Supplement to “Higher-Order Total Variation Classes on Grids: Minimax Theory and Trend Filtering Methods”

Veeranjaneyulu Sadhanala
Carnegie Mellon University
Pittsburgh, PA 15213
vsadhana@cs.cmu.edu

Yu-Xiang Wang
Carnegie Mellon University/Amazon AI
Pittsburgh, PA 15213/Palo Alto, CA 94303
yuxiangw@amazon.com

James Sharpnack
University of California, Davis
Davis, CA 95616
jsharpna@ucdavis.edu

Ryan J. Tibshirani
Carnegie Mellon University
Pittsburgh, PA 15213
ryantibs@stat.cmu.edu

We provide proofs and additional details for the results in “Higher-Order Total Variation Classes on Grids: Minimax Theory and Trend Filtering Methods”.

A.1 Proof of Lemma 1

The proof is straightforward. Note that the nullity of $\widetilde{\Delta}^{(k+1)}$ is the number of nonzero singular values of $\widetilde{\Delta}^{(k+1)}$, or equivalently, the number of nonzero eigenvalues of $(\widetilde{\Delta}^{(k+1)})^T\widetilde{\Delta}^{(k+1)}$. But, following from (9), and abbreviating $D = D_{\mathrm{1d}}^{(k+1)}$,

$$(\widetilde{\Delta}^{(k+1)})^T\widetilde{\Delta}^{(k+1)} = D^TD \otimes I \otimes \cdots \otimes I \; + \; I \otimes D^TD \otimes \cdots \otimes I \; + \; \ldots \\ + \; I \otimes I \otimes \cdots \otimes D^TD,$$

the Kronecker sum of D^TD with itself, a total of d times. Using a standard fact about Kronecker sums, if $\rho_1, \ldots, \rho_N$ denote the eigenvalues of D^TD then

$$\rho_{i_1} + \rho_{i_2} + \cdots + \rho_{i_d}, \quad i_1, \ldots, i_d \in \{1, \ldots, N\},$$

are the eigenvalues of $(\widetilde{\Delta}^{(k+1)})^T\widetilde{\Delta}^{(k+1)}$. By counting the multiplicity of the zero eigenvalue, we arrive at a nullity for $\widetilde{\Delta}^{(k+1)}$ of $(k+1)^d$. One can now directly check that each of the polynomials specified in the lemma is indeed in the null space (e.g., simply by inspection of the formula in (8)), and that these polynomial functions are linearly independent, which completes the proof. $\square$

A.2 Proof of Lemma 2

Let us define

$$\widetilde{D} = \begin{bmatrix} C_{\mathrm{1d}}^{(k+1)} \\ D_{\mathrm{1d}}^{(k+1)} \end{bmatrix} \in \mathbb{R}^{N \times N},$$

where the first $k+1$ rows are given by a matrix $C^{(k+1)} \in \mathbb{R}^{(k+1)\times N}$ that completes the row space, as in Lemma 2 of Wang et al. [10]. And now, again by Lemma 2 of Wang et al. [10],

$$\big(H_{\mathrm{1d}}^{(k)}\big)^{-1} = \frac{1}{k!}\widetilde{D}, \tag{A.1}$$

where $H_{\mathrm{1d}}^{(k)} \in \mathbb{R}^{N\times N}$ is the falling factorial basis matrix of order k, which has elements

$$\big(H_{\mathrm{1d}}^{(k)}\big)_{ij} = h_j(i/N), \quad i,j = 1, \ldots, N,$$

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

with h_i, $i = 1, \dots, N$ denoting the falling factorial basis functions in (10).

Let us write the KTF problem in (5), (9) explicitly as

$$\min_{\theta \in \mathbb{R}^n} \; \frac{1}{2}\|y-\theta\|_2^2 + \lambda \left\| \begin{bmatrix} D_{1\mathrm{d}}^{(k+1)} \otimes I \otimes \cdots \otimes I \\ I \otimes D_{1\mathrm{d}}^{(k+1)} \otimes \cdots \otimes I \\ \vdots \\ I \otimes I \otimes \cdots \otimes D_{1\mathrm{d}}^{(k+1)} \end{bmatrix} \theta \right\|_1 . \tag{A.2}$$

We now transform variables in this problem by defining $\theta = (H_{1\mathrm{d}}^{(k)} \otimes \cdots \otimes H_{1\mathrm{d}}^{(k)})\alpha$ and using (A.1), which turns (A.2) into an equivalent basis form,

$$\min_{\alpha \in \mathbb{R}^n} \; \frac{1}{2}\Big\|y - \Big(H_{1\mathrm{d}}^{(k)} \otimes \cdots \otimes H_{1\mathrm{d}}^{(k)}\Big)\alpha\Big\|_2^2 + \lambda k! \left\| \begin{bmatrix} I^0 \otimes H_{1\mathrm{d}}^{(k)} \otimes \cdots \otimes H_{1\mathrm{d}}^{(k)} \\ H_{1\mathrm{d}}^{(k)} \otimes I^0 \otimes \cdots \otimes H_{1\mathrm{d}}^{(k)} \\ \vdots \\ H_{1\mathrm{d}}^{(k)} \otimes H_{1\mathrm{d}}^{(k)} \otimes \cdots \otimes I^0 \end{bmatrix} \alpha \right\|, \tag{A.3}$$

where $I^0 = [\, 0_{(N-k-1)\times(k+1)} \;\; I_{(N-k-1)} \,]$.

Interestingly, the penalty in (A.3) is not a pure sparsity penalty on the coefficients α (as it is in basis form in 1d) but a sparsity penalty on aggregated (sums of) coefficients. This makes the penalty a little hard to interpret, but to glean intuition, we can rewrite the problem once more via the transformation

$$f = \sum_{i_1,\dots,i_d=1}^{N} \alpha_{i_1,\dots,i_d} (h_{i_1} \otimes h_{i_2} \otimes \cdots \otimes h_{i_d}), \tag{A.4}$$

where recall we are indexing the components of α by $\alpha_{i_1,\dots,i_d}$, for $i_1, \dots, i_d = 1, \dots, N$ (and the summands above use tensor products of univariate functions). To be concrete, note that the function f defined in (A.4) evaluates to

$$f(x) = \sum_{i_1,\dots,i_d=1}^{N} \alpha_{i_1,\dots,i_d} h_{i_1}(x) h_{i_2}(x_2) \cdots h_{i_d}(x_d), \quad x \in [0,1]^d.$$

