Computing asymptotic eigenvectors and eigenvalues of perturbed symmetric matrices

Konstantin Usevich, Simon Barthelme

COMPUTING ASYMPTOTIC EIGENVECTORS AND EIGENVALUES OF PERTURBED SYMMETRIC MATRICES*

KONSTANTIN USEVICH† AND SIMON BARTHELMÉ‡

Abstract. Computing the eigenvectors and eigenvalues of a perturbed matrix can be remarkably difficult when the unperturbed matrix has repeated eigenvalues. In this work we show how the limiting eigenvectors and eigenvalues of a symmetric matrix $\boldsymbol{K}(\varepsilon)$ as $\varepsilon \to 0$ can be obtained relatively easily from successive Schur complements, provided that the entries scale in different orders of ε. If the matrix does not directly exhibit this structure, we show that putting the matrix into a "generalised kernel form" can be very informative. The resulting formulas are much simpler than classical expressions obtained from complex integrals involving the resolvent.

We apply our results to the problem of computing the eigenvalues and eigenvectors of kernel matrices in the "flat limit", a problem that appears in many applications in statistics and approximation theory. In particular, we prove a conjecture from [SIAM J. Matrix Anal. Appl., 2021, 42(1):17–57] which connects the eigenvectors of kernel matrices to multivariate orthogonal polynomials.

Key words. matrix perturbations, kernel matrices, eigenvectors, eigenvalues, flat limit, radial basis functions, tropical algebra

MSC codes. 15A18,15B57,15A80,47A55,47A75,47B34,65F15

1. Introduction.

The original impetus for this work lies in the problem of finding the *flat limit* of kernel matrices [5, 12]. A instance of that problem reads as follows: given a set $\{x_1, \dots, x_n\}$ of points on the line, we form a $n \times n$ matrix

$$\boldsymbol{K}(\varepsilon) = \left[\exp\left(-(\varepsilon(x_i - x_j))^2\right)\right]_{i,j=1}^{n}. \tag{1.1}$$

Such matrices are ubiquitous in physics, statistics, numerical analysis, and other fields. The function $k_\varepsilon(x, y) = \exp(-\varepsilon^2(x-y)^2)$ is called a kernel, here specifically the Gaussian kernel, and the matrix $\boldsymbol{K}(\varepsilon)$ is a kernel matrix. The parameter ε plays the role of an inverse scale parameter: the lower ε is, the slower the kernel function decays as a function of distance. The limit $\varepsilon \to 0$ is therefore called the "flat limit". We are interested in characterising the eigenvectors and eigenvalues of $\boldsymbol{K}(\varepsilon)$ in small ε (see figure 1.1 for an illustration.).

To do so, we may expand $\boldsymbol{K}(\varepsilon)$ as a power series in ε, using the expansion of $\exp(\varepsilon)$ in small ε:

$$\boldsymbol{K}(\varepsilon) = \left[1 - (\varepsilon(x_i - x_j))^2 + \frac{1}{2}(\varepsilon(x_i - x_j))^4 - \dots\right]_{i,j=1}^{n}. \tag{1.2}$$

Equivalently, we may write $\boldsymbol{K}(\varepsilon)$ as:

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{K}_0 + \varepsilon^2 \boldsymbol{K}_2 + \varepsilon^4 \boldsymbol{K}_4 + \dots$$

where $\boldsymbol{K}_0 = \mathbb{1}\mathbb{1}^\top$, and $\boldsymbol{K}_{2m} = \left[\frac{1}{m!}(\varepsilon(x_i - x_j))^{2m}\right]_{i,j=1}^{n}$. We can treat $\boldsymbol{K}(\varepsilon)$ as an *analytic matrix perturbation* in the sense of Kato [19], and hope to extract some information about the small ε regime by treating $\boldsymbol{K}(\varepsilon)$ as a perturbation around $\boldsymbol{K}_0$. Unfortunately, an issue arises immediately. While $\boldsymbol{K}(\varepsilon)$ has full rank for any $\varepsilon > 0$, $\boldsymbol{K}_0$ has rank one and gives us no information about the remaining $n - 1$ eigenvectors and eigenvalues. $\boldsymbol{K}(\varepsilon)$ is an analytic perturbation, but it is a *singular* one and the eigenvectors of the limit tell us little about the limit of the eigenvectors. The problem was tackled in [5], but for eigenvectors the most general result is given only as a conjecture, and, in addition, the proofs are long and convoluted.

For a simpler example of the kind of questions we are trying to address, consider the following matrix in small ε:

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix} + \varepsilon \begin{pmatrix} 1 & 1 & \\ 1 & 1 & \\ & & 0 \end{pmatrix} + \varepsilon^2 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1+\varepsilon+\varepsilon^2 & \varepsilon+\varepsilon^2 & \varepsilon^2 \\ \varepsilon+\varepsilon^2 & \varepsilon+\varepsilon^2 & \varepsilon^2 \\ \varepsilon^2 & \varepsilon^2 & \varepsilon^2 \end{pmatrix} \tag{1.3}$$

*Submitted to the editors DATE.

Funding: This work was supported by the ANR projects MIAI@Grenoble Alpes (ANR-19-P3IA-0003) and LeaFleT (ANR-19-CE23-0021-01).

†Université de Lorraine and CNRS, CRAN (Centre de Recherche en Automatique en Nancy), UMR 7039, Campus Sciences, BP 70239, 54506 Vandœuvre-lès-Nancy cedex, France (konstantin.usevich@cnrs.fr).

‡CNRS, Univ. Grenoble Alpes, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (simon.barthelme@gipsa-lab.fr).

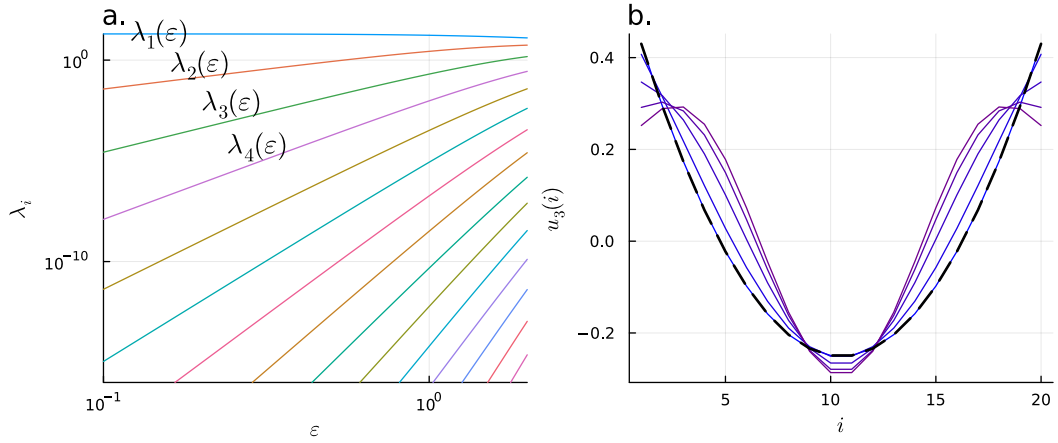

FIG. 1.1. *Eigenvalues and eigenvectors of a kernel matrix in small ε. We take 20 equispaced points $x_1, \dots, x_n$ in $[0,1]$ and form the matrix defined in eq. 1.1.* ***a.*** *The eigenvalues of $\boldsymbol{K}(\varepsilon)$ (computed numerically) as a function of ε, on log-log axes. All eigenvalues except the largest one go to 0 as $\varepsilon \to 0$, and they do so at increasing rate.* ***b.*** *The eigenvector $u_3(i)$ corresponding to $\lambda_3(\varepsilon)$, plotted as a function of i. We show the eigenvector for different values of ε, going from blue to purple as $\varepsilon \to 0$. We see that the eigenvector converges to a well defined limit (shown as the dotted black line) even though the associated eigenvalue goes to 0. Our goal in this paper is to analyse the asymptotic spectral behaviour of matrices that depend analytically on a parameter ε, and are singular at $\varepsilon = 0$. Theorem 5.12 is applicable to the kernel matrix shown here, and provides an expansion for the eigenvalues, as well as an expression for the limiting eigenvectors.*

The limit of $\boldsymbol{K}(\varepsilon)$ as $\varepsilon \to 0$ has rank 1, and therefore a single nonzero eigenpair with $\lambda = 1$ and eigenvector $\boldsymbol{u} = \begin{pmatrix}1 & 0 & 0\end{pmatrix}^\top$. On the other hand, $\boldsymbol{K}(\varepsilon)$ is the sum of three SPD matrices and clearly $\det(\boldsymbol{K}(\varepsilon)) > 0$ for $\varepsilon > 0$. This implies that for all ε, $\boldsymbol{K}(\varepsilon)$ has three non-zero eigenvalues. Two of these eigenvalues must then go to 0 as $\varepsilon \to 0$, and cannot be recovered from $\boldsymbol{A}(0)$. In fact, as we will see (section 3), one eigenvalue goes to zero as $\mathcal{O}(\varepsilon)$, the other as $\mathcal{O}(\varepsilon^2)$.

In this work we describe a set of tools for determining the limiting eigenvectors and eigenvalues of singular perturbations. We restrict ourselves to symmetric matrices, since they are much easier to handle.

We show the following:

1. If the matrix $\boldsymbol{K}(\varepsilon)$ has entries with a particular pattern of orders of magnitude in ε (induced by so-called diagonal scaling), then information about the eigenvalues and eigenvectors can be easily extracted from the Schur complements of the matrix of leading coefficients. These results have a close kinship with the approaches in the tropical algebra literature [1, 3, 16] (in particular, the diagonal scalings we use are strongly related to the so-called "Hungarian scalings" used in tropical algebra [17]).
2. Otherwise, there is a relatively large class of matrices that can be directly "rotated" into a diagonally scaled form, for which all limiting eigenvectors and eigenvalues can be also found from Schur complements. This class of matrices (called "generalised kernel matrices") include kernel matrices [5] as a special case, and in particular we prove the conjecture formulated in [5] on eigenvectors of kernel matrices in the multivariate case (both for unisolvent and non-unisolvent case, thus also generalising results in [25]). All the results from [5] can be obtained using the techniques from this paper, but with simpler and much more straightforward proofs.
3. The generalised kernel form may not be sufficient to characterise all limiting eigenvectors and eigenvalues. In this case, we propose an iterative algorithm, again based on Schur complements, to extract the relevant information at increasing orders of ε.

Compared to classical approaches in analytic perturbation theory we do not use complex integrals at all but rely on basic linear-algebraic tools. Our results use regularised inverses (as in [6]), and Schur complements, which show up in other works as well (for instance, see [8] on differentiability of eigenvectors).

The paper is structured as follows. In section 2 we recall the definition and main properties of analytic matrix perturbations and introduce the notion of *asymptotic spectral equivalent*. The asymptotic spectral equivalent is one of the key notions for this paper, and encodes the information on limiting eigenvalues and eigenvectors. We show how the asymptotic spectral equivalent is linked to regularised inverses. The main results on limiting eigenvalues and eigenvectors are contained in section 3 for diagonally-scaled matrices and in section 4 for matrices in the generalised kernel form. Section 5 contains application of the results to the case of kernel matrices and shows how to treat in a unified way tsmooth and finitely smooth kernels, in the

unisolvent as well as in the non-unisolvent case (thus proving and generalising results and conjectures from [6]). Finally, we discuss in section 6 what can be done in the case when the generalised kernel form is not sufficiently informative.

2. Analytic eigenvalue decompositions and related tools. Matrix perturbation theory is an old and large field, and often surprisingly intricate. Studying general perturbations of general matrices (or worse, operators) can be very difficult. We focus on *symmetric* matrices, which are more tractable. We study analytic matrix perturbations, of the form:

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{K}_0 + \varepsilon \boldsymbol{K}_1 + \varepsilon^2 \boldsymbol{K}_2 + \dots, \tag{2.1}$$

a special case of which is naturally the linear "matrix pencil" $\boldsymbol{K}_0 + \varepsilon \boldsymbol{K}_1$.

Our goal is to characterise the spectral behaviour of $\boldsymbol{K}(\varepsilon)$ as $\varepsilon \to 0$, i.e., what are the eigenvalues and eigenvectors like for small ε? If $\boldsymbol{K}_0$ has full rank, and simple eigenvalues, then the question is easy to answer using regular perturbation theory [9]. If $\boldsymbol{K}_0$ has repeated eigenvalues, for instance when it is rank-deficient (as for kernel matrices), a lot more work is involved. Fortunately, we have the following result, due to Rellich, described in the next subsection.

2.1. Analytic eigenvalue decomposition of symmetric matrices.

THEOREM 2.1 ([22], th. I.1.1). *Let $\boldsymbol{K}(\varepsilon) = \boldsymbol{K}_0 + \varepsilon \boldsymbol{K}_1 + \varepsilon^2 \boldsymbol{K}_2 + \dots$, with $\boldsymbol{K}(\varepsilon)$ symmetric. The eigenvalues $\lambda_1(\varepsilon) \dots \lambda_n(\varepsilon)$ and the corresponding normalized eigenvectors $\boldsymbol{u}_1(\varepsilon), \dots, \boldsymbol{u}_n(\varepsilon)$ (i.e., satisfying $\|\boldsymbol{u}_k(\varepsilon)\|_2$) may be chosen analytic in a (complex) neighbourhood of 0, and satisfy*

Remark 2.2. In matrix notation, the analytic eigenvalue decomposition can be written as

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{U}(\varepsilon) \boldsymbol{\Lambda}(\varepsilon) \boldsymbol{U}^{\mathsf{T}}(\varepsilon),$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1(\varepsilon) & & \\ & \ddots & \\ & & \lambda_n(\varepsilon) \end{pmatrix}, \quad \boldsymbol{U}(\varepsilon) = \begin{pmatrix} \boldsymbol{u}_1(\varepsilon) & \cdots & \boldsymbol{u}_n(\varepsilon) \end{pmatrix}.$$

The orthogonality and normalization of eigenvectors imply that the eigenvector matrix satisfies the constraints

$$\boldsymbol{U}^{\mathsf{T}}(\varepsilon) \boldsymbol{U}(\varepsilon) = \boldsymbol{U}(\varepsilon) \boldsymbol{U}^{\mathsf{T}}(\varepsilon) = \boldsymbol{I}. \tag{2.2}$$

Remark 2.3. Note (2.2) requires that the eigenvectors are of norm 1 for all ε under consideration. This constraint can be relaxed to require that $\boldsymbol{U}^{\mathsf{T}}(\varepsilon) \boldsymbol{U}(\varepsilon)$ is a diagonal matrix (of the form $\boldsymbol{I} + \mathcal{O}(\varepsilon)$), see [15] for a related discussion.

We give an example of the analytic eigenvalue decomposition below.

Example 2.4. Let

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 2\varepsilon^2 + \varepsilon^3 \end{pmatrix}.$$

The eigenvalues of $\boldsymbol{K}(\varepsilon)$ can be found by solving $\det(\boldsymbol{K}(\varepsilon) - \lambda \boldsymbol{I}) = 0$ for z, which in this case works out to:

$$\lambda = \frac{1}{2}(1 + 2\varepsilon + \varepsilon^3 \pm \sqrt{D}), \text{ where } D = 1 - 2\varepsilon^3 + 4\varepsilon^4 + 4\varepsilon^5 + \varepsilon^6,$$

Note that $\sqrt{D}$ expands to $\sqrt{D} = 1 - \varepsilon^3 + \mathcal{O}(\varepsilon^4)$ and is analytic in a neighborhood of 0. Therefore, the eigenvalues are also analytic with the Taylor expansions

$$\lambda_0(\varepsilon) = 1 + \varepsilon^2 + \mathcal{O}(\varepsilon^3), \quad \lambda_1(\varepsilon) = \varepsilon^2 + \varepsilon^3 + \mathcal{O}(\varepsilon^4).$$

As in the rest of this paper, we order eigenvalues such that λ_0 is asymptotically larger than λ_1. The eigenvectors can be found by solving

$$\begin{pmatrix} 1 - \lambda(\varepsilon) & \varepsilon \\ \varepsilon & 2\varepsilon^2 + \varepsilon^3 - \lambda(\varepsilon) \end{pmatrix} \boldsymbol{x} = 0$$

for $\boldsymbol{x}$, where $\lambda(\varepsilon)$ is either of the eigenvalues. For λ_1, if we pick $\boldsymbol{x} = \begin{pmatrix} 1 \\ u \end{pmatrix}$ and solve for u we find $u = \frac{\lambda_1(\varepsilon)-1}{\varepsilon} = \varepsilon + \mathcal{O}(\varepsilon^2)$. For $\lambda_2(\varepsilon)$, if we pick $\boldsymbol{x} = \begin{pmatrix} v \\ 1 \end{pmatrix}$ and solve for v, we find $v = \frac{\lambda_2(\varepsilon)}{\varepsilon} - 2\varepsilon + \varepsilon^3 = -\varepsilon + \mathcal{O}(\varepsilon^2)$. After dividing by the norm of $\boldsymbol{x}$, we get the following system of eigenvectors:

$$\boldsymbol{U}(\varepsilon) = \begin{pmatrix} 1 + \mathcal{O}(\varepsilon) & -\varepsilon + \mathcal{O}(\varepsilon^2) \\ \varepsilon + \mathcal{O}(\varepsilon^2) & 1 + \mathcal{O}(\varepsilon) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \mathcal{O}(\varepsilon),$$

which is indeed analytic near 0.

In our work we will be concerned with finding the limiting eigenvalues and eigenvectors, that is finding $\boldsymbol{U}(0)$, as well as the leading terms in the expansion of $\lambda(\varepsilon)$ (see subsection 2.3 for a precise definition). The classical analytic perturbation theory (see e.g. the book by Kato [19]) provides an exhaustive construction of perturbation series for $\boldsymbol{U}(\varepsilon)$ and $\boldsymbol{\Lambda}(\varepsilon)$, by using the tools of complex analysis and an expansion of the the *resolvent* $\boldsymbol{R}_\varepsilon(z) = (z\boldsymbol{I} - \boldsymbol{K}(\varepsilon))^{-1}$. However, the resulting perturbation series are often complicated and difficult to work with. Moreover, we are dealing with rank-deficient $\boldsymbol{K}(0)$, and all the eigenvalues may have different leading exponents in ε, as shown in the following example. The approach [19] is not well adapted to our case, as it typically proceeds by recursion over the groups of eigenvalues.

In order to find the leading terms, we follow a different approach (related to [1, 20]), as we need only the leading terms in the expansions. The approach consists in bringing the matrix into so-called diagonally scaled form, and then uses regularized inverses and Schur complements, and is described in detail in section 3. To give a preview of the results, we consider the following example, where the strategy from Example 2.4 can no longer be applied.

Example 2.5. Let us study a 5×5 matrix as an example:

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \frac{\varepsilon}{2} & \varepsilon^4 & 0 & 0 \\ \frac{\varepsilon}{2} & \frac{1}{4}\varepsilon^2 & \frac{\varepsilon^2}{2} & 0 & 0 \\ \varepsilon^4 & \varepsilon^2 & \varepsilon^2 & \frac{\varepsilon^3}{2} & 0 \\ 0 & 0 & \frac{\varepsilon^3}{2} & \frac{1}{8}\varepsilon^4 & \frac{\varepsilon^4}{2} \\ 0 & 0 & 0 & \frac{\varepsilon^4}{2} & \varepsilon^4 \end{pmatrix}$$

To compute the eigenvalues of such a matrix, the naïve approach which consists in finding the roots of the characteristic polynomial is no longer possible, since there is no closed-form formula for the roots of a degree 5 polynomial. The tools described in section 3 are applicable however, and tell us that the eigenvalues have expansion:

$$\begin{aligned} &\lambda_1(\varepsilon) = \tilde{\lambda}_1 + \mathcal{O}(\varepsilon), \\ &\lambda_2(\varepsilon) = \varepsilon^2(\tilde{\lambda}_2 + \mathcal{O}(\varepsilon)), \quad \lambda_3(\varepsilon) = \varepsilon^2(\tilde{\lambda}_3 + \mathcal{O}(\varepsilon)), \\ &\lambda_4(\varepsilon) = \varepsilon^4(\tilde{\lambda}_4 + \mathcal{O}(\varepsilon)), \quad \lambda_5(\varepsilon) = \varepsilon^4(\tilde{\lambda}_5 + \mathcal{O}(\varepsilon)) \end{aligned}$$

where $\tilde{\lambda}_1 = 1, \tilde{\lambda}_2 = \frac{1+\sqrt{2}}{2}, \tilde{\lambda}_3 = \frac{1-\sqrt{2}}{2}, \tilde{\lambda}_4 = \frac{9+\sqrt{113}}{16}, \tilde{\lambda}_5 = \frac{9-\sqrt{113}}{16}$. There are thus three groups of eigenvalues: one eigenvalue that does not go to 0 (leading exponent ε^0), 2 eigenvalues that go to 0 at rate ε^2 (leading exponent ε^2), and two other eigenvalues that go to 0 at rate ε^4 (leading exponent ε^4). The matrix $\boldsymbol{U}^{(0)}$ of asymptotic eigenvectors is given by:

$$\boldsymbol{U}(\varepsilon) = \left(\begin{array}{c|cc|cc} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.38 & -0.92 & 0 & 0 \\ 0 & 0.92 & 0.38 & 0 & 0 \\ 0 & 0 & 0 & 0.41 & -0.91 \\ 0 & 0 & 0 & -0.91 & 0.41 \end{array}\right) + \mathcal{O}(\varepsilon),$$

We report numerical values (up to two digits), exact expressions are available but lengthy. The vertical bars separate the three groups of eigenvectors. The calculations are explained in Example 3.13.

