M s + s)m^2\|\Im[M_{1,1}]\|.$$

Increasing κ_+ such that $(6c_M + 1)sm^2 < 1$ for every $|z| \geq \kappa_+$, we obtain

$$\|\Im[M_{1,1}]\| \leq \frac{6c_z s + 1}{1 - (6c_M + 1)sm^2}m^2\Im[z] + \frac{6c_\tau s + 2}{1 - (6c_M + 1)sm^2}m^2\tau.$$

We conclude by using the definition of m. □

A.2.3. Stieltjes transform representation

Similarly to the classical Nevanlinna representation theorem for scalar function mapping the upper-half complex plane $\mathbb{H}$ to itself, we may represent any solution to the (R)MDE using a matrix Nevanlinna-Riesz-Herglotz representation. Doing so, we are going to use Lemma A.5 to show that the probability measure in such representation has bounded support when $\tau = 0$.

Lemma A.6 (Nevanlinna-Riesz-Herglotz representation)**.** *Assume that $M \in \mathscr{M}$ such that, for all $z \in \mathbb{H}$, $M(z)$ solves the MDE (19). Then,*

$$M(z) = M_\infty + \int_{\mathbb{R}} \frac{\Omega(d\lambda)}{\lambda - z}$$

for all $z \in \mathbb{H}$, where Ω is a real Borel $\ell \times \ell$ positive semidefinite measure satisfying $\int_{\mathbb{R}} \frac{v^*\Omega(d\lambda)v}{1+\lambda^2} < \infty$ for all $v \in \mathbb{C}^\ell$. Furthermore, $\Omega_{1,1}$ is compactly supported with $\operatorname{supp}(\Omega) \subseteq \operatorname{supp}(\Omega_{1,1})$ *and*

$$\int_{\mathbb{R}} \Omega(d\lambda) = \begin{bmatrix} I_n & -\mathbb{E}[B^T]Q^{-1} \\ -Q^{-1}\mathbb{E}[B] & Q^{-1}\mathbb{E}[BB^T]Q^{-1}. \end{bmatrix}.$$

Proof. Since M_∞ is real, the matrix-valued function $z \in \mathbb{H} \mapsto M(z) - M_\infty \in \mathscr{A}$ is a matrix-valued Herglotz function [GT97, Definition 5.2]. In particular, for every $v \in \mathbb{C}^\ell$, $z \in \mathbb{H} \mapsto q(z) := v^*\left(M(z) - M_\infty\right)v$ is a scalar Herglotz function. Recall the notation $C = Q - \mathcal{S}_{2,2}(M)$ in (27). Using (27a) and rearranging $M_{1,1}(\mathbb{E}A - (\mathbb{E}B^T - \mathcal{S}_{1,2}(M))C^{-1}(\mathbb{E}B - \mathcal{S}_{2,1}(M)) - \mathcal{S}_{1,1}(M) - zI_n) = I_n$, we obtain

$$zM_{1,1} = M_{1,1}\left(\mathbb{E}A - (\mathbb{E}B^T - \mathcal{S}_{1,2}(M))C^{-1}(\mathbb{E}B - \mathcal{S}_{2,1}(M)) - \mathcal{S}_{1,1}(M)\right) - I_n.$$

Taking the limit as the imaginary part of the spectral parameter goes to infinity using Lemma A.3, we get that $\lim_{\Im[z]\uparrow\infty} zM_{1,1} = -I_n$. By (27c), (27d) along with our flatness assumption, $\lim_{\Im[z]\uparrow\infty} zM_{1,2} = \mathbb{E}[B^T]Q^{-1}$ and $\lim_{\Im[z]\uparrow\infty} zM_{2,1} = Q^{-1}\mathbb{E}[B]$. Furthermore, by the resolvent trick, $C^{-1} - M_\star = C^{-1}\mathcal{S}_{2,2}(M)M_\star$. Therefore, by (20), (27e),

$$\lim_{\Im[z]\uparrow\infty} z(M_{2,2} - M_\star) = -Q^{-1}\mathcal{S}_{2,2}(I_\ell)Q^{-1} - Q^{-1}\mathbb{E}[B]\mathbb{E}[B^T]Q^{-1} = -Q^{-1}\mathbb{E}[BB^T]Q^{-1}.$$

By Nevanlinna-Riesz-Herglotz representation theorem [GT97, Part (iii) of Theorem 2.2 along with Part (iii) of Theorem 2.3] and Lemma A.3, there exists a Borel measure ω on $\mathbb{R}$ satisfying

$$\int_{\mathbb{R}} (1 + \lambda^2)^{-1}\omega(d\lambda) < \infty \quad \text{and} \quad v^*\left(M(z) - M_\infty\right)v = \int_{\mathbb{R}} \frac{\omega(d\lambda)}{\lambda - z}$$

for every $z \in \mathbb{H}$. Furthermore, the measure ω is finite with

$$\int_{\mathbb{R}} \omega(d\lambda) = v^* \begin{bmatrix} I_n & -\mathbb{E}[B^T]Q^{-1} \\ -Q^{-1}\mathbb{E}[B] & Q^{-1}\mathbb{E}[BB^T]Q^{-1}. \end{bmatrix} v.$$

Since this is true for all $v \in \mathbb{C}^\ell$, it follows that $M(z) = M_\infty + \int_{\mathbb{R}} \frac{\Omega(d\lambda)}{\lambda - z}$ for all $z \in \mathbb{H}$ where Ω is a real Borel $\ell \times \ell$ positive semidefinite measure satisfying $\int_{\mathbb{R}} \frac{v^*\Omega(d\lambda)v}{1+\lambda^2} < \infty$ for all $v \in \mathbb{C}^\ell$ and with the given normalization.

By Lemma A.5, there exists $\kappa_+ \in \mathbb{R}_{>0}$ such that $\|\Im[M]\|$ converges uniformly to 0 on as $\Im[z]$ approaches 0 on $\{z \in \mathbb{H} : \Re[z] \geq \kappa_+\}$. By the Stieltjes inversion formula for Ω [GT97, Theorem 5.4],

$$\pi^{-1} \lim_{\epsilon \downarrow 0} \int_{\lambda_1}^{\lambda_2} \Im[M(\lambda + i\epsilon)]d\lambda = \Omega\left((\lambda_1, \lambda_2)\right) + 2^{-1}\Omega\left(\{\lambda_1\}\right) + 2^{-1}\Omega\left(\{\lambda_2\}\right).$$

Thus, the fact that Ω is compactly supported is a consequence of uniform convergence of $\|\Im[M](z)\|$ to 0 as $\Im[z] \downarrow 0$ with $\Re[z] \geq \kappa_+$. We may also observe from the proof of Lemma A.5 that $\Im[M] = 0$ whenever $\Im[M_{1,1}] = 0$, implying that $\mathrm{supp}(\Omega) \subseteq \mathrm{supp}(\Omega_{1,1})$. □

We can interpret Lemma A.6 as a matrix-valued Stieltjes transform. Hence, we will refer to it using this terminology. Furthermore, given the normalization of Ω in Lemma A.6, we say that $\Omega_{1,1}$ is a matrix-valued probability measure in the sense that $v_*\Omega_{1,1}v$ is a real Borel measure satisfying $\int_{\mathbb{R}} v_*\Omega_{1,1}(d\lambda)v = 1$ for every $v \in \mathbb{C}^n$.

Lemma A.6 also provides an explicit bound on the solution to (19). Indeed, if $M \in \mathscr{M}$ solves (19) for every $z \in \mathbb{H}$, then

$$\|M(z)\| \leq \|M_\infty\| + \mathrm{dist}(z, \mathrm{supp}(\Omega))^{-1} \left\| \int_{\mathbb{R}} \Omega(d\lambda) \right\| \tag{28}$$

on $\mathbb{H}$.

When considering the regularized matrix Dyson equation (22) and the solution $M^{(\tau)}(z)$, we encounter a challenge in directly applying the same procedure to obtain a bound on $\|M^{(\tau)}(z)\|$. The issue arises from the fact that M_∞ has a positive semidefinite imaginary part, which implies that the function $z \mapsto M^{(\tau)}(z) - M_\infty$ may not be a Herglotz function. One potential alternative approach is to utilize a multivariate Herglotz representation, as discussed in [LN17]. This representation provides an integral representation for the function $(z, i\tau) \mapsto M^{(\tau)}(z)$ involving a multivariate measure. However, it should be noted that in such representations, the measure cannot be finite unless it is trivial. Nonetheless, an analogue of Lemma A.6 holds for the upper-left $n \times n$ block of the solution to the RMDE. The result is obtained via a similar argument, so we omit the proof.

Lemma A.7. *Fix $\tau \in \mathbb{R}_{\geq 0}$ and assume that $M \in \mathscr{M}$ such that, for all $z \in \mathbb{H}$, $M(z)$ solves the RMDE (22). Then, $M_{1,1}(z) = \int_{\mathbb{R}} \frac{\Omega_{1,1}(d\lambda)}{\lambda - z}$ for all $z \in \mathbb{H}$, where $\Omega_{1,1}$ is a real Borel $n \times n$ positive semidefinite measure satisfying $\int_{\mathbb{R}} \Omega_{1,1}(d\lambda) = I_n$.*

Aside from their inherent value as results, Lemmas A.6 and A.7 hold particular significance because they enable us to treat the solution of the MDE as the limit of solutions to the RMDE as τ approaches zero. The key factor in this step is the tightness of the family of measures induced by the Stieltjes representation of RMDE solutions. We present the following result as a corollary, as it can be derived almost directly from a combination of the preceding lemmas.

Corollary A.2. *For every $\tau \in \mathbb{R}_{>0}$, let $M^{(\tau)} \in \mathscr{M}_+$ such that, for all $z \in \mathbb{H}$, $M^{(\tau)}(z)$ solves the RMDE (22). Denote by $\Omega_{1,1}^{(\tau)}$ the positive semidefinite measure in the Nevanlinna-Riesz-Herglotz representation of $M_{1,1}^{(\tau)}$ for all $\tau \in \mathbb{R}_{>0}$. Then, for every $v \in \mathbb{R}^n$ and $\tau_+ \in \mathbb{R}_{>0}$, the family of measures $\{v^*\Omega_{1,1}^{(\tau)}v : \tau \in [0, \tau_+]\}$ is tight.*

Proof. Let $\tau \in \mathbb{R}_{>0}$. By Lemmas A.4 and A.5, there exists $\kappa, c, \kappa_+ \in \mathbb{R}_{>0}$ such that $\|\Im[M_{1,1}(z)]\| \leq c(|z| - \kappa)^{-2}(\tau + \Im[z])$ for every $z \in \mathbb{H}$ with $|z| \geq \kappa_+$. Then,

$$\|\Im[M_{1,1}(\lambda + i\epsilon)]\| \leq c(\sqrt{\lambda^2 + \epsilon^2} - \kappa)^{-2}(\tau + \epsilon) \leq c(\lambda - \kappa)^{-2}(\tau + \epsilon)$$

for every $\lambda > \kappa_+$ and $\epsilon \in [0, 1]$. Here, c is some constant independent of λ and τ.

Hence, for every $\lambda_+ > \kappa_+$, by the Stieltjes inversion formula for $\Omega^{(\tau)}$ [GT97, Theorem 5.4],

$$\begin{aligned} \Omega_{1,1}^{(\tau)}((\lambda_+, \infty)) &\preceq \pi^{-1} \lim_{\epsilon \downarrow 0} \int_{\lambda_+}^{\infty} \Im[M_{1,1}^{(\tau)}(\lambda + i\epsilon)] d\lambda \\ &\preceq \pi^{-1} \lim_{\epsilon \downarrow 0} \int_{\lambda_+}^{\infty} \|\Im[M_{1,1}^{(\tau)}(\lambda + i\epsilon)]\| d\lambda \\ &\preceq c\pi^{-1}\tau \int_{\lambda_+}^{\infty} (\lambda - \kappa)^{-2} d\lambda. \end{aligned}$$

Therefore, if τ is bounded, we may pick $\lambda_+ > \kappa_+$ arbitrarily large to ensure that $\int_{\lambda_+}^{\infty}(\lambda - \kappa)^{-2} d\lambda$ is arbitrarily small. □

A.2.4. Power series representation

As the set of admissible solutions $\mathscr{M}$ consists of analytic matrix-valued functions, any solution to (19) can be expressed as a power series. By employing the Stieltjes transform representation provided in Lemma A.6, we can derive a recurrence relation that determines the coefficients in such an expansion. This recurrence relation will enable us to systematically compute the coefficients of the power series representation of the solution.

Lemma A.8. *Let $M \in \mathscr{M}$ such that, for all $z \in \mathbb{H}$, $M(z)$ solves the MDE (19) and let Ω be the positive semidefinite measure in the Stieltjes transform representation of M. Then, there exists $\lambda_+ > \sup\{|\lambda| \in \operatorname{supp}(\Omega)\}$ such that*

$$M(z) = \sum_{j \in \mathbb{N}} z^{-j} M_j = (\mathbb{E}L - z\Lambda)^{-1} \sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} z^{-k} \mathcal{S}(M_k)(\mathbb{E}L - z\Lambda)^{-1} \right)^j$$

for every $z \in \mathbb{H}$ with $|z| \geq \lambda_+$. Here, $M_0 = M_\infty$ and $M_j = -\int_{\mathbb{R}} \lambda^{j-1}\Omega(d\lambda)$ for every $j \in \mathbb{N}$.