Thus we can equivalently write the basis form in (A.3) in functional form

$$\min_{f \in H_d} \; \frac{1}{2} \sum_{x \in Z_d} \big(y(x) - f(x)\big)^2 + \lambda \sum_{j=1}^{d} \sum_{x_{-j} \in Z_{d-1}} \mathrm{TV}\bigg(\frac{\partial^k f(\cdot, x_{-j})}{\partial x_j^k}\bigg), \tag{A.5}$$

where recall $f(\cdot, x_{-j})$ denotes f as function of the jth dimension with all other dimensions fixed at x_{-j}, $\partial^k/\partial x_j^k(\cdot)$ denotes the kth partial weak derivative operator with respect to x_j, for $j = 1, \dots, d$, and $\mathrm{TV}(\cdot)$ denotes the total variation operator. To see the equivalence between the penalty terms in (A.3) and (A.5), it can be directly checked that

$$k!\Big(I^0 \otimes H_{1\mathrm{d}}^{(k)} \otimes \cdots \otimes H_{1\mathrm{d}}^{(k)}\Big)\alpha$$

contains the differences of the function $\partial^k f/\partial x_1^k$ over all pairs of grid positions that are adjacent in the x_1 direction, where f is as in (A.4). This, combined with the fact that $\partial^k f/\partial x_1^k$ is constant in between lattice positions, means that

$$k!\Big\|\Big(I^0 \otimes H_{1\mathrm{d}}^{(k)} \otimes \cdots \otimes H_{1\mathrm{d}}^{(k)}\Big)\alpha\Big\|_1 = \sum_{x_{-1} \in Z_{d-1}} \mathrm{TV}\bigg(\frac{\partial^k f(\cdot, x_{-1})}{\partial x_1^k}\bigg),$$

the total variation of $\partial^k f/\partial x_1^k$ added up over all slices of the lattice Z_d in the x_1 direction. Similar arguments apply to the penalty terms corresponding to dimensions $j = 2, \dots, d$, and this completes the proof. □

A.3 Proof of Theorem 2

For $d = 2$, it is shown in the proof of Corollary 8 in Wang et al. [11] that the GTF operator $\Delta^{(k+1)}$ satisfies the incoherence property, as defined in Theorem 1, for any choice of cutoff $i_0 \geq 1$, and with a constant $\mu = 4$ when k is even and $\mu = 2$ when k is odd. It suffices to upper bound the partial sum term $\sum_{i=i_0+1}^{n-1} \xi_i^{-2}$. Lemma A.1 gives the key calculation, where it is shown that for large enough n and each $i_0 \geq 1$,

$$\sum_{i=i_0+1}^{n-1} \frac{1}{\xi_i^2} \leq c \cdot \begin{cases} n \log(n/i_0) & \text{for } k = 0 \\ n^{k+1} i_0^{-k} & \text{for } k \geq 1, \end{cases}$$

where $c > 0$ is a constant that depends only on k. For $k = 0$, to minimize to the upper bound given in Theorem 1, we want to balance

$$\frac{i_0}{n} \quad \text{with} \quad \frac{\mu}{n} \sqrt{\log n \log(n/i_0)} C_n.$$

This leads us to choose $i_0 \asymp C_n \log n$, and plugging this in gives the result for $k = 0$.

For $k \geq 1$, we want to balance

$$\frac{i_0}{n} \quad \text{with} \quad \frac{\mu}{n} \sqrt{\log n (n/i_0)^k} C_n.$$

This leads us to take $i_0 \asymp n^{k/(k+2)} (\log n)^{1/(k+2)} C_n^{2/(k+2)}$, and plugging this in completes the proof for $k \geq 1$. ☐

A.4 Lemma A.1

The next lemma is the key driver for the sharp rate established in Theorem 2. Here and henceforth, denote $[i] = \{1, \ldots, i\}$ for an integer $i \geq 1$.

Lemma A.1. *Let $\xi_1 \leq \ldots \leq \xi_{n-1}$ be the nonzero singular values of the GTF operator $\Delta^{(k+1)}$ of order $k + 1$. If $k = 0$, then for any $i_0 \in [n-1]$,*

$$\sum_{i=i_0+1}^{n-1} \frac{1}{\xi_i^2} \leq cn \log(n/i_0).$$

for large enough n, where $c > 0$ is an absolute constant. If $k > 1$, then for any $i_0 \in [n-1]$,

$$\sum_{i=i_0+1}^{n-1} \frac{1}{\xi_i^2} \leq cn^{k+1}/i_0^k,$$

for large enough n, where now $c > 0$ is a constant depending only on k.

Proof. In the following, we denote by $c > 0$ a constant whose value may change from line to line, as needed.

Let us denote by $\lambda_1 \leq \ldots \leq \lambda_{n-1}$ the nonzero eigenvalues of the Laplacian of the 2d grid graph of size $N \times N$. As shown in Wang et al. [11], the GTF operator $\Delta^{(k+1)}$ has squared singular values $\xi_i^2 = \lambda_i^{k+1}$, $i \in [n-1]$. We can index the eigenvalues of the Laplacian by 2d grid positions, and we note (as, e.g., in the proof of Corollary 8 in Wang et al. [11]) that they may be written as

$$\lambda_{i_1,i_2} = 4 \sin^2 \Big(\frac{\pi(i_1 - 1)}{2N} \Big) + 4 \sin^2 \Big(\frac{\pi(i_2 - 1)}{2N} \Big), \quad i_1, i_2 \in [N].$$

For the first claim in the lemma, take $j_0 = \lfloor \sqrt{i_0} \rfloor$. Observe, using $\sin(x) \geq x/2$ for $x \in [0, \pi/2]$,

$$\begin{aligned} \sum_{i=i_0+1}^{n-1} \frac{1}{\lambda_i} &\leq \sum_{\min\{i_1,i_2\} \geq j_0+1} \frac{1}{\lambda_{i_1,i_2}} \\ &\leq cn \sum_{\min\{i_1,i_2\} \geq j_0+1} \frac{1}{(i_1-1)^2 + (i_2-1)^2} \end{aligned}$$

$$
\begin{aligned}
&\leq cn \sum_{i_1=j_0}^{N-1} \sum_{i_2=1}^{N-1} \frac{1}{i_1^2 + i_2^2} \\
&\leq cn \sum_{i_1=j_0}^{N-1} \int_0^{N-1} \frac{1}{i_1^2 + x^2}\, dx \\
&= cn \sum_{i_1=j_0}^{N-1} \frac{1}{i_1} \tan^{-1}\left(\frac{N-1}{i_1}\right) \\
&\leq cn \sum_{i_1=j_0}^{N-1} \frac{1}{i_1} \frac{\pi}{2} \\
&\leq cn \log(N/j_0),
\end{aligned}
$$

for sufficiently large n.

