SIAM J. MATRIX ANAL. APPL.
Vol. 42, No. 1, pp. 17–57

SPECTRAL PROPERTIES OF KERNEL MATRICES IN THE FLAT LIMIT*

SIMON BARTHELMÉ† AND KONSTANTIN USEVICH‡

Abstract. Kernel matrices are of central importance to many applied fields. In this manuscript, we focus on spectral properties of kernel matrices in the so-called "flat limit," which occurs when points are close together relative to the scale of the kernel. We establish asymptotic expressions for the determinants of the kernel matrices, which we then leverage to obtain asymptotic expressions for the main terms of the eigenvalues. Analyticity of the eigenprojectors yields expressions for limiting eigenvectors, which are strongly tied to discrete orthogonal polynomials. Both smooth and finitely smooth kernels are covered, with stronger results available in the finite smoothness case.

Key words. kernel matrices, eigenvalues, eigenvectors, radial basis functions, perturbation theory, flat limit, discrete orthogonal polynomials

AMS subject classifications. 15A18, 47A55, 47A75, 47B34, 60G15, 65D05

DOI. 10.1137/19M129677X

1. Introduction.

For an ordered set of points $\mathcal{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$, $\boldsymbol{x}_k \in \Omega \subset \mathbb{R}^d$, not lying in general on a regular grid, and a kernel $K : \Omega \times \Omega \to \mathbb{R}$, the kernel matrix $\boldsymbol{K}$ is defined as

$$\boldsymbol{K} = \boldsymbol{K}(\mathcal{X}) = [K(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1}^{n,n}\,.$$

These matrices occur in approximation theory (kernel-based approximation and interpolation, [33, 43]), statistics and machine learning (Gaussian process models [44], support vector machines, and kernel PCA [34]). Often, a positive scaling parameter is introduced,[1] and the scaled kernel matrix becomes

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

where typically $K_\varepsilon(\boldsymbol{x}, \boldsymbol{y}) = K(\varepsilon\boldsymbol{x}, \varepsilon\boldsymbol{y})$. If the kernel is radial (the most common case), then its value depends only on the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, and ε determines how quickly the kernel decays with distance.

Understanding spectral properties of kernel matrices is essential in statistical applications (e.g., for selecting hyperparameters), as well as in scientific computing (e.g., for preconditioning [13, 41]). Because the spectral properties of kernel matrices are not directly tractable in the general case, one usually needs to resort to asymptotic results. The most common form of asymptotic analysis takes $n \to \infty$. Three cases are typically considered: (a) when the distribution of points in $\mathcal{X}$ converges to some

*Received by the editors October 31, 2019; accepted for publication (in revised form) September 11, 2020; published electronically January 13, 2021.
https://doi.org/10.1137/19M129677X
Funding: The work of the author was supported by the ANR project GenGP, grant ANR-16-CE23-0008 and project LeaFleT, grant ANR-19-CE23-0021-01.
†CNRS, Université Grenoble Alpes, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (simon.barthelme@gipsa-lab.fr).
‡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@univ-lorraine.fr).
[1] In this paper, for simplicity, we consider only the case of isotropic scaling (i.e., all the variables are scaled with the same parameter ε). The results should hold for the constant anisotropic case, by rescaling the set of points $\mathcal{X}$ in advance.

continuous measure on Ω, the kernel matrix tends in some sense to a linear operator in a Hilbert space, whose spectrum is then studied [39]; (b) recently, some authors have obtained asymptotic results in a regime where both n and the dimension d tend to infinity [11, 7, 40], using the tools of random matrix theory; (c) in a special case of $\mathcal{X}$ lying on a regular grid, stationary kernel matrices become Toeplitz or more generally multilevel Toeplitz, whose asymptotic spectral distribution is determined by their symbol (or the Fourier transform of the sequence) [18, 2, 38, 29, 3].

Driscoll and Fornberg [10] pioneered a new form of asymptotic analysis for kernel methods, in the context of radial basis function (RBF) interpolation. The point set $\mathcal{X}$ is considered fixed, with arbitrary geometry (i.e., not lying in general on a regular grid), and the scaling parameter ε approaches 0. Driscoll and Fornberg called this the "flat limit," as kernel functions become flat over the range of $\mathcal{X}$ as $\varepsilon \to 0$. Very surprisingly, they showed that for certain kernels the RBF interpolant stays well-defined in the flat limit, and tends to the Lagrange polynomial interpolant. Later, a series of papers extended their results to the multivariate case [31, 24, 32] and established similar convergence results for various types of radial functions (see [35, 27] and references therein). In particular, [35] showed that for kernels of finite smoothness the limiting interpolant is a spline rather than a polynomial.

The flat limit is interesting for several reasons. In contrast to other asymptotic analyses, it is deterministic ($\mathcal{X}$ is fixed), and makes very few assumptions on the geometry of the point set. In addition, kernel methods are plagued by the problem of picking a scale parameter [34]. One either uses burdensome procedures like cross validation or maximum likelihood [44] or suboptimal but cheap heuristics like the median distance heuristic [16]. The flat limit analysis may shed some light on the problem. Finally, the results derived here can be thought of as perturbation results, in the sense that they are formally exact in the limit, but useful approximations when the scale is not too small.

Despite its importance, little was known until recently about the eigenstructure of kernel matrices in the flat limit. The difficulty comes from the fact that $\boldsymbol{K}_\varepsilon = K(0,0)\mathbf{1}\mathbf{1}^\mathsf{T} + \mathcal{O}(\varepsilon)$, where $\mathbf{1} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^\mathsf{T}$, i.e., we are dealing with a singular perturbation problem.[2]

Schaback [31, Theorem 6], and, more explicitly, Wathen and Zhu [42] obtained results on the orders of eigenvalues of kernel matrices for smooth analytic radial basis kernels, based on the Courant–Fischer minimax principle. A heuristic analysis of the behavior of the eigenvalues in the flat limit was also performed in [15]. However, the main terms in the expansion of the eigenvalues have never been obtained, and the results in [31, 42] apply only to smooth kernels. In addition, they hold no direct information on the limiting eigenvectors.

In this paper, we try filling this gap by characterizing both the eigenvalues and eigenvectors of kernel matrices in the flat limit. We consider both completely smooth kernels and finitely smooth kernels. The latter (Matérn-type kernels) are very popular in spatial statistics. For establishing asymptotic properties of eigenvalues, we use the expression for the limiting determinants of $\boldsymbol{K}_\varepsilon$ (obtained only for the smooth case), and Binet-Cauchy formulas. As a special case, we recover the results of Schaback [31] and Wathen and Zhu [42], but for a wider class of kernels.

1.1. Overview of the results.

Some of the results are quite technical, so the goal of this section is to serve as a reader friendly summary of the contents.

[2]Seen from the point of view of the characteristic polynomial, the equation $\det(\boldsymbol{K}_\varepsilon - \lambda \boldsymbol{I}) = 0$ has a solution of multiplicity $n-1$ at $\varepsilon = 0$, but these roots immediately separate when $\varepsilon > 0$.

1.1.1. Types of kernels. We begin with some definitions.

A kernel is called *translation invariant* if

$$K(\boldsymbol{x}, \boldsymbol{y}) = \varphi(\boldsymbol{x} - \boldsymbol{y}) \tag{2}$$

for some function φ. A kernel is *radial*, or *stationary* if, in addition, we have

$$K(\boldsymbol{x}, \boldsymbol{y}) = f(\|\boldsymbol{x} - \boldsymbol{y}\|_2), \tag{3}$$

i.e., $\varphi(t) = f(|t|)$ and the value of the kernel depends only on the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. The function f is an *RBF*. Finally, $K(\boldsymbol{x}, \boldsymbol{y})$ may be positive (semi)definite, in which case the kernel matrix is positive (semi)definite [9] for all sets of distinct points $\mathcal{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and all $n > 0$.

All of our results are valid for stationary, positive semidefinite kernels. In addition, some are also valid for translation-invariant kernels or even general, nonradial kernels. For simplicity, we focus on radial kernels in this introductory section.

An important property of a radial kernel is its order of smoothness, which we call r throughout this paper. The definition is at first glance not very enlightening: formally, if the first p odd-order derivatives of the RBF g are zero, and the $(p+1)$th is nonzero, then $r = p + 1$. To understand the definition, some Fourier analysis is required [37], but for the purposes of this article we will just note two consequences. When interpolating using a radial kernel of smoothness r, the resulting interpolant is $r - 1$ times differentiable. When sampling from a Gaussian process with covariance function of smoothness order r, the sampled process is also $r - 1$ times differentiable (almost surely). r may equal ∞, which is the case we call *infinitely smooth*. If r is finite we talk about a *finitely smooth* kernel. We treat the two cases separately due to the importance of infinitely smooth kernels, and because proofs are simpler in that case.

Finally, the points are assumed to lie in some subset of $\mathbb{R}^d$, and if $d = 1$ we call this the *univariate* case, as opposed to the *multivariate* case ($d > 1$).

1.1.2. Univariate results. In the univariate case, we can give simple closed-form expressions for the eigenvalues and eigenvectors of kernel matrices as $\varepsilon \to 0$. What form these expressions take depends essentially on the order of smoothness.

We shall contrast two kernels that are at opposite ends of the smoothness spectrum. One, the Gaussian kernel, is infinitely smooth, and is defined as

$$K_\varepsilon(x, y) = \exp\left(-\varepsilon(x - y)^2\right).$$

The other has smoothness order 1, and is known as the "exponential" kernel (and is also a Matérn kernel):

$$K_\varepsilon(x, y) = \exp\left(-\varepsilon|x - y|\right).$$

Both kernels are radial and positive definite. However, the small-ε asymptotics of these two kernels are strikingly different.

In the case of the Gaussian kernel, the eigenvalues go to 0 extremely fast, except for the first one, which goes to n. Specifically, the first eigenvalue is $\mathcal{O}(1)$, the second is $\mathcal{O}(\varepsilon^2)$, the third is $\mathcal{O}(\varepsilon^4)$, etc.[3] Figure 1 shows the eigenvalues of the Gaussian kernel for a fixed set $\mathcal{X}$ of randomly chosen nodes in the unit interval ($n = 10$ here). The eigenvalues are shown as a function of ε, under log-log scaling. As expected from Theorem 4.2 (see also [31, 42]), for each i, $\log \lambda_i$ is approximately linear as a function

[3] In fact, Theorem 4.2 guarantees the same behavior for general kernels of sufficient smoothness.

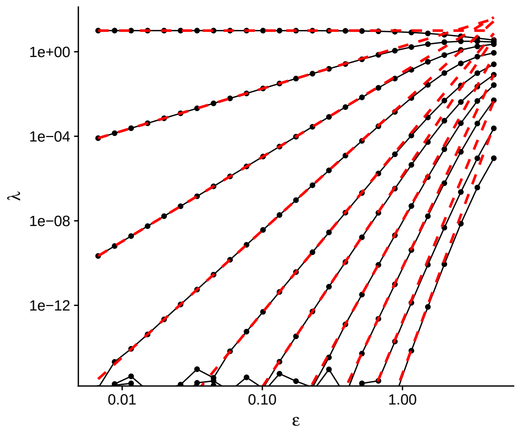

FIG. 1. *Eigenvalues of the Gaussian kernel ($d = 1$). The set of $n = 10$ nodes was drawn uniformly from the unit interval. In black, eigenvalues of the Gaussian kernel, for different values of ε. The dashed red curves are our small-ε expansions. Note that both axes are scaled logarithmically. The noise apparent for small ε values in the low range is due to loss of precision in the numerical computations.*

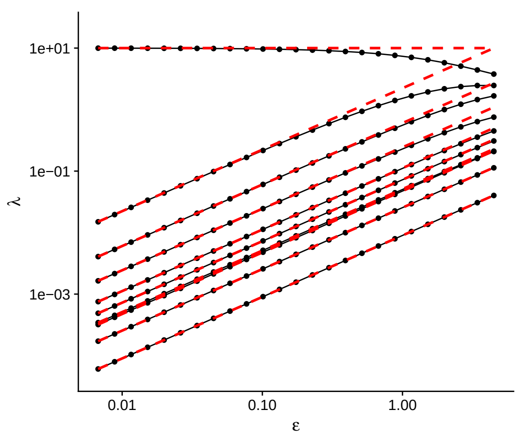

FIG. 2. *Eigenvalues of the exponential kernel ($d = 1$), using the same set of points as in Figure 1. The largest eigenvalue has a slope of 0 for small ε, the others have unit slope, as in Theorem 4.5.*

of $\log \varepsilon$. In addition, the main term in the scaling of $\log \lambda_i$ in ε (i.e., the offsets of the various lines) is also given by Theorem 4.2, and the corresponding asymptotic approximations are plotted in red. They show very good agreement with the exact eigenvalues, up to $\varepsilon \approx 1$.

Contrast that behavior with the one exhibited by the eigenvalues of the exponential kernel. Theorem 4.6 describes the expected behavior: the top eigenvalue is again $\mathcal{O}(1)$ and goes to n, while all remaining eigenvalues are $\mathcal{O}(\varepsilon)$. Figure 2 is the counterpart of the previous figure, and shows clearly that all eigenvalues except for the top one go to 0 at unit rate. The main term in the expansions of eigenvalues determines again the offsets shown in Figure 2, which can be computed from the eigenvalues of the centered distance matrix as shown in Theorem 4.6.

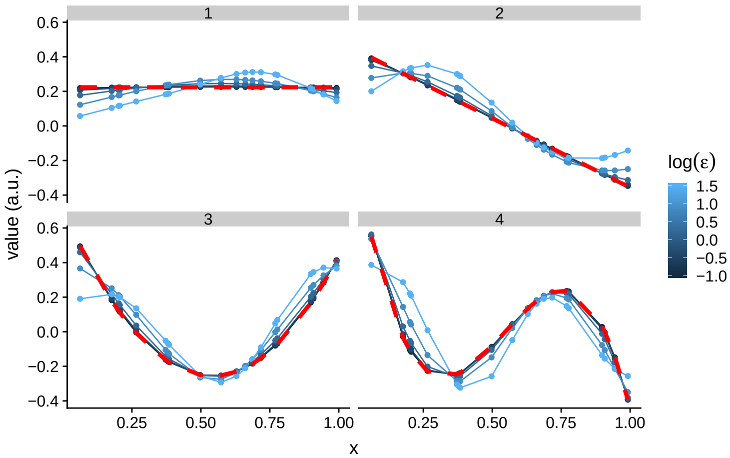

FIG. 3. *First four eigenvectors of the Gaussian kernel. We used the same set of points as in the previous two figures. In blue, eigenvectors of the Gaussian kernel, for different values of ε. The dashed red curves show the theoretical limit as $\varepsilon \to 0$ (i.e., the first four orthogonal polynomials of the discrete measure).*

To sum up: except for the top eigenvalue, which behaves in the same way for both kernels, the rest scale quite differently. More generally, Theorem 4.6 states that for kernels of smoothness order $r < n$ ($r = 1$ for the exponential, $r = \infty$ for the Gaussian), the eigenvalues are divided into two groups. The first group of eigenvalues is of size r, and have orders 1, ε^2, ε^4, etc. The second group is of size $n - r$, and all have the same order, ε^{2r-1}.

The difference between the two kernels is also reflected in how the eigenvectors behave. For the Gaussian kernel, the limiting eigenvectors (shown in Figure 3) are columns of the $\boldsymbol{Q}$ matrix of the QR factorization of the Vandermonde matrix (i.e., the orthogonal polynomials with respect to the discrete uniform measure on $\mathcal{X}$). For instance, the top eigenvector equals the constant vector $\frac{1}{\sqrt{n}}\mathbf{1}$, and the second eigenvector equals $\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\mathsf{T} - \frac{\sum x_i}{n}$ (up to normalization). Each successive eigenvector depends on the geometry of $\mathcal{X}$ via the moments $m_p(\mathcal{X}) = \sum_{i=1}^n x_i^p$. In fact, this result is valid for any positive definite smooth analytic in ε kernel as shown by Corollary 4.3.

In the case of finite smoothness, the two groups are associated with different groups of eigenvectors. The first group of r eigenvectors is again orthogonal polynomials. The second group is splines of order $2r - 1$. Convergence of eigenvectors is shown in Figure 4. This general result for eigenvectors is shown in Theorem 8.1.

1.1.3. The multivariate case. The general, multivariate case requires more care. Polynomials continue to play a central role in the flat limit, but when $d > 1$ they appear naturally in groups of equal degree. For instance, in $d = 2$, we may write $\boldsymbol{x} = \begin{bmatrix} y & z \end{bmatrix}^\mathsf{T}$ and the first few monomials are as follows:

- Degree 0: 1;
- Degree 1: y, z;
- Degree 2: y^2, yz, z^2;
- Degree 3: y^3, y^2z, yz^2, z^3,

etc. Note that there is one monomial of degree 0, two degree-1 monomials, three degree-2 monomials, and so on. If $d = 1$ there is a single monomial in each group,

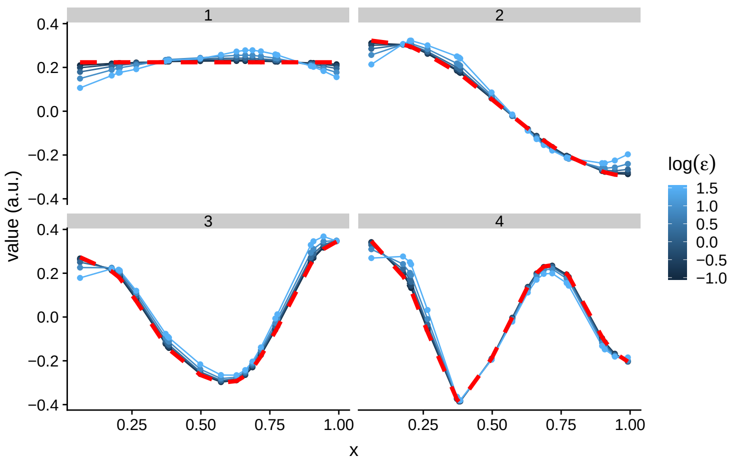

FIG. 4. *First four eigenvectors of the exponential kernel. The same set of nodes is used as in Figure 3. In blue, the eigenvectors of the kernel, for different values of ε. The dashed red curves show the theoretical limit as $\varepsilon \to 0$. From Theorem 8.1, these are (1) the vector $\frac{\mathbf{1}}{\sqrt{n}}$ and (2)–(4) the first three eigenvectors of $(\boldsymbol{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n})\boldsymbol{D}_{(1)}(\boldsymbol{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n})$, where $\boldsymbol{D}_{(1)}$ is the Euclidean distance matrix.*

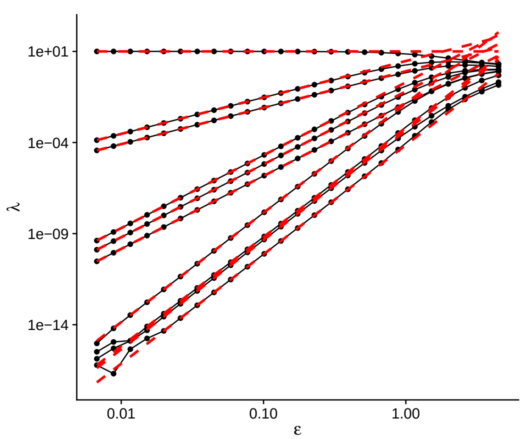

FIG. 5. *Eigenvalues of the Gaussian kernel in the multivariate case ($d = 2$). The set of $n = 10$ nodes was drawn uniformly from the unit square. In black, eigenvalues of the Gaussian kernel, for different values of ε. The dashed red curves are our small-ε expansions. Eigenvalues appear in groups of different orders, recognizable here as having the same slopes as $\varepsilon \to 0$. A single eigenvalue is of order $\mathcal{O}(1)$. Two eigenvalues are of order $\mathcal{O}(\varepsilon^2)$, as many as there are monomials of degree 1 in two dimensions. Three eigenvalues are of order $\mathcal{O}(\varepsilon^4)$, as many as there are monomials of degree 2 in two dimensions, etc.*

and here lies the essence of the difference between the univariate and multivariate cases.

In infinitely smooth kernels like the Gaussian kernel, as shown in [31, 42], there are as many eigenvalues of order $\mathcal{O}(\varepsilon^{2k})$ as there are monomials of degree k in dimension d: for instance, there are 4 monomials of degree 3 in dimension 2, and so 4 eigenvalues of order $\mathcal{O}(\varepsilon^8)$. An example is shown in Figure 5. In finitely smooth kernels like the exponential kernel, there are r successive groups of eigenvalues with

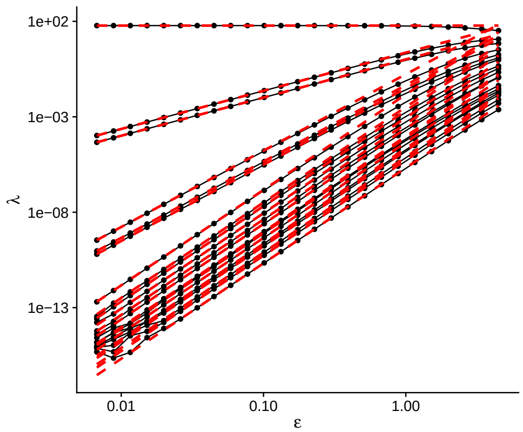

FIG. 6. *Eigenvalues of a finitely smooth kernel in $d = 2$. Here we picked a kernel with smoothness index $r = 3$ (the exponential kernel has order $r = 1$). The set of $n = 20$ nodes was drawn uniformly from the unit square. The first $r = 3$ groups of eigenvalues have the same behavior as in the Gaussian kernel: one is $\mathcal{O}(1)$, two are $\mathcal{O}(\varepsilon^2)$, three are $\mathcal{O}(\varepsilon^4)$. All the rest are $\mathcal{O}(\varepsilon^5)$.*

order $\mathcal{O}(1)$, $\mathcal{O}(\varepsilon^2), \ldots$, up to $\mathcal{O}(\varepsilon^{2r-2})$. Following that, all remaining eigenvalues have order $\mathcal{O}(\varepsilon^{2r-1})$, just like in the one-dimensional case.

We show in Theorem 6.2, that the main terms of these eigenvalues in each group can be computed from the QR factorization of the Vandermonde matrix and a Schur complement associated with the Wronskian matrix of the kernel. In the finite smoothness case (see Figure 6), the same expansions are valid until the smoothness order, and the last group of eigenvalues is given by the eigenvalues of the projected distance matrix, as in the one-dimensional case.

Finally, in the multivariate case, Theorem 6.2 characterizes the eigenprojectors. In a nutshell, the invariant subspaces associated with each group of eigenvalues are spanned by orthogonal polynomials of a certain order. The eigenvectors are the subject of a conjecture given in section 8, which we believe to be quite solid.

1.2. Overview of tools used in the paper. For finding the orders of the eigenvalues, as in [31, 42], we use the Courant–Fischer minimax principle (see Theorem 3.7). However, unlike [31, 42], we do not use directly the results of Micchelli [28], but rather rotate the kernel matrices using the Q factor in the QR factorization of the Vandermonde matrix, and find the expansion of rotated matrices from the Taylor expansion of the kernel.

