On a Phase Transition in General Order Spline Regression

Yandi Shen, Qiyang Han, and Fang Han

***Abstract*— In the Gaussian sequence model $Y = \theta_0 + \varepsilon$ in $\mathbb{R}^n$, we study the fundamental limit of statistical estimation when the signal θ_0 belongs to a class $\Theta_n(d, d_0, k)$ of (generalized) splines with free knots located at equally spaced design points. Here d is the degree of the spline, d_0 is the order of differentiability at each inner knot, and k is the maximal number of pieces. We show that, given any integer $d \geq 0$ and $d_0 \in \{-1, 0, \ldots, d-1\}$, the minimax rate of estimation over $\Theta_n(d, d_0, k)$ exhibits the following phase transition: $\inf_{\widetilde{\theta}} \sup_{\theta \in \Theta_n(d,d_0,k)} \mathbb{E}_\theta \|\widetilde{\theta} - \theta\|^2 \asymp_d \begin{cases} k \log\log(16n/k), & 2 \leq k \leq k_0, \\ k \log(en/k), & k \geq k_0 + 1. \end{cases}$ The transition boundary k_0, which takes the form $\lfloor (d+1)/(d-d_0) \rfloor + 1$, demonstrates the critical role of the regularity parameter d_0 in the separation between a faster $\log\log(16n)$ and a slower $\log(en)$ rate. We further show that, once encouraging an additional 'd-monotonicity' shape constraint (including monotonicity for $d = 0$ and convexity for $d = 1$), the above phase transition is removed and the faster $k \log\log(16n/k)$ rate can be achieved for all k. These results provide theoretical support for developing ℓ_0-penalized (shape-constrained) spline regression procedures as useful alternatives to ℓ_1- and ℓ_2-penalized ones.**

***Index Terms*— Splines, minimax rate, phase transition, law of iterated logarithm, shape constraint.**

I. Introduction

CONSIDER the regression model with equally spaced design points

$$Y_i = f_0(i/n) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $f_0 : [0, 1] \to \mathbb{R}$ is an unknown function and ε_i's are independent normal random variables with mean zero and variance σ^2. Throughout the paper, we reserve the notation θ_0 for the truth in (1), i.e., $(\theta_0)_i \equiv f_0(i/n)$. The main goal of this paper is to study the fundamental limit of statistical estimation of θ_0 when f_0 is a general order spline with free knots.

Manuscript received December 9, 2020; revised September 20, 2021; accepted January 19, 2022. Date of publication February 11, 2022; date of current version May 20, 2022. The work of Qiyang Han was supported by NSF under Grant DMS-1916221 and Grant DMS-2143468. The work of Fang Han was supported by NSF under Grant DMS-1712536 and Grant SES-2019363. *(Corresponding author: Yandi Shen.)*

Yandi Shen is with the Department of Statistics, The University of Chicago, Chicago, IL 60615 USA (e-mail: ydshen@uchicago.edu).

Qiyang Han is with the Department of Statistics, Rutgers University, Piscataway, NJ 08854 USA (e-mail: qh85@stat.rutgers.edu).

Fang Han is with the Department of Statistics, University of Washington, Seattle, WA 98105 USA (e-mail: fanghan@uw.edu).

Communicated by M. Raginsky, Associate Editor for Probability and Statistics.

Digital Object Identifier 10.1109/TIT.2022.3150253

Consider the (generalized) spline space with the following three parameters: d, the degree of the spline; d_0, the level of continuity; k, the maximal number of pieces. More formally, (d, d_0, k)-splines are defined as (exact definition in Section II):

$$\Big\{ f : [0,1] \to \mathbb{R} : f \text{ has at most } k+1 \text{ knots, is a degree } d \text{ polynomial between knots, and is } d_0\text{-times differentiable at each inner knot} \Big\}. \tag{2}$$

For any fixed degree d, d_0 takes value in $\{-1, 0, \ldots, d-1\}$, with $d_0 = d-1$ being the smoothest case and $d_0 = -1$ allowing for discontinuity between pieces. To avoid degeneracy to global polynomials, we only consider the case $k \geq 2$ in this paper. The corresponding sequence space is defined as

$$\Theta_n(d, d_0, k) \equiv \Big\{ \theta \in \mathbb{R}^n : \theta_i = f(i/n) \text{ for some } (d, d_0, k)\text{-spline } f \Big\}, \tag{3}$$

with free knots further assumed to be located at design points (see Section II for the precise definition and discussions). Compared to splines in more classical settings [1]–[3], the above parameter space does not fix the knots a priori and thus provides more flexibility. Previously, general order splines with free knots have been studied in, e.g., [4]–[6].

Splines of the forms (2) and (3) have frequently emerged in nonparametric curve estimation problems. For example, the classical smoothing splines [3] arise from minimizing the least squares criterion with an ℓ_2 roughness penalty. In the ℓ_1 world, splines are closely related to total variation regularization or denoising studied in, e.g., [4], [7]–[15]. In recent years, these methods with the spline space (3) received a revival of interest under the name *trend filtering*; cf. [5], [16]–[18].

A. Phase Transition in Minimax Rates

Despite the long history and large volume of works related to the spline spaces (2)-(3), their fundamental statistical limits have remained largely unexplored. Our first main result in this paper reveals the following intriguing phase transition in the minimax rate of estimation error over $\Theta_n(d, d_0, k)$:

$$\inf_{\widetilde{\theta}} \sup_{\theta \in \Theta_n(d,d_0,k)} \mathbb{E}_\theta \|\widetilde{\theta} - \theta\|^2 \asymp_d \begin{cases} \sigma^2 k \log\log(16n/k), & 2 \leq k \leq k_0, \\ \sigma^2 k \log(en/k), & k \geq k_0 + 1. \end{cases} \tag{4}$$

Here, $\|\cdot\|$ denotes the Euclidean norm and $\asymp_d$ denotes equivalence in order up to some positive constant that only depends on d. The transition boundary k_0, which takes the form $\lfloor (d+1)/(d-d_0) \rfloor + 1$ with $\lfloor \cdot \rfloor$ denoting the floor function, governs the maximal number of pieces above which the optimal dependence of the estimation error on the sample size n changes from the faster $\log\log(16n)$ rate to the slower $\log(en)$ rate. Notably, for any fixed degree d, k_0 is an increasing function of the regularity parameter d_0. In the two extreme cases, we have $k_0 = d+2$ if $d_0 = d-1$ (smoothest) and $k_0 = 2$ if $d_0 = -1$ (roughest). In other words, *the driving factor behind the phase transition in (4) is the regularity due to the differentiability structure encoded in* d, d_0.

The minimax rate in (4) is achieved by the ℓ_0-constrained spline least squares estimator (LSE) $\widehat{\theta} \equiv \widehat{\theta}(\Theta_n(d, d_0, k), Y)$, with $Y \equiv (Y_1, \ldots, Y_n)^\top$ and

$$\widehat{\theta}(\Theta, Y) \equiv \underset{\theta \in \Theta}{\arg\min} \|Y - \theta\|_2^2 \quad \text{for any } \Theta \subset \mathbb{R}^n. \tag{5}$$

In fact, a more general oracle inequality allowing for arbitrary model mis-specification can be proved for $\widehat{\theta}$. Due to the non-convexity of $\Theta_n(d, d_0, k)$, the solution to (5) with $\Theta = \Theta_n(d, d_0, k)$ may not be unique and we choose any $\widehat{\theta}$ that achieves the minimum. Among the three parameters, we take d and d_0 to be fixed in advance and consider k as a tuning parameter to balance the approximation error of θ_0 in (1) by $\Theta_n(d, d_0, k)$ and the complexity of the latter space. The estimator in (5) with $\Theta = \Theta_n(d, d_0, k)$ can therefore be viewed as a class of ℓ_0-splines in their constrained form.

The minimax rates in (4) suggest some interesting comparison between ℓ_0- and ℓ_1-regularizers in spline regression. For expository purpose, let us consider the simplest piecewise constant class $\Theta_n(0, -1, k)$, where the transition boundary is given by $k_0 = 2$. There, while the ℓ_0-constrained spline LSE, as defined in (5) with $\Theta = \Theta_n(0, -1, 2)$, is able to achieve the faster $\log\log(16n)$ rate with 2 pieces, the same rate has been proven to be un-attainable by the ℓ_1 trend filtering, even with an additional minimum spacing condition that could be substantially improved with ℓ_0-splines [18]–[20]. These theoretical discoveries thus motivate further algorithmic developments for the general ℓ_0-constrained spline LSEs. Some notable progress along this line includes dynamic programming algorithms in the discontinuous case ($d_0 = -1$) developed in [21]–[24], and in the first-order continuous case ($d_0 = 0$) developed in the recent work [25]. Generic efficient algorithms suited for ℓ_0-penalized general-order splines remain a challenging and important open research problem of great practical value.

B. Removal of the Phase Transition Under Shape Constraints

To motivate the second main result of this paper, we recall the following minimax result from [26]: for all $k \geq 2$,

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta_n^*(0,k)} \mathbb{E}_{\theta^*} \|\widetilde{\theta}^* - \theta^*\|^2 \asymp \sigma^2 k \log\log(16n/k), \tag{6}$$

where $\Theta_n^*(0, k)$ is the sub-class of $\Theta_n(0, -1, k)$ with non-decreasing signals. Comparing (4) with $d = 0, d_0 = -1$ and (6) above, we see that the phase transition from the faster rate $\log\log(16n)$ to the slower rate $\log(en)$ in (4) is removed in (6) under the additional monotonicity shape constraint. This raises the natural questions of whether a similar gain by shape constraints applies to higher-order splines, and if so, which type of shape constraints should be encouraged. As shape-constrained models repeatedly prove their usefulness in various applications, answering the above questions is of both practical and theoretical interests.

To this end, following [27], [28], we consider the following sub-class of (d, d_0, k)-splines with an additional 'd-monotone' shape constraint (exact definition in Section III):

$$\Big\{ f : [0,1] \to \mathbb{R} : f \text{ is a } (d, d-1, k)\text{-spline with non-decreasing highest-order polynomial coefficients} \Big\}. \tag{7}$$

Two canonical examples are $d = 0$ and 1, with the former corresponding to non-decreasing signals with at most k constant pieces, and the latter corresponding to convex signals with at most k linear pieces. Both classes have been extensively studied in the literature; cf. [26], [28]–[30] for the case $d = 0$ and [28], [30], [31] for the case $d = 1$. Define the sequence space corresponding to (7) as $\Theta_n^*(d, k)$.

As a special case of our second main result, we show an analogue of (6) under the convexity (=1-monotone) shape constraint: for all $k \geq 2$,

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta_n^*(1,k)} \mathbb{E}_{\theta^*} \|\widetilde{\theta}^* - \theta^*\|^2 \asymp \sigma^2 k \log\log(16n/k). \tag{8}$$

The same upper bound actually holds for the general d-monotone class $\Theta_n^*(d, k)$, with a complementary lower bound showing that the $\log\log(16n)$ rate cannot be further improved even with only two pieces. Comparing (4) and (8), it is hence clear that *a higher-order 'd-monotonicity' shape constraint removes the phase transition in (4) for general d in that the faster $k\log\log(16n/k)$ rate can now be achieved for all k*. The d-monotonicity therefore offers an attractive non-parametric sub-class $\Theta_n^*(d, k)$ of the general $\Theta_n(d, d-1, k)$ over which additional gain can be obtained in estimating the underlying signal.

C. Proof Techniques

We now remark on the technical challenges in proving (4) and (8). After a standard reduction step of the mean squared error of the LSE, the key is to control the Gaussian complexity width terms

$$\mathbb{E} \sup_{\theta \in \Theta : \|\theta\| \leq 1} \big(\varepsilon \cdot \theta\big)^2, \quad \Theta \in \Big\{\Theta_n(d, d_0, k), \Theta_n^*(d, k)\Big\}, \tag{9}$$

where $\cdot$ denotes the standard inner product in $\mathbb{R}^n$.

Unlike the relatively straightforward control of the above term that leads to the $\log(en)$ part in (4), non-trivial analytic and probabilistic efforts are required for the derivation of the correct transition boundary k_0 and the faster $\log\log(16n)$ rate. The analytic step is to derive, under the unit norm constraint on the spline vector, sharp enough controls for the magnitudes of the polynomial coefficients of signals in $\Theta_n(d, d_0, k)$ and $\Theta_n^*(d, k)$. On a given piece, these estimates should, apart from depending on the length of the current piece, be 'tied' to either the left-most or the right-most knot of the signal. In the unshaped class $\Theta_n(d, d_0, k)$, this is possible due to the strong regularity inherited in the differentiability structure in the regime $k \leq k_0$. In the shaped class $\Theta_n^*(d,k)$, it is the global regularity within the d-monotonicity shape constraint that plays the key role. Once the above controls are obtained, a generalized version of the law of iterated logarithm (LIL), which we developed in Section IV, can be applied to obtain the $\log\log(16n)$ rates in (4) and (8). In fact, this generalized LIL can be used for controlling the Gaussian complexity width (9) with a general piecewise polynomial structured weight vector θ, and therefore maybe of independent interest.

Although the proof techniques mentioned above are tailored to the derivation of the minimax rates in (4) and (8) in the Gaussian sequence model (1), these techniques may be of broader interest in other contexts as well, in particular in view of the fundamental role of the Gaussian complexity width (9) in various statistical learning settings. For example, the techniques developed in this paper have been recently applied in [32] in the context of change-point estimation with piecewise polynomial signals.

D. Organization

The rest of the paper is organized as follows. Sections II and III are devoted to the study of unshaped splines $\Theta_n(d,d_0,k)$ and shaped splines $\Theta_n^*(d,k)$, respectively. A general version of the LIL in expectation is developed in Section IV. Main proofs of the results are presented in Sections V and VI, with the remaining technical lemmas collected in the Appendix.

E. Notation

For any $x\in\mathbb{R}$, write $(x)_+\equiv\max\{x,0\}$. Let $\mathbf{1}.$ denote the indicator function. For any non-negative integers a,b, we use $[a;b]$ to denote the set $\{a,\ldots,b\}$ and $(a;b]$ to denote the set $\{a+1,\ldots,b\}$. For any two positive integers a,b, let $\mathrm{Mod}(a;b)$ be the remainder of a divided by b. For any two real numbers a,b, define $a\vee b\equiv\max\{a,b\}$ and $a\wedge b\equiv\min\{a,b\}$. For any positive integers $m\geq n$, let $\underline{\odot}(m;n)\equiv m(m-1)\ldots(m-n+1)$ and $\overline{\odot}(m;n)\equiv m(m+1)\ldots(m+n-1)$. Let $\mathbb{Z}_+$ denote the set of positive integers and $\mathbb{Z}_{\geq 0}\equiv\mathbb{Z}_+\cup\{0\}$. For any $d\in\mathbb{Z}_+$, let $\mathbb{S}^d\subset\mathbb{R}^{d+1}$ stand for the unit sphere. We write $\mathbb{E}_{\theta_0}$ as expectation under the experiment (1) with truth θ_0.