As for the second claim in the lemma, observe, again using $\sin(x) \geq x/2$ for $x \in [0, \pi/2]$,

$$
\begin{aligned}
\sum_{i=i_0+1}^{n-1} \frac{1}{\lambda_i^{k+1}} &\leq \sum_{(i_1-1)^2+(i_2-1)^2 \geq i_0}^{n} \frac{1}{\lambda_{i_1,i_2}^{k+1}} \\
&\leq cn^{k+1} \sum_{(i_1-1)^2+(i_2-1)^2 \geq i_0} \frac{1}{((i_1-1)^2 + (i_2-1)^2)^{k+1}} \\
&\leq cn^{k+1} \left(\int_{i_0 \leq x^2+y^2 \leq 2(n-1),\, x,y \geq 0} \frac{1}{(x^2+y^2)^{k+1}}\, dx\, dy + \sum_{(i_1-1)^2+(i_2-1)^2 = i_0} \frac{1}{i_0^{k+1}} \right) \\
&\leq cn^{k+1} \left(\int_0^{\pi/2} \int_{\sqrt{i_0}}^{\sqrt{2(n-1)}} \frac{1}{r^{2(k+1)}} r\, dr\, d\theta + \frac{1}{i_0^{k+1/2}} \right) \\
&\leq cn^{k+1} \left(\frac{\pi}{2} \int_{i_0}^{2(n-1)} \frac{1}{u^{k+1}}\, du + \frac{1}{i_0^{k+1/2}} \right) \\
&= cn^{k+1} \left(\frac{\pi}{2} \left(\frac{1}{i_0^k} - \frac{1}{(2(n-1))^k} \right) + \frac{1}{i_0^{k+1/2}} \right) \\
&\leq cn^{k+1}/i_0^k.
\end{aligned}
$$

□

A.5 Proof of Theorem 3

The KTF operator (9), in the current case that $d = 2$, is simply

$$
\widetilde{\Delta}^{(k+1)} = \begin{bmatrix} D_{1\text{d}}^{(k+1)} \otimes I \\ I \otimes D_{1\text{d}}^{(k+1)} \end{bmatrix}.
$$

Abbreviate $N' = N - k - 1$. Let β_i, u_i, v_i be a triplet of nonzero singular value, left singular vector, and right singular vector of $D_{1\text{d}}^{(k+1)}$, for $i \in [N']$. We seek the left singular values of $\widetilde{\Delta}^{(k+1)}$, i.e, the eigenvectors (corresponding to nonzero eigenvalues) of

$$
\widetilde{\Delta}^{(k+1)} (\widetilde{\Delta}^{(k+1)})^T = \begin{bmatrix} DD^T \otimes I & D \otimes D^T \\ D^T \otimes D & I \otimes DD^T \end{bmatrix},
$$

where we abbreviate $D = D_{1\text{d}}^{(k+1)}$. The vectors

$$
\begin{bmatrix} \beta_i \cdot u_i \otimes v_j \\ \beta_j \cdot v_i \otimes u_j \end{bmatrix}, \quad i, j \in [N'], \tag{A.6}
$$

for $i, j \in [N']$ are $(N')^2$ such eigenvectors. The other eigenvectors are given by

$$\begin{bmatrix} u_i \otimes p_j \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ p_j \otimes u_i \end{bmatrix}, \quad i \in [N'], j \in [k+1], \tag{A.7}$$

where p_j, $j \in [k+1]$ form an orthogonal basis for the null space of $D_{\rm 1d}^{(k+1)}$.

The main technical challenge is in establishing incoherence of the vectors in (A.6), (A.7). We note that, based on the Kronecker product form of these vectors, it suffices to show incoherence of u_i, v_i, $i \in [N']$, and p_i, $i \in [k+1]$. Incoherence of u_i, $i \in [N']$ is established in Lemma A.3 and of v_i, $i \in [N']$ in Lemma A.4, using specialized approximations for eigenvectors of Toeplitz matrices from Bottcher et al. [2]. Incoherence of p_i, $i \in [k+1]$ may be seen by choosing, e.g., these vectors to be the discrete Legendre orthogonal polynomials as in Neuman and Schonbach [6]. We have thus shown that $\widetilde{\Delta}^{(k+1)}$ satisfies the incoherence property, as defined in Theorem 1, for any choice of $i_0 \geq 1$.

Now we address the partial sum term $\sum_{i=i_0+1}^{n-1} \xi_i^{-2}$. Lemma A.2 shows that for large enough n and a constant $c > 0$ depending only on k,

$$\sum_{i=i_0+1}^{n-1} \frac{1}{\xi_i^2} \leq c \cdot \begin{cases} n \log(n/i_0) & \text{for } k = 0 \\ n^{k+1} i_0^{-k} & \text{for } k \geq 1, \end{cases}$$

just as was the case for GTF. (In fact, this result is proved by tying the singular values of the KTF operator to those of the GTF operator.) Repeating the same arguments as in the proof of Theorem 2 gives the desired result. ☐

A.6 Lemma A.2

This lemma proves a result analogous to Lemma A.1, by tying together the singular values of the KTF and GTF operators.

Lemma A.2. *Let $\xi_1 \leq \ldots \leq \xi_{n-(k+1)^2}$ be the nonzero singular values of the KTF operator $\widetilde{\Delta}^{(k+1)}$ of order $k+1$. If $k = 0$, then for any $i_0 \in [n-(k+1)^2 - 1]$,*

$$\sum_{i=i_0+1}^{n-(k+1)^2} \frac{1}{\xi_i^2} \leq cn \log(n/i_0).$$

for large enough n, where $c > 0$ is an absolute constant. If $k > 1$, then for $i_0 \in [n-(k+1)^2 - 1]$,

$$\sum_{i=i_0+1}^{n-(k+1)^2} \frac{1}{\xi_i^2} \leq cn^{k+1}/i_0^k,$$

for large enough n, where now $c > 0$ is a constant depending only on k.

Proof. Abbreviate $D = D_{\rm 1d}^{(k+1)}$, and write G for the GTF operator of order $k+1$ defined over a 1d chain of length N. Also let $N' = N - k - 1$, and $k' = \lfloor (k+1)/2 \rfloor$. Then D is given by removing the first k_1 rows and last k_2 rows of G, i.e.,

$$D = PG, \quad \text{where } P = \begin{bmatrix} 0_{N' \times k'} & I_{N'} & 0_{N' \times k'} \end{bmatrix}.$$

This means

$$DD^T = PGG^TP^T.$$

Let β_i, $i \in [N']$ be the eigenvalues of DD^T, and let α_i, $i \in [N]$ be the eigenvalues of GG^T. The Cauchy interlacing theorem now tells us that

$$\beta_i \geq \alpha_i^{k+1}, \quad i \in [N']. \tag{A.8}$$

This key property will allow us to relate the nonzero singular values of the KTF operator to those of the GTF operator.