2.2. Notation and assumptions. All matrix perturbations $\boldsymbol{K}(\varepsilon)$ considered here are symmetric, real and analytic: $\forall \varepsilon \in \mathbb{R}, \boldsymbol{K}(\varepsilon) \in \mathbb{R}^{n\times n}, \boldsymbol{K}(\varepsilon)^\top = \boldsymbol{K}(\varepsilon)$. We do *not* assume that $\boldsymbol{K}(\varepsilon)$ is positive definite. Our results can be extended to linear operators in Hilbert spaces by treating $\boldsymbol{K}(\varepsilon)$ as a $\infty \times \infty$ pseudo-matrix, but we take n finite for simplicity.

We need some notation related to power series.

DEFINITION 2.6. *Let* $p(\varepsilon) = \sum_{i=0}^{\infty} a_i \varepsilon^i$ *a power series in* ε. *Then:*

- *The* leading term, *noted* $\mathrm{lt}(p)$ *is the first non-zero term*
- *The* leading coefficient, *noted* $\mathrm{lc}(p)$ *is the coefficient of* $\mathrm{lt}(p)$
- *The* valuation, *noted* $\mathrm{val}(p)$ *is the degree of* $\mathrm{lt}(p)$
- *The* leading monomial *is* $\varepsilon^{\mathrm{val}(p)}$
- *The* truncation *of* p *to degree* k *is the series* $\mathrm{trunc}_{\varepsilon^k}(p) = \sum_{i=0}^{k} a_i \varepsilon^i$.

Example 2.7. Let $p(\varepsilon) = 2\varepsilon^3 + 3\varepsilon^5 + \varepsilon^7$. Then $\mathrm{lt}(p) = 2\varepsilon^3$, $\mathrm{lc}(p) = 2$, $\mathrm{val}\, p = 3$, the leading monomial is ε^3 and $\mathrm{trunc}_{\varepsilon^5}(p) = 2\varepsilon^3 + 3\varepsilon^5$.

Some of the facts from [19] are essential and will let us set up notation and assumptions. We assume throughout that $\boldsymbol{K}(\varepsilon)$ has dimension $n \times n$ and is symmetric. Its limit $\boldsymbol{K}(0)$ is rank-deficient with rank $c_0 < n$. In such a case results from [19] tell us that the eigenvalues of $\boldsymbol{K}(\varepsilon)$ have the following behaviour in small ε:

- c_0 eigenvalues have valuation 0 in ε, i.e. an expansion of the form $\lambda(\varepsilon) = \tilde{\lambda} + \mathcal{O}(\varepsilon)$, with $\tilde{\lambda} \neq 0$. These go to the non-zero eigenvalues of $\boldsymbol{K}(0)$ in the limit (i.e. $\tilde{\lambda}$ is a non-zero eigenvalue of $\boldsymbol{K}_0$)
- The other eigenvalues come in groups with increasing valuation; depending on the other terms of $\boldsymbol{K}(\varepsilon)$ as a power series, there may be a group with valuation 1, a group with valuation 2, etc.

We group eigenvalues asymptotically by valuation. We note the valuations $\alpha_0, \alpha_1, \ldots, \alpha_p$, so that there are $p+1$ groups of eigenvalues (generally, $\alpha_0 = 0$). The valuations are increasing: $\alpha_i \leq \alpha_{i-1}$. The number of eigenvalues in group i is denoted c_i. The eigenvalues in group i $(\lambda_{i,k}(\varepsilon), k = 1, \ldots, c_i)$ have expansion

$$\lambda_{i,k} = \varepsilon^{\alpha_i} \left(\tilde{\lambda}_{i,k} + O(\varepsilon) \right), \tag{2.3}$$

where within each group we order eigenvalues in decreasing $\tilde{\lambda}_{i,k}$, so that for small enough ε, $\lambda_{i,k}(\varepsilon) \geq \lambda_{i,k+1}(\varepsilon)$. Note that some of these eigenvalues can be negative.

The eigenvectors expand as $\boldsymbol{U}(\varepsilon) = \boldsymbol{U}^{(0)} + \varepsilon \boldsymbol{U}^{(1)} + \ldots$. Here we are only interested in computing $\boldsymbol{U}^{(0)}$, which we partition as

$$\boldsymbol{U}^{(0)} = \begin{pmatrix} \boldsymbol{U}_0 & \boldsymbol{U}_1 & \ldots & \boldsymbol{U}_p \end{pmatrix} \tag{2.4}$$

according to the eigenvalues they are associated with. $\boldsymbol{U}_i \in \mathbb{R}^{n \times c_i}$ contains the c_i limiting eigenvectors associated with the i-th group of eigenvalues. We also use the following compact notation for the expansion of the i-th group of eigenvalues (2.3) and their leading terms

$$\boldsymbol{\lambda}_i(\varepsilon) = \varepsilon^{\alpha_i} \left(\tilde{\boldsymbol{\lambda}}_i + O(\varepsilon) \right).$$

Example 2.8. In the 5x5 matrix of example Example 2.5, we get

$$(\alpha_0, c_0) = (0, 1), (\alpha_1, c_1) = (2, 2), (\alpha_2, c_2) = (4, 2).$$

meaning that there is 1 eigenvalue with valuation 0, 2 eigenvalues with valuation 2, and 2 eigenvalues of valuation 4.

The notation for the block of eigenvalues becomes

$$\tilde{\boldsymbol{\lambda}}_0 = 1, \tilde{\boldsymbol{\lambda}}_1 = \begin{pmatrix} 2.08 \\ 0.57 \end{pmatrix}, \tilde{\boldsymbol{\lambda}}_2 = \begin{pmatrix} 2.18 \\ 0.77 \end{pmatrix},$$

(truncated to to 2 digits of accuracy) and the eigenvector blocks, respectively

$$\boldsymbol{U}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{U}_1 = \begin{pmatrix} 0 & 0 \\ 0.38 & -0.92 \\ 0.92 & 0.38 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{U}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0.41 & -0.91 \\ -0.91 & 0.41 \end{pmatrix}.$$

2.3. The Asymptotic Spectral Equivalent. We now introduce an operator that preserves all the asymptotic spectral information in a matrix $\boldsymbol{K}(\varepsilon)$. Given $\boldsymbol{K}(\varepsilon)$, we can form another matrix, noted $\underline{\boldsymbol{K}(\epsilon)}$, which we call the "Asymptotic Spectral Equivalent". $\underline{\boldsymbol{K}(\epsilon)}$ is also a matrix perturbation, which shares the asymptotic spectral properties of $\boldsymbol{K}(\varepsilon)$, but whose particular form makes those properties easy to read out.

DEFINITION 2.9 (Asymptotic Spectral Equivalent). *Let $\boldsymbol{K}(\varepsilon) = \boldsymbol{U}(\varepsilon)\boldsymbol{\Lambda}(\varepsilon)\boldsymbol{U}(\varepsilon)^\top$ a (symmetric) analytic matrix perturbation. We define the Asymptotic Spectral Equivalent of $\boldsymbol{K}(\varepsilon)$ as*

$$\underline{\boldsymbol{K}(\epsilon)} \stackrel{def}{=} \boldsymbol{U}^{(0)} \operatorname{lt}(\boldsymbol{\Lambda})\boldsymbol{U}^{(0)} = \sum_{i=0}^{p} \varepsilon^{\alpha_i} \boldsymbol{U}_i \operatorname{diag}(\tilde{\boldsymbol{\lambda}}_i)\boldsymbol{U}_i^\top \tag{2.5}$$

We often write

$$\underline{\boldsymbol{K}(\epsilon)} = \sum_{i=0}^{p} \varepsilon^{\alpha_i} \underline{\boldsymbol{K}}_i \tag{2.6}$$

and the theorems given below provide ways to identify the terms $\underline{\boldsymbol{K}}_0, \underline{\boldsymbol{K}}_1, \dots$. From these terms, and eq. (2.5) it is easy to go back to the asymptotic eigenvalues and eigenvectors of $\boldsymbol{K}(\varepsilon)$. If all the eigenvalues are asymptotically simple (all the $\tilde{\boldsymbol{\lambda}}_i$ distinct within each block), then there is no ambiguity in the eigenvectors either. If the eigenvalues are not asymptotically simple, then there is an ambiguity that can only be lifted by continuing the perturbation series to higher orders.

Let us now list a few properties of the ASE, most of which are very easy to prove.

LEMMA 2.10. *The ASE has the following properties.*

1. *Every term in $\underline{\boldsymbol{K}(\epsilon)}$ is symmetric, i.e. $\underline{\boldsymbol{K}}_i = \underline{\boldsymbol{K}}_i^\top$ for all i.*
2. *The terms in $\underline{\boldsymbol{K}(\epsilon)}$ are orthogonal, $\underline{\boldsymbol{K}}_i^\top \underline{\boldsymbol{K}}_j = 0$ if $i \neq j$.*
3. *If $\boldsymbol{K}(\varepsilon)$ has full rank for $\varepsilon > 0$, then so does $\underline{\boldsymbol{K}(\epsilon)}$.*
4. *Let $\boldsymbol{Q}$ an orthogonal matrix ($\boldsymbol{Q}^{-1} = \boldsymbol{Q}^\top$). Then $\underline{\boldsymbol{Q}^\top \boldsymbol{K}(\varepsilon)\boldsymbol{Q}} = \boldsymbol{Q}^\top \underline{\boldsymbol{K}(\epsilon)} \boldsymbol{Q}$.*

Proof. (1) and (2) follow directly from the definition.

For (3), if $\underline{\boldsymbol{K}(\epsilon)}$ is not full rank, then there is some $\boldsymbol{x}$ such that $\underline{\boldsymbol{K}(\epsilon)}\boldsymbol{x} = 0$. $\boldsymbol{U}^0$ is full rank by construction, so $\boldsymbol{K}(\epsilon)\boldsymbol{x} = 0$ implies that at least one of the eigenvalues is 0 for all ε, which contradicts the hypothesis that $\boldsymbol{K}(\varepsilon)$ is invertible for $\varepsilon > 0$.

(4) follows from applying the change of basis to $\boldsymbol{K}(\varepsilon)$, which leaves the eigenvalues intact but changes $\boldsymbol{U}^{(0)}$ to $\boldsymbol{Q}\boldsymbol{U}^{(0)}\boldsymbol{Q}^\top$. ∎

Example 2.11. Consider the 2×2 matrix studied in Example 2.4, where $\boldsymbol{K}(\varepsilon)$ has eigenvalues $\lambda_0(\varepsilon) = 1 + \mathcal{O}(\varepsilon)$ and $\lambda_1(\varepsilon) = \varepsilon^2(1 + \mathcal{O}(\varepsilon))$, with associated eigenvectors $\begin{pmatrix}1\\0\end{pmatrix} + \mathcal{O}(\varepsilon)$ and $\begin{pmatrix}0\\1\end{pmatrix} + \mathcal{O}(\varepsilon)$.

Therefore

$$\underline{\boldsymbol{K}(\epsilon)} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & \varepsilon^2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} + \varepsilon^2 \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}.$$

Example 2.12. We return to the matrix treated in Examples 2.5 to 2.8. In this case we had three groups of eigenvalues, of order $1, \varepsilon^2, \varepsilon^4$, we get:

$$\underline{\boldsymbol{K}(\epsilon)} = \underline{\boldsymbol{K}}_0 + \varepsilon^2 \underline{\boldsymbol{K}}_1 + \varepsilon^4 \underline{\boldsymbol{K}}_2 = \boldsymbol{U}_0\boldsymbol{U}_0^\top + \varepsilon^2 \boldsymbol{U}_1 \begin{pmatrix}2.08 & 0\\ 0 & 0.57\end{pmatrix} \boldsymbol{U}_1^\top + \varepsilon^4 \boldsymbol{U}_2 \begin{pmatrix}2.18 & 0\\ 0 & 0.77\end{pmatrix} \boldsymbol{U}_2^\top.$$

where all numerical values are truncated to two digits.

Our method consists in obtaining formulas for the ASE, from which asymptotic eigenvalues and eigenvectors can then be read out. In short, for the k-th block the limiting spectral information can be retrieved from $\underline{\boldsymbol{K}}_k$. The following example explain this process as well as the possible ambiguities that arise when there are multiple eigenvalues.

Example 2.13 (Limiting eigenvalues and eigenvectors from ASE). Let the ASE be given by

$$\underline{\boldsymbol{K}(\epsilon)} = \begin{pmatrix}1 & \frac{1}{2} & & & \\ \frac{1}{2} & 1 & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & \ddots\end{pmatrix} + \varepsilon \begin{pmatrix}0 & & & & \\ & 0 & & & \\ & & 2 & & \\ & & & 1 & \\ & & & & \ddots\end{pmatrix} = \underline{\boldsymbol{K}}_0 + \varepsilon \underline{\boldsymbol{K}}_1$$

This tells us that $\boldsymbol{K}(\varepsilon)$ has two groups of eigenvalues, one group of two eigenvalues with valuation 0, one group with two eigenvalues with valuation 2:

- The first group has the expansion $\lambda_{0,i}(\varepsilon) = \tilde{\lambda}_{0,i} + \mathcal{O}(\varepsilon)$ where $i \in \{0,1\}$ and $\tilde{\lambda}_{0,0} = \frac{3}{2}, \tilde{\lambda}_{0,1} = \frac{1}{2}$ are the eigenvalues of $\underline{\boldsymbol{K}}_0$. The corresponding asymptotic eigenvectors are the eigenvectors of $\underline{\boldsymbol{K}}_0$.
- The second group has the expansion $\lambda_{1,i}(\varepsilon) = \varepsilon(\eta_{1,i} + \mathcal{O}(\varepsilon))$ where $i \in \{0,1\}$ and $\tilde{\lambda}_{1,0} = 2, \tilde{\lambda}_{1,1} = 1$ are the eigenvalues of $\underline{\boldsymbol{K}}_1$. The corresponding asymptotic eigenvectors are the eigenvectors of $\underline{\boldsymbol{K}}_1$.

In this case the asymptotic eigenvectors can be retrieved unambiguously from the ASE because there are no eigenvalue multiplicities in $\underline{\boldsymbol{K}}_0$ and $\underline{\boldsymbol{K}}_1$.

If we modify the example to

$$\underline{\boldsymbol{K}(\epsilon)} = \begin{pmatrix} 1 & \frac{1}{2} & & \\ \frac{1}{2} & 1 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} = \underline{\boldsymbol{K}}_0 + \varepsilon \underline{\boldsymbol{K}}_1$$

then $\underline{\boldsymbol{K}}_1$ has repeated eigenvalues and the individual eigenvectors at order ε cannot be resolved. We can only obtain the asymptotic eigenprojector on the subspace spanned by $\boldsymbol{u}_3(0), \boldsymbol{u}_4(0)$, further terms in the expansion of the eigenprojector are necessary to disambiguate the eigenvectors themselves.

2.4. Regularised inverses and their asymptotics. One of the key tools of our proofs are regularized inverses which serve as a probing device to obtain ASEs. The "regularised inverse" of a matrix $\boldsymbol{K}(\varepsilon)$ is the matrix

$$\boldsymbol{M}(z) = \boldsymbol{K}(\varepsilon)(\boldsymbol{K}(\varepsilon) + z\boldsymbol{I})^{-1} = \boldsymbol{I} - (\boldsymbol{K}(\varepsilon) + z\boldsymbol{I})^{-1},$$

defined for those z such that $\boldsymbol{K}(\varepsilon) + z\boldsymbol{I}$ is invertible. Regularised inverses often turn up in the theory of kernel methods in statistics [6].

As we will show later, by scaling z in ε, the regularised inverse can be used to probe eigenvalues that tend to 0 slower than ε^s. The next lemma explains how.

LEMMA 2.14 (Asymptotics of regularised inverses). *Let $\boldsymbol{K}(\varepsilon)$ a symmetric matrix with asymptotic spectral equivalent $\underline{\boldsymbol{K}(\epsilon)} = \sum_{i=0}^{p} \varepsilon^{\alpha_i} \underline{\boldsymbol{K}}_i$. Let $s > 0$, $j = \underset{s.t.\ \alpha_i \leq s}{\operatorname{argmax}}\ i$. Then for any $\tau \notin \{-\tilde{\lambda}_{ij}\}$, the regularised inverse*

$$\boldsymbol{M}_{s,\tau}(\varepsilon) \overset{def}{=} \boldsymbol{M}(\tau\varepsilon^s) \tag{2.7}$$

is well-defined for small ε and its limit at 0 is completely determined by the ASE:

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \begin{cases} \sum_{i=0}^{j} \boldsymbol{U}_i \boldsymbol{U}_i^\top + o(1), & \text{if } \alpha_j < s, \\ \sum_{i=0}^{j-1} \boldsymbol{U}_i \boldsymbol{U}_i^\top + \underline{\boldsymbol{K}}_j (\underline{\boldsymbol{K}}_j + \tau \boldsymbol{I})^{-1} + o(1), & \text{otherwise.} \end{cases} \tag{2.8}$$

Proof. By Theorem 2.1 and from (2.2) we have that

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \boldsymbol{U}(\varepsilon)\boldsymbol{\Lambda}(\varepsilon)(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^s \boldsymbol{I})^{-1}\boldsymbol{U}(\varepsilon)^\top = (\boldsymbol{U}^{(0)} + o(1))\boldsymbol{\Lambda}(\varepsilon)(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^s \boldsymbol{I})^{-1}(\boldsymbol{U}^{(0)} + o(1)), \tag{2.9}$$

so we can focus on the limit of the middle term. The matrix $\boldsymbol{\Lambda}(\varepsilon)(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^s \boldsymbol{I})^{-1}$ is diagonal, and we are going to find the limits of its diagonal elements. Consider an eigenvalue $\lambda_{i,k}(\varepsilon)$ in the i-th group (2.3) . Then, if $\tau \neq -\tilde{\lambda}_{i,k}$, the corresponding diagonal element of $\boldsymbol{\Lambda}(\varepsilon)(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^s \boldsymbol{I})^{-1}$ is equal to

$$\frac{\lambda_{i,k}(\varepsilon)}{\lambda_{i,k}(\varepsilon) + \tau\varepsilon^s} = \begin{cases} 1 + o(1), & \text{if } s > \alpha, \\ \frac{\tilde{\lambda}_{i,k}}{\tau + \tilde{\lambda}_{i,k}} + o(1), & \text{if } s = \alpha, \\ 0 + o(1), & \text{otherwise,} \end{cases} \tag{2.10}$$

where these three cases follow from the power series expansion of $\frac{1}{1+x}$ at 0.