Let $C^m([0,1])$ denote the set of all m-times differentiable functions on $[0,1]$. For any $f\in C^m([0,1])$ and integer $0\leq\ell\leq m$, let $f^{(0)}(x)\equiv f(x)$ and $(D^{(\ell)}f)(x)\equiv f^{(\ell)}(x)$ be the ℓ-th derivative of f at point x. For any function f defined on $[0,1]$, $\tau\in[0,1]$, and real number c, define the first-order integral $(I^1_{c;\tau}f)(x)\equiv\int_\tau^x f(y)\,\mathrm{d}y+c$ for $x\in[0,1]$, and the m-th order integral iteratively as $(I^m_{c_0,\ldots,c_{m-1};\tau}f)(x)\equiv\big(I^1_{c_0;\tau}(I^{m-1}_{c_1,\ldots,c_{m-1};\tau}f)\big)(x)$ for any positive integer $m\geq 2$ and real sequence $\{c_\ell\}_{\ell=0}^{m-1}$. For any real function f, let $f(x_-)$ and $f(x_+)$ denote the left and right limits at x, respectively.

For two non-negative sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n\lesssim_d b_n$ (resp. $a_n\gtrsim_d b_n$) if $a_n\leq Cb_n$ (resp. $a_n\geq cb_n$) for some $C,c>0$ that only depend on d. We also write $a_n\asymp_d b_n$ if both $a_n\lesssim_d b_n$ and $a_n\gtrsim_d b_n$ hold. In the following, we will suppress d in $\lesssim_d$, $\gtrsim_d$, and $\asymp_d$ when no confusion is possible. For any given constants $a_1,a_2,\ldots$, we write $C(a_1,a_2,\ldots)$ and $c(a_1,a_2,\ldots)$ to denote positive constants that only depend on $a_1,a_2,\ldots$.

II. General-Order Spline Regression

We start with an exact definition of the general-order spline space in (2):

$$\begin{aligned}\mathcal{F}_n(d,d_0,k)\equiv\Big\{&f:[0,1]\to\mathbb{R}:\text{there exist}\\ &0\equiv n_0\leq\ldots\leq n_k\equiv n\text{ such that }n_0,\ldots,n_k\in\mathbb{Z}_{\geq0},\\ &n_i-n_{i-1}\geq(d+1)\mathbf{1}_{n_i>n_{i-1}},f\text{ is a }d\text{-degree polynomial}\\ &\text{on each interval }(n_{i-1}/n,n_i/n],\text{ and }f^{(\ell)}\big((n_i/n)_-\big)\\ &=f^{(\ell)}\big((n_i/n)_+\big)\text{ for all }i\in[1;k-1]\text{ and }\ell\in[0;d_0]\Big\}.\end{aligned}$$

For any fixed degree $d\geq 0$, the range of d_0 is $[-1;d-1]$, with $d_0=-1$ allowing the spline f to be completely discontinuous. The numbers $n_0/n,\ldots,n_k/n$ are the *knots* of f, with the middle $(k-1)$ ones as *inner knots*. Define the corresponding sequence space

$$\Theta_n(d,d_0,k)\equiv\Big\{\theta\in\mathbb{R}^n:\theta_i=f(i/n)\quad\text{for some }f\in\mathcal{F}_n(d,d_0,k)\Big\};\tag{10}$$

in what follows, we suppress the subscript n of $\Theta_n(d,d_0,k)$ when no confusion is possible and name $n_0,\ldots,n_k$ in its corresponding spline $f\in\mathcal{F}_n(d,d_0,k)$ the *knots* of θ. The gap $d+1$ between n_i and n_{i-1} in the above definition is necessary for the identifiability of f in the discontinuous case.

The function space $\mathcal{F}_n(d,d_0,k)$ enforces the inner knots of the spline to be positioned among the design points. This is due to two reasons:

(i) It ensures the existence of the LSE as defined in (5) with $\Theta=\Theta(d,d_0,k)$. Indeed, the minimization can be first taken over at most $(n+1)^{k-1}$ configurations of the inner knots, after which the problem becomes strictly convex with respect to the rest of the polynomial coefficients and thus has a unique solution.

(ii) It facilitates fast computation of the LSE via dynamic programming algorithms; see [25] for detailed illustration of the piecewise linear case.

For any fixed $d\in\mathbb{Z}_{\geq0}$ and $d_0\in[-1;d-1]$, let

$$k_0\equiv k_0(d,d_0)\equiv\left\lfloor\frac{d+1}{d-d_0}\right\rfloor+1.\tag{11}$$

Our first main result is the following oracle inequality. Recall that we only consider the case $k\geq 2$ in this paper and the analysis of global polynomials (corresponding to $k=1$) is rather straightforward.

Theorem 1: Fix any $\theta_0\in\mathbb{R}^n$. Let $\widehat{\theta}\equiv\widehat{\theta}(\Theta(d,d_0,k),Y)$ be the LSE as defined in (5) under the experiment (1) with truth θ_0. Then, for any $\delta>0$, there exists some $C=C(d,\delta)$ such that the following statements hold for any $n\geq\underline{n}$ with some $\underline{n}=\underline{n}(d)$. If $2\leq k\leq k_0$,

$$\mathbb{E}_{\theta_0}\|\widehat{\theta}-\theta_0\|^2\leq(1+\delta)\inf_{\theta\in\Theta(d,d_0,k)}\|\theta-\theta_0\|^2+C\sigma^2k\log\log(16n/k),$$

and if $k \geq k_0 + 1$,

$$\mathbb{E}_{\theta_0}\|\widehat{\theta} - \theta_0\|^2 \leq (1+\delta) \inf_{\theta\in\Theta(d,d_0,k)} \|\theta - \theta_0\|^2 + C\sigma^2 k \log(en/k).$$

The following lower bound result shows that Theorem 1 is optimal in the minimax sense.

Proposition 2: Under the experiment (1), there exists some $c = c(d)$ such that the following statements hold for all $n \geq \underline{n}$ with some $\underline{n} = \underline{n}(d)$. If $2 \leq k \leq k_0$,

$$\inf_{\widetilde{\theta}} \sup_{\theta\in\Theta(d,d_0,k)} \mathbb{E}_\theta\|\widetilde{\theta} - \theta\|^2 \geq c\sigma^2 k \log\log(16n/k),$$

and if $k \geq k_0 + 1$,

$$\inf_{\widetilde{\theta}} \sup_{\theta\in\Theta(d,d_0,k)} \mathbb{E}_\theta\|\widetilde{\theta} - \theta\|^2 \geq c\sigma^2 k \log(en/k),$$

where the infimum over $\widetilde{\theta}$ in both displays is taken over all measurable functions of Y.

The proof of Theorem 1 is presented in Section V, and the proof of Proposition 2 can be found in Appendix A-A.

Remark 3: The above two results imply, in particular, the minimax rates in (4). There, the upper bound $k\log(en/k)$ above the transition boundary k_0 is not essentially new and can be proved via straightforward modifications of the classical arguments in, e.g., [33], [34]. Rather, our main contribution lies in establishing the sharp transition boundary k_0 and the faster $\log\log(16n)$ rate below this boundary.

Given the above minimax rates, a remaining question of significant interest is adaptive estimation with respect to the unknown piece number k. Motivated by the classical results in [34]–[37], a promising candidate is the penalized spline LSE defined by

$$\widehat{\theta}_{\text{adapt}} \equiv \widehat{\theta}(\Theta(d, d_0, \widehat{k}), Y), \tag{12}$$

where the data-driven $\hat{k}$ is given by

$$\widehat{k} \equiv \arg\min_{1\leq k\leq n} \Big\{\|Y - \widehat{\theta}(\Theta(d,d_0,k),Y)\|^2 + \text{pen}(k; d, d_0)\Big\}, \tag{13}$$

with the penalty function $\text{pen}(\cdot; d, d_0)$ chosen to be proportional to the minimax rates in (4). In the piecewise constant case $(d, d_0) = (0, -1)$, the adaptive property of $\widehat{\theta}_{\text{adapt}}$ has been shown by [26, Theorems 3.1 and 4.1] using a special representation of the penalized LSE, cf. [26, Lemma 5.1]. We leave the general case as future work.

Remark 4: It is important to mention here one crucial difference between our perspective for the phase transition results and the $\log\log(16n)$ rates and the one taken in [26]. There, the faster $\log\log(16n)$ rate for $\Theta(0, -1, 2)$ follows immediately from the general iterated logarithmic rates for $\Theta^*(0, k)$, the class of piecewise constant and non-decreasing signals with at most k pieces (formally defined in Section III). In other words, the $\log\log(16n)$ rate for $\Theta(0, -1, 2)$ is perceived in [26] as a consequence of the monotonicity shape constraint. In contrast, the $\log\log(16n)$ rate for $\Theta(d, d_0, k)$ in (4) in the regime $k \leq k_0$ is inherited from the strong regularity in the signal parametrized by the degree d and the level of continuity d_0, rather than any explicit shape constraint. In the regime $k > k_0$, the $\log\log(16n)$ rate is not possible due to insufficient regularity in $\Theta(d, d_0, k)$, unless additional shape constraints are enforced; see Section III ahead for more details.

Remark 5: Recently, [18] studied the theoretical properties of trend filtering (TF), a class of ℓ_1-regularized discrete spline methods. More precisely, under the experiment (1), the d-th order TF estimator is

$$\widehat{\theta}^d_{\text{TF}} \equiv \min_{\theta\in\mathbb{R}^n} \Big\{\|Y - \theta\|^2 + \lambda\|D^{(d)}\theta\|_1\Big\}, \tag{14}$$

where $\|\cdot\|_1$ denotes the vector ℓ_1 norm, $\lambda > 0$ is a tuning parameter, and $D^{(d)} : \mathbb{R}^n \to \mathbb{R}^{n-d}$, when applied to vectors, represents the d-th order discrete difference operator defined as $D^{(0)}\theta \equiv \theta$, $D^{(1)}\theta \equiv (\theta_2 - \theta_1, \ldots, \theta_n - \theta_{n-1})^\top$, and $D^{(r)}\theta \equiv D^{(1)}(D^{(r-1)}\theta)$ for $r \geq 2$. Equation (14) is a convex problem and can be solved efficiently via algorithms designed for lasso-type problems [5].

For any $\theta_0 \in \Theta(d, d-1, k)$ in (1), Corollary 2.11 in [18] proved that, upon choosing the tuning parameter λ properly and assuming a minimum spacing condition to be detailed below,

$$\mathbb{E}_{\theta_0}\|\widehat{\theta}^{d+1}_{\text{TF}} - \theta_0\|^2 \leq C\sigma^2\Big(k\log(en/k) + k^{2(d+1)}\mathbf{1}_{d>0}\Big). \tag{15}$$

for some $C = C(d)$. Comparing (15) with our Theorem 1 and Proposition 2, we see the following distinctions between ℓ_0-regularized splines and their ℓ_1 counterparts.

(i) The bound (15) requires a minimum spacing condition that regulates, for non-vanishing pieces $(n_i; n_{i+1}]$ between knots with different signs (see Page 210 of [18] for their definition for the signs of knots), $n_{i+1} - n_i \geq cn/k$ for some $c = c(d)$. This is stronger than the constant gap condition assumed in $\Theta(d, d_0, k)$. Moreover, Theorem 4.2 in [20] suggests that this minimum spacing condition is essential to the TF estimators, namely, the performance of (14) could deteriorate to $\sqrt{n}$ (up to some polylogarithmic factors) without it.

(ii) Over the class $\Theta(d, d-1, k)$ with transition boundary $k_0 = d+2$, the ℓ_1 TF estimator in (14) is in general rate sub-optimal below the boundary, even with the additional minimum spacing condition mentioned above. Specifically, in the constant space $\Theta(0, -1, k)$, the minimax rate of estimation is $\log\log(16n)$ with $k = 2$ pieces, but the TF estimator in (14) with $d = 0$ can only achieve the slower $\log(en)$ rate in view of Lemma 2.4 of [18].

Remark 6: For the computation of the ℓ_0-constrained spline LSE $\widehat{\theta}$ and its adaptive version (12), the major difficulty in the development of efficient algorithms is measured by the regularity parameter d_0. For $d_0 = -1$, both estimators can be computed efficiently using standard dynamic programming algorithms [21]–[24] along with more refined pruning arguments [38], [39]. For the first-order continuous case ($d_0 = 0$), [25] recently introduced for the linear case ($d = 1$) a novel dynamic programming algorithm with empirical linear to quadratic time complexity, which can be readily extended to arbitrary order $d \in \mathbb{Z}_+$. We expect that the above method could potentially be extended to the case of general d and d_0, but this has to be left as the subject of future research.

Remark 7: We have focused in this paper the minimax rates with respect to the ℓ_2 loss on the design points; it would also

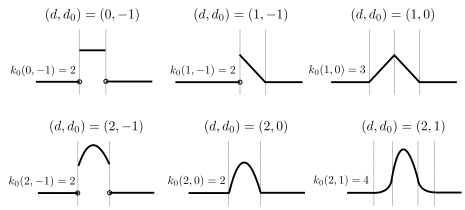

Fig. 1. Minimum number of $k = k_0 + 1$ pieces required to construct 1-sparse vectors with general position in $\Theta(d, d_0, k)$ for $d \in [0; 2]$ and $d_0 \in [-1; d-1]$.

be of interest to investigate minimax rates in the global L_2 loss over $\mathcal{F}_n(d, d_0, k)$. However, it is well-known that using the global L_2 loss may lead to difficulties near boundaries in some higher order shape constrained problems (see [40]); we leave it as future research.