The key results are the expansions for the determinants of $\boldsymbol{K}_\varepsilon$, which use the expansions of rotated kernel matrices. Our results on determinants (Theorems 4.1, 4.5, 6.1, and 6.3) generalize those of Lee and Micchelli [26]. The next key observation is that principal submatrices of $\boldsymbol{K}_\varepsilon$ are also kernel matrices, hence the results on determinants imply the results on expansions of elementary symmetric polynomials of eigenvalues (via the correspondence between elementary symmetric polynomials (see Theorem 3.1), and the Binet–Cauchy formula). Finally, the main terms of the eigenvalues can be retrieved from the main terms of the elementary symmetric polynomials, as shown in Lemma 3.3. An important tool for the multivariate and finite smoothness case is Lemma 3.12 on low-rank perturbation of elementary symmetric polynomials that we could not find elsewhere in the literature

To study the properties of the eigenvectors, we use analyticity of the eigenvalues and eigenprojectors for Hermitian analytic matrix-valued functions [22]. By using an extension of the Courant–Fischer principle (see Lemma 3.10), we can fully characterize the limiting eigenvectors in the univariate case, and obtain the limiting invariant subspaces for the groups of the limiting eigenvectors in the multivariate case. Moreover, by using the perturbation expansions from [22], we can find the last individual eigenvectors in the finitely smooth case.

We note that the multivariate case requires a number of technical assumptions on the arrangement of points, which are typical for multivariate polynomial interpolation. For the results on determinants, no assumptions are required. However, for getting the results on the eigenvalues or eigenvectors at a certain order of ε, we need a relaxed version to the well-known unisolvency condition [31, 14], namely, the multivariate Vandermonde matrix up to the corresponding degree to be full rank.

1.3. Organization of the paper. In an attempt to make the paper reader friendly, it is organized as follows. In section 2 we recall the main terminology for one-dimensional (1D) kernels. In section 3 we gather well-known (or not so well-known) results on eigenvalues, determinants, elementary symmetric polynomials, and their perturbations, which are key tools used in the paper. Section 4 contains the main results on determinants, eigenvalues, and eigenvectors in the univariate ($d = 1$) case. While these results are special cases of the multivariate case ($d > 1$), the latter is burdened with heavier notation due to the complexity of dealing with multivariate polynomials. To get a gist of the results and techniques, the reader is advised to first consult the case $d = 1$. In section 5, we introduce all the needed notation to handle the multivariate case. Section 6 contains the main results of the paper on determinants, eigenvalues, and groups of eigenvectors in the multivariate case. Thus sections 2 to 6 are quite self-contained and contain most of the results in the paper, except the result on precise locations of the last group of eigenvectors in the finite smoothness case. In section 7, we provide a brief summary on analytic perturbation theory, needed only for proving the stronger result on eigenvectors for finitely smooth kernels in section 8.

2. Background and main notation. This section contains main definitions and examples of kernels, mostly in the 1D case (with multivariate extensions given in section 5). We assume that the kernel $K : \Omega \times \Omega \to \mathbb{R}$ is in the class $\mathcal{C}^{(\ell,\ell)}(\Omega)$, $\Omega = (-a; a)$, i.e., all the partial derivatives $\frac{\partial^2}{\partial x^i \partial y^j} K$ exist and are continuous for $0 \le i, j \le \ell$ on $\Omega \times \Omega$.

We will often need the following short-hand notation for partial derivatives:

$$K^{(i,j)} \stackrel{\text{def}}{=} \frac{\partial^{i+j}}{\partial x^i \partial y^j} K,$$

and we introduce the so-called Wronskian matrix

$$\boldsymbol{W}_{\le k} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{K^{(0,0)}(0,0)}{0!0!} & \cdots & \frac{K^{(0,k)}(0,0)}{0!k!} \\ \vdots & & \vdots \\ \frac{K^{(k,0)}(0,0)}{k!0!} & \cdots & \frac{K^{(k,k)}(0,0)}{k!k!} \end{bmatrix}. \tag{4}$$

2.1. Translation invariant and radial kernels. Let us consider an important example of a translation invariant (2) kernel, which in the univariate case becomes $K(x, y) = \varphi(x - y)$. We assume that $\varphi \in \mathcal{C}^{2r}(-2a, 2a)$; hence, φ has a Taylor expansion around 0:

$$\varphi(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_{2r} t^{2r} + o(t^{2r}), \quad \text{where } \alpha_k \stackrel{\text{def}}{=} \frac{\varphi^{(k)}(0)}{k!}.$$

Therefore $K \in \mathcal{C}^{(r,r)}(\Omega \times \Omega)$ and its derivatives at 0 are

$$\frac{K^{(i,j)}(0,0)}{i!j!} = \frac{(-1)^j \varphi^{(i+j)}(0)}{i!j!} = (-1)^j \binom{i+j}{j} \alpha_{i+j}, \tag{5}$$

i.e., the Wronskian matrix has the form:

$$\boldsymbol{W}_{\leq k} = \begin{bmatrix} \alpha_0 & -\alpha_1 & \alpha_2 & -\alpha_3 & \cdots \\ \alpha_1 & -2\alpha_2 & 3\alpha_3 & -4\alpha_4 & \cdots \\ \alpha_2 & -3\alpha_3 & 6\alpha_4 & -10\alpha_5 & \cdots \\ \alpha_3 & -4\alpha_4 & 10\alpha_5 & -20\alpha_6 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

A special case of the translational kernels are smooth and radial, which will be considered in the next subsection. The simplest example is the Gaussian kernel with

$$\varphi(t) = \exp(-t^2) = 1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} + \cdots,$$

i.e., for all integers $\nu \geq 0$

$$\alpha_{2\nu} = \frac{(-1)^\nu}{\nu!}, \quad \alpha_{2\nu+1} = 0.$$

In this case, the Wronskian matrix becomes

$$\boldsymbol{W}_{\leq k} = \begin{bmatrix} 1 & 0 & -1 & 0 & \cdots \\ 0 & 2 & 0 & -2 & \cdots \\ -1 & 0 & 3 & 0 & \cdots \\ 0 & -2 & 0 & -\frac{10}{3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

An important subclass consists of radial kernels (3), i.e., $K(x,y) = f(|x-y|)$ in the univariate case. We put the following assumptions on f:

- $f \in \mathcal{C}^{2r}(-2a, 2a)$;
- the highest derivative of odd order is not zero, i.e., $f^{(2r-1)}(0) \neq 0$;
- the lower derivatives with odd order vanish, i.e., $f^{(2\ell-1)}(0) = 0$ for $\ell < r$.

If the function $f(t)$ satisfies these assumptions, then r is called the *order of smoothness* of K. Note that f admits a Taylor expansion

$$f(t) = f_0 + f_2 t^2 + \cdots + f_{2r-2} t^{2r-2} + t^{2r-1}(f_{2r-1} + \mathcal{O}(t)), \tag{6}$$

where $f_k = f^{(k)}(0)/k!$ is a shorthand notation for the scaled derivative at 0. For example, for the exponential kernel $\exp(-|x-y|)$, we have $f_0 = 1$ and $f_1 = -1$, so the smoothness order is $r = 1$. For the $\mathcal{C}^2$ Matérn kernel $(1 + |x-y|)\exp(-|x-y|)$, we have $f_0 = 1, f_1 = 0, f_2 = -1/2, f_3 = 1/3$, so the smoothness order is $r = 2$.

2.2. Distance matrices, Vandermonde matrices, and their properties. In the general multivariate case, we define $\boldsymbol{D}_{(k)}$ to be the kth Hadamard power of the Euclidean distance matrix $\boldsymbol{D}_{(1)}$, i.e.,

$$\boldsymbol{D}_{(k)} = \left[\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^k\right]_{i,j=1}^{n,n}.$$

Next, we focus on the univariate case (the multivariate case will be discussed later in section 5) and denote by $\boldsymbol{V}_{\le k}$ the univariate Vandermonde matrix up to degree k:

$$\boldsymbol{V}_{\le k} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix}, \tag{7}$$

which has rank $\min(n, k+1)$ if the nodes are distinct. In particular, the matrix $\boldsymbol{V}_{\le n-1}$ is square and invertible for distinct nodes.

For even k, the Hadamard powers $\boldsymbol{D}_{(k)}$ of the distance matrix can be expressed via the columns of the Vandermonde matrix using the binomial expansion

$$\boldsymbol{D}_{(2\ell)} = \sum_{j=0}^{2\ell} (-1)^j \binom{2\ell}{j} \boldsymbol{v}_j \boldsymbol{v}_{2\ell-j}^{\mathsf{T}}, \tag{8}$$

where $\boldsymbol{v}_k \stackrel{\text{def}}{=} \begin{bmatrix} x_1^k & \cdots & x_n^k \end{bmatrix}^{\mathsf{T}}$ are columns of Vandermonde matrices. Therefore, $\operatorname{rank} \boldsymbol{D}_{(2\ell)} = \min(2\ell+1, n)$ if all points x_k are distinct.

On the other hand, for k odd, the matrices $\boldsymbol{D}_{(k)}$ exhibit an entirely different set of properties. Of most interest here is conditional positive-definitess, which guarantees that the distance matrices are positive definite when projected onto a certain subspace. The following result appears, e.g., in [12, Chapter 8], but follows directly from an earlier paper by Micchelli [28].

LEMMA 2.1. *For a distinct node set $\mathcal{X} \subset \mathbb{R}$ of size $n > r \ge 1$, we let $\boldsymbol{B}$ be a full column rank matrix such that $\boldsymbol{B}^{\mathsf{T}} \boldsymbol{V}_{\le r-1} = 0$. Then the matrix $(-1)^r \boldsymbol{B}^{\mathsf{T}} \boldsymbol{D}_{(2r-1)} \boldsymbol{B}$ is positive definite.*

For instance, if $r = 1$, we may pick any basis $\boldsymbol{B}$ orthogonal to the vector $\mathbf{1}$, and the lemma implies that $(-\boldsymbol{B}^{\mathsf{T}} \boldsymbol{D}_{(1)} \boldsymbol{B})$ has $n-1$ positive eigenvalues. We note for future reference that the result generalizes almost unchanged to the multivariate case.

2.3. Scaling and expansions of kernel matrices. In this subsection, we consider the general multivariate case. Given a general kernel K, we define its scaled version as

$$K_\varepsilon(\boldsymbol{x}, \boldsymbol{y}) = K(\varepsilon \boldsymbol{x}, \varepsilon \boldsymbol{y}), \tag{9}$$

while for the specific case of radial kernels we use the following form:

$$K_\varepsilon(\boldsymbol{x}, \boldsymbol{y}) = f(\varepsilon \|\boldsymbol{x} - \boldsymbol{y}\|_2). \tag{10}$$

Note that the definitions (9) and (10) coincide for nonnegative ε, but differ if we formally take other values (e.g., complex) of ε. Depending on context, we will use one or the other of the definitions later on, especially when we talk about analyticity of kernel matrices.

Using (6) and (10), we may write the scaled kernel matrix (1) as

$$\boldsymbol{K}_\varepsilon = f_0\boldsymbol{D}_{(0)}+\varepsilon^2 f_2\boldsymbol{D}_{(2)}+\cdots+\varepsilon^{2r-2}f_{2r-2}\boldsymbol{D}_{(2r-2)}+\varepsilon^{2r-1}f_{2r-1}\boldsymbol{D}_{(2r-1)}+\mathcal{O}(\varepsilon^{2r}). \tag{11}$$

In the univariate case, (8) gives a way to rewrite the expansion (11) as

$$\boldsymbol{K}_\varepsilon = \sum_{\ell=0}^{r-1}\varepsilon^{2j}\underbrace{\boldsymbol{V}_{\leq 2\ell}\boldsymbol{W}_{\diagup 2\ell}\boldsymbol{V}_{\leq 2\ell}^{\mathsf{T}}}_{f_{2\ell}\boldsymbol{D}_{(2\ell)}}+\varepsilon^{2r-1}f_{2r-1}\boldsymbol{D}_{(2r-1)}+\mathcal{O}(\varepsilon^{2r}), \tag{12}$$

where $\boldsymbol{W}_{\diagup s}\in\mathbb{R}^{(s+1)\times(s+1)}$ has nonzero elements only on its antidiagonal:

$$\Big(\boldsymbol{W}_{\diagup s}\Big)_{i+1,s-i+1} \stackrel{\text{def}}{=} f_s(-1)^i\binom{s}{i}. \tag{13}$$

For example,

$$\boldsymbol{W}_{\diagup 2} = f_2\begin{bmatrix}0 & 0 & 1\\ 0 & -2 & 0\\ 1 & 0 & 0\end{bmatrix}.$$

In fact, from (5), nonzero elements of $\boldsymbol{W}_{\diagup s}$ are scaled derivatives of the kernel:

$$\Big(\boldsymbol{W}_{\diagup s}\Big)_{i+1,s-i+1} = \frac{K^{(i,s-i)}(0,0)}{i!(s-i)!};$$

this justifies the notation $\boldsymbol{W}_{\diagup s}$ (i.e., an antidiagonal of the Wronskian matrix).

3. Determinants and elementary symmetric polynomials.

In this paper, we will heavily use the elementary symmetric polynomials of eigenvalues, and we collect in this section some useful (more or less known) facts about them.

3.1. Eigenvalues, principal minors, and elementary symmetric polynomials.

The kth elementary symmetric polynomial of $\lambda_1,\ldots,\lambda_n$ is defined as

$$e_k(\lambda_1,\ldots,\lambda_n) = \sum_{\substack{\#\mathcal{Y}=k\\ \mathcal{Y}\subset\{1,\ldots,n\}}}\prod_{i\in\mathcal{Y}}\lambda_i, \tag{14}$$

i.e., the sum is running over all possible subsets of $\{1,\ldots,n\}$ of size k. In particular, $e_1(\lambda_1,\ldots,\lambda_n)=\sum_{i=1}^n\lambda_i$, $e_2(\lambda_1,\ldots,\lambda_n)=\sum_{i<j}\lambda_i\lambda_j$, and $e_n(\lambda_1,\ldots,\lambda_n)=\prod_{i=1}^n\lambda_i$.

Next, we consider the elementary symmetric polynomials applied to eigenvalues of matrices, and define (with some abuse of notation)

$$e_k(\boldsymbol{A}) \stackrel{\text{def}}{=} e_k(\lambda_1(\boldsymbol{A}),\ldots,\lambda_n(\boldsymbol{A})).$$

Then the first and the last polynomials are the trace and determinant of $\boldsymbol{A}$:

$$e_1(\boldsymbol{A}) = \operatorname{Tr}\boldsymbol{A},\quad e_n(\boldsymbol{A}) = \det\boldsymbol{A}.$$

This fact is a special case of a more general result on sums of principal minors.

THEOREM 3.1 (see [19, Theorem 1.2.12]).

$$e_k(\boldsymbol{A}) = \sum_{\substack{\mathcal{Y}\subset\{1,\ldots,n\}\\ \#\mathcal{Y}=k}}\det(\boldsymbol{A}_{\mathcal{Y},\mathcal{Y}}), \tag{15}$$

where $\boldsymbol{A}_{\mathcal{Y},\mathcal{Y}}$ is a submatrix of $\boldsymbol{A}$ with rows and columns indexed by $\mathcal{Y}$, i.e., the sum runs over all principal minors of size $k\times k$.

Remark 3.2. The scaled symmetric polynomials $(-1)^k e_k(\boldsymbol{A})$ are the coefficients of the characteristic polynomial $\det(\lambda\boldsymbol{I}-\boldsymbol{A})$ at the coefficient λ^{n-k}.

3.2. Orders of elementary symmetric polynomials. Next, we assume that $\lambda_1(\varepsilon), \ldots, \lambda_n(\varepsilon)$ are functions of some small parameter ε and we are interested in the orders of the corresponding elementary symmetric polynomials

$$e_s(\varepsilon) \stackrel{\text{def}}{=} e_s(\lambda_1(\varepsilon), \ldots, \lambda_n(\varepsilon)), \quad 1 \le s \le n,$$

as $\varepsilon \to 0$. The following obvious observation will be important.

LEMMA 3.3. *Assume that*

$$\lambda_1(\varepsilon) = \mathcal{O}(\varepsilon^{L_1}),\ \lambda_2(\varepsilon) = \mathcal{O}(\varepsilon^{L_2}),\ \ldots,\ \lambda_n(\varepsilon) = \mathcal{O}(\varepsilon^{L_n}),$$

as $\varepsilon \to 0$ *and* $0 \le L_1 \le \cdots \le L_n$ *are some integers. Then it holds that*

$$e_1(\varepsilon) = \mathcal{O}(\varepsilon^{L_1}),\ e_2(\varepsilon) = \mathcal{O}(\varepsilon^{L_1+L_2}),\ \ldots,\ e_n(\varepsilon) = \mathcal{O}(\varepsilon^{L_1+\cdots+L_n}).$$

Proof. The proof follows from the definition of e_s, the fact that the product $f(\varepsilon) = \mathcal{O}(\varepsilon^a)$ and $g(\varepsilon) = \mathcal{O}(\varepsilon^b)$ is of the order$f(\varepsilon)g(\varepsilon) = \mathcal{O}(\varepsilon^{a+b})$. ∎

We will need a refinement of Lemma 3.3 concerning the main terms of such functions. We distinguish two situations: when the orders of λ_k are separated, and when they form groups. For example, if $\lambda_1(\varepsilon) = a + \mathcal{O}(\varepsilon)$, $\lambda_2(\varepsilon) = \varepsilon(b + \mathcal{O}(\varepsilon))$, $\lambda_3(\varepsilon) = \varepsilon^2(c + \mathcal{O}(\varepsilon))$, then the main terms of elementary symmetric polynomials are products of the main terms of λ_k:

$$e_1(\varepsilon) = a + \mathcal{O}(\varepsilon), \quad e_2(\varepsilon) = \varepsilon(ab + \mathcal{O}(\varepsilon)), \quad e_3(\varepsilon) = \varepsilon^3(abc + \mathcal{O}(\varepsilon)).$$

On the other hand, in the case $\lambda_1(\varepsilon) = a+\mathcal{O}(\varepsilon)$, $\lambda_{k+1}(\varepsilon) = \varepsilon(b_k+\mathcal{O}(\varepsilon))$, $k \in \{1, 2, 3\}$, the behavior of the main terms of elementary symmetric polynomials is different:

$$\begin{aligned}&e_1(\varepsilon) = a + \mathcal{O}(\varepsilon), \quad e_2(\varepsilon) = \varepsilon(a(b_1 + b_2 + b_3) + \mathcal{O}(\varepsilon)),\\ &e_3(\varepsilon) = \varepsilon^2(a(b_1b_2 + b_1b_3 + b_2b_3) + \mathcal{O}(\varepsilon)), \quad e_4(\varepsilon) = \varepsilon^3(ab_1b_2b_3 + \mathcal{O}(\varepsilon)).\end{aligned}$$

The following two lemmas generalize these observations to arbitrary orders.

LEMMA 3.4. *Suppose that* $\lambda_1(\varepsilon), \ldots, \lambda_n(\varepsilon)$ *have the form*

$$\lambda_1(\varepsilon) = \varepsilon^{L_1}(\widetilde{\lambda}_1 + \mathcal{O}(\varepsilon)), \ldots, \lambda_n(\varepsilon) = \varepsilon^{L_n}(\widetilde{\lambda}_n + \mathcal{O}(\varepsilon)) \tag{16}$$

for some integers $0 \le L_1 \le \cdots \le L_r$. *Then*

1. *the elementary symmetric polynomials have the form*

$$\begin{aligned} e_1(\varepsilon) &= \varepsilon^{L_1}(\widetilde{e}_1 + \mathcal{O}(\varepsilon))\\ e_2(\varepsilon) &= \varepsilon^{L_1+L_2}(\widetilde{e}_2 + \mathcal{O}(\varepsilon))\\ &\ \vdots\\ e_n(\varepsilon) &= \varepsilon^{L_1+\cdots+L_n}(\widetilde{e}_n + \mathcal{O}(\varepsilon)); \end{aligned} \tag{17}$$

2. *if either* $s = n$ *or* $L_s < L_{s+1}$, *the main term* $\widetilde{e}_s$ *can be expressed as*

$$\widetilde{e}_s = \widetilde{\lambda}_1 \cdots \widetilde{\lambda}_s.$$

In particular, if $s > 1$ *and* $\widetilde{e}_{s-1} \neq 0$, *then the main term* $\widetilde{\lambda}_s$ *can be found as*

$$\widetilde{\lambda}_s = \frac{\widetilde{e}_s}{\widetilde{e}_{s-1}}.$$

We also need a generalization of Lemma 3.4 to the case of a group of equal L_k.

LEMMA 3.5. *Let* $\lambda_k(\varepsilon)$ *and* L_k *be as in Lemma* 3.4 *(i.e.,* $\lambda_k(\varepsilon)$ *have the form* (16) *and the corresponding* $e_k(\varepsilon)$ *have the form* (17)*), and define* $L_0 = -1$, $L_{n+1} = +\infty$, *for an easier treatment of border cases.*

If for $1 \leq m \leq n-s$, *there is a separated group of* m *functions*

$$\lambda_{s+1}(\varepsilon), \dots, \lambda_{s+m}(\varepsilon)$$

of repeating degree, i.e.,

$$L_s < L_{s+1} = \cdots = L_{s+m} < L_{s+m+1}, \tag{18}$$

then the main terms $\widetilde{e}_{s+k}$, $1 \leq k \leq m$ *in* (17)*, are connected with elementary symmetric polynomials of the main terms* $\widetilde{\lambda}_{s+k}$, $1 \leq k \leq m$ *as follows:*

$$\widetilde{e}_{s+k} = \begin{cases} e_k(\widetilde{\lambda}_{s+1}, \dots, \widetilde{\lambda}_{s+m}), & s = 0, \\ \widetilde{\lambda}_1 \cdots \widetilde{\lambda}_s e_k(\widetilde{\lambda}_{s+1}, \dots, \widetilde{\lambda}_{s+m}) = \widetilde{e}_s e_k(\widetilde{\lambda}_{s+1}, \dots, \widetilde{\lambda}_{s+m}), & s > 0. \end{cases}$$

In particular, if $s > 1$ *and* $\widetilde{e}_s \neq 0$, *the elementary symmetric polynomials for the main terms are equal to*

$$e_k(\widetilde{\lambda}_{s+1}, \dots, \widetilde{\lambda}_{s+m}) = \frac{\widetilde{e}_{s+k}}{\widetilde{e}_s},$$

hence $\widetilde{\lambda}_{s+1}, \dots, \widetilde{\lambda}_{s+m}$ *are the roots of the polynomial (see Remark* 3.2*)*

$$q(\lambda) = \widetilde{e}_s \lambda^m - \widetilde{e}_{s+1} \lambda^{m-1} + \widetilde{e}_{s+2} \lambda^{m-2} + \cdots + (-1)^m \widetilde{e}_{s+m}.$$

Proofs of Lemmas 3.4 to 3.5 are contained in Appendix A. Note that for analytic $\lambda_k(\varepsilon)$, Lemmas 3.4 and 3.5 follow from the Newton–Puiseux theorem (see, for example, [36, section 2.1]), but we prefer to keep a more general formulation in this paper.

Remark 3.6. Assumptions in Lemmas 3.4 to 3.5 can be relaxed (when expansions (16) are valid up to a certain order), but we keep the current statement for simplicity.

3.3. Orders of eigenvalues and eigenvectors. First, we recall a corollary of the Courant–Fischer "min-max" principle giving a bound on smallest eigenvalues.

THEOREM 3.7 (see [19, Theorem 4.3.21]). *Let* $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *be symmetric, and its eigenvalues arranged in nonincreasing order* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. *If there exist an* m*-dimensional subspace* $\mathcal{L} \subset \mathbb{R}^n$ *and a constant* c_1 *such that*

$$\frac{\boldsymbol{u}^\mathsf{T} \boldsymbol{A} \boldsymbol{u}}{\boldsymbol{u}^\mathsf{T} \boldsymbol{u}} \leq c_1$$

for all $\boldsymbol{u} \in \mathcal{L} \setminus \{\boldsymbol{0}\}$, *then the smallest* m *eigenvalues are bounded as*

$$\lambda_n \leq \lambda_{n-1} \leq \cdots \leq \lambda_{n-m+1} \leq c_1.$$

For determining the orders of eigenvalues, we will need the following corollary of Theorem 3.7 (see the proof of [42, Theorem 8]). We provide a short proof for completeness.