Proof. Since $\operatorname{supp}(\Omega)$ is compact by Lemma A.6, $\sup\{|\lambda| \in \operatorname{supp}(\Omega)\}$ is finite. Let $z \in \mathbb{H}$ with $|z| > \sup\{|\lambda| \in \operatorname{supp}(\Omega)\}$ and write

$$M(z) = M_\infty + \int_{\mathbb{R}} \frac{\Omega(d\lambda)}{\lambda - z} = M_\infty - z^{-1} \int_{\mathbb{R}} \frac{\Omega(d\lambda)}{1 - \lambda/z}.$$

We recognize $(1 - \lambda/z)^{-1}$ as a geometric series and write $(1 - \lambda/z)^{-1} = \sum_{j=0}^{\infty} \frac{\lambda^j}{z^j}$. By Fubini's theorem,

$$\int_{\mathbb{R}} \frac{\Omega(d\lambda)}{1 - \lambda/z} = \sum_{j=0}^{\infty} z^{-j} \int_{\mathbb{R}} \lambda^j \Omega(d\lambda)$$

which implies that

$$M(z) = M_\infty - \sum_{j=0}^{\infty} z^{-j-1} \int_{\mathbb{R}} \lambda^j \Omega(d\lambda).$$

On the other hand, by definition, $M(z)$ solves (19), and we may write

$$M(z) = \mathcal{F}(M(z)) = (\mathbb{E}L - \mathcal{S}(M(z)) - z\Lambda)^{-1}.$$

Using the Schur complement formula, we decompose

$$(\mathbb{E}L - z\Lambda)^{-1} = \begin{bmatrix} R & -R\mathbb{E}[B^T]Q^{-1} \\ -Q^{-1}\mathbb{E}[B]R & Q^{-1} + Q^{-1}\mathbb{E}[B]R\mathbb{E}[B^T]Q^{-1} \end{bmatrix} \tag{29}$$

with $R = \mathcal{R}(z; \mathbb{E}[B^T]Q^{-1}\mathbb{E}[B] - \mathbb{E}[A])$. Since

$$\|R\| = \|\mathcal{R}(z; \mathbb{E}[B^T]Q^{-1}\mathbb{E}[B] - \mathbb{E}[A])\| \leq \operatorname{dist}(z, \sigma\left(\mathbb{E}[A] - \mathbb{E}[B^T]Q^{-1}\mathbb{E}[B]\right))^{-1},$$

we obtain $(\mathbb{E}L - z\Lambda)^{-1} \xrightarrow{|z|\to\infty} M_\infty$. Because M_∞ is non-zero only in its lower-right $d \times d$ block, it follows from Lemma A.4 and the flatness of the super-operator that

$$\|\mathcal{S}(M(z))(\mathbb{E}L - z\Lambda)^{-1}\| \xrightarrow{|z|\to\infty} 0.$$

Let $\lambda_+ > \max\{|\lambda| \in \operatorname{supp}(\Omega)\}$ such that $\|\mathcal{S}(M(z))(\mathbb{E}L - z\Lambda)^{-1}\| < 1$ for all $z \in \mathbb{H}$ with $|z| \geq \lambda_+$. Then, $I_\ell - \mathcal{S}(M(z))(\mathbb{E}L - z\Lambda)^{-1}$ is non-singular with Neumann series

$$\left(I_\ell - \mathcal{S}(M(z))(\mathbb{E}L - z\Lambda)^{-1}\right)^{-1} = \sum_{j=0}^{\infty} \left(\mathcal{S}(M(z))(\mathbb{E}L - z\Lambda)^{-1}\right)^j.$$

In particular,

$$(\mathbb{E}L - \mathcal{S}(M(z)) - z\Lambda)^{-1} = (\mathbb{E}L - z\Lambda)^{-1} \sum_{j=0}^{\infty} \left(\mathcal{S}(M(z))(\mathbb{E}L - z\Lambda)^{-1}\right)^j.$$

We obtain the result by plugging the series expansion for $M(z)$ and using linearity of the super-operator. □

A.3. Existence and Uniqueness

In this subsection, we finally prove the existence and uniqueness of a solution to the MDE (19). As we discussed above, we will first prove existence of a solution to the RMDE for every $\tau > 0$ and use a continuity argument to take vanishing τ.

A.3.1. Solution to the RMDE

For every $\tau > 0$, the existence and uniqueness of a unique $M^{(\tau)} \in \mathscr{M}_+$ satisfying (22) for every $z \in \mathbb{H}$ follows directly from [HFS07]. At a high-level, the proof of existence and uniqueness of a solution to (22) in [HFS07] is based on the fact that the mapping $\mathcal{F}^{(\tau)}$ satisfies the conditions of the Earle-Hamilton fixed-point theorem [EH70], which states that every strict holomorphic function is automatically a contraction with respect to the *Carathéodory-Riffen-Finsler (CRF) pseudometric*. We will discuss and use the CRF-pseudometric in order to show stability in Appendix A.4.2.

For now, we define

$$\mathscr{M}_b := \{f : \{z \in \mathbb{H} : |z| \le b\} \mapsto \mathscr{A}_b \text{ analytic}\} \tag{30}$$

with

$$\mathscr{A}_b := \{W \in \mathbb{C}^{\ell\times\ell} : \Im[W] \succ 0, \|W\| < b\} \cap \mathscr{A}. \tag{31}$$

for every $b > 0$. Indeed, for every $b > 0$, $\mathscr{M}_b$ is a domain in the Banach space of matrix-valued bounded holomorphic functions on $\mathbb{H}$ with the canonical supremum norm. Also, $\mathscr{A}_b$ is a domain in the Banach space of complex symmetric $\ell \times \ell$ matrices with the operator norm. Using the work we did above, we can easily show that $\mathcal{F}^{(\tau)}$ is indeed a strict holomorphic function on $\mathscr{A}_b$ for every $\tau > 0$. The following lemma is a direct adaptation of [HFS07, Proposition 3.2].

Lemma A.9. *Let $z \in \mathbb{H}$, $\tau, b \in \mathbb{R}_{>0}$ and define $m_b := \|\mathbb{E}L\| + (s+1)b + \tau$. Then, for every $W \in \mathscr{A}_b$, $\|\mathcal{F}^{(\tau)}(W)\| \le \tau^{-1}$ and $\Im[\mathcal{F}^{(\tau)}(W)] \succeq \tau m_b^{-2} I_\ell \succ 0$.*[8] *In particular, if $b > \tau^{-1}$, then $\mathcal{F}^{(\tau)}$ maps $\mathscr{A}_b$ strictly into itself.*

Proof. Let $W \in \mathscr{A}_b$. By Lemma A.2, $\|\mathcal{F}^{(\tau)}(W)\| \le \tau^{-1}$ and $\Im[\mathcal{F}^{(\tau)}(W)] \succeq \tau \mathcal{F}^{(\tau)}(W)[\mathcal{F}^{(\tau)}(W)]^*$. Let $v \in \mathbb{C}^\ell$ such that $\|v\| = 1$. By Cauchy-Schwarz inequality,

$$1 = v^*(\mathcal{F}^{(\tau)}(W))^{-1}\mathcal{F}^{(\tau)}(W)v \le \|\mathcal{F}^{(\tau)}(W)v\|\|(\mathcal{F}^{(\tau)}(W))^{-*}v\|$$

which implies that $\|(\mathcal{F}^{(\tau)}(W))^{-1}\|^{-2} \le \|(\mathcal{F}^{(\tau)}(W))^{-*}v\|^{-2} \le \|\mathcal{F}^{(\tau)}(W)v\|^2$. Additionally,

$$\|(\mathcal{F}^{(\tau)}(W))^{-1}\| = \|\mathbb{E}L - \mathcal{S}(W) - z\Lambda - i\tau I_\ell\| \le m_b.$$

Thus, $\Im[\mathcal{F}^{(\tau)}(W)] \succeq \tau m_b^{-2} I_\ell$. □

The existence of a unique solution to the RMDE then follows directly from an application of the Earle-Hamilton fixed-point theorem. Indeed, for every $b \in \mathbb{R}_{>0}$, $\mathcal{F}^{(\tau)}$ has exactly one fixed point on $\mathscr{M}_b$. Since $\mathscr{M}_+ = \bigcup_{b\in\mathbb{R}_{>0}} \mathscr{M}_b$, we obtain the following result.

Lemma A.10 ([HFS07, Theorem 2.1])**.** *There exists a unique solution $M \in \mathscr{M}_+$ such that $M^{(\tau)}(z)$ solves (22) for every $\tau \in \mathbb{R}_{>0}$ and $z \in \mathbb{H}$. Furthermore, for every $W_0 \in \mathscr{M}_+$, the iterates $W_{k+1} = \mathcal{F}^{(\tau)}(W_k)$ converge in norm to $M^{(\tau)}$.*

In what follows, we will denote the unique solution of the RMDE with $\tau > 0$ by $M^{(\tau)}$.

A.3.2. Solution to the MDE

We are finally ready to prove the existence and the uniqueness of a solution to the MDE (19).

Theorem A.1 (Existence and Uniqueness)**.** *There exists a unique analytic matrix-valued function $M \in \mathscr{M}$ such that $M(z)$ solves the MDE (19) for every $z \in \mathbb{H}$.*

For clarity, we will separate the proof into two distinct sub-proofs: proof of existence and proof of uniqueness. Together, the following two proofs prove Theorem A.1.

[8] The inequality $\Im[\mathcal{F}^{(\tau)}(W)] \succeq \tau m_b^{-2} I_\ell$ is strict whenever $s \neq 0$.

Proof of existence in Theorem A.1. For every $k \in \mathbb{N}$, let $M^{(k^{-1})}$ be the unique solution to the RMDE and write

$$M_{1,1}^{(k^{-1})}(z) = \int_{\mathbb{R}} \frac{\Omega_{1,1}^{(k^{-1})}(d\lambda)}{\lambda - z}$$

the Stieltjes transform representation guaranteed by Lemma A.7. Additionally, let $\{v_j : j \in \mathbb{N}\} \subseteq \mathbb{C}^n$ be a countable dense subset of the ball of n-dimensional complex unit vectors.

By Corollary A.2, the family of measures $\{v_1^T \Omega_{1,1}^{(k^{-1})} v_1 : k \in \mathbb{N}\}$ is tight. Consequently, by Prokhorov's theorem, there exists a probability measure ω_1 and a subsequence $\{\tau_{1,k} : k \in \mathbb{N}\} \subseteq \{k^{-1} : k \in \mathbb{N}\}$ such that $v_1^* \Omega_{1,1}^{(\tau_{1,k})} v_1$ converges weakly to ω_1 as k approaches infinity.

We now proceed inductively. Assume that there exists $m \in \mathbb{N}$ and collection of compactly supported measures $\{\omega_j : 1 \le j \le m\}$ such that $v_j^* \Omega_{1,1}^{(\tau_{m,k})} v_j$ converges weakly to ω_j for all $1 \le j \le m$ as k approaches infinity. By Corollary A.2 and Prokhorov's theorem, there exists a probability measure ω_{m+1} and a subsequence $\{\tau_{m+1,k} : k \in \mathbb{N}\} \subseteq \{\tau_{m,k} : k \in \mathbb{N}\}$ such that $v_{m+1}^* \Omega_{1,1}^{(\tau_{m+1,k})} v_{m+1}$ converges weakly to ω_{m+1} as k approaches infinity. Also, by construction of the subsequence, $v_j^* \Omega_{1,1}^{(\tau_{m+1,k})} v_j$ converges weakly to ω_j for all $1 \le j \le m+1$ as k approaches infinity.

Let $\tau_k = \tau_{k,k}$ for all $k \in \mathbb{N}$. By construction, $v_j^* \Omega^{(\tau_k)} v_j$ converges weakly to a probability measure ω_j for every $j \in \mathbb{N}$ as $k \to \infty$. Furthermore, by Lemma A.2, $\{M_{1,1}^{(\tau_k)} : k \in \mathbb{N}\}$ is a locally uniformly bounded sequence of analytic functions. Hence, Montel's theorem guarantees the existence of a subsequence, which we will assume WLOG to be $\{\tau_k : k \in \mathbb{N}\}$ up to renaming, such that $M_{1,1}^{(\tau_k)}$ converges to an analytic function $M_{1,1}$.

By the proof of Corollary A.2, there exists $\kappa_+ \in \mathbb{R}_{>0}$ and a constant $c \in \mathbb{R}_{>0}$ such that

$$\int_{\lambda_+}^{\infty} \omega_j(d\lambda) = \lim_{k\to\infty} \int_{\lambda_+}^{\infty} v_j^* \Omega_{1,1}^{(\tau_k)}(d\lambda) v_j \le c \lim_{k\to\infty} \tau_k \int_{\lambda_+}^{\infty} (\lambda - \kappa)^{-2} d\lambda = 0$$

for every $\lambda_+ \ge \kappa_+$ and $j \in \mathbb{N}$. By Lemma A.7,

$$v_j^* \Im[M_{1,1}] v_j = \lim_{k\to\infty} v_j^* \Im[M_{1,1}^{(\tau_k)}] v_j = \Im[z] \int_{\mathbb{R}} \frac{\omega_j(d\lambda)}{|\lambda - z|^2}.$$

Since ω_j is a probability measure,

$$\int_{\mathbb{R}} \frac{\omega_j(d\lambda)}{|\lambda - z|^2} = \int_{[-\kappa_+,\kappa_+]} \frac{\omega_j(d\lambda)}{|\lambda - z|^2} \ge \left(\max_{\lambda \in [-\kappa_+,\kappa_+]} |\lambda - z| \right)^{-2}$$

which implies that $v_j^* \Im[M_{1,1}] v_j \ge \Im[z] \left(\max_{\lambda\in[-\kappa_+,\kappa_+]} |\lambda - z| \right)^{-2}$ for every $j \in \mathbb{N}$.

Fix $z \in \mathbb{H}$, $\epsilon = 3^{-1}(\Im[z])^2 \left(\max_{\lambda\in[-\kappa_+,\kappa_+]} |\lambda - z| \right)^{-2} \in \mathbb{R}_{>0}$. Let $v \in \mathbb{C}^n$ be any unit vector and let $j \in \mathbb{N}$ such that $\|v - v_j\| \le \epsilon$. Then,

$$u^* \Im[M_{1,1}] u = (v - v + u)^* \Im[M_{1,1}](v - v + u) \ge v^* \Im[M_{1,1}] v - 2\|u - v\| \|M_{1,1}\| \|v\| \ge \frac{\epsilon}{3\Im[z]} > 0.$$

In particular, $\Im[M_{1,1}(z)] \succ 0$ for all $z \in \mathbb{H}$.

Define $M_{1,2}, M_{2,1}$ and $M_{2,2}$ as functions of $M_{1,1}$ using (27c),(27d),(27e) respectively and let M be the block matrix with i,j block given by $M_{i,j}$ for all $(i,j) \in \{1,2\}^2$. It follows from Lemma A.1, that $Q - \mathcal{S}_{2,2}(M)$ is non-singular and that M is well-defined. By construction, it is clear that $M \in \mathscr{M}$ and that $M(z)$ solves (19) for all $z \in \mathbb{H}$. ☐

As mentioned earlier, removing $\tilde{S}$ from the super-operator has the advantage that each block in the block decomposition of the MDE is solely determined by the upper-left $n \times n$ block. This upper-left block exhibits favorable properties, including an a priori norm bound due to the position of the spectral parameter. By leveraging these properties, we can establish the existence of a solution and subsequently construct the remaining part of the solution.

Proof of uniqueness in Theorem A.1. Uniqueness of the solution follows from analyticity and the power series representation in Lemma A.8. Let $\lambda_+ \in \mathbb{R}_{>0}$ such that

$$M(z) = \sum_{j \in \mathbb{N}} z^{-j} M_j = (\mathbb{E}L - z\Lambda)^{-1} \sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} z^{-k} \mathcal{S}(M_k)(\mathbb{E}L - z\Lambda)^{-1} \right)^j$$

for every $z \in \mathbb{H}$ with $|z| \geq \lambda_+$.