Injecting eq. (2.10) into eq. (2.9), we find the following:

- if s does not match any of the valuations ($\alpha_j \neq s$),

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \sum_{i=0}^{j} \boldsymbol{U}_i (\boldsymbol{I}_{c_i} + \operatorname{diag}(o(1)))\boldsymbol{U}_i^\top + o(1),$$

  where c_i is the size of the block of eigenvalues with valuation α_i;

- if $\alpha_j = s$, then by denoting

$$\widehat{\Lambda}_j = \operatorname{diag}\left(\frac{\tilde{\lambda}_{j,1}}{\tilde{\lambda}_{j,1}+\tau}, \dots, \frac{\tilde{\lambda}_{j,c_j}}{\tilde{\lambda}_{j,c_j}+\tau}\right),$$

we have that

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \sum_{i=0}^{j-1} \boldsymbol{U}_i(\boldsymbol{I}_{c_i} + \operatorname{diag}(o(1)))\boldsymbol{U}_i^\top + \boldsymbol{U}_j(\widehat{\Lambda}_j + o(1))\boldsymbol{U}_j^\top.$$

Elementary calculations show that $\boldsymbol{U}_j\widehat{\Lambda}_j\boldsymbol{U}_j^\top = \underline{\boldsymbol{K}}_j(\underline{\boldsymbol{K}}_j + \tau\boldsymbol{I})^{-1}$, which concludes the proof. ∎

Remark 2.15. Lemma 2.14 shows that regularised inverses have a well-defined limit, and importantly that the scaling in ε can be used as a probing device for finding blocks of eigenvalues of order s. Consider the function

$$r(s) = \operatorname{rank} \lim_{\varepsilon\to 0} \boldsymbol{M}_s(\varepsilon) \tag{2.11}$$

for $s > 0$. The above lemma implies that $r(s)$ is a piecewise constant function, with jumps at $\nu_0, \nu_1, \dots, \nu_p$. Indeed,

$$r(s) = \sum_{i,\alpha_i \le s} c_i \tag{2.12}$$

so that $r(s)$ counts the number of eigenvalues with valuation $\le s$.

Example 2.16. Take

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 2\varepsilon^2 + \varepsilon^3 \end{pmatrix},$$

as in example 2.4. Let us pretend for now that we do not know what the eigenvalues of $\boldsymbol{K}(\varepsilon)$ are. We will see that asymptotic eigenvalues and eigenvectors can be recovered using regularised inverses.

Using the standard formulas for inversion of 2×2 matrices,

$$\boldsymbol{M}(z) = \boldsymbol{K}(\varepsilon)(\boldsymbol{K}(\varepsilon) + z\boldsymbol{I})^{-1} = \boldsymbol{I} - \frac{z}{\det(\boldsymbol{K}(\varepsilon)+z)}\begin{pmatrix} z+2\varepsilon^2+\varepsilon^3 & -\varepsilon \\ -\varepsilon & 1+z \end{pmatrix}$$

where $\frac{z}{\det(\boldsymbol{K}(\varepsilon)+z)} = \frac{z}{(1+z)(z+2\varepsilon^2+\varepsilon^3)-\varepsilon^2}$. For fixed z, direct computation gives

$$\lim_{\varepsilon\to 0} \boldsymbol{M}(z) = \begin{pmatrix} \frac{1}{1+z} & \\ & 0 \end{pmatrix}$$

By lemma 2.14, this implies that one eigenvalue has valuation 0 and leading coefficient 1. Scaling $z = \tau\varepsilon$ and taking the limit again, we obtain:

$$\lim_{\varepsilon\to 0} \boldsymbol{M}(\varepsilon\tau) = \begin{pmatrix} 1 & \\ & 0 \end{pmatrix}$$

By lemma 2.14, no eigenvalue is of valuation 1. Setting $z = \tau\varepsilon^2$ on the other hand gives:

$$\lim_{\varepsilon\to 0} \boldsymbol{M}(\varepsilon^2\tau) = \begin{pmatrix} 1 & \\ & \frac{1}{1+\tau} \end{pmatrix}$$

By lemma 2.14, there is an eigenvalue of valuation 2 with leading term 1, which agrees with the computation in example 2.4.

In this case, computation of the eigenvalues via regularised inverses is more tedious that computing eigenvalues directly. If the matrix has a specific form, as in section 3, going through the regularised inverse is much easier.

We are now ready to present our main results.

3. Matrices in diagonal-scaling form. The reader may have noticed that in many examples we have given so far the entries of the matrix are of different orders: some entries of order 1, some entries of order ε, some entries of order ε^2, etc. When such structure exists, it can be exploited to obtain asymptotic expansions more easily (an idea that appears in the tropical algebra literature, see [1]).

Example 3.1. Take the matrix defined by eq. (1.3)

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1+\varepsilon+\varepsilon^2 & \varepsilon+\varepsilon^2 & \varepsilon^2 \\ \varepsilon+\varepsilon^2 & \varepsilon+\varepsilon^2 & \varepsilon^2 \\ \varepsilon^2 & \varepsilon^2 & \varepsilon^2 \end{pmatrix}$$

Not only are the entries of different magnitudes in ε, but here they are ordered such that the valuation is non-decreasing across rows and columns ($\operatorname{val} K_{i+c,j} \geq \operatorname{val} K_{i,j}$ for $c \geq 0$). We can take advantage of this property to re-express $\boldsymbol{K}(\varepsilon)$ as:

$$\begin{aligned} \boldsymbol{K}(\varepsilon) &= \begin{pmatrix} \varepsilon^0 & & \\ & \varepsilon^{\frac{1}{2}} & \\ & & \varepsilon^1 \end{pmatrix} \begin{pmatrix} 1+\varepsilon+\varepsilon^2 & \varepsilon^{\frac{1}{2}}+\varepsilon^{\frac{3}{2}} & \varepsilon \\ \varepsilon^{\frac{1}{2}}+\varepsilon^{\frac{3}{2}} & 1+\varepsilon & \varepsilon^{\frac{1}{2}} \\ \varepsilon & \varepsilon^{\frac{1}{2}} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon^0 & & \\ & \varepsilon^{\frac{1}{2}} & \\ & & \varepsilon^1 \end{pmatrix} \\ &= \begin{pmatrix} \varepsilon^0 & & \\ & \varepsilon^{\frac{1}{2}} & \\ & & \varepsilon^1 \end{pmatrix} \left(\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} + o(1) \right) \begin{pmatrix} \varepsilon^0 & & \\ & \varepsilon^{\frac{1}{2}} & \\ & & \varepsilon^1 \end{pmatrix} \\ &= \boldsymbol{\Delta}(\varepsilon)(\boldsymbol{I} + \mathcal{O}(\varepsilon))\boldsymbol{\Delta}(\varepsilon). \end{aligned} \tag{3.1}$$

The diagonal matrix $\boldsymbol{\Delta}(\varepsilon) = \operatorname{diag}(\varepsilon^0, \varepsilon^{\frac{1}{2}}, \varepsilon^1)$ that appears in this equation is what we call a "scaling matrix".

For matrices which admit a non-trivial diagonal scaling, we can derive precise results on the asymptotic eigenvalues and eigenvectors (Theorem 3.2). In the next subsection, we set the notation and assumptions and formulate the main result of this section. To make the result more user-friendly, section subsection 3.2 explains how to find and use diagonal scalings in computations and highlight links to tropical algebra. The proof of Theorem 3.2 is presented in subsection 3.3, and relies on regularised inverses.

3.1. ASE of diagonally-scaled matrices. In this section, we address the following case:

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{\Delta}(\varepsilon)(\boldsymbol{H} + o(1))\boldsymbol{\Delta}(\varepsilon) \tag{3.2}$$

where $\boldsymbol{\Delta}(\varepsilon)$ is a scaling matrix. Any matrix perturbation can be put into the form of equation of eq. 3.2, if only under the trivial scaling $\boldsymbol{\Delta}(\varepsilon) = \boldsymbol{I}$. In such a case our theorem will have nothing of much interest to say - it will only describe the eigenvalues and eigenvectors of order ε^0. Our results begin to be interesting if the scaling matrix is non-trivial, which means that $\boldsymbol{K}(\varepsilon)$ needs to have entries with different orders of magnitude in ε.

We need to set up some notation to describe the scaling matrix $\boldsymbol{\Delta}(\varepsilon)$ and the block structure it induces in $\boldsymbol{H}$. We write:

$$\boldsymbol{\Delta}(\varepsilon) = \boldsymbol{\Delta}_{\boldsymbol{b}}(\varepsilon) = \operatorname{diag}(\underbrace{\varepsilon^{\nu_0}, \ldots, \varepsilon^{\nu_0}}_{b_0}, \underbrace{\varepsilon^{\nu_1}, \ldots, \varepsilon^{\nu_1}}_{b_1}, \ldots, \underbrace{\varepsilon^{\nu_p}, \ldots, \varepsilon^{\nu_p}}_{b_p}) \tag{3.3}$$

where $\nu_0 < \nu_1 < \cdots < \nu_r$ and each valuation ν_i (not necessarily integer) is repeated b_i times.

We partition $\boldsymbol{H}$ according to the valuations in $\boldsymbol{\Delta}(\varepsilon)$, as

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_{0,0} & \boldsymbol{H}_{0,1} & \ldots & \boldsymbol{H}_{0,p} \\ \boldsymbol{H}_{1,0} & \boldsymbol{H}_{1,1} & \ldots & \boldsymbol{H}_{1,p} \\ \vdots & \vdots & \ldots & \ldots \\ \boldsymbol{H}_{p,0} & \boldsymbol{H}_{p,1} & \ldots & \boldsymbol{H}_{p,p} \end{pmatrix}, \tag{3.4}$$

with $\boldsymbol{H}_{i,j} \in \mathbb{R}^{b_i \times b_j}$ (b_i are positive integers). For convenience, we define:

- $\boldsymbol{H}_{\leq i, \leq j}$ to be the submatrix of $\boldsymbol{H}$ with row blocks up to i and column blocks up to j;
- $\boldsymbol{H}_{i, \leq j}$ to be the submatrix of the i-th block row

$$\boldsymbol{H}_{i,\leq j} = \begin{pmatrix} \boldsymbol{H}_{0,0} & \boldsymbol{H}_{0,1} & \ldots & \boldsymbol{H}_{0,j} \end{pmatrix},$$

  and similar notation $\boldsymbol{H}_{\leq i,j}$ for the submatrix of the j-th block column;

- shortcuts $\boldsymbol{H}_{<i,<j} = \boldsymbol{H}_{\leq i-1,\leq j-1}$, $\boldsymbol{H}_{i,<j} = \boldsymbol{H}_{i,\leq j-1}$, $\boldsymbol{H}_{<i,j} = \boldsymbol{H}_{\leq i-1,j}$.

Whenever $\boldsymbol{H}_{<i,<i}$ is invertible, we define the i-th Schur complement as

$$\boldsymbol{S}_i = \boldsymbol{H}_{i,i} - \boldsymbol{H}_{i,<i}(\boldsymbol{H}_{<i,<i})^{-1}\boldsymbol{H}_{<i,i}, \tag{3.5}$$

and formally define $\boldsymbol{S}_0 = \boldsymbol{H}_{0,0}$. If there exists a j such that $\boldsymbol{S}_j$ is not invertible, then the sequence of Schur complements stops at this $\boldsymbol{S}_j$ (non-invertibility of $\boldsymbol{S}_j$ implies non-invertibility of $\boldsymbol{H}_{<j+1,<j+1}$).

Armed with the above notation, we can formulate the following theorem.

THEOREM 3.2. *Let $\boldsymbol{K}(\varepsilon)$ be as in eq. (3.2). Assume that the matrix $\boldsymbol{H}_{\leq j,\leq j}$, for $j \leq p$ is invertible (i.e., all Schur complements up to $\boldsymbol{S}_j$ exist). Then:*

1. *the ASE of $\boldsymbol{K}(\varepsilon)$ has the following block-diagonal form*

$$\underline{\boldsymbol{K}}(\varepsilon) = \begin{pmatrix} \varepsilon^{2\nu_0}\boldsymbol{S}_0 & & & \\ & \ddots & & \\ & & \varepsilon^{2\nu_j}\boldsymbol{S}_j & \\ & & & o(\varepsilon^{2\nu_j}) \end{pmatrix}, \tag{3.6}$$

2. *if $\boldsymbol{H}$ is invertible (i.e., $j = p$), there is no $o(\varepsilon^{2\nu_p})$ term and the ASE is completely determined by $\boldsymbol{H}$:*

$$\underline{\boldsymbol{K}}(\varepsilon) = \begin{pmatrix} \varepsilon^{2\nu_0}\boldsymbol{S}_0 & & & \\ & \varepsilon^{2\nu_1}\boldsymbol{S}_1 & & \\ & & \ddots & \\ & & & \varepsilon^{2\nu_p}\boldsymbol{S}_p \end{pmatrix}, \tag{3.7}$$

Proof. The proof of Theorem 3.2 is deferred to subsection 3.3, and uses regularised inverses. ∎

Remark 3.3 (Related results). We know of related, but not equivalent, results in the literature. Tropicalisation of the characteristic polynomial of $\boldsymbol{K}(\varepsilon)$ can be used to lower-bound the valuation of the eigenvalues [2, 1, 3], and obtain the leading coefficients of eigenvalues in certain cases. Schur complements appear in the Lidskii–Vishik–Lyusternik approach to perturbation theory of non-symmetric matrices [20, 21], and in [8].

How to work with and interpret the results of theorem 3.2 will hopefully become clearer with the tools we introduce in section 3.2. Let us give two simple examples of its application (more can be found in the next subsection).

Example 3.4. Let us apply theorem 3.2 to the 2×2 matrix of example 2.4. We first need to re-express the matrix in a diagonal-scaling form:

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 2\varepsilon^2 + \varepsilon^3 \end{pmatrix} = \begin{pmatrix} 1 & \\ & \varepsilon \end{pmatrix} \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} + \mathcal{O}(\varepsilon) \right\} \begin{pmatrix} 1 & \\ & \varepsilon \end{pmatrix}.$$

In the notation used above, the scaling has valuations $\nu_0 = 0$, $\nu_1 = 1$, and $\boldsymbol{H} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. By theorem 3.2, the eigenvalues of $\boldsymbol{K}(\varepsilon)$ come in two "blocks" of valuation $2\nu_0 = 0$ and $2\nu_1 = 2$. To find the leading terms of the eigenvalues, we partition $\boldsymbol{H}$ according to the structure given by the valuations, which here is:

$$\boldsymbol{H} = \left[\begin{array}{c|c} 1 & 1 \\ \hline 1 & 2 \end{array}\right]$$

and compute the sequence of Schur complements:

$$\boldsymbol{S}_0 = H_{0,0} = 1$$

$$\boldsymbol{S}_1 = H_{0,0} - \frac{H_{1,1}}{H_{1,2}^2} = 1.$$

$\boldsymbol{S}_0$ and $\boldsymbol{S}_1$ are of size 1×1 with eigenvalue 1. These blocks are of size 1. The ASE is therefore

$$\underline{\boldsymbol{K}(\epsilon)} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \varepsilon^2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

This result agrees with the calculations performed earlier.

Example 3.5. We apply Theorem 3.2 to the matrix in eq. (3.1). Here $\boldsymbol{H} = \boldsymbol{I}$, $\nu_0 = 0, \nu_1 = \frac{1}{2}, \nu_2 = 1$, $b_0 = b_1 = b_2 = 1$. The Schur complements are $S_0 = S_1 = S_2 = 1$. According to the theorem, we have

$$\underline{\boldsymbol{K}(\epsilon)} = \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 & & \\ & 1 & \\ & & 0 \end{pmatrix} + \varepsilon^2 \begin{pmatrix} 0 & & \\ & 0 & \\ & & 1 \end{pmatrix}$$

We can infer from this expression that $\boldsymbol{K}(\varepsilon)$ has three eigenvalues with expansion $\lambda_0(\varepsilon) = 1 + o(1), \lambda_1(\varepsilon) = \varepsilon + o(1), \lambda_2(\varepsilon) = \varepsilon^2 + o(1)$. The associated limiting eigenvectors are $[1,0,0]^\top, [0,1,0]^\top, [0,0,1]^\top$.

Remark 3.6. Let us explain how the asymptotic eigenvectors and eigenvalues can be recovered from the ASE, as given in eq. (3.6) in the general case (see also ex. 2.13 for generalities on interpreting the ASE in terms of asymptotic eigenvalues and eigenvectors).

Take j the maximum index such that $\boldsymbol{S}_j$ exists, and denote for each $i \leq j$ the block matrix

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{0} & \cdots & \boldsymbol{I}_{b_i} & \cdots & \boldsymbol{0} \end{pmatrix}^\mathsf{T}.$$

Note that in this case eqs. (3.6) and (3.7), can be written as

$$\underline{\boldsymbol{K}}(\varepsilon) = \sum_{i=0}^{j} \varepsilon^{2\nu_i} \boldsymbol{Z}_i \boldsymbol{S}_i \boldsymbol{Z}_i^\mathsf{T} + o(\varepsilon^{2\nu_j}), \quad \text{and} \tag{3.8}$$

$$\underline{\boldsymbol{K}}(\varepsilon) = \sum_{i=0}^{p} \varepsilon^{2\nu_i} \boldsymbol{Z}_i \boldsymbol{S}_i \boldsymbol{Z}_i^\mathsf{T}, \tag{3.9}$$

respectively.

For $i < j$, Theorem 3.2 states that there exists a block of eigenvalues with valuation $2\nu_i$. Since $\boldsymbol{Z}_i$ is orthonormal, the eigenvalues and eigenvectors of the term $\boldsymbol{Z}_i \boldsymbol{S}_i \boldsymbol{Z}_i^\top$ in the ASE can be obtained from the eigenvalues and eigenvectors of $\boldsymbol{S}_i$. By assumption, $i < j$ so that $\boldsymbol{S}_i$ is invertible, and has b_i non-zero eigenvalues $\eta_{i,1} \dots \eta_{i,b_i}$. Then $\boldsymbol{K}(\varepsilon)$ has a block of asymptotic eigenvalues of the form $\lambda(\varepsilon) = \varepsilon^{2\nu_i}(\eta_{i,k} + \mathcal{O}(\varepsilon))$ for $k \in \{1, \dots, b_i\}$. The corresponding eigenvectors can be obtained from the eigenvectors of $\boldsymbol{S}_i$, if all eigenvalues of $\boldsymbol{S}_i$ are simple. If there are repeated eigenvalues in $\boldsymbol{S}_i$, then the asymptotic eigenvectors cannot be identified (a further expansion is needed to make the eigenvalues distinct).

In interpreting the last block j in the expansion of the ASE, we need to be careful. If $j < p$ then the expansion is truncated early, we are in case 1 of the theorem, and the ASE is only identified up to valuation $2\nu_j$. The Schur complement $\boldsymbol{S}_j$ is non-invertible. Its non-zero eigenvalues (and eigenvectors) give the leading coefficients of asymptotic eigenvalues of $\boldsymbol{K}(\varepsilon)$ with valuation $2\nu_j$, and its zero eigenvalues (and the associated null space) correspond to eigenvalues of $\boldsymbol{K}(\varepsilon)$ with valuation strictly higher than $2\nu_j$. If, on the other hand, $\boldsymbol{H}$ is invertible, then the last Schur complement $\boldsymbol{S}_j$ has full rank, and all asymptotic eigenvalues of $\boldsymbol{K}(\varepsilon)$ can be identified using the theorem.

3.2. Theorem 3.2: a user's guide. To make theorem 3.2 more useful in calculations (either by hand or on a computer), let us explain how to easily compute $\boldsymbol{H}$ from $\boldsymbol{K}(\varepsilon)$ given a candidate scaling. The right tools to use come from tropical algebra (see e.g. [18] for an introduction). Fortunately, they are easy to understand and can be described with minimal background.

In computations by hand, it is useful to write down a *valuation matrix* for the entries of $\boldsymbol{K}(\varepsilon)$:

DEFINITION 3.7 (Valuation matrix). *The valuation matrix $\boldsymbol{\Omega}$ of $\boldsymbol{K}(\varepsilon)$ is a matrix with entries in $\mathbb{Z}_+ \cup \{\infty\}$ that contains the element-wise valuations of $\boldsymbol{K}(\varepsilon)$, i.e. $\boldsymbol{\Omega} = [\operatorname{val} K_{ij}(\varepsilon)]_{i,j}$*

Example 3.8. If

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \varepsilon^2 + \varepsilon^3 & 0 \\ \varepsilon^2 + \varepsilon^3 & 5\varepsilon^3 & 0 \\ 0 & 0 & \varepsilon^1 \end{pmatrix}$$

then

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & 2 & \infty \\ 2 & 3 & \infty \\ \infty & \infty & 1 \end{pmatrix}.$$

Recall that $\operatorname{val}(0) = \infty$.