The squared nonzero singular values of $\widetilde{\Delta}^{(k+1)}$ are the nonzero eigenvalues of $(\widetilde{\Delta}^{(k+1)})^T \widetilde{\Delta}^{(k+1)}$. We can index the eigenvalues of $(\widetilde{\Delta}^{(k+1)})^T \widetilde{\Delta}^{(k+1)}$ by 2d grid positions, as in

$$\psi_{i_1,i_2} = \rho_{i_1} + \rho_{i_2}, \quad i_1, i_2 \in [N],$$

where ρ_i, $i \in [N]$ denote the eigenvalues of $D^T D$, i.e., $\rho_1 = \cdots = \rho_{k+1} = 0$ and $\rho_{i+k+1} = \beta_i$, $i \in [N']$, where D, β_i, $i \in [N']$ are as above. Also, as in the proof of Lemma A.1, we can write the eigenvalues of the Laplacian matrix of the 2d grid graph as

$$\lambda_{i_1,i_2} = \alpha_{i_1} + \alpha_{i_2}, \quad i_1, i_2 \in [N]$$

where α_i, $i \in [N]$ is as above. For arbitrary $i_1, i_2 \in [N]$, with at least one $i_1 > k+2$ or $i_2 > k+2$, note that

$$\frac{1}{\psi_{i_1,i_2}} = \frac{1}{\beta_{i_1-k-1} + \beta_{i_2-k-1}} \leq \frac{1}{\alpha_{i_1-k-1}^{k+1} + \alpha_{i_2-k-1}^{k+1}} \leq \frac{2^{k+1}}{\lambda_{i_1-k-1,i_2-k-1}^{k+1}},$$

where we use the convention $\beta_{-i} = 0$ and $\alpha_{-i} = 0$ for $i \leq 0$, the first inequality was due to the key property (A.8), and the second was due to the simple fact $(a+b)^k \leq 2^k a^k + 2^k b^k$. The last display shows that to bound the sum of squared reciprocal nonzero singular values of the KTF operator, it suffices to bound the reciprocal kth power of Laplacian eigenvalues, as was the case for the GTF operator. Proceeding as in the proof of Lemma A.1 gives the result. □

A.7 Lemmas A.3 and A.4

The next two lemmas establish incoherence of the left and right singular vectors of $D_{\mathrm{1d}}^{(k+1)}$. They rely heavily on the work on symmetric banded Toeplitz matrices in Bogoya et al. [1].

Lemma A.3. *The left singular vectors $u_1, \cdots, u_{N-k-1}$ of $D_{\mathrm{1d}}^{(k+1)}$ are incoherent, i.e., there exists a constant $\mu > 0$ depending only on k such that*

$$\max_{i \in [N-k-1]} \|u_i\|_\infty \leq \frac{\mu}{\sqrt{N-k-1}}.$$

Proof. For $k = 0$, the result has already been proved in Wang et al. [11]. Assume $k \geq 1$ henceforth. Abbreviate $D = D_{\mathrm{1d}}^{(k+1)}$, and $N' = N - k - 1$. The left singular vectors of D are the eigenvectors of DD^T, which is a symmetric banded Toeplitz matrix with entries

$$(DD^T)_{ij} = t_{|i-j|}, \quad \text{where } t_\ell = (-1)^\ell \binom{2k+2}{k+1+\ell} \; \ell = 0, \ldots, k+1.$$

Let $\beta_1 \leq \ldots \leq \beta_{N'}$ be the eigenvalues of DD^T. Observe that $\beta_{N'} \leq 4^{k+1}$ by the Gershgorin circle theorem. We intend to apply approximation results from [1], but DD^T does not satisfy their simple-loop assumption. Fortunately,

$$\Delta := 4^{k+1} I_{N'} - \mathrm{abs}(DD^T)$$

does and it is sufficient to show the incoherence of eigenvectors of Δ, as by Lemma A.5, it implies incoherence of eigenvectors of DD^T. Δ is a symmetric banded Toeplitz matrix with entries

$$\Delta_{ij} = a_{|i-j|}, \quad \text{where } a_\ell = 4^{k+1} \mathbb{1}(\ell = 0) - \binom{2k+2}{k+1+\ell} \text{ for } \ell = 0, \ldots, k+1.$$

To apply results from [1], we introduce some notation. The symbol of Δ is given by

$$a(t) = \sum_{\ell=-(k+1)}^{k+1} a_\ell t^\ell = 4^{k+1} - \left(\frac{1}{t} + 2 + t\right)^{k+1}.$$

Define $g(\sigma) = a(e^{i\sigma}) = 4^{k+1} - (2 + 2\cos\sigma)^{k+1}$ for $\sigma \in \mathbb{R}$. In the language of [1], a satisfies the simple-loop conditions; in particular, $a \in SL^\alpha$ for any $\alpha \geq 4$.

For an eigenvalue β of Δ, the characteristic polynomial is given by

$$p_\beta(t) = a(t) - \beta \tag{A.9}$$

and its $2k+2$ roots are given by $z_0(\beta), z_1(\beta), z_2(\beta), \ldots, z_k(\beta)$ and their inverses, where $|z_0(\beta)| = 1$, $|z_\kappa(\beta)| > 1$ for $\kappa \in [k]$. The positive Wiener-Hopf function (see Chapter 1 in [3]) corresponding to our symbol is

$$b_+(t, \beta) = (-1)^k \, \Pi_{\kappa=1}^k (t - z_\kappa(\beta)).$$

Let $\tilde{\beta}_j, s_j, j \in [N']$ be the eigenvalues and eigenvectors of Δ. From Theorems 2.5, 4.1 and Lemma 4.2 in [1], s_j can be written as

$$s_j = M_j + L_j + R_j + \epsilon_j \tag{A.10}$$

where for $m \in [N']$, suppressing the argument $\tilde{\beta}_j$,

$$\begin{aligned} M_{j,m} &= \frac{z_0^{\frac{N'-1}{2}-m}}{|b_+(z_0)|} + (-1)^{N'-j} \frac{\bar{z}_0^{\frac{N'-1}{2}-m}}{|b_+(\bar{z}_0)|}, \\ L_{j,m} &= \frac{z_0^{\frac{N'+1}{2}} b_+(z_0)}{|b_+(z_0)|} \sum_{\kappa=1}^{k} \frac{z_\kappa^{-m-1}}{\frac{\partial b_+}{\partial t}(z_\kappa)(z_\kappa - z_0)(z_\kappa - \bar{z}_0)} \\ R_{j,m} &= \overline{L_{j,N'-1-m}}. \end{aligned}$$

and $\|\epsilon_j\|_\infty \leq \frac{c}{N^2}$ where c does not depend on j, N. Write using triangle inequality,

$$\begin{aligned} \frac{s_{jm}}{\|s_j\|} &\leq \frac{|M_{jm}|}{\|s_j\|} + \frac{|s_{jm} - M_{jm}|}{\|s_j\|} \\ &\leq \frac{1/|b_+(z_0)| + 1/|b_+(\bar{z}_0)|}{\|s_j\|} + \frac{|s_{jm} - M_{jm}|}{\|s_j\|} \end{aligned}$$