Since resolvent of Hermitian matrices are analytic when the spectral parameter is away from the support, is follows from the decomposition in (29) that $(\mathbb{E}L - z\Lambda)^{-1}$ is analytic. Write $(\mathbb{E}L - z\Lambda)^{-1} = \sum_{j=0}^{\infty} z^{-j} L_j$ for some complex matrices $\{L_j : j \in \mathbb{N}\} \subseteq \mathbb{C}^{\ell \times \ell}$. Plugging this in the power series expansion of M and gathering coefficients of z^{-1}, we get that $M_1 = L_1 + L_0 \mathcal{S}(M_0) L_1 + L_0 \mathcal{S}(M_1) L_0$. We computed above that $(\mathbb{E}L - z\Lambda)^{-1} \to M_\infty$ as $|z| \to \infty$ and similarly for $M(z)$. In other words, $L_0 = M_0 = M_\infty$. Looking at the structure of the super-operator, $\mathcal{S}_{2,2}(M_0) = 0$, which gives us $L_0 \mathcal{S}(M_0) = 0$. In particular, $M_1 = L_1 + L_0 \mathcal{S}(M_0) L_1$ is expressible solely in terms of L_0 and L_1.

Let $k \in \mathbb{N}$. Gathering the coefficients for $z^{-(k+1)}$ in the power series expansion, we get that

$$M_{k+1} = f(M_0, M_1, \ldots, M_k) + L_0 \mathcal{S}(M_{k+1}) L_0$$

for some function analytic f. By induction hypothesis, we assume that $\{M_j : j \in \{0, 1, \ldots, k\}\}$ are fully determined by $\{L_j : j \in \mathbb{N}_0\}$. Furthermore, since $L_0 = M_\infty$ is 0 everywhere outside its lower $d \times d$ block,

$$L_0 \mathcal{S}(M_{k+1}) L_0 = \begin{bmatrix} 0_{n \times n} & 0_{n \times d} \\ 0_{d \times n} & M_\star \mathcal{S}_{2,2}(M) M_\star . \end{bmatrix}$$

Therefore, extracting the upper-left $n \times n$ block, we obtain that the upper-left $n \times n$ block along with both off-diagonal blocks of M_{k+1} are determined by the coefficient matrices $\{L_j : j \in \mathbb{N}_0\}$. Since $\mathcal{S}_{2,2}(M)$ does not depend on the lower-right block of M, we may also determine the lower-right block of M_{k+1}.

Inducting, we get that any two solution to (19), M and $\tilde{M}$ must be equal for all $z \in \mathbb{H}$ with $|z| > \lambda_+$ for some $\lambda_+ \in \mathbb{R}_{>0}$. By analytic continuation, it follows that $M(z) = \tilde{M}(z)$ for all $z \in \mathbb{H}$. □

For the rest of this document, we will denote the unique solution of the MDE with by M.

Remarkably, during the proof of the uniqueness of the solution to the linearized matrix Dyson equation, we come across the stability operator evaluated at $z = \infty$. The significance lies in the fact that our demonstration is equivalent to establishing the invertibility of the stability operator at infinity. This outcome effectively allows us to recursively determine the terms within the power series expansion of M.

In contrast to the guarantee offered by Lemma A.10, it is also important to consider that Theorem A.1 does not ensure pointwise convergence for the fixed-point iteration $f_{k+1} = \mathcal{F}(f_k)$ with an initial condition $f_0 \in \mathscr{M}$ to the solution of the matrix Dyson equation. This highlights one of the primary reason we rely on the solution to the regularized MDE as a means to establish stability, effectively treating it as a surrogate for the solution to the MDE.

A.4. Stability

Continuing our analysis, we introduce the following assumption, akin to what was presented in the main text.

Assumption 4. For every $z \in \mathbb{H}$, there exists a function f and subsequence $\{\tau_k\} \subseteq \mathbb{R}_{>0}$ such that $\tau_k \to 0$, $f(\tau_k) \to 0$ and $\|M^{(\tau_k)}(z) - M(z)\| \leq f(\tau_k) + o_\ell(1)$ for all $k \in \mathbb{N}$ and every $\ell \in \mathbb{N}$ large enough.

It is noteworthy that Assumption 4 is fulfilled within the frameworks based on the matrix Dyson equation for linearization as detailed in [EKN20, And13, FKN23]. This satisfaction is explicitly indicated by [EKN20, Equation 4.11], [And13, Estimates 6.3.3.], and [FKN23, Equation A.25]. In general, the validity of Assumption 4 in these cases stems from the ability to construct a dimension-independent representation of the solution to the (R)MDE using tools from free probability. As asserted by [HT05, Lemma 5.4], such a representation exists whenever L takes the form $L = A_0 \otimes I_n + \sum_{j=1}^k A_i \otimes X_j$, where $\{A_j\}_{j=0}^k$ forms a collection of complex $d \times d$ self-adjoint matrices, and $\{X_j\}_{j=1}^k$ forms a collection of independent random matrices with $\{(X_j)_{a,a}\}_{a=1}^n \cup \{(\sqrt{2}\Re X_j)_{a,b}\}_{a<b} \cup \{(\sqrt{2}\Im X_j)_{a,b}\}_{a<b}$ being a collection of n^2 i.i.d. centered Gaussian random variables for every $j \in \{1, 2, \dots, k\}$.

Furthermore, Assumption 4 is related to the *stability operator*. Following the notation in [AEKN19], the stability operator is defined as $\mathcal{L} : W \in \mathbb{C}^{\ell \times \ell} \mapsto W - M\mathcal{S}(W)M$. The concept of the stability operator is inherently connected to the analysis of the matrix Dyson equation [AEKN19, Erd19, AEK19b, FKN23]. The term stability operator is aptly chosen because, when it is both invertible and its inverse is bounded, it provides a means to establish the stability of the matrix Dyson equation through techniques like an implicit function theorem such as the one in [AEK19b, Lemma 4.10] as demonstrated in the work of [Erd19, EKN20]. The stability operator organically appears in the uniqueness argument, where its invertibility at infinity allows us to uniquely and recursively determine the power series expansion of the solution. The connection between the stability operator and Assumption 4 becomes apparent when we consider the derivative of $M^{(\tau)}(z)$ with respect to $i\tau$, which yields $\mathcal{L}(\partial_{i\tau} M(z)) = (M(z))^2$. Because $M(z)$ is bounded in operator norm, we can conclude that Assumption 4 is implied by the requirement of having an invertible stability operator with a bounded inverse.

We proceed to prove the asymptotic stability of the MDE. To this end, let $F(z) = \mathbb{E}(L - z\Lambda)^{-1} \in \mathscr{M}$ be the expected pseudo-resolvent. It is inconvenient to work directly with the expected pseudo-resolvent, and we will systematically prefer working with a regularized version of the same object. For each $\tau \in \mathbb{R}_{>0}$, we consider the expected regularized pseudo-resolvent $F^{(\tau)}(z) = \mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1} \in \mathscr{M}_+$ which satisfies

$$(\mathbb{E}L - \mathcal{S}(F^{(\tau)}(z)) - z\Lambda - i\tau I_\ell)F^{(\tau)}(z) = I_\ell + D^{(\tau)}, \tag{32}$$

where $D^{(\tau)}$ is a regularized perturbation term explicitly given by

$$D^{(\tau)} = \mathbb{E}\left[\left(\mathbb{E}L - L - \mathcal{S}(\mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1})\right)(L - z\Lambda - i\tau I_\ell)^{-1}\right]. \tag{33}$$

Essentially, we consider $F^{(\tau)}$ as a function that almost satisfies the MDE, up to an additive perturbation term $D^{(\tau)}$. By stability, we mean the property of the MDE that implies $F(z)$ is close — pointwise in spectral norm — to the solution $M(z)$ of (19) for every $z \in \mathbb{H}$ whenever the perturbation $D^{(\tau)}$ and the regularization parameter τ are small.

Since our primary objective is to investigate the behavior in the high-dimensional limit, it is essential for the super-operator $\mathcal{S}$, among other objects, to remain bounded as the problem dimension increases. We make the following assumption.

Assumption 5. Suppose that there exists $s \in \mathbb{R}_{>0}$ such that $\|\mathcal{S}(W)\| \leq s\|W\|$ for every $W \in \mathbb{C}^{\ell\times\ell}$ and $\limsup_{\ell\to\infty} s < \infty$. Furthermore, assume that $\limsup_{\ell\to\infty} \|\mathbb{E}L\| < \infty$ and $\limsup_{\ell\to\infty} \mathbb{E}\|(L - z\Lambda)^{-1}\|^2 < \infty$.

As denoted in (32), let us consider the expected regularized pseudo-resolvent $F^{(\tau)} = \mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1}$ as a function that approximately solve (22) up to an additive perturbation matrix $D^{(\tau)}$ explicitly provided by (33). For a fixed $z \in \mathbb{H}$, let $E_\tau = \mathcal{F}^{(\tau)}(F^{(\tau)})D^{(\tau)}$ for every $\tau \in \mathbb{R}_{\geq 0}$, defining the *error matrix*, and $\epsilon_\tau = \|E_\tau\|$ representing the *magnitude of the error* at τ. In the subsequent discussions, it will be convenient to fixed $z \in \mathbb{H}$ and write $F^{(\tau)} \equiv F^{(\tau)}(z)$ as well as $M^{(\tau)} \equiv M^{(\tau)}(z)$.

We want to compare the pseudo-resolvent with the solution of the MDE using the following pairwise comparisons:

$$(L - z\Lambda)^{-1} - M(z) = (L - z\Lambda)^{-1} - \mathbb{E}(L - z\Lambda)^{-1} \tag{34a}$$
$$+ \mathbb{E}(L - z\Lambda)^{-1} - \mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1} \tag{34b}$$
$$+ \mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1} - M^{(\tau)}(z) \tag{34c}$$
$$+ M^{(\tau)}(z) - M(z). \tag{34d}$$

The first comparison in (34a) corresponds to the concentration step of our argument. Although this difference may not generally be controlled in norm, we have the capability to demonstrate concentration, either in probability or almost surely, of generalized trace entries of the regularized pseudo-resolvent around its mean. By separating the concentration step from the rest of the method, we adopt a strategy that enables us to primarily work with *deterministic* objects throughout the analysis. This approach offers significant simplifications in various steps and allows us to work with norm bounds.

The second comparison in (34a) assesses the proximity of the pseudo-resolvent to its regularized counterpart, measured in norm. We show in Lemma A.11 that this difference can be easily controlled by the parameter τ and the norm of the pseudo-resolvent $(L - z\Lambda)^{-1}$. Consequently, if the norm of $(L - z\Lambda)^{-1}$ is bounded, we can employ the regularized pseudo-resolvent with small $\tau \in \mathbb{R}_{>0}$ as a valid approximation for the pseudo-resolvent.

The third comparison is directly linked with the stability properties of the RMDE. We will use the Carathéodory-Riffen-Finsler pseudometric to control the distance between $F^{(\tau)}$ and $M^{(\tau)}$ in term of ϵ_τ, and eventually the norm of the perturbation matrix, as the dimension of the problem increases. The convergence of the expected regularized pseudo-resolvent to the solution of the regularized matrix Dyson equation depends intricately on the rate at which τ approaches zero while ℓ increases to infinity.

The fourth and final comparison, (34d), simply states that the solution to (22) should be a good approximation for (19) for small τ. For a fixed $\tau \in \mathbb{R}_{>0}$, it follows from the construction of M that $\|M^{(\tau)}(z) - M(z)\| \to 0$ as $\tau \to 0$. However, because we are taking $\ell \to \infty$ and $\tau \to 0$, we rely on Assumption 4 to control this term.

In this section, we will show that (34b) becomes negligible as τ vanishes. Then, we will take care of (34c) by arguing that the RMDE is asymptotically stable for every $\tau \in \mathbb{R}_{>0}$. Our argument will imply that (19) is asymptotically stable.

A.4.1. Regularization

In this subsection, we analyze the discrepancy between the expected pseudo-resolvent and the expected regularized pseudo-resolvent in terms of the regularization parameter τ. We aim to establish bounds that quantify how close these two pseudo-resolvents are as τ varies. We have the following result.

Lemma A.11. *For every $\tau \in \mathbb{R}_{\geq 0}$ and $z \in \mathbb{H}$, $\|(L - z\Lambda - i\tau I_\ell)^{-1} - (L - z\Lambda)^{-1}\| \leq \tau\|(L - z\Lambda)^{-1}\|^2$.*

Proof. By the resolvent trick, $\|(L-z\Lambda-i\tau I_\ell)^{-1}-(L-z\Lambda)^{-1}\| \leq \tau\|(L-z\Lambda-i\tau I_\ell)^{-1}\|\|(L-z\Lambda)^{-1}\|$. Let $v \in \mathbb{C}^\ell$ be arbitrary and decompose $L - z\Lambda - i\tau I_\ell = X + iY - i\tau I_\ell$ with $X = \Re[L - z\Lambda]$ and $Y = \Im[L - z\Lambda]$. Then, using the fact that $(L - z\Lambda - i\tau I_\ell)^* = X - iY + i\tau I_\ell$,

$$
\begin{aligned}
v^* (X + iY - i\tau I_\ell)^* (X + iY - i\tau I_\ell)\, v &\geq v^*(X + iY)^*(X + iY)v + \tau^2 v^* v - 2\tau v^* Y v \\
&\geq v^*(X + iY)^*(X + iY)v.
\end{aligned}
$$

Because taking the inverse reverses the Loewner partial ordering, it follows that $\|(L - z\Lambda - i\tau I_\ell)^{-1}\| \leq \|(L - z\Lambda)^{-1}\|$. □

By Lemma A.11 and Jensen's inequality, the expected pseudo-resolvent $\mathbb{E}(L-z\Lambda)^{-1}$ is well-approximated by its regularized version for small τ as long as $\mathbb{E}\|(L - z\Lambda)^{-1}\|^2$ is bounded.

A.4.2. Bounds in the Carathéodory-Riffen-Finsler pseudometric

Given a domain $\mathscr{D}$ in a complex Banach space $(\mathscr{X}, \|\cdot\|)$, the *infinitesimal Carathéodory-Riffen-Finsler (CRF)-pseudometric* [Har03, Har79] for $\mathscr{D}$ is defined as

$$
\alpha : (x, v) \in \mathscr{D} \times \mathscr{X} \mapsto \sup\{|\mathrm{D}g(x)v| \,:\, g \in \mathrm{Hol}(\mathscr{D}, \Delta)\} \in \mathbb{R}
$$

where Δ is the complex open unit disk and $\mathrm{D}g(x)$ is the Fréchet derivative of g at x. We set

$$
\mathscr{L}(\gamma) = \int_0^1 \alpha(\gamma(y), \gamma'(t))\mathrm{d}t
$$

for *admissible curve* γ in the set Γ of all curves in $\mathscr{D}$ with piecewise continuous derivative. The infinitesimal pseudometric is a seminorm at each point in $\mathscr{D}$, and we view $\mathscr{L}(\gamma)$ as the length of the curve γ measured with respect to α. The *CRF-pseudometric* [Har03, Har79] of $\mathscr{D}$ is defined as

$$
\rho : (x, y) \in \mathscr{D}^2 \mapsto \inf\{\mathscr{L}(\gamma) \,:\, \gamma \in \Gamma,\, \gamma(0) = x,\, \gamma(1) = y\} \in \mathbb{R}_{\geq 0}.
$$

For every $b \in \mathbb{R}_{>0}$, we will consider the domain $\mathscr{A}_b$ defined in (31) in the Banach space of $\ell \times \ell$ complex matrices equipped with the operator norm.