Diagonal scalings need to be designed carefully so that:

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{\Delta}(\varepsilon)(\boldsymbol{H} + o(1))\boldsymbol{\Delta}(\varepsilon) \tag{3.10}$$

as in theorem 3.2. Equation (3.10) implies:

$$\operatorname{val} K_{ij}(\varepsilon) \geq \operatorname{val} \Delta(\varepsilon)_{ii} + \operatorname{val} \Delta(\varepsilon)_{jj} \tag{3.11}$$

This leads to the following definition:

Definition 3.9 (Valid scaling). *We say $\boldsymbol{\Delta}(\varepsilon)$ is a valid scaling for $\boldsymbol{K}(\varepsilon)$ if eq. (3.11) is verified for all i, j. We say that $\boldsymbol{\Delta}(\varepsilon)$ is* tight *at entry i, j if:*

$$\operatorname{val} K_{ij}(\varepsilon) = \operatorname{val} \Delta(\varepsilon)_{ii} + \operatorname{val} \Delta(\varepsilon)_{jj}$$

Given a candidate scaling $\boldsymbol{\Delta}(\varepsilon)$, one can form the matrix $\tilde{\boldsymbol{\Omega}}$ with entries

$$\tilde{\Omega}_{i,j} = \operatorname{val} \Delta(\varepsilon)_{ii} + \operatorname{val} \Delta(\varepsilon)_{jj}$$

Then $\boldsymbol{\Delta}(\varepsilon)$ is a valid scaling iff

$$\boldsymbol{\Omega} - \tilde{\boldsymbol{\Omega}} \geq 0 \tag{3.12}$$

element-wise.

Example 3.10 (Example 3.8, continued). Let us try the candidate scaling $\boldsymbol{\Delta}(\varepsilon) = \operatorname{diag}(\varepsilon^0, \varepsilon^1)$ on the matrix given in example 3.8. With this scaling,

$$\boldsymbol{\Omega} - \tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

This scaling is tight only at entry $(0, 0)$. Let us try to tighten the scaling, with $\boldsymbol{\Delta}(\varepsilon) = \operatorname{diag}(\varepsilon^0, \varepsilon^{\frac{3}{2}})$. Then:

$$\boldsymbol{\Omega} - \tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} - \begin{pmatrix} 0 & \frac{3}{2} \\ \frac{3}{2} & 3 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}.$$

The scaling is now tight at $(0, 0)$ and $(1, 1)$. Finally, we can easily see that raising the valuation of the scaling any further makes it invalid: for instance $\boldsymbol{\Delta}(\varepsilon) = \operatorname{diag}(\varepsilon^0, \varepsilon^2)$ is not a valid scaling, since

$$\boldsymbol{\Omega} - \tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} - \begin{pmatrix} 0 & 2 \\ 2 & 4 \end{pmatrix}. = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

and eq. (3.12) is not verified anymore.

There is a systematic way of finding valid scalings that are maximally tight, via the Hungarian algorithm, see e.g. [17, 18] for an introduction.

Once a valid scaling has been found, the matrix $\boldsymbol{H}$ is easy to compute:

Proposition 3.11. *Given a valid scaling $\boldsymbol{\Delta}(\varepsilon)$, the decomposition*

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{\Delta}(\varepsilon)(\boldsymbol{H} + o(1))\boldsymbol{\Delta}(\varepsilon)$$

is verified for

$$H_{ij} = \begin{cases} \operatorname{lc} K_{ij}(\varepsilon) \text{ if } \boldsymbol{\Delta}(\varepsilon) \text{ is tight at } (i,j) \\ 0 \text{ otherwise} \end{cases} \tag{3.13}$$

Recall that $\operatorname{lc} K_{ij}$ is notation for the leading coefficient of the entry K_{ij}.

Proof. Follows directly by verifying that eq. (3.10) holds entry-wise for both tight and non-tight entries. □

Example 3.12 (Example 3.8, continued). Recall that with the scaling $\boldsymbol{\Delta}(\varepsilon) = \operatorname{diag}(\varepsilon^0, \varepsilon^1)$, the only tight entry is $(0,0)$. Then we have

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \varepsilon^2 + \varepsilon^3 \\ \varepsilon^2 + \varepsilon^3 & 5\varepsilon^3 \end{pmatrix} = \begin{pmatrix} 1 & \\ & \varepsilon \end{pmatrix} \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + o(1) \right\} \begin{pmatrix} 1 & \\ & \varepsilon \end{pmatrix}$$

Applying theorem 3.2 with this scaling is not very informative, since the second Schur complement $\boldsymbol{S}_1 = 0$. All we can deduce is that the first eigenvalue has expansion $\lambda_0(\varepsilon) = 1 + \mathcal{O}(\varepsilon)$, and that the second eigenvalue has valuation larger than 2. If we improve the scaling to $\boldsymbol{\Delta}(\varepsilon) = \operatorname{diag}(\varepsilon^0, \varepsilon^{\frac{3}{2}})$, entry $(1,1)$ becomes tight, and we can write

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \\ & \varepsilon^{\frac{3}{2}} \end{pmatrix} \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} + o(1) \right\} \begin{pmatrix} 1 & \\ & \varepsilon^{\frac{3}{2}} \end{pmatrix}$$

Applying theorem 3.2 with this scaling now gives us the asymptotics of both eigenpairs, with $\lambda_1(\varepsilon) = \varepsilon^3(5 + \mathcal{O}(\varepsilon))$.

Example 3.13 (Example 2.5, continued). We can compute the asymptotics of the matrix given in ex. 2.5 using the tools of this section. Some of the computations are best done with the help of a Computer Algebra System. For convenience we reproduce the matrix here:

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \frac{\varepsilon}{2} & \varepsilon^4 & 0 & 0 \\ \frac{\varepsilon}{2} & \frac{1}{4}\varepsilon^2 & \frac{\varepsilon^2}{2} & 0 & 0 \\ \varepsilon^4 & \varepsilon^2 & \varepsilon^2 & \frac{\varepsilon^3}{2} & 0 \\ 0 & 0 & \frac{\varepsilon^3}{2} & \frac{1}{8}\varepsilon^4 & \frac{\varepsilon^4}{2} \\ 0 & 0 & 0 & \frac{\varepsilon^4}{2} & \varepsilon^4 \end{pmatrix}$$

The corresponding valuation matrix is:

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & 1 & 4 & \infty & \infty \\ 1 & 2 & 2 & \infty & \infty \\ 4 & 2 & 2 & 3 & \infty \\ \infty & \infty & 3 & 4 & 4 \\ \infty & \infty & \infty & 4 & 4 \end{pmatrix}$$

Take $\boldsymbol{\Delta}(\varepsilon) = \operatorname{diag}(\varepsilon^0, \varepsilon^0, \varepsilon^1, \varepsilon^1, \varepsilon^2, \varepsilon^2)$. Then:

$$\tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 0 & 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 3 & 3 \\ 2 & 3 & 3 & 4 & 4 \\ 2 & 3 & 3 & 4 & 4 \end{pmatrix}$$

We highlight in blue the entries for which the scaling is tight ($\tilde{\Omega}_{i,j} = \Omega_{i,j}$). One can check that the scaling is valid by computing $\boldsymbol{\Omega} - \tilde{\boldsymbol{\Omega}}$, which should have non-negative entries.

$\boldsymbol{H}$ is found by applying proposition 3.11:

$$\boldsymbol{H} = \left(\begin{array}{c|cc|cc} 1 & \frac{1}{2} & & & \\ \hline \frac{1}{2} & \frac{1}{4} & \frac{1}{2} & & \\ & \frac{1}{2} & 1 & \frac{1}{2} & \\ \hline & & \frac{1}{2} & \frac{1}{8} & \frac{1}{2} \\ & & & \frac{1}{2} & 1 \end{array}\right)$$

The block structure in $\boldsymbol{H}$ corresponding to the valuations is highlighted. The Schur complements for this block structure are

$$\boldsymbol{S}_0 = 1, \quad \boldsymbol{S}_1 = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \quad \boldsymbol{S}_2 = \begin{pmatrix} \frac{1}{8} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}.$$

Theorem 3.2 gives us the ASE of $\boldsymbol{K}$ as:

$$\underline{\boldsymbol{K}(\epsilon)} = \begin{pmatrix} 1 & & \\ & & \\ & & \end{pmatrix} + \varepsilon^2 \begin{pmatrix} 0 & \frac{1}{2} & \\ \frac{1}{2} & 1 & \\ & & \end{pmatrix} + \varepsilon^4 \begin{pmatrix} & & \\ & \frac{1}{8} & \frac{1}{2} \\ & \frac{1}{2} & 1 \end{pmatrix}$$

The limiting eigenvalues and eigenvectors can be computed by diagonalising $\boldsymbol{S}_0, \boldsymbol{S}_1$ and $\boldsymbol{S}_2$ using the classical formulas for 2×2 matrices (we do not detail these calculations).

Example 3.14. Our final example for this section concerns a matrix for which 3.2 fails to characterise all eigenvalues, even with the optimal scaling (we end up in case (1) of the theorem). Take

$$\boldsymbol{K} = \begin{pmatrix} 1 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon^3 & -\varepsilon^3 \\ \varepsilon & -\varepsilon^3 & \varepsilon^3 \end{pmatrix}$$

with valuation

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 3 \end{pmatrix}$$

With the scaling $\boldsymbol{D} = \operatorname{diag}(\varepsilon^0, \varepsilon^1, \varepsilon^1)$, we have

$$\tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{pmatrix}.$$

Again the tight entries are highlighted in blue, and it can easily be checked that the scaling cannot be improved. Prop. 3.11 gives:

$$\boldsymbol{H} = \left(\begin{array}{c|cc} 1 & 1 & 1 \\ \hline 1 & 0 & 0 \\ 1 & 0 & 0 \end{array}\right)$$

The Schur complements are $\boldsymbol{S}_0 = 1$, $\boldsymbol{S}_1 = -\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Since $\boldsymbol{S}_1$ is of rank one, there is a single eigenvalue of valuation 2. Theorem 3.2 gives:

$$\underline{\boldsymbol{K}(\epsilon)} = \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix} - \varepsilon^2 \begin{pmatrix} 0 & & \\ & 1 & 1 \\ & 1 & 1 \end{pmatrix} + \mathcal{O}(\varepsilon^3)$$

The last eigenvalue has valuation > 2 but is not identified. For this we need a stronger theorem, specifically theorem 4.2. We revisit this computation in example 4.3.

The rest of this section contains the proof of theorem 3.2, and readers can skip ahead to section 4 for a generalisation of the theorem that is more widely applicable.

3.3. Regularised inverses in diagonally-scaled matrices. We shall now study the asymptotic spectral behaviour of $\boldsymbol{K}(\varepsilon)$ using regularised inverses, as introduced in section 2.4. The regularised inverses inherit a block structure from the diagonal scaling, as summarised in the following lemma:

LEMMA 3.15. *Let s be a real positive number and*

$$j = \max_{2\nu_i \leq s} i$$

be the index of the last block in $\boldsymbol{\Delta}(\varepsilon)$ to have valuation less than or equal to $\frac{s}{2}$. Then the leading term of the regularised inverse $\boldsymbol{M}_{s,\tau}(\varepsilon)$ of $\boldsymbol{K}(\varepsilon)$ at order s defined as in (2.7) has the following expression depending on whether $2\nu_j \neq s$:

- *in case $2\nu_j \neq s$ and $\boldsymbol{H}_{\leq j, \leq j}$ invertible, then for all τ*

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \begin{pmatrix} \boldsymbol{I}_b & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + o(1), \tag{3.14}$$

where $b = \sum_{i=0}^{j} b_i$.

- *in case $2\nu_j = s$ and $\boldsymbol{H}_{<j,<j}$ invertible (so that the Schur complement $\boldsymbol{S}_j$ is well defined), then*

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \begin{pmatrix} \boldsymbol{I}_{b-b_j} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_j(\boldsymbol{S}_j + \tau\boldsymbol{I})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + o(1), \tag{3.15}$$

for all τ such that the inverse exists (e.g., for all positive τ in the SPD case).

Proof. We compute directly the leading term of $\boldsymbol{M}_s$. It is important to keep in mind throughout that there are two cases, depending on whether $\frac{s}{2}$ matches one of the valuations ν_i. Recall the definition of j, $j = \underset{2\nu_i \le s}{\operatorname{argmax}}\, i$, whereby the matching case has $\nu_j = \frac{s}{2}$ and the non-matching case has $\nu_j < \frac{s}{2}$.

First, note that:

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \boldsymbol{K}(\varepsilon)(\boldsymbol{K}(\varepsilon) + \tau\varepsilon^s\boldsymbol{I})^{-1} = \boldsymbol{I} - \tau\varepsilon^s(\boldsymbol{K}(\varepsilon) + \tau\varepsilon\boldsymbol{I})^{-1}$$

We prefer the latter form because it is symmetric. Injecting the definition of $\boldsymbol{K}(\varepsilon)$ (eq. (3.2)),

$$\boldsymbol{M}_{s,\tau}(\varepsilon) = \boldsymbol{I} - \tau\varepsilon^s\left(\boldsymbol{\Delta}(\varepsilon)(\boldsymbol{H} + o(1))\boldsymbol{\Delta}(\varepsilon) + \tau\varepsilon^s\boldsymbol{I}\right)^{-1}$$

To compute the leading term, we define a modified diagonal scaling where each valuation greater than ν_j is clipped to $\frac{s}{2}$:

$$\tilde{\boldsymbol{\Delta}}(\varepsilon) = \operatorname{diag}(\boldsymbol{\varepsilon}^{\min(\nu_0, \frac{s}{2})}, \ldots, \ldots, \boldsymbol{\varepsilon}^{\min(\nu_r, \frac{s}{2})}) \tag{3.16}$$

Note that it may be the case that $\nu_j = \frac{s}{2}$ (matching case), or not (non-matching). This lets us rewrite $\boldsymbol{M}_s$ as:

$$\boldsymbol{M}_s(\varepsilon) = \boldsymbol{I} - \tau\left(\tilde{\boldsymbol{\Delta}}(\varepsilon)(\tilde{\boldsymbol{H}} + o(1))\tilde{\boldsymbol{\Delta}}(\varepsilon)\right)^{-1}, \tag{3.17}$$

where

$$\tilde{\boldsymbol{H}} = \begin{cases} \begin{pmatrix} \boldsymbol{H}_{\le j,\le j} & \boldsymbol{0} \\ \boldsymbol{0} & \tau\boldsymbol{I} \end{pmatrix}, & \text{if } \nu_j < \frac{s}{2}, \\ \begin{pmatrix} \boldsymbol{H}_{<j,<j} & \boldsymbol{H}_{<j,j} & \boldsymbol{0} \\ \boldsymbol{H}_{j,<j} & \boldsymbol{H}_{j,j} + \tau\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \tau\boldsymbol{I} \end{pmatrix}, & \nu_j = \frac{s}{2}. \end{cases}$$

The two cases are non-matching and matching, respectively. We use the fact that $\tilde{\boldsymbol{\Delta}}(\varepsilon)$ is square to pull it out of the inverse in eq. (3.17), and obtain:

$$\boldsymbol{M}_s(\varepsilon) = \boldsymbol{I} - \tau\varepsilon^{\frac{s}{2}}\tilde{\boldsymbol{\Delta}}(\varepsilon)^{-1}\left(\tilde{\boldsymbol{H}} + o(1)\right)^{-1}\tilde{\boldsymbol{\Delta}}(\varepsilon)^{-1}\varepsilon^{\frac{s}{2}} \tag{3.18}$$

and we note that block i in $\tilde{\boldsymbol{\Delta}}(\varepsilon)^{-1}\varepsilon^{\frac{s}{2}}$ is either $o(1)$ if $\nu_i < \frac{s}{2}$, or $1 + o(1)$ otherwise.

Note that the matrix $\tilde{\boldsymbol{H}}$ is invertible under the assumptions in the theorem (i.e., invertibility of $\boldsymbol{H}_{\le j,\le j}$ for the non-matching case and invertibility of both $\boldsymbol{H}_{<j,<j}$ and $\boldsymbol{S}_j + \tau\boldsymbol{I}_{c_j}$ for the matching case), so that:

$$\boldsymbol{M}_s(\varepsilon) = \boldsymbol{I} - \tau\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}\left(\tilde{\boldsymbol{H}}^{-1} + o(1)\right)\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} + o(1) \tag{3.19}$$

Here multiplication to the left and right by the matrix $\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}$ selects the blocks with valuation $\le \frac{s}{2}$. We therefore only need to compute the relevant part in $\tilde{\boldsymbol{H}}^{-1}$, which we can do by block matrix inversion:

$$\tilde{\boldsymbol{H}}^{-1} = \begin{cases} \begin{pmatrix} * & * \\ * & \tau^{-1}\boldsymbol{I} \end{pmatrix}, & \text{if } \nu_j < \frac{s}{2}, \\ \begin{pmatrix} * & * & * \\ * & (\boldsymbol{S}_j + \tau\boldsymbol{I})^{-1} & 0 \\ * & 0 & \tau^{-1}\boldsymbol{I} \end{pmatrix} & \text{otherwise.} \end{cases}$$

Inserting into eq. (3.19) and simplifying, we obtain (3.14) and (3.15) as claimed. ∎

Now we are ready to prove Theorem 3.2.

Proof of Theorem 3.2. Lemma 3.15 tells us that the limit-rank function (eq. (2.11)) for $s \leq j$ equals

$$r(s) = \sum_{i:2\nu_i \leq s} b_i \tag{3.20}$$

so that new dimensions appear every time $\frac{s}{2}$ matches one of the scalings ν_i. By matching the rank function with the that of (2.11), and comparing results in Lemma 3.15 to that in Lemma 2.14, we obtain (3.6). Note that if $\boldsymbol{H}$ is invertible, then the equality (3.20) holds for all s, and therefore the matrix in (3.7) is full rank and is equal to the ASE. ∎

4. Generalised kernel form. The results of the previous section are only directly useful if the matrix has entries with different orders of magnitude in ε, so that a non-trivial scaling matrix can be used. An example of a matrix that only has the trivial scaling is the kernel matrix given in eq. (1.2). Recall that this matrix can be expanded as a series in ε^2, where each term is a matrix of distances raised to some power:

$$\boldsymbol{K}_{2l} = \frac{1}{l!}\left[(x_i - y_i)^{2l}\right]_{i=1,j=1}^{n}$$

Using the binomial theorem, we can expand the distances in terms of monomials:

$$\boldsymbol{K}_{2l} = \frac{1}{l!}\left[\sum_{q=0}^{2l}\binom{2l}{q}(-1)^q x_i^q y_i^{2l-q}\right]_{i=1,j=1}^{n}$$

If we note $\boldsymbol{v}_i = [x_j^i]_{j=1}^n$, this results in the expansion:

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{v}_0\boldsymbol{v}_0^\top - \varepsilon^2\left(\boldsymbol{v}_2\boldsymbol{v}_0^\top - 2\boldsymbol{v}_1\boldsymbol{v}_1^\top + \boldsymbol{v}_0\boldsymbol{v}_2^\top\right) + \dots$$

Recall that $\boldsymbol{v}_0 = \mathbb{1}$, the constant vector, so that every entry in $\boldsymbol{K}(\varepsilon)$ is $\mathcal{O}(1)$ (more precisely, has zero valuation). We can only use the trivial scaling matrix, and so theorem 3.2 only tells us about the eigenvalues and eigenvectors of valuation 0. There is only one such eigenvalue, since $\boldsymbol{K}_0 = \boldsymbol{v}_0\boldsymbol{v}_0^\top$ is of rank one.

Thus, theorem 3.2 is not powerful enough to directly characterise the ASE of all analytic perturbations. Better results are needed, and this section we will consider a generalisation of (3.2), and characterise the ASE for matrices of the form:

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V}\boldsymbol{\Delta}(\varepsilon)\left(\boldsymbol{W} + o(1)\right)\boldsymbol{\Delta}(\varepsilon)\boldsymbol{V}^\top \tag{4.1}$$

Although this form may seem abstract, it includes very general kernel matrices in the flat limit. We will begin by defining these matrices in more detail, explaining what $\boldsymbol{V}$ and $\boldsymbol{W}$ correspond to in eq. (4.1), and then derive the ASE of matrices in this form.

4.1. ASE of matrices in generalised kernel form. We need to make the notation more precise. Again, we define the scaling matrix similarly to (3.3)

$$\boldsymbol{\Delta}(\varepsilon) = \boldsymbol{\Delta}_{\boldsymbol{a}}(\varepsilon) = \operatorname{diag}(\underbrace{\varepsilon^{\nu_0},\dots,\varepsilon^{\nu_0}}_{a_0},\underbrace{\varepsilon^{\nu_1},\dots\varepsilon^{\nu_1}}_{a_1},\dots,\underbrace{\varepsilon^{\nu_p}\dots\varepsilon^{\nu_p}}_{a_p}), \tag{4.2}$$

where each valuation ν_i is repeated a_i times and $\nu_0 < \nu_1 < \dots < \nu_p$.