Lastly, we provide some intuition for the form of k_0 defined in (11). This will mostly be clear from the perspective of minimax lower bounds in Proposition 2. There, the situation is somewhat similar to the derivation of minimax lower bounds in the sparse linear regression setting [33], [41], [42], in that we only have to find, for each fixed d and d_0, the minimum value of k such that a subset S of 1-sparse vectors can be constructed in $\Theta(d, d_0, k)$ with cardinality $|S| \geq cn$ for some $c = c(d)$. Heuristically, this value can be found via the following degree-of-freedom (DOF) calculation:

$$(k-2)(d+1) \geq (k-1)(d_0+1) + 1. \tag{16}$$

Here, the left-hand side is the DOF for any 1-sparse $\theta \in \Theta(d, d_0, k)$ with the two end pieces being constantly zero, as each of the middle $(k-2)$ pieces has $(d+1)$ DOF arising from the d-degree polynomial. On the right-hand side, the first term $(k-1)(d_0+1)$ results from the (d_0+1) continuity constraints at each of the $(k-1)$ inner knots, and the additional 1 DOF excludes the possibility that $\theta \equiv 0$. Solving (16) yields that $k \geq 1 + \lceil (d+2)/(d-d_0) \rceil$, which indeed holds for $k = k_0 + 1$ as defined in (11), with equality when $d_0 = d - 1$.

Figure 1 demonstrates the minimum number k of pieces needed for $d \in [0; 2]$ and $d_0 \in [-1; d-1]$ so that a 1-sparse vector can be constructed in general position. The minimum value of k in each scenario matches $k_0 + 1$ as defined in (11).

III. General-Order Splines With Shape Constraint

As mentioned in (6) in the introduction, in contrast to the phase transition in (4), the faster $\log\log(16n)$ rate of estimation becomes universal in the class $\Theta^*(0, k)$ that contains all piecewise constant non-decreasing signals. This section derives higher-order analogues of this result. We start with the convexity constraint in the linear case in Section III-A, and then generalize to higher-order splines in Section III-B.

A. Convex Piecewise Linear Regression

Convex regression is one of the central topics in shape constrained regression; see, e.g., [28], [30], [31] for global risk bounds and adaptation properties of the convex LSE.

We start by defining the function space of convex piecewise linear functions:

$$\mathcal{F}_n^*(1, k) \equiv \Big\{ f \in \mathcal{F}_n(1, 0, k) : f \text{ has non-decreasing slopes on } [0,1] \Big\}, \tag{17}$$

and the space on the sequence level:

$$\Theta_n^*(1, k) \equiv \Big\{ \theta^* \in \mathbb{R}^n : \theta_i^* = f^*\Big(\frac{i}{n}\Big) \text{ for some } f^* \in \mathcal{F}_n^*(1, k) \Big\}, \tag{18}$$

with the subscript n in $\Theta_n^*(1, k)$ suppressed in the sequel. The following two results show that the convexity shape constraint eliminates the phase transition in $\Theta(1, 0, k)$.

Proposition 8: Fix any $\theta_0 \in \mathbb{R}^n$. Let $\widehat{\theta}^* \equiv \widehat{\theta}(\Theta^*(1, k), Y)$ be the LSE as defined in (5) under the experiment (1) with truth θ_0. Then, for any $\delta > 0$, there exists some $C = C(\delta)$ such that for any $n \geq \underline{n}$ with some universal $\underline{n}$ and $k \geq 2$,

$$\mathbb{E}_{\theta_0} \|\widehat{\theta}^* - \theta_0\|^2 \leq (1+\delta) \inf_{\theta^* \in \Theta^*(1,k)} \|\theta^* - \theta_0\|^2 + C\sigma^2 k \log\log(16n/k). \tag{19}$$

Proposition 9: Under the experiment (1), there exists some universal constant c such that for all $n \geq \underline{n}$ with some universal $\underline{n}$ and $k \geq 2$,

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta^*(1,k)} \mathbb{E}_{\theta^*} \|\widetilde{\theta}^* - \theta^*\|^2 \geq c\sigma^2 k \log\log(16n/k),$$

where the infimum over $\widetilde{\theta}^*$ is taken over all measurable functions of Y.

Proposition 8 follows from its more general version in Theorem 11 ahead. The proof of Proposition 9 will be presented in Appedix A-B.

Remark 10: The in-expectation version of Theorem 4.3 in [30] proved a similar oracle inequality for the convex LSE:

$$\mathbb{E}_{\theta_0} \|\widehat{\theta}(\Theta^*, Y) - \theta_0\|^2 \leq \inf_{\theta^* \in \Theta^*} \Big(\|\theta^* - \theta_0\|^2 + Ck(\theta^*) \log(en/k(\theta^*)) \Big) \tag{20}$$

for some universal constant $C > 0$, where $\Theta^* \equiv \Theta^*(1, n)$ is the larger class of equispaced realizations of general convex functions on $[0,1]$, and $k(\theta^*)$ is the number of linear pieces of θ^*, i.e., $k(\theta^*) \equiv \sum_{i=2}^n \mathbf{1}_{2\theta_i^* < \theta_{i-1}^* + \theta_{i+1}^*}$. Note that Θ^*, as opposed to $\Theta^*(1, k)$, is a closed convex cone in $\mathbb{R}^n$. The bounds (19) and (20) are complementary in nature: the bound (20) exploits the convexity of Θ^* to obtain a sharp oracle inequality (in the sense of leading constant 1 before $\inf_{\theta^* \in \Theta^*} \|\theta^* - \theta_0\|^2$), but only achieves a slower worst-case $k \log(en/k)$ rate over the smaller class $\Theta^*(1, k)$; the bound (19), or its adaptive version modified in a similar way as (12), is minimax optimal over $\Theta^*(1, k)$ but loses the sharp leading constant 1. We conjecture that the $\log(en)$ rate in (20) for the convex LSE cannot be improved to a $\log\log(16n)$ rate, similar to the isotonic LSE that provably only achieves a $\log(en)$ rate for piecewise constant signals, cf. [26, Proposition 2.1]. The main bottleneck to formally proving such a result lies in the lack of an $\mathcal{O}(\log(en))$ lower bound for the Gaussian complexity width associated with $\Theta^* = \Theta^*(1, n)$, which still remains an interesting open problem in convex geometry.

B. General-Order Spline Regression With Shape Constraint

Following [27], [28], we consider the class of d-monotone splines defined as follows. Let

$$\mathcal{F}_n^*(0,k) \equiv \Big\{ f \in \mathcal{F}_n(0,-1,k) : f \text{ is non-decreasing on } [0,1] \Big\}$$

be the 0-monotone class. Next, for any $d \in \mathbb{Z}_+$, define

$$\mathcal{F}_n^*(d,k) \equiv \Big\{ f : [0,1] \to \mathbb{R} : \; f(x) = (I^d_{r_0,\ldots,r_{d-1};0} f_\circ)(x) \text{ for some } f_\circ \in \mathcal{F}_n^*(0,k) \text{ and real sequence } \{r_\ell\}_{\ell=0}^{d-1} \Big\}.$$

Define the sequence version of the above space as

$$\Theta_n^*(d,k) \equiv \Big\{ \theta^* \in \mathbb{R}^n : \theta_i^* = f^*\Big(\frac{i}{n}\Big) \text{ for some } f^* \in \mathcal{F}_n^*(d,k) \Big\}, \tag{21}$$

shorthanded as $\Theta^*(d,k)$. One can readily check that for $d = 0$, $\Theta^*(0,k)$ is the class of k-piece isotonic signals studied in [26]; for $d = 1$, $\Theta^*(1,k)$ coincides with the convex piecewise linear class in (18). Moreover, two facts follow immediately from the above definitions: (i) For any $d \geq 1$, $f^* \in \mathcal{F}_n^*(d,k) \subset C^{d-1}([0,1])$ so that $\Theta^*(d,k) \subset \Theta(d,d-1,k)$ with the latter defined in (10); (ii) For any $d \geq 1$ and $\ell \in [1;d]$, it holds that $(f^*)^{(\ell)} \in \mathcal{F}_n^*(d-\ell,k)$.

The following result, with Proposition 8 as a special case, shows that d-monotonicity eliminates the phase transition in the general spline space $\Theta(d,d-1,k)$. Its proof is given in Section VI.

Theorem 11: Fix any $\theta_0 \in \mathbb{R}^n$. Let $\widehat{\theta}^* \equiv \widehat{\theta}(\Theta^*(d,k), Y)$ be the LSE as defined in (5) under the experiment (1) with truth θ_0. Then, for any $\delta > 0$, there exists some $C = C(d,\delta)$ such that for any $n \geq \underline{n}$ with some $\underline{n} = \underline{n}(d)$ and $k \geq 2$,

$$\mathbb{E}_{\theta_0} \|\widehat{\theta}^* - \theta_0\|^2 \leq (1+\delta) \inf_{\theta^* \in \Theta^*(d,k)} \|\theta^* - \theta_0\|^2 + C\sigma^2 k \log\log(16n/k).$$

Moreover, there exists some $c = c(d)$ such that for all $n \geq \underline{n}$ and $k \geq 2$,

$$\inf_{\widetilde{\theta}^*} \sup_{\theta^* \in \Theta^*(d,k)} \mathbb{E}_{\theta^*} \|\widetilde{\theta}^* - \theta^*\|^2 \geq c\sigma^2 \log\log(16n),$$

where the infimum over $\widetilde{\theta}^*$ is taken over all measurable functions of Y.

Remark 12: The essential technical difficulties in proving Theorem 11 and Proposition 9 over the oracle inequality version of (6) (cf. Theorem 2.1 of [26]) rest in the additional regularity of $\Theta^*(d,k)$ over $\Theta^*(0,k)$.

(i) For the upper bound, [26] made essential use of the fact that $\widehat{\theta}(\Theta^*(0,k))$ is the sample average given the estimated knots; cf. Lemma 5.1 therein. The analogous property is, unfortunately, not true even for $\widehat{\theta}(\Theta^*(1,k))$. Instead, we provide a completely different proof which is based on a new parametrization for general-order splines with shape constraint (cf. Lemma 23 ahead). We further observe that this new proof technique, when applied to the setting of [26], significantly simplifies their proof; see Section VI-C for details.

(ii) For the lower bound in Proposition 9, the continuity constraint in $\Theta^*(1,k)$ requires a much more delicate construction of least favorable signals that achieves the $k \log\log(16n/k)$ rate, compared to $\Theta^*(0,k)$; see Appendix A-B for more details. This lower bound construction can actually be extended to yield the optimal $k \log\log(16n/k)$ rate over the quadratic class $\Theta^*(2,k)$, but a general lower bound of the order $k \log\log(16n/k)$ is still lacking for higher-order d-monotone splines.

IV. A Generalized Law of Iterated Logarithm

In this section, we present a generalized law of iterated logarithm (LIL) in expectation that underlies the $\log\log(16n)$ rates derived in Sections II and III. Recall that a centered random variable X is said to be *sub-Gaussian* with parameter τ, if there exists some $K > 0$ such that $\mathbb{E}\exp(\lambda X) \leq K \exp(\lambda^2 \tau^2/2)$ for any $\lambda \in \mathbb{R}$.

Theorem 13: Fix positive integers $d \geq 1$ and $n \geq 2$. Let $\{\varepsilon_i\}_{i=1}^n$ be a sequence of independent and identically distributed centered sub-Gaussian random variables with parameter 1. Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a strictly increasing continuous function with inverse ψ^{-1}. Let

$$Z \equiv \max_{1 \leq n_1 < n_2 \leq n} \frac{\big|\sum_{i \in (n_1;n_2]} (i-n_1)^d \varepsilon_i\big|}{(n_2-n_1)^d (n_2 \wedge (n-n_1))^{1/2}}.$$

Then, provided that

$$\int_1^\infty e^{-c_0 (\psi^{-1}(t))^2} \, \mathrm{d}t < \infty \tag{22}$$

for some sufficiently small $c_0 = c_0(d)$, there exist some $C_1 = C_1(\psi,d) > 0$ and $C_2 = C_2(d) > 0$ such that

$$\mathbb{E}\psi(Z) \leq C_1 \big[\psi\big((C_2 \log\log(16n))^{1/2}\big) \vee 1\big].$$

Compared to the classical LIL for (partial) sums of i.i.d. variables [43, Theorem 8.5.2], Theorem 13 does not provide an almost sure limit but makes the following two generalizations: (i) the independent summands are equipped with polynomial weights of arbitrary degree; (ii) both ends of the sum n_1 and n_2 are allowed to change, which leads to the factor $(n_2-n_1)^d (n_2 \wedge (n-n_1))^{1/2}$ in the denominator of Z. These generalizations are essential in the bounds of the complexity widths in (9) (see the proof of Proposition 17 for details) and potentially could find applications in other related problems involving splines of general order.

The proof of the above theorem shares certain similarity to the proof of the classical LIL; details can be found in Appendix B. Here are some choices of ψ's that will be relevant in the proofs of results in Sections II and III.

Example 14: Let $\psi(t) = t^\alpha$ where $\alpha > 0$. Then $\psi^{-1}(t) = t^{1/\alpha}$, so clearly (22) holds.

Example 15: Let $\psi(t) = e^{ct^\alpha} - 1$ where $\alpha, c > 0$. Then $\psi^{-1}(t) = \big(\log(1+t)/c\big)^{1/\alpha}$. So

$$\int_1^\infty e^{-(c_0/c^{2/\alpha})(\log(1+t))^{2/\alpha}} \, \mathrm{d}t \begin{cases} < \infty, & \alpha \in (0,2], c \in (0, c_0 \mathbf{1}_{\alpha=2} + \infty \mathbf{1}_{\alpha \in (0,2)}), \\ = \infty, & \text{otherwise.} \end{cases}$$

Note that a law of iterated logarithm in expectation fails in general for the choice $\psi(t) = e^{ct^\alpha} - 1$ whenever $\alpha > 2$, as $\alpha = 2$ corresponds to the maximal integrability of Gaussian random variables.

V. Proof of Theorem 1

Starting from this section, unless otherwise specified, we will focus on the case $\sigma^2 = 1$; the extension to an arbitrary $\sigma^2 > 0$ is straightforward and hence not recorded here. We will also omit the proof for the $k \log(en/k)$ part of Theorem 1 as it follows essentially from the classical arguments in [33], [34] by completely ignoring the regularity constraints. For the rest of the section, we focus on illustrating the form of k_0 in (11) from the upper bound perspective and proving the faster $\log\log(16n)$ rate below the transition boundary. Section V-A provides a proof outline with illustrative simple cases discussed at first. Section V-B reduces the proof of Theorem 1 to the bound of complexity width in Proposition 17. The key ingredients to the proof of this proposition will be presented in Sections V-C and V-D, followed by the main proof in Section V-E.