As we already stated multiple times, the map $\mathcal{F}^{(\tau)}$ is a strict contraction with respect to the CRF-pseudometric.

Lemma A.12. *Fix $z \in \mathbb{H}$ and $\tau \in \mathbb{R}_{>0}$. For every $b \in \mathbb{R}_{>0}$, let $m_b = \|\mathbb{E}L\| + sb + |z| + \tau + 1$ and $\xi = (m_b^2\tau^{-2} - 1)^{-1}$. Suppose that $\tau^{-1}(1+2\xi) < b$ and let ρ denotes the CRF-pseudometric on $\mathscr{A}_b$. Then, for every $X, Y \in \mathscr{A}_b$, $\rho(\mathcal{F}^{(\tau)}(X), \mathcal{F}^{(\tau)}(Y)) \leq (1 + \xi)^{-1}\rho(X, Y)$.*

Proof. Define $\mathcal{G} : W \in \mathscr{A}_b \mapsto \mathcal{F}^{(\tau)}(W) + \xi(\mathcal{F}^{(\tau)}(W) - \mathcal{F}^{(\tau)}(X))$. By Lemmas A.2 and A.9, we have $\Im[\mathcal{F}^{(\tau)}(W)] \succ \tau m_b^{-2}$ and $\|\mathcal{F}^{(\tau)}(W)\| \leq \tau^{-1}$ for every $W \in \mathscr{A}_b$. Hence, $\Im[\mathcal{G}(W)] \succ (1+\xi)\tau m_b^{-2} - \xi\tau^{-1} \succeq 0$ and $\|\mathcal{G}(W)\| \leq \tau^{-1} + 2\tau^{-1}\xi < b$. Therefore, $\mathcal{G}$ is a strictly holomorphic function in the sense that it is an holomorphic function mapping $\mathscr{A}_b$ strictly into itself. The result follows from the proof of Earle-Hamilton fixed-point theorem [Har79, Theorem 4]. □

It is important to note that Lemma A.12 does not imply that the function $\mathcal{F}^{(\tau)}$ is a contraction with respect to the operator norm. Indeed, with respect to the operator norm, the map $\mathcal{F}^{(\tau)}$ is merely an analytic function mapping *a* domain strictly into itself.

Our main tool to extrapolate results about norms is the following Schwarz-Pick inequality, which we state here for completeness.

Proposition A.1 ([Har79, Proposition 3]). *Let $\mathscr{D}_1$ and $\mathscr{D}_2$ be domains in complex Banach spaces and let ρ_1 and ρ_2 be the associated CRF-pseudometric. If $h : \mathscr{D}_1 \mapsto \mathscr{D}_2$ is holomorphic, then $\rho_2(h(x), h(y)) \le \rho_1(x, y)$ for all $x, y \in \mathscr{D}_1$.*

In fact, it can be showed that the inequality in Proposition A.1 may be replaced by an equality when the function h is biholomorphic, making the CRF-pseudometric biholomorphically invariant. Proposition A.1 states that the CRF-pseudometric is a contraction in the sense that it is decreasing under holomorphic mappings.

Let ρ_Δ denote the CRF-pseudometric on the complex open unit disk Δ. Proposition A.1 is particularly useful because ρ_Δ, which is called the Poincaré metric, admits the closed form

$$\rho_\Delta(z_1, z_2) = \operatorname{arctanh}\left|\frac{z_1 - z_2}{1 - \bar{z_1} z_2}\right|. \tag{35}$$

See [Har79, Example 2] for a derivation of (35). Using Proposition A.1 and equation (35), we may show that the CRF-pseudometric dominates the operator norm.

Lemma A.13. *Fix $z \in \mathbb{H}$ and $\tau \in \mathbb{R}_{>0}$. For every $b \in \mathbb{R}_{>0}$ with $b > \tau^{-1}$ and ρ the CRF-pseudometric on $\mathscr{A}_b$, $\|M^{(\tau)} - F^{(\tau)}\| \le (b + \tau^{-1}) \tanh(\rho(M^{(\tau)}, F^{(\tau)}))$.*

Proof. Let m_b be defined as in the proof of Lemma A.12 and recall that $\|M^{(\tau)}\| \le \tau^{-1}$ as well as $\Im[M^{(\tau)}] \succ \tau m_b^{-2}$. Similarly, we have $\|(L - z\Lambda - i\tau I_\ell)^{-1}\| \le \tau^{-1}$ and $\Im[(L - z\Lambda - i\tau I_\ell)^{-1}] \succeq \tau(\|L\| + |z| + \tau)^{-2}$. Hence, by monotonicity of the expectation and Jensen's inequality, $\|F^{(\tau)}\| \le \tau^{-1}$ and $\Im[F^{(\tau)}] \succeq \tau(\mathbb{E}\|L\| + |z| + \tau)^{-2} \succ 0$. In particular, $M^{(\tau)}, F^{(\tau)} \in \mathscr{A}_b$.

Let $U \in \mathbb{C}^{\ell \times \ell}$ with $\|U\|_* \le 1$ and define the holomorphic function $f : W \in \mathscr{A}_b \mapsto \operatorname{tr}(U(W - M^{(\tau)}))(b + \tau^{-1})^{-1} \in \Delta$. By Proposition A.1 and equation (35),

$$\operatorname{arctanh}\left|\frac{\operatorname{tr}(U(W - M^{(\tau)}))}{(b + \tau^{-1})}\right| = \rho_\Delta\left(f(M^{(\tau)}), f(W)\right) \le \rho\left(M^{(\tau)}, W\right)$$

for every $W \in \mathscr{A}_b$. Plugging $W = F^{(\tau)}$ and rearranging,

$$|\operatorname{tr}(U(F^{(\tau)} - M^{(\tau)}))| \le (b + \tau^{-1}) \tanh\left(\rho(M^{(\tau)}, F^{(\tau)})\right).$$

We obtain the result by taking the supremum over U and using duality. □

In Lemma A.10, we established that the solution to the regularized matrix Dyson equation can be obtained using a fixed-point iteration scheme. Using this idea, we will recursively define a sequence of matrices and use the contraction property in Lemma A.12 to control the distance between $M^{(\tau)}$ and $F^{(\tau)}$ in the CRF-pseudometric. Since the CRF-pseudometric dominates the operator norm, we will obtain convergence in norm. The only remaining ingredients are control of $\rho(M^{(\tau)}, F^{(\tau)})$ and $\rho(\mathcal{F}^{(\tau)}(F^{(\tau)}), F^{(\tau)})$. While the norm $\|M^{(\tau)} - M_0\|$ may be easily bounded uniformly in ℓ, transferring this bound to the CRF-pseudometric poses additional difficulties which we address in the following lemma.

Lemma A.14. *Under the settings of Lemma A.12, further assume that* $b > \tau^{-1} + \tau m_b^{-2}$ *and* $\epsilon_\tau < \tau m_b^{-2}$*. Then,* $\rho(M^{(\tau)}, F^{(\tau)}) < 4\tau^{-1}(\tau m_b^{-2} - \epsilon_\tau)^{-1}$.

Proof. By (32), $F^{(\tau)} = \mathcal{F}^{(\tau)}(F^{(\tau)}) + E_\tau$ which implies that $\Im[F^{(\tau)}] \succeq \Im[\mathcal{F}^{(\tau)}(F^{(\tau)})] - \epsilon_\tau \succ \tau m_b^{-2} - \epsilon_\tau$.

Let $W_t := tM^{(\tau)} + (1-t)F^{(\tau)}$ be a linear interpolation of $M^{(\tau)}$ and $F^{(\tau)}$ for $t \in [0,1]$.[9] Then, $\|W_t\| \le \tau^{-1} < b$ and $\Im[W_t] \succ \tau m_b^{-2} - \epsilon_\tau$ for every $t \in [0,1]$. It is clear that $\mathbb{B}_{\tau m_b^{-2}-\epsilon_\tau}(W_t) \subseteq \mathscr{A}_b$, where $\mathbb{B}_r(X)$ denotes the open ball in $\mathbb{C}^{\ell\times\ell}$ of radius r centered around X. Define a sequence $\{t_j\}_{j\in\mathbb{N}_0} \subseteq [0,1]$ such that $t_0 = 0$ and $t_{j+1} = \big(t_j + 4^{-1}\tau(\tau m_b^{-2} - \epsilon_\tau)\big) \wedge 1$ for every $j \in \mathbb{N}_0$. Indeed,

$$\begin{aligned}\|W_{t_{j+1}} - W_{t_j}\| &= \|(t_{j+1} - t_j)M^{(\tau)} + (t_j - t_{j+1})F^{(\tau)}\| \\ &\le 2\tau^{-1}(t_{j+1} - t_j) \le 2^{-1}(\tau m_b^{-2} - \epsilon_\tau) < \tau m_b^{-2} - \epsilon_\tau\end{aligned}$$

for every $j \in \mathbb{N}_0$. To summarize, we constructed a sequence of complex matrices $\{W_{t_j}\}_{j\in\mathbb{N}_0} \subseteq \mathscr{A}_b$ interpolating $M^{(\tau)}$ and $F^{(\tau)}$ such that $W_{t_{j+1}} \in \mathbb{B}_{\tau m_b^{-2}-\epsilon_\tau}(W_{t_j}) \subseteq \mathscr{A}_b$. The sequence $\{t_j\}_{j\in\mathbb{N}_0}$ attains 1 in at most $\eta := 4\tau^{-1}(\tau m_b^{-2} - \epsilon_\tau)^{-1}$ steps.

Fix $j \in \mathbb{N}$ such that $t_j \ne 1$ and define the holomorphic function

$$g : w \in \Delta \mapsto W_{t_j} + \frac{w(\tau m_b^{-2} - \epsilon_\tau)}{\|W_{t_{j+1}} - W_{t_j}\|}(W_{t_{j+1}} - W_{t_j}) \in \mathscr{A}_b$$

as in [Har79, proof of Theorem 5]. By Proposition A.1 and equation (35),

$$\rho\left(W_{t_j}, W_{t_{j+1}}\right) = \rho\left(g(0), g\left(\frac{\|W_{t_{j+1}} - W_{t_j}\|}{\tau m_b^{-2} - \epsilon_\tau}\right)\right) \le \rho_\Delta\left(0, \frac{\|W_{t_{j+1}} - W_{t_j}\|}{\tau m_b^{-2} - \epsilon_\tau}\right) \le \operatorname{arctanh}(2^{-1}) < 1.$$

Finally, applying the triangle inequality, $\rho(M^{(\tau)}, M_0) \le \sum_{j=0}^{\eta-1} \rho(W_{t_j}, W_{t_{j+1}}) < \eta$. This concludes the proof. ☐

Using a similar argument, we can bound $\rho(\mathcal{F}^{(\tau)}(F^{(\tau)}), F^{(\tau)})$ using ϵ_τ.

Lemma A.15. *Under the settings of Lemma A.14,* $\rho(\mathcal{F}^{(\tau)}(F^{(\tau)}), F^{(\tau)}) \le \operatorname{arctanh}(\epsilon_\tau m_b^2/\tau)$.

Proof. Using (32), $\mathcal{F}^{(\tau)}(F^{(\tau)}) - F^{(\tau)} = -E_\tau$ which implies that $\|\mathcal{F}^{(\tau)}(F^{(\tau)}) - F^{(\tau)}\| \le \epsilon_\tau$. Furthermore, $F^{(\tau)} \in \mathscr{A}_b$ by the proof of Lemma A.13. Define the holomorphic function

$$g : w \in \Delta \mapsto \mathcal{F}^{(\tau)}(F^{(\tau)}) + \frac{w\tau m_b^{-2}}{\|\mathcal{F}^{(\tau)}(F^{(\tau)}) - F^{(\tau)}\|}(F^{(\tau)} - \mathcal{F}^{(\tau)}(F^{(\tau)})) \in \mathscr{A}_b.$$

Assuming that $\epsilon_\tau < \tau m_b^{-2}$,

$$\begin{aligned}\rho\left(\mathcal{F}^{(\tau)}(F^{(\tau)}), F^{(\tau)}\right) = \rho\left(g(0), g\left(\frac{\|\mathcal{F}^{(\tau)}(F^{(\tau)}) - F^{(\tau)}\|}{\tau m_b^{-2}}\right)\right) &\le \rho_\Delta\left(0, \frac{\|\mathcal{F}^{(\tau)}(F^{(\tau)}) - F^{(\tau)}\|}{\tau m_b^{-2}}\right) \\ &\le \operatorname{arctanh}\left(\frac{\|\mathcal{F}^{(\tau)}(F^{(\tau)}) - F^{(\tau)}\|}{\tau m_b^{-2}}\right),\end{aligned}$$

where we have used the Schwarz-Pick inequality stated Proposition A.1 as well as (35). ☐

[9]Since the CRF-pseudometric on $\mathscr{A}_b$ can be seen as a generalization of the hyperbolic metric, the specific choice of linear interpolation in our analysis is not expected to be optimal. Although alternative interpolation schemes could potentially yield more optimal results, the focus of our analysis lies elsewhere, and the linear interpolation is not a limiting factor in terms of the convergence rate we obtain.

A.4.3. Main stability result

Combining Lemmas A.12 to A.15, we obtain the following asymptotic stability result.

Theorem A.2 (Theorem 3.2 from the main text). *Suppose that* $\|D^{(\tau)}\| \xrightarrow{\ell\to\infty} 0$ *for every* $\tau \in \mathbb{R}_{>0}$. *Then, under Assumptions 4 and 5,* $\|M(z) - \mathbb{E}(L - z\Lambda)^{-1}\| \xrightarrow{\ell\to\infty} 0$ *for every* $z \in \mathbb{H}$.

Proof. Fix $z \in \mathbb{H}$, $\tau \in \mathbb{R}_{>0}$. Let $b = \tau^{-1}+2\tau$, $m_b = \|\mathbb{E}L\|+sb+|z|+\tau+1$ and $\xi = (m_b^2\tau^{-2}-1)^{-1}$. In order to apply Lemmas A.12 and A.14, we first have to show that b satisfies $\tau^{-1}(1+2\xi) < b$ and $b > \tau^{-1} + \tau m_b^{-2}$. To this end, notice that $m_b^2\tau^{-2} > \tau^{-2} + 1$. Therefore, $\xi < \tau^2$ and $\tau^{-1}(1+2\xi) < b$. Furthermore, $m_b^{-2} < 1$ so $\tau^{-1} + \tau m_b^{-2} < b$.