LEMMA 3.8. *Suppose that $\boldsymbol{K}(\varepsilon) \in \mathbb{R}^{n\times n}$ is symmetric positive semidefinite for $\varepsilon \in [0, \varepsilon_0]$ and its eigenvalues are ordered as*

$$\lambda_1(\varepsilon) \geq \lambda_2(\varepsilon) \geq \cdots \geq \lambda_n(\varepsilon) \geq 0.$$

Suppose that there exists a matrix $\boldsymbol{U} \in \mathbb{R}^{n\times m}$, $\boldsymbol{U}^\mathsf{T}\boldsymbol{U} = \boldsymbol{I}_m$, such that

$$\boldsymbol{U}^\mathsf{T}\boldsymbol{K}(\varepsilon)\boldsymbol{U} = \mathcal{O}(\varepsilon^L) \tag{19}$$

in $[0, \varepsilon_0]$, $\varepsilon_0 > 0$. Then the last m eigenvalues of $\boldsymbol{K}(\varepsilon)$ are of order at least L, i.e.,

$$\lambda_j(\varepsilon) = \mathcal{O}(\varepsilon^L) \text{ for } n - m < j \leq n. \tag{20}$$

Proof. Assumption (19) and equivalence of matrix norms implies that

$$\|\boldsymbol{U}^\mathsf{T}\boldsymbol{K}(\varepsilon)\boldsymbol{U}\|_2 \leq C\varepsilon^L$$

for some constant C. Hence, we have that for any $\boldsymbol{z} \in \mathbb{R}^m \setminus \{0\}$,

$$\frac{\boldsymbol{z}^\mathsf{T}\boldsymbol{U}^\mathsf{T}\boldsymbol{K}(\varepsilon)\boldsymbol{U}\boldsymbol{z}}{\boldsymbol{z}^\mathsf{T}\boldsymbol{U}^\mathsf{T}\boldsymbol{U}\boldsymbol{z}} \leq \|\boldsymbol{U}^\mathsf{T}\boldsymbol{K}(\varepsilon)\boldsymbol{U}\|_2 \leq C\varepsilon^L.$$

By choosing the range of $\boldsymbol{U}$ as $\mathcal{L}$ and applying Theorem 3.7, we complete the proof. ☐

Next, we recall a classic result on eigenvalues/eigenvectors for analytic perturbations.

THEOREM 3.9 (see [22, Chapter II, Theorem 1.10]). *Let $\boldsymbol{K}(\varepsilon)$ be an $n \times n$ matrix depending analytically on ε in the neighborhood of 0 and symmetric for real ε. Then all the eigenvalues $\lambda_k(\varepsilon)$, $1 \leq k \leq n$, can be chosen analytic; moreover, the orthogonal projectors $\boldsymbol{P}_k(\varepsilon)$ on the corresponding rank-one eigenspaces can be also chosen analytic, so that*

$$\boldsymbol{K}(\varepsilon) = \sum_{k=1}^{n} \lambda_k(\varepsilon)\boldsymbol{P}_k(\varepsilon) \tag{21}$$

is the eigenvalue decomposition of $\boldsymbol{K}(\varepsilon)$ in a neighborhood of 0.

We will be interested in finding the limiting rank-one projectors in Theorem 3.9, i.e.,

$$\boldsymbol{P}_k = \lim_{\varepsilon\to 0} \boldsymbol{P}_k(\varepsilon) = \boldsymbol{P}_k(0),$$

where the last equality follows from analyticity of $\boldsymbol{P}_k(\varepsilon)$ at 0. Note that, for kernel matrices, it is impossible to retrieve the information about the limiting projectors just from $\boldsymbol{K}(0)$ (which is rank-one). In what follows, instead of $\boldsymbol{P}_k$ we will talk about limiting eigenvectors $\boldsymbol{p}_k$ (i.e., $\boldsymbol{P}_k = \boldsymbol{p}_k\boldsymbol{p}_k^\mathsf{T}$), although the latter are defined only up to a change of sign.

Armed with Theorem 3.9, we can obtain an extension of Lemma 3.8, which also gives us information about limiting eigenvectors.

LEMMA 3.10. *Let $\boldsymbol{K}(\varepsilon)$ and $\boldsymbol{U} \in \mathbb{R}^{n\times m}$ be as in Lemma 3.8 and, moreover, $\boldsymbol{K}(\varepsilon)$ be analytic in the neighborhood of 0. Then $\operatorname{span}(\boldsymbol{U})$ contains all the limiting eigenvectors corresponding to the eigenvalues of the order at least L (i.e., $\mathcal{O}(\varepsilon^L)$).*

Proof. Let $L_1 \leq \cdots \leq L_k \leq +\infty$ be the exact orders of the eigenvalues, i.e., either

$$\lambda_k(\varepsilon) = \varepsilon^{L_k}(\widetilde{\lambda}_k + \mathcal{O}(\varepsilon)), \quad \widetilde{\lambda}_k > 0,$$

or $L_k = +\infty$ in the case $\lambda_k(\varepsilon) \equiv 0$. Then, from orthogonality of the eigenvectors, we only need to prove that $\boldsymbol{p}_k^\mathsf{T}\boldsymbol{U} = 0$ for $L_k < L$. Note that $\widetilde{\lambda}_k > 0$ due to positive semidefiniteness of $\boldsymbol{K}(\varepsilon)$.

Suppose that, on the contrary, there exists k with $L_k < L$ such that $\boldsymbol{p}_k^\mathsf{T}\boldsymbol{U} \neq 0$, and let it be the minimal such k. Then from the factorization (21) we have that

$$\boldsymbol{U}^\mathsf{T}\boldsymbol{K}(\varepsilon)\boldsymbol{U} = \varepsilon^{L_k}\Big(\sum_{\ell: L_\ell = L_k} \widetilde{\lambda}_\ell \boldsymbol{U}^\mathsf{T}\boldsymbol{p}_k\boldsymbol{p}_k^\mathsf{T}\boldsymbol{U} + \mathcal{O}(\varepsilon)\Big)$$

with a nonzero main term (from $\widetilde{\lambda}_k > 0$), which contradicts the assumption (19). ∎

3.4. Saddle point matrices and elementary symmetric polynomials. In this subsection, we will be interested in determinants and elementary symmetric polynomials for so-called saddle point matrices [4]. Let $\boldsymbol{V} \in \mathbb{R}^{n\times r}$ be a full column rank matrix. For a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we define the *saddle-point matrix* as

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{V} \\ \boldsymbol{V}^\mathsf{T} & 0 \end{bmatrix}.$$

Consider a full QR decomposition of $\boldsymbol{V} \in \mathbb{R}^{n\times r}$, i.e.,

$$\boldsymbol{V} = \boldsymbol{Q}_{\text{full}}\boldsymbol{R}_{\text{full}} = \boldsymbol{Q}\boldsymbol{R}, \quad \boldsymbol{Q}_{\text{full}} = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{Q}_\perp \end{bmatrix}, \quad \boldsymbol{R}_{\text{full}} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}, \tag{22}$$

where $\boldsymbol{Q}_{\text{full}} \in \mathbb{R}^{n\times n}$, $\boldsymbol{Q}_\perp \in \mathbb{R}^{n\times(n\text{-}r)}$, $\boldsymbol{Q}_{\text{full}}^\mathsf{T}\boldsymbol{Q}_{\text{full}} = \boldsymbol{I}$, $\boldsymbol{R}$ is upper triangular, and $\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{V} = 0$.

LEMMA 3.11. *For any $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times r}$, it holds that*

$$\det\begin{bmatrix} \boldsymbol{A} & \boldsymbol{V} \\ \boldsymbol{V}^\mathsf{T} & 0 \end{bmatrix} = (-1)^r \det(\boldsymbol{V}^\mathsf{T}\boldsymbol{V})\det(\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp) = (-1)^r(\det \boldsymbol{R})^2\det(\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp),$$

where the matrices $\boldsymbol{R}$ and $\boldsymbol{Q}_\perp$ are from the full QR factorization given in (22).

The proof of Lemma 3.11 is contained in Appendix B.

We will also also need to evaluate the elementary symmetric polynomials of matrices of the form $\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp$. For a power series (or a polynomial)

$$a(t) = a_0 + a_1 t + \cdots + a_N t^N + \ldots,$$

we use the following notation, standard in combinatorics, for its coefficients:

$$[t^r]\left\{a(t)\right\} \stackrel{\text{def}}{=} a_r = \frac{a^{(r)}(0)}{r!}.$$

With this notation, the following lemma on low-rank perturbations of $\boldsymbol{A}$ holds true.

LEMMA 3.12. *Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times r}$ with the QR decomposition of $\boldsymbol{V}$ given as before by* (22). *Then for $k \geq r$ the polynomial $a(t) = e_k(\boldsymbol{A} + t\boldsymbol{V}\boldsymbol{V}^\mathsf{T})$ has degree at most r, and its leading coefficient is given by*

$$[t^r]\left\{e_k(\boldsymbol{A} + t\boldsymbol{V}\boldsymbol{V}^\mathsf{T})\right\} = \det(\boldsymbol{V}^\mathsf{T}\boldsymbol{V})e_{k-r}(\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp).$$

In particular, if $k = n$, we get

$$[t^r]\left\{\det(\boldsymbol{A} + t\boldsymbol{V}\boldsymbol{V}^\mathsf{T})\right\} = \det(\boldsymbol{V}^\mathsf{T}\boldsymbol{V})\det(\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp) = (-1)^r\det\begin{bmatrix} \boldsymbol{A} & \boldsymbol{V} \\ \boldsymbol{V}^\mathsf{T} & 0 \end{bmatrix}.$$

The proof of Lemma 3.12 is also contained in Appendix B.

Remark 3.13. Alternative expressions for perturbations of elementary symmetric polynomials are available in [20, Theorem 2.16], and [21, Corollary 3.3], but they do not lead directly to the compact expression in Lemma 3.12 that we need.

4. Results in the 1D case.

4.1. Smooth kernels. We begin this section by generalizing the result of [26, Corollary 2.9] on determinants of scaled kernel matrix $\boldsymbol{K}_\varepsilon$ in the smooth case.

THEOREM 4.1. *Let* $K \in \mathcal{C}^{(n,n)}(\Omega\times\Omega)$ *be the kernel, and* (9) *be its scaled version. Then for small* ε,

1. *the determinant of* $\boldsymbol{K}_\varepsilon$ *from* (1) *has the expansion*

$$\det(\boldsymbol{K}_\varepsilon) = \varepsilon^{n(n-1)}(\det(\boldsymbol{V}_{\le n-1})^2 \det \boldsymbol{W}_{\le n-1} + \mathcal{O}(\varepsilon));$$

2. *if* $\boldsymbol{K}_\varepsilon$ *is positive semidefinite on* $[0,\varepsilon_0], \varepsilon_0 > 0$, *then eigenvalues have orders*

$$\lambda_k(\varepsilon) = \mathcal{O}(\varepsilon^{2(k-1)}).$$

While the proof of 1 is given in [26, Corollary 2.9], and the proof of 2 for the radial analytic kernels is contained in [31, 42], we provide a short proof of Theorem 4.1 in subsection 4.3, which also illustrates the main ideas behind other proofs in the paper.

Theorem 4.1, together with Theorem 3.1, allows us to find the main terms of the eigenvalues for analytic-in-parameter $\boldsymbol{K}_\varepsilon$.

THEOREM 4.2. *Let* $K \in \mathcal{C}^{(n,n)}(\Omega)$ *be the kernel such that* $\boldsymbol{K}_\varepsilon$ *is symmetric positive semidefinite on* $[0,\varepsilon_0], \varepsilon_0 > 0$, *and analytic in* ε *in the neighborhood of* 0. *Then for* $s \le n$ *it holds that*

$$\lambda_s = \varepsilon^{2(s-1)}(\widetilde{\lambda}_s + \mathcal{O}(\varepsilon)),$$

where the main terms satisfy

$$\widetilde{\lambda}_1 \dots \widetilde{\lambda}_s = \det(\boldsymbol{V}_{\le s-1}^{\mathsf{T}}\boldsymbol{V}_{\le s-1}) \det(\boldsymbol{W}_{\le s-1}). \tag{23}$$

Proof. First, due to analyticity and Theorem 4.1, we have that expansions (16) are valid for $L_s = 2(s-1)$. Second, the submatrices of $\boldsymbol{K}_\varepsilon$ are also kernel matrices (of smaller size), which, in turn can be found from Theorem 4.1. More precisely,

$$\begin{aligned}
e_s(\boldsymbol{K}_\varepsilon) &= \sum_{1\le k_1<\dots<k_s\le n} \det(\boldsymbol{K}_\varepsilon(x_{k_1},\dots,x_{k_s}))\\
&= \varepsilon^{s(s-1)}\Big(\det \boldsymbol{W}_{\le s-1} \sum_{1\le k_1<\dots<k_s\le n} (\det \boldsymbol{V}_{\le s-1}(x_{k_1},\dots,x_{k_s}))^2 + \mathcal{O}(\varepsilon)\Big)\\
&= \varepsilon^{s(s-1)}\left(\det \boldsymbol{W}_{\le s-1} \det(\boldsymbol{V}_{\le s-1}^{\mathsf{T}}\boldsymbol{V}_{\le s-1}) + \mathcal{O}(\varepsilon)\right), \qquad (24)
\end{aligned}$$

where the penultimate equality follows from Theorem 4.1, and the last equality follows from the Binet–Cauchy formula. ∎

Finally, we employ Lemma 3.4 on relations between the main terms in (16) and (17).

COROLLARY 4.3. *Let* $\boldsymbol{K}_\varepsilon$ *be as in Theorem* 4.2, *and the points in* $\mathcal{X}$ *be distinct.*

1. *If for* $1 < s \leq n$, $\det \boldsymbol{W}_{\leq s-2} \neq 0$, *the main term of the sth eigenvalue can be obtained as*

$$\widetilde{\lambda}_s = \frac{\det(\boldsymbol{V}_{\leq s-1}^{\mathsf{T}} \boldsymbol{V}_{\leq s-1})}{\det(\boldsymbol{V}_{\leq s-2}^{\mathsf{T}} \boldsymbol{V}_{\leq s-2})} \cdot \frac{\det(\boldsymbol{W}_{\leq s-1})}{\det(\boldsymbol{W}_{\leq s-2})}. \tag{25}$$

2. *If* $1 \leq s < n$, $\det \boldsymbol{W}_{\leq s-1} \neq 0$, *then the limiting eigenvectors* $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$ *are the first s columns of the* $\boldsymbol{Q}_{\text{full}}$ *factor of the QR factorization* (22) *of* $\boldsymbol{V}_{\leq s-1}$.
3. *In particular, if* $\det \boldsymbol{W}_{\leq n-2} \neq 0$, *then all the main terms of the eigenvalues are given by* (25), *and all the limiting eigenvectors are given by the columns of the* $\boldsymbol{Q}$ *matrix in the QR factorization of* $\boldsymbol{V}_{\leq n-1}$.

The proof of Corollary 4.3 is also contained in subsection 4.3.

Remark 4.4. In Corollary 4.3, by Cramer's rule, the individual ratios in (25) can be computed in the following way:

$$\frac{\det(\boldsymbol{V}_{\leq s-1}^{\mathsf{T}} \boldsymbol{V}_{\leq s-1})}{\det(\boldsymbol{V}_{\leq s-2}^{\mathsf{T}} \boldsymbol{V}_{\leq s-2})} = ((\boldsymbol{V}_{\leq s-1}^{\mathsf{T}} \boldsymbol{V}_{\leq s-1})^{-1})_{s,s}^{-1} = R_{s,s}^2,$$

where $R_{s,s}$ is the last $((s,s)$th) diagonal element of the $\boldsymbol{R} \in \mathbb{R}^{s \times s}$ matrix in the thin QR decomposition of $\boldsymbol{V}_{\leq s-1}$. Similarly,

$$\frac{\det(\boldsymbol{W}_{\leq s-1})}{\det(\boldsymbol{W}_{\leq s-2})} = ((\boldsymbol{W}_{\leq s-1})^{-1})_{s,s}^{-1} = C_{s,s}^2,$$

where $C_{s,s}$ is the last diagonal element of the Cholesky factor of $\boldsymbol{W}_{\leq s-1} = \boldsymbol{C}\boldsymbol{C}^{\mathsf{T}}$.

4.2. Finite smoothness. Next, we provide an analogue of Theorem 4.1 for a radial kernel with the order of smoothness r, which is smaller or equal to the number of points (i.e., Theorem 4.1 cannot be applied).

THEOREM 4.5. *For a radial kernel* (10) *with order of smoothness* $r \leq n$,

1. *the determinant can be expressed as*

$$\det(\boldsymbol{K}_\varepsilon) = \varepsilon^{n(2r-1)-r^2} \left(\widetilde{k} + \mathcal{O}(\varepsilon)\right),$$

where the main term is given by

$$\widetilde{k} = (-1)^r \det \boldsymbol{W}_{\leq r-1} \det \begin{bmatrix} f_{2r-1}\boldsymbol{D}_{(2r-1)} & \boldsymbol{V}_{\leq r-1} \\ \boldsymbol{V}_{\leq r-1}^{\mathsf{T}} & 0 \end{bmatrix} \tag{26}$$

$$= \det \boldsymbol{W}_{\leq r-1} \det(\boldsymbol{V}_{\leq r-1}^{\mathsf{T}} \boldsymbol{V}_{\leq r-1}) \det(f_{2r-1} \boldsymbol{Q}_{\perp}^{\mathsf{T}} \boldsymbol{D}_{(2r-1)} \boldsymbol{Q}_{\perp}), \tag{27}$$

where $\boldsymbol{Q}_\perp \in \mathbb{R}^{n \times (n-r)}$ *is the semiorthogonal matrix such that* $\boldsymbol{Q}_\perp^{\mathsf{T}} \boldsymbol{V}_{\leq r-1} = 0$ *(e.g., the matrix* $\boldsymbol{Q}_\perp$ *in the full QR decomposition* (22) *of* $\boldsymbol{V} = \boldsymbol{V}_{\leq r-1}$*);*

2. *if* $\boldsymbol{K}_\varepsilon$ *is positive semidefinite on* $[0, \varepsilon_0], \varepsilon_0 > 0$, *the eigenvalues have orders*

$$\lambda_s(\varepsilon) = \begin{cases} \mathcal{O}(\varepsilon^{2(s-1)}), & s \leq r, \\ \mathcal{O}(\varepsilon^{2r-1}), & s > r. \end{cases}$$

The proof of Theorem 4.5 again is postponed to subsection 4.4 in order to present a more straightforward corollary on eigenvalues.

As an example, we have for $r = 1$ (exponential kernel)

$$\det(\boldsymbol{K}_\varepsilon) = -\varepsilon^{n-1} n K(0,0) \det \begin{bmatrix} \boldsymbol{D}_{(1)} & \boldsymbol{1} \\ \boldsymbol{1}^\top & 0 \end{bmatrix}, \tag{28}$$

where $\boldsymbol{1}$ denotes a vector with all entries equal to 1.

Combining Theorem 4.5 with Lemmas 3.5 and 3.10, we get the following result.

THEOREM 4.6. *Let K be a kernel satisfying the assumptions of Theorem 4.5, where $\boldsymbol{K}_\varepsilon$ is positive semidefinite on $[0, \varepsilon_0], \varepsilon_0 > 0$, and analytic in ε. Then it holds that*

1. *the main terms of the first r eigenvalues $\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_r$ satisfy (23) for $1 \le s \le r$. In particular, if for $1 < s \le r$, $\det \boldsymbol{W}_{\le s-2} \ne 0$, then $\widetilde{\lambda}_s$ is given by (25);*
2. *if the points in $\mathcal{X}$ are distinct and for $1 \le s \le r$, $\det \boldsymbol{W}_{\le s-1} \ne 0$, then the first s limiting eigenvectors are as in Corollary 4.3. In particular, for the case $\det \boldsymbol{W}_{\le r-1} \ne 0$, the last $n - r$ limiting eigenvectors span the column space of $\boldsymbol{Q}_\perp$;*
3. *if $\det \boldsymbol{W}_{\le r-1} \ne 0$ and the points in $\mathcal{X}$ are distinct, then $\widetilde{\lambda}_{r+1}, \ldots, \widetilde{\lambda}_n$ are the eigenvalues of*

$$f_{2r-1}(\boldsymbol{Q}_\perp^\top \boldsymbol{D}_{(2r-1)} \boldsymbol{Q}_\perp).$$

The proof of Theorem 4.6 is also contained in subsection 4.4. Note that we obtain results on the precise locations of the last $n - r$ limiting eigenvectors in section 8.