A. Proof Outline

1) Piecewise Linear Case: We first consider the piecewise linear case $d = 1, d_0 \in \{-1, 0\}$, and assume $\theta_0 = 0$ in (1) for simplicity of discussion. Here, the transition boundary in (11) is $k_0 = 2$ for $d_0 = -1$ and $k_0 = 3$ for $d_0 = 0$, beyond which the $\log\log(16n)$ rate cannot be attained. We focus on the case of $k = 3$ pieces and illustrate the difference between $d_0 = -1$ and $d_0 = 0$. To start, a standard reduction to complexity width in Proposition 16 ahead yields that for some universal constant $C > 0$,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\|\widehat{\theta} - \theta_0\|^2 - C\|\theta_{\mathrm{oracle}} - \theta_0\|^2 \\
&\quad \leq C \cdot \mathbb{E} \sup_{\theta \in \Theta(1,d_0,3): \|\theta\| \leq 1} \big(\varepsilon \cdot \theta\big)^2 \equiv C \cdot \mathbb{E}Z^2,
\end{aligned}$$

where θ_{oracle} is any oracle in $\Theta(1, d_0, 3)$ such that $\inf_{\theta \in \Theta(1,d_0,3)}\|\theta - \theta_0\|^2$ is achieved, and $\mathbb{E}Z^2$ is termed the 'complexity width' of $\Theta(1, d_0, 3)$. To bound $\mathbb{E}Z^2$, we use the following parametrization for any given $f \in \mathcal{F}_n(1, d_0, 3)$ with knots $0 = n_0/n \leq n_1/n \leq n_2/n \leq n_3/n = 1$: for $i \in \{0, 1, 2\}$,

$$f(x) = a_i + b_i(x - n_i/n), \quad x \in \left(\frac{n_i}{n}, \frac{n_{i+1}}{n}\right]. \tag{23}$$

Under the additional continuity constraint when $d_0 = 0$, one has

$$a_1 = a_0 + b_0\frac{n_1 - n_0}{n} \text{ and } a_2 = a_1 + b_1\frac{n_2 - n_1}{n}. \tag{24}$$

Under the parametrization (23), the supremum within the complexity width can be bounded by

$$\begin{aligned}
Z \leq \sup_{\theta \in \Theta(1,d_0,3): \|\theta\| \leq 1} \sum_{i=0}^{2} \Bigg(&|a_i|\bigg|\sum_{j \in (n_i; n_{i+1}]} \varepsilon_j\bigg| \\
&+ \left|\frac{b_i}{n}\right|\bigg|\sum_{j \in (n_i; n_{i+1}]} (j - n_i)\varepsilon_j\bigg|\Bigg).
\end{aligned}$$

The magnitudes of $\{a_i\}$ and $\{b_i\}$ can be drastically different for $d_0 = -1$ and $d_0 = 0$. We illustrate this on the middle piece $(n_1; n_2]$.

- ($d_0 = -1$). The constraint $1 \geq \|\theta\| \geq \|\theta\|_{(n_1; n_2]}$ directly yields the following estimates for a_1 and b_1 with some universal $C > 0$:

  $$|a_1| \leq C(n_2 - n_1)^{-1/2} \text{ and } |b_1/n| \leq C(n_2 - n_1)^{-3/2}. \tag{25}$$

  Such estimates cannot be improved for, e.g., $f(x) = c(L^{-1/2} - nL^{-3/2}(x - 1/2))\mathbf{1}_{(1/2, 1/2+L/n]}(x)$ for small $c > 0$ and $L \geq 2$.
- ($d_0 = 0$). With the additional continuity constraint in (24), refined estimates can be obtained:

  $$|a_1| \leq Cn_2^{-1/2} \text{ and } |b_1/n| \leq C(n_2 \wedge (n - n_1))^{-3/2}. \tag{26}$$

  These estimates only hold up to $k = 3$ pieces. For $k \geq 4$, the best possible estimates are of type (25) by considering, e.g., $f(x) = c\big(nL^{3/2}\big(x - (1/2 - L/n)\big)\mathbf{1}_{(1/2-L/n, 1/2]}(x) - nL^{3/2}\big(x - (1/2 + L/n)\big)\mathbf{1}_{(1/2, L/n+1/2]}(x)\big)$ for small $c > 0$ and $L \geq 2$.

The crucial difference here is that estimates of type (26) enable a law of iterated logarithm (cf. Theorem 13) with $\mathbb{E}Z^2 \lesssim \log\log(16n)$, while those of (25) correspond to the maxima of $O(n)$ independent Gaussian random variables with $\mathbb{E}Z^2 \lesssim \log(en)$.

2) General Case: Similar to the linear case discussed above, the key step is to prove

$$\mathbb{E} \sup_{\theta \in \Theta(d, d_0, k_0), \|\theta\| \leq 1} (\varepsilon \cdot \theta)^2 \leq C \log\log(16n), \tag{27}$$

and we need to obtain estimates of type (26). For simplicity, we consider the smoothest case $d_0 = d - 1$ so that $k_0 = d + 2$.

Fix a degree d, and any $f \in \mathcal{F}_n(d, d-1, d+2)$ along with the corresponding $\theta \in \Theta(d, d-1, d+2)$ of unit norm and knots $0 = n_0 \leq n_1 \leq \ldots \leq n_{d+2} = n$. We use the following parametrization of f:

$$f(x) = \sum_{\ell=1}^{d+1} a_\ell^i \left(x - \frac{n_i}{n}\right)^{\ell - 1}, \quad x \in \left(\frac{n_i}{n}, \frac{n_{i+1}}{n}\right], \tag{28}$$

and focus on a generic piece $(n_i; n_{i+1}]$ at the sequence level. Here the superscript i represents 'the $(i+1)$-th piece $(n_i; n_{i+1}]$' and the subscript ℓ represents 'the ℓ-th coefficient' in the polynomial. We aim at obtaining the following estimates: for all $\ell \in [1; d+1]$,

$$1 \geq c \cdot (a_\ell^i)^2((n_{i+1} - n_i)/n)^{2(\ell-1)}(n_{i+1} \wedge (n - n_i)), \tag{29}$$

with some $c = c(d)$. Once these estimates are obtained, one can immediately apply Theorem 13 to obtain a $\log\log(16n)$ bound on the complexity width on $(n_i; n_{i+1}]$.

In (29), the $(d-1)$-th order differentiability at each inner knot naturally divides the coefficients into two groups, the 'shared coefficients' $\{a_\ell^i\}_{\ell \in [1;d]}$ and the 'nuisance coefficient' a_{d+1}^i. This suggests the following two-step proof strategy:

(i) First, we show that the estimate for the second group, a_{d+1}^i, follows from that of the first group; cf. Lemma 21 ahead.

(ii) Second, we obtain estimates in (29) for $\ell \in [1; d]$ with the choice $k_0 = d + 2$; cf. Lemma 22 ahead.

In the proof below, we will see clearly why $k_0 = d+2$ is the maximal number of pieces where the estimates in (29) are achievable. At a high level, the coefficient estimates $\{a^i_\cdot\}$ on the piece $(n_i; n_{i+1}]$ necessarily depend on coefficient estimates at locations to the both sides of i. The passage of such information, for example from the rightmost knot, is precisely characterized in Lemma 19 ahead through a set of quadratic forms, which are obtained via 'iterative cancellation' to be detailed in Section V-C. The transition boundary k_0 is then determined via 'counting of quadratic forms' (cf. (37) in the main proof ahead) that mirrors the DOF calculation in (16), thereby unifying the heuristics in the upper and lower bounds.

B. Reduction to Complexity Width

We first introduce some notation. For any fixed $\theta_0 \in \mathbb{R}^n$, let $\theta_{\mathrm{oracle}} \equiv \theta_{\mathrm{oracle}}(\theta_0) \in \Theta(d, d_0, k)$ be an oracle such that $\inf_{\theta\in\Theta(d,d_0,k)} \|\theta - \theta_0\|$ is achieved, with knots $0 = n_0 \le n_1 \le \ldots \le n_k = n$. For each $\theta \in \mathbb{R}^n$, define $\theta_{[j]}$ as the sub-vector $(\theta_i)_{i\in(n_j;n_{j+1}]}$ and $v_j(\theta) \equiv v_j(\theta;\theta_{\mathrm{oracle}}) \equiv (\theta-\theta_{\mathrm{oracle}})_{[j]}/\|(\theta-\theta_{\mathrm{oracle}})_{[j]}\|$.

The following result is a standard reduction principle for the LSE tailored to the class of splines. Its proof can be found in Appendix C.

Proposition 16: Fix any $\theta_0 \in \mathbb{R}^n$. Let $\widehat{\theta} \equiv \widehat{\theta}(\Theta(d,d_0,k), Y)$ be the LSE as defined in (5) under the experiment (1) with truth θ_0. Then, for any $\delta > 0$, there exists some $C = C(\delta) > 0$ such that

$$\mathbb{E}_{\theta_0}\|\widehat{\theta} - \theta_0\|^2 \le (1+\delta)\|\theta_{\mathrm{oracle}} - \theta_0\|^2 + C\cdot \mathbb{E}\sup_{\theta\in\Theta(d,d_0,k)} \sum_{j=0}^{k-1}\big(\varepsilon_{[j]}\cdot v_j(\theta)\big)^2.$$

Now, note that each $v_j(\theta)$ is also a spline with unit norm and the same parameters (d, d_0, k) (rigorously speaking, the two end pieces of $v_j(\theta)$ may have length smaller than $d+1$, but these pieces are negligible since there are at most $2k$ of them and each only contributes a constant (up to d) factor to the complexity width). Therefore, in view of Proposition 16, the $\log\log(16n)$ part of Theorem 1 for $k \le k_0$ is immediately implied by the following result by noticing that $\Theta(d,d_0,k) \subset \Theta(d,d_0,k_0)$ for all $k \le k_0$.

Proposition 17: There exists some $C = C(d)$ such that

$$\mathbb{E}\sup_{\theta\in\Theta(d,d_0,k_0),\|\theta\|\le 1}(\varepsilon\cdot\theta)^2 \le C\log\log(16n).$$

The following two subsections present the main ingredients to the proof of Proposition 17, whose details will be presented in Section V-E.

C. Groundwork

Fix any $f \in \mathcal{F}_n(d, d_0, k_0)$ with knots $0 = n_0/n \le n_1/n \le \ldots \le n_{k_0}/n = 1$ and recall the parametrization (28). Due to the regularity constraints, similar relations as the linear equations of the type (24) exist between adjacent knots. We use the notation $\mathrm{Coef}[a^i_p; a^{i-1}_q]$ to denote the coefficient of a^{i-1}_q in the linear equation of a^i_p, i.e., $a^i_p = \sum_q \mathrm{Coef}[a^i_p; a^{i-1}_q]a^{i-1}_q$. The following lemma makes explicit this dependence. Its proof and proofs for other lemmas in this subsection are contained in Appendix C. We write

$$n_{i;j} \equiv (n_i - n_j)/n. \tag{30}$$

Lemma 18: For any $i \in [1; k_0-1]$, $p \in [1; d_0+1]$, and $q \in [1; d+1]$,

$$\mathrm{Coef}[a^i_p; a^{i-1}_q] = \binom{q-1}{p-1} n_{i;i-1}^{q-p}\mathbf{1}_{q\ge p}.$$

The next Lemma 19 provides, as described in the proof outline in Section V-A.2, the exact forms of the quadratic forms obtained by 'iterative cancellation' from right. These quadratic forms lay the foundation for coefficient estimates of type (29). For the rest of this section, we reserve the notation s for the number of 'iterative cancellation' performed.

Before stating the general formulation in Lemma 19, we first present the illustrative case of cubic spline ($d = 3, k_0 = 5$) in the sequence space with unit norm. We detail below the starting point ($s = 0$) and the first two steps of cancellation ($s \in \{1, 2\}$). Following the proof outline in Section V-A.2, we separate the quadratic forms that only involve the 'shared coefficients' $\{a^i_\ell\}_{\ell\in[1;3]}$ and those that also involve the 'nuisance coefficient' a^i_4.

- ($s = 0$). The ℓ_2 constraint on $(n_4; n_5]$ for the signal ($\|\theta\|_{(n_4;n_5]} \le \|\theta\| = 1$) provides control on the following 4 quadratic forms of length 1:

$$1 \ge c\cdot\bigg[\Big\{(n-n_4)(a^4_1)^2 + \frac{(n-n_4)^3}{n^2}(a^4_2)^2 + \Big\} \frac{(n-n_4)^5}{n^4}(a^4_3)^2 + \frac{(n-n_4)^7}{n^6}(a^4_4)^2\bigg].$$

- ($s = 1$). For the first cancellation, we have, by Lemma 18,

$$\begin{pmatrix} a^4_1 \\ a^4_2 \\ a^4_3 \end{pmatrix} = \begin{pmatrix} 1 & n_{4;3} & n^2_{4;3} & n^3_{4;3} \\ 0 & 1 & 2n_{4;3} & 3n^2_{4;3} \\ 0 & 0 & 1 & 3n_{4;3} \end{pmatrix}\begin{pmatrix} a^3_1 \\ a^3_2 \\ a^3_3 \\ a^3_4 \end{pmatrix}. \tag{31}$$

  The identity (31) enables us to first find a linear combination of (a^4_2, a^4_3) to cancel a^3_4, and then to find another linear combination of (a^4_1, a^4_2, a^4_3) to cancel both a^3_3 and a^3_4. These, along with direct expansion of the term $(a^4_3)^2(n-n_4)^5/n^4$ using (31), leave us with 3 quadratic forms of length 2:

$$\begin{aligned}1 \ge c\cdot\bigg[&\Big\{(n-n_4)\big(3a^3_1 + n_{4;3}a^3_2\big)^2 \\ &+ \frac{(n-n_4)^3}{n^2}\big(a^3_2 + n_{4;3}a^3_3\big)^2\Big\} \\ &+ \frac{(n-n_4)^5}{n^4}\big(a^3_3 + 3n_{4;3}a^3_4\big)^2\bigg].\end{aligned}$$

- ($s = 2$). For the second cancellation, we have, by Lemma 18 again,

$$\begin{pmatrix} a^3_1 \\ a^3_2 \\ a^3_3 \end{pmatrix} = \begin{pmatrix} 1 & n_{3;2} & n^2_{3;2} & n^3_{3;2} \\ 0 & 1 & 2n_{3;2} & 3n^2_{3;2} \\ 0 & 0 & 1 & 3n_{3;2} \end{pmatrix}\begin{pmatrix} a^2_1 \\ a^2_2 \\ a^2_3 \\ a^2_4 \end{pmatrix}.$$

Then, finding a linear combination of (a_1^3, a_2^3, a_3^3) to cancel a_4^2 and directly expanding $\big(a_2^3+n_{4;3}a_3^3\big)^2(n-n_4)^3/n^2$, we obtain 2 quadratic forms of length 3:

$$1 \geq c \cdot \Bigg[(n-n_4) \cdot \bigg(3a_1^2 + (2n_{3;2}+n_{4;3})a_2^2 + (n_{3;2}^2 + n_{3;2}n_{4;3})a_3^2\bigg)^2 + \frac{(n-n_4)^3}{n^2} \cdot \bigg(a_2^2 + (2n_{3;2}+n_{4;3})a_3^2 + (3n_{3;2}^2+3n_{3;2}n_{4;3})a_4^2\bigg)^2\Bigg].$$

To state the above cancellation scheme for general d and d_0, some further notation is introduced. Fix d, d_0, and the resulting k_0 as defined in (11). Define the sequence $\{\overline{\beta}_j^s\}$, $s \in [0; \lfloor (d_0+1)/(d-d_0)\rfloor]$ recursively as follows. Let $\overline{\beta}_0^s \equiv 1$,

$$\overline{\beta}_j^s \equiv \sum_{\ell=0}^{j} \binom{s(d-d_0)-\ell}{j-\ell} n_{k_0-s;k_0-1-s}^{j-\ell} \overline{\beta}_\ell^{s-1} \tag{32}$$

for $j \in [1; s(d-d_0)]$, and $\overline{\beta}_j^s \equiv 0$ for $j > s(d-d_0)$. Further define, for every $i \in [1; (s+1)d_0 - sd + 1]$ and $j \in [0; s(d-d_0)]$,

$$D(i,0) \equiv 1, \quad D(i,j) \equiv \frac{\overline{\odot}(i;j)}{\underline{\odot}(d+1-i;j)} \quad \text{for } j \geq 1.$$

Lastly, let $\overline{\beta}_{i,j}^s \equiv D(i,j)\overline{\beta}_j^s$.