By definition of the error matrix, we have $\epsilon_\tau \le \tau^{-1}\|D^{(\tau)}\|$. Assume that ℓ is large enough such that $\epsilon_\tau < \tau m_b^{-2}$ and let ρ denote the CRF-pseudometric on $\mathscr{A}_b$. By Lemma A.13,

$$\|M^{(\tau)} - F^{(\tau)}\| \le (b+\tau^{-1})\tanh\left(\rho(M^{(\tau)}, F^{(\tau)})\right).$$

Recursively define a sequence $\{M_k : k \in \mathbb{N}_0\} \subseteq \mathscr{A}_+$ such that $M_0 = F^{(\tau)}(z)$ and $M_{k+1} = \mathcal{F}^{(\tau)}(M_k)(z)$ for every $k \in \mathbb{N}_0$. Hence, by Lemmas A.12, A.14 and A.15,

$$\begin{aligned}
\rho(M^{(\tau)}, F^{(\tau)}) &\le \rho(M^{(\tau)}, M_k) + \sum_{j=1}^{k} \rho(M_j, M_{j-1}) \\
&\le (1+\xi)^{-k}\rho(M^{(\tau)}, F^{(\tau)}) + \rho\left(\mathcal{F}^{(\tau)}(F^{(\tau)}), F^{(\tau)}\right)\sum_{j=0}^{\infty}(1+\xi)^{-j} \\
&\le \frac{4}{\tau(\tau m_b^{-2} - \epsilon_\tau)(1+\xi)^k} + \frac{\operatorname{arctanh}(\epsilon_\tau/(\tau m_b^{-2}))}{1-(1+\xi)^{-1}}.
\end{aligned}$$

Since the above inequalities hold for every $k \in \mathbb{N}$, we may take the limit as $k \to \infty$ to obtain

$$\rho(M^{(\tau)}, F^{(\tau)}) \le \frac{(1+\xi)}{\xi}\operatorname{arctanh}\left(\frac{m_b^2\|D^{(\tau)}\|}{\tau^2}\right).$$

As $\tau \to 0$ and $\ell \to \infty$, assuming that $\|\mathbb{E}L\|$ and s remains bounded, we have $b \asymp m_b \asymp \tau^{-1}$ and $\xi \asymp \tau^2$. Thus, asymptotically, we have

$$\|M^{(\tau)} - F^{(\tau)}\| \lesssim \tau^{-1}\tanh\left(\tau^{-2}\operatorname{arctanh}\left(\tau^{-4}\|D^{(\tau)}\|\right)\right).$$

For instance, letting $\tau = \|D^{(\tau)}\|^{1/8}$, we get $\|M^{(\tau)} - F^{(\tau)}\| \lesssim \|D^{(\tau)}\|^{1/8}$ for every ℓ large enough.

Combining this with Assumptions 4 and 5 and lemma A.11, there exists a function f and a subsequence $\{\tau_k\} \subseteq \mathbb{R}_{>0}$ such that $\tau_k \to 0$ and $f(\tau_k) \to 0$ and

$$\begin{aligned}
\|M(z) - F(z)\| &\le \|M(z) - M^{(\tau)}(z)\| + \|M^{(\tau_k)}(z) - F^{(\tau)}(z)\| + \|F^{(\tau)}(z) - F(z)\| \\
&\lesssim f(\tau_k) + \tau_k^{-1}\tanh\left(\tau_k^{-2}\operatorname{arctanh}(\tau_k^{-4}\|D^{(\tau)}\|)\right) + \tau_k + o_\ell(1)
\end{aligned}$$

for every $k \in \mathbb{N}$ and $\ell \in \mathbb{N}$ large enough. □

The proof of Theorem A.2 highlights a contrast between τ and ℓ in terms of their effect on the convergence behavior. Furthermore, it's important to highlight that the proof of Theorem A.2 essentially simplifies the stability analysis of the MDE to demonstrating that the difference between the expected regularized pseudo-resolvent before and after a single application of the RMDE map is small in the natural CRF-pseudometric.

A.5. Resolvent Approximation

Now that we have existence of a unique solution M to (19) as well as an asymptotic stability property, we want to show that $M(z)$ serves as a favorable asymptotic approximation for the pseudo-resolvent $(L - z\Lambda)^{-1}$. In view of Theorem A.2, the focus shifts to proving that the perturbation matrix vanishes in norm as the problem dimension grows for every regularization parameter. There are various methods to establish this, depending on the assumptions about the linearization L. To apply our framework and study random features ridge regression, we naturally choose a route based on Gaussian concentration inequalities. This choice confines our theoretical considerations to linearizations characterized by Gaussian-concentrated entries.

Assumption 6. Suppose that $\gamma \in \mathbb{N}$, $g \sim \mathcal{N}(0, I_\gamma)$ and that there exists a map $\mathcal{C} : \mathbb{R}^\gamma \mapsto \mathbb{R}^{\ell\times\ell}$ such that $L \equiv L(g) = \mathcal{C}(g) + \mathbb{E}L$. Furthermore, assume that $\mathcal{C}$ is symmetric in the sense that $\mathcal{C}(x) = (\mathcal{C}(x))^T$ for every $x \in \mathbb{R}^\gamma$.

With reference to (34), our remaining tasks are to demonstrate the vanishing of the perturbation matrix in ℓ and establish the concentration of functionals of $(L - z\lambda)^{-1}$ around their mean. The next two subsections are dedicated to these objectives.

A.5.1. Perturbation

We focus on the expected pseudo-resolvent denoted as $F \equiv F(z) := \mathbb{E}(L - z\Lambda)^{-1}$, and more generally, the regularized expected pseudo-resolvent denoted as $F^{(\tau)} \equiv F^{(\tau)}(z) := \mathbb{E}(L - z\Lambda - i\tau I_\ell)^{-1}$ for every $\tau \in \mathbb{R}_{>0}$. In view of our stability argument, we only have to show that the perturbation matrix defined in (33) is vanishing.

Several methods can be employed to demonstrate that the perturbation matrix is vanishing, each tailored to specific assumptions about the linearization. Based on our motivating example, we leverage a Gaussian concentration argument inspired by works such as [LLC18, Cho22]. To utilize this Gaussian concentration argument, we operate under Assumption 6. This allows us to derive straightforward conditions on the function $\mathcal{C}$, ensuring $D^{(\tau)} \to 0$ as $\ell \to \infty$ for all $\tau \in \mathbb{R}_{>0}$. Additionally, we employ a Gaussian concentration inequality to show that Lipschitz functionals of the regularized pseudo-resolvent $(L - z\Lambda - i\tau I_\ell)^{-1}$ concentrate around their mean. We present a Gaussian concentration inequality for Lipschitz functions and direct the reader to [Led01] and [Tao12] for further details. Alternatively, we can utilize the Nash-Poincaré inequality [Pas05, Proposition 2.4], a consequence of Stein's lemma, to establish concentration.

Proposition A.2 (Gaussian concentration inequality for Lipschitz functions [Tao12, Theorem 2.1.12] and [Led01, Proposition 1.10])**.** *Suppose that $x \sim \mathcal{N}(0, I_\gamma)$ and let $f : \mathbb{R}^\gamma \mapsto \mathbb{R}$ be a λ-Lipschitz function. Then, for every $t \in \mathbb{R}$,*

$$\mathbb{P}\left(|f(x) - \mathbb{E}[f(x)]| \geq t\right) \leq c_1 e^{-\frac{c_2 t^2}{\lambda^2}}$$

for some absolute constants $c_1, c_2 \in \mathbb{R}_{>0}$.

Under Assumption 6, we aim to decompose the perturbation matrix $D^{(\tau)}$ into terms that are amenable to analysis. To achieve this, define

$$\begin{aligned}\Delta(L, \tau; z) &= \mathbb{E}[(L - \mathbb{E}L)(L - z\Lambda - i\tau I_\ell)^{-1}] \\ &\quad + \mathbb{E}[(\tilde{L} - \mathbb{E}L)(L - z\Lambda - i\tau I_\ell)^{-1}(\tilde{L} - \mathbb{E}L)(L - z\Lambda - i\tau I_\ell)^{-1}] \qquad (36)\end{aligned}$$

where $\tilde{L}$ is an i.i.d. copy of L,

$$\tilde{\mathcal{S}} : M \in \mathbb{C}^{\ell \times \ell} \mapsto \mathbb{E}\left[(L - \mathbb{E}L)M(L - \mathbb{E}L)\right] - \mathcal{S}(M) \in \mathbb{C}^{\ell \times \ell},\text{[10]} \tag{37}$$

and consider the decomposition

$$D^{(\tau)} = \mathbb{E}\left[\mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1})(L - z\Lambda - i\tau I_\ell)^{-1}\right] - \mathcal{S}(F^{(\tau)})F^{(\tau)} \tag{38a}$$

$$+ \mathbb{E}\left[\tilde{\mathcal{S}}((L - z\Lambda - i\tau I_\ell)^{-1})(L - z\Lambda - i\tau I_\ell)^{-1}\right] \tag{38b}$$

$$- \Delta(L, \tau). \tag{38c}$$

The first perturbation term in (38a) arises from the use of the expected pseudo-resolvent in Theorem A.2. To ensure that this perturbation term is asymptotically small, we require the super-operator $\mathcal{S}$ to be *averaging*. This implies that $\mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1})$ should exhibit a "law of large numbers" behavior and converge to a deterministic limit. While working directly with the pseudo-resolvent would eliminate this specific perturbation term from the expectation of $D^{(\tau)}$, such an approach would have its disadvantages. Utilizing the expected pseudo-resolvent, on the other hand, allows us to work with deterministic objects and leverage norm bounds. We derive a condition for $\mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1})$ to concentrate around its mean based on Gaussian concentration.

The second perturbation term, as expressed in (38b), arises from our specific definition of the super-operator and would not be present if we defined the super-operator as $\mathbb{E}[(L-\mathbb{E}L)M(L-\mathbb{E}L)]$. However, our chosen definition of the super-operator, coupled with the assumption $Q = \mathbb{E}Q$, ensures that the MDE can be determined by the upper-left $n \times n$ block. This distinction allows us to establish the existence of a solution to (19). Consequently, we view $\tilde{\mathcal{S}}$ as a correction term that be vanishing in ℓ.

Finally, (38c) posits that the matrix L should approximate a Gaussian distribution in the sense that it should asymptotically satisfy a matrix Stein lemma with a vanishing error. The quantity $\|\Delta(L, \tau)\|$ serves informally as a metric characterizing the distance between L and a matrix with Gaussian entries. Notably, the subsequent result demonstrate that $\Delta(L, \tau) = 0$ holds whenever L has Gaussian entries.

In order to maintain a certain level of abstraction, we will directly assume that the mapping $g \mapsto \mathcal{S}(L(g) - z\Lambda - i\tau I_\ell)^{-1}$ is λ-Lipschitz with respect to the operator norm and employ an ϵ-net argument to obtain bounds $\mathbb{E}_{\tilde{L}}[\|(\tilde{L} - \mathbb{E}L)((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z))(\tilde{L} - \mathbb{E}L)\|]$ for $k \in \mathbb{N}$.

Lemma A.16. *Fix $z \in \mathbb{H}$ and $\tau \in \mathbb{R}_{>0}$. Assume that the mapping $g \mapsto \mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1})$ is λ-Lipschitz with respect to the operator norm. Then, for every $k \in \mathbb{N}$, there exists an absolute constant $c \in \mathbb{R}_{>0}$ such that*

$$\mathbb{E}\left[\|\mathcal{S}\left((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z)\right)\|^k\right] \le c\ell^{k/2}\lambda^k.$$

Proof. Let $u, v \in \mathbb{C}^\ell$ be arbitrary unit vectors. By Proposition A.2,

$$\mathbb{P}\left(\lambda^{-1}\left|u^*\mathcal{S}\left((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z)\right)v\right| \ge t\right) \le c_1 e^{-c_2 t^2}$$

for some absolute constant $c_1, c_2 \in \mathbb{R}_{>0}$. Suppose that $\epsilon \in (0, 8^{-1})$ and let $\mathscr{N}$ be an ϵ-net for the unit ball of ℓ-dimensional real vectors. Then, given $u \in \mathbb{C}^\ell$, we may find $v_1, v_2 \in \mathscr{N}$ such

[10] We may also remove any term in the upper-left block of $\mathbb{E}\left[(L - \mathbb{E}L)M(L - \mathbb{E}L)\right]$ from $\mathcal{S}$ and add them to $\tilde{\mathcal{S}}$ without changing any of our arguments.

that $\|u - v_1 - iv_2\| \leq \|\Re[u] - v_1\| + \|\Im[u] - v_2\| \leq 2\epsilon$. In particular, $\mathscr{N} + i\mathscr{N} := \{v_1 + iv_2 : v_1, v_2 \in \mathscr{N}\}$ forms a 2ϵ-net for the unit sphere of ℓ-dimensional complex unitary vectors. By [Ver18, Corollary 4.2.13], $|\mathscr{N} + i\mathscr{N}| \leq (2\epsilon^{-1} + 1)^{2\ell}$.

Let $u, v \in \mathbb{C}^\ell$ be unitary and let $u_0, v_0 \in \mathscr{N} + i\mathscr{N}$ such that $\|u - u_0\| \leq 2\epsilon$ and $\|v - v_0\| \leq 2\epsilon$. Let $W \in \mathbb{C}^{\ell\times\ell}$ be any matix. Using the identity $u^*Wv = u_0^*Wv_0 + (u^* - u_0^*)Wv + u_0^*W(v - v_0)$, we obtain $|u^*Wv| \leq \sup_{u_0,v_0\in\mathscr{N}+i\mathscr{N}} |u_0^*Wv_0| + 8\epsilon\|W\|$. Taking the supremum over unitary complex vectors u and v, we get that $\|W\| \leq (1 - 8\epsilon)^{-1} \sup_{u_0,v_0\in\mathscr{N}+i\mathscr{N}} |u_0^*Wv_0|$. Therefore, using a union bound,

$$\mathbb{P}\left(\lambda^{-1}\|\mathcal{S}\left((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z)\right)\| \geq u\right) \leq c_1(2\epsilon^{-1} + 1)^{4\ell}e^{-c_2u^2}$$

for every $u \in \mathbb{R}$. Let $u = c(2\sqrt{\ell} + t)$ for some $t \in \mathbb{R}_{\geq 0}$ such that $u^2 \geq c^2(4\ell + t^2)$. Choosing $c \in \mathbb{R}_{>0}$ large enough such that $c_2^2u^2 \geq \ln(2\epsilon^{-1} + 1)4\ell + t^2$,

$$c_1(2\epsilon^{-1} + 1)^{4\ell}e^{-c_2u^2} \leq c_1(2\epsilon^{-1} + 1)^{4\ell}e^{-\ln(2\epsilon^{-1}+1)4\ell}e^{-t^2} = c_1e^{-t^2}.$$

Let $k \in \mathbb{N}$ be arbitrary. Then,

$$\begin{aligned}\mathbb{E}&\left[\lambda^{-k}\|\mathcal{S}\left((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z)\right)\|^k\right] \\ &= \int_0^\infty \mathbb{P}\left(\lambda^{-k}\|\mathcal{S}\left((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z)\right)\|^k \geq u\right) \mathrm{d}u\end{aligned}$$

On one hand,

$$\int_0^{(2c\sqrt{\ell})^k} \mathbb{P}\left(\lambda^{-k}\|\mathcal{S}\left((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z)\right)\|^k \geq u\right) \mathrm{d}u \leq (2c\sqrt{\ell})^k.$$