4.3. Proofs for the 1D smooth case. We need the following technical lemma.

LEMMA 4.7. *For any upper triangular matrix $\boldsymbol{R} \in \mathbb{R}^{(k+1)\times(k+1)}$ it holds that*

$$\boldsymbol{\Delta}^{-1} \boldsymbol{R} \boldsymbol{\Delta} = \operatorname{diag}(\boldsymbol{R}) + \mathcal{O}(\varepsilon),$$

where $\operatorname{diag}(\boldsymbol{R})$ *is the diagonal part of* $\boldsymbol{R}$ *and* $\boldsymbol{\Delta} = \boldsymbol{\Delta}_k(\varepsilon) \in \mathbb{R}^{(k+1)\times(k+1)}$ *is defined as*

$$\boldsymbol{\Delta}_k(\varepsilon) \stackrel{def}{=} \begin{bmatrix} 1 & & & \\ & \varepsilon & & \\ & & \ddots & \\ & & & \varepsilon^k \end{bmatrix}. \tag{29}$$

Proof. A direct calculation gives

$$\boldsymbol{\Delta}^{-1} \boldsymbol{R} \boldsymbol{\Delta} = \begin{bmatrix} R_{1,1} & \varepsilon R_{1,2} & \cdots & \varepsilon^k R_{1,k+1} \\ 0 & R_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \varepsilon R_{k,k+1} \\ 0 & \cdots & 0 & R_{k+1,k+1} \end{bmatrix}.$$

∎

Proof of Theorem 4.1. We will use a special form of the Maclaurin expansion (i.e., the Taylor expansion at 0) for bivariate functions, differentiable with respect to a "rectangular" set of multi-indices. Let us take $x, y \in \Omega$ and apply first the Maclaurin expansion with respect to x:

$$K(x,y) = K(0,y) + xK^{(1,0)}(0,y) + \cdots + \frac{x^{n-1} K^{(n-1,0)}(0,y)}{(n-1)!} + r_n(x,y). \tag{30}$$

Then the Maclaurin expansion of (30) with respect to y yields

$$
\begin{array}{rccccccc|c}
K(x,y) = & K^{(0,0)} & + \cdots + & \frac{y^{n\text{-}1}K^{(0,n\text{-}1)}}{(n\text{-}1)!} & + & \frac{y^{n}K^{(0,n)}(0,\theta_{y,1}y)}{n!} & & & = K(0,y) \\
+ & xK^{(1,0)} & + \cdots + & \frac{xy^{n\text{-}1}K^{(1,n\text{-}1)}}{(n\text{-}1)!} & + & \frac{xy^{n}K^{(1,n)}(0,\theta_{y,2}y)}{n!} & & & = xK^{(1,0)}(0,y) \\
 & \vdots & & \vdots & & \vdots & & & \vdots \\
+ & \frac{x^{n\text{-}1}K^{(n\text{-}1,0)}}{(n\text{-}1)!} & + \cdots + & \frac{x^{n\text{-}1}y^{n\text{-}1}K^{(n\text{-}1,n\text{-}1)}}{(n\text{-}1)!(n\text{-}1)!} & + & \frac{x^{n\text{-}1}y^{n}K^{(n\text{-}1,n)}(0,\theta_{y,n}y)}{(n\text{-}1)!n!} & & & = \frac{x^{n\text{-}1}K^{(n\text{-}1,0)}(0,y)}{(n\text{-}1)!} \\
+ & r_n(x,y), & & & & & & &
\end{array}
$$

where $0 \le \theta_{y,1}, \ldots, \theta_{y,n} \le 1$ depend only on y, the corresponding terms of (30) are on the right, and a shorthand notation $K^{(i,j)} = K^{(i,j)}(0,0)$ is used. From the integral form of $r_n(x,y)$ and Taylor expansion for $g(y) \stackrel{\text{def}}{=} K^{(n,0)}(t,y)$ (as in (30)), we get

$$
\begin{aligned}
r_n(x,y) &= \int_0^x \frac{(x-t)^{n-1}K^{(n,0)}(t,y)}{(n-1)!}dt \\
&= \sum_{\ell=0}^{n-1}\left(\int_0^x \frac{(x-t)^{n-1}}{(n-1)!}\frac{y^{\ell}K^{(n,\ell)}(t,0)}{\ell!}dt\right) + \int_0^x\int_0^y \frac{(x-t)^{n-1}(y-s)^{n-1}K^{(n,n)}(t,s)}{(n-1)!(n-1)!}dtds.
\end{aligned}
$$

By the mean value theorem (as in [45, section 6.3.3]), there exist $0 \le \eta_{x,1}, \ldots, \eta_{x,n} \le 1$ (depending only on x) and $0 \le \zeta_{x,y}, \xi_{x,y} \le 1$, such that $r_n(x,y) =$

$$
\frac{x^nK^{(n,0)}(\eta_{x,1}x,0)}{n!} + \cdots + \frac{x^ny^{n\text{-}1}K^{(n,n\text{-}1)}(\eta_{x,n}x,0)}{n!(n\text{-}1)!} + \frac{x^ny^nK^{(n,n)}(\zeta_{x,y}x,\xi_{x,y}y)}{n!n!}.
$$

Next, with some abuse of notation, let $\varepsilon_0 > 0$ be such that $\varepsilon_0 x, \varepsilon_0 y \in \Omega$. Then, after replacing (x,y) with $(\varepsilon x, \varepsilon y)$, $\varepsilon \in [0,\varepsilon_0]$ in the expansion of $K(x,y)$, we obtain

$$
\begin{aligned}
&K(\varepsilon x, \varepsilon y) = [1, \varepsilon x, \ldots, (\varepsilon x)^{n-1}]\boldsymbol{W}_{\le n-1}[1, \varepsilon y, \ldots, (\varepsilon y)^{n-1}]^{\mathsf{T}} \\
&\quad + \varepsilon^n([1, \varepsilon x, \ldots, (\varepsilon x)^{n-1}]\boldsymbol{w}_{1,y}(\varepsilon) + \boldsymbol{w}_{2,x}^{\mathsf{T}}(\varepsilon)[1, \varepsilon y, \ldots, (\varepsilon y)^{n-1}]^{\mathsf{T}}) + \varepsilon^{2n}w_{3,x,y}(\varepsilon),
\end{aligned} \tag{31}
$$

where $\boldsymbol{W}_{\le n-1}$ is the Wronskian matrix defined in (4),

$$
w_{3,x,y}(\varepsilon) \stackrel{\text{def}}{=} \frac{x^ny^nK^{(n,n)}(\zeta_{\varepsilon x,\varepsilon y}(\varepsilon x), \xi_{\varepsilon x,\varepsilon y}(\varepsilon y))}{n!n!},
$$

such that $w_{3,x,y} : [0,\varepsilon_0] \to \mathbb{R}^n$ is bounded and $\boldsymbol{w}_{1,y}, \boldsymbol{w}_{2,x} : [0,\varepsilon_0] \to \mathbb{R}^n$ are bounded vector functions (depending only on y and x, respectively), defined as

$$
\begin{aligned}
\boldsymbol{w}_{1,y}(\varepsilon) &\stackrel{\text{def}}{=} \frac{y^n}{n!}\begin{bmatrix} K^{(0,n)}(0,\theta_{\varepsilon y,1}(\varepsilon y)) & K^{(1,n)}(0,\theta_{\varepsilon y,2}(\varepsilon y)) & \cdots & \frac{K^{(n-1,n)}(0,\theta_{\varepsilon y,n}(\varepsilon y))}{(n-1)!}\end{bmatrix}^{\mathsf{T}}, \\
\boldsymbol{w}_{2,x}(\varepsilon) &\stackrel{\text{def}}{=} \frac{x^n}{n!}\begin{bmatrix} K^{(n,0)}(\eta_{\varepsilon x,1}(\varepsilon x),0) & K^{(n,1)}(\eta_{\varepsilon x,2}(\varepsilon x),0) & \cdots & \frac{K^{(n,n-1)}(\eta_{\varepsilon x,n}(\varepsilon x),0)}{(n-1)!}\end{bmatrix}^{\mathsf{T}}.
\end{aligned}
$$

Let $\varepsilon_0 > 0$, such that $\{\varepsilon_0 x_1, \ldots, \varepsilon_0 x_n\} \subset \Omega$. From (31), the scaled kernel matrix admits for $\varepsilon \in [0,\varepsilon_0]$ the expansion

$$
\boldsymbol{K}_{\varepsilon} = \boldsymbol{V\Delta W \Delta V}^{\mathsf{T}} + \varepsilon^{k+1}(\boldsymbol{V\Delta W}_1(\varepsilon) + \boldsymbol{W}_2(\varepsilon)\boldsymbol{\Delta V}^{\mathsf{T}}) + \varepsilon^{2(k+1)}\boldsymbol{W}_3(\varepsilon) \tag{32}
$$

with $k = n-1$, $\boldsymbol{W} = \boldsymbol{W}_{\le k}$, $\boldsymbol{V} = \boldsymbol{V}_{\le k}$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}_k(\varepsilon)$ as in (29), and $\boldsymbol{W}_1(\varepsilon), \boldsymbol{W}_2(\varepsilon), \boldsymbol{W}_3(\varepsilon)$ are $\mathcal{O}(1)$ matrices defined, respectively, as $\boldsymbol{W}_3(\varepsilon) \stackrel{\text{def}}{=} \left[w_{3,x_i,x_j}(\varepsilon)\right]_{i,j=1}^{n,n}$,

$$\boldsymbol{W}_1(\varepsilon) \stackrel{\text{def}}{=} \begin{bmatrix}\boldsymbol{w}_{1,x_1}(\varepsilon) & \cdots & \boldsymbol{w}_{1,x_n}(\varepsilon)\end{bmatrix}, \boldsymbol{W}_2(\varepsilon) \stackrel{\text{def}}{=} \begin{bmatrix}\boldsymbol{w}_{2,x_1}(\varepsilon) & \cdots & \boldsymbol{w}_{2,x_n}(\varepsilon)\end{bmatrix}^{\mathsf{T}}.$$

Let $\boldsymbol{V} = \boldsymbol{QR}$ be the (full) QR factorization. Then from Lemma 4.7, we have that

$$\boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{V}\boldsymbol{\Delta} = \widetilde{\boldsymbol{R}} + \mathcal{O}(\varepsilon), \quad \text{where } \widetilde{\boldsymbol{R}} = \operatorname{diag}(\boldsymbol{R}).$$

By pre-/post-multiplying (32) by $\boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{\mathsf{T}}$ and its transpose, we get

$$\begin{aligned}\boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}\boldsymbol{\Delta}^{-1} &= (\widetilde{\boldsymbol{R}} + \mathcal{O}(\varepsilon))\boldsymbol{W}(\widetilde{\boldsymbol{R}}^{\mathsf{T}} + \mathcal{O}(\varepsilon))\\ &\quad + \varepsilon^{k+1}(\widetilde{\boldsymbol{R}} + \mathcal{O}(\varepsilon))\boldsymbol{W}_1(\varepsilon)\boldsymbol{Q}\boldsymbol{\Delta}^{-1} + \varepsilon^{k+1}\boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{W}_2(\varepsilon)(\widetilde{\boldsymbol{R}}^{\mathsf{T}} + \mathcal{O}(\varepsilon))\\ &\quad + (\varepsilon^{k+1}\boldsymbol{\Delta}^{-1})\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{W}_3(\varepsilon)\boldsymbol{Q}(\varepsilon^{k+1}\boldsymbol{\Delta}^{-1})\\ &= \widetilde{\boldsymbol{R}}\boldsymbol{W}\widetilde{\boldsymbol{R}}^{\mathsf{T}} + \mathcal{O}(\varepsilon),\end{aligned} \tag{33}$$

where the last equality follows from $\varepsilon^{k+1}\boldsymbol{\Delta}^{-1} = \mathcal{O}(\varepsilon)$. Now we are ready to prove the statements of the theorem.

1. From (33) we immediately get

$$\varepsilon^{-n(n-1)} \det \boldsymbol{K}_\varepsilon = (\det \widetilde{\boldsymbol{R}})^2 \det \boldsymbol{W} + \mathcal{O}(\varepsilon) = (\det \boldsymbol{R})^2 \det \boldsymbol{W} + \mathcal{O}(\varepsilon).$$

2. Since $k = n-1$, we can also rewrite (33) as

$$\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q} = \begin{bmatrix}\mathcal{O}(1) & \mathcal{O}(\varepsilon) & \cdots & \mathcal{O}(\varepsilon^k)\\ \mathcal{O}(\varepsilon) & \mathcal{O}(\varepsilon^2) & \cdots & \mathcal{O}(\varepsilon^{k+1})\\ \vdots & \vdots & & \vdots\\ \mathcal{O}(\varepsilon^k) & \mathcal{O}(\varepsilon^{k+1}) & \cdots & \mathcal{O}(\varepsilon^{2k})\end{bmatrix}, \tag{34}$$

whose lower-right submatrices have orders

$$(\boldsymbol{Q}_{s:k})^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_{s:k} = \mathcal{O}(\varepsilon^{2s}), \tag{35}$$

where $\boldsymbol{Q}_{s:k}$ denotes the matrix of the last $n-s$ columns of $\boldsymbol{Q}$. By Lemma 3.8, (35) implies the required orders of the eigenvalues. ∎

Proof of Corollary 4.3. 1. Part 1 follows from (23) and Lemma 3.4.

2. From Theorem 4.2, we have $\widetilde{\lambda}_\ell \neq 0$ for $1 \le \ell \le s$. Let $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ be the factor of the QR factorization of $\boldsymbol{V}_{\le n-1}$ (which can also be taken as $\boldsymbol{Q}_{\text{full}}$ factor in the full QR factorization for any $\boldsymbol{V}_{\le \ell-1}$, $\ell < n$). From (35) and Lemma 3.10, $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_\ell$ are orthogonal to $\operatorname{span}(\boldsymbol{Q}_{:,\ell:k})$ for each $1 \le \ell \le s$. Due to orthonormality of the columns of $\boldsymbol{Q}$, the vectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$ must coincide with its first s columns.

3. The result on eigenvalues follows from 1. For the result on eigenvectors, we note that if the first $n-1$ columns of $\boldsymbol{Q}$ are limiting eigenvectors, then the last column should be the remaining limiting eigenvector. ∎

4.4. Proofs for the 1D finite smoothness case.

Proof of Theorem 4.5. First, we will rewrite the expansion (11) in a convenient form. We will group the elements in (12) to get

$$\boldsymbol{K}_\varepsilon = \boldsymbol{V}_{\le 2r-2}\boldsymbol{\Delta}_{2r-2}(\varepsilon)\boldsymbol{W}_{\blacktriangleright}\boldsymbol{\Delta}_{2r-2}(\varepsilon)\boldsymbol{V}_{\le 2r-2}^{\mathsf{T}} + \varepsilon^{2r-1}(f_{2r-1}\boldsymbol{D}_{(2r-1)} + \mathcal{O}(\varepsilon)),$$

where $\boldsymbol{W}_{\blacktriangleright} \in \mathbb{R}^{(2r-1)\times(2r-1)}$ is the antitriangular matrix defined as

$$\boldsymbol{W}_{\blacktriangleright} = \boldsymbol{W}_{\diagup 0} + \boldsymbol{W}_{\diagup 2} + \cdots + \boldsymbol{W}_{\diagup 2(r-2)}, \tag{36}$$

and $\boldsymbol{W}_{\diagup s}$ are defined[4] in (13). For example, in the case when $r = 2$

$$\boldsymbol{W}_{\blacktriangleright} = \begin{bmatrix} f_0 & 0 & f_1 \\ 0 & -2f_1 & 0 \\ f_1 & 0 & 0 \end{bmatrix}.$$

Next, we note that $\boldsymbol{W}_{\blacktriangleright}$ can be split as

$$\boldsymbol{W}_{\blacktriangleright} = \begin{bmatrix} \boldsymbol{W}_{\le r-1} & \widetilde{\boldsymbol{W}}_1 \\ \widetilde{\widetilde{\boldsymbol{W}}}_2 & \end{bmatrix},$$

where $\boldsymbol{W}_{\le r-1}$ is exactly the Wronskian matrix defined in (4). Therefore, since the matrices $\boldsymbol{V}_{\le 2r-2}$ and $\boldsymbol{\Delta}_{2r-2}(\varepsilon)$ can be partitioned as

$$\boldsymbol{V}_{\le 2r-2} = \begin{bmatrix} \boldsymbol{V}_{\le r-1} & \boldsymbol{V}_{rest} \end{bmatrix}, \quad \boldsymbol{\Delta}_{2r-2}(\varepsilon) = \begin{bmatrix} \boldsymbol{\Delta}_{r-1}(\varepsilon) & \\ & \varepsilon^r \boldsymbol{\Delta}_{r-2}(\varepsilon) \end{bmatrix},$$

we get

$$\begin{aligned} \boldsymbol{V}_{\le 2r-2}\boldsymbol{\Delta}_{2r-2}(\varepsilon)\widetilde{\boldsymbol{W}}\boldsymbol{\Delta}_{2r-2}(\varepsilon)\boldsymbol{V}_{\le 2r-2}^{\mathsf{T}} &= \boldsymbol{V}_{\le r-1}\boldsymbol{\Delta}_{r-1}(\varepsilon)\boldsymbol{W}_{\le r-1}\boldsymbol{\Delta}_{r-1}(\varepsilon)\boldsymbol{V}_{\le r-1}^{\mathsf{T}} \\ + \varepsilon^r \boldsymbol{V}_{\le r-1}\boldsymbol{\Delta}_{r-1}(\varepsilon) &\underbrace{\widetilde{\boldsymbol{W}}_1\boldsymbol{\Delta}_{r-2}(\varepsilon)\boldsymbol{V}_{rest}^{\mathsf{T}}}_{\boldsymbol{W}_1(\varepsilon)} + \varepsilon^r \underbrace{\boldsymbol{V}_{rest}\boldsymbol{\Delta}_{r-2}(\varepsilon)\widetilde{\boldsymbol{W}}_2}_{\boldsymbol{W}_2(\varepsilon)}\boldsymbol{\Delta}_{r-1}(\varepsilon)\boldsymbol{V}_{\le r-1}^{\mathsf{T}}, \end{aligned}$$

which, after denoting $\boldsymbol{W} = \boldsymbol{W}_{\le r-1}$, $\boldsymbol{V} = \boldsymbol{V}_{\le r-1}$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}_{r-1}(\varepsilon)$, gives

$$\boldsymbol{K}_\varepsilon = \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{V}^{\mathsf{T}} + \varepsilon^r(\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{W}_1(\varepsilon) + \boldsymbol{W}_2(\varepsilon)\boldsymbol{\Delta}\boldsymbol{V}^{\mathsf{T}}) + \varepsilon^{2r-1}(f_{2r-1}\boldsymbol{D}_{(2r-1)} + \mathcal{O}(\varepsilon)).$$

Next, we take the QR decomposition $\boldsymbol{V}$ (22) and consider a diagonal scaling matrix

$$\widetilde{\boldsymbol{\Delta}} = \begin{bmatrix} \boldsymbol{\Delta} & \\ & \varepsilon^{r-1}\boldsymbol{I}_{n-r} \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

After pre-/post-multiplying $\boldsymbol{K}_\varepsilon$ by $\widetilde{\boldsymbol{\Delta}}^{-1}\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}$ and its transpose, we get

$$\widetilde{\boldsymbol{\Delta}}^{-1}\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_{\text{full}}\widetilde{\boldsymbol{\Delta}}^{-1} = \begin{bmatrix} \boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}\boldsymbol{\Delta}^{-1} & \varepsilon^{1-r}\boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_\perp \\ \varepsilon^{1-r}\boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}\boldsymbol{\Delta}^{-1} & \varepsilon^{2-2r}\boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_\perp \end{bmatrix}, \tag{37}$$

where $\boldsymbol{Q}_{\text{full}} = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{Q}_\perp \end{bmatrix}$ as in (22). For the upper-left block we get, by Lemma 4.7,

$$\boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}\boldsymbol{\Delta}^{-1} = \operatorname{diag}(\boldsymbol{R})\boldsymbol{W}\operatorname{diag}(\boldsymbol{R}) + \mathcal{O}(\varepsilon).$$

The lower-left block (which is a transpose of the upper-right one) becomes

$$\varepsilon^{1-r}\boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}\boldsymbol{\Delta}^{-1} = \varepsilon(\boldsymbol{Q}_\perp^{\mathsf{T}}(\boldsymbol{W}_2(\varepsilon)\operatorname{diag}(\boldsymbol{R}) + \varepsilon^{r-1}f_{2r-1}\boldsymbol{D}_{(2r-1)}\boldsymbol{Q}\boldsymbol{\Delta}^{-1}) + \mathcal{O}(\varepsilon)).$$

Finally, the lower-right block is

$$\varepsilon^{2-2r}\boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_\perp = \varepsilon(f_{2r-1}\boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{D}_{(2r-1)}\boldsymbol{Q}_\perp + \mathcal{O}(\epsilon)).$$

[4] In the sum (36), $\boldsymbol{W}_{\diagup 2\ell}$ are padded by zeros to $(2r-1)\times(2r-1)$ matrices.

1. Combining the blocks in (37) gives

$$\begin{aligned}
&\varepsilon^{-r(r-1)-2(n-r)(r-1)} \det \boldsymbol{K}_{\varepsilon} = \det(\widetilde{\boldsymbol{\Delta}}^{-1}\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{K}_{\varepsilon}\boldsymbol{Q}_{\text{full}}\widetilde{\boldsymbol{\Delta}}^{-1}) \\
&= \varepsilon^{n-r}((\det \boldsymbol{R})^2 \det \boldsymbol{W} \det(f_{2r-1}\boldsymbol{Q}_{\perp}^{\mathsf{T}}\boldsymbol{D}_{(2r-1)}\boldsymbol{Q}_{\perp}) + \mathcal{O}(\varepsilon)) \\
&= (-1)^r \varepsilon^{n-r}\left(\det \boldsymbol{W} \det \begin{bmatrix} f_{2r-1}\boldsymbol{D}_{(2r-1)} & \boldsymbol{V}_{\le r-1} \\ \boldsymbol{V}_{\le r-1}^{\mathsf{T}} & 0 \end{bmatrix} + \mathcal{O}(\varepsilon)\right),
\end{aligned}$$

where the last equality follows by Lemma 3.11.

2. From (37) it follows that

$$\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{K}_{\varepsilon}\boldsymbol{Q}_{\text{full}} = \begin{bmatrix} \mathcal{O}(1) & \cdots & \mathcal{O}(\varepsilon^{r-1}) & \mathcal{O}(\varepsilon^{r}) & \cdots & \mathcal{O}(\varepsilon^{r}) \\ \vdots & & \vdots & \vdots & & \vdots \\ \mathcal{O}(\varepsilon^{r-1}) & \cdots & \mathcal{O}(\varepsilon^{2(r-1)}) & \mathcal{O}(\varepsilon^{2r-1}) & \cdots & \mathcal{O}(\varepsilon^{2r-1}) \\ \mathcal{O}(\varepsilon^{r}) & \cdots & \mathcal{O}(\varepsilon^{2r-1}) & \mathcal{O}(\varepsilon^{2r-1}) & \cdots & \mathcal{O}(\varepsilon^{2r-1}) \\ \vdots & & \vdots & \vdots & & \vdots \\ \mathcal{O}(\varepsilon^{r}) & \cdots & \mathcal{O}(\varepsilon^{2r-1}) & \mathcal{O}(\varepsilon^{2r-1}) & \cdots & \mathcal{O}(\varepsilon^{2r-1}) \end{bmatrix}, \tag{38}$$

thus Lemma 3.8 implies the orders of the eigenvalues, as in the proof of Theorem 4.1. ∎

Proof of Theorem 4.6. 1. Note that $K \in \mathcal{C}^{(r,r)}$, hence for $s \le r$ we can proceed as in Theorem 4.2, taking into account the fact that the orders of the eigenvalues are given by Theorem 4.5. Therefore, (23) holds true for $1 \le s \le n$.

2. For $k = n-1$, from (38), we have that

$$(\boldsymbol{Q}_{s:k})^{\mathsf{T}}\boldsymbol{K}_{\varepsilon}\boldsymbol{Q}_{s:k} = \begin{cases} \mathcal{O}(\varepsilon^{2s}), & s < r, \\ \mathcal{O}(\varepsilon^{2r-1}), & s \ge r. \end{cases} \tag{39}$$

Then the statement follows from Lemma 3.10, as in the proof of Corollary 4.3.

3. Now let us consider the case $s > r$. In this case, we have

$$\begin{aligned}
&\varepsilon^{-n(2r-1)+r^2} e_s(\boldsymbol{K}_{\varepsilon}) = \varepsilon^{-n(2r-1)+r^2} \sum_{\substack{\mathcal{Y}\subset\{1,\ldots,n\} \\ |\mathcal{Y}|=s}} \det(\boldsymbol{K}_{\varepsilon,\mathcal{Y}}) \\
&= \det \boldsymbol{W}_{\le r-1} \sum_{|\mathcal{Y}|=s} \det(f_{2r-1}\boldsymbol{Q}_{\perp,\mathcal{Y}}^{\mathsf{T}}\boldsymbol{D}_{(2r-1),\mathcal{Y}}\boldsymbol{Q}_{\perp,\mathcal{Y}}) \det(\boldsymbol{V}_{\le r-1,\mathcal{Y}}^{\mathsf{T}}\boldsymbol{V}_{\le r-1,\mathcal{Y}}) + \mathcal{O}(\varepsilon) \\
&= \det \boldsymbol{W}_{\le r-1} \sum_{|\mathcal{Y}|=s} [t^r]\left\{\det(f_{2r-1}\boldsymbol{D}_{(2r-1),\mathcal{Y}} + t\boldsymbol{V}_{\le r-1,\mathcal{Y}}\boldsymbol{V}_{\le r-1,\mathcal{Y}}^{\mathsf{T}})\right\} + \mathcal{O}(\varepsilon) \\
&= \det \boldsymbol{W}_{\le r-1}[t^r]\left\{e_s(f_{2r-1}\boldsymbol{D}_{(2r-1)} + t\boldsymbol{V}_{\le r-1}\boldsymbol{V}_{\le r-1}^{\mathsf{T}})\right\} + \mathcal{O}(\varepsilon) \\
&= \det \boldsymbol{W}_{\le r-1} \det(\boldsymbol{V}_{\le r-1}^{\mathsf{T}}\boldsymbol{V}_{\le r-1}) e_{s-r}(f_{2r-1}\boldsymbol{Q}_{\perp}^{\mathsf{T}}\boldsymbol{D}_{(2r-1)}\boldsymbol{Q}_{\perp}) + \mathcal{O}(\varepsilon),
\end{aligned}$$

where the individual steps follow from Theorem 4.5 and Lemma 3.12. Therefore, by Lemma 3.5 we get that for all $1 \le j \le n-r$

$$\widetilde{e}_j(\widetilde{\lambda}_{r+1},\ldots,\widetilde{\lambda}_n) = e_j(f_{2r-1}\boldsymbol{Q}_{\perp}^{\mathsf{T}}\boldsymbol{D}_{(2r-1)}\boldsymbol{Q}_{\perp}),$$

which together with Remark 3.2 completes the proof. ∎

5. Multidimensional case: Preliminary facts and notations. The multidimensional case requires introducing heavier notation, which we review in this section.