From Theorem 2.6 [1], we have $\|s_j\| = \sqrt{N'}\left(|b_+(z_0)|^{-2} + |b_+(\bar{z}_0)|^{-2}\right)^{1/2} + O(1)$ and $\frac{|s_{jm} - M_{jm}|}{\|s_j\|} = O(1/\sqrt{N'})$. Therefore, for large enough N, there exists a constant μ independent of $j, \tilde{\beta}_j, N$ such that

$$\frac{\|s_j\|_\infty}{\|s_j\|} \leq \frac{\mu}{\sqrt{N'}}.$$

This completes the proof as the above holds for all $j \in [N']$. □

Lemma A.4. *The right singular vectors v_i, $i \in [N-k-1]$ of $D_{\text{1d}}^{(k+1)}$ are incoherent, i.e., there exists a constant $\mu > 0$ depending only on k such that*

$$\max_{i \in [N-k-1]} \|v_i\|_\infty \leq \frac{\mu}{\sqrt{n}}.$$

Proof. As in the the proof of Lemma A.3, abbreviate $D = D_{\text{1d}}^{(k+1)}$, and $N' = N - k - 1$. Let the eigenvectors of $4^{k+1} I_{N'} - \text{DD}^{\text{T}}$ be $s_j, j \in [N']$. By Lemma A.5, the left singular vectors of D are given by $u_j = S s_j / \|s_j\|, j \in [N']$ where S is the alternating sign diagonal matrix in Lemma A.5.

Bounding the interior elements. The right singular vectors are given by $v_j = D^T u_j / \sqrt{\beta_j}$ for $j \in [N']$. Denote $k' = (k+1)/2$. Let $w \in \mathbb{R}^{k+2}$ denote the forward difference vector of order $k+1$, given by, $w_\ell = D_{1\ell}, \ell \in [k+2]$. Consider the approximation in (A.10). For interior elements of v_j, that is, for $m = k+2, \cdots, N-k-1$, abbreviating $m' = m - k - 2$,

$$\begin{aligned} \sqrt{\beta_j} v_{jm} &= \sum_{\ell=1}^{k+2} w_\ell \cdot u_{j,m'+\ell} \\ &= \frac{1}{\|s_j\|} \sum_{\ell=1}^{k+2} (-1)^{m'+\ell} w_\ell s_{j,m'+\ell} \\ &= \frac{(-1)^{m'+1}}{\|s_j\|} \sum_{\ell=1}^{k+2} |w_\ell| \left(M_{j,m'+\ell} + L_{j,m'+\ell} + R_{j,m'+\ell} + \epsilon_{j,m'+\ell}\right) \end{aligned}$$

Now note that $\sum_{\ell=1}^{k+2} |w_\ell| t^{\ell-1} = (t+1)^{k+1} = t^{(k+1)/2} p(t)$ where $p(t) = (2 + t + 1/t)^{(k+1)/2}$. Let $M_{j,m}^{(1)} = z_0^{(N'-1)/2-m} / |b_+(z_0)|, M_{j,m}^{(2)} = (-1)^{N'-j} \bar{z}_0^{(N'-1)/2-m} / |b_+(\bar{z}_0)|$ denote the two terms of $M_{j,m}$ and $L_{j,m}^\kappa, R_{j,m}^\kappa$ denote the κth term in the summation defining L, R for $j \in [N'], m \in [N'], \kappa \in [k]$. Continuing from the above display,

$$\sqrt{\beta_j} v_{jm} = \frac{(-1)^{m'+1}}{\|s_j\|} \Big(p(z_0) M_{j,m'+k'}^{(1)} + p(\bar{z}_0) M_{j,m'+k'}^{(2)} + \sum_{\kappa=1}^{k} p(z_\kappa) L_{j,m'+k'}^\kappa + p(\bar{z}_\kappa) R_{j,m'+k'}^\kappa$$

$$+ |w_\ell|\epsilon_{j,m'+\ell}\Big)$$

As $z_\kappa, \bar{z}_\kappa$ are roots of $p_\beta(t)$ in (A.9) with $\beta = 4^{k+1} - \beta_j$, note that $p(z_\kappa), p(\bar{z}_\kappa) \in \{-\sqrt{\beta_j}, \sqrt{\beta_j}\}$ for $\kappa \in \{0\} \cup [k]$ and further $p(z_0) = p(\bar{z}_0) = \sqrt{\beta_j}$. Therefore, we have

$$v_{jm} = \frac{(-1)^{m'+1}}{\|s_j\|}\left(M_{j,m'+k'} + \sum_{\kappa=1}^{k} \pm L^{\kappa}_{j,m'+k'} \pm R^{\kappa}_{j,m'+k'} + \sum_{\ell=1}^{k+2} |w_\ell|\epsilon_{j,m'+\ell}\right)$$

Again from Theorem 2.6 [1], as shown in Lemma A.3, $|M_{j,m'+k'}|/\|s_j\| = O(1/\sqrt{N})$ and the rest of the terms contribute $O(1/\sqrt{N})$. Therefore, $v_{jm} \leq \mu/\sqrt{N}$ for some constant μ independent of j, N, β_j. This completes the proof for the interior elements.

Bounding the boundary elements. We bound the remaining elements as follows. For each $j \in [N']$, observe that $Dv_j = \sqrt{\beta_j}u_j$ where the singular value $\sqrt{\beta_j} \leq c$ for some constant c depending only on k, by Lemma A.2. Therefore, by the incoherence of left singular vectors (Lemma A.3), there exists a constant μ such that

$$\|Dv_j\|_\infty \leq \frac{\mu}{\sqrt{N}}, \quad j \in [N'].$$

We know that for all $\ell \in [k+2]$, $|w_\ell| = \theta(1)$. So $v_j[k]$ should satisfy

$$|\sum_{\ell=1}^{k+2} w_\ell v_j[k+\ell-1]| \leq \frac{\mu}{\sqrt{N}}.$$

Note that $|w_1| = 1$. By the triangle inequality,

$$|v_j[k]| \leq \frac{\mu}{\sqrt{N}} + |\sum_{\ell=2}^{k+2} ||w_\ell||v_i[k+\ell-1]|.$$

Using the bound $|v_i[k+j-1]| \leq \mu/\sqrt{N}$ for all $j = 2, 3, \ldots, k+1$ for interior elements of right singular vectors, this implies that

$$|v_i[k]| \leq c(k)\frac{\mu}{\sqrt{N}}.$$

Repeat the same argument recursively to $v_j[k-1]$, $v_j[k-2]$ and all the way to $v_j[1]$. Similarly, we can get the same result for the last k components of v_j. This completes the proof. □

Lemma A.5. *Denote* $D = D^{(k+1)}_{\text{1d}} \in \mathbb{R}^{(N-k-1)\times N}$. *Let* $\Delta = DD^T, \Delta_+ = \text{abs}(DD^T)$ *where* $\text{abs}(A)_{ij} = |A_{ij}|$ *for a matrix. Let* $\Delta = U\Lambda U^T, \Delta_+ = U_+\Lambda_+U_+^T$ *be their singular value decompositions. Then*

(a) Δ, Δ_+ *are similar, that is,* $\Lambda = \Lambda_+$*;*

(b) U *and* U_+ *are same upto signs, that is,* $\text{abs}(U) = \text{abs}(U_+)$.