We assume that $\boldsymbol{V} \in \mathbb{R}^{n\times\sum a_i}$ and is partitioned according to the valuations, as

$$\boldsymbol{V} = \begin{pmatrix}\boldsymbol{V}_0 & \boldsymbol{V}_1 & \dots & \boldsymbol{V}_p\end{pmatrix}, \tag{4.3}$$

so that $\boldsymbol{V}_i \in \mathbb{R}^{n\times a_i}$. We will use the QR factorisation of $\boldsymbol{V}$, which we arrange in the block-upper triangular form:

$$\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{R} = \begin{pmatrix}\boldsymbol{Q}_0 & \boldsymbol{Q}_1 & \dots & \boldsymbol{Q}_p\end{pmatrix}\begin{pmatrix}\boldsymbol{R}_{0,0} & \boldsymbol{R}_{0,1} & \dots & \dots & \boldsymbol{R}_{0,p}\\ & \boldsymbol{R}_{1,1} & \boldsymbol{R}_{1,2} & \dots & \boldsymbol{R}_{1,p}\\ & & \boldsymbol{R}_{2,2} & \dots & \boldsymbol{R}_{2,p}\\ & & & \ddots & \vdots\\ & & & & \boldsymbol{R}_{p,p}\end{pmatrix}, \tag{4.4}$$

so that the blocks $\boldsymbol{R}_{i,j} \in \mathbb{R}^{b_i\times a_j}$ and $\boldsymbol{Q}_r \in \mathbb{R}^{n\times b_i}$.

Remark 4.1. Under the assumption that $\operatorname{rank} \boldsymbol{V} = n$, the numbers b_i sum to n, and $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is a square matrix. Moreover b_i measures the new dimensions introduced by $\boldsymbol{V}_i$, i.e.,

$$b_i = \begin{cases} \operatorname{rank} \boldsymbol{V}_0, & i = 0, \\ \operatorname{rank} \boldsymbol{V}_{\leq i} - \operatorname{rank} \boldsymbol{V}_{<i}, & i > 0, \end{cases}$$

where we assume that $b_i > 0$ (i.e. $\operatorname{rank} \boldsymbol{V}_{\leq i} > \operatorname{rank} \boldsymbol{V}_{<i}$).

The matrix $\boldsymbol{W} \in \mathbb{R}^{\sum a_i \times \sum a_i}$ in eq. (4.1) is also divided into blocks

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_{0,0} & \boldsymbol{W}_{0,1} & \dots & \boldsymbol{W}_{0,p} \\ \boldsymbol{W}_{1,0} & \boldsymbol{W}_{1,1} & \dots & \boldsymbol{W}_{1,p} \\ \vdots & \vdots & \dots & \dots \\ \boldsymbol{W}_{p,0} & \boldsymbol{W}_{p,1} & \dots & \boldsymbol{W}_{p,p} \end{pmatrix}.$$

according to the structure of the valuations in (4.2).

Finally, we define the following matrix:

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{R}_{0,0} & & & \\ & \boldsymbol{R}_{1,1} & & \\ & & \ddots & \\ & & & \boldsymbol{R}_{p,p} \end{pmatrix} \boldsymbol{W} \begin{pmatrix} \boldsymbol{R}_{0,0}^\top & & & \\ & \boldsymbol{R}_{1,1}^\top & & \\ & & \ddots & \\ & & & \boldsymbol{R}_{p,p}^\top \end{pmatrix} \tag{4.5}$$

The sequence of Schur complements in $\boldsymbol{H}$ are defined in the same way as for thm. 3.2, see (3.5).

With this notation, we obtain the following generalisation of theorem 3.2:

THEOREM 4.2. *Let $\boldsymbol{K}(\varepsilon)$ be given in (4.1), with $\boldsymbol{W}$ invertible, $\boldsymbol{H}$ be as in (4.8), and $\boldsymbol{V}$ satisfying assumptions of Remark 4.1. Let $\boldsymbol{S}_0, \dots, \boldsymbol{S}_p$ denote the Schur complements in $\boldsymbol{H}$.*

1. *Then the ASE is given by*

$$\underline{\boldsymbol{K}(\epsilon)} = \sum_{i=0}^{p} \varepsilon^{2\nu_i} \boldsymbol{Q}_i \boldsymbol{S}_i \boldsymbol{Q}_i^\top. \tag{4.6}$$

2. *We have $\boldsymbol{S}_0 = \boldsymbol{W}_{0,0}$, and for any $j > 0$ such that $\boldsymbol{V}_{\leq j-1}$ is full column rank, $\boldsymbol{S}_j$ admits a simpler expression via Schur complements of $\boldsymbol{W}$*

$$\boldsymbol{S}_j = \boldsymbol{R}_{j,j} \left(\boldsymbol{W}_{j,j} - \boldsymbol{W}_{j,<j} (\boldsymbol{W}_{<j,<j})^{-1} \boldsymbol{W}_{<j,j} \right) \boldsymbol{R}_{j,j}^\top, \tag{4.7}$$

where the matrices $\boldsymbol{W}_{j,<j}$ and $\boldsymbol{W}_{<j,j} = \boldsymbol{W}_{j,<j}^\mathsf{T}$ are defined as

$$\boldsymbol{W}_{j,<j} = \begin{pmatrix} \boldsymbol{W}_{j,0} & \boldsymbol{W}_{j,1} & \cdots & \boldsymbol{W}_{j,j-1} \end{pmatrix}.$$

The proof is deferred to the next subsection.

Example 4.3 (Ex. 3.14 revisited). Ex. 3.14 is a case where thm. 3.2 fails to characterise all eigenvalues. We show that thm. 4.2 succeeds. In this example we have:

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon^3 & -\varepsilon^3 \\ \varepsilon & -\varepsilon^3 & \varepsilon^3 \end{pmatrix}$$

One can check that $\boldsymbol{K}(\varepsilon)$ can also be written as

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V} \begin{pmatrix} 1 & & \\ & \varepsilon & \\ & & \varepsilon^{\frac{3}{2}} \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & \varepsilon & \\ & & \varepsilon^{\frac{3}{2}} \end{pmatrix}$$

with

$$\boldsymbol{V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}.$$

This is a form compatible with theorem 4.2. Since $\boldsymbol{V}$ already has orthogonal columns we can write

$$\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{R} = \left(\begin{array}{c|c|c} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{array}\right) \left(\begin{array}{c|c|c} 1 & \sqrt{2} & \\ & & \sqrt{2} \end{array}\right).$$

We highlight the block structure corresponding to the successive valuations $0, 1, \frac{3}{2}$. In the notation of this section, we have (eq. (4.5)):

$$\boldsymbol{H} = \left(\begin{array}{c|c|c} 1 & & \\ \hline & \sqrt{2} & \\ \hline & & \sqrt{2} \end{array}\right) \left(\begin{array}{c|c|c} 1 & 1 & \\ \hline 1 & & \\ \hline & & 1 \end{array}\right) \left(\begin{array}{c|c|c} 1 & & \\ \hline & \sqrt{2} & \\ \hline & & \sqrt{2} \end{array}\right) = \begin{pmatrix} 1 & \sqrt{2} & \\ \sqrt{2} & & \\ & & 2 \end{pmatrix}$$

The Schur complements in $\boldsymbol{H}$ are $\boldsymbol{S}_0 = 1, \boldsymbol{S}_1 = -2, \boldsymbol{S}_2 = 2$. Applying theorem 4.2, the ASE of $\boldsymbol{K}$ equals:

$$\underline{\boldsymbol{K}(\epsilon)} = \boldsymbol{q}_0\boldsymbol{q}_0^\top - 2\varepsilon^2\boldsymbol{q}_1\boldsymbol{q}_1^\top + 2\varepsilon^3\boldsymbol{q}_2\boldsymbol{q}_2^\top = \begin{pmatrix} 1 & & \\ & & \\ & & \end{pmatrix} - \varepsilon^2 \begin{pmatrix} & & \\ & 1 & 1 \\ & 1 & 1 \end{pmatrix} + \varepsilon^3 \begin{pmatrix} & & \\ & 1 & -1 \\ & -1 & 1 \end{pmatrix}$$

Compared to the previous attempt, we have managed to identify the eigenvalue of order $\mathcal{O}(\varepsilon^3)$.

4.2. Proof of theorem 4.2. We begin with a lemma that allows us to convert the form of eq. (4.1) to the simpler form we used previously in section 3, that of eq. (3.2). The lemma already appears in a different form in [5].

LEMMA 4.4. *Let $\boldsymbol{V}$, $\boldsymbol{Q}$, $\boldsymbol{R}$ be as in (4.3), (4.4), and $\boldsymbol{H}$ as in (4.5). Then we have the following asymptotic equivalence:*

$$\boldsymbol{R}\boldsymbol{\Delta}_{\boldsymbol{a}}(\varepsilon)(\boldsymbol{W} + o(1))\boldsymbol{\Delta}_{\boldsymbol{a}}(\varepsilon)\boldsymbol{R}^\top = \boldsymbol{\Delta}_{\boldsymbol{b}}(\varepsilon)(\boldsymbol{H} + o(1))\boldsymbol{\Delta}_{\boldsymbol{b}}(\varepsilon) \tag{4.8}$$

where $\boldsymbol{\Delta}_{\boldsymbol{a}}(\varepsilon)$ and $\boldsymbol{\Delta}_{\boldsymbol{b}}(\varepsilon)$ are defined in (4.2) and (3.3), respectively, and

Proof. We denote $\boldsymbol{A}(\varepsilon) = \boldsymbol{R}\boldsymbol{\Delta}(\varepsilon)(\boldsymbol{W} + o(1))\boldsymbol{\Delta}(\varepsilon)\boldsymbol{R}^\top$, and $\boldsymbol{A}_{i,j}$ the block i, j in the partitioning induced by the valuations in the scaling matrix $\boldsymbol{\Delta}(\varepsilon)$.

$$\begin{aligned} \boldsymbol{A}_{i,j}(\varepsilon) &= \sum_{k \geq i, l \geq j} \boldsymbol{R}_{i,k}\varepsilon^{\nu_i}(\boldsymbol{W}_{k,l} + o(1))(\boldsymbol{R})_{l,j}^\top\varepsilon^{\nu_j} \\ &= \varepsilon^{\nu_i}(\boldsymbol{R}_{i,i}\boldsymbol{W}_{i,i}(\boldsymbol{R}_{i,i})^\top + o(1))\varepsilon^{\nu_j} \end{aligned}$$

since the valuations in $\boldsymbol{\Delta}(\varepsilon)$ are increasing. Eq. (4.5) is exactly this result expressed for all blocks. ∎

Given this lemma, the proof of theorem 4.2 is straightforward.

Proof of Theorem 4.2. For the first part of the theorem (statement (1)), consider the matrix

$$\boldsymbol{K}'(\varepsilon) = \boldsymbol{Q}^\top\boldsymbol{K}(\varepsilon)\boldsymbol{Q} = \boldsymbol{R}\boldsymbol{\Delta}(\varepsilon)(\boldsymbol{H} + o(1))^{-1}\boldsymbol{\Delta}(\varepsilon)\boldsymbol{R}^\top$$

and note that by lemma 2.10, $\underline{\boldsymbol{K}'(\varepsilon)} = \boldsymbol{Q}^\top\underline{\boldsymbol{K}(\epsilon)}\boldsymbol{Q}$. Lemma 4.4 lets us apply theorem Theorem 3.2 to $\boldsymbol{K}'$. We obtain eq. (4.6) via $\underline{\boldsymbol{K}(\epsilon)} = \boldsymbol{Q}\underline{\boldsymbol{K}'(\varepsilon)}\boldsymbol{Q}^\top$.

For the second part (statement (2)), note that in this case since $\operatorname{rank}\boldsymbol{R} = \operatorname{rank}\boldsymbol{V} = n$, all the diagonal blocks $\boldsymbol{R}_{i,i}$ are full row rank. This implies that $\boldsymbol{H}$ is invertible (thanks to invertibility of $\boldsymbol{W}$). The simplified expression for the Schur complement $\boldsymbol{S}_i$ (eq. (4.7)) can be obtained as follows. If $\boldsymbol{V}_{\leq j-1}$ is full column rank, then $\boldsymbol{R}_{i,i}$ are square for $j < i$ and the matrix

$$\boldsymbol{A} = \operatorname{diag}(\boldsymbol{S}_0, \ldots, \boldsymbol{S}_{j-1})$$

is invertible. Then the Schur complement is obtained as

$$\begin{aligned} \boldsymbol{S}_j &= \boldsymbol{R}_{j,j}\boldsymbol{W}_{j,j}\boldsymbol{R}_{j,j}^\mathsf{T} - \boldsymbol{R}_{j,j}\boldsymbol{W}_{j,<j}\boldsymbol{A}^\mathsf{T}(\boldsymbol{A}\boldsymbol{W}_{<j,<j}\boldsymbol{A}^\mathsf{T})^{-1}\boldsymbol{A}\boldsymbol{W}_{<j,j}\boldsymbol{R}_{j,j}^\mathsf{T} \\ &= \boldsymbol{R}_{j,j}\left(\boldsymbol{W}_{j,j} - \boldsymbol{W}_{j,<j}(\boldsymbol{W}_{<j,<j})^{-1}\boldsymbol{W}_{<j,j}\right)\boldsymbol{R}_{j,j}^\top, \end{aligned}$$

which completes the proof. ∎

5. Application to kernel matrices. In this section, we apply theorem 4.2 to kernel matrices, to settle a conjecture from [5]. The only difficulty is to show that kernel matrices can indeed be written in the form required by theorem 4.2, and explicitate the matrices involved.

5.1. Kernel matrices: background and notation. We need to briefly recall some definitions and notation on kernel matrices. For more information on these matrices, we refer the reader to [23, 26, 13]. We follow the notation used in [5].

Kernel matrices are formed from a set of n points in $\mathbb{R}^d$, noted $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\}$. We define

$$\boldsymbol{K}(\varepsilon) = [k(\varepsilon\boldsymbol{x}_i, \varepsilon\boldsymbol{x}_j)]_{i,j=1}^{n,n} \tag{5.1}$$

where $k(\boldsymbol{x}, \boldsymbol{y})$ is a positive-definite *kernel function* and ε is a (spatial) non-negative scaling parameter. We seek to characterise the ASE of kernel matrices in the flat limit $\varepsilon \to 0$.

The class of kernel functions is very wide, but the most commonly-used are *radial*, meaning that $k(\boldsymbol{x}, \boldsymbol{y})$ only depends on the (Euclidean) distance $\|\boldsymbol{x} - \boldsymbol{y}\|$:

$$k(\boldsymbol{x}, \boldsymbol{y}) = \psi(\|\boldsymbol{x} - \boldsymbol{y}\|) \tag{5.2}$$

Radial kernels are particularly easy to work with because the flat limit expansion of $k(\varepsilon\boldsymbol{x}, \varepsilon\boldsymbol{y})$ can be obtained from an expansion of $\psi(s)$ at 0, i.e.:

$$k(\varepsilon\boldsymbol{x}, \varepsilon\boldsymbol{y}) = \psi(\varepsilon \|\boldsymbol{x} - \boldsymbol{y}\|) = \psi_0 + \varepsilon\psi_1 \|\boldsymbol{x} - \boldsymbol{y}\| + \varepsilon^2\psi_2 \|\boldsymbol{x} - \boldsymbol{y}\|^2 + \varepsilon^3\psi_3 \|\boldsymbol{x} - \boldsymbol{y}\|^3 + \dots \tag{5.3}$$

In this section we assume that $\psi(s)$ is analytic at 0, so that we are dealing with analytic perturbations.

Remark 5.1. While we consider radial kernels in this section, the result will also hold for other smooth kernels (as long as the kernel matrices analytic in the scaling parameter ε), similarly to [5].

The following criterion, called the "regularity index", is key for characterising the flat limit of kernel matrices:

Definition 5.2 (Regularity index). *Let $k(\boldsymbol{x}, \boldsymbol{y}) = \psi(\|\boldsymbol{x} - \boldsymbol{y}\|)$ a radial kernel, and $\psi(s)$ have the following expansion at 0:*

$$\psi(s) = \psi_0 + \psi_1 s + \psi_2 s^2 + \dots \tag{5.4}$$

We say k has regularity index r if $\psi_{2r-1} \neq 0$, and $\psi_{2c-1} = 0$ for $c < r$, i.e. ψ_{2r-1} is the first non-zero odd term in the expansion. Kernels with $r < \infty$ are said to be "finitely smooth", kernels with $r = \infty$ are said to be "completely smooth".

The spectral behaviour of kernel matrices depends in the most part on the regularity index, although we cannot provide a concise explanation of why this is the case (see [7] for a discussion).

Two kernels with widely different flat limit behaviour are the Gaussian kernel and the exponential kernel, with $r = \infty$ and $r = 1$, respectively:

Example 5.3 (Examples of kernels with different regularity coefficients). The Gaussian kernel (eq. (1.1)) corresponds to $\psi(s) = \exp(-s^2) = 1 - s^2 + \frac{1}{2}s^4 - \frac{1}{6}s^6 + \dots$, which contains only *even* monomials in s ($s^0, s^2, s^4, \dots$). This leads to a small-ε expansion that contains only even powers of ε, so that the Gaussian kernel has therefore regularity $r = \infty$.

Contrast this to the so-called *exponential* kernel, which has

$$\psi(s) = \exp(-s) = 1 - s + \frac{1}{2}s^2 - \frac{1}{6}s^3 + \dots \tag{5.5}$$

where the first odd power of s is s^1. The regularity of the exponential kernel is therefore $r = 1$.

The following kernel is a special case of the Matérn family of kernels [24] with $r = 2$:

$$\psi(s) = (1 + s)\exp(-s) = (1 + s)(1 - s + \frac{1}{2}s^2 - \frac{1}{6}s^3 \dots) = 1 - \frac{s^2}{2} + \frac{s^3}{3} + \dots \tag{5.6}$$

Compared to the exponential kernel, the term in s drops out, but the term in s^3 remains, which increases r from 1 to 2.

5.2. Vandermonde, Wronskian and distance matrices. To express kernel matrices in the requisite form we need some standard notation for multivariate polynomials (see [5] for details).

Let $\boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & \dots & x_d \end{pmatrix}^\top \in \mathbb{R}^d$. A monomial in $\boldsymbol{x}$ is a function of the form:

$$\boldsymbol{x}^{\boldsymbol{\alpha}} = \prod_{i=1}^{d} x_i^{\alpha_i}$$

for $\boldsymbol{\alpha} \in \mathbb{N}^d$ (a multi-index). The degree of a monomial is defined $|\boldsymbol{\alpha}| = \sum_{i=1}^d \alpha_i$. For instance: $\boldsymbol{x}^{(1,3,1)} = x_1^1 x_2^3 x_3^1$ has degree 5, so does $\boldsymbol{x}^{(2,2,1)} = x_1^2 x_2^2 x_3^1$. The numbers of monomials of degree $\leq s$ and degree $= s$ in dimension d are given by

$$\mathbb{P}_{s,d} = \binom{k+d}{d}, \quad \mathbb{H}_{s,d} = \binom{k+d-1}{d-1} \tag{5.7}$$

respectively. By evaluating monomials of degree $\leq s$ on a discrete set of nodes $\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$, we form a matrix:

Definition 5.4 (Vandermonde matrix). *The Vandermonde matrix of degree $\leq s$ on nodes $\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$*

$$\boldsymbol{V}_{\leq s} = [\boldsymbol{x}_i^{\boldsymbol{\alpha}}]_{i \in \{1,\dots,n\}, |\boldsymbol{\alpha}| \leq s} \tag{5.8}$$

The nodes vary along the rows, the monomials (indexed by $\boldsymbol{\alpha}$) along the columns. Which (degree-graded) monomial order is used is irrelevant for our results.