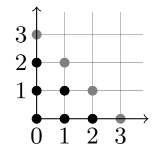

FIG. 7. *Sets of multi-indices, $d = 2$. Black dots: $\mathbb{P}_2$; gray dots: $\mathbb{H}_3$.*

5.1. Multi-indices and sets. For a multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}_+^d$, denote

$$\boldsymbol{\alpha}! \stackrel{\text{def}}{=} \alpha_1! \cdots \alpha_d!, \qquad |\boldsymbol{\alpha}| \stackrel{\text{def}}{=} \sum_{k=1}^{d} \alpha_k,$$

where $0! = 1$ by convention. For example, $|(2,1,3)| = 6$ and $(2,1,3)! = 12$.

We will frequently use the following notations,

$$\mathbb{P}_k \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in \mathbb{Z}_+^d : |\boldsymbol{\alpha}| \leq k\}, \qquad \mathbb{H}_k \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in \mathbb{Z}_+^d : |\boldsymbol{\alpha}| = k\} = \mathbb{P}_k \setminus \mathbb{P}_{k-1}.$$

The cardinalities of these sets are given by the following well-known formulas:

$$\#\mathbb{P}_k = \binom{k+d}{d}, \qquad \#\mathbb{H}_k = \binom{k+d-1}{d-1} = \#\mathbb{P}_k - \#\mathbb{P}_{k-1},$$

and will be used throughout this paper.

Example 1. For $d = 1$, we have $\mathbb{H}_k = \{k\}$ and $\mathbb{P}_k = \{0, 1, \ldots, k\}$. For $d = 2$, an example is shown in Figure 7.

An important class of multi-index sets is the lower sets. An $\mathcal{A} \subset \mathbb{Z}_+^d$ is called a *lower set* [17] if for any $\boldsymbol{\alpha} \in \mathcal{A}$ all "lower" multi-indices are also in the set, i.e.,

$$\boldsymbol{\alpha} \in \mathcal{A},\ \boldsymbol{\beta} \leq \boldsymbol{\alpha} \ \Rightarrow\ \boldsymbol{\beta} \in \mathcal{A},$$

$$\text{where} \quad (\beta_1, \ldots, \beta_d) \leq (\alpha_1, \ldots, \alpha_d) \iff \beta_1 \leq \alpha_1, \ldots, \beta_d \leq \alpha_d.$$

Note that all $\mathbb{P}_k$ are lower sets.

5.2. Monomials and orderings. For a vector of variables $\boldsymbol{x} = \begin{bmatrix} x_1 & \cdots & x_d \end{bmatrix}^{\mathsf{T}}$, the monomial $\boldsymbol{x}^{\boldsymbol{\alpha}}$ is defined as

$$\boldsymbol{x}^{\boldsymbol{\alpha}} \stackrel{\text{def}}{=} x_1^{\alpha_1} \cdots x_d^{\alpha_d}.$$

Remark 5.1. Note that $|\boldsymbol{\alpha}|$ is the total degree of the monomial $\boldsymbol{x}^{\boldsymbol{\alpha}}$. The sets of multi-indices $\mathbb{P}_k$ and $\mathbb{H}_k$ therefore correspond to the sets of monomials of degree $\leq k$ and $= k$, respectively:

$$\{\boldsymbol{x}^{\boldsymbol{\alpha}} : |\boldsymbol{\alpha}| \leq k\}, \qquad \{\boldsymbol{x}^{\boldsymbol{\alpha}} : |\boldsymbol{\alpha}| = k\}.$$

In what follows, we assume that an ordering of multi-indices, i.e., all the elements in $\mathbb{Z}_+$ are linearly ordered, i.e. the relation $\prec$ is defined for all pairs of multi-indices. For example, an ordering for $d = 2$ is given by

$$(0,0) \prec (1,0) \prec (0,1) \prec (2,0) \prec (1,1) \prec (0,2) \prec (3,0) \prec (2,1) \prec (1,2) \prec \cdots . \tag{40}$$

In this paper, the ordering will not be important, as the results will not depend on the ordering. The only requirement is that the order is graded [8, Chapter 2, section 2], i.e.,

$$|\boldsymbol{\alpha}| < |\boldsymbol{\beta}| \quad \Rightarrow \quad \boldsymbol{\alpha} \prec \boldsymbol{\beta}.$$

Remark 5.2. For convenience, in the case $d \geq 2$, we can use an ordering satisfying

$$(1, 0 \ldots, 0) \prec (0, 1 \ldots, 0) \prec \cdots \prec (0, 0, \ldots, 1),$$

such that the matrix $\boldsymbol{V}_1$ defined in the next subsection is equal to

$$\boldsymbol{V}_1 = \begin{bmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_n \end{bmatrix}^{\mathsf{T}}.$$

This is not the case for graded lexicographic or reverse lexicographic [8, Chapter 2, section 2] orderings. Instead, a graded reflected lexicographic order [1] can be used (see (40)).

5.3. Multivariate Vandermonde matrices. Next, for an ordered set of points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and set of multi-indices $\mathcal{A} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m\} \subset \mathbb{Z}_+^d$ ordered according to the chosen ordering, we define the multivariate Vandermonde matrix as

$$\boldsymbol{V}_{\mathcal{A}} = \boldsymbol{V}_{\mathcal{A}}(\mathcal{X}) = [(\boldsymbol{x}_i)^{\boldsymbol{\alpha}_j}]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}.$$

We will introduce a special notation $\boldsymbol{V}_{\leq k} \stackrel{\text{def}}{=} \boldsymbol{V}_{\mathbb{P}_k}$ and $\boldsymbol{V}_k \stackrel{\text{def}}{=} \boldsymbol{V}_{\mathbb{H}_k}$. Since the ordering is graded, the matrix $\boldsymbol{V}_{\leq k}$ can be split into blocks $\boldsymbol{V}_k$ arranged by increasing degree:

$$\boldsymbol{V}_{\leq k} = \begin{bmatrix} \boldsymbol{V}_0 & \boldsymbol{V}_1 & \cdots & \boldsymbol{V}_k \end{bmatrix}. \tag{41}$$

It is easy to see that in the case $d = 1$ the definition coincides with the previous definition of the Vandermonde matrix (7).

An example of the Vandermonde matrix for $d = 2$ and the set of points

$$\mathcal{X} = \{\left[\begin{smallmatrix} y_1 \\ z_1 \end{smallmatrix}\right], \left[\begin{smallmatrix} y_2 \\ z_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} y_3 \\ z_3 \end{smallmatrix}\right]\}$$

with the ordering as in (40) is given below:

$$\boldsymbol{V}_{\leq 2} = \left[\begin{array}{c|cc|ccc} 1 & y_1 & z_1 & y_1^2 & y_1 z_1 & z_1^2 \\ 1 & y_2 & z_2 & y_2^2 & y_2 z_2 & z_2^2 \\ 1 & y_3 & z_3 & y_3^2 & y_3 z_3 & z_3^2 \end{array}\right].$$

A special case is when the Vandermonde matrix is square, i.e., the number of monomials of degree $\leq k$ is equal to the number of points:

$$n = \binom{k+d}{d} = \#\mathbb{P}_k. \tag{42}$$

For example, $n = k + 1$ if $d = 1$ and $n = \frac{(k+2)(k+1)}{2}$ if $d = 2$.

Remark 5.3. Unlike the 1D case, even if all the points are different, the Vandermonde matrix $\boldsymbol{V}_{\leq k}$ is not necessarily invertible. For example, take the set of points on one of the axes

$$\mathcal{X} = \{\left[\begin{smallmatrix} -1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]\}$$

for which the Vandermonde matrix is rank deficient:

$$\boldsymbol{V}_{\leq 1} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

This effect is well known in approximation theory [17]. If the square Vandermonde matrix is nonsingular, then the set of points $\mathcal{X}$ is called unisolvent. It is known [30, Prop. 4] that a general configuration of points (e.g., $\mathcal{X}$ are drawn from an absolutely continuous probability distribution with respect to the Lebesgue measure), is unisolvent almost surely.

5.4. Kernels and smoothness classes. For a function $f : \mathcal{U} \to \mathbb{R}$, $\mathcal{U} \in \mathbb{R}^d$, and a multi-index $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{Z}_+^d$ we use a shorthand notation for its partial derivatives (if they exist):

$$f^{(\boldsymbol{\alpha})}(\boldsymbol{x}) = \frac{\partial f^{|\boldsymbol{\alpha}|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}(\boldsymbol{x}).$$

It makes sense to define the smoothness classes with respect to lower sets. For a lower set $\mathcal{A} \subset \mathbb{Z}_+^d$ we define the class of functions $\mathcal{U} \to \mathbb{R}$ which have on $\mathcal{U}$ all continuous derivatives $f^{(\boldsymbol{\alpha})}$, $\alpha \in \mathcal{A}$. This class is denoted by $\mathcal{C}^{\mathcal{A}}(\mathcal{U})$.

We will consider kernels $K : \Omega \times \Omega \to \mathbb{R}$ in the class $\mathcal{C}^{(k,k)}(\Omega) \stackrel{\text{def}}{=} \mathcal{C}^{\mathbb{P}_k \times \mathbb{P}_k}(\Omega)$, i.e., which has all partial derivatives up to order k for $\boldsymbol{x}$ and $\boldsymbol{y}$ separately.

Next, assume that we are given a kernel $K \in \mathcal{C}^{\mathcal{A} \times \mathcal{B}}(\Omega)$ for lower sets $\mathcal{A}$ and $\mathcal{B}$. We will define the Wronskian matrix for this function as

$$\boldsymbol{W}_{\mathcal{A},\mathcal{B}} = \left[\frac{K^{(\boldsymbol{\alpha},\boldsymbol{\beta})}(\boldsymbol{0},\boldsymbol{0})}{\boldsymbol{\alpha}!\boldsymbol{\beta}!}\right]_{\boldsymbol{\alpha}\in\mathcal{A},\boldsymbol{\beta}\in\mathcal{B}}, \tag{43}$$

where the rows and columns are indexed by multi-indices in $\mathcal{A}$ and $\mathcal{B}$, according to the chosen ordering.

As a special case, we will denote $\boldsymbol{W}_{\leq k} = \boldsymbol{W}_{\mathbb{P}_k} \stackrel{\text{def}}{=} \boldsymbol{W}_{\mathbb{P}_k,\mathbb{P}_k}$. For example, for $d = 2$ and $k = 2$ (Example in Figure 7), and ordering (40) we have $\boldsymbol{W}_{\leq 2} =$

$$\begin{bmatrix}
K^{((0,0),(0,0))} & K^{((0,0),(1,0))} & K^{((0,0),(0,1))} & \frac{K^{((0,0),(2,0))}}{2} & K^{((0,0),(1,1))} & \frac{K^{((0,0),(0,2))}}{2} \\
K^{((1,0),(0,0))} & K^{((1,0),(1,0))} & K^{((1,0),(0,1))} & \frac{K^{((1,0),(2,0))}}{2} & K^{((1,0),(1,1))} & \frac{K^{((1,0),(0,2))}}{2} \\
K^{((0,1),(0,0))} & K^{((0,1),(1,0))} & K^{((0,1),(0,1))} & \frac{K^{((0,1),(2,0))}}{2} & K^{((0,1),(1,1))} & \frac{K^{((0,1),(0,2))}}{2} \\
\frac{K^{((2,0),(0,0))}}{2} & \frac{K^{((2,0),(1,0))}}{2} & \frac{K^{((2,0),(0,1))}}{2} & \frac{K^{((2,0),(2,0))}}{4} & \frac{K^{((2,0),(1,1))}}{2} & \frac{K^{((2,0),(0,2))}}{4} \\
K^{((1,1),(0,0))} & K^{((1,1),(1,0))} & K^{((1,1),(0,1))} & \frac{K^{((1,1),(2,0))}}{2} & K^{((1,1),(1,1))} & \frac{K^{((1,1),(0,2))}}{2} \\
\frac{K^{((0,2),(0,0))}}{2} & \frac{K^{((0,2),(1,0))}}{2} & \frac{K^{((0,2),(0,1))}}{2} & \frac{K^{((0,2),(2,0))}}{4} & \frac{K^{((0,2),(1,1))}}{2} & \frac{K^{((0,2),(0,2))}}{4}
\end{bmatrix},$$

where we omit the arguments of $K^{(\boldsymbol{\alpha},\boldsymbol{\beta})}$. We will also need block-antidiagonal matrices $\boldsymbol{W}_{\diagup s} \in \mathbb{R}^{\#\mathbb{P}_s \times \#\mathbb{P}_s}$ defined as follows:

$$\boldsymbol{W}_{\diagup s} = \begin{bmatrix} & & \boldsymbol{W}_{\mathbb{H}_0,\mathbb{H}_s} \\ & \iddots & \\ \boldsymbol{W}_{\mathbb{H}_s,\mathbb{H}_0} & & \end{bmatrix}, \tag{44}$$

where $\boldsymbol{W}_{\mathcal{A},\mathcal{B}}$ are blocks of the Wronskian matrix defined in (43). For example

$$\boldsymbol{W}_{\diagup 0} = \begin{bmatrix} W_{0,0} \end{bmatrix}, \quad \boldsymbol{W}_{\diagup 1} = \begin{bmatrix} & \boldsymbol{W}_{\mathbb{H}_0,\mathbb{H}_1} \\ \boldsymbol{W}_{\mathbb{H}_1,\mathbb{H}_0} & \end{bmatrix},$$

and in general $\boldsymbol{W}_{\diagup s}$ contains the main block antidiagonal of $\boldsymbol{W}_{\leq s}$.

5.5. Taylor expansions. The standard Taylor expansion (at $\mathbf{0}$, i.e., Maclaurin expansion) in the multivariate case is as follows [45, section 8.4.4]. Let $f \in \mathcal{C}^{k+1}(\Omega)$, where Ω is an open neighborhood of $\mathbf{0}$ containing a line segment from 0 to $\boldsymbol{x}$, denoted as $[\mathbf{0}, \boldsymbol{x}]$. Then the following the following Taylor expansion holds:

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{P}_k} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} f^{(\boldsymbol{\alpha})}(\mathbf{0}) + r_k(\boldsymbol{x}), \tag{45}$$

where the remainder can be expressed in the Lagrange or integral forms:

$$r_k(\boldsymbol{x}) = \int_0^1 (k+1)(1-t)^k \Big(\sum_{\boldsymbol{\beta} \in \mathbb{H}_{k+1}} \frac{\boldsymbol{x}^{\boldsymbol{\beta}}}{\boldsymbol{\beta}!} f^{(\boldsymbol{\beta})}(t\boldsymbol{x}) \Big) dt = \sum_{\boldsymbol{\beta} \in \mathbb{H}_{k+1}} \frac{\boldsymbol{x}^{\boldsymbol{\beta}}}{\boldsymbol{\beta}!} f^{(\boldsymbol{\beta})}(\theta \boldsymbol{x})$$

with $\theta \in [0,1]$. A more general Taylor (Maclaurin) expansion has a remainder in the Peano form, and requires smoothness of order one less, i.e., if $f \in \mathcal{C}^k(\Omega)$, we have

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{P}_k} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} f^{(\boldsymbol{\alpha})}(\mathbf{0}) + o(\|\boldsymbol{x}\|_2^k). \tag{46}$$

We also need a "bivariate" version for a function $f : \Omega \times \Omega \to \mathbb{R}$ (the arguments are split into two groups) such that $f \in \mathcal{C}^{\mathbb{P}_{k_1+1} \times \mathbb{P}_{k_2+1}}(\Omega \times \Omega)$. Then we can take $\boldsymbol{x}, \boldsymbol{y}$ such that $[0, \boldsymbol{x}], [0, \boldsymbol{y}] \subset \Omega$ and apply the same steps as in the proof of Theorem 4.1 to get

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y}) = &\sum_{\boldsymbol{\alpha} \in \mathbb{P}_{k_1}, \boldsymbol{\beta} \in \mathbb{P}_{k_2}} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\beta}}}{\boldsymbol{\alpha}! \boldsymbol{\beta}!} f^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}(\mathbf{0}, \mathbf{0}) + \sum_{\boldsymbol{\alpha} \in \mathbb{P}_{k_1}, \boldsymbol{\beta} \in \mathbb{H}_{k_2+1}} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\beta}}}{\boldsymbol{\alpha}! \boldsymbol{\beta}!} f^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}(\mathbf{0}, \theta_{\boldsymbol{y}, \boldsymbol{\alpha}} \boldsymbol{y}) \\ + &\sum_{\boldsymbol{\alpha} \in \mathbb{H}_{k_1+1}, \boldsymbol{\beta} \in \mathbb{P}_{k_2}} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\beta}}}{\boldsymbol{\alpha}! \boldsymbol{\beta}!} f^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}(\eta_{\boldsymbol{x}, \boldsymbol{\beta}} \boldsymbol{x}, \mathbf{0}) + \sum_{\boldsymbol{\alpha} \in \mathbb{H}_{k_1+1}, \boldsymbol{\beta} \in \mathbb{H}_{k_2+1}} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\beta}}}{\boldsymbol{\alpha}! \boldsymbol{\beta}!} f^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}(\zeta_{\boldsymbol{x}, \boldsymbol{y}} \boldsymbol{x}, \xi_{\boldsymbol{x}, \boldsymbol{y}} \boldsymbol{y}), \end{aligned}$$

where $\{\eta_{\boldsymbol{x}, \boldsymbol{\beta}}\}_{\boldsymbol{\beta} \in \mathbb{P}_{k_2}} \subset [0,1]$ depend on $\boldsymbol{x}$, $\{\theta_{\boldsymbol{y}, \boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathbb{P}_{k_1}} \subset [0,1]$ depend on $\boldsymbol{y}$, and $\zeta_{x,y}, \xi_{x,y} \in [0,1]$ depend on both $\boldsymbol{x}$ and $\boldsymbol{y}$.

5.6. Distance matrices and expansions of radial kernels. Next, we consider the radial kernel (10) with order of smoothness r. For even k, as in the univariate case, we will need an expansion in a form similar to (12), which was obtained in the univariate case via the binomial expansion (8). Although we can also use the same approach in the multivariate case, we prefer to derive this expansion directly from Taylor's formula. Let $K \in \mathcal{C}^{2r-2}(\Omega \times \Omega)$ (not necessarily radial). Then the Taylor expansion in Peano's form (46) yields an expansion in ε,

$$K(\varepsilon \boldsymbol{x}, \varepsilon \boldsymbol{y}) = \sum_{k=0}^{2r-2} \varepsilon^k \sum_{\substack{\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{P}_k \\ |\boldsymbol{\alpha}| + |\boldsymbol{\beta}| = k}} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\beta}}}{\boldsymbol{\alpha}! \boldsymbol{\beta}!} K^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}(\mathbf{0}, \mathbf{0}) + o(\varepsilon^{2r-2}),$$

which in matrix form can be written as

$$\boldsymbol{K}_{\varepsilon} = \sum_{k=0}^{2r-2} \varepsilon^k \boldsymbol{V}_{\le k} \boldsymbol{W}_{\diagup k} \boldsymbol{V}_{\le k}^{\mathsf{T}} + o(\varepsilon^{2r-2}). \tag{47}$$

For a radial kernel, the two expansions (47) and (11) coincide on $[0, \varepsilon_0]$; therefore, the distance matrices of even order $\boldsymbol{D}_{(2\ell)}$ have a compact expression as

$$f_{2\ell}\boldsymbol{D}_{(2\ell)} = \boldsymbol{V}_{\leq 2\ell}\boldsymbol{W}_{\diagup 2\ell}\boldsymbol{V}_{\leq 2\ell}^{\mathsf{T}}$$

and, moreover, the expansion of $\boldsymbol{K}_\varepsilon$ given in (12) is also valid in the multivariate case.[5]

Remark 5.4. For k odd, the matrices $\boldsymbol{D}_{(k)}$ in the multivariate case also have the conditional positive-definiteness property (as in Lemma 2.1), except that the number of points should be $n > \#\mathbb{P}_{r-1}$.

6. Results in the multivariate case.

6.1. Determinants in the smooth case. For a degree k, we will introduce a notation for the sum of all total degrees of monomials with degrees in $\mathbb{P}_k$:

$$M = M(k,d) \stackrel{\text{def}}{=} \sum_{\boldsymbol{\alpha}\in\mathbb{P}_k} |\boldsymbol{\alpha}| \quad \text{such that} \quad \prod_{\boldsymbol{\alpha}\in\mathbb{P}_k} \varepsilon^{|\boldsymbol{\alpha}|} = \varepsilon^{M(k,d)},$$

which is given by[6]

$$M(k,d) = d\binom{k+d}{d+1}.$$

For example, if $d = 1$, then $M(k,1) = \binom{k+2}{2}$. With this notation, we can formulate the result on determinants in the multivariate case.

THEOREM 6.1. *Assume that the kernel is in $\mathcal{C}^{(k+1,k+1)}$, the scaled kernel matrix is defined by* (9) *and* (1), *and also*

$$\#\mathbb{P}_{k-1} < n \leq \#\mathbb{P}_k.$$

1. *If $n = \#\mathbb{P}_k$, then*

$$\det \boldsymbol{K}_\varepsilon = \varepsilon^{2M(k,d)}(\det \boldsymbol{W}_{\leq k}(\det(\boldsymbol{V}_{\leq k}))^2 + \mathcal{O}(\varepsilon)).$$

2. *If $n < \#\mathbb{P}_k$, for $\ell = \#\mathbb{P}_k - n$, we have*

$$\det \boldsymbol{K}_\varepsilon = \varepsilon^{2(M(k,d)-k\ell)}(\det(\boldsymbol{Y}\boldsymbol{W}_{\leq k}\boldsymbol{Y}^{\mathsf{T}})\det(\boldsymbol{V}_{\leq k-1}^{\mathsf{T}}\boldsymbol{V}_{\leq k-1}) + \mathcal{O}(\varepsilon)),$$

where $\boldsymbol{Y} \in \mathbb{R}^{n\times\#\mathbb{P}_k}$ is defined as

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{I}_{\#\mathbb{P}_{k-1}} & \\ & \boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{V}_k \end{bmatrix},$$

$\boldsymbol{V}_k$ is the Vandermonde matrix (41) *for monomials of degree k, and $\boldsymbol{Q}_\perp \in \mathbb{R}^{n\times\ell}$ comes from the full QR decomposition of $\boldsymbol{V} = \boldsymbol{V}_{\leq k-1}$ (see* (22)*).*

3. *If $\boldsymbol{K}_\varepsilon$ is positive semidefinite on $[0,\varepsilon_0]$, the eigenvalues split into $k+1$ groups*

$$\underbrace{\widetilde{\lambda}_{0,0}}_{\mathcal{O}(1)}, \underbrace{\{\widetilde{\lambda}_{1,j}\}_{j=1}^{d}}_{\mathcal{O}(\varepsilon^2)}, \ldots, \underbrace{\{\widetilde{\lambda}_{s,j}\}_{j=1}^{\#\mathbb{H}_s}}_{\mathcal{O}(\varepsilon^{2s})}, \ldots, \underbrace{\{\widetilde{\lambda}_{k,j}\}_{j=1}^{\#\mathbb{H}_k-\ell}}_{\mathcal{O}(\varepsilon^{2k})}. \tag{48}$$

[5]Note that there is an equivalent way of obtaining the expansion of the distance matrices in terms of monomials, simply by writing $\|\boldsymbol{x}-\boldsymbol{y}\|^{2p} = (\sum_{i=1}^{d}(x_i - 2y_ix_i + y_i))^p$ and expanding.