On the other hand,

$$\begin{aligned}&\int_{(2c\sqrt{\ell})^k}^\infty \mathbb{P}\left(\lambda^{-k}\|\mathcal{S}\left((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z)\right)\|^k \geq u\right) \mathrm{d}u \\ &\leq c^kc_1k\int_0^\infty (2\sqrt{\ell} + t)^{k-1}e^{-t^2}\mathrm{d}t \\ &\leq c^kc_1k\left((4\sqrt{\ell})^{k-1}\int_0^{2\sqrt{\ell}} e^{-t^2}\mathrm{d}t + \int_{2\sqrt{\ell}}^\infty (2t)^{k-1}e^{-t^2}\mathrm{d}t\right)\end{aligned}$$

Since $\int_0^{2\sqrt{\ell}} e^{-t^2}\mathrm{d}t \leq \int_0^\infty e^{-t^2}\mathrm{d}t = \sqrt{\frac{\pi}{2}}$, the first term in on the RHS of the inequality above grows with ℓ like $\ell^{(k-1)2}$. For the second term, simply notice that

$$\begin{aligned}\int_{2\sqrt{\ell}}^\infty (2t)^{k-1}e^{-t^2}\mathrm{d}t &= (2t)^{k-2}\left(\int_{-\infty}^t (2u)e^{-u^2}\mathrm{d}u\right)\Big|_{2\sqrt{\ell}}^\infty \\ &\quad - 2(k-2)\int_{2\sqrt{\ell}}^\infty (2t)^{k-3}\left(\int_{-\infty}^t (2u)e^{-u^2}du\right)\mathrm{d}t\end{aligned}$$

where $\int_{-\infty}^t (2u)e^{-u^2}\mathrm{d}u = -e^{-t^2}$. Hence,

$$\int_{2\sqrt{\ell}}^\infty (2t)^{k-1}e^{-t^2}\mathrm{d}t = (4\sqrt{\ell})^{k-2}e^{-4\ell} + 2(k-2)\int_{2\sqrt{\ell}}^\infty (2t)^{k-3}e^{-t^2}\mathrm{d}t.$$

Unfolding the recurrence, we observe that as ℓ grows, $\int_{2\sqrt{\ell}}^{\infty}(2t)^{k-1}e^{-t^2}\mathrm{d}t$ is dominated by $(4\sqrt{\ell})^{k-1}$ which in turn is dominated by $(2c\sqrt{\ell})^k$. Combining everything, it follows that for every $k \in \mathbb{N}$ there exists a constant $c \in \mathbb{R}_{>0}$, up to renaming, such that $\mathbb{E}[\lambda^{-k}\|\mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z))\|^k] \leq c\ell^{k/2}$. □

The practicality of Lemma A.16 relies on the Lipschitz constant λ satisfying $\lim_{\ell\to\infty} \lambda\sqrt{\ell} = 0$. Under this condition, we may prove Theorem A.3.

Theorem A.3 (Theorem 3.3 from the main text)**.** *Let $\tau \in \mathbb{R}_{>0}$, $z \in \mathbb{H}$ and $D^{(\tau)}$ be the perturbation matrix in (33). Under Assumption 6, assume that the mapping $g \in (\mathbb{R}^\gamma, \|\cdot\|_2) \mapsto \mathcal{S}((L(g) - z\Lambda - i\tau I_\ell)^{-1}) \in (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_2)$ is λ-Lipschitz with respect to the operator norm. Then, there exists an absolute constant $c \in \mathbb{R}_{>0}$ such that*

$$\|D^{(\tau)}\| \leq c\tau^{-1}\sqrt{\ell}\lambda + \tau^{-2}\|\tilde{\mathcal{S}}\| + \|\Delta(L,\tau)\|.$$

Proof. By (38),

$$\begin{aligned}\|D^{(\tau)}\| \leq{}& \|\mathbb{E}[\mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z))(L - z\Lambda - i\tau I_\ell)^{-1}]\| \\ &+ \|\mathbb{E}[\tilde{\mathcal{S}}((L - z\Lambda - i\tau I_\ell)^{-1})(L - z\Lambda - i\tau I_\ell)^{-1}]\| + \|\Delta(L,\tau)\|.\end{aligned}$$

By Jensen's inequality, submultiplicativity of the operator norm, Lemma A.2 and Lemma A.16,

$$\|\mathbb{E}[\mathcal{S}((L - z\Lambda - i\tau I_\ell)^{-1} - F^{(\tau)}(z))(L - z\Lambda - i\tau I_\ell)^{-1}]\| \leq c\tau^{-1}\sqrt{\ell}\lambda.$$

Similarly,

$$\|\mathbb{E}[\tilde{\mathcal{S}}((L - z\Lambda - i\tau I_\ell)^{-1})(L - z\Lambda - i\tau I_\ell)^{-1}]\| \leq \tau^{-2}\|\tilde{\mathcal{S}}\|$$

The theorem follows. □

As a direct outcome of Theorem A.3, it follows that $\|D^{(\tau)}\|$ tends towards zero as the dimension ℓ approaches infinity under the conditions $\lim_{\ell\to\infty}\sqrt{\ell}\lambda = 0$, $\lim_{\ell\to\infty}\|\tilde{\mathcal{S}}\| = 0$ and $\lim_{\ell\to\infty}\|\Delta(L,\tau)\| = 0$ for every $\tau \in \mathbb{R}_{>0}$ small enough. In our application to the test error of random features ridge regression, we upper bound the Lipschitz constant λ in Theorem A.3 by $\lambda \leq \tau^{-2}\|\mathcal{S}\|_{F\to 2}\lambda_{\mathcal{C}}$ where $\|\mathcal{S}\|_{F\to 2}$ denote the operator norm of the map $\mathcal{S} : (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_F) \mapsto (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_2)$ and $\lambda_{\mathcal{C}}$ is the Lipschitz constant associated with the map $\mathcal{C} : (\mathbb{R}^\gamma, \|\cdot\|_2) \mapsto (\mathbb{R}^{\ell\times\ell}, \|\cdot\|_F)$. Then, $\lim_{\ell\to\infty}\sqrt{\ell}\lambda = 0$ follows from $\|\mathcal{S}\|_{\|\cdot\|_F\to\|\cdot\|_2} \lesssim \ell^{-\frac{1}{2}}$ and $\lambda_{\mathcal{C}} \lesssim \ell^{-\frac{1}{2}}$.

It is trivial to control $\|\Delta(L,\tau)\|$ when L has Gaussian entries. Alternatively, an interpolation approach based on cumulant bounds in the spirit of [LP09, Proposition 3.1] appears to be a suitable avenue to extend the result to other distributions. In Appendix B.2, we employ a leave-one-out strategy to demonstrate that $\|\Delta(L,\tau)\|$ is vanishing in ℓ for every $\tau \in \mathbb{R}_{>0}$.

The culmination of Theorem A.2 and Theorem A.3 along with these specified conditions signifies that $M(z)$ becomes a deterministic equivalent for the expected pseudo-resolvent $\mathbb{E}(L - z\Lambda)^{-1}$ across all $z \in \mathbb{H}$.

Corollary A.3. *Let $z \in \mathbb{H}$ and λ be defined as in Theorem A.3. Under Assumptions 4 to 6, suppose that $\lim_{\ell\to\infty}\sqrt{\ell}\lambda = \lim_{\ell\to\infty}\|\tilde{\mathcal{S}}\| = \lim_{\ell\to\infty}\|\Delta(L,\tau)\| = 0$ for every $\tau \in \mathbb{R}_{>0}$ small enough. Then, $\|\mathbb{E}(L - z\Lambda)^{-1} - M(z)\| \xrightarrow{\ell\to\infty} 0$.*

In certain scenarios, it is feasible to alleviate the reliance of Corollary A.3 on Assumption 6. Notably, by employing a universality result such as the one presented in [BvH23, Lemma 6.11], one may directly argue that certain functionals of resolvent of random matrices do not depend on the distribution of the input.

The utilization of the operator norm in Theorem A.3 is a direct consequence of our previous decision to work with the expected regularized pseudo-resolvent while deferring the concentration step. Intuitively, we anticipate that the expected pseudo-resolvent converges pointwise in operator norm to the solution of the MDE. However, the concentration of the resolvent around its expectation is typically only valid in the context of generalized trace entries.

A.5.2. Concentration

The only remaining task is to establish that the expected pseudo-resolvent is itself a deterministic equivalent for the true pseudo-resolvent — a widely acknowledged fact that stems from a variety of methodologies. We present one such result, based on the assumptions used above.

Lemma A.17 (Lemma 3.2 from the main text)**.** *Let $U \in \mathbb{C}^{\ell\times\ell}$ with $\|U\|_F \leq 1$ and assume that the map $g \in (\mathbb{R}^\gamma, \|\cdot\|_2) \mapsto (L(g) - z\Lambda)^{-1} \in (\mathbb{C}^{\ell\times\ell}, \|\cdot\|_F)$ is λ-Lipschitz with $\lambda \asymp \ell^{-r}$ for some $r > 0$. Under Assumption 6,* $\operatorname{tr}(U((L - z\Lambda)^{-1} - \mathbb{E}(L - z\Lambda)^{-1})) \xrightarrow[\ell\to\infty]{a.s.} 0$.

Proof. By Cauchy-Schwarz inequality, $|\operatorname{tr} U((L(g_1)-z\Lambda)^{-1}-(L(g_2)-z\Lambda)^{-1})| \leq \|U\|_F\|(L(g_1)-z\Lambda)^{-1} - (L(g_2) - z\Lambda)^{-1}\|_F \leq \lambda\|g_1 - g_2\|$ for every $g_1, g_2 \in \mathbb{R}^\gamma$. Hence, by Proposition A.2, there exists absolute constants $c_1, c_2 \in \mathbb{R}_{>0}$ such that

$$\mathbb{P}\left(\operatorname{tr}\left(U\left((L - z\Lambda)^{-1} - \mathbb{E}(L - z\Lambda)^{-1}\right)\right) \geq t\right) \leq c_1 e^{-\frac{c_2 t^2}{\lambda^2}}.$$

Assuming that $\lambda^2 \leq c_3\ell^{-r}$ for some $r \in \mathbb{R}_{>0}$, we may take an infinite sum over ℓ and obtain

$$\sum_{\ell=1}^{\infty}\mathbb{P}\left(\operatorname{tr}\left(U\left((L - z\Lambda)^{-1} - \mathbb{E}(L - z\Lambda)^{-1}\right)\right) \geq t\right) \leq c_1\sum_{\ell=1}^{\infty} e^{-c_4 t^2\ell^r} < \infty$$

Thus, $\operatorname{tr} U((L - z\Lambda)^{-1} - \mathbb{E}(L - z\Lambda)^{-1}) \to 0$ as $\ell \to \infty$ by Borel–Cantelli lemma. □

Combining Corollary A.3 and Lemma A.17 through the utilization of Von Neumann's trace inequality, we derive the ensuing anisotropic law, presented here for the sake of comprehensiveness.

Corollary A.4. *Under the settings of Corollary A.3 and lemma A.17,* $\operatorname{tr}(U((L - z\Lambda)^{-1} - M(z))) \xrightarrow[\ell\to\infty]{a.s.} 0$ *for every $U \in \mathbb{C}^{\ell\times\ell}$ with $\|U\|_* \leq 1$.*

Appendix B: Empirical test error of random features ridge regression

In this section of the supplement, we conclude the proof of our main theorem, which establishes an asymptotically exact expression for the empirical test error of random features ridge regression. For a detailed presentation of the theorem and related discussions, we refer the reader to the main paper. Within this supplement section, we first provide a brief recap of the settings. Subsequently, we demonstrate that $\|\Delta(L, \tau; z)\|$, as defined in (36), converges to zero as $n \to \infty$ for every regularization parameter $\tau \in \mathbb{R}_{>0}$ (Lemma 4.1 from the main text). Finally, leveraging the stability property of the matrix Dyson equation, we then establish that Assumption 4 holds (Lemma 4.4 from the main text).

B.1. Settings

Let $A = n^{-\frac{1}{2}}\sigma(XW) \in \mathbb{R}^{n_{\text{train}}\times d}$, where $X \in \mathbb{R}^{n_{\text{train}}\times n_0}$ is a deterministic matrix, $W \in \mathbb{R}^{n_0\times d}$ is a random matrix, $\delta \in \mathbb{R}_{>0}$ is the ridge parameter, and σ is a λ_σ-Lipschitz activation function. Similarly, let $\widehat{A} = n^{-\frac{1}{2}}\sigma(\widehat{X}W) \in \mathbb{R}^{n_{\text{test}}\times d}$, where $\widehat{X} \in \mathbb{R}^{n_{\text{test}}\times n_0}$ is a deterministic matrix. Following the setup of [LLC18], we assume that $W = \varphi(Z)$ for some $Z \in \mathbb{R}^{n_0\times d}$ with independent standard normal entries, and φ is a λ_φ-Lipschitz function. The Lipschitz constants λ_σ and λ_φ are required to be independent of the problem dimension. As $n \to \infty$ with $n_{\text{train}} \propto n_{\text{test}} \propto n_0 \propto d$, we expect $\limsup_{n\to\infty}(\lambda_\varphi \vee \lambda_\sigma) < \infty$. Additionally, we stipulate that $\limsup_{n\to\infty} \max\{\|X\|, \|\widehat{X}\|\} < \infty$ and similarly for the label vectors y and $\widehat{y}$.

We assume that $\{(a_j^T, \hat{a}_j^T)^T\}_{j=1}^d$, representing the columns of A and $\widehat{A}$, are i.i.d. random vectors with

$$\mathbb{E}[(a_1^T, \hat{a}_1^T)^T] = 0 \quad \text{and} \quad \mathbb{E}[(a_1^T, \hat{a}_1^T)^T(a_1^T, \hat{a}_1^T)] = \begin{bmatrix} K_{AA^T} & K_{A\widehat{A}^T} \\ K_{\widehat{A}A^T} & K_{\widehat{A}\widehat{A}^T} \end{bmatrix}.$$

Here, K_{AA^T}, $K_{A\widehat{A}^T}$, $K_{\widehat{A}A^T}$, and $K_{\widehat{A}\widehat{A}^T}$ encode the covariance between the entries of A and $\widehat{A}$. In addition to assuming that the random features matrices A and $\widehat{A}$ are centered, we also assume that both $\|A\|$ and $\|\widehat{A}\|$ are bounded in L^4.

To summarize, we make the following assumption throughout this section.

Assumption 7. Suppose that $n_{\text{train}}, d, n_{\text{test}}, n_0 \propto n$ and $\limsup_{n\to\infty} \|X\| \vee \|\widehat{X}\| \vee \|y\| \vee \|\widehat{y}\| \vee \mathbb{E}[\|A\|^4] \vee \mathbb{E}[\|\widehat{A}\|^4] \vee \lambda_\sigma \vee \lambda_\varphi < \infty$. Furthermore, suppose that A and $\widehat{A}$ are centered.