Example 5.5. In dimension one, the generalised Vandermonde matrix simplifies to the classical Vandermonde matrix:

$$\boldsymbol{V}_{\leq s} = \left[x_i^j\right]_{i \in \{1,\dots,n\}, j \in \{0,\dots,s\}} \tag{5.9}$$

By eq. (5.7), the matrix $\boldsymbol{V}_{\leq s}$ has dimension $n \times \mathbb{P}_{s,d} = n \times \binom{s+d}{d}$ in dimension d. One needs to keep in mind two important differences between the univariate and the multivariate case:

1. In the univariate case, the matrix $\boldsymbol{V}_{\leq s}$ always has full column rank, if $s \leq n-1$ and if the nodes in $\mathcal{X}$ are distinct. This is no longer the case in the multivariate case. For instance, if the nodes lie on a line, then $\boldsymbol{V}_{\leq 1}$ has rank 2 instead of rank $1 + d$. When the matrix $\boldsymbol{V}_{\leq s}$ is rank-deficient, we say the nodes are *non-unisolvent* at degree s.
2. In the univariate case, by picking $s = n - 1$ we obtain a square matrix $\boldsymbol{V}_{\leq n-1}$. In the multivariate case this may or may not be possible depending on n and d. For instance, in dimension $d = 2$, the size of $\boldsymbol{V}_{\leq s}$ is $1, 3, 6, 10, \dots$ with $s = 0, 1, 2, 3, \dots$. If $n = 5$ then $\boldsymbol{V}_{\leq 1}$ is too narrow and $\boldsymbol{V}_{\leq 2}$ too wide.

We also split the Vandermonde matrices in blocks, which correspond to fixed degrees of monomials:

$$\boldsymbol{V}_{\leq s} = \begin{pmatrix} \boldsymbol{V}_0 & \boldsymbol{V}_1 & \cdots & \boldsymbol{V}_s \end{pmatrix}, \quad \boldsymbol{V}_s = [\boldsymbol{x}_i^{\boldsymbol{\alpha}}]_{i \in \{1,\dots,n\}, |\boldsymbol{\alpha}| = s}$$

The particular order of monomials of the same degree will not be important for what follows, but we assume it fixed.

Example 5.6. In dimension 2, the generalised Vandermonde matrix has blocks $\boldsymbol{V}_s$ with $1, 2, 3, \dots$ columns:

$$\boldsymbol{V}_{\leq 2} = \left[\begin{array}{c|cc|ccc} 1 & y_1 & z_1 & y_1^2 & y_1 z_1 & z_1^2 \\ 1 & y_2 & z_2 & y_2^2 & y_2 z_2 & z_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_n & z_n & y_n^2 & y_n z_n & z_n^2 \end{array}\right], \tag{5.10}$$

where

$$\mathcal{X} = \{\left[\begin{smallmatrix} y_1 \\ z_1 \end{smallmatrix}\right], \left[\begin{smallmatrix} y_2 \\ z_2 \end{smallmatrix}\right], \dots, \left[\begin{smallmatrix} y_n \\ z_n \end{smallmatrix}\right]\},$$

with a particular chosen ordering of monomials.

Alongside Vandermonde matrices, we need to define Wronskian matrices, which are matrices of partial derivatives of the kernel:

DEFINITION 5.7. *Let k a radial kernel with regularity index r, and let d be fixed. Then for $i, j \geq 0$ such that $i + j \leq 2r - 2$ we define the Wronskian submatrix $\boldsymbol{W}_{i,j} \in \mathbb{R}^{\mathbb{H}_{i,d} \times \mathbb{H}_{j,d}}$ as*

$$\boldsymbol{W}_{i,j} = \left[\frac{\frac{\partial^q}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}} \partial \boldsymbol{y}^{\boldsymbol{\beta}}} k(\boldsymbol{x}, \boldsymbol{y})\Big|_{\boldsymbol{x},\boldsymbol{y}=0}}{\boldsymbol{\alpha}!\boldsymbol{\beta}!} \right]_{|\boldsymbol{\alpha}|=i, |\boldsymbol{\beta}|=j},$$

where the columns and the rows and columns of $\boldsymbol{W}_{i,j}$ are indexed by the multi-induces whose order is consistent with ordering the degrees in the Vandermonde matrices (5.8) (i.e., rows and columns of $\boldsymbol{W}_{i,j}$ are in the same order as columns for $\boldsymbol{V}_i$ and $\boldsymbol{V}_j$ respectively).

See the appendix of [7] for convenient formulas for Wronskians of common kernels.

We can also define stacked Wronskian matrices $\boldsymbol{W}_{\leq i, \leq i}$, for $i < r$ as

$$\boldsymbol{W}_{\leq i, \leq i} = \begin{pmatrix} \boldsymbol{W}_{(0,0)} & \boldsymbol{W}_{(0,1)} & \cdots & \boldsymbol{W}_{(0,i)} \\ \boldsymbol{W}_{(1,0)} & \boldsymbol{W}_{(1,1)} & \cdots & \boldsymbol{W}_{(1,i)} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{W}_{(i,0)} & \boldsymbol{W}_{(i,1)} & \cdots & \boldsymbol{W}_{(i,i)} \end{pmatrix} \tag{5.11}$$

This matrix is of size $\mathbb{P}_{i,d} \times \mathbb{P}_{i,d}$ and contain all partial derivatives up to order $2i$. Moreover, it possesses the following nice property:

LEMMA 5.8. *For a (strictly) positive definite kernel, all the Wronskian matrices are (strictly) positive definite.*

Proof. See [7]. ∎

Finally, the following matrices are required in the expansion of finitely-smooth kernels:

$$\boldsymbol{D}^{(q)} = \left[\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^q \right]_{i,j=1}^n. \tag{5.12}$$

We conclude this subsection by a result on conditional positive definiteness of $\boldsymbol{D}^{(q)}$ for odd order d.

LEMMA 5.9. *If $q = 2r - 1$ for integer $r \geq 1$, and $\boldsymbol{V}_{\leq r-1}$ is full column rank, then the matrix*

$$(-1)^r \boldsymbol{A}^t \boldsymbol{D}^{(q)} \boldsymbol{A}$$

is strictly positive definite for any full column rank matrix $\boldsymbol{A}$ with $\boldsymbol{A}^t \boldsymbol{V}_{\leq r-1} = 0$.

Proof. See [13], ch. 8. ∎

We can now show that kernel matrices have the form required to apply theorem 4.2. We separate the completely smooth and finitely smooth cases.

5.3. Results in for smooth radial kernels. We first analyse the smooth case, where the kernel function is differentiable sufficiently many times. With some abuse of notation, denote by

$$\boldsymbol{\Delta}_k = \boldsymbol{\Delta}_k(\varepsilon) = \begin{bmatrix} 1 & & & \\ & \varepsilon I_{\mathbb{H}_1} & & \\ & & \ddots & \\ & & & \varepsilon^k I_{\mathbb{H}_k} \end{bmatrix}, \tag{5.13}$$

which is a particular case of the matrix (4.2) (corresponding to the choice of with $a_j = \mathbb{H}_{j,d}$ and $\nu_j = j$, i.e the number of repetitions of ε^j is according to the number of homogeneous polynomials of degree d).

LEMMA 5.10. *Let $k(\boldsymbol{x}, \boldsymbol{y})$ be a positive definite kernel function, such that $k(\boldsymbol{x}, \boldsymbol{y})$ is a radial kernel with regularity r, and $\mathcal{X}$ be a node set such that* $\operatorname{rank} \boldsymbol{V}_{\leq p-1} = n$ *for $p \leq r$.*

Then the kernel matrix has the following asymptotic form in $\varepsilon \to 0$:

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V}_{\leq p-1} \boldsymbol{\Delta}_{p-1}(\varepsilon)(\boldsymbol{W}_{\leq p-1} + o(1)) \boldsymbol{\Delta}_{p-1}(\varepsilon) \boldsymbol{V}_{\leq p-1}^\top. \tag{5.14}$$

where $\boldsymbol{V} = \boldsymbol{V}_{\leq p}$ is the Vandermonde matrix described in (5.8), $\boldsymbol{W} = \boldsymbol{W}_{\leq p, \leq p}$ is the Wronskian matrix from (5.11), and $\boldsymbol{\Delta}(\varepsilon) = \boldsymbol{\Delta}_k$ is the diagonal matrix defined in (5.13).

Proof. The proof is given in section 8. ∎

Note that the same expansion holds for general smooth kernels (not necessarily radial).

LEMMA 5.11. *$k(\boldsymbol{x}, \boldsymbol{y})$ be a positive definite kernel function with $k \in \mathcal{C}^{p,p}$ and $\mathcal{X}$ be a node set such that* $\operatorname{rank} \boldsymbol{V}_{\leq p-1} = n$. *Then $\boldsymbol{K}(\varepsilon)$ has the same expansion as in (5.14)*

Proof. The proof is given in section 8. ∎

Lemma 5.10 and Lemma 5.11 help us to establish find the ASE for the completely smooth case.

THEOREM 5.12. *Let $\boldsymbol{K}(\varepsilon)$ satisfy the conditions of Lemma 5.10 or Lemma 5.11, and, in addition p is the smallest such number (i.e.,* $\operatorname{rank} \boldsymbol{V}_{\leq p-1} < n$ *but* $\operatorname{rank} \boldsymbol{V}_{\leq p} = n$*). Let $\boldsymbol{Q}_0, \dots, \boldsymbol{Q}_p$ and $\boldsymbol{R}_{0,0}, \dots, \boldsymbol{R}_{p,p}$ come from the block QR factorization of the matrix*

$$\boldsymbol{V} = \boldsymbol{V}_{\leq p} = \begin{pmatrix} \boldsymbol{V}_0 & \boldsymbol{V}_1 & \cdots & \boldsymbol{V}_p \end{pmatrix}.$$

Then the ASE of $\boldsymbol{K}(\varepsilon)$ is given by

$$\underline{\boldsymbol{K}(\epsilon)} = \sum_{i=0}^{p} \boldsymbol{Q}_i \boldsymbol{S}_i \boldsymbol{Q}_i^\top \varepsilon^{2p}, \tag{5.15}$$

where $\boldsymbol{S}_i$ are the Schur complements of the block matrix $\boldsymbol{H}$ defined in (4.5).

Proof. By Lemma 5.10, $\boldsymbol{K}(\varepsilon)$ has factorization (5.14), where the matrix $\boldsymbol{W}$ is strictly positive definite by Lemma 5.8. Therefore, we can apply Theorem 4.2, and the statements follow from Theorem 4.2. ∎

In the case when the Vandermonde matrices are full rank, then we can use the simplified expression for the Schur complements. These expressions lead to the following corollary, settling Conjecture 1 in [5] in the smooth case.

COROLLARY 5.13. *Let $j \leq p$ be such that the matrix $\boldsymbol{V}_{\leq j}$ is full-column rank. Then*

1. *the number of eigenvalues in the j-th block (with the order ε^{2j}) is exactly equal to $\mathbb{H}_{j,d}$ (the number of monomials of degree j);*
2. *the Schur complement $\boldsymbol{S}_j$ admits the simplified expression (4.7), and therefore, the leading coefficients of the eigenvalues are given by the eigenvalues of $\boldsymbol{S}_j$ and the "leading eigenvectors" (in the sense of Definition 2.9) are given by the eigenvectors of*

$$\boldsymbol{Q}_j \boldsymbol{S}_j \boldsymbol{Q}_j^t = \begin{cases} \boldsymbol{V}_0 \boldsymbol{W}_{0,0} \boldsymbol{V}_0^\top, & j = 0, \\ \boldsymbol{Q}_j \boldsymbol{Q}_j^t \boldsymbol{V}_j \left(\boldsymbol{W}_{j,j} - \boldsymbol{W}_{j,<j} (\boldsymbol{W}_{<j,<j})^{-1} \boldsymbol{W}_{<j,j} \right) \boldsymbol{V}_j^\top \boldsymbol{Q}_j \boldsymbol{Q}_j^t, & j > 0. \end{cases} \tag{5.16}$$

Proof. This follows from the fact that the full rank property $\boldsymbol{V}_{\leq j}$ implies that all the diagonal blocks in the QR decomposition ($\boldsymbol{R}_{i,i}$ for $i \leq j$) are square (i.e. $b_i = a_i$ in (4.4)). Therefore we obtain the simplified expression (4.7). The rest follows from the fact that $\boldsymbol{Q}_j^t \boldsymbol{V}_j = \boldsymbol{R}_{j,j}$. ∎

Note that Theorem 5.12 applies also in the the non-unisolvent case (the case when the matrices $\boldsymbol{V}_{\leq j-1}$ may not be of full column rank). However, the expressions for the Schur complements are more complicated than the ones in Corollary 5.13, see subsection 5.5.

5.4. Finitely smooth case . The finitely smooth case appears when there is no such p that satisfies the conditions of Lemma 5.10. This happens, for example, if $\operatorname{rank} \boldsymbol{V}_{\leq r-1} < n$ for a radial kernel of the regularity index r. This case can be treated with the following proposition

PROPOSITION 5.14. *Let ψ be with regularity r and* $\operatorname{rank} \boldsymbol{V}_{\leq r-1} < n$. *Then the kernel matrix has the following asymptotic form in $\varepsilon \to 0$:*

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V} \boldsymbol{\Delta}(\varepsilon) (\boldsymbol{W} + o(1)) \boldsymbol{\Delta}(\varepsilon) \boldsymbol{V}^\top, \tag{5.17}$$

- $\boldsymbol{V} = [\boldsymbol{V}_{\leq r-1} \ \boldsymbol{A}]$ *where $\boldsymbol{A} \in \mathbb{R}^{n \times a_j}$, is an arbitrary full column rank matrix such that* $\operatorname{span} \boldsymbol{V} = \mathbb{R}^n$ *(which implies (with $c = n - \operatorname{rank} \boldsymbol{V}_{\leq r-1}$).*
- $\boldsymbol{\Delta}(\varepsilon)$ *is a diagonal scaling matrix, with block structure*

$$\boldsymbol{\Delta}(\varepsilon) = \begin{pmatrix} 1 & & & & \\ & \varepsilon^1 \boldsymbol{I}_{\mathbb{H}_{1,d}} & & & \\ & & \ddots & & \\ & & & \varepsilon^{r-1} \boldsymbol{I}_{\mathbb{H}_{r-1,d}} & \\ & & & & \varepsilon^{r-\frac{1}{2}} \boldsymbol{I}_c \end{pmatrix},$$

where each block of integer valuation t is of size $\mathbb{H}_{i,d}$. The only block with fractional valuation is the last one.

- $\boldsymbol{W}$ *is an extended "Wronskian" matrix with the following structure:*

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_{\leq r-1,\leq r-1} & \boldsymbol{0} \\ \boldsymbol{0} & \psi_{2r-1}\boldsymbol{A}^\dagger \boldsymbol{D}^{(2r-1)}(\boldsymbol{A}^\dagger)^\mathsf{T} \end{pmatrix}. \tag{5.18}$$

Proof. The proof is given in section 8. ∎

THEOREM 5.15. *Let $\boldsymbol{K}(\varepsilon)$ satisfy the conditions of Proposition 5.14, and, in addition* $\operatorname{rank} \boldsymbol{V}_{\leq r-1} < n$. *Let $\boldsymbol{Q}_0, \dots, \boldsymbol{Q}_{r-1}$ and $\boldsymbol{R}_{0,0}, \dots, \boldsymbol{R}_{r-1,r-1}$ come from the block QR factorization of the matrix*

$$\boldsymbol{V} = \boldsymbol{V}_{\leq r-1} = \begin{pmatrix} \boldsymbol{V}_0 & \boldsymbol{V}_1 & \cdots & \boldsymbol{V}_{r-1} \end{pmatrix}.$$

Then the ASE of $\boldsymbol{K}(\varepsilon)$ is given by

$$\underline{\boldsymbol{K}(\epsilon)} = \sum_{i=0}^{r-1} \boldsymbol{Q}_i \boldsymbol{S}_i \boldsymbol{Q}_i^\top \varepsilon^{2i} + \varepsilon^{2r-1}\psi_{2r-1}\boldsymbol{A}\boldsymbol{A}^\dagger \boldsymbol{D}^{(2r-1)}\boldsymbol{A}\boldsymbol{A}^\dagger, \tag{5.19}$$

where the terms $\boldsymbol{S}_0, \boldsymbol{S}_1, \dots$ have exactly the same form as in Theorem 5.12.

Proof. By Proposition 5.14, $\boldsymbol{K}(\varepsilon)$ has factorization (5.14), where the matrix $\boldsymbol{W}$ is strictly positive definite by Lemma 5.8. Finally we note that, since $\boldsymbol{A}^t \boldsymbol{V}_{\leq r-1}$, the QR decomposition (4.4) of the matrix $\boldsymbol{V}$ in Proposition 5.14 from given by

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{Q}_0 & \dots & \boldsymbol{Q}_{r-1} & \boldsymbol{A} \end{pmatrix} \begin{pmatrix} \boldsymbol{R}_{0,0} & \dots & \boldsymbol{R}_{0,r-1} & \\ & \ddots & \vdots & \\ & & \boldsymbol{R}_{p,p} & \\ & & & \boldsymbol{I}_c \end{pmatrix},$$

with $c = n - \operatorname{rank} \boldsymbol{V}_{\leq r-1}$ hence, we can again apply Theorem 4.2, and the statements of the theorem follows from Theorem 4.2, 1–2. ∎

As a corollary of Theorem 5.15, we recover both [5, Conjecture 1] and [5, Theorem 6.3] in the finite smoothness case.

COROLLARY 5.16.
1. *For $j \leq r-1$, if $\boldsymbol{V}_{\leq j}$ if full column rank, then the number of eigenvalues of the order ε^{2j} is exactly equal to $\mathbb{H}_{j,d}$, and the expressions for the Schur complement are as in Corollary 5.13.*
2. *If, in addition $\boldsymbol{V}_{\leq r-1}$ is full column rank, then there are exactly $n - \mathbb{P}_{r-1,d}$ eigenvalues of order ε^{2r-1} corresponding to the ASE term $\psi_{2r-1}\boldsymbol{A}\boldsymbol{A}^\dagger \boldsymbol{D}^{(2r-1)}\boldsymbol{A}\boldsymbol{A}^\dagger$.*

Proof. The first part of the corollary is proved similarly to Corollary 5.13. The second part follows from Lemma 5.9, which implies that the $\boldsymbol{A}^\dagger \boldsymbol{D}^{(2r-1)}\boldsymbol{A}$ has rank c (and so has the matrix $\boldsymbol{A}\boldsymbol{A}^\dagger \boldsymbol{D}^{(2r-1)}\boldsymbol{A}\boldsymbol{A}^\dagger$). ∎

5.5. Non-unisolvent case. Finally, we make some remarks on the non-unisolvent case, i.e. the case where the Vandermonde matrix $\boldsymbol{V}_{\leq s}$ are rank-deficient for some s. We make first the remark on the ranks on the Vandermonde matrices.

LEMMA 5.17. *If all the points $\boldsymbol{x}_i$, $i \in \{1, \dots, n\}$ are distinct, then we have that* $\operatorname{rank} \boldsymbol{V}_{\leq s} > \operatorname{rank} \boldsymbol{V}_{\leq s-1}$ *for all $1 \leq s < n$.*

Proof. If all the points are distinct, then there exists a vector $\boldsymbol{a} \in \mathbb{R}^n$ such that all $t_i = \boldsymbol{a}^t \boldsymbol{x}_i$ are distinct. Note that the univariate Vandermonde matrices $\widetilde{\boldsymbol{V}}_{\leq s} = [t_i^j]_{i=1,j=0}^{n,s}$ are full column rank for $s \leq n-1$. Note that the last column of $\widetilde{\boldsymbol{V}}_{\leq s}$ lies in $\operatorname{span} \boldsymbol{V}_s$. Hence, for all $1 \leq s \leq n-1$, $\operatorname{span} \boldsymbol{V}_s$ contains at least one vector that does not belong to $\operatorname{span} \boldsymbol{V}_{\leq s-1}$, and therefore $\operatorname{rank} \boldsymbol{V}_{\leq s} > \operatorname{rank} \boldsymbol{V}_{\leq s-1}$. ∎

Note that Theorems 5.12 and 5.15 still apply, and as a special case, we generalize the results in [25] (see [25, Theorem 8] where the number of eigenvalues of given order is provided for analytic kernels).