[6]See [27, Eqs. (3.19)–(3.20)], where $M(k,d)$ is given in a slightly different form.

The proof of Theorem 6.1 is postponed to subsection 6.3.

From Theorem 6.1, we can also get a result on eigenvalues and eigenvectors. For this, we partition the $\boldsymbol{Q}_{\text{full}}$ matrix in the full QR factorization of $\boldsymbol{V}_{\leq k-1}$ as

$$\boldsymbol{Q}_{\text{full}} = \begin{bmatrix} \boldsymbol{Q}_0 & \boldsymbol{Q}_1 & \dots & \boldsymbol{Q}_k \end{bmatrix} \tag{49}$$

with $\boldsymbol{Q}_s \in \mathbb{R}^{n \times \#\mathbb{H}_s}$, $0 \leq s < k$, and $\boldsymbol{Q}_k \in \mathbb{R}^{n \times (\#\mathbb{H}_k - \ell)}$.

THEOREM 6.2. *Let K be as in Theorem 6.1, such that $\boldsymbol{K}_\varepsilon$ is symmetric positive semidefinite on $[0, \varepsilon_0]$ and analytic in ε in a neighborhood of 0. Then the eigenvalues in the groups have the form*

$$\lambda_{s,j} = \varepsilon^{2s}(\widetilde{\lambda}_{s,j} + \mathcal{O}(\varepsilon)),$$

where $\widetilde{\lambda}_{0,0} = nK^{(0,0)}$ and the other main terms are given as follows.

1. *For $1 \leq s < k$, if $\det \boldsymbol{W}_{\leq s-1} \neq 0$ and $\boldsymbol{V}_{\leq s-1}$ is full rank, then*

$$\widetilde{\lambda}_{s,1} \cdots \widetilde{\lambda}_{s,\#\mathbb{H}_s} = \frac{\det(\boldsymbol{V}_{\leq s}^{\mathsf{T}} \boldsymbol{V}_{\leq s}) \det(\boldsymbol{W}_{\leq s})}{\det(\boldsymbol{V}_{\leq s-1}^{\mathsf{T}} \boldsymbol{V}_{\leq s-1}) \det(\boldsymbol{W}_{\leq s-1})}. \tag{50}$$

2. *For any $1 \leq s \leq k$, if $\det \boldsymbol{W}_{\leq s-1} \neq 0$ and $\boldsymbol{V}_{\leq s-1}$ is full rank, the main terms $\widetilde{\lambda}_{s,1}, \dots, \widetilde{\lambda}_{s,\#\mathbb{H}_s}$ (or $\widetilde{\lambda}_{k,1}, \dots, \widetilde{\lambda}_{k,\#\mathbb{H}_k - \ell}$ if $s = k$) are the eigenvalues of*

$$\boldsymbol{Q}_s^{\mathsf{T}} \boldsymbol{V}_s \widetilde{\boldsymbol{W}} \boldsymbol{V}_s^{\mathsf{T}} \boldsymbol{Q}_s, \tag{51}$$

where $\widetilde{\boldsymbol{W}} = \boldsymbol{W}_{\lrcorner} - \boldsymbol{W}_{\llcorner}(\boldsymbol{W}_{\leq s-1})^{-1}\boldsymbol{W}_{\urcorner}$ is the Schur complement coming from the following partition of the Wronskian:

$$\boldsymbol{W}_{\leq s} = \begin{bmatrix} \boldsymbol{W}_{\leq s-1} & \boldsymbol{W}_{\urcorner} \\ \boldsymbol{W}_{\llcorner} & \boldsymbol{W}_{\lrcorner} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W}_{\leq s-1} & \boldsymbol{W}_{\urcorner,s} \\ \boldsymbol{W}_{\llcorner,s} & \boldsymbol{W}_{\lrcorner,s} \end{bmatrix}.$$

3. *For $0 \leq s < k$, if $\det \boldsymbol{W}_{\leq s} \neq 0$ and $\boldsymbol{V}_{\leq s}$ is full rank, then the limiting eigenvectors from the sth group $\boldsymbol{p}_{s,1}, \dots, \boldsymbol{p}_{s,\#\mathbb{H}_s}$ span the column space of $\boldsymbol{Q}_s$. Moreover, if $\det \boldsymbol{W}_{\leq k-1} \neq 0$, the remaining eigenvectors span the column space of $\boldsymbol{Q}_k$.*

The proof of Theorem 6.2 is postponed to subsection 6.4.

Theorem 6.2 does not give information on the precise location of limiting eigenvectors in each group. We formulate the following conjecture, which we validated numerically.

CONJECTURE 1. *For $1 \leq s \leq k$, if $\det \boldsymbol{W}_{\leq s} \neq 0$ and $\boldsymbol{V}_{\leq s}$ is full rank, the limiting eigenvectors in the $\mathcal{O}(\varepsilon^{2s})$ group are the columns of $\boldsymbol{Q}_s \boldsymbol{U}_s$, where $\boldsymbol{U}_s$ contains the eigenvectors*[7] *of the matrix $\boldsymbol{Q}_s^{\mathsf{T}} \boldsymbol{V}_s \widetilde{\boldsymbol{W}} \boldsymbol{V}_s^{\mathsf{T}} \boldsymbol{Q}_s$ from (51).*

6.2. Finite smoothness case. We prove a generalization of Theorem 4.5 to the multivariate case.

THEOREM 6.3. *For small ε and a radial kernel (10) with order of smoothness r,*

1. *the determinant of $\boldsymbol{K}_\varepsilon$ in the case $n = \#\mathbb{P}_{r-1} + N$ given in (1) has the expansion*

$$\det(\boldsymbol{K}_\varepsilon) = \varepsilon^{2M(r-1,d)+(2r-1)N} \left(\widetilde{k} + \mathcal{O}(\varepsilon)\right),$$

[7]There is a usual issue of ambiguous definition of $\boldsymbol{U}_s$ if the matrix has repeating eigenvalues.

where $\widetilde{k}$ *has exactly the same expression as in* (26) *or* (27) *(with* $\boldsymbol{V}_{\leq r}$ *and* $\boldsymbol{D}_{(2r-1)}$ *replaced with their multivariate counterparts);*

2. *if* $\boldsymbol{K}_\varepsilon$ *is positive semidefinite on* $[0, \varepsilon_0]$*, the eigenvalues are split into* $r+1$ *groups*

$$\underbrace{\widetilde{\lambda}_{0,0}}_{\mathcal{O}(1)}, \underbrace{\{\widetilde{\lambda}_{1,j}\}_{j=1}^{d}}_{\mathcal{O}(\varepsilon^2)}, \ldots, \underbrace{\{\widetilde{\lambda}_{r-1,j}\}_{j=1}^{\#\mathbb{H}_r-1}}_{\mathcal{O}(\varepsilon^{2(r-1)})}, \underbrace{\{\widetilde{\lambda}_{k,j}\}_{j=1}^{N}}_{\mathcal{O}(\varepsilon^{2r-1})};$$

3. *in the analytic in the* ε *case, the main terms for the first* r *groups are the same as in Theorem* 6.2*. For the last group, if* $\det \boldsymbol{W}_{\leq r-1} \neq 0$ *and* $\boldsymbol{V}_{\leq r-1}$ *is full rank, the main terms are the eigenvalues of*

$$f_{2r-1}(\boldsymbol{Q}_\perp^\top \boldsymbol{D}_{(2r-1)} \boldsymbol{Q}_\perp),$$

where $\boldsymbol{Q}_\perp$ *comes from the full QR factorization* (22) *of* $\boldsymbol{V}_{\leq r-1}$*;*

4. *for* $0 \leq s < r$*, the subspace spanned by the limiting eigenvectors for the* $\mathcal{O}(\varepsilon^{2s})$ *group of eigenvalues are as in Theorem* 6.2*. If* $\det \boldsymbol{W}_{\leq r-1} \neq 0$ *and* $\boldsymbol{V}_{\leq r-1}$ *is full rank, the eigenvectors for the last group of* $\mathcal{O}(\varepsilon^{2r-1})$ *eigenvalues span the column space of* $\boldsymbol{Q}_\perp$*.*

The proof of Theorem 6.3 is postponed to subsection 6.4. Note that we obtain a stronger result on the precise location of the last group of eigenvectors in section 8.

6.3. Determinants in the smooth case.

Before proving Theorem 6.1, we again need a technical lemma, which is an analogue of (4.7).

LEMMA 6.4. *Let* $\boldsymbol{R}$ *be an upper-block-triangular matrix*

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{1,1} & \boldsymbol{R}_{1,k} & \cdots & \boldsymbol{R}_{1,k+1} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{R}_{k,k+1} \\ 0 & \cdots & 0 & \boldsymbol{R}_{k+1,k+1} \end{bmatrix},$$

where the blocks $\boldsymbol{R}_{i,j} \in \mathbb{R}^{M_i \times N_j}$ *are not necessarily square. Then it holds that*

$$\begin{bmatrix} I_{M_1} & & & \\ & \varepsilon^{-1} I_{M_2} & & \\ & & \ddots & \\ & & & \varepsilon^{-k} I_{M_{k+1}} \end{bmatrix} \boldsymbol{R} \begin{bmatrix} I_{N_1} & & & \\ & \varepsilon I_{N_2} & & \\ & & \ddots & \\ & & & \varepsilon^{k} I_{N_{k+1}} \end{bmatrix} = \operatorname{blkdiag}(\boldsymbol{R}) + \mathcal{O}(\varepsilon),$$

where $\operatorname{blkdiag}(\boldsymbol{R})$ *is just the block-diagonal part of* $\boldsymbol{R}$*:*

$$\operatorname{blkdiag}(\boldsymbol{R}) = \begin{bmatrix} \boldsymbol{R}_{1,1} & & \\ & \ddots & \\ & & \boldsymbol{R}_{k+1,k+1} \end{bmatrix}.$$

Proof. The proof is analogous to that of Lemma 4.7. ∎

Proof of Theorem 6.1. First, we fix a degree-compatible ordering of multi-indices

$$\mathbb{P}_k = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{\#\mathbb{P}_k}\},$$

and denote the $\#\mathbb{P}_k \times \#\mathbb{P}_k$ matrix (note that $\#\mathbb{P}_k = n + \ell$)

$$\boldsymbol{\Delta} = \boldsymbol{\Delta}_k(\varepsilon) \stackrel{\text{def}}{=} \begin{bmatrix} \varepsilon^{|\boldsymbol{\alpha}_1|} & & \\ & \ddots & \\ & & \varepsilon^{|\boldsymbol{\alpha}_{\#\mathbb{P}_k}|} \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & \varepsilon I_{\#\mathbb{H}_1} & & \\ & & \ddots & \\ & & & \varepsilon^k I_{\#\mathbb{H}_k} \end{bmatrix}, \tag{52}$$

and by $\boldsymbol{E}_n$ the principal $n \times n$ submatrix of $\boldsymbol{\Delta}$. Note that their determinants are

$$\det(\boldsymbol{\Delta}) = \varepsilon^{M(k,d)} \text{ and } \det(\boldsymbol{E}_n) = \varepsilon^{M(k,d)-\ell k}. \tag{53}$$

From the bimultivariate Taylor expansion, as in the proof of Theorem 4.1, we get

$$\begin{aligned} &K(\varepsilon\boldsymbol{x}, \varepsilon\boldsymbol{y}) = [(\varepsilon\boldsymbol{x})^{\boldsymbol{\alpha}_1}, \ldots, (\varepsilon\boldsymbol{x})^{\boldsymbol{\alpha}_{n+\ell}}]\boldsymbol{W}[(\varepsilon\boldsymbol{y})^{\boldsymbol{\alpha}_1}, \ldots, (\varepsilon\boldsymbol{y})^{\boldsymbol{\alpha}_{n+\ell}}]^{\mathsf{T}} \\ &+ \varepsilon^{k+1}[(\varepsilon\boldsymbol{x})^{\boldsymbol{\alpha}_1}, \ldots, (\varepsilon\boldsymbol{x})^{\boldsymbol{\alpha}_{n+\ell}}]\boldsymbol{w}_{1,\boldsymbol{y}}(\varepsilon) + \varepsilon^{k+1}\boldsymbol{w}_{2,\boldsymbol{x}}(\varepsilon)^{\mathsf{T}}[(\varepsilon\boldsymbol{y})^{\boldsymbol{\alpha}_1}, \ldots, (\varepsilon\boldsymbol{y})^{\boldsymbol{\alpha}_{n+\ell}}]^{\mathsf{T}} \\ &+ \varepsilon^{2(k+1)}w_{3,\boldsymbol{x},\boldsymbol{y}}(\varepsilon), \end{aligned} \tag{54}$$

where $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ are bounded (and continuous) $[0, \varepsilon_0] \to \mathbb{R}^n$ vector functions depending on $\boldsymbol{y}$ and $\boldsymbol{x}$, respectively, and w_3 is a bounded (and continuous) function $[0, \varepsilon_0] \to \mathbb{R}$.

Let $\varepsilon_0 > 0$, such that $\{\varepsilon_0\boldsymbol{x}_1, \ldots, \varepsilon_0\boldsymbol{x}_n\} \in \Omega$ for all i. From (54), the scaled kernel matrix admits for $0 \le \varepsilon \le \varepsilon_0$ the expansion

$$\boldsymbol{K}_\varepsilon = \boldsymbol{V}_{\le k}\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{V}_{\le k}^{\mathsf{T}} + \varepsilon^n(\boldsymbol{V}_{\le k}\boldsymbol{\Delta}\boldsymbol{W}_1(\varepsilon) + \boldsymbol{W}_2(\varepsilon)\boldsymbol{\Delta}\boldsymbol{V}_{\le k}^{\mathsf{T}}) + \varepsilon^{2n}\boldsymbol{W}_3(\varepsilon), \tag{55}$$

where $\boldsymbol{W}_1(\varepsilon), \boldsymbol{W}_2(\varepsilon), \boldsymbol{W}_3(\varepsilon) = \mathcal{O}(1)$, $\boldsymbol{W} = \boldsymbol{W}_{\le k}$, $\boldsymbol{V} = \boldsymbol{V}_{\le k}$.

Next, we take the full QR decomposition $\boldsymbol{V} = \boldsymbol{Q}_{\text{full}}\boldsymbol{R}_{\text{full}}$ of $\boldsymbol{V} = \boldsymbol{V}_{\le k-1}$, partitioned as in (22), so that $\boldsymbol{R} \in \mathbb{R}^{\#\mathbb{P}_{k-1} \times \#\mathbb{P}_{k-1}}$ and $\boldsymbol{Q}_\perp \in \mathbb{R}^{n \times (n - \#\mathbb{P}_{k-1})}$. Note that

$$\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{V}_{\le k} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{V}_k \\ \boldsymbol{O} & \boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{V}_k \end{bmatrix},$$

and by Lemma 6.4 we have

$$\boldsymbol{E}_n^{-1}\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{V}_{\le k}\boldsymbol{\Delta} = \underbrace{\begin{bmatrix} \text{blkdiag}(\boldsymbol{R}) & \\ & \boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{V}_k \end{bmatrix}}_{\widetilde{\boldsymbol{R}}} + \mathcal{O}(\varepsilon).$$

Next, we pre-/post-multiply (55) by $\boldsymbol{E}_n^{-1}\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}$ and its transpose, to get (as in (33))

$$\boldsymbol{E}_n^{-1}\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_{\text{full}}\boldsymbol{E}_n^{-1} = \widetilde{\boldsymbol{R}}\boldsymbol{W}\widetilde{\boldsymbol{R}}^{\mathsf{T}} + \mathcal{O}(\varepsilon). \tag{56}$$

Finally, we prove the statements of the theorem.

1. From (53) we have that

$$\varepsilon^{-2(M(k,d)-\ell k)} \det \boldsymbol{K}_\varepsilon = \det(\widetilde{\boldsymbol{R}}\boldsymbol{W}\widetilde{\boldsymbol{R}}^{\mathsf{T}}).$$

Thus, if $n = \#\mathbb{P}_k$, then we have

$$\det(\widetilde{\boldsymbol{R}}\boldsymbol{W}\widetilde{\boldsymbol{R}}^{\mathsf{T}}) = (\det(\widetilde{\boldsymbol{R}}))^2 \det(\boldsymbol{W}) = \det(\boldsymbol{V}_{\le k})^2 \det(\boldsymbol{W}),$$

where the last equality follows from the fact that

$$\det(\widetilde{\boldsymbol{R}}) = \det(\text{blkdiag}(\boldsymbol{R})) \det(\boldsymbol{Q}_\perp^{\mathsf{T}}\boldsymbol{V}_k) = \det(\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{V}_{\le k}) = \det(\boldsymbol{V}_{\le k}),$$

because $\boldsymbol{Q}_{\text{full}}^{\mathsf{T}}\boldsymbol{V}_{\le k}$ is block triangular.

2. For $n < \#\mathbb{P}_k$, we note that

$$\widetilde{\boldsymbol{R}} = \begin{bmatrix} \operatorname{blkdiag} \boldsymbol{R} & \\ & \boldsymbol{I}_\ell \end{bmatrix} \boldsymbol{Y},$$

hence

$$\det(\widetilde{\boldsymbol{R}}\boldsymbol{W}\widetilde{\boldsymbol{R}}^\mathsf{T}) = (\det(\boldsymbol{R}))^2 \det(\boldsymbol{Y}\boldsymbol{W}\boldsymbol{Y}^\mathsf{T}) = \det(\boldsymbol{V}_{\le k-1}^\mathsf{T}\boldsymbol{V}_{\le k-1}) \det(\boldsymbol{Y}\boldsymbol{W}\boldsymbol{Y}^\mathsf{T}).$$

3. Finally, as in the proof of Theorem 4.1, (56) implies that (35) holds as well in the multivariate case; this, together with Theorem 3.7, completes the proof. ∎

6.4. Individual eigenvalues, eigenvectors, and finite smoothness.

Proof of Theorem 6.2. 1–2. Choose a subset $\mathcal{Y}$ of $\mathcal{X}$ of size m, $\#\mathbb{P}_{s-1} < m \le \#\mathbb{P}_s$, $s \le k$. Then we have that

$$\boldsymbol{K}_{\varepsilon,\mathcal{Y}} = \varepsilon^{2(M-s(\#\mathbb{P}_s-m))}(\det(\boldsymbol{Y}\boldsymbol{W}_{\le s}\boldsymbol{Y}^\mathsf{T})\det(\boldsymbol{V}_{\le s-1,\mathcal{Y}}^\mathsf{T}\boldsymbol{V}_{\le s-1,\mathcal{Y}}) + \mathcal{O}(\varepsilon)),$$

and

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{I}_{\#\mathbb{P}_{s-1}} & \\ & \boldsymbol{R}_{s,\mathcal{Y}} \end{bmatrix},$$

where $\boldsymbol{R}_{s,\mathcal{Y}} = \boldsymbol{Q}_{\perp,\mathcal{Y}}^\mathsf{T}\boldsymbol{V}_s$. In particular

$$\begin{aligned}\det(\boldsymbol{Y}\boldsymbol{W}_{\le s}\boldsymbol{Y}^\mathsf{T}) &= \det \begin{bmatrix} \boldsymbol{W}_{\le s-1} & \boldsymbol{W}_{\neg}\boldsymbol{R}_{s,\mathcal{Y}}^\mathsf{T} \\ \boldsymbol{R}_{s,\mathcal{Y}}\boldsymbol{W}_{\llcorner} & \boldsymbol{R}_{s,\mathcal{Y}}\boldsymbol{W}_{\lrcorner}\boldsymbol{R}_{s,\mathcal{Y}}^\mathsf{T} \end{bmatrix} \\ &= \det \boldsymbol{W}_{\le s-1} \det(\boldsymbol{R}_{s,\mathcal{Y}}\boldsymbol{W}_{\lrcorner}\boldsymbol{R}_{s,\mathcal{Y}}^\mathsf{T} - \boldsymbol{R}_{s,\mathcal{Y}}\boldsymbol{W}_{\llcorner}(\boldsymbol{W}_{\le s-1})^{-1}\boldsymbol{W}_{\neg}\boldsymbol{R}_{s,\mathcal{Y}}^\mathsf{T}) \\ &= \det \boldsymbol{W}_{\le s-1} \det(\boldsymbol{Q}_{\perp,\mathcal{Y}}^\mathsf{T}\boldsymbol{V}_s\widetilde{\boldsymbol{W}}\boldsymbol{V}_s^\mathsf{T}\boldsymbol{Q}_{\perp,\mathcal{Y}}),\end{aligned}$$

hence by Lemma 3.11

$$\begin{aligned}&\det(\boldsymbol{Y}\boldsymbol{W}_{\le s}\boldsymbol{Y}^\mathsf{T})\det(\boldsymbol{V}_{\le s-1,\mathcal{Y}}^\mathsf{T}\boldsymbol{V}_{\le s-1,\mathcal{Y}}) \\ &= \det \boldsymbol{W}_{\le s-1}[\gamma^{\#\mathbb{P}_{s-1}}]\left\{\det(\boldsymbol{V}_{s,\mathcal{Y}}\widetilde{\boldsymbol{W}}\boldsymbol{V}_{s,\mathcal{Y}}^\mathsf{T} + \gamma\boldsymbol{V}_{\le s-1,\mathcal{Y}}\boldsymbol{V}_{\le s-1,\mathcal{Y}}^\mathsf{T})\right\}\end{aligned}$$

and therefore

$$\begin{aligned}\widetilde{e}_m &= \det \boldsymbol{W}_{\le s-1}[\gamma^{\#\mathbb{P}_{s-1}}]\left\{e_m(\boldsymbol{V}_s\widetilde{\boldsymbol{W}}\boldsymbol{V}_s^\mathsf{T} + \gamma\boldsymbol{V}_{\le s-1}\boldsymbol{V}_{\le s-1}^\mathsf{T})\right\} \\ &= \det \boldsymbol{W}_{\le s-1}\det(\boldsymbol{V}_{\le s-1}^\mathsf{T}\boldsymbol{V}_{\le s-1})e_{m-\#\mathbb{P}_{s-1}}(\boldsymbol{Q}_{s:k}^\mathsf{T}\boldsymbol{V}_s\widetilde{\boldsymbol{W}}\boldsymbol{V}_s^\mathsf{T}\boldsymbol{Q}_{s:k}) \\ &= \det \boldsymbol{W}_{\le s-1}\det(\boldsymbol{V}_{\le s-1}^\mathsf{T}\boldsymbol{V}_{\le s-1})e_{m-\#\mathbb{P}_{s-1}}(\boldsymbol{Q}_{s}^\mathsf{T}\boldsymbol{V}_s\widetilde{\boldsymbol{W}}\boldsymbol{V}_s^\mathsf{T}\boldsymbol{Q}_{s}),\end{aligned} \tag{57}$$

where $\boldsymbol{Q}_{s:k} \stackrel{\text{def}}{=} \begin{bmatrix}\boldsymbol{Q}_s & \cdots & \boldsymbol{Q}_k\end{bmatrix}$, the penultimate equality follows from Lemma 3.11, and the last equality follows from the fact that only the top block of $\boldsymbol{Q}_{s:k}^\mathsf{T}\boldsymbol{V}_s$ is nonzero. The rest of the proof follows from Lemma 3.5 to (57).