B.2. Proof of Lemma 4.1 from the main text

As part of the proof of our main theorem, we use our matrix Dyson equation framework to derive an anisotropic global law for the pseudo-resolvent $(L - z\Lambda)^{-1}$ with $\Lambda := \text{BlockDiag}\{I_{n_{\text{train}}+d}, 0_{2n_{\text{test}}\times 2n_{\text{test}}}\}$ and linearization

$$L = \begin{bmatrix} \delta I_{n_{\text{train}}} & A & 0_{n_{\text{train}}\times n_{\text{test}}} & 0_{n_{\text{train}}\times n_{\text{test}}} \\ A^T & -I_{d\times d} & 0_{d\times n_{\text{test}}} & \widehat{A}^T \\ 0_{n_{\text{test}}\times n_{\text{train}}} & 0_{n_{\text{test}}\times d} & 0_{n_{\text{test}}\times n_{\text{test}}} & -I_{n_{\text{test}}} \\ 0_{n_{\text{test}}\times n_{\text{train}}} & \widehat{A} & -I_{n_{\text{test}}} & 0_{n_{\text{test}}\times n_{\text{test}}} \end{bmatrix} \in \mathbb{R}^{\ell\times\ell}. \tag{39}$$

In the application of our framework, it is crucial to demonstrate that $\|\Delta(L, \tau; z)\|$, as defined in (36), tends to zero as $n \to \infty$ for every regularization parameter $\tau \in \mathbb{R}_{>0}$. This task is the focus of this section.

Fix $z \in \mathbb{H}$, let $\{a_j\}_{j=1}^d$, $\{\hat{a}_j\}_{j=1}^d$ denote the columns of A and $\widehat{A}$ respectively. Suppose that $l_j^T = (a_j^T, 0, 0, \hat{a}_j^T)$ and $L_j = l_j e_{n_{\text{train}}+j}^T + e_{n_{\text{train}}+j} l_j^T$ for every $j \in \{1, 2, \ldots, d\}$, where $\{e_j\}_{j=1}^\ell$ is the canonical basis of $\mathbb{R}^\ell$. In particular, we may write $L = \mathbb{E}L + \sum_{j=1}^d L_j$. For every $j \in \{1, 2, \ldots, d\}$, let $P_j \in \mathbb{R}^{\ell\times\ell}$ be the orthogonal matrix permuting the first and $n_{\text{train}} + j$th entries exclusively and $C_j \in \mathbb{R}^{(\ell-1)\times(\ell-1)}$ be the matrix cycling from position $n_{\text{train}} + j - 1$ to 1. For instance, if $v = (v_k)_{k=1}^{\ell-1}$, then

$$v^T C_j^{-1} = (v_2, v_3, \ldots, v_{j-1}, v_1, v_j, v_{j+1}, \ldots, v_{\ell-1}).$$

We will rely heavily on a Schur complement decomposition of $(P_j L P_j - zI_\ell)^{-1}$. For every $j \in \{1, 2, \ldots, d\}$, let $l_{-j} \in \mathbb{R}^{\ell-1}$ be obtained by removing the $n_{\text{train}} + j$th entry of l_j and

$L_{-j} \in \mathbb{R}^{(\ell-1)\times(\ell-1)}$ be obtained by removing the $n_{\text{train}} + j$th columns and $n_{\text{train}} + j$th row from L. Define the scalar $\xi_j := (1 + z + l_{-j}^T (L_{-j} - zI_{\ell-1})^{-1} l_{-j})^{-1}$ and the matrix

$$\Xi_j := C_j (L_{-j} - zI_{\ell-1})^{-1} C_j - \xi_j C_j (L_{-j} - zI_{\ell-1})^{-1} l_j l_j^T (L_{-j} - zI_{\ell-1})^{-1} C_j.$$

We have the following block inversion formula.

Lemma B.1. *For every* $j \in \{1, 2, \ldots, d\}$ *and* $z \in \mathbb{H}$*,*

$$(P_j L P_j - zI_\ell)^{-1} = \begin{bmatrix} -\xi_j & \xi_j l_{-j}^T (L_{-j} - zI_{\ell-1})^{-1} C_j \\ \xi_j C_j (L_{-j} - zI_{\ell-1})^{-1} l_{-j} & \Xi_j \end{bmatrix}.$$

Proof. The lemma follows directly from the observation

$$P_j L P_j = \begin{bmatrix} -1 & l_{-j}^T C_j^{-1} \\ C_j^{-1} l_{-j} & C_j^{-1} L_{-j} C_j^{-1} \end{bmatrix}$$

and an application of the block matrix inversion lemma. □

For every $j \in \{1, 2, \ldots, d\}$, let $q_j = l_{-j}^T R_{-j} l_{-j}$ and $R_{-j} := (L_{-j} - zI_\ell)^{-1}$. Concentration of bilinear forms is a central ingredient of many random matrix theory proof. We obtain a concentration result for q_j by adapting [LLC18, Lemma 4].

Lemma B.2. *Under Assumption 7,* $\lim_{n\to\infty} \mathbb{E}[\max_{1\le j\le d} |q_j - \mathbb{E}q_j|^2] = 0$ *for every* $z \in \mathbb{H}$*.*

Proof. Adapting [LLC18, Lemma 4], there exists some absolute constants $c_1, c_2 \in \mathbb{R}_{>0}$ such that

$$\mathbb{P}\left(|q_j - \mathbb{E}q_j| > t\right) \le c_1 e^{-c_2 n \min\{t, t^2\}}$$

for every $t \in \mathbb{R}_{\ge 0}$. Then,

$$\mathbb{E}[\max_{1\le j\le d} |q_j - \mathbb{E}q_j|^2] \le n^{-\frac{1}{2}} + \int_{n^{-\frac{1}{2}}}^{1} \mathbb{P}(\max_{1\le j\le d} |q_j - \mathbb{E}q_j|^2 > t)\mathrm{d}t + \int_{1}^{\infty} \mathbb{P}(\max_{1\le j\le d} |q_j - \mathbb{E}q_j|^2 > t)\mathrm{d}t$$

Using a union bound,

$$\int_{n^{-\frac{1}{2}}}^{1} \mathbb{P}(\max_{1\le j\le d} |q_j - \mathbb{E}q_j|^2 > t)\mathrm{d}t \le c_1 n \int_{n^{-\frac{1}{2}}}^{1} e^{-c_2 n t}\mathrm{d}t = \frac{c_1}{c_2}\left(e^{-c_2\sqrt{n}} - e^{-c_2 n}\right)$$

Also,

$$\begin{aligned} \int_{1}^{\infty} \mathbb{P}(\max_{1\le j\le d} |q_j - \mathbb{E}q_j|^2 > t)\mathrm{d}t \le c_1 n \int_{1}^{\infty} e^{-c_2 n\sqrt{t}}\mathrm{d}t = 2c_1 n \int_{1}^{\infty} t e^{-c_2 n t}\mathrm{d}t \\ = \frac{2c_1}{c_2} e^{-c_2 n}(1 + \frac{1}{c_2 n}) \end{aligned}$$

Taking $n \to \infty$ concludes the proof. □

We need one additional tool in order to show universality, which we state here. We omit the proof, as it follows directly from Hölder's inequality.

Lemma B.3. *If* $\limsup_{n\to\infty} \max\{\mathbb{E}[\|A\|^4], \mathbb{E}[\|\widehat{A}\|^4]\} < \infty$ *then* $\limsup_{n\to\infty} \mathbb{E}[\|L - \mathbb{E}L\|^4] < \infty$*.*

We are ready to show universality.

Lemma B.4 (Lemma 4.1 from the main text)**.** *Fix $z \in \mathbb{H}$ and $\tau \in \mathbb{R}_{>0}$. Under Assumption 7, $\lim_{\ell\to\infty} \|\Delta(L,\tau)\| = 0$.*

Proof. For simplicity, we aim to demonstrate that $\lim_{n\to\infty} \|\mathbb{E}[(L-\mathbb{E}L)(L-zI_\ell)^{-1}] + \mathbb{E}[(\tilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}(\tilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}]\| = 0$ for every $z \in \mathbb{H}$. This adjustment streamlines notation without altering any steps in the proof. For every $j \in \{1,2,\dots,d\}$,

$$P_jL_jP_j = \begin{bmatrix} 0 & l_{-j}^T C_j^{-1} \\ C_j^{-1} l_{-j} & 0 \end{bmatrix}$$

and, by Lemma B.1,

$$\begin{aligned}
\mathbb{E}\left[(L-\mathbb{E}L)(L-zI_\ell)^{-1}\right] &= \sum_{j=1}^d P_j \mathbb{E}\left[P_jL_jP_j(P_jLP_j - zI_\ell)^{-1}\right]P_j \\
&= \sum_{j=1}^d P_j \mathbb{E}\begin{bmatrix} \xi_j l_{-j}^T R_{-j} l_{-j} & l_{-j}^T R_{-j} C_j - \xi_j l_{-j}^T R_{-j} l_{-j} l_{-j}^T R_{-j} C_j \\ -\xi_j C_j^{-1} l_{-j} & \xi_j C_j^{-1} l_{-j} l_{-j}^T R_{-j} C_j \end{bmatrix} P_j \\
&= \sum_{j=1}^d P_j \mathbb{E}\begin{bmatrix} \xi_j l_{-j}^T R_{-j} l_{-j} & -\xi_j l_{-j}^T R_{-j} l_{-j} l_{-j}^T R_{-j} C_j \\ -\xi_j C_j^{-1} l_{-j} & \xi_j C_j^{-1} l_{-j} l_{-j}^T R_{-j} C_j \end{bmatrix} P_j
\end{aligned}$$

where we recall that $R_{-j} = (L_{-j} - zI_{\ell-1})^{-1}$. On the other hand,

$$\begin{aligned}
&\mathbb{E}\left[(\widetilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}(\widetilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}\right] \\
&= \sum_{j=1}^d P_j \mathbb{E}\left[P_j\widetilde{L}_jP_j(P_jLP_j - zI_\ell)^{-1}P_j\widetilde{L}_jP_j(P_jLP_j - zI_\ell)^{-1}\right]P_j \\
&= \sum_{j=1}^d P_j \mathbb{E}\begin{bmatrix} \xi_j \tilde{l}_{-j}^T R_{-j} l_{-j} & \tilde{l}_{-j}^T R_{-j} C_j - \xi_j \tilde{l}_{-j}^T R_{-j} l_{-j} l_{-j}^T R_{-j} C_j \\ -\xi_j C_j^{-1} \tilde{l}_{-j} & \xi_j C_j^{-1} \tilde{l}_{-j} l_{-j}^T R_{-j} C_j \end{bmatrix}^2 P_j.
\end{aligned}$$

Thus,

$$\mathbb{E}\left[(L-\mathbb{E}L)(L-zI_\ell)^{-1}\right] + \mathbb{E}\left[(\widetilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}(\widetilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}\right] = \sum_{j=1}^d P_j \mathbb{E}[\xi_j \Psi_j] P_j$$

where $q_j = l_{-j}^T R_{-j} l_{-j}$, $\tilde{q}_j = \tilde{l}_{-j}^T R_{-j} \tilde{l}_{-j}$, $\rho_j = \tilde{l}_{-j}^T R_{-j} l_{-j}$ and

$$\Psi_j = \begin{bmatrix} q_j - \tilde{q}_j + 2\xi_j\rho_j^2 & \rho_j \tilde{l}_{-j}^T R_{-j} C_j - 2\xi_j \rho_j^2 l_{-j}^T R_{-j} C_j + (\tilde{q}_j - q_j) l_{-j}^T R_{-j} C_j \\ -2\xi_j\rho_j C_j^{-1}\tilde{l}_{-j} - C_j^{-1} l_{-j} & C_j^{-1}(l_{-j}l_{-j}^T - \tilde{l}_{-j}\tilde{l}_{-j}^T) R_{-j} C_j + 2\xi_j\rho_j C_j^{-1} \tilde{l}_{-j} l_{-j}^T R_{-j} C_j \end{bmatrix}.$$

We will consider the upper blocks of $\sum_{j=1}^d P_j \mathbb{E}[\xi_j \Psi_j] P_j$ separately.

First, for the upper-left corner, the sum along with the permutation matrices P_j are simply tiling the diagonal. Thus, we may use Cauchy-Schwarz inequality and Jensen's inequality to obtain

$$\begin{aligned}\left\|\sum_{j=1}^{d} P_j \mathbb{E}\begin{bmatrix}\xi_j(q_j-\tilde{q}_j)+2\xi_j^2\rho_j^2 & 0\\ 0 & 0\end{bmatrix} P_j\right\| &\le \max_{1\le j\le d} |\mathbb{E}[\xi_j(q_j-\tilde{q}_j)+2\xi_j^2\rho_j^2]|\\ &\le \frac{2}{\Im[z]}\mathbb{E}[|q-\tilde{q}|]+2|\mathbb{E}[\xi_1^2 l_{-1}^T R_{-1} K R_{-1} l_{-1}]|\\ &\le \frac{4}{\Im[z]}\mathbb{E}[|q-\mathbb{E}q|]+\frac{2\mathbb{E}[\|L-\mathbb{E}L\|^2]\|K\|}{(\Im[z])^4}.\end{aligned}$$

Here, we introduced the correlation matrix $K=\mathbb{E}[l_{-1}l_{-1}^T]$. Using Jensen's inequality and Cauchy Schwarz,

$$\begin{aligned}\|K\| &\le \|\mathbb{E}[a_1a_1^T]\|+\|\mathbb{E}[\hat{a}_1a_1^T]\|+\|\mathbb{E}[a_1\hat{a}_1^T]\|+\|\mathbb{E}[\hat{a}_1\hat{a}_1^T]\|\\ &= d^{-1}(\|\mathbb{E}[AA^T]\|+\|\mathbb{E}[\widehat{A}A^T]\|+\|\mathbb{E}[A\widehat{A}^T]\|+\|\mathbb{E}[\widehat{A}\widehat{A}^T]\|)\\ &\le d^{-1}(\mathbb{E}[\|A\|^2]+2\sqrt{\mathbb{E}[\|\widehat{A}\|^2]\mathbb{E}[\|A\|^2]}+\mathbb{E}[\|\widehat{A}\|^2])\lesssim d^{-1}. \qquad (40)\end{aligned}$$

Here, we used the fact that $\mathbb{E}[\|A\|^2]$ and $\mathbb{E}[\|\widehat{A}\|^2]$ are bounded by assumption. Additionally, by Lemma B.2, it is clear that $\mathbb{E}[|q-\mathbb{E}q|]\to 0$ as $n\to\infty$.