COROLLARY 5.18. *Assume that the conditions of Theorem 5.12 or Theorem 5.15 hold and $j \leq r-1$, and all the points $\boldsymbol{x}_i$ are distinct.*

1. *If* $\operatorname{rank} \boldsymbol{V}_{\leq j}$ *is rank deficient but* $\operatorname{rank} \boldsymbol{V}_{\leq j-1}$ *is full column rank (rank* $\mathbb{P}_{j-1,d}$*), then there are* $\operatorname{rank} \boldsymbol{V}_{\leq j} - \mathbb{P}_{j-1,d}$ *eigenvalues of the degree* ε^{2j} *and the formula* (5.16) *is still valid for the corresponding term of the ASE.*
2. *If* $\operatorname{rank} \boldsymbol{V}_{\leq j}$ *and* $\operatorname{rank} \boldsymbol{V}_{\leq j-1}$ *are both rank deficient (i.e.,* $\operatorname{rank} \boldsymbol{V}_{\leq j} < \mathbb{P}_{j,d}, \operatorname{rank} \boldsymbol{V}_{\leq j-1} < \mathbb{P}_{j-1,d}$*, then there are* $\operatorname{rank} \boldsymbol{V}_{\leq j} - \operatorname{rank} \boldsymbol{V}_{\leq j-1}$ *eigenvalues of degree* ε^{2j} *and the matrix* $\boldsymbol{S}_j$ *is obtained from the Schur complement (with respect to the last block) of*

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{R}_{0,0}\boldsymbol{W}_{0,0}\boldsymbol{R}_{0,0}^t & \dots & \boldsymbol{R}_{0,0}\boldsymbol{W}_{0,j}\boldsymbol{R}_{j,j}^t \\ \vdots \dots & \dots & \\ \boldsymbol{R}_{j,j}\boldsymbol{W}_{j,0}\boldsymbol{R}_{0,0}^t & \dots & \boldsymbol{R}_{j,j}\boldsymbol{W}_{j,j}\boldsymbol{R}_{j,j}^t \end{pmatrix}.$$

Proof. The statements about the numbers of the eigenvalues follow from Remark 4.1, where b_i gives the number of rows in the diagonal block $\boldsymbol{Q}_{i,i}$ of the QR decomposition. Note that b_i is nonzero by Lemma 5.17. Finally, the expressions for the Schur complement follow from combining (3.5) and (4.5). ∎

Remark 5.19. The Vandermonde matrices $\boldsymbol{V}_{\leq j}$ are rank-deficient if the points $x_1 \dots x_n$ are sampled from an algebraic variety having polynomial equations of degree $\leq j$. For example, consider points sampled on a circle (i.e., $x_1^2 + x_2^2 = 1$), or, in general a conic section, in $d = 2$. Then we have that

$$\operatorname{rank} \boldsymbol{V}_{\leq j} = 2j + 1,$$

which is smaller than $\mathbb{P}_{j,2} = \binom{j+2}{2}$ as long as $j \geq 2$. In this case, Corollary 5.18 gives the limiting eigenvector and eigenvalues for kernel matrices corresponding to nodes on a circle.

The example of the circle can be generalized to the case when the points $\boldsymbol{x}_j$ lie on an algebraic variety. In this case, the ranks of the Vandermonde matrix are connected to the Hilbert function [10, Ch. 9] of the corresponding polynomial ideal (see also [4] for examples of Hilbert functions). Note that for some particular algebraic varieties (e.g., spheres) it may be more beneficial to use some predefined basis (e.g., spherical harmonics [14]) instead of multivariate monomials.

5.6. Numerical illustration. We illustrate our results with two examples in dimension 2, and contrast unisolvent to non-unisolvent sets.

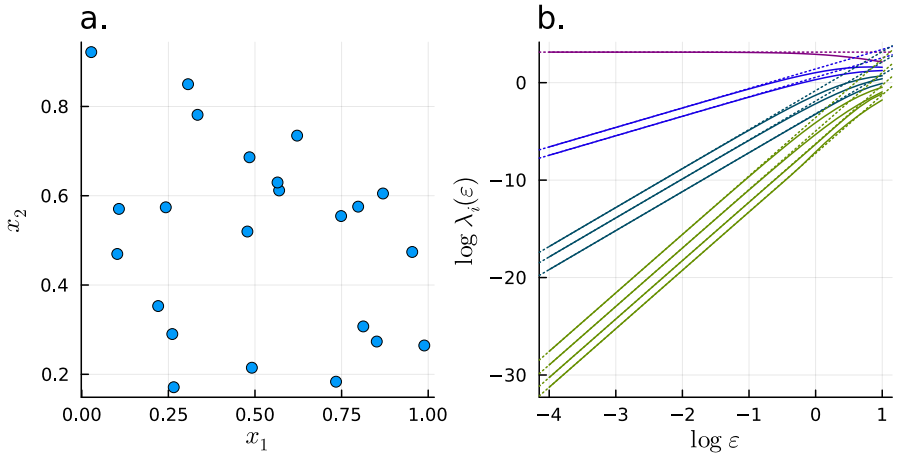

FIG. 5.1. *Eigenvalues of a Gaussian kernel matrix for a unisolvent node set.* ***a.*** *Nodes drawn i.i.d. from the unit square* ***b.*** *The* 10 *largest eigenvalues of* $\boldsymbol{K}(\varepsilon)$ *for this node set, as a function of* ε *(in log-log scale). Groups of eigenvalues with different valuations appear in different colors. There is* 1 *eigenvalue with valuation* 0, 2 *eigenvalues with valuation* 2, 3 *eigenvalues with valuation* 4, 4 *eigenvalues with valuation* 6, *etc. The asymptotic approximation of an eigenvalue, keeping only the leading term, corresponds to a line in log-log space. These approximations are shown here as dotted lines (note the varying slopes, correspond to different valuations).*

In our first example (fig. 5.1), nodes are drawn i.i.d. uniformly from the unit square. In such configurations, the node set is unisolvent almost surely. The Gaussian kernel (eq. (1.1)) is completely smooth, and cor. 5.13 applies. As a consequence, the successive blocks of eigenvalues of with asymptotic behaviour in $\varepsilon^0, \varepsilon^2, \varepsilon^4, \dots$ have size equal to $\mathbb{H}_{0,2}, \mathbb{H}_{1,2}, \mathbb{H}_{2,d}, \mathbb{H}_{3,d}, \mathbb{H}_{4,d}, \dots = 1, 2, 3, 4, \dots$. In general the i-th block has valuation $2i$ and is of size $\mathbb{H}_{i,d} = i + 1$. The leading coefficients can be obtained from eq. (5.16). In fig. 5.1, we show the eigenvalues of the example matrix as a function of ε, along with the asymptotic approximations.

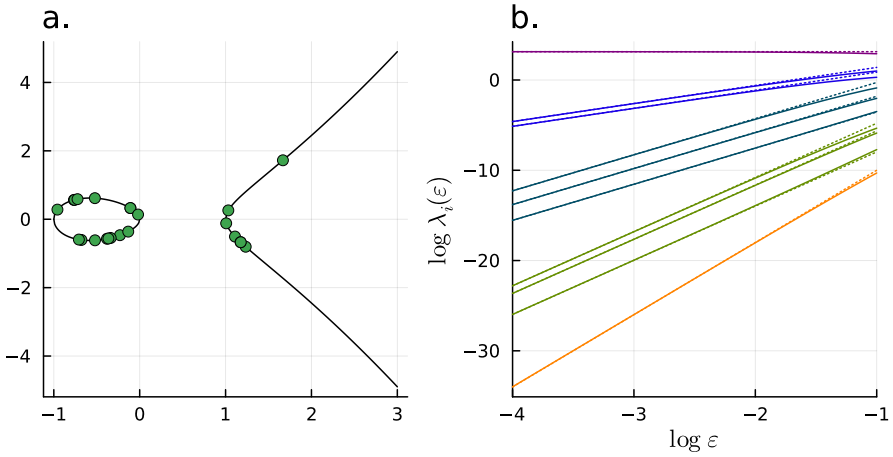

FIG. 5.2. *Eigenvalues of a Gaussian kernel matrix for a non-unisolvent node set.* ***a.*** *Nodes drawn i.i.d. from the affine variety $x_2^2 = x_1^3 - x_1$ (the underlying curve is shown as a solid line). This node set is non-unisolvent at degree 3 and higher, see text.* ***b.*** *The 10 largest eigenvalues of $\boldsymbol{K}(\varepsilon)$ for this node set, as a function of ε (in log-log scale). Groups of eigenvalues with different valuations appear in different colors. There is 1 eigenvalue with valuation 0, 2 eigenvalues with valuation 2, 3 eigenvalues with valuation 4, but only 3 eigenvalues with valuation 6. All further groups of eigenvalues are also of size ≤ 3. Compared to the unisolvent set, the eigenvalue groups are smaller starting at valuation 8, which corresponds to polynomials of degree 4.*

In our second example (fig. 5.2), the nodes are sampled i.i.d. from the algebraic curve $x_2^2 = x_1^3 - x_1$. This is an affine variety of dimension one and degree 3, and the node sets is non-unisolvent for polynomials of degree 3 and higher. More precisely, the blocks $\boldsymbol{R}_{i,i}$ in the QR decomposition of the Vandermonde matrix (which have size $\mathbb{H}_{i,2} = i - 1, i > 0$) have rank ≤ 3 for $i \geq 3$. Compared to the unisolvent case, we still have blocks of eigenvalues with asymptotic behaviour in $\varepsilon^0, \varepsilon^2, \varepsilon^4, \ldots$ but starting at valuation 8 (which corresponds to degree 4) these blocks are all of size ≤ 3. The leading coefficients of the eigenvalues can be obtained from cor. 5.18.

Remark 5.20. Just before the submission of this manuscript, we became aware of the paper by Diab and Batenkov [11] that investigated the asymptotic eigenvalues in the non-unisolvent case, using the tools similar to the ones in [5].

6. ASE in the degenerate case. The goal of this section is to address the general case of positive definite $\boldsymbol{K}(\varepsilon)$. In fact, not every matrix $\boldsymbol{K}(\varepsilon)$ can be reduced to the generalized kernel form (4.1). As an example, consider the following matrix:

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & \varepsilon^1 & \varepsilon^1 \\ \varepsilon^1 & 2\varepsilon^2 & \varepsilon^2 \\ \varepsilon^1 & \varepsilon^2 & \varepsilon^2 + \varepsilon^3 \end{pmatrix}.$$

Indeed, we can represent the matrix in the diagonal scaling form as

$$\boldsymbol{K}(\varepsilon) = \begin{pmatrix} 1 & & \\ & \varepsilon^1 & \\ & & \varepsilon^1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 + \varepsilon^1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & \varepsilon^1 & \\ & & \varepsilon^1 \end{pmatrix}. \tag{6.1}$$

We see that the last Schur complement of the last block is

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

therefore, by theorem Theorem 3.2 the first two eigenvalues are $O(1)$, $O(\varepsilon^2)$ and the last eigenvalue is at least $O(\varepsilon^3)$. In addition, the first two limiting eigenvectors are

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

respectively, and therefore the limiting eigenvectors are given by $\boldsymbol{U}_0 = \boldsymbol{I}$.

However, we cannot bring the matrix to the form (4.1) with the diagonal scaling

$$\begin{pmatrix} 1 & & \\ & \varepsilon^1 & \\ & & \varepsilon^{\frac{3}{2}} \end{pmatrix},$$

thus we cannot use Theorem 3.2 to get all the information on the ASE. In order to deal with such cases, we are going to propose an iterative reduction strategy, also based on Schur complements.

6.1. Schur complement in the diagonally scaled form. Then we are able to derive the following result, that helps us to continue the reduction beyond the case in Theorem 3.2.

THEOREM 6.1. *Let $\boldsymbol{K}(\varepsilon)$ be partitioned as*

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{\Delta}(\varepsilon)\boldsymbol{H}(\varepsilon)\boldsymbol{\Delta}(\varepsilon),$$

$$\boldsymbol{H}(\varepsilon) = \begin{pmatrix} \boldsymbol{H}_{11}(\varepsilon) & \boldsymbol{H}_{12}(\varepsilon) \\ \boldsymbol{H}_{21}(\varepsilon) & \boldsymbol{H}_{22}(\varepsilon) \end{pmatrix}$$

where $\boldsymbol{H}_{11}(\varepsilon)\,\boldsymbol{H}_{22}(\varepsilon)$ are $m \times m$ and $(n-m)\times(n-m)$ symmetric, $\boldsymbol{H}_{21}(\varepsilon) = \boldsymbol{H}_{21}^{\mathsf{T}}(\varepsilon)$, $\boldsymbol{H}_{11}(0)$ is non-singular (rank m), and

$$\boldsymbol{\Delta}(\varepsilon) = \begin{pmatrix} \boldsymbol{\Delta}_m(\varepsilon) & \\ & \varepsilon^s \boldsymbol{I}_{n-m} \end{pmatrix}, \quad \boldsymbol{\Delta}_m(\varepsilon) = \operatorname{diag}(\varepsilon^{\gamma_1}, \dots, \varepsilon^{\gamma_m}). \tag{6.2}$$

where $\gamma_1 \le \dots \le \gamma_m < s$. Then $\underline{\boldsymbol{K}}(\varepsilon) = \underline{\boldsymbol{K}'}(\varepsilon)$, with

$$\boldsymbol{K}'(\varepsilon) = \begin{pmatrix} \boldsymbol{\Delta}_m(\varepsilon)\boldsymbol{H}_{11}(\varepsilon)\boldsymbol{\Delta}_m(\varepsilon) & 0 \\ 0 & \varepsilon^{2s}(\boldsymbol{H}_{22}(\varepsilon) - \boldsymbol{H}_{21}(\varepsilon)\boldsymbol{H}_{11}^{-1}(\varepsilon)\boldsymbol{H}_{12}(\varepsilon)) \end{pmatrix}, \tag{6.3}$$

Before proving Theorem 6.1, we show an example of such a reduction.

Example 6.2. We continue the example from the beginning of the section 6. Applying Theorem 6.1 in (6.1) with $\Delta(\varepsilon)$ as in (6.1), we get that the Schur complement becomes

$$\begin{pmatrix} 2 & 1 \\ 1 & 1+\varepsilon \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix},$$

hence the ASE of the matrix in (6.1) is equal to the ASE of

$$\begin{pmatrix} 1 & & \\ & \varepsilon^2 & \\ & & \varepsilon^3 \end{pmatrix},$$

which is already in the ASE form.

Remark 6.3. Note that Theorem 6.1 can be used to obtain an ASE of any symmetric analytic matrix in an iterative fashion. Indeed, take the leading term of the right lower block in (6.3)

$$\boldsymbol{A}(\varepsilon) = \varepsilon^{2s}(\boldsymbol{H}_{22}(\varepsilon) - \boldsymbol{H}_{21}(\varepsilon)\boldsymbol{H}_{11}^{-1}(\varepsilon)\boldsymbol{H}_{12}(\varepsilon)),$$

and assume that $\operatorname{val}(\boldsymbol{A}(\varepsilon)) = \gamma \ge 2s$. Then the leading term of the matrix $\boldsymbol{A}(\varepsilon)/\varepsilon^\gamma$ will describe the term of the ASE for the group of eigenvalues of the next valuation ε^γ. By choosing an appropriate rotation $\boldsymbol{Q}$, this matrix can be brought to

$$\boldsymbol{Q}\boldsymbol{A}(\varepsilon)\boldsymbol{Q}^t = \begin{pmatrix} \mathcal{O}(\varepsilon^\gamma) & \mathcal{O}(\varepsilon^{\gamma+1}) \\ \mathcal{O}(\varepsilon^{\gamma+1}) & \mathcal{O}(\varepsilon^{\gamma+1}) \end{pmatrix},$$

which is in a diagonally scaled form and thus Theorem 6.1 can be applied again (combined with Theorem 3.2 if necessary).

6.2. Proof of Theorem 6.1. In order to prove Theorem 6.1, we will need several lemmas. The first lemma slightly generalizes Lemma 2.14.

LEMMA 6.4. *Let $\boldsymbol{\Delta}(\varepsilon)$ be as in (6.2), with $\gamma_1 \le \dots \le \gamma_m < s$, the matrix $\widetilde{\Lambda}(\varepsilon)$ be an analytic diagonal matrix*

$$\widetilde{\boldsymbol{\Lambda}}(\varepsilon) = \begin{pmatrix} \widetilde{\lambda}_1 + \mathcal{O}(\varepsilon) & & \\ & \ddots & \\ & & \widetilde{\lambda}_n + \mathcal{O}(\varepsilon) \end{pmatrix},$$

so that $\widetilde{\lambda}_k \neq 0$ for $k \le m$, and define

$$\boldsymbol{\Lambda}(\varepsilon) = \boldsymbol{\Delta}(\varepsilon)\widetilde{\boldsymbol{\Lambda}}(\varepsilon)\boldsymbol{\Delta}(\varepsilon).$$

Then for general τ (not equal to any of $\{-\widetilde{\lambda_k}\}_{k=m+1}^n$) and for any matrix $\boldsymbol{A}(\varepsilon) = \mathcal{O}(\varepsilon)$, the regularized inverse matrix has the leading term

$$\boldsymbol{\Lambda}(\varepsilon)(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^{2s}(\boldsymbol{I} + \boldsymbol{A}(\varepsilon)))^{-1} = \boldsymbol{I} - \tau\varepsilon^{2s}(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^{2s}(\boldsymbol{I} + \boldsymbol{A}(\varepsilon)))^{-1} = \underline{\boldsymbol{\Lambda}} + \mathcal{O}(\varepsilon),$$

where the matrix $\underline{\boldsymbol{\Lambda}}$ is equal to

$$\underline{\boldsymbol{\Lambda}} = \operatorname{diag}\left(1, \dots, 1, \frac{\widetilde{\lambda}_{m+1}}{\widetilde{\lambda}_{m+1} + \tau}, \dots, \frac{\widetilde{\lambda}_n}{\widetilde{\lambda}_n + \tau}\right).$$

Note that there may be zeros among $\{\widetilde{\lambda_k}\}_{k=m+1}^n$ in which case the zeros also appear on the diagonal of $\underline{\boldsymbol{\Lambda}}$.