3. We repeat the same steps as in the proof of Corollary 4.3 (for groups of limiting eigenvectors). Since, from the proof of Theorem 6.1, for any $0 \le \ell \le s$, the order of $(\boldsymbol{Q}_{\ell+1:k})^\mathsf{T}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_{\ell+1:k}$ is $\mathcal{O}(\varepsilon^{2(\ell+1)})$ (as in (35)), and the order of eigenvalues in $\mathcal{O}(1), \dots, \mathcal{O}(\varepsilon^{2s})$ groups is exact, all the eigenvectors

$$\{\boldsymbol{p}_{\ell,1}, \dots, \boldsymbol{p}_{\ell,\#\mathbb{H}_\ell}\}_{0\le\ell\le s}$$

must be orthogonal to $\operatorname{span}(\boldsymbol{Q}_{\ell+1:k})$, which proves the first part of the statement. If, moreover, $s = k-1$, then the last block of eigenvectors (corresponding to eigenvalues of order $\mathcal{O}(\varepsilon^{2k})$) must be contained in $\operatorname{span}(\boldsymbol{Q}_k)$. ∎

Proof of Theorem 6.3. 1. The proof repeats that of Theorem 4.5 with the following minor modifications (in order to take into account the multivariate case):

- the matrix $\boldsymbol{W}_{\blacktriangleright} \in \mathbb{R}^{\#\mathbb{P}_{2r-2}\times\#\mathbb{P}_{2r-2}}$ is defined as

$$W_{\blacktriangleright,\boldsymbol{\alpha},\boldsymbol{\beta}} = \begin{cases} \frac{K^{(\boldsymbol{\alpha},\boldsymbol{\beta})}(0,0)}{\boldsymbol{\alpha}!\boldsymbol{\beta}!}, & |\boldsymbol{\alpha}|+|\boldsymbol{\beta}| \le 2r, \\ 0, & |\boldsymbol{\alpha}|+|\boldsymbol{\beta}| > 2r; \end{cases}$$

  i.e., in the sum (36) the matrices $\boldsymbol{W}_{\diagup s}$ are defined according to (44);
- the matrix $\boldsymbol{\Delta}_k(\varepsilon)$ is defined in (52), and Lemma 6.4 is used instead of Lemma 4.7;
- the extended diagonal scaling matrix is

$$\widetilde{\boldsymbol{\Delta}} = \begin{bmatrix} \boldsymbol{\Delta}_{r-1} & \\ & \varepsilon^{r-1}\boldsymbol{I}_N \end{bmatrix} \in \mathbb{R}^{n\times n},$$

  where $N = n - \#\mathbb{P}_{r-1}$;
- the last displayed formula in the proof of Theorem 4.5 becomes

$$\begin{aligned} \varepsilon^{-2M(r-1,d)-2N(r-1)} \det \boldsymbol{K}_\varepsilon &= \det(\widetilde{\boldsymbol{\Delta}}^{-1}\boldsymbol{Q}_{\text{full}}^\mathsf{T}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_{\text{full}}\widetilde{\boldsymbol{\Delta}}^{-1}) \\ &= \varepsilon^N((\det \boldsymbol{R})^2 \det \boldsymbol{W} \det(f_{2r-1}\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{D}_{(2r-1)}\boldsymbol{Q}_\perp) + \mathcal{O}(\varepsilon)). \end{aligned}$$

2. Again, as in the proof of Theorem 4.5, the matrix $\boldsymbol{Q}_{s:k}^\mathsf{T}\boldsymbol{K}_\varepsilon\boldsymbol{Q}_{s:k}$ has the order given in (39), hence the orders of the eigenvalues follow from Theorem 3.7.

3. The proof of this statement repeats the proof of Theorem 4.6 without changes.

4. The last statement follows from combining (39) with Theorem 3.7, and proceeding as in the proof of the corresponding statement in Theorem 6.2. ∎

7. Perturbation theory: A summary for Hermitian matrices. This section contains a summary of facts from [22, Chapter II] to deal with analytic perturbations of self-adjoint operators in a finite-dimensional vector space (i.e., Hermitian matrices). Formally, and in keeping with the notation used in [22], we assume that we are given a matrix-valued function depending on ε such that

$$\boldsymbol{T}(\varepsilon) = \boldsymbol{T}^{(0)} + \varepsilon\boldsymbol{T}^{(1)} + \varepsilon^2\boldsymbol{T}^{(2)} + \cdots, \tag{58}$$

where we assume that the matrices $\boldsymbol{T}^{(k)} \in \mathbb{C}^{n\times n}$ are Hermitian. In a neighborhood of 0, $0 \in \mathcal{D}_0 \subset \mathbb{C}$, $\boldsymbol{T}(\varepsilon)$ has $s \le n$ semisimple eigenvalues generically (i.e., except a finite number of exceptional points). For simplicity of presentation,[8] we assume that $s = n$, which is the case if there exists $\varepsilon_1 \in \mathcal{D}_0$ having all distinct eigenvalues. The interesting case (considered in this paper) is when $\varepsilon = 0$ is an exceptional point, i.e., $\boldsymbol{T}(0) = \boldsymbol{T}^{(0)}$ has multiple eigenvalues (e.g., a low-rank matrix with a multiple eigenvalue 0).

7.1. Perturbation of eigenvalues and group eigenprojectors. Since all matrices are Hermitian, by [22, Theorem 1.10, Chapter II] (see Theorem 3.9), the eigenvalues $\lambda_1(\varepsilon), \dots, \lambda_n(\varepsilon)$ and the rank-one projectors on the corresponding eigenspaces $\boldsymbol{P}_1(\varepsilon), \dots, \boldsymbol{P}_n(\varepsilon)$ are holomorphic functions of ε in a neighborhood of 0, $0 \in \mathcal{D} \subset \mathbb{C}$.

[8] We can also consider the general case if needed.

Remark 7.1. If the matrix $\boldsymbol{T}^{(0)}$ has d multiple eigenvalues $\{\mu_k\}_{k=1}^d$ with multiplicities $m_1, \ldots, m_d$, i.e., after proper reordering, at $\varepsilon = 0$,

$$(\lambda_1(0), \ldots, \lambda_n(0)) = (\underbrace{\mu_1, \ldots, \mu_1}_{m_1 \text{ times}}, \underbrace{\mu_2, \ldots, \mu_2}_{m_2 \text{ times}}, \ldots, \underbrace{\mu_d, \ldots, \mu_d}_{m_d \text{ times}}),$$

then the projectors on the invariant subspaces associated with $\mu_1, \ldots, \mu_s$ are sums of the corresponding rank-one projectors on the eigenspaces:

$$\begin{aligned}
\boldsymbol{P}_{\mu,1} &= \boldsymbol{P}_1(0) + \cdots + \boldsymbol{P}_{m_1}(0),\\
\boldsymbol{P}_{\mu,2} &= \boldsymbol{P}_{m_1+1}(0) + \cdots + \boldsymbol{P}_{m_1+m_2}(0),\\
&\vdots\\
\boldsymbol{P}_{\mu,d} &= \boldsymbol{P}_{m_1+\cdots+m_{d-1}+1}(0) + \cdots + \boldsymbol{P}_n(0).
\end{aligned}$$

In this paper, our aim is to obtain a limiting eigenstructure at $\varepsilon = 0$. In the case of multiple eigenvalues, this information cannot be retrieved from the spectral decompositon of $\boldsymbol{T}^{(0)}$ alone (we can only retrieve the group projectors $\boldsymbol{P}_{\mu,j}$ from the spectral decomposition of $\boldsymbol{T}^{(0)}$). In what follows, we look in details at perturbation expansions in order to find individual $\boldsymbol{P}_k(0)$.

As shown in [22, Chapter II], we can consider perturbations of a possibly multiple eigenvalue. Let λ be an eigenvalue of $\boldsymbol{T}^{(0)}$ of multiplicity m, and $\boldsymbol{P}$ is the corresponding orthogonal projector on the m-dimensional eigenspace. The projector on the perturbed m-dimensional invariant subspace is an analytic matrix-valued function

$$\boldsymbol{P}(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \boldsymbol{P}^{(k)}$$

with the coefficients given by

$$\boldsymbol{P}^{(0)} = \boldsymbol{P}, \quad \boldsymbol{P}^{(1)} = -\boldsymbol{P}\boldsymbol{T}^{(1)}\boldsymbol{S} - \boldsymbol{S}\boldsymbol{T}^{(1)}\boldsymbol{P},$$

$$\boldsymbol{P}^{(k)} = -\sum_{p=1}^{k} (-1)^p \sum_{\substack{\nu_1+\cdots+\nu_p=k\\ k_1+\cdots+k_{p+1}=p\\ \nu_j\geq 1, k_j\geq 0}} \boldsymbol{S}^{(k_1)}\boldsymbol{T}^{(\nu_1)}\boldsymbol{S}^{(k_2)}\cdots\boldsymbol{S}^{(k_p)}\boldsymbol{T}^{(\nu_p)}\boldsymbol{S}^{(k_{p+1})}, \tag{59}$$

where $\boldsymbol{S} = (\boldsymbol{T} - \lambda\boldsymbol{I})^\dagger$, $\boldsymbol{S}^{(0)} = -\boldsymbol{P}$, and $\boldsymbol{S}^{(k)} = \boldsymbol{S}^k$.

7.2. Reduction and splitting the groups. In order to find the individual projectors of the eigenspaces corresponding to a multiple λ, and the expansion of the corresponding eigenvalues, the following reduction (or splitting) procedure [22, Chapter II, section 2.3] can be applied, which localizes the matrix to the m-dimensional subspace corresponding to the perturbations of λ.

We first define the eigennilpotent matrix $\boldsymbol{D}(\varepsilon)$ as

$$\boldsymbol{D}(\varepsilon) = (\boldsymbol{T}(\varepsilon) - \lambda\boldsymbol{I})\boldsymbol{P}(\varepsilon) = \boldsymbol{P}(\varepsilon)(\boldsymbol{T}(\varepsilon) - \lambda\boldsymbol{I}) = \boldsymbol{P}(\varepsilon)(\boldsymbol{T}(\varepsilon) - \lambda\boldsymbol{I})\boldsymbol{P}(\varepsilon),$$

which from [22, Chapter II, section 2.2] has an expansion

$$\boldsymbol{D}(\varepsilon) = 0 + \sum_{k=1}^{\infty} \varepsilon^k \widetilde{\boldsymbol{T}}^{(k)},$$

where the expressions for $\widetilde{\boldsymbol{T}}^{(k)}$ are as follows:

$$
\begin{aligned}
\widetilde{\boldsymbol{T}}^{(1)} =& \boldsymbol{P}\boldsymbol{T}^{(1)}\boldsymbol{P},\\
\widetilde{\boldsymbol{T}}^{(2)} =& \boldsymbol{P}\boldsymbol{T}^{(2)}\boldsymbol{P} - \boldsymbol{P}\boldsymbol{T}^{(1)}\boldsymbol{P}\boldsymbol{T}^{(1)}\boldsymbol{S} - \boldsymbol{P}\boldsymbol{T}^{(1)}\boldsymbol{S}\boldsymbol{T}^{(1)}\boldsymbol{P} - \boldsymbol{S}\boldsymbol{T}^{(1)}\boldsymbol{P}\boldsymbol{T}^{(1)}\boldsymbol{P},\\
\widetilde{\boldsymbol{T}}^{(k)} =& -\sum_{p=1}^{k}(-1)^p \sum_{\substack{\nu_1+\cdots+\nu_p=k\\ k_1+\cdots+k_{p+1}=p-1\\ \nu_j\geq 1, k_j\geq 0}} \boldsymbol{S}^{(k_1)}\boldsymbol{T}^{(\nu_1)}\boldsymbol{S}^{(k_2)}\cdots\boldsymbol{S}^{(k_p)}\boldsymbol{T}^{(\nu_p)}\boldsymbol{S}^{(k_{p+1})}. \qquad (60)
\end{aligned}
$$

Remark 7.2. Note that the matrices $\widetilde{\boldsymbol{T}}^{(k)}$ are self-adjoint, which follows from the fact that for real ε the matrices $\boldsymbol{T}(\varepsilon)$ and $\boldsymbol{P}(\varepsilon)$ (and hence $\boldsymbol{D}(\varepsilon)$) are self-adjoint.

Next, we define the matrix $\widetilde{\boldsymbol{T}}(\varepsilon)$ as

$$
\widetilde{\boldsymbol{T}}(\varepsilon) = \frac{1}{\varepsilon}\boldsymbol{D}(\varepsilon) = \sum_{k=0}^{\infty}\varepsilon^k\widetilde{\boldsymbol{T}}^{(k+1)},
$$

such that $\widetilde{\boldsymbol{T}}(0) = \widetilde{\boldsymbol{T}}^{(1)}$. Note that by Remark 7.2, all the matrices $\widetilde{\boldsymbol{T}}^{(k+1)}$ are Hermitian and all the eigenvalues of $\widetilde{\boldsymbol{T}}^{(1)}(\varepsilon)$ are holomorphic functions of ε. The idea of the reduction process is to apply the perturbation theory to the matrix $\widetilde{\boldsymbol{T}}^{(1)}(\varepsilon)$.

Let $\widetilde{\boldsymbol{T}}^{(1)}$ have s eigenvalues $\widetilde{\lambda}_1,\ldots,\widetilde{\lambda}_s$ with multiplicities

$$
\ell_1+\cdots+\ell_s = m,
$$

where we take into account only the m eigenvalues[9] in the subspace spanned by $\boldsymbol{P}(\varepsilon)$.

Then $\widetilde{\lambda}_1,\ldots,\widetilde{\lambda}_s$ determine the splitting of λ in the following way.

LEMMA 7.3 (a summary of [22, Chapter II, section 2.3]). *Let*

$$
\underbrace{\widetilde{\lambda}_{1,1}(\varepsilon),\ldots,\widetilde{\lambda}_{1,\ell_1}(\varepsilon)}_{pert.\ of\ \widetilde{\lambda}_1},\ldots,\underbrace{\widetilde{\lambda}_{s,1}(\varepsilon),\ldots,\widetilde{\lambda}_{s,\ell_s}(\varepsilon)}_{pert.\ of\ \widetilde{\lambda}_s}
$$

be the holomorphic functions for the perturbations of the eigenvalues $\widetilde{\lambda}_1,\ldots,\widetilde{\lambda}_s$ *of* $\widetilde{\boldsymbol{T}}(\varepsilon)$. *Then the holomorphic functions corresponding to perturbations of the eigenvalue* λ *of the original matrix* $\boldsymbol{T}(\varepsilon)$ *are given by*

$$
\{\lambda_1(\varepsilon),\ldots,\lambda_m(\varepsilon)\} = \bigcup_{k=1}^{s}\{\lambda+\varepsilon\widetilde{\lambda}_{k,1}(\varepsilon),\ldots,\lambda+\varepsilon\widetilde{\lambda}_{k,\ell_k}(\varepsilon)\}.
$$

Moreover, the expansions $\widetilde{\boldsymbol{P}}_{k,j}(\varepsilon)$ *of the projectors on the eigenspaces of* $\widetilde{\boldsymbol{T}}(\varepsilon)$ *(corresponding to* $\widetilde{\lambda}_{k,j}(\varepsilon)$*) give the expansions of the projectors on the eigenspaces of* $\boldsymbol{T}(\varepsilon)$ *corresponding to* $\lambda_1(\varepsilon),\ldots,\lambda_m(\varepsilon)$.

Lemma 7.3 can be applied recursively: for each individual eigenvalue $\widetilde{\lambda}_k$ (of multiplicity $\ell_k > 1$) we can consider the corresponding reduced matrix

$$
{}^{2}\widetilde{\boldsymbol{T}}(\varepsilon) = \frac{1}{\varepsilon}(\widetilde{\boldsymbol{T}}^{(1)}(\varepsilon) - \widetilde{\lambda}_k\boldsymbol{I})\widetilde{\boldsymbol{P}}_k(\varepsilon),
$$

[9]The other $n-m$ eigenvalues are 0.

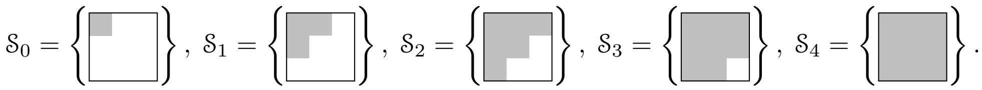

FIG. 8. *Classes of staircase matrices. White color stands for zero blocks.*

where $\widetilde{\boldsymbol{P}}_k(\varepsilon)$ is the perturbation of the total projection on the ℓ_k-dimensional eigenspace corresponding to $\widetilde{\lambda}_k$ (which can be computed as in the previous subsection). Depending on the eigenvalues of the main term of the reduced matrix, either the splitting will occur again, or there will be no splitting; in any case, after a finite number of steps, all the individual eigenvalues will be split into simple (multiplicity 1) eigenvalues.[10]

8. Results on eigenvectors for finitely smooth kernels. In this section we are going to prove the following result for the multivariate case.

THEOREM 8.1. *Let the radial kernel be as in Theorem* 6.3 *with* $\boldsymbol{K}_\varepsilon$ *positive semidefinite on* $[0, \varepsilon_0], \varepsilon_0 > 0$ *and analytic in* ε*. If* $\det \boldsymbol{W}_{\le r-1} \neq 0$*,* $\boldsymbol{V}_{\le r-1}$ *is full rank, and* $\boldsymbol{Q}_\perp^\mathsf{T} \boldsymbol{D}_{(2r-1)} \boldsymbol{Q}_\perp$ *is invertible, then the eigenvectors corresponding to the last group of* $\mathcal{O}(\varepsilon^{2r-1})$ *eigenvalues are given by the columns of*

$$\boldsymbol{Q}_\perp \boldsymbol{U},$$

where the columns of $\boldsymbol{U}$ *are eigenvectors of* $\boldsymbol{Q}_\perp^\mathsf{T} \boldsymbol{D}_{(2r-1)} \boldsymbol{Q}_\perp$*, and the matrix* $\boldsymbol{Q}_\perp$ *comes from the full QR factorization* (22) *of* $\boldsymbol{V}_{\le r-1}$ *(as in Theorem* 6.3*).*

We conjecture that for finitely smooth kernels as well, the individual eigenvectors for the $\mathcal{O}(\varepsilon^{2s})$ groups, $0 \le s \le r-1$, can be obtained as in Conjecture 1.

8.1. Block staircase matrices. We first need some facts about a class of so-called block staircase matrices. Let $\boldsymbol{n} = (n_0, \dots, n_s) \in \mathbb{N}^{s+1}$ such that

$$n = n_0 + \cdots + n_s$$

and consider the following block partition of a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times m}$:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}^{(0,0)} & \cdots & \boldsymbol{A}^{(0,s)} \\ \vdots & & \vdots \\ \boldsymbol{A}^{(s,0)} & \cdots & \boldsymbol{A}^{(s,s)} \end{bmatrix},$$

where the blocks are of size $\boldsymbol{A}^{(k,l)} \in \mathbb{C}^{n \times n}$. Assuming that the partition is fixed, we define the classes $\mathcal{S}_p$ of "staircase" matrices

$$\{\boldsymbol{0}\} = \mathcal{S}_{-1} \subset \mathcal{S}_0 \subset \mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_{2s-1} = \mathcal{S}_{2s} = \mathcal{S}_{2s+1} = \cdots = \mathbb{C}^{n \times n},$$

such that the matrices in $\mathcal{S}_p$ have nonzero blocks only up to the sth antidiagonal

$$\mathcal{S}_p = \{\boldsymbol{A} \in \mathbb{C}^{m \times n} \mid \boldsymbol{A}^{(k,l)} = 0 \text{ for all } k + l > p\}.$$

In Figure 8, we illustrate the classes for $s = 2$.

The following obvious property will be useful.

LEMMA 8.2.
1. $\boldsymbol{A} \in \mathcal{S}_p, \boldsymbol{B} \in \mathcal{S}_q \Rightarrow \boldsymbol{AB} \in \mathcal{S}_{p+q}$.
2. *For any* $\boldsymbol{A} \in \mathcal{S}_p$ *and upper triangular* $\boldsymbol{R}$*, it follows that* $\boldsymbol{RAR}^\mathsf{T} \in \mathcal{S}_p$.
3. *For any* $\boldsymbol{A} \in \mathcal{S}_p$ *and block-diagonal matrix* $\boldsymbol{\Lambda}$ *it holds that* $\boldsymbol{\Lambda A}, \boldsymbol{A\Lambda} \in \mathcal{S}_p$.

Proof. The proof follows from straighforward verification. ∎

[10]This follows from our assumption that the eigenvalues are generically simple.

8.2. Proof of Theorem 8.1.

Proof. The kernel matrix has an expansion with only even powers of ε until $2r-2$

$$\boldsymbol{K}_\varepsilon = \sum_{j=0}^{r-1} \varepsilon^{2j} \boldsymbol{V}_{\leq 2j} \boldsymbol{W}_{\diagup 2j} \boldsymbol{V}_{\leq 2j}^{\mathsf{T}} + \varepsilon^{2r-1} f_{2r-1} \boldsymbol{D}_{(2r-1)} + \mathcal{O}(\varepsilon^{2r}),$$

since odd block diagonals $\boldsymbol{W}_{\diagup 2j-1}$, $j < r$ vanish. We look at the transformed matrix

$$\boldsymbol{T}(\varepsilon) = \boldsymbol{Q}_{\text{full}}^{\mathsf{T}} \boldsymbol{K}_\varepsilon \boldsymbol{Q}_{\text{full}} = \boldsymbol{T}_0 + \varepsilon \boldsymbol{T}_1 + \cdots + \varepsilon^{2r-1} \boldsymbol{T}_{2r-1} + \mathcal{O}(\varepsilon^{2r}),$$

where $\boldsymbol{Q}_{\text{full}}$ is the matrix of the full QR decomposition of $\boldsymbol{V}_{\leq 2r-2}$:

$$\boldsymbol{T}_s = \begin{cases} \boldsymbol{Q}_{\text{full}}^{\mathsf{T}} \boldsymbol{V}_{\leq s} \boldsymbol{W}_{\diagup s} \boldsymbol{V}_{\leq s}^{\mathsf{T}} \boldsymbol{Q}_{\text{full}}, & s \text{ even}, s < 2r-1, \\ 0, & s \text{ odd}, s < 2r-1, \\ f_{2r-1} \boldsymbol{Q}_{\text{full}}^{\mathsf{T}} \boldsymbol{D}_{(2r-1)} \boldsymbol{Q}_{\text{full}}, & s = 2r-1. \end{cases}$$

Due to the fact that that $\boldsymbol{W}_{\diagup s}$ is block antidiagonal, $\boldsymbol{Q}_{\text{full}}^{\mathsf{T}} \boldsymbol{V}_{\leq s}$ is upper triangular and by Lemma 8.2, we have that $\boldsymbol{T}_s$ is block staircase for $s < 2r-1$, i.e., $\boldsymbol{T}_{\diagup s} \in \mathcal{S}_s$.