We now turn our attention to the upper-right $1\times(\ell-1)$ corner of $\sum_{j=1}^d P_j\mathbb{E}[\xi_j\Psi_j]P_j$. For every unit vector $x\in\mathbb{C}^\ell$,

$$\begin{aligned}&\left\|\sum_{j=1}^{d} P_j\mathbb{E}\begin{bmatrix}0 & \xi_j\rho_j\tilde{l}_{-j}^T R_{-j}C_j-2\xi_j^2\rho_j^2 l_{-j}^T R_{-j}C_j+\xi_j(\tilde{q}_j-q_j)l_{-j}^T R_{-j}C_j\\ 0 & 0\end{bmatrix}P_j x\right\|_2\\ &=\left\|\mathbb{E}\begin{pmatrix}\begin{pmatrix}0 & \xi_1\rho_1\tilde{l}_{-1}^T R_{-1}C_1-2\xi_1^2\rho_1^2 l_{-1}^T R_{-1}C_1+\xi_1(\tilde{q}_1-q_1)l_{-1}^T R_{-1}C_1\end{pmatrix}P_1x\\ \vdots\\ \begin{pmatrix}0 & \xi_d\rho_d\tilde{l}_{-d}^T R_{-d}C_d-2\xi_d^2\rho_d^2 l_{-d}^T R_{-d}C_d+\xi_d(\tilde{q}_d-q_d)l_{-d}^T R_{-d}C_d\end{pmatrix}P_dx\end{pmatrix}\right\|_2\\ &\le\sqrt{\ell}\max_{1\le j\le d}\|\mathbb{E}[\xi_j\rho_j\tilde{l}_{-j}^T R_{-j}-2\xi_j^2\rho_j^2 l_{-j}^T R_{-j}+\xi_j(\tilde{q}_j-q_j)l_{-j}^T R_{-j}]\|_2.\end{aligned}$$

On one hand, since

$$\mathbb{E}[\xi_j\rho_j\tilde{l}_{-j}^T R_{-j}-2\xi_j^2\rho_j^2 l_{-j}^T R_{-j}]=\mathbb{E}[\xi_j l_{-j}^T R_{-j}KR_{-j}-2\xi_j^2 l_{-j}^T R_{-j}KR_{-j}l_{-j}l_{-j}^T R_{-j}]$$

and $|\xi_j|\le(\Im[z])^{-1}$,

$$\max_{1\le j\le d}\|\mathbb{E}[\xi_j\rho_j\tilde{l}_{-j}^T R_{-j}-2\xi_j^2\rho_j^2 l_{-j}^T R_{-j}]\|\le\frac{\mathbb{E}[\|l_{-1}\|]\|K\|}{(\Im[z])^3}+\frac{2\mathbb{E}[\|l_{-1}\|^3]\|K\|}{(\Im[z])^5}.$$

Furthermore, by Cauchy-Schwarz for complex random variables,

$$\begin{aligned}\|\mathbb{E}[\xi_j(\tilde{q}_j-q_j)l_{-j}^T R_{-j}]\|_2 &= \sup_{\|y\|\le 1}|\mathbb{E}[\xi_j(\tilde{q}_j-q_j)l_{-j}^T R_{-j}y]|\\ &\le(\Im[z])^{-1}\sup_{\|y\|\le 1}\sqrt{\mathbb{E}[|q-\tilde{q}|^2]\mathbb{E}[|l_{-j}^T R_{-j}y|^2]}\\ &=(\Im[z])^{-1}\sup_{\|y\|\le 1}\sqrt{\mathbb{E}[|q-\tilde{q}|^2]\mathbb{E}[y^*R_{-j}^*KR_{-j}y]}\\ &\le\frac{\sqrt{\mathbb{E}[|q-\tilde{q}|^2]\|K\|}}{(\Im[z])^2}.\end{aligned}$$

Combining everything, we obtain that the upper-right $1\times(\ell-1)$ corner of $\sum_{j=1}^d P_j\mathbb{E}[\xi_j\Psi_j]P_j$ is bounded, in norm, by $\sqrt{\ell}\mathbb{E}[\|l_{-1}\|]\|K\|/(\Im[z])^3+2\sqrt{\ell}\mathbb{E}[\|l_{-1}\|^3]\|K\|/(\Im[z])^5+\sqrt{\ell\mathbb{E}[|q-\tilde{q}|^2]\|K\|}/(\Im[z])^2$. We conclude that this bound vanishes as n increases using $\mathbb{E}[\|l_{-1}\|]\le\mathbb{E}[\|L-\mathbb{E}L\|]$, (40) and Lemma B.2.

We consider the two lower blocks together. For notational convenience, let

$$\underline{\Psi}_j=\begin{bmatrix}0 & 0\\ -2\xi_j\rho_jC_j^{-1}\tilde{l}_{-j}-C_j^{-1}l_{-j} & C_j^{-1}(l_{-j}l_{-j}^T-\tilde{l}_{-j}\tilde{l}_{-j}^T)R_{-j}C_j+2\xi_j\rho_jC_j^{-1}\tilde{l}_{-j}l_{-j}^TR_{-j}C_j\end{bmatrix}$$

for every $j\in[\![d]\!]$. Since we expect q_j to concentrate around its mean, we write $\xi_j=(1+z+q_j)^{-1}=(1+z+\mathbb{E}q_j)^{-1}+\frac{\mathbb{E}q_j-q_j}{(1+z+\mathbb{E}q_j)}\xi_j$ and

$$\sum_{j=1}^d P_j\mathbb{E}[\xi_j\underline{\Psi}_j]P_j=(1+z+\mathbb{E}q)^{-1}\sum_{j=1}^d P_j\mathbb{E}[\underline{\Psi}_j]P_j-(1+z+\mathbb{E}q)^{-1}\sum_{j=1}^d P_j\mathbb{E}[(q_j-\mathbb{E}q_j)\xi_j\underline{\Psi}_j]P_j.$$

Using independence of R_{-j}, l_{-j} and $\tilde{l}_{-j}$,

$$\mathbb{E}\underline{\Psi}_j=\begin{bmatrix}0 & 0\\ -2\xi_jC_j^{-1}KR_{-j}l_{-j} & 2\xi_jC_j^{-1}KR_{-j}l_{-j}l_{-j}^TR_{-j}C_j\end{bmatrix}.$$

Using a similar argument as above,

$$\left\|\sum_{j=1}^d P_j\mathbb{E}\begin{bmatrix}0 & 0\\ -2\xi_jC_j^{-1}KR_{-j}l_{-j} & 0\end{bmatrix}P_j\right\|\le\frac{2\sqrt{\ell}\mathbb{E}[\|l_{-1}\|]\|K\|}{(\Im[z])^2}\xrightarrow{n\to\infty}0.$$

Moreover, further decomposing the lower-right corner,

$$\begin{aligned}\sum_{j=1}^d P_j\mathbb{E}\begin{bmatrix}0 & 0\\ 0 & 2\xi_jC_j^{-1}KR_{-j}l_{-j}l_{-j}^TR_{-j}C_j\end{bmatrix}P_j&=(1+z+\mathbb{E}q)^{-1}\sum_{j=1}^d P_j\mathbb{E}\begin{bmatrix}0 & 0\\ 0 & 2C_j^{-1}KR_{-j}KR_{-j}C_j\end{bmatrix}P_j\\&\quad-(1+z+\mathbb{E}q)^{-1}\sum_{j=1}^d P_j\mathbb{E}\begin{bmatrix}0 & 0\\ 0 & 2(q_j-\mathbb{E}q_j)\xi_jC_j^{-1}KR_{-j}l_{-j}l_{-j}^TR_{-j}C_j\end{bmatrix}P_j\end{aligned}$$

with $|(1+z+\mathbb{E}q)^{-1}|\le(\Im[z])^{-1}$,

$$\left\|\sum_{j=1}^d P_j\mathbb{E}\begin{bmatrix}0 & 0\\ 0 & 2C_j^{-1}KR_{-j}KR_{-j}C_j\end{bmatrix}P_j\right\|\le\frac{2d\|K\|^2}{(\Im[z])^2}$$

and

$$\left\|\sum_{j=1}^d P_j\mathbb{E}\begin{bmatrix}0 & 0\\ 0 & 2(q_j-\mathbb{E}q_j)\xi_jC_j^{-1}KR_{-j}l_{-j}l_{-j}^TR_{-j}C_j\end{bmatrix}P_j\right\|\le\frac{2d\|K\|\sqrt{\mathbb{E}[|q-\mathbb{E}q|^2]\mathbb{E}[\|l_{-1}\|^4]}}{(\Im[z])^3}.$$

In particular, $\|(1+z+\mathbb{E}q)^{-1}\sum_{j=1}^d P_j\mathbb{E}[\underline{\Psi}_j]P_j\|\xrightarrow{n\to\infty}0$.

It only remains to show that $\|(1+z+\mathbb{E}q)^{-1}\sum_{j=1}^d P_j\mathbb{E}[(q_j-\mathbb{E}q_j)\xi_j\underline{\Psi}_j]P_j\|$ vanishes. To this end, we undo the decomposition and notice that

$$\sum_{j=1}^d P_j\mathbb{E}[(q_j-\mathbb{E}q_j)\xi_j\underline{\Psi}_j]P_j = \mathbb{E}\left[(\underline{L}-\mathbb{E}\underline{L})\Omega(L-zI_\ell)^{-1}\right] + \mathbb{E}\left[(\underline{\widetilde{L}}-\mathbb{E}\underline{L})\Omega(L-zI_\ell)^{-1}(\widetilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}\right]$$

where

$$\underline{L} = \begin{bmatrix} \delta I_{n_{\text{train}}} & A & 0 & 0 \\ 0 & -I_{d\times d} & 0 & 0 \\ 0 & 0 & 0 & -I_{n_{\text{test}}} \\ 0 & \widehat{A} & -I_{n_{\text{test}}} & 0 \end{bmatrix}$$

and

$$\Omega = \text{BlockDiag}\{0_{n_{\text{train}}\times n_{\text{train}}}, \text{Diag}\{q_j-\mathbb{E}q_j\}_{j=1}^d, 0_{2n_{\text{test}}\times 2n_{\text{test}}}\}.$$

Using the bound $\|(L-zI_\ell)^{-1}\| \leq (\Im[z])^{-1}$, it follows from Jensen's and Cauchy-Schwarz inequalities that

$$\left\|\mathbb{E}\left[(\underline{L}-\mathbb{E}\underline{L})\Omega(L-zI_\ell)^{-1}\right]\right\| \leq \frac{\sqrt{\mathbb{E}[\|L-\mathbb{E}L\|^2]\mathbb{E}[\max_{1\leq j\leq d}|q_j-\mathbb{E}q_j|^2]}}{\Im[z]}$$

and

$$\left\|\mathbb{E}\left[(\underline{\widetilde{L}}-\mathbb{E}\underline{L})\Omega(L-zI_\ell)^{-1}(\widetilde{L}-\mathbb{E}L)(L-zI_\ell)^{-1}\right]\right\| \leq \frac{\sqrt{\mathbb{E}[\|L-\mathbb{E}L\|^4]\mathbb{E}[\max_{1\leq j\leq d}|q_j-\mathbb{E}q_j|^2]}}{(\Im[z])^2}.$$

This term gives us the bottleneck conditions on the norm of the matrix $L-\mathbb{E}L$ and the concentration of q around its mean. By Lemmas B.2 and B.3, both of the RHS bounds vanish as n diverges to infinity. □

Appendix C: Intermediate lemmas

This section introduces some small lemmas that are regularly used throughout this paper.

C.1. General matrix identities

Lemma C.1 (Resolvent trick)**.** *If $W_1, W_2 \in \mathbb{C}^{\ell\times\ell}$ are non-singular, $W_1^{-1}-W_2^{-1} = W_1^{-1}(W_2-W_1)W_2^{-1}$.*

Proof. Multiply on the left by W_1 and on the right by W_2. □

Lemma C.2. *Let $w\in\mathbb{C}$, $W\in\mathbb{C}^{\ell\times\ell}$ and assume that $|w|\geq a>b\geq\|W\|$. Then, $wI_\ell-W$ is non-singular and $\|(wI_\ell-W)^{-1}\|\leq(a-b)^{-1}$.*

Proof. For any $v\in\mathbb{C}^\ell$, we have $\|(wI_\ell-W)v\|\geq\|wv\|-\|Wv\|\geq(a-b)\|v\|$. Therefore, $wI_\ell-W$ is invertible. Letting $v=(wI_\ell-W)^{-1}$ gives the result. □

Lemma C.3. *For every commensurable matrices $W_1, W_2^T\in\mathbb{C}^{n\times d}$ and $z\in\mathbb{C}\setminus\{0\}$ with $z\notin\sigma(W_1W_2)\cup\sigma(W_2W_1)$, $W_1(W_2W_1-zI_d)^{-1} = (W_1W_2-zI_n)^{-1}W_1$.*

Proof. Left-multiply the equation on both sides by $W_1W_2 - zI_n$ and right-multiply by $W_2W_1 - zI_n$. □

Lemma C.4. *For every $z \in \mathbb{H}$ and hermitian matrix $H \in \mathbb{R}^{\ell\times\ell}$, $\|(H - zI_\ell)^{-1}\| \leq (\Im[z])^{-1}$.*

C.1.1. Real and imaginary parts of matrices

A recurrent theme in this document is the use of matrix real and imaginary part. Given a matrix $W \in \mathbb{C}^{\ell\times\ell}$, we decompose $W = \Re[W] + i\Im[W]$ where $2\Re[W] = W + W^*$ and $2i\Im[W] = W - W^*$.

Lemma C.5. $\max\{\|\Re[W]\|, \|\Im[W]\|\} \leq \|W\|$ *for every $W \in \mathbb{C}^{\ell\times\ell}$.*

Proof. Let $v \in \mathbb{C}^\ell$ be unitary. By Cauchy-Schwarz's inequality, $\|Wv\|^2 = \|Wv\|^2\|v\|^2 \geq |v^*Wv|^2 = (v^*\Re[W]v)^2 + (v^*\Im[W]v)^2$. Since both $\Re[W]$ and $\Im[W]$ are Hermitian by definition, $\sup_{\|v\|=1} |v^*\Re[W]v| = \|\Re[W]\|$ and $\sup_{\|v\|=1} |v^*\Im[W]v| = \|\Im[W]\|$ respectively, where the supremum is taken over unitary vectors $v \in \mathbb{C}$. Taking the supremum in the equation above and using the definition of operator norm concludes the proof. □

Corollary C.1. *A matrix $W \in \mathbb{C}^{\ell\times\ell}$ is non-singular whenever either $\Re[W]$ or $\Im[W]$ is non-singular.*

Lemma C.6. *Assume that $W \in \mathbb{C}^{\ell\times\ell}$ is invertible. Then, $\Re[W^{-1}] = W^{-1}\Re[W]W^{-*}$ and $\Im[W^{-1}] = -W^{-1}\Im[W]W^{-*}$.*

Proof. Write $W = \Re[W] + i\Im[W]$. By Lemma C.1 and the definition of matrix imaginary part, $2i\Im[W^{-1}] = W^{-1} - W^{-*} = W^{-1}(W^* - W)W^{-*}$. Since $W^* = \Re[W] - i\Im[W]$, $W^* - W = -2i\Im[W]$. A similar argument applies to the real part. □