Proof of Lemma A.5. Denote $N' = N - k - 1$. Let $S \in \mathbb{R}^{N'\times N'}$ be the alternating sign diagonal matrix with the diagonal $[1, -1, 1, -1, \cdots]$. Note that $S^{-1} = S^T = S$. From the relation

$$\Delta = S^{-1}\Delta_+ S$$

we observe that Δ and Δ_+ are similar, that is, $\Lambda = \Lambda_+$. From their singular value decompositions,

$$U\Lambda U^T = SU_+\Lambda U_+^T S^T$$

we also see that $U = SU_+$ which implies $\text{abs}(U) = \text{abs}(U_+)$. □

A.8 Proof of Lemma 3

Denote

$$\widetilde{Z}_d = \big\{x = (x_1, \ldots, x_d) \in Z_d : x_j \le 1 - (k+1)/N, j = 1, \ldots, d\big\}.$$

Pick an arbitrary $\theta \in \mathcal{H}_d^{k+1}(L)$, corresponding to discretizations of $f \in H(k+1, L; [0,1]^d)$. The bound (15) holds at any $x \in \widetilde{Z}_d$, and the fact that $\delta(N) \le cL/N$ is verified by Lemma A.6. The KTF penalty is then

$$\|\widetilde{\Delta}^{(k+1)}\theta\|_1 = \sum_{x \in \widetilde{Z}_d} \big|\big(D_{x_j^{k+1}}\theta\big)(x)\big| \le cnLN^{k-1} = cLn^{1-(k+1)/d},$$

recalling $N = n^{1/d}$.

A.9 Lemma A.6

The following lemma follows standard calculations in numerical analysis, e.g., as in Strikwerda [8].

Lemma A.6. *Let $f \in H(k+1, L; [0,1]^d)$. The kth order forward discrete difference along a unit direction $v \in \mathbb{R}^d$, with step size $h > 0$, obeys at any point $x \in [0,1]^d$,*

$$\left|\frac{1}{h^k}(D_{v^k}\theta)(x) - \frac{\partial^k}{\partial v^k} f(x)\right| \le cLh,$$

where $c > 0$ is a constant depending only on k, provided that $x + khv \in [0,1]^d$ (so that the discrete approximation is well-defined).

Proof. By Taylor expanding f around x at $x, x + hv, x + 2hv, \ldots, x + khv$, we have

$$\begin{aligned}
f(x+hv) &= f(x) + \frac{\partial}{\partial v} f(x)h + \frac{1}{2}\frac{\partial^2}{\partial v^2} f(x)h^2 + \ldots + \frac{1}{k!}\frac{\partial^k}{\partial v^k} f(x)h^k + r(h), \\
f(x+2hv) &= f(x) + \frac{\partial}{\partial v} f(x)(2h) + \frac{1}{2}\frac{\partial^2}{\partial v^2} f(x)(2h)^2 + \ldots + \frac{1}{k!}\frac{\partial^k}{\partial v^k} f(x)(2h)^k + r(2h), \\
&\;\;\vdots \\
f(x+khv) &= f(x) + \frac{\partial}{\partial v} f(x)(kh) + \frac{1}{2}\frac{\partial^2}{\partial v^2} f(x)(kh)^2 + \ldots + \frac{1}{k!}\frac{\partial^k}{\partial v^k} f(x)(kh)^k + r(kh),
\end{aligned}$$

where $r(ih)$ is integral form of the remainder in the expansion for $x + ihv$, satisfying

$$|r(ih)| = \left|\frac{1}{k!}\int_0^{ih} \frac{\partial^{k+1}}{\partial v^{k+1}} f(x+tv)t^k \, dt\right| \le \frac{k^{k+1}}{(k+1)!} Lh^{k+1}, \quad i = 1, \ldots, k.$$

(Note that such integrals are well-defined since Lipschitz continuity of $\partial^k f/\partial v^k$ implies that the $(k+1)$st derivative $\partial^{k+1} f/\partial v^{k+1}$ exists almost everywhere and is Lebesgue integrable, by Rademacher's theorem.) In the inequality above, we invoked the Holder property, recalling $f \in H(k+1, L; [0,1]^d)$.

Now denote the vector of kth order forward difference coefficients by $w = (w_1, \ldots, w_{k+1})$, i.e.,

$$w_i = (-1)^i \binom{k}{i-1}, \quad i = 1, \ldots, k+1.$$

Inverting the above system of equations, and inspecting the last equality in the inverted system, gives

$$\frac{\partial^k}{\partial v^k} f(x)h^k = \sum_{i=1}^{k+1} w_i\Big(f(x + (i-1)hv) - r((i-1)h)\Big) = (D_{v^k}\theta)(x) - \sum_{i=1}^{k+1} w_i r((i-1)h).$$

Using our previous bound on the magnitude of remainders, we see

$$\left|(D_{v^k}\theta)(x) - \frac{\partial^k}{\partial v^k} f(x)h^k\right| \le \frac{k^{k+1}}{(k+1)!}\sum_{i=1}^{k+1} |w_i| Lh^{k+1},$$

and dividing through by h^k gives the claimed result. □

A.10 Proof of Lemma 4

We need only to construct a single counterexample for each $k, d \geq 1$. We give such a construction for $d = 2$ and $k = 1$; all other cases follows similarly. Consider a function $f : [0,1]^d \to \mathbb{R}$ defined by $f(x) = Mx_1 + x_2$, and let $\theta \in \mathbb{R}^n$ contain the evaluations of f over the grid Z_2. As f is linear, it is clearly an element of $H(2, 1; [0,1]^2)$. But, for any x on the left boundary of Z_2,

$$\|\Delta^{(2)}\theta\|_1 \geq \left| f\left(x+\frac{e_1}{N}\right)+f\left(x-\frac{e_2}{N}\right)+f\left(x+\frac{e_2}{N}\right)-3f(x)\right| = \left|f\left(x+\frac{e_1}{N}\right)-f(x)\right| = Mn^{1/2},$$

Since M can be arbitrary, this proves the result.

A.11 Proof of Theorem 4

We will show that

$$R\big(\mathcal{H}_d^{k+1}(L_n)\big) = \Omega(n^{-\frac{2k+2}{2k+2+d}} L_n^{\frac{2d}{2k+2+d}}). \tag{A.11}$$

Taking $L_n = C_n/n^{1-(k+1)/d}$ and applying Lemma 3 would then establish the result.