We work under the extra condition that

$$n_{1;0} \wedge n_{k_0;k_0-1} \geq \max\{n_{2;1}, \ldots, n_{k_0-1;k_0-2}\}. \tag{33}$$

We remark that condition (33) is made merely for presentational simplicity; see the comments after Lemma 22 ahead for detailed discussion of this condition.

Lemma 19: Suppose (33) holds. Fix d, d_0, and k_0 as defined in (11), and any $\theta \in \Theta(d, d_0, k_0)$ such that $\|\theta\| \leq 1$. Then, there exists some $c = c(d)$ such that, for any $s \in [0; \lfloor (d_0+1)/(d-d_0)\rfloor]$,

$$1 \geq c\Bigg\{\sum_{i=1}^{(s+1)d_0-sd+1} + \sum_{i=(s+1)d_0-sd+2}^{sd_0-(s-1)d+1}\Bigg\} \frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\Bigg(\sum_{j=0}^{s(d-d_0)} \overline{\beta}_{i,j}^s a_{i+j}^{k_0-1-s}\Bigg)^2. \tag{34}$$

Remark 20: Several remarks for the quadratic forms above are in order.

(i) The quadratic forms in (34) are obtained via iterative cancellation from knot n_{k_0-1}.

(ii) In a generic $\overline{\beta}_{i,j}^s$, the superscript s marks the counts of cancellations already performed, i indicates the i-th quadratic form, and j indicates the coefficient for the j-th component in this quadratic form.

(iii) In (34), we intentionally separate the indices $i \in [1; (s+1)d_0 - sd+1]$ and $i \in [(s+1)d_0 - sd+2; sd_0-(s-1)d+1]$ since the first set of quadratic forms only involves the 'shared coefficients' $a_j^{\cdot}$ with $j \in [1; d_0+1]$.

(iv) Every time s grows by 1, the first summand of (34) has $(d-d_0)$ fewer quadratic forms with each one comprising of $(d-d_0)$ more components.

D. Key Estimates

Recall the coefficient sequence $\{a_\ell^i\}_{i\in[0;k_0-1],\ell\in[1;d+1]}$ defined in (28). As described in Section V-A, we aim to obtain sharp estimates of type (29). For any $a, b \in [1; n]$, define

$$M(a,b) \equiv (a \wedge (n-b))^{1/2}.$$

The first result below reduces the task of obtaining (29) for all the coefficients down to estimating only the 'shared coefficients' $\{a_\ell^{\cdot}\}_{\ell\in[1;d_0+1]}$, from which the estimates for 'nuisance coefficients' $\{a_\ell^{\cdot}\}_{\ell\in[d_0+2;d+1]}$ can be derived. Its proof can be found in Appendix C.

Lemma 21: Fix any $i \in [1; k_0-2]$. Suppose there exists some $c = c(d)$ such that for every $\ell \in [1; d_0+1]$, it holds that $1 \geq c(a_\ell^i)^2 n_{i+1,i}^{2(\ell-1)} M^2(n_{i+1}, n_i)$. Then, there exists some $c' = c'(d)$ such that

$$1 \geq c'(a_\ell^i)^2 n_{i+1,i}^{2(\ell-1)} M^2(n_{i+1}, n_i)$$

for every $\ell \in [d_0+2; d+1]$.

Following the preceding lemma, the next result, which builds on the groundwork derived in Lemma 19, makes use of an inductive argument to derive sharp estimates of the type (29) for $\{a_\ell^{i+1}\}_{\ell\in[1;d_0+1]}$ on a fixed target piece $(n_{i+1}; n_{i+2}]$. To make the notation more accessible, we present here the special case $d_0 = d-1$ (so that $k_0 = d+2$) and defer the case of general d_0 to Appendix C-E.

Lemma 22: Suppose $d_0 = d-1$ and (33) holds. Fix $i \in [0; d-1]$. For some $c = c(d)$, the following estimates hold for all locations $1 \leq j \leq i+1$:

$$1 \geq c \max_{1\leq \ell \leq d}\Bigg\{(a_\ell^j)^2 \cdot n_{i+2;j}^{2\{(d-i)\wedge(\ell-1)\}} \cdot \Bigg(\prod_{k=d-i+2}^{(d-j+2)\wedge \ell} n_{d+3-k;j}^2\Bigg) \cdot n_{j+1;j}^{2(\ell-(d-j+2))_+} \cdot M^2(n_{j+1}, n_j)\Bigg\}.$$

Here $\prod_{k=k_1}^{k_2} \equiv 1$ for $k_2 < k_1$. In particular, for $j = i+1$:

$$1 \geq c \max_{1\leq \ell\leq d}\Bigg\{(a_\ell^{i+1})^2 \cdot n_{i+2;i+1}^{2(\ell-1)} \cdot M^2(n_{i+2}, n_{i+1})\Bigg\}. \tag{35}$$

The proof of the above lemma is presented in the next subsection. We emphasize that the condition (33) is made only for presentational simplicity, as we explain below. If it does not hold, we can adopt the following partition of the pieces $\{(n_0; n_1], \ldots, (n_{d+1}; n_{d+2}]\}$ via general length constraints. Fix a target piece $(n_{i+1}; n_{i+2}]$ with $i \in [0; d-1]$.

S1. First locate among all pieces the longest one denoted as $(n_{i_1^*}; n_{i_1^*+1}]$ with $i_1^* \in [0; d+1]$. If this is the target piece, then we can directly apply Lemma 27 in Appendix E to this piece to obtain the desired estimates in (35).

S2. If not, assume without loss of generality that the target piece is to the left of this longest piece, i.e., $i+1 < i_1^*$. Then, we can locate the longest piece among $\{(n_0; n_1], \ldots, (n_{i_1^*-1}; n_{i_1^*}]\}$, which we denote as $(n_{i_2^*}; n_{i_2^*+1}]$ with $i_2^* \in [0; i_1^*-1]$. If the target piece is among $\{(n_{i_2^*}; n_{i_2^*+1}], \ldots, (n_{i_1^*-1}; n_{i_1^*}]\}$, we can then make the following two modifications of Lemmas 19 and 22: (i) choose location $n_{i_1^*}$ (instead of the current n_{d+1}) as the starting point for the cancellation of the quadratic forms; (ii) choose location i_2^*+1 (instead of the current location 1) as the starting point for the induction in Lemma 22. These two modifications will yield the desired estimates for $\{a_\ell^{i+1}\}$ in (35).

S3. If this is not the case, i.e., $(n_{i+1}; n_{i+2}] \in \{(n_1; n_2], \ldots, (n_{i_2^*-1}; n_{i_2^*}]\}$, we can then iterate S2 with i_2^* in place of i_1^*. This partitioning will terminate in a finite number of steps.

Condition (33) (with $n_{1;0} \le n_{d+2;d+1}$), along with the current versions of Lemmas 19 and 22, correspond to the above partitioning scheme with an early stop at S2 with $i_1^* = d+1$ and $i_2^* = 0$. On the other hand, condition (33) represents the most difficult case in the sense that the maximal gap $i_1^* - i_2^* = d+1$ activates the condition $k \le k_0 = d+2$ as seen in (37) in the proof ahead.

E. Main Proof

The main step in the proof of Proposition 17 is the set of coefficient estimates in Lemma 22, with its more general version stated in Appendix C-E. We present the proof of this lemma in the special case $d_0 = d-1$; the proof for the general case is completely analogous.

Proof of Lemma 22: Let

$$Q_j^2(\ell) = n_{i+2;j}^{2\{(d-i)\wedge(\ell-1)\}} \cdot \Bigg(\prod_{k=d-i+2}^{(d-j+2)\wedge\ell} n_{d+3-k;j}^2 \Bigg) \cdot n_{j+1;j}^{2(\ell-(d-j+2))_+}.$$

For the rest of the proof, empty $\prod$ is to be understood as 1 and empty $\bigvee$ is to be understood as 0. We will prove (a slightly stronger version with $M(n_1, n_0)$ instead of $M(n_{j+1}, n_j)$)

$$1 \gtrsim \max_{1\le\ell\le d} \big\{ (a_\ell^j)^2 Q_j^2(\ell) \big\} \cdot M^2(n_1, n_0) \tag{36}$$

by induction on $j \in [1; i+1]$. The baseline case $j=1$ clearly holds by the condition (33) and application of Lemma 27 to the piece $(n_0; n_1]$. Now, suppose the induction holds up to some location $j \in [1; i]$, and we will prove the iteration at location $j+1$.

(Part I). We deal with $\{a_\ell^{j+1}\}_{\ell=1}^{d-j+1}$ in this part. For this, we first obtain estimates for a_{d+1}^j and then use triangle inequality. Applying Lemma 19 with $d_0 = d-1$ and $s = d-j$, the j-th term in the first summand therein yields that

$$\begin{aligned} 1 &\gtrsim \frac{(n-n_{d+1})^{2j-1}}{n^{2(j-1)}} \Bigg(\sum_{\ell=0}^{d-j} \overline{\beta}_{j,\ell}^{d-j} a_{j+\ell}^{j+1} \Bigg)^2 \\ &= \frac{(n-n_{d+1})^{2j-1}}{n^{2(j-1)}} \\ &\quad \Bigg(\sum_{\ell=0}^{d-j} \overline{\beta}_{j,\ell}^{d-j} \sum_{k=\ell}^{d-j+1} \binom{k+j-1}{\ell+j-1} n_{j+1;j}^{k-\ell} a_{k+j}^j \Bigg)^2 \\ &\equiv \frac{(n-n_{d+1})^{2j-1}}{n^{2(j-1)}} \Bigg(\sum_{k=0}^{d-j+1} \bar{\gamma}_{j,k}^{d-j+1} a_{k+j}^j \Bigg)^2, \end{aligned}$$

where we used Lemma 18 and $\bar{\gamma}_{j,k}^{d-j+1} \equiv \sum_{q=0}^{(d-j)\wedge k} \overline{\beta}_{j,q}^{d-j} \binom{k+j-1}{q+j-1} n_{j+1;j}^{k-q}$. Note that for a generic number of k pieces, when $j=1$, we need to take $d_0 = d-1$ and $s = (k-1)$-$(j+1) = k-3$ in Lemma 19, in which case the first summand is non-void if and only if

$$d - s = d - k + 3 \ge 1 \iff k \le d+2. \tag{37}$$

This explains the transition boundary $k_0 = d+2$ as in (11).

Combining the above estimate with the estimates for $\{a_k^j\}_{k=j}^d$ from the induction assumption, and using Lemma 28 to cancel everything but a_{d+1}^j, we have

$$\begin{aligned} 1 &\gtrsim \frac{(n-n_{d+1})^{2j-1}}{n^{2(j-1)}} \Bigg(\sum_{k=0}^{d-j+1} \bar{\gamma}_{j,k}^{d-j+1} a_{k+j}^j \Bigg)^2 \\ &\quad + \sum_{k=j}^{d} \Big[(a_k^j)^2 \cdot Q_j^2(k) \cdot M^2(n_1, n_0) \Big] \\ &\gtrsim (a_{d+1}^j)^2 \Bigg\{ \frac{(n-n_{d+1})^{2j-1} (\bar{\gamma}_{j,d-j+1}^{d-j+1})^2}{n^{2(j-1)}} \wedge \\ &\quad \bigwedge_{k=j}^{d} \Bigg[Q_j^2(k) M^2(n_1, n_0) \frac{(\bar{\gamma}_{j,d-j+1}^{d-j+1})^2}{(\bar{\gamma}_{j,k-j}^{d-j+1})^2} \Bigg] \Bigg\} \\ &\equiv (a_{d+1}^j)^2 \Bigg\{ A_j \wedge \bigwedge_{k=j}^{d} B_{j,k} \Bigg\}. \end{aligned}$$