Proof. With some abuse of notation we denote $\widetilde{\boldsymbol{\Lambda}}_0 = \widetilde{\boldsymbol{\Lambda}}(0)$, and take the matrix $\widehat{\boldsymbol{\Lambda}}(\varepsilon) := \widetilde{\boldsymbol{\Lambda}}(\varepsilon) + \tau\varepsilon^{2s}\boldsymbol{A}(\varepsilon)$. Then, it is easy to see that we can put the matrix $\widehat{\boldsymbol{\Lambda}}(\varepsilon)$ in the diagonally scaled form:

$$\widehat{\boldsymbol{\Lambda}}(\varepsilon) = \boldsymbol{\Delta}(\varepsilon)(\widetilde{\boldsymbol{\Lambda}}(\varepsilon) + \underbrace{\tau\varepsilon^{2s}\boldsymbol{\Delta}(\varepsilon)^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{\Delta}(\varepsilon)^{-1}}_{\mathcal{O}(\varepsilon)})\boldsymbol{\Delta}(\varepsilon) = \boldsymbol{\Delta}(\varepsilon)(\widetilde{\boldsymbol{\Lambda}}_0 + \mathcal{O}(\varepsilon))\boldsymbol{\Delta}(\varepsilon).$$

Then we have that

$$\boldsymbol{I} - \tau\varepsilon^{2s}(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^{2s}(\boldsymbol{I} + \boldsymbol{A}(\varepsilon)))^{-1} = \boldsymbol{I} - \tau\varepsilon^{2s}(\widehat{\boldsymbol{\Lambda}}(\varepsilon) + \tau\varepsilon^{2s}\boldsymbol{I})^{-1} = \widehat{\boldsymbol{\Lambda}}(\varepsilon)(\widehat{\boldsymbol{\Lambda}}(\varepsilon) + \tau\varepsilon^{2s}\boldsymbol{I})^{-1},$$

hence we can use Lemma 3.15 (applied to $\boldsymbol{K}(\varepsilon) = \widetilde{\boldsymbol{\Lambda}}(\varepsilon)$) in diagonally scaled form

$$\widehat{\boldsymbol{\Lambda}}(\varepsilon)(\widehat{\boldsymbol{\Lambda}}(\varepsilon) + \tau\varepsilon^{2s}\boldsymbol{I})^{-1} = \underline{\boldsymbol{\Lambda}} + \mathcal{O}(\varepsilon),$$

which completes the proof. ∎

6.2.1. Main result. Then Lemma 6.4 implies the following.

COROLLARY 6.5. *Let $\boldsymbol{Q}(\varepsilon) = \boldsymbol{I} + O(\varepsilon)$. Then the congruence with $\boldsymbol{Q}(\varepsilon)$ preserves the asymptotic spectral equivalent for any matrix:*

$$\underline{\boldsymbol{Q}(\varepsilon)\boldsymbol{K}(\varepsilon)\boldsymbol{Q}^{\mathsf{T}}(\varepsilon)} = \underline{\boldsymbol{K}(\epsilon)}.$$

Proof. Denote $\boldsymbol{K}'(\varepsilon) = \boldsymbol{Q}(\varepsilon)\boldsymbol{K}(\varepsilon)\boldsymbol{Q}^{\mathsf{T}}(\varepsilon)$. Then we have that

$$\begin{aligned}
\boldsymbol{M}_s'(\varepsilon) &= \boldsymbol{I} - \tau\varepsilon^s(\boldsymbol{K}'(\varepsilon) + \tau\varepsilon^s\boldsymbol{I})^{-1} = \boldsymbol{I} - \tau\varepsilon^s(\boldsymbol{Q}(\varepsilon)\boldsymbol{U}(\varepsilon)\boldsymbol{\Lambda}(\varepsilon)\boldsymbol{U}^{\mathsf{T}}(\varepsilon)\boldsymbol{Q}^{\mathsf{T}}(\varepsilon) + \tau\varepsilon^s\boldsymbol{I})^{-1} \\
&= \boldsymbol{I} - \tau\varepsilon^s(\boldsymbol{Q}(\varepsilon)\boldsymbol{U}(\varepsilon)(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^s(\boldsymbol{I} + O(\varepsilon)))\boldsymbol{U}^{\mathsf{T}}(\varepsilon)\boldsymbol{Q}^{\mathsf{T}}(\varepsilon))^{-1} \\
&= \boldsymbol{I} - \tau\varepsilon^s\boldsymbol{Q}^{-\mathsf{T}}(\varepsilon)\boldsymbol{U}(\varepsilon)(\boldsymbol{\Lambda}(\varepsilon) + \tau\varepsilon^s(\boldsymbol{I} + O(\varepsilon)))^{-1}\boldsymbol{U}^{\mathsf{T}}(\varepsilon)\boldsymbol{Q}^{-1}(\varepsilon) \\
&= \boldsymbol{I} - (\boldsymbol{I} + \mathcal{O}(\varepsilon)(\boldsymbol{U}_0 + O(\varepsilon))(\boldsymbol{I} - \underline{\boldsymbol{\Lambda}} + O(\varepsilon))(\boldsymbol{U}_0^{\mathsf{T}} + O(\varepsilon))(\boldsymbol{I} + \mathcal{O}(\varepsilon)) \\
&= \boldsymbol{U}_0\underline{\boldsymbol{\Lambda}}\boldsymbol{U}_0^{\mathsf{T}} + O(\varepsilon) = \boldsymbol{M}_s(\varepsilon) + O(\varepsilon),
\end{aligned}$$

where the last but one equality is by Lemma 6.4. By matching the main terms of $\boldsymbol{M}_s(\varepsilon)$ and $\boldsymbol{M}_s'(\varepsilon)$ for all possible s, we conclude that the ASE of $\boldsymbol{K}'(\varepsilon)$ and $\boldsymbol{K}(\varepsilon)$ are equal. ∎

Proof of Theorem 6.1. Consider the following analytic matrix

$$\boldsymbol{Q}(\varepsilon) = \begin{pmatrix} \boldsymbol{I}_m & \boldsymbol{0} \\ -\varepsilon^s \boldsymbol{H}_{21}(\varepsilon)\boldsymbol{H}_{11}^{-1}(\varepsilon)\boldsymbol{\Delta}_m^{-1}(\varepsilon) & \boldsymbol{I}_{n-m} \end{pmatrix}$$

Then we can verify that

$$\begin{aligned}
&\boldsymbol{Q}(\varepsilon)\boldsymbol{K}(\varepsilon)\boldsymbol{Q}^{\mathsf{T}}(\varepsilon) = \\
&= \begin{pmatrix} \boldsymbol{I}_m & \boldsymbol{0} \\ -\varepsilon^s \boldsymbol{H}_{21}(\varepsilon)\boldsymbol{H}_{11}^{-1}(\varepsilon)\boldsymbol{\Delta}_m^{-1}(\varepsilon) & \boldsymbol{I}_{n-m} \end{pmatrix} \boldsymbol{\Delta}(\varepsilon) \begin{pmatrix} \boldsymbol{H}_{11}(\varepsilon) & \boldsymbol{H}_{12}(\varepsilon) \\ \boldsymbol{H}_{21}(\varepsilon) & \boldsymbol{H}_{22}(\varepsilon) \end{pmatrix} \boldsymbol{\Delta}(\varepsilon) \begin{pmatrix} \boldsymbol{I}_m & -\varepsilon^s \boldsymbol{\Delta}_m^{-1}(\varepsilon)\boldsymbol{H}_{11}^{-1}(\varepsilon)\boldsymbol{H}_{12}(\varepsilon) \\ \boldsymbol{0} & \boldsymbol{I}_{n-m} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Delta}_m(\varepsilon)\boldsymbol{H}_{11}(\varepsilon)\boldsymbol{\Delta}_m(\varepsilon) & 0 \\ 0 & \varepsilon^{2s}(\boldsymbol{H}_{22}(\varepsilon) - \boldsymbol{H}_{21}(\varepsilon)\boldsymbol{H}_{11}^{-1}(\varepsilon)\boldsymbol{\Delta}_m^{-1}(\varepsilon)\boldsymbol{\Delta}_m(\varepsilon)\boldsymbol{H}_{11}(\varepsilon)\boldsymbol{\Delta}_m(\varepsilon)\boldsymbol{\Delta}_m^{-1}(\varepsilon)\boldsymbol{H}_{12}(\varepsilon)) = \boldsymbol{K}'(\varepsilon), \end{pmatrix}
\end{aligned}$$

and the proof is complete by Corollary 6.5. ∎

7. Conclusion. We hope to have convinced the reader that theorems 3.2, 4.2 and 6.1, can be used to simplify the analysis of matrix perturbations. One noteworthy limitation is that we have assumed that the perturbations are analytic, i.e. the classical framework used by Rellich and Kato. This limitation can be lifted, if one instead looks at the matrix $\boldsymbol{K}(\varepsilon)$ as admitting an asymptotic series (which need not be a power series). We intend to extend our results in this direction in future work.

8. Appendix.

Proof of Lemma 5.10. As shown in the proofs of [5, Theorems 4.5 and 6.3], under such assumptions, the kernel matrix has expansion

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V}_{\le p-1}\boldsymbol{\Delta}_{p-1}\boldsymbol{W}_{\le p-1}\boldsymbol{\Delta}_{p-1}\boldsymbol{V}_{\le p-1}^{\mathsf{T}} + \varepsilon^p(\boldsymbol{V}_{\le p-1}\boldsymbol{\Delta}_{p-1}\boldsymbol{W}_1(\varepsilon) + \boldsymbol{W}_2(\varepsilon)\boldsymbol{\Delta}_{p-1}\boldsymbol{V}_{\le p-1}^{\mathsf{T}}) + \varepsilon^{2p-1}(\boldsymbol{W}_3(\varepsilon)), \tag{8.1}$$

where $\boldsymbol{W}_3(\varepsilon) = \varepsilon^{2p-1}(\psi_{2p-1}\boldsymbol{D}^{(2p-1)} + \mathcal{O}(\varepsilon))$ and $\psi_{2p-1} = 0$ if $p < r$. Note that since $\boldsymbol{V} = \boldsymbol{V}_{\le p-1}$ is full row rank, we have

$$\boldsymbol{I}_n = \boldsymbol{V}\boldsymbol{\Delta}_{p-1}\boldsymbol{\Delta}_{p-1}^{-1}\boldsymbol{V}^{\dagger}, \tag{8.2}$$

hence, we can rewrite

$$\varepsilon^p(\boldsymbol{V}_{\le p-1}\boldsymbol{\Delta}_{p-1}\boldsymbol{W}_1(\varepsilon) + \boldsymbol{W}_2(\varepsilon)\boldsymbol{\Delta}_{p-1}\boldsymbol{V}_{\le p-1}^{\mathsf{T}}) = \boldsymbol{V}\boldsymbol{\Delta}_{p-1}(\widetilde{\boldsymbol{W}}_1(\varepsilon) + \widetilde{\boldsymbol{W}}_2(\varepsilon)+)\boldsymbol{\Delta}_{p-1}\boldsymbol{V}^t,$$

where

$$\widetilde{\boldsymbol{W}}_2(\varepsilon) = \varepsilon^p\boldsymbol{\Delta}(\varepsilon)^{-1}\boldsymbol{V}^{\dagger}\boldsymbol{W}_2(\varepsilon) = \mathcal{O}(\varepsilon), \quad \widetilde{\boldsymbol{W}}_1(\varepsilon) = \varepsilon^p\boldsymbol{W}_1(\varepsilon)(\boldsymbol{V}^{\dagger})^{\mathsf{T}}\boldsymbol{\Delta}(\varepsilon)^{-1} = \mathcal{O}(\varepsilon).$$

Similarly, for $\boldsymbol{W}_3(3)$, we have

$$\varepsilon^{2p-1}\boldsymbol{W}_3(3) = \boldsymbol{V}\boldsymbol{\Delta}_{p-1}\underbrace{\varepsilon^{p-1}\boldsymbol{\Delta}_{p-1}^{-1}\boldsymbol{V}^{\dagger}(\varepsilon\boldsymbol{W}_3(\varepsilon))(\boldsymbol{V}^{\dagger})^t\varepsilon^{p-1}\boldsymbol{\Delta}_{p-1}^{-1}}_{\widetilde{\boldsymbol{W}}_3(\varepsilon)=\mathcal{O}(\varepsilon)}\boldsymbol{\Delta}_{p-1}\boldsymbol{V}.$$

Combining it all together, we obtain

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V}\boldsymbol{\Delta}(\varepsilon)(\boldsymbol{W} + \underbrace{\widetilde{\boldsymbol{W}}_1(\varepsilon) + \widetilde{\boldsymbol{W}}_2(\varepsilon) + \widetilde{\boldsymbol{W}}_3(\varepsilon)}_{\mathcal{O}(\varepsilon)})\boldsymbol{\Delta}(\varepsilon)\boldsymbol{V}^{\top},$$

which completes the proof. ∎

Proof of Lemma 5.11. The proof repeats that of Lemma 5.10, but instead of (8.1) we use another expansion from [5, eqn. (32),(55)], which reads

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V}\boldsymbol{\Delta}(\varepsilon)\boldsymbol{W}\boldsymbol{\Delta}(\varepsilon)\boldsymbol{V}^{\top} + \varepsilon^p\boldsymbol{V}\boldsymbol{\Delta}(\varepsilon)\boldsymbol{W}_1(\varepsilon) + \varepsilon^p\boldsymbol{W}_2(\varepsilon)\boldsymbol{\Delta}(\varepsilon)\boldsymbol{V}^t + \varepsilon^{2p}\boldsymbol{W}_3(\varepsilon).$$

∎

Proof of Proposition 5.14. We use the expansion (8.1) for $p = r$, and an idea similar to the one in (8.2), but for $\boldsymbol{V} = [\boldsymbol{V}_{\le r-1}\ \boldsymbol{A}]$. Define the matrix $\widetilde{\boldsymbol{V}}$ as

$$\widetilde{\boldsymbol{V}} = \begin{bmatrix} \boldsymbol{V}_{\le r-1}^{\dagger} \\ \boldsymbol{A}^{\dagger} \end{bmatrix}.$$

Then, since $\boldsymbol{V}$ is full row rank and the matrices $\boldsymbol{V}_{\leq r-1}$ and $\boldsymbol{A}$ span orthogonal subspaces, we have

$$\boldsymbol{I}_n = \boldsymbol{V}\widetilde{\boldsymbol{V}} = \boldsymbol{V}\boldsymbol{\Delta}(\varepsilon)\boldsymbol{\Delta}(\varepsilon)^{-1}\widetilde{\boldsymbol{V}},$$

hence, we can rewrite

$$\varepsilon^r \boldsymbol{V}_{\leq r-1}\boldsymbol{\Delta}_{r-1}\boldsymbol{W}_1(\varepsilon) = \boldsymbol{V}_{\leq r-1}\boldsymbol{\Delta}_{r-1}\widetilde{\boldsymbol{W}}_1(\varepsilon)\boldsymbol{\Delta}(\varepsilon)\boldsymbol{V}^t,$$
$$\varepsilon^r \boldsymbol{W}_2(\varepsilon)\boldsymbol{\Delta}_{r-1}\boldsymbol{V}_{\leq r-1}^{\mathsf{T}} = \boldsymbol{V}\boldsymbol{\Delta}(\varepsilon)\widetilde{\boldsymbol{W}}_2(\varepsilon)\boldsymbol{\Delta}_{r-1}\boldsymbol{V}_{\leq r-1}^t,$$

where

$$\widetilde{\boldsymbol{W}}_1(\varepsilon) = \varepsilon^r \boldsymbol{W}_1(\varepsilon)(\widetilde{\boldsymbol{V}})^{\mathsf{T}}\boldsymbol{\Delta}(\varepsilon)^{-1} = o(1).\widetilde{\boldsymbol{W}}_2(\varepsilon) = \varepsilon^r \boldsymbol{\Delta}(\varepsilon)^{-1}\widetilde{\boldsymbol{V}}\boldsymbol{W}_2(\varepsilon) = o(1).$$

Similarly, for $\boldsymbol{W}_3(3)$, we have

$$\varepsilon^{2r-1}\boldsymbol{W}_3(3) = \boldsymbol{V}(\varepsilon^{r-\frac{1}{2}}\boldsymbol{I})\widetilde{\boldsymbol{V}}\boldsymbol{W}_3(\varepsilon)(\widetilde{\boldsymbol{V}})^t(\varepsilon^{r-\frac{1}{2}}\boldsymbol{I})\boldsymbol{V}^t.$$

Note that

$$\widetilde{\boldsymbol{V}}\boldsymbol{W}_3(\varepsilon)(\widetilde{\boldsymbol{V}})^t = \begin{bmatrix} \boldsymbol{V}_{\leq r-1}^{\dagger}\boldsymbol{D}^{(2r-1)}(\boldsymbol{V}_{\leq r-1}^{\dagger})^{\mathsf{T}} & \boldsymbol{V}_{\leq r-1}^{\dagger}\boldsymbol{D}^{(2r-1)}(\boldsymbol{A}^{\dagger})^{\mathsf{T}} \\ \boldsymbol{A}^{\dagger}\boldsymbol{D}^{(2r-1)}(\boldsymbol{V}_{\leq r-1}^{\dagger})^{\mathsf{T}} & \boldsymbol{A}^{\dagger}\boldsymbol{D}^{(2r-1)}(\boldsymbol{A}^{\dagger})^{\mathsf{T}} \end{bmatrix} + \mathcal{O}(\varepsilon),$$

and hence

$$(\varepsilon^{r-\frac{1}{2}}\boldsymbol{I})\widetilde{\boldsymbol{V}}\boldsymbol{W}_3(\varepsilon)(\widetilde{\boldsymbol{V}})^t(\varepsilon^{r-\frac{1}{2}}\boldsymbol{I}) = \boldsymbol{\Delta}(\varepsilon)\left(\begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{A}^{\dagger}\boldsymbol{D}^{(2r-1)}(\boldsymbol{A}^{\dagger})^{\mathsf{T}} \end{bmatrix} + o(1)\right)\boldsymbol{\Delta}(\varepsilon)$$

Combining it all together, we obtain

$$\boldsymbol{K}(\varepsilon) = \boldsymbol{V}\boldsymbol{\Delta}(\varepsilon)(\left(\begin{bmatrix} \boldsymbol{W}_{\leq r-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{A}^{\dagger}\boldsymbol{D}^{(2r-1)}(\boldsymbol{A}^{\dagger})^{\mathsf{T}} \end{bmatrix} + o(1)\right))\boldsymbol{\Delta}(\varepsilon)\boldsymbol{V}^{\top},$$

which completes the proof. ∎

References.

[1] Marianne Akian, Ravindra Bapat, and Stephane Gaubert. Min-plus methods in eigenvalue perturbation theory and generalised Lidskii-Vishik-Ljusternik theorem. *arXiv preprint math/0402090*, 2004.

[2] Marianne Akian, Stephane Gaubert, and Andrea Marchesini. Tropical bounds for eigenvalues of matrices. *Linear Algebra and its Applications*, 446:281–303, 2014.

[3] Marianne Akian, Ravindra Bapat, and Stéphane Gaubert. Non-archimedean valuations of eigenvalues of matrix polynomials. *Linear Algebra and its Applications*, 498:592–627, 2016.

[4] Jason M. Altschuler and Pablo A. Parrilo. Kernel approximation on algebraic varieties. *SIAM Journal on Applied Algebra and Geometry*, 7(1):1–28, 2023. doi: 10.1137/21M1425050.

[5] Simon Barthelmé and Konstantin Usevich. Spectral properties of kernel matrices in the flat limit. *SIAM Journal on Matrix Analysis and Applications*, 42(1):17–57, 2021.

[6] Simon Barthelmé, Nicolas Tremblay, Konstantin Usevich, and Pierre-Olivier Amblard. Determinantal point processes in the flat limit. *arXiv preprint arXiv:2107.07213*, 2021.

[7] Simon Barthelmé, Pierre-Olivier Amblard, Nicolas Tremblay, and Konstantin Usevich. Gaussian process regression in the flat limit. *arXiv preprint arXiv:2201.01074*, 2022.

[8] Marcus Carlsson. Perturbation theory for the spectral decomposition of hermitian matrices. *arXiv preprint arXiv:1809.09480*, 2018.

[9] Richard Courant and David Hilbert. *Methods of mathematical physics.* John Wiley & Sons, 1995.

[10] David Cox, John Little, and Donald O'Shea. *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.* Springer, 2nd edition, 1997.

[11] Nuha Diab and Dmitry Batenkov. Spectral properties of infinitely smooth kernel matrices in the single cluster limit, with applications to multivariate super-resolution, 2024. URL https://arxiv.org/abs/2407.10600.

[12] Tobin A Driscoll and Bengt Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Computers & Mathematics with Applications*, 43(3-5):413–422, 2002.

[13] Gregory E Fasshauer. *Meshfree approximation methods with MATLAB*, volume 6. World Scientific, 2007.

[14] Bengt Fornberg, Elisabeth Larsson, and Natasha Flyer. Stable computations with Gaussian radial basis functions. *SIAM Journal on Scientific Computing*, 33(2):869–892, 2011.

[15] Anne Greenbaum, Ren-cang Li, and Michael L Overton. First-order perturbation theory for eigenvalues and eigenvectors. *SIAM review*, 62(2):463–482, 2020.
[16] James Hook. Max-plus singular values. *Linear Algebra and its Applications*, 486:419–442, 2015.
[17] James Hook, Jennifer Pestana, Francoise Tisseur, and Jonathan Hogg. Max-balanced Hungarian scalings. *SIAM Journal on Matrix Analysis and Applications*, 40(1):320–346, 2019.
[18] Michael Joswig. *Essentials of tropical combinatorics*, volume 219. American Mathematical Society, 2021.
[19] T. Kato. *Perturbation theory for linear operators*. Springer-Verlag, 2nd corrected edition, 1995.
[20] Viktor Borisovich Lidskii. Perturbation theory of non-conjugate operators. *USSR Computational Mathematics and Mathematical Physics*, 6(1):73–85, 1966.
[21] Julio Moro, James V Burke, and Michael L Overton. On the Lidskii–Vishik–Lyusternik perturbation theory for eigenvalues of matrices with arbitrary jordan structure. *SIAM Journal on Matrix Analysis and Applications*, 18(4):793–817, 1997.
[22] Franz Rellich and Joan Berkowitz. *Perturbation theory of eigenvalue problems*. CRC Press, 1969.
[23] Robert Schaback and Holger Wendland. Kernel techniques: from machine learning to meshless methods. *Acta Numerica*, 15:543–639, 2006.
[24] Michael L Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.
[25] Andrew J. Wathen and Shengxin Zhu. On spectral distribution of kernel matrices related to radial basis functions. *Numerical Algorithms*, 70(4):709–726, Dec 2015. ISSN 1572-9265. doi: 10.1007/s11075-015-9970-0.
[26] Holger Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2004. doi: 10.1017/CBO9780511617539.