The result is "nearly" a textbook result on Holder classes in nonparametric regression. A standard result (e.g., see Chapter 2.8 of Korostelev and Tsybakov [5]) is that, in a model

$$y_i = f_0(x_i) + \epsilon_i, \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2),\ i = 1,\dots,n$$

where the design points $x_i \in [0,1]^d$, $i = 1,\dots,n$ are fixed and arbitrary, we have

$$\inf_{\hat{f}} \sup_{f_0 \in H(k+1, L_n; [0,1]^d)} \mathbb{E}\|\hat{f} - f_0\|_2^2 = \Omega(n^{-\frac{2k+2}{2k+2+d}} L_n^{\frac{2d}{2k+2+d}}), \tag{A.12}$$

where $\|\cdot\|_2$ denotes the L_2 norm on functions, defined as

$$\|f\|_2^2 = \int_{[0,1]^d} f(x)^2\, dx.$$

Note that in this notation, we can rewrite the desired result (A.11) as

$$\inf_{\hat{f}} \sup_{f_0 \in H(k+1, L_n; [0,1]^d)} \mathbb{E}\|\hat{f} - f_0\|_n^2 = \Omega(n^{-\frac{2k+2}{2k+2+d}} L_n^{\frac{2d}{2k+2+d}}), \tag{A.13}$$

where $\|\cdot\|_n$ the empirical norm on functions, defined as

$$\|f\|_n^2 = \frac{1}{n}\sum_{i=1}^n f(x_i)^2.$$

The proof of (A.12) reduces the estimation problem to a multiple hypothesis testing problem, and then constructs a sufficiently hard set of hypothesis by taking linear combinations of kernel "bump" functions and applying the Varshamov–Gilbert lemma (e.g., see Sections 2.7, 2.8 of Korostelev and Tsybakov [5], or Section 2.6 of Tsybakov [9]). But when the design points are $\{x_1,\dots,x_n\} = Z_d$, the regular lattice on $[0,1]^d$, the constructed bump functions are still orthogonal with respect to the empirical inner product (just as they are with respect to the L_2 inner product) because their supports are nonoverlapping. Therefore the exact same sequence of arguments leads to (A.13), i.e., leads to (A.11), if the empirical norm a bump function is at least of the same order as its L_2 norm as verified below.

Let the $[0,1]^d$ be partitioned into $1/h^d$ cubes where $h \asymp n^{-1/(2k+2+d)}$ and the bump function centered around 0 be

$$\varphi(x) = h^{k+1} K\left(\frac{2\|x\|_2}{h}\right) \mathbb{I}\left(x \in [-h/2, h/2]^d\right), \text{ where } K(u) = \exp\left(\frac{-1}{1-u^2}\right)\mathbb{I}(|u|<1).$$

We know that $\|\varphi\|_2^2 = \Theta(h^{2k+2+d})$. Therefore it is sufficient to show that $\|\varphi\|_n^2 = \Theta(h^{2k+2+d})$. Consider the set U of grid points which lie in a sphere of radius $ch/2$ where $c = \sqrt{\log_{2e} 2}$ around the center of the bump, say x_0. As $K(u) \geq \frac{1}{2}$ for $|u| \leq c$, we have $\varphi(x - x_0) \geq h^{k+1}.\frac{1}{2}$ for $x \in U$. The number of elements in U is $\asymp nh^d$ as we consider all points lying in a sphere of radius $ch/2$. Therefore $\|\varphi\|_n^2 \gtrsim \frac{1}{n} \cdot nh^d \cdot h^{2k+2} \asymp h^{2k+2+d}$. □

A.12 Proof of Theorem 5

Define a class

$$\mathcal{S}_d^{k+1} = \{\theta \in \mathbb{R}^n : \|\Delta^{(k+1)}\theta\|_2 \le B_n\} = \{\theta \in \mathbb{R}^n : \theta^T L^{k+1}\theta \le B_n^2\}.$$

Notice that $\mathcal{S}_d^{k+1}(B_n) \subseteq \mathcal{T}_d^{k+1}(C_n)$ provided $B_n = C_n/\sqrt{r}$, where $r \asymp n$ is the number of rows of of $\Delta^{(k+1)}$, owing to the simple inequality $\|x\|_1 \le \sqrt{r}\|x\|_2$ for $x \in \mathbb{R}^n$. We will show that

$$R\big(\mathcal{S}_d^{k+1}(B_n)\big) = \Omega(n^{-\frac{d}{2k+2+d}} B_n^{\frac{2d}{2k+2+d}}). \tag{A.14}$$

Taking $B_n \asymp C_n/\sqrt{n}$ would then give the result.

Letting $L = U\Lambda U^T$ be an eigendecomposition, and note that for any estimator $\hat{\theta}$ of θ_0,

$$\|\hat{\theta} - \theta_0\|_2 = \|U^T\hat{\theta} - U^T\theta_0\|_2,$$

which means that we may rotate the parameter space and equivalently consider the minimax error over the rotated class

$$\widetilde{\mathcal{S}}_d^{k+1} = \bigg\{\gamma \in \mathbb{R}^n : \sum_{i=1}^n \lambda_i^{k+1}\gamma_i^2 \le B_n^2\bigg\},$$

where we have denoted the eigenvalues (diagonal elements of Λ) as λ_i, $i \in [n]$. We will now seek to embed a hyperrectangle in the above class and make use of results of Donoho et al. [4].

Write $\gamma = (\alpha, \beta) \in \mathbb{R} \times \mathbb{R}^{n-1}$, and order $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$, so the above class becomes

$$\widetilde{\mathcal{S}}_d^{k+1} = \bigg\{(\alpha,\beta) \in \mathbb{R} \times \mathbb{R}^{n-1} : \sum_{i=2}^n \lambda_i^{k+1}\beta_i^2 \le B_n^2\bigg\} := \mathbb{R} \times \mathcal{E}(B_n),$$

where we have used the fact that $\lambda_1 = 0$. (Here and henceforth, although unconventional, we will index β according to components $i = 2, \ldots, n$, rather than $i = 1, \ldots, n-1$, because it simplifies notation later.) The minimax risk (writing $\gamma_0 = U^T\theta_0$, and $\gamma_0 = (\alpha_0, \beta_0)$) satisfies

$$\inf_{\hat{\gamma}} \sup_{\gamma_0 \in \widetilde{\mathcal{S}}_d^{k+1}} \frac{1}{n}\mathbb{E}\|\hat{\gamma} - \gamma_0\|_2^2 = \frac{\sigma^2}{n} + \inf_{\hat{\beta}} \sup_{\beta_0 \in \mathcal{E}(B_n)} \frac{1}{n}\mathbb{E}\|\hat{\beta} - \beta_0\|_2^2.$$