As $A_j / (\bar{\gamma}_{j,d-j+1}^{d-j+1})^2 = n_{d+2;d+1}^{2(j-1)} (n - n_{d+1}) \gtrsim B_{j,j} / (\bar{\gamma}_{j,d-j+1}^{d-j+1})^2$ by the assumption that the two end pieces are longer than any middle pieces, we only need to bound from below $\wedge_{k=j}^d B_{j,k}$. By definition of $\bar{\gamma}_{\cdot,\cdot}^{\cdot}$ and non-negativity of $\overline{\beta}_{\cdot,\cdot}^{\cdot}$, for any $j \le k \le d$,

$$\begin{aligned} \frac{(\bar{\gamma}_{j,d-j+1}^{d-j+1})^2}{(\bar{\gamma}_{j,k-j}^{d-j+1})^2} &\asymp \frac{\big(\sum_{q=0}^{d-j} \overline{\beta}_{j,q}^{d-j} n_{j+1;j}^{(d-j+1)-q} \big)^2}{\big(\sum_{q=0}^{k-j} \overline{\beta}_{j,q}^{d-j} n_{j+1;j}^{k-j-q} \big)^2} \\ &\quad \text{(by definition)} \\ &\asymp \bigvee_{p=0}^{d-k} \frac{\bigvee_{q=0}^{k-j} (\overline{\beta}_{j,p+q}^{d-j})^2 n_{j+1;j}^{2\{(d-j+1)\text{-}(p+q)\}}}{\bigvee_{q=0}^{k-j} (\overline{\beta}_{j,q}^{d-j})^2 n_{j+1;j}^{2(k-j-q)}} \\ &\quad \text{(by rearranging the numerator)} \\ &\ge \bigvee_{p=0}^{d-k} \Bigg\{ n_{j+1;j}^{2(d-k+1-p)} \bigwedge_{q=0}^{k-j} \Bigg(\frac{\overline{\beta}_{j,p+q}^{d-j}}{\overline{\beta}_{j,q}^{d-j}} \Bigg)^2 \Bigg\} \\ &\quad \text{(by Lemma 28)} \\ &\gtrsim \bigvee_{p=0}^{d-k} \Bigg\{ n_{j+1;j}^{2(d-k+1-p)} \bigwedge_{q=0}^{k-j} \prod_{r=1+q}^{p+q} n_{d+2-r;j+1}^2 \Bigg\} \\ &\quad \text{(by Lemma 33)} \\ &= \bigvee_{p=0}^{d-k} \Bigg\{ n_{j+1;j}^{2(d-k+1-p)} \prod_{r=1+k-j}^{p+k-j} n_{d+2-r;j+1}^2 \Bigg\} \\ &\quad \text{(minimum at } q = k-j). \end{aligned}$$

Hence

$$\begin{aligned} 1 &\gtrsim (a_{d+1}^j)^2 \Bigg[\bigwedge_{k=j}^{d} Q_j^2(k) \\ &\quad \cdot \bigvee_{p=0}^{d-k} \Bigg\{ n_{j+1;j}^{2(d-k+1-p)} \prod_{r=1+k-j}^{p+k-j} n_{d+2-r;j+1}^2 \Bigg\} \Bigg] M^2(n_1, n_0). \end{aligned}$$

This implies that for $1 \le \ell \le d$, by taking $p = (\ell - k)_+$ above and Lemma 18,

$$M(n_1,n_0)|a_\ell^{j+1}| \lesssim M(n_1,n_0)\Big[\sum_{k=\ell}^{d} n_{j+1;j}^{k-\ell}|a_k^j| + n_{j+1;j}^{d+1-\ell}|a_{d+1}^j|\Big]$$

$$\begin{aligned}
&\lesssim \sum_{k=\ell}^{d} n_{j+1;j}^{k-\ell} Q_j^{-1}(k) \\
&\quad + \bigvee_{k=j}^{d}\Big\{Q_j^{-1}(k) n_{j+1;j}^{k-\ell+(\ell-k)_+} \prod_{r=1}^{(\ell-k)_+} n^{-1}_{d+2+j-k-r;j+1}\Big\} \\
&= \sum_{k=\ell}^{d} n_{j+1;j}^{k-\ell} Q_j^{-1}(k) \\
&\quad + \bigvee_{j\le k<\ell\vee j}\Big\{Q_j^{-1}(k)\prod_{r=1}^{\ell\vee j-k} n^{-1}_{d+2+j-k-r;j+1}\Big\} \\
&\quad + \bigvee_{\ell\vee j\le k\le d}\Big\{Q_j^{-1}(k) n_{j+1;j}^{k-\ell}\Big\}.
\end{aligned}$$

Using that $k \mapsto n_{j+1;j}^{k-\ell} Q_j^{-1}(k)$ is non-increasing, the first and third terms in the above display are on the same order as $Q_j^{-1}(\ell) + Q_j^{-1}(\ell\vee j) n_{j+1;j}^{\ell\vee j-\ell} \asymp Q_j^{-1}(\ell)$. Hence we only need to verify for all $1 \le \ell \le d-j+1$, $1 \le j \le i$,

$$\begin{aligned}
&\mathfrak{Q}_{j,1}(\ell) + \mathfrak{Q}_{j,2}(\ell) \equiv Q_j^{-1}(\ell) \\
&\quad + \bigvee_{j\le k<\ell\vee j}\Big\{Q_j^{-1}(k)\prod_{r=1}^{\ell\vee j-k} n^{-1}_{d+2+j-k-r;j+1}\Big\} \lesssim Q_{j+1}^{-1}(\ell).
\end{aligned} \tag{38}$$

(Case 1). If $1 \le \ell \le d-i+1$, $Q_j^{-1}(\ell) = n_{i+2;j}^{-(\ell-1)}$ and $Q_j^{-1}(\ell) = n_{i+2;j+1}^{-(\ell-1)}$, so:

- (first term) $\mathfrak{Q}_{j,1}(\ell) = n_{i+2;j}^{-(\ell-1)} \le n_{i+2;j+1}^{-(\ell-1)} = Q_{j+1}^{-1}(\ell)$.
- (second term) without loss of generality we assume $\ell > j$ (otherwise this term does not exist):

$$\begin{aligned}
\mathfrak{Q}_{j,2}(\ell) &= \bigvee_{j\le k<\ell}\Big\{n_{i+2;j}^{-(k-1)}\prod_{r=1}^{\ell-k} n^{-1}_{d+2+j-k-r;j+1}\Big\} \\
&\le \bigvee_{j\le k<\ell}\Big\{n_{i+2;j}^{-(k-1)} n^{-(\ell-k)}_{d+2+j-\ell;j+1}\Big\} \\
&\le \bigvee_{j\le k<\ell}\Big\{n_{i+2;j+1}^{-(k-1)} n^{-(\ell-k)}_{i+2;j+1}\Big\} \\
&= n_{i+2;j+1}^{-(\ell-1)} = Q_{j+1}^{-1}(\ell),
\end{aligned}$$

where the first equality follows since $k < \ell \le d-i+1$ so that $Q_j^{-1}(k) = n_{i+2;j}^{-(k-1)}$, and the second inequality follows by noting that $\ell \le d-i+1$ implies $d+2+j-\ell \ge i+2$.

(Case 2). If $d-i+2 \le \ell \le d-j+1$, $Q_j^{-1}(\ell) = n_{i+2;j}^{-(d-i)}\prod_{s=d-i+2}^{\ell} n^{-1}_{d+3-s;j}$ and $Q_{j+1}^{-1}(\ell) = n_{i+2;j+1}^{-(d-i)}\prod_{s=d-i+2}^{\ell} n^{-1}_{d+3-s;j+1}$, so:

- (first term) similarly as above,

$$\mathfrak{Q}_{j,1}(\ell) = n_{i+2;j}^{-(d-i)}\prod_{s=d-i+2}^{\ell} n^{-1}_{d+3-s;j} \le n_{i+2;j+1}^{-(d-i)}\prod_{s=d-i+2}^{\ell} n^{-1}_{d+3-s;j+1} = Q_{j+1}^{-1}(\ell).$$

- (second term) similarly as above we assume $\ell > j$, then

$$\begin{aligned}
&\mathfrak{Q}_{j,2}(\ell) \\
&= \bigvee_{j\le k<\ell}\Big\{n_{i+2;j}^{-(d-i)}\prod_{s=d-i+2}^{k} n^{-1}_{d+3-s;j}\prod_{r=1}^{\ell-k} n^{-1}_{d+2+j-k-r;j+1}\Big\} \\
&\le n_{i+2;j+1}^{-(d-i)}\bigvee_{j\le k<\ell}\Big\{\prod_{u=d+3-k}^{i+1} n^{-1}_{u;j+1}\prod_{u=d+2+j-\ell}^{d+1+j-k} n^{-1}_{u;j+1}\Big\}.
\end{aligned}$$

Note that $\prod_{u=d+2+j-\ell}^{d+1+j-k} n^{-1}_{u;j+1} \le \prod_{u=d+2+j-\ell-(j-1)}^{d+1+j-k-(j-1)} n^{-1}_{u;j+1} = \prod_{u=d+3-\ell}^{d+2-k} n^{-1}_{u;j+1}$, where the inequality follows by $j \ge 1$ and $\ell \le d-j+1$, so the above display can be further bounded by

$$\mathfrak{Q}_{j,2}(\ell) \le n_{i+2;j+1}^{-(d-i)}\prod_{u=d+3-\ell}^{i+1} n^{-1}_{u;j+1} = Q_{j+1}^{-1}(\ell).$$

Hence (38) is verified and we have finished the proof for Part I.

(Part II). We deal with $\{a_\ell^{j+1}\}_{\ell=d-j+2}^{d}$ in this step. Applying Lemma 19 with $d_0 = d-1$ and $s = d-j$, the last $(j-1)$ terms in the first summand therein take the form

$$1 \gtrsim \frac{(n-n_{d+1})^3}{n^2}\Big(\overline{\beta}_{2,0}^{d-j} a_2^{j+1} + \ldots + \overline{\beta}_{2,d-j}^{d-j} a_{d-j+2}^{j+1}\Big)^2 \tag{R.2}$$

$$+ \frac{(n-n_{d+1})^5}{n^4}\Big(\overline{\beta}_{3,0}^{d-j} a_3^{j+1} + \ldots + \overline{\beta}_{3,d-j}^{d-j} a_{d-j+3}^{j+1}\Big)^2 \tag{R.3}$$

$\ldots$

$$+ \frac{(n-n_{d+1})^{2j-1}}{n^{2(j-1)}}\Big(\overline{\beta}_{j,0}^{d-j} a_j^{j+1} + \ldots + \overline{\beta}_{j,d-j}^{d-j} a_d^{j+1}\Big)^2. \tag{R.j}$$

Combining (R.2) with the estimates for $\{a_\ell^{j+1}\}_{\ell=2}^{d-j+1}$ obtained in Part I, and using Lemma 29 iteratively to cancel everything but a_{d-j+2}^{j+1}, we obtain

$$\begin{aligned}
1 &\gtrsim \frac{(n-n_{d+1})^3}{n^2}\Big(\sum_{k=0}^{d-j}\overline{\beta}_{2,k}^{d-j} a_{k+2}^{j+1}\Big)^2 \\
&\qquad + \sum_{k=2}^{d-j+1}\Big[(a_k^{j+1})^2 \cdot Q_{j+1}^2(k)\cdot M^2(n_1,n_0)\Big] \\
&\gtrsim (a_{d-j+2}^{j+1})^2\Big\{\frac{(n-n_{d+1})^3(\overline{\beta}_{2,d-j}^{d-j})^2}{n^2}\wedge \\
&\qquad \bigwedge_{k=2}^{d-j+1}\Big[Q_{j+1}^2(k) M^2(n_1,n_0)\frac{(\overline{\beta}_{2,d-j}^{d-j})^2}{(\overline{\beta}_{2,k-2}^{d-j})^2}\Big]\Big\} \\
&\equiv (a_{d-j+2}^{j+1})^2\Big\{A_j^{(2)} \wedge \bigwedge_{k=2}^{d-j+1} B_{j,k}^{(2)}\Big\}.
\end{aligned}$$

Similar to Part I, we only need to get a lower bound for $\bigwedge_{k=2}^{d-j+1} B_{j,k}^{(2)}$. As $(\overline{\beta}_{2,d-j}^{d-j})^2/(\overline{\beta}_{2,k-2}^{d-j})^2 \gtrsim \prod_{r=k-1}^{d-j} n^2_{d+2-r;j+1}$ by Lemma 33, it follows that

$$1 \gtrsim (a_{d-j+2}^{j+1})^2\bigwedge_{k=2}^{d-j+1}\Big[Q_{j+1}^2(k)\prod_{r=k-1}^{d-j} n^2_{d+2-r;j+1}\Big] M^2(n_1,n_0).$$

As $k \mapsto Q_{j+1}^2(k)\prod_{r=k-1}^{d-j} n_{d+2-r;j+1}^2 = Q_{j+1}^2(d-j+1)n_{d+3-k;j+1}^2$ is non-increasing on $k \in [2; d-j+1]$, the minimum is taken at $k = d-j+1$ in the above display. Since $Q_{j+1}^2(d-j+1)n_{j+2;j+1}^2 = Q_{j+1}^2(d-j+2)$, we arrive at

$$1 \gtrsim \left(a_{d-j+2}^{j+1}\right)^2 Q_{j+1}^2(d-j+2)M^2(n_1, n_0),$$

which is the desired estimate for a_{d-j+2}^{j+1}. Now iterate along (R.3)-(R.j) to complete the proof for Part II. This completes the proof. $\square$

Proof of Proposition 17: We shorthand $\Theta(d, d_0, k_0)$ as Θ, and the sample points will be indexed using ι. For any $\theta \in \Theta$, let $\{n_j\}_{j=0}^{k_0}$ be its knots: $0 = n_0 \le n_1 \le \ldots \le n_{k_0} = n$. The overall complexity width can then be bounded piece by piece:

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\Theta}(\varepsilon\cdot\theta)^2 &= \mathbb{E}\sup_{\theta\in\Theta}\left(\sum_{i=1}^{k_0}(\varepsilon\cdot\theta)_{(n_{i-1};n_i]}\right)^2\\ &\le C\sum_{i=1}^{k_0}\mathbb{E}\sup_{\theta\in\Theta}(\varepsilon\cdot\theta)^2_{(n_{i-1};n_i]}.\end{aligned}$$

We will prove that each summand in the above display can be bounded by a constant multiple of $\log\log(16n)$.