We proceed by a series of Kato's reductions, according to Lemma 7.3. At each order of ε, a multiple eigenvalue 0 is split into a group of nonzero eigenvalues and eigenvalue 0 of smaller multiplicity. Formally, we consider a sequence of reduced matrices

$${}^{s+1}\widetilde{\boldsymbol{T}}(\varepsilon) = \frac{1}{\varepsilon} {}^{s}\widetilde{\boldsymbol{T}}(\varepsilon) \boldsymbol{P}_s(\varepsilon), \quad {}^{0}\widetilde{\boldsymbol{T}}(\varepsilon) = \boldsymbol{T}(\varepsilon),$$

where $\boldsymbol{P}_k(\varepsilon)$ is the projector onto the perturbation of the nullspace of ${}^{s}\widetilde{\boldsymbol{T}}(0)$ (i.e., the invariant subspace associated with the eigenvalue 0). Its power series expansion

$${}^{s}\widetilde{\boldsymbol{T}}(\varepsilon) = \sum_{j=s}^{\infty} \varepsilon^{j-s} \cdot {}^{s}\widetilde{\boldsymbol{T}}_j$$

can be computed according to (60). For each matrix we will be interested only in the first $2r-s$ terms, as summarized below,

$$\begin{array}{ccccc} 1 & \varepsilon & \varepsilon^2 & \cdots & \varepsilon^{2r-1} \\ \hline \boldsymbol{T}_0 & & & & \\ \boldsymbol{T}_1 & \widetilde{\boldsymbol{T}}_1 & & & \\ \boldsymbol{T}_2 & \widetilde{\boldsymbol{T}}_2 & {}^{2}\widetilde{\boldsymbol{T}}_2 & & \\ \vdots & & & \ddots & \\ \boldsymbol{T}_{2r-1} & \widetilde{\boldsymbol{T}}_{2r-1} & {}^{2}\widetilde{\boldsymbol{T}}_{2r-1} & & {}^{2r-1}\widetilde{\boldsymbol{T}}_{2r-1} \end{array},$$

since we are interested only in the terms

$$\boldsymbol{T}_0, \widetilde{\boldsymbol{T}}_1, {}^{2}\widetilde{\boldsymbol{T}}_2, \ldots, {}^{2k}\widetilde{\boldsymbol{T}}_{2k},$$

whose eigenvectors give limiting eigenvectors for the original matrix $\boldsymbol{T}(\varepsilon)$.

Next, we will look in detail at the form of coefficients of the reduced matrices. The projector on the image space of $\boldsymbol{T}_0$ is $\boldsymbol{\Pi}_0 = \boldsymbol{e}_0 \boldsymbol{e}_0^{\mathsf{T}}$ (where $\boldsymbol{e}_0 = (1, 0, \ldots, 0)^{\mathsf{T}}$), the projector on the nullspace is

$$\boldsymbol{P}_0 = \begin{bmatrix} 0 & \\ & \boldsymbol{I}_{n-1} \end{bmatrix},$$

and the matrix $\boldsymbol{S} = \boldsymbol{S}_0$ in (59). By examining the terms in (59), we have that the coefficients of the reduced matrices preserve the staircase class, i.e., $\widetilde{\boldsymbol{T}}_s \in \mathcal{S}_s$. This can be seen by verifying that if $\boldsymbol{A}_s \in \mathcal{S}_s$ and $\boldsymbol{S}^{(j)}$ are diagonal, then the products

$$\boldsymbol{S}^{(k_1)}\boldsymbol{A}^{(\nu_1)}\boldsymbol{S}^{(k_2)}\cdots\boldsymbol{S}^{(k_p)}\boldsymbol{A}^{(\nu_p)}\boldsymbol{S}^{(k_{p+1})} \in \mathcal{S}_s \tag{61}$$

if $\nu_1 + \cdots + \nu_p = s$. Next, we note that since $\boldsymbol{T}_1 \in \mathcal{S}_1$, then we have

$$\widetilde{\boldsymbol{T}}_1 = \boldsymbol{P}_0\boldsymbol{T}_1\boldsymbol{P}_0 = 0.$$

Hence we have that $\boldsymbol{P}_1 = \boldsymbol{I}$ and $\boldsymbol{S}_1 = 0$, and the second step of reduction does not change the matrices, i.e. ${}^{2}\widetilde{\boldsymbol{T}}_s = \widetilde{\boldsymbol{T}}_s$.

Proceeding by induction, at the step of reduction $(2s-2) \to (2s-1)$ the staircase order of the matrices is preserved due to (61), and block diagonality of $\boldsymbol{P}_{2s-1}$ and ${}^{2s\text{-}2}\widetilde{\boldsymbol{T}}_{2s-2}$. Since the reduction step $(2s-1) \to 2s$ does not change anything, we get

$$\boldsymbol{P}_{2s} = \begin{bmatrix} 0_{(\#\mathbb{P}_s)\times(\#\mathbb{P}_s)} & \\ & \boldsymbol{I}_{(n-\#\mathbb{P}_s)\times(n-\#\mathbb{P}_s)} \end{bmatrix}, \tag{62}$$

which we know from Theorem 6.3 and

$${}^{2s}\widetilde{\boldsymbol{T}}_{2s} = \begin{bmatrix} 0_{(\#\mathbb{P}_s)\times(\#\mathbb{P}_s)} & & \\ & *_{(\#\mathbb{H}_{s+1})\times(\#\mathbb{H}_{s+1})} & \\ & & 0_{(n-\#\mathbb{P}_{s+1})\times(n-\#\mathbb{P}_{s+1})} \end{bmatrix}. \tag{63}$$

The last reduction step is different, as we get ${}^{2r\text{-}1}\widetilde{\boldsymbol{T}}_{2r-1}$ which is not equal to zero. In order to obtain ${}^{2r\text{-}1}\widetilde{\boldsymbol{T}}_{2r-1}$, we note the following: at the first step of the reduction the matrices $\widetilde{\boldsymbol{T}}_{2j-1}$ defined by (60), are the sums of the terms running over multi-indices

$$\nu_1 + \cdots + \nu_p = 2j-1,$$

where at least one of ν_ℓ should be odd and all $\nu_\ell \leq 2j-1$. Therefore, we have $\widetilde{\boldsymbol{T}}_{2j-1} = 0$ if $j < r$ and $\widetilde{\boldsymbol{T}}_{2r-1} = \boldsymbol{P}\boldsymbol{T}_{2r-1}\boldsymbol{P}$. Proceeding by induction, we get that

$${}^{s}\widetilde{\boldsymbol{T}}_{2j-1} = \begin{cases} 0, & j < r \text{ and } s < 2r-1, \\ f_{2r-1}\boldsymbol{P}_{2r-2}\boldsymbol{T}_{2r-1}\boldsymbol{P}_{2r-2}, & j = r \text{ and } s = 2r-1, \end{cases}$$

where $\boldsymbol{P}_{2r-2}$ is defined in (62).

Thus we have that the limiting eigenvectors of $\boldsymbol{Q}_{\text{full}}\boldsymbol{T}_\varepsilon\boldsymbol{Q}_{\text{full}}^\top$ for the order $\mathcal{O}(\varepsilon^{2r-1})$ are the limiting eigenvectors (corresponding to nonzero eigenvalues) of the matrix

$$\boldsymbol{Q}_\perp\boldsymbol{Q}_\perp^\top\boldsymbol{D}_{(2r-1)}\boldsymbol{Q}_\perp\boldsymbol{Q}_\perp^\top,$$

where $\boldsymbol{Q}_{\text{full}} = \begin{bmatrix}\boldsymbol{Q} & \boldsymbol{Q}_\perp\end{bmatrix}$ is the splitting of $\boldsymbol{Q}_{\text{full}}$ such that $\boldsymbol{Q}_\perp \in \mathbb{R}^{n\times(n-\#\mathbb{P}_{r-1})}$. ∎

9. Discussion. We have shown that kernel matrices become tractable in the flat limit, and exhibit deep ties to orthogonal polynomials. We would like to add some remarks and highlight some open problems.

First, we expect our analysis to generalize in a mostly straightforward manner to the "continuous" case, i.e., to kernel integral operators. This should make it possible to examine a double asymptotic in which $n \to \infty$ as $\varepsilon \to 0$. One could then leverage recent results on the asymptotics of orthogonal polynomials, for instance, [23].

Second, our results may be used empirically to create preconditioners for kernel matrices. There is already a vast literature on approximate kernel methods, including in the flat limit (e.g., [13, 25]), and future work should examine how effective polynomial preconditioners are compared to other available methods.

Third, many interesting problems (e.g., spectral clustering [39]) involve not the kernel matrix itself but some rescaled variant. We expect that multiplicative perturbation theory could be brought to bear here [36].

Finally, while multivariate polynomials are relatively well-understood objects, our analysis also shows that in the finite smoothness case, a central role is played by a different class of objects, namely, multivariate polynomials are replaced by the eigenvectors of distance matrices of an odd power. To our knowledge, very little has been proved about such objects but some literature from statistical physics [6] points to a link to "Anderson localization." Anderson localization is a well-known phenomenon in physics whereby eigenvectors of certain operators are localized, in the sense of having fast decay over space. This typically does not hold for orthogonal polynomials, which tend rather to be localized in frequency. Thus, we conjecture that eigenvectors of completely smooth kernels are localized in frequency, contrary to eigenvectors of finitely smooth kernels, which (at low energies) are localized in space. The results in [6] are enough to show that this holds for the exponential kernel in $d = 1$, but extending this to $d > 1$ and higher regularity orders is a fascinating and probably nontrivial problem.

Appendix A. Proofs for elementary symmetric polynomials.

Proof of Lemma 3.4.

1. By definition, the elementary symmetric polynomials can be expanded as

$$
\begin{aligned}
e_s(\varepsilon) &= \sum_{1 \le t_1 < \cdots < t_s \le n} \prod_{j=1}^{s} \varepsilon^{L_{t_j}} (\widetilde{\lambda}_{t_j} + \mathcal{O}(\varepsilon)) \\
&= \varepsilon^{L_1 + \cdots + L_s} \Big(\underbrace{\sum_{\substack{1 \le t_1 < \cdots < t_s \le n \\ L_1 + \cdots + L_s = L_{t_1} + \cdots + L_{t_s}}} \prod_{j=1}^{s} \widetilde{\lambda}_{t_j}}_{\widetilde{e}_s} + \mathcal{O}(\varepsilon) \Big),
\end{aligned}
\tag{64}
$$

which follows from the fact that $L_{t_1} + \cdots + L_{t_s}$ is minimized at $t_j = j$.

2. The case $s = n$ is obvious, because there is only one possible tuple $(t_1, \ldots, t_s)$. Consider the case $L_s < L_{s+1}$, $1 \le s < n$. If $t_s > s$, then the sum is increased,

$$L_1 + \cdots + L_s < L_{t_1} + \cdots + L_{t_s},$$

hence $(1, \ldots, s)$ is the only possible choice for $(t_1, \ldots, t_s)$ in (64). ∎

Proof of Lemma 3.5. When (18) is satisfied, we need to find

$$\widetilde{e}_{s+k} = \sum_{\substack{1 \le t_1 < \cdots < t_{s+k} \le n \\ L_1 + \cdots + L_{s+k} = L_{t_1} + \cdots + L_{t_{s+k}}}} \prod_{j=1}^{s+k} \widetilde{\lambda}_{t_j},$$

where, as in the previous case, the minimum sum $L_{t_1} + \cdots + L_{t_{s+k}}$ is achieved by

$$L_1 + \cdots + L_s + L_{s+1} + \cdots + L_{s+k},$$

and the sum increases if $t_s > s$ or if $t_{s+k} > s+m$. Therefore, $(t_1, \ldots, t_s) = (1, \ldots, s)$ and the main term becomes

$$\widetilde{e}_{s+k} = \prod_{i=1}^{s} \widetilde{\lambda}_i \left(\sum_{s+1 \le t_{s+1} < \cdots < t_{s+k} \le s+m} \prod_{j=1}^{k} \widetilde{\lambda}_{t_{s+j}} \right). \qquad \square$$

Appendix B. Proofs for saddle point matrices.

Proof of Lemma 3.11. We note that $(\det \boldsymbol{R})^2 = \det(\boldsymbol{V}^\mathsf{T}\boldsymbol{V})$ and

$$\begin{aligned}\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{V} \\ \boldsymbol{V}^\mathsf{T} & 0 \end{bmatrix} &= \det \left(\begin{bmatrix} \boldsymbol{Q}_{\text{full}}^\mathsf{T} & \\ & \boldsymbol{I}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{V} \\ \boldsymbol{V}^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_{\text{full}} & \\ & \boldsymbol{I}_r \end{bmatrix} \right) \\ &= \det \begin{bmatrix} \boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}^\mathsf{T} & \boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp & \boldsymbol{R} \\ \boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q} & \boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp & \\ \boldsymbol{R}^\mathsf{T} & & 0 \end{bmatrix} = (-1)^r (\det \boldsymbol{R})^2 \det(\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp). \qquad \square\end{aligned}$$

Before proving Lemma 3.12, we need a technical lemma first.

LEMMA B.1. *For* $\boldsymbol{G} \in \mathbb{R}^{r\times r}$ *and* $\boldsymbol{M} = \left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{smallmatrix}\right) \in \mathbb{R}^{n\times n}$, $\boldsymbol{A} \in \mathbb{R}^{r\times r}$, *it holds that*

$$[t^r] \left\{ \det \left(\boldsymbol{M} + t \begin{pmatrix} \boldsymbol{G} & 0 \\ 0 & 0 \end{pmatrix} \right) \right\} = \det \boldsymbol{G} \det \boldsymbol{D}.$$

Proof. From [5, Theorem 1], we have that[11]

$$\frac{1}{r!} \frac{d^r}{dt^r} \det \left(\boldsymbol{M} + t \begin{pmatrix} \boldsymbol{G} & 0 \\ 0 & 0 \end{pmatrix} \right) = \det \begin{pmatrix} \boldsymbol{G} & \boldsymbol{B} \\ 0 & \boldsymbol{D} \end{pmatrix} = \det \boldsymbol{G} \det \boldsymbol{D}. \qquad \square$$

Proof of Lemma 3.12. Due to invariance under similarity transformations of the elementary polynomials, we get

$$\begin{aligned}&[t^r] \left\{ e_k(\boldsymbol{A} + t\boldsymbol{V}\boldsymbol{V}^\mathsf{T}) \right\} = [t^r] \left\{ e_k(\boldsymbol{Q}_{\text{full}}^\mathsf{T}(\boldsymbol{A} + t\boldsymbol{V}\boldsymbol{V}^\mathsf{T})\boldsymbol{Q}_{\text{full}}) \right\} \\ &= [t^r] \left\{ e_k\Big(\underbrace{\begin{bmatrix} \boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q} + t\boldsymbol{R}\boldsymbol{R}^\mathsf{T} & \boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp \\ \boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q} & \boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp \end{bmatrix}}_{\boldsymbol{B}(t)} \Big) \right\} \\ &= \sum_{\substack{|\mathcal{J}|=k \\ \mathcal{J}\subseteq\{1,\ldots,n\}}} [t^r] \left\{ \det\left(\boldsymbol{B}_{\mathcal{J},\mathcal{J}}(t)\right) \right\} = \sum_{\substack{|\mathcal{J}|=k \\ \{1,\ldots,r\}\subseteq\mathcal{J}\subseteq\{1,\ldots,n\}}} [t^r] \left\{ \det\left(\boldsymbol{B}_{\mathcal{J},\mathcal{J}}(t)\right) \right\},\end{aligned}$$

where the last equality follows from the fact that the polynomial $\det\left(\boldsymbol{B}_{\mathcal{J},\mathcal{J}}(t)\right)$ has degree that is equal to the cardinality of the intersection $\mathcal{J} \cap \{1, \ldots, r\}$. Any such $\mathcal{J}$ can be written as $\{1, \ldots, r\} \cup \mathcal{J}'$, where $\mathcal{J}' \subseteq \{r+1, \ldots, n\}$. Applying Lemma B.1 to each term individually, we get

$$[t^r] \left\{ \det\left(\boldsymbol{B}_{\mathcal{J},\mathcal{J}}(t)\right) \right\} = \det(\boldsymbol{R}\boldsymbol{R}^\mathsf{T}) \det\left((\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp)_{\mathcal{J}',\mathcal{J}'} \right),$$

thus, summing over all $\mathcal{J}' \subseteq \{r+1, \ldots, n\}$ yields

$$\det(\boldsymbol{R}\boldsymbol{R}^\mathsf{T}) \sum_{\substack{\mathcal{J}'\subseteq\{r+1,\ldots,n\} \\ |\mathcal{J}'|=k-r}} \det\left((\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp)_{\mathcal{J}',\mathcal{J}'} \right) = \det(\boldsymbol{R}\boldsymbol{R}^\mathsf{T}) e_{k-r}(\boldsymbol{Q}_\perp^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}_\perp). \qquad \square$$

[11] Note that in the sum in [5, Eq. (7)] only one term is nonzero.

Acknowledgments. The authors would like to acknowledge Pierre Comon for his help at the beginning of this project, and Cosme Louart and Malik Tiomoko for their help in improving the presentation of the results. We would also like to thank two anonymous referees for their careful reading of our work and many suggestions.

REFERENCES

[1] OEIS *Orderings Foundation*, https://oeis.org/wiki/Orderings.
[2] B. Baxter, *Norm estimates for inverses of Toeplitz distance matrices*, J. Approx. Theory, 79 (1994), pp. 222–242.
[3] B. Baxter, *Preconditioned conjugate gradients, radial basis functions, and Toeplitz matrices*, Comput. Math. Appl., 43 (2002), pp. 305–318.
[4] M. Benzi and V. Simoncini, *On the eigenvalues of a class of saddle point matrices*, Numer. Math., 103 (2006), pp. 173–196.
[5] R. Bhatia and T. Jain, *Higher order derivatives and perturbation bounds for determinants*, Linear Algebra Appl., 431 (2009), pp. 2102–2108, https://doi.org/10.1016/j.laa.2009.07.003.
[6] E. Bogomolny, O. Bohigas, and C. Schmit, *Spectral properties of distance matrices*, J. Phys. A, 36 (2003), 3595.
[7] R. Couillet and F. Benaych-Georges, *Kernel spectral clustering of large dimensional data*, Electron. J. Stat., 10 (2016), pp. 1393–1454.
[8] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 2nd ed., Springer, New York, 1997.
[9] S. De Marchi and R. Schaback, *Nonstandard kernels and their applications*, Dolomites Res. Notes Approx., 2 (2009), pp. 16–43.
[10] T. A. Driscoll and B. Fornberg, *Interpolation in the limit of increasingly flat radial basis functions*, Comput. Math. Appl., 43 (2002), pp. 413–422.
[11] N. El Karoui, *The spectrum of kernel random matrices*, Ann. Statist., 38 (2010), pp. 1–50.
[12] G. E. Fasshauer, *Meshfree approximation methods with MATLAB*, Vol. 6, World Scientific, Singapore, 2007.
[13] B. Fornberg, E. Larsson, and N. Flyer, *Stable computations with Gaussian radial basis functions*, SIAM J. Sci. Comput., 33 (2011), pp. 869–892.
[14] B. Fornberg, G. Wright, and E. Larsson, *Some observations regarding interpolants in the limit of flat radial basis functions*, Comput. Math. Appl., 47 (2004), pp. 37–55.
[15] B. Fornberg and J. Zuev, *The runge phenomenon and spatially variable shape parameters in rbf interpolation*, Comput. Math. Appl., 54 (2007), pp. 379–398.
[16] D. Garreau, W. Jitkrittum, and M. Kanagawa, *Large Sample Analysis of the Median Heuristic*, preprint https://arxiv.org/abs/1707.07269, 2017.
[17] M. Gasca and T. Sauer, *Polynomial interpolation in several variables*, Adv. Comput. Math., 12 (2000), 377.
[18] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, Chelsea, New York, 1984.
[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1990.
[20] I. C. F. Ipsen and R. Rehman, *Perturbation bounds for determinants and characteristic polynomials*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 762–776, https://doi.org/10.1137/070704770.
[21] T. Jain, *Derivatives for antisymmetric tensor powers and perturbation bounds*, Linear Algebra Appl., 435 (2011), pp. 1111–1121, https://doi.org/10.1016/j.laa.2011.02.026.
[22] T. Kato, *Perturbation Theory for Linear Operators*, 2nd corrected ed., Springer, Berlin, 1995.
[23] A. Kroó and D. Lubinsky, *Christoffel functions and universality in the bulk for multivariate orthogonal polynomials*, Canad. J. Math., 65 (2013), pp. 600–620.
[24] E. Larsson and B. Fornberg, *Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions*, Comput. Math. Appl., 49 (2005), pp. 103–130.
[25] S. Le Borne, *Factorization, symmetrization, and truncated transformation of radial basis function-GA stabilized Gaussian radial basis functions*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 517–541.
[26] Y. J. Lee and C. A. Micchelli, *On collocation matrices for interpolation and approximation*, J. Approx. Theory, 174 (2013), pp. 148–181, https://doi.org/10.1016/j.jat.2013.06.007.

[27] Y. J. Lee, C. A. Micchelli, and J. Yoon, *A study on multivariate interpolation by increasingly flat kernel functions*, J. Math. Anal. Appl., 427 (2015), pp. 74–87.
[28] C. A. Micchelli, *Interpolation of scattered data: Distance matrices and conditionally positive definite functions*, Constr. Approx., 2 (1986), pp. 11–22.
[29] M. Miranda and P. Tilli, *Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 867–881.
[30] T. Sauer, *Polynomial interpolation in several variables: Lattices, differences, and ideals*, Stud. Comput. Math., 12 (2006), pp. 191–230.
[31] R. Schaback, *Multivariate interpolation by polynomials and radial basis functions*, Constr. Approx., 21 (2005), pp. 293–317.
[32] R. Schaback, *Limit problems for interpolation by analytic radial basis functions*, J. Comput. Appl. Math., 212 (2008), pp. 127–149.
[33] R. Schaback and H. Wendland, *Kernel techniques: From machine learning to meshless methods*, Acta Numer., 15 (2006), pp. 543–639.
[34] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, 2002.
[35] G. Song, J. Riddle, G. E. Fasshauer, and F. J. Hickernell, *Multivariate interpolation with increasingly flat radial basis functions of finite smoothness*, Adv. Comput. Math., 36 (2012), pp. 485–501.
[36] F. Sosa and J. Moro, *First order asymptotic expansions for eigenvalues of multiplicatively perturbed matrices*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 1478–1504.
[37] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York, 1999.
[38] E. E. Tyrtyshnikov, *A unifying approach to some old and new theorems on distribution and clustering*, Linear Algebra Appl., 232 (1996), pp. 1–43, https://doi.org/10.1016/0024-3795(94)00025-5.
[39] U. Von Luxburg, M. Belkin, and O. Bousquet, *Consistency of spectral clustering*, Ann. Statist., 36 (2008), pp. 555–586.
[40] R. Wang, Y. Li, and E. Darve, *On the numerical rank of radial basis function kernels in high dimensions*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 1810–1835.
[41] A. J. Wathen, *Realistic eigenvalue bounds for Galerkin mass matrix*, IMA J. Numer. Anal., 7 (1987), pp. 449–457.
[42] A. J. Wathen and S. Zhu, *On spectral distribution of kernel matrices related to radial basis functions*, Numer. Algorithms, 70 (2015), pp. 709–726, https://doi.org/10.1007/s11075-015-9970-0.
[43] H. Wendland, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math., Cambridge University Press, Cambridge, 2004, https://doi.org/10.1017/CBO9780511617539.
[44] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, Vol. 2, MIT Press, Cambridge, MA, 2006.
[45] V. A. Zorich, *Mathematical Analysis* I, Springer, Berlin, 2004.