We focus on the second term. The ellipsoid $\mathcal{E}(B_n)$ is compact, convex, orthosymmetric and quadratically convex, the latter property as defined in Donoho et al. [4]. We can therefore use Lemma 6 and Theorem 7 in their work to conclude that the minimax risk over $\mathcal{E}(B_n)$ is at least four-fifths of the minimax linear risk of its hardest hyperrectangle,

$$\inf_{\hat{\beta}} \sup_{\beta_0 \in \mathcal{E}(B_n)} \frac{1}{n}\mathbb{E}\|\hat{\beta} - \beta_0\|_2^2 \ge \frac{4}{5} \sup_{H \subseteq \mathcal{E}(B_n)} \inf_{\hat{\beta} \text{ linear}} \sup_{\beta_0 \in H} \frac{1}{n}\mathbb{E}\|\hat{\beta} - \beta_0\|_2^2, \tag{A.15}$$

where the outer sup on the right-hand side is over hyperrectangles H contained in $\mathcal{E}(B_n)$. Consider hyperrectangles parametrized by a threshold τ,

$$H(\tau) = \{\beta \in \mathbb{R}^{n-1} : |\beta_i| \le t_i(\tau),\ i = 2, \ldots, n\},$$

where for all $i = 2, \ldots, n$, using multi-index notation $i = (i_1, \ldots, i_d)$, we let

$$t_{i+1}(\tau) = \begin{cases} B_n/(\sum_{i_1,\ldots,i_d \le \tau} \lambda_i^{k+1})^{1/2} & \text{if } i_1, \ldots, i_d \le \tau \\ 0 & \text{else.} \end{cases}$$

It is not hard to check that $H(\tau) \subseteq \mathcal{E}(B_n)$. The minimax linear risk over $H(\tau)$ decomposes, and can be evaluated exactly, as in Donoho et al. [4],

$$\inf_{\hat{\beta} \text{ linear}} \sup_{\beta_0 \in H(\tau)} \frac{1}{n}\mathbb{E}\|\hat{\beta} - \beta_0\|_2^2 = \frac{1}{n}\sum_{i=2}^n \frac{t_i(\tau)^2\sigma^2}{t_i(\tau)^2 + \sigma^2} = \frac{1}{n}\frac{(\tau^d - 1)\sigma^2 B_n^2}{B_n^2 + \sum_{i_1,\ldots,i_d \le \tau} \lambda_i^{k+1}}.$$

Lemma A.7 provides an upper bound on the sum in the denominator above, and plugging this in, we get

$$\inf_{\hat{\beta} \text{ linear}} \sup_{\beta_0 \in H(\tau)} \frac{1}{n}\mathbb{E}\|\hat{\beta} - \beta_0\|_2^2 \ge \frac{1}{n}\frac{(\tau^d - 1)\sigma^2 B_n^2}{B_n^2 + c\frac{\tau^{2k+2+d}}{N^{2k+2}}},$$

for a constant $c > 0$. This lower bound is maximized at $\tau \asymp (B_n^2 N^{2k+2})^{\frac{1}{2k+2+d}}$, in which case, we see

$$\inf_{\hat{\beta} \text{ linear}} \sup_{\beta_0 \in H(\tau)} \frac{1}{n}\mathbb{E}\|\hat{\beta} - \beta_0\|_2^2 = \Omega(n^{-\frac{d}{2k+2+d}} B_n^{\frac{2d}{2k+2+d}}).$$

Recalling (A.15), we have hence shown (A.14), and this completes the proof.

A.13 Lemma A.7

This result slightly generalizes Lemma A.3 of Sadhanala et al. [7].

Lemma A.7. *Let $L \in \mathbb{R}^{n \times n}$ denote the Laplacian matrix of the d-dimensional grid graph with equal side lengths $N = n^{1/d}$, and let*

$$\lambda_{i_1,\ldots,i_d} = 4 \sum_{j=1}^{d} \sin^2\Big(\frac{\pi(i_j - 1)}{2N}\Big), \quad i_1, \ldots, i_d \in [N]$$

denote its eigenvalues. Then for any integer $k \geq 0$ and $\tau \in [N]$,

$$\sum_{i_1,\ldots,i_d \leq \tau} \lambda_{i_1,\ldots,i_d}^{k+1} \leq c \frac{\tau^{2k+2+d}}{N^{2k+2}},$$

for a constant $c > 0$ depending only on k and d.

Proof. The proof follows the same chain of arguments as that for Lemma A.3 in Sadhanala et al. [7]. Using the fact that $\sin(x) \leq x$ for all $x \geq 0$,

$$\begin{aligned}\sum_{i_1,\ldots,i_d \leq \tau} \lambda_{i_1,\ldots,i_d}^{k+1} &\leq \frac{\pi^{2k+2}}{4^k N^{2k+2}} \sum_{i_1,\ldots,i_d \leq \tau} \Big((i_1-1)^2 + \ldots + (i_d-1)^2\Big)^{k+1} \\ &\leq \frac{\pi^{2k+2}}{4^k N^{2k+2}} \tau^{d-1} \sum_{i=1}^{\tau} (i-1)^{2k+2} \\ &\leq c \frac{\tau^{2k+2+d}}{N^{2k+2}}.\end{aligned}$$

□

References

[1] J.M. Bogoya, A. Böttcher, S.M. Grudsky, and E.A. Maximenko. Eigenvectors of hermitian toeplitz matrices with smooth simple-loop symbols. *Linear Algebra and its Applications*, 493 (Supplement C):606 – 637, 2016. ISSN 0024-3795. doi: https://doi.org/10.1016/j.laa.2015.12.017.

[2] Albrecht Bottcher, Sergei M. Grudsky, and Egor A. Maksimenko. On the structure of the eigenvectors of large hermitian Toeplitz band matrices. In *Recent Trends in Toeplitz and Pseudodifferential Operators*, volume 210, pages 15–36. Springer, 2010.

[3] A. Böttcher. *Spectral Properties of Banded Toeplitz Matrices*, pages 1–29. 2005.

[4] David Donoho, Richard Liu, and Brenda MacGibbon. Minimax risk over hyperrectangles, and implications. *Annals of Statistics*, 18(3):1416–1437, 1990.

[5] Aleksandr P. Korostelev and Alexandre B. Tsybakov. *Minimax Theory of Image Reconstructions*. Springer, 2003.

[6] Charles P. Neuman and Dave I. Schonbach. Discrete (Legendre) orthogonal polynomials—a survey. *International Journal for Numerical Methods in Engineering*, 8(4):743–770, 1974.

[7] Veeranjaneyulu Sadhanala, Yu-Xiang Wang, and Ryan J. Tibshirani. Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. *Advances in Neural Information Processing Systems*, 29, 2016.

[8] John C. Strikwerda. *Finite difference schemes and partial differential equations*. SIAM, 2004.

[9] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

[10] Yu-Xiang Wang, Alexander Smola, and Ryan J. Tibshirani. The falling factorial basis and its statistical applications. *International Conference on Machine Learning*, 31, 2014.

[11] Yu-Xiang Wang, James Sharpnack, Alex Smola, and Ryan J. Tibshirani. Trend filtering on graphs. *Journal of Machine Learning Research*, 17(105):1–41, 2016.