We start with the first piece $(n_0; n_1]$. Let $f \in \mathcal{F}_n(d, d_0, k)$ be a generating spline of θ, i.e., $\theta_\iota = f(\iota/n)$ for $\iota \in [1; n]$. For this piece, we use the following parametrization of $f(\cdot)$ slightly different from (28): for any $x \in (0, n_1/n]$,

$$f(x) = \sum_{\ell=1}^{d_0+1}\widetilde{a}_\ell^1\left(x - \frac{n_1}{n}\right)^{\ell-1} + \sum_{\ell=d_0+2}^{d+1} a_\ell^0\left(x - \frac{n_1}{n}\right)^{\ell-1}. \quad (39)$$

Then, the complexity width in question can be written as

$$\begin{aligned}(\varepsilon\cdot\theta)_{(n_0;n_1]} &= \sum_{\ell=1}^{d_0+1}\sum_{\iota\in(n_0;n_1]}\widetilde{a}_\ell^1\left(\frac{\iota-n_1}{n}\right)^{\ell-1}\varepsilon_\iota\\ &\quad + \sum_{\ell=d_0+2}^{d+1}\sum_{\iota\in(n_0;n_1]} a_\ell^0\left(\frac{\iota-n_1}{n}\right)^{\ell-1}\varepsilon_\iota.\end{aligned}$$

Applying Lemma 27 to the piece $(n_0; n_1]$, we have

$$\sum_{\ell=1}^{d_0+1}(\widetilde{a}_\ell^1)^2\frac{n_1^{2\ell-1}}{n^{2(\ell-1)}} + \sum_{\ell=d_0+2}^{d+1}(a_\ell^0)^2\frac{n_1^{2\ell-1}}{n^{2(\ell-1)}} \lesssim 1. \quad (40)$$

Thus the complexity width over the first piece $(n_0; n_1]$ can be bounded by

$$\begin{aligned}&\mathbb{E}\sup_{\theta\in\Theta}(\varepsilon\cdot\theta)^2_{(n_0;n_1]}\\ &\lesssim \sum_{\ell=1}^{d_0+1}\mathbb{E}\sup_{1\le n_1\le n}\ \sup_{(\widetilde{a}_\ell^1)^2\frac{n_1^{2\ell-1}}{n^{2(\ell-1)}}\le 1}\frac{(\widetilde{a}_\ell^1)^2}{n^{2(\ell-1)}}\\ &\qquad\left(\sum_{\iota\in(n_0;n_1]}(\iota-n_1)^{\ell-1}\varepsilon_\iota\right)^2\\ &\quad + \sum_{\ell=d_0+2}^{d+1}\mathbb{E}\sup_{1\le n_1\le n}\ \sup_{(a_\ell^0)^2\frac{n_1^{2\ell-1}}{n^{2(\ell-1)}}\le 1}\frac{(a_\ell^0)^2}{n^{2(\ell-1)}}\\ &\quad\left(\sum_{\iota\in(n_0;n_1]}(\iota-n_1)^{\ell-1}\varepsilon_\iota\right)^2\\ &\le C\log\log(16n),\end{aligned}$$

where the second inequality is due to Theorem 13 with $\psi(x) = x^2$ therein. The complexity width over the last piece $(n_{k_0-1}; n_{k_0}]$ can be handled similarly.

Starting from the second until the second last piece, we use the parametrization (28) on the piece $(n_{i+1}; n_{i+2}]$, yielding

$$(\varepsilon\cdot\theta)_{(n_{i+1};n_{i+2}]} = \sum_{\ell=1}^{d+1}\sum_{\iota\in(n_{i+1};n_{i+2}]} a_\ell^{i+1}\left(\frac{\iota-n_{i+1}}{n}\right)^{\ell-1}\varepsilon_\iota.$$

Thus the complexity width in question can be bounded by

$$\begin{aligned}&\mathbb{E}\sup_{\theta\in\Theta}(\varepsilon\cdot\theta)^2_{(n_{i+1};n_{i+2}]}\\ &\lesssim \sum_{\ell=1}^{d+1}\mathbb{E}\sup_{\theta\in\Theta}\frac{(a_\ell^{i+1})^2}{n^{2(\ell-1)}}\left(\sum_{\iota\in(n_{i+1};n_{i+2}]}(\iota-n_{i+1})^{\ell-1}\varepsilon_\iota\right)^2\\ &\lesssim \sum_{\ell=1}^{d+1}\mathbb{E}\sup_{\substack{n_{i+1}<n_{i+2},\\ (a_\ell^{i+1})^2 n_{i+2;i+1}^{2(\ell-1)}M^2(n_{i+2},n_{i+1})\le 1}}\\ &\qquad\frac{(a_\ell^{i+1})^2}{n^{2(\ell-1)}}\left(\sum_{\iota\in(n_{i+1};n_{i+2}]}(\iota-n_{i+1})^{\ell-1}\varepsilon_\iota\right)^2\\ &\le C\log\log(16n),\end{aligned}$$

where the second inequality is by plugging in the estimates a_ℓ^{i+1}, $\ell \in [1; d+1]$ from Lemma 25 (the general version of Lemma 22 with general $d_0 \in [-1; d-1]$), and the third inequality is by applying Theorem 13 with $\psi(x) = x^2$ therein. The proof is thus complete. $\square$

VI. Proof of Theorem 11, Upper Bound

A. Proof Outline

For expository purpose, we focus on the convex linear case $\Theta^*(1, k)$ with truth $\theta_0 = 0$ in (1). Using the reduction Proposition 16, the key ingredient is to show

$$\mathbb{E}\sup_{\theta^*\in\Theta^*(1,k):\|\theta^*\|\le 1}\left(\varepsilon\cdot\theta^*\right)^2 \le C\log\log(16n). \quad (41)$$

To control the complexity width, we may parametrize any $\theta^* \in \Theta^*(1, k)$ by

$$\theta_i^* = c_0 + \sum_{j=1}^{j^*} a_j\left(\frac{n_j - i}{n}\right)_+ + \sum_{j=j^*}^{k-1} b_j\left(\frac{i - n_j}{n}\right)_+, \quad (42)$$

where

- j^* is the index of the knot where the slope of the underlying convex function f^* crosses zero if it does, and is otherwise set to be k;
- $\{a_j\}$ and $\{b_j\}$ are two *non-negative* real sequences parametrizing the *change of slope*, in the two regions where f^* has negative and positive slopes, respectively.

With the parametrization (42), proving (41) then reduces to obtaining sharp estimates for $\{a_j\}, \{b_j\}$, and c_0. These estimates are obtained in rather different ways:

- For the coefficients $\{a_j\}, \{b_j\}$, the non-negativity property turns out to be the key in obtaining sharp estimates for their magnitudes. Combined with the LIL (cf. Theorem 13), these coefficients contribute the desired $\log\log(16n)$ factor to the complexity width (41).

- For the coefficient c_0, an *a priori* estimate $|c_0| \le C/\sqrt{n}$ is obtained (cf. Lemma 24) under the assumed (convexity) shape constraint and the ℓ_2 constraint on the signal. This means that the coefficient c_0 only contributes a constant factor to the complexity width (41).

It should be noted that for the larger class $\Theta(1,0,k)$ without the convexity shape constraint, a parametrization in the form of (42) still holds but *without the non-negativity constraint on* $\{a_j\}, \{b_j\}$. The lack of such sign constraints unfortunately makes this representation not quite useful in obtaining LIL for $\Theta(1,0,3)$, so a different representation (cf. (23)) and a different proof strategy (cf. Section V-A) are adopted for $\Theta(1,0,3)$.

B. Groundwork

The first result establishes a canonical parametrization for general-order d-monotone splines. By definition, the polynomial coefficient of the highest order for a d-monotone spline is increasing and thus crosses zero at most once. In the following parametrization, we choose this cross point as the pivot.

Lemma 23: For any $f^* \in \mathcal{F}_n^*(d,k)$, there exists some integer $j^* \in [0;k]$ and real sequences $\{a_j\}_{j=1}^{j^*}$, $\{b_j\}_{j=j^*}^{k-1}$, and $\{c_\ell\}_{\ell=0}^{d-1}$ such that $a_j(-1)^{d+1} \ge 0$, $b_j \ge 0$, and

$$f^*(x) = \sum_{j=1}^{j^*} a_j\Big(\frac{n_j}{n} - x\Big)_+^d + \sum_{j=j^*}^{k-1} b_j\Big(x - \frac{n_j}{n}\Big)_+^d + \sum_{\ell=0}^{d-1}\frac{c_\ell}{\ell!}x^\ell \tag{43}$$

for $x \in (0,1]$, where $\{n_j/n\}_{j=0}^k$ are the knots of f^*. On the sequence level, we have for every $\theta^* \in \Theta^*(d,k)$:

$$\theta_i^* = \sum_{j=1}^{j^*} a_j\left(\frac{n_j - i}{n}\right)_+^d + \sum_{j=j^*}^{k-1} b_j\left(\frac{i-n_j}{n}\right)_+^d + \sum_{\ell=0}^{d-1}\frac{c_\ell}{\ell!}(i/n)^\ell. \tag{44}$$

The next result generalizes the bound $|c_0| \le C/\sqrt{n}$ in the previous proof outline, indicating that all lower-order polynomial coefficients of a d-monotone spline can be well-controlled.

Lemma 24: For any $\theta^* \in \Theta^*(d,k)$ with $\|\theta^*\|^2 \le 1$, there exists some $C = C(d)$ such that, in its canonical form (44), $|c_\ell| \le C/\sqrt{n}$ for every $\ell \in [0;d-1]$.

The proof of the above lemmas can be found in Appendix D.

C. Main Proof

Proof of Theorem 11 (Upper Bound): Throughout the proof, we will shorthand $\Theta^*(d,k)$ as Θ^*. We start with a slight modification of the reduction principle in Proposition 16.

Let $L_0 \equiv n/k$ be an integer without loss of generality. Let $\theta^*_{\mathrm{oracle}}$ be an oracle in Θ^* that achieves the infimum. Let $n_j \equiv n_j(\theta^*_{\mathrm{oracle}})$, $0 \le j \le k$ be the knots of $\theta^*_{\mathrm{oracle}}$: $0 = n_0 \le n_1 \le \ldots \le n_k = n$. For each $j \in [0;k-1]$, let $m_j \equiv m_j(\theta^*_{\mathrm{oracle}}) \equiv \lceil (n_{j+1}-n_j)/L_0 \rceil$, $n_{j,p} \equiv n_{j,p}(\theta^*_{\mathrm{oracle}}) \equiv n_j + p\cdot L_0$ for $p \in [0;m_j-1]$ so that $n_{j,0} = n_j$ and $n_{j,m_j} \equiv n_{j,m_j}(\theta^*_{\mathrm{oracle}}) \equiv n_{j+1}$. Lastly, for any $\theta^* \in \Theta^*$, let $s_{j,p} \equiv s_{j,p}(\theta^*, \theta^*_{\mathrm{oracle}})$ be the number of knots of $\theta^* - \theta^*_{\mathrm{oracle}}$ on the segment $(n_{j,p}, n_{j,p+1}]$, so that $\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1} s_{j,p} \le k$. Under the above notation, define, for each $\theta \in \mathbb{R}^n$, $(\theta)_{[j,p]}$ as the sub-vector $(\theta_i)_{i\in(n_{j,p},n_{j,p+1}]}$.

Following the same line of proof as Proposition 16 on this finer resolution $\{n_{j,p}\}$, we have, for any $\delta > 0$ and then some $C = C(\delta)$,

$$\begin{aligned}\mathbb{E}_{\theta_0}\|\widehat{\theta} - \theta_0\|^2 &\le (1+\delta)\|\theta^*_{\mathrm{oracle}} - \theta_0\|^2 \\ &\quad + C\cdot\mathbb{E}\sup_{\theta^*\in\Theta^*}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\big(\varepsilon_{[j,p]}\cdot v_{j,p}(\theta^*)\big)^2,\end{aligned}$$

where $v_{j,p}(\theta^*) \equiv v_{j,p}(\theta^*;\theta^*_{\mathrm{oracle}}) \equiv (\theta^* - \theta^*_{\mathrm{oracle}})_{[j,p]}/\|(\theta^* - \theta^*_{\mathrm{oracle}})_{[j,p]}\|$.

We now prove that the second term on the right side can be bounded by a constant multiple of $k\log\log(16n/k)$. Some extra notation is hence needed. For any $\theta^* \in \Theta^*$, denote the set of $s_{j,p}$ knots of $v_{j,p}(\theta^*)$ as $n_{j,p,1},\ldots,n_{j,p,s_{j,p}}$. Also define $n_{j,p,0} \equiv n_{j,p,0}(\theta^*_{\mathrm{oracle}}) \equiv n_{j,p}$ and $n_{j,p,s_{j,p}+1} \equiv n_{j,p,s_{j,p}+1}(\theta^*_{\mathrm{oracle}}) \equiv n_{j,p+1}$. Moreover, in view of the canonical parametrization of shape-constrained splines in Lemma 23, let for each fixed $j \in [0;k-1]$ and $p \in [0;m_j]$ the index $q \equiv q(\theta^*, \theta^*_{\mathrm{oracle}}) \in [0;s_{j,p}]$ be such that, on $(n_{j,p}, n_{j,p+1}]$, $(n_{j,p,q^*-1}, n_{j,p,q^*}]$ is the last piece on which the sign of the highest order polynomial component of $\theta^* - \theta^*_{\mathrm{oracle}}$ is negative.

Under the above notation, we have $v_{j,p}(\theta^*) \in \Theta^*_{n_{j,p+1}-n_{j,p}}(d, s_{j,p}+1)$ (here we assume without loss of generality that the two end pieces of $\theta^* - \theta^*_{\mathrm{oracle}}$ adjacent to $n_{j,p}$ and $n_{j,p+1}$ also have length at least $d+1$ since there are at most $2k$ such pieces and each only contributes a constant factor to the complexity width). Thus Lemma 23 entails that there exist real sequences $\{c_{j,p,\ell}\} \equiv \{c_{j,p,\ell}(\theta^*, \theta^*_{\mathrm{oracle}})\}$, and some $q^* \in [1;s_{j,p}]$ along with sequences of equal sign $\{a_{j,p,q}\}_{q=1}^{q^*} \equiv \{a_{j,p,q}(\theta^*,\theta^*_{\mathrm{oracle}})\}_{q=1}^{q^*}$, $\{b_{j,p,q}\}_{q=q^*}^{s_{j,p}} \equiv \{b_{j,p,q}(\theta^*,\theta^*_{\mathrm{oracle}})\}_{q=q^*}^{s_{j,p}}$ such that

$$\begin{aligned}\big(v_{j,p}(\theta^*)\big)_i &= \sum_{q=1}^{q^*} a_{j,p,q}\left(\frac{n_{j,p,q} - (i+n_{j,p})}{n}\right)_+^d \\ &\quad + \sum_{q=q^*}^{s_{j,p}} b_{j,p,q}\left(\frac{(i+n_{j,p}) - n_{j,p,q}}{n}\right)_+^d \\ &\quad + \sum_{\ell=0}^{d-1}\frac{c_{j,p,\ell}}{\ell!}\left(\frac{i-n_{j,p}}{n}\right)^\ell \\ &\equiv \big(v^1_{j,p}(\theta^*)\big)_i + \big(v^2_{j,p}(\theta^*)\big)_i,\end{aligned} \tag{45}$$

where $(v^2_{j,p}(\theta^*))_i \equiv \sum_{\ell=0}^{d-1} c_{j,p,\ell}\big((i-n_{j,p})/n\big)^\ell/\ell!$. Therefore, we have

$$\begin{aligned}&\mathbb{E}\sup_{\theta^*\in\Theta^*}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\big(\varepsilon_{[j,p]}\cdot v_{j,p}(\theta^*)\big)^2 \\ &\le 2\Big(\mathbb{E}\sup_{\theta^*\in\Theta^*}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\big(\varepsilon_{[j,p]}\cdot v^1_{j,p}(\theta^*)\big)^2 \\ &\qquad + \mathbb{E}\sup_{\theta^*\in\Theta^*}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\big(\varepsilon_{[j,p]}\cdot v^2_{j,p}(\theta^*)\big)^2\Big) \\ &\equiv 2\Big((I) + (II)\Big).\end{aligned}$$

We first upper bound (II). Since for each j,p and $\theta^* \in \Theta^*$, $v_{j,p}(\theta^*) \in \Theta^*_{n_{j,p+1}-n_{j,p}}(d, s_{j,p}+1)$ and has unit norm, Lemma 24 entails that there exists some $C = C(d)$ such that $|c_{j,p,\ell}| \le C/\sqrt{n_{j,p+1}-n_{j,p}}$ for $j \in [0;k-1]$, $p \in [0;m_j]$, and $\ell \in [0;d-1]$. Let $\Delta n_{j,p} \equiv n_{j,p+1}-n_{j,p}$. Then, we have

$$\begin{aligned}(II) &\le C \cdot \mathbb{E} \sup_{|c_{j,p,\ell}| \le C/\sqrt{\Delta n_{j,p}}} \sum_{j=0}^{k-1} \sum_{p=0}^{m_j-1} \sum_{\ell=0}^{d-1} \\ &\qquad \frac{c_{j,p,\ell}^2}{n^{2\ell}(\ell!)^2} \bigg(\sum_{i \in (n_{j,p}; n_{j,p+1}]} (i-n_{j,p})^\ell \varepsilon_i \bigg)^2 \\ &\le C \cdot \sum_{j=0}^{k-1} \sum_{p=0}^{m_j-1} \sum_{\ell=0}^{d-1} (\Delta n_{j,p})^{-1} \\ &\qquad \frac{\mathbb{E}\big[\sum_{i \in (n_{j,p}; n_{j,p+1}]} (i-n_{j,p})^\ell \varepsilon_i\big]^2}{n^{2\ell}} \\ &\le C \cdot \sum_{j=0}^{k-1} \sum_{p=0}^{m_j-1} 1 = C \cdot \sum_{j=0}^{k-1} m_j \le Ck.\end{aligned}$$

Next, we bound (I). Some extra notation is needed. Define the following partition of $(n_{j,p}; n_{j,p+1}]$ with intervals

$$\begin{aligned}I^B_{j,p,\ell} \equiv \bigg(& n_{j,p} + \Big\lceil (1-2^{-(\ell-1)})\Delta n_{j,p} \Big\rceil ; \\ & n_{j,p} + \big\lceil (1-2^{-\ell})\Delta n_{j,p} \big\rceil \bigg]\end{aligned}$$

for $\ell \in [1; t_{j,p}]$ and $t_{j,p} \equiv \lceil \log_2 \Delta n_{j,p} \rceil$, and similarly,

$$\begin{aligned}I^A_{j,p,\ell} \equiv \bigg(& n_{j,p+1} - \big\lceil (1-2^{-\ell})\Delta n_{j,p} \big\rceil ; \\ & n_{j,p+1} - \Big\lceil (1-2^{-(\ell-1)})\Delta n_{j,p} \Big\rceil \bigg].\end{aligned}$$

From this definition, we immediately have (with analogous conclusions for $I^A_{j,p,\ell}$): (i) $|I^B_{j,p,\ell}| \le \lceil 2^{-\ell}\Delta n_{j,p} \rceil$; (ii) $2(\sum_{\ell > \ell_0} |I^B_{j,p,\ell}| + 1) \ge \sum_{\ell \ge \ell_0} |I^B_{j,p,\ell}|$ for any $\ell_0 \in [1; t_{j,p}]$. Then, let

$$\begin{aligned}B_{j,p,\ell} &\equiv B_{j,p,\ell}(\theta^*, \theta^*_{\text{oracle}}) \equiv \sum_{q=q^*}^{s_{j,p}} b_{j,p,\ell} \mathbf{1}_{n_{j,p,q} \in I^B_{j,p,\ell}}, \\ \delta^B_{j,p,\ell} &\equiv \delta^B_{j,p,\ell}(\theta^*, \theta^*_{\text{oracle}}) \equiv \max \Big\{ q^* \le q \le s_{j,p} : \mathbf{1}_{n_{j,p,q} \in I^B_{j,p,\ell}} \Big\}.\end{aligned}$$

In words, $\delta^B_{j,p,\ell}$ equals to 1 if and only if among the knots $\{n_{j,p,q}\}_{q=q^*}^{s_{j,p}}$, there is at least one that lies in the interval $I^B_{j,p,\ell}$, and if such is the case, $B_{j,p,\ell}$ returns the block sum. We omit the similar definitions for $A_{j,p,\ell}$ and $\delta^A_{j,p,\ell}$. By definition, we immediately have $\sum_{\ell=1}^{t_{j,p}} \delta^B_{j,p,\ell} \le s_{j,p}$.

In the parametrization (45), using the constraint $\|v_{j,p}(\theta^*)\| \le 1$ and the bounds $|c_{j,p,\ell}| \le C/\sqrt{n_{j,p+1}-n_{j,p}}$ for $\ell \in [0; d-1]$, we have $\|v^1_{j,p}(\theta^*)\| \le C$ (recall the definition of $v^1_{j,p}$ in (45)) for some $C = C(d)$. Hence for some sufficiently small $c = c(d)$,

$$\begin{aligned}1 &\ge c \cdot \sum_{i \in (n_{j,p}; n_{j,p+1}]} \bigg[\sum_{q=1}^{q^*} a_{j,p,q} \Big(\frac{n_{j,p,q}-i}{n} \Big)_+^d \\ &\qquad + \sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \Big(\frac{i-n_{j,p,q}}{n} \Big)_+^d \bigg]^2 \\ &\ge c \cdot \sum_{i \in (n_{j,p}; n_{j,p+1}]} \bigg[\sum_{q=1}^{q^*} a_{j,p,q} \Big(\frac{n_{j,p,q}-i}{n} \Big)_+^d \bigg]^2 \\ &\qquad \vee \bigg[\sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \Big(\frac{i-n_{j,p,q}}{n} \Big)_+^d \bigg]^2,\end{aligned}$$

where the second inequality follows from the fact that the interaction term between the two summands in the first inequality is 0 for each i.

Now, starting from the constraint $1 \ge c \cdot \sum_{i \in (n_{j,p}; n_{j,p+1}]} \big[\sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \big(\frac{i-n_{j,p,q}}{n} \big)_+^d \big]^2$, we will obtain estimates for $B_{j,p,\ell}$. Fix j,p. By the disjointness of $I^B_{j,p,\ell}$ and the non-negativeness of $\{b_{j,p,q}\}$, we have

$$\begin{aligned}1 &\ge c \cdot \sum_{i \in (n_{j,p}; n_{j,p+1}]} \\ &\qquad \bigg[\sum_{\ell=1}^{t_{j,p}} \sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \mathbf{1}_{n_{j,p,q} \in I^B_{j,p,\ell}} \Big(\frac{i-n_{j,p,q}}{n} \Big)_+^d \bigg]^2 \\ &\ge c \cdot \sum_{i \in (n_{j,p}; n_{j,p+1}]} \\ &\qquad \bigg[\sum_{\ell=1}^{t_{j,p}} \sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \mathbf{1}_{n_{j,p,q} \in I^B_{j,p,\ell}} \Big(\frac{i-(I^B_{j,p,\ell})_+}{n} \Big)_+^d \bigg]^2 \\ &= c \cdot \sum_{i \in (n_{j,p}; n_{j,p+1}]} \bigg[\sum_{\ell=1}^{t_{j,p}} B_{j,p,\ell} \Big(\frac{i-(I^B_{j,p,\ell})_+}{n} \Big)_+^d \bigg]^2 \\ &\ge c \cdot \sum_{\ell=1}^{t_{j,p}} B^2_{j,p,\ell} \sum_{i \in (n_{j,p}; n_{j,p+1}]} \Big(\frac{i-(I^B_{j,p,\ell})_+}{n} \Big)_+^{2d} \\ &\ge c \cdot \sum_{\ell=1}^{t_{j,p}} B^2_{j,p,\ell} \frac{(n_{j,p+1}-(I^B_{j,p,\ell})_+)^{2d+1}}{n^{2d}} \\ &\ge c \cdot \sum_{\ell=1}^{t_{j,p}} B^2_{j,p,\ell} \frac{(n_{j,p+1}-(I^B_{j,p,\ell})_-)^{2d+1}}{n^{2d}}, \end{aligned} \tag{46}$$

where $(I^B_{j,p,\ell})_+$ ($(I^B_{j,p,\ell})_-$) is defined to be the right (left) endpoint of $I^B_{j,p,\ell}$, and the last inequality follows from property (ii) of the partition $I^B_{j,p,\ell}$.

We are now ready to bound the term (I). First by the vanishing of interaction terms, we have $(I) = (I_1) + (I_2)$, where

$$\begin{aligned}(I_1) &\equiv \mathbb{E} \sup_{\theta^* \in \Theta^*} \sum_{j=0}^{k-1} \sum_{p=0}^{m_j-1} \\ &\qquad \bigg[\sum_{q=0}^{q^*} a_{j,p,q} \bigg(\sum_{i \in (n_{j,p}; n_{j,p,q}]} \Big(\frac{n_{j,p,q}-i}{n} \Big)_+^d \varepsilon_i \bigg) \bigg]^2, \\ (I_2) &\equiv \mathbb{E} \sup_{\theta^* \in \Theta^*} \sum_{j=0}^{k-1} \sum_{p=0}^{m_j-1} \\ &\qquad \bigg[\sum_{q=q^*}^{s_{j,p}} b_{j,p,q} \bigg(\sum_{i \in (n_{j,p,q}; n_{j,p+1}]} \Big(\frac{i-n_{j,p,q}}{n} \Big)_+^d \varepsilon_i \bigg) \bigg]^2.\end{aligned}$$

Due to symmetry, we only bound (I_2) as follows:

$$(I_2) = \mathbb{E}\sup_{\theta^*}\sum_{j,p}\Bigg[\sum_{q=q^*}^{s_{j,p}}\sum_{\ell=1}^{t_{j,p}}\mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}}b_{j,p,q}\Bigg(\sum_{i\in(n_{j,p,q};n_{j,p+1}]}\left(\frac{i-n_{j,p,q}}{n}\right)^d\varepsilon_i\Bigg)\Bigg]^2$$
$$\le \mathbb{E}\sup_{\theta^*}\sum_{j,p}\Bigg[\sum_{\ell=1}^{t_{j,p}}\Bigg\{\sum_{q=q^*}^{s_{j,p}}\mathbf{1}_{n_{j,p,q}\in I^B_{j,p,\ell}}b_{j,p,q}\Bigg\}\max_{\tau\in I^B_{j,p,\ell}}\Bigg|\sum_{i\in(\tau;n_{j,p+1}]}\left(\frac{i-\tau}{n}\right)^d\varepsilon_i\Bigg|\Bigg]^2$$
$$= \mathbb{E}\sup_{\theta^*}\sum_{j,p}\Bigg[\sum_{\ell=1}^{t_{j,p}}B_{j,p,\ell}\max_{\tau\in I^B_{j,p,\ell}}\Bigg|\sum_{i\in(\tau;n_{j,p+1}]}\left(\frac{i-\tau}{n}\right)^d\varepsilon_i\Bigg|\Bigg]^2$$
$$\le \mathbb{E}\max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in(\tau;n_{j,p+1}]}(i-\tau)^d\varepsilon_i\right)^2}{(n_{j,p+1}-(I^B_{j,p,\ell})_-)^{2d+1}}.$$

Here, the first inequality follows from the non-negativity of $\{b_{j,p,q}\}$, the second equality follows from the definition of $B_{j,p,\ell}$, and the last inequality follows from Cauchy-Schwarz along with the estimates for $B_{j,p,\ell}$ in (46). Furthermore, we define

$$\Delta^B \equiv \Bigg\{\{\delta^B_{j,p,\ell}\}:\delta^B_{j,p,\ell}\in\{0,1\},\sum_{j=1}^{k}\sum_{p=1}^{m_j}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\le k\Bigg\}$$

to be the admissible set for the sequence $\{\delta^B_{j,p,\ell}\}$. As $\sum_{j=1}^k\sum_{p=1}^{m_j}\sum_{\ell=1}^{t_{j,p}}1 = \sum_{j=1}^k\sum_{p=1}^{m_j}\lceil\log_2(n_{j,p+1}-n_{j,p})\rceil \le Ck\lceil\log_2(n/k)\rceil$, a combinatorial estimate yields that $|\Delta^B| \le \binom{Ck\lceil\log_2(n/k)\rceil}{k} \le (Ce\lceil\log_2(n/k)\rceil)^k$.

Now, using the basic inequality $(a+b)^2\le 2(a^2+b^2)$, it suffices to bound by the order $k\log\log(16n/k)$ the following two terms:

$$\mathbb{E}\max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\cdot\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\right)^2}{\left(n_{j,p+1}-(I^B_{j,p,\ell})_-\right)^{2d+1}} \tag{47}$$

and

$$\mathbb{E}\max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\cdot\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in((I^B_{j,p,\ell})_+;n_{j,p+1}]}(i-\tau)^d\varepsilon_i\right)^2}{\left(n_{j,p+1}-(I^B_{j,p,\ell})_-\right)^{2d+1}}. \tag{48}$$

From here on, in view of Theorem 13, the proof is essentially the same as that of Lemma 5.2 in [26] (our (47) and (48) correspond to their (42) and (43)). For the sake of completeness, we will present the proof for the bound of (47); the bound for (48) follows from essentially the proof of (43) in [26].

Denote the variable in (47) as Z, i.e.,

$$Z \equiv \max_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\cdot\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\right)^2}{\left(n_{j,p+1}-(I^B_{j,p,\ell})_-\right)^{2d+1}}.$$

We bound the tail probability of Z as follows. For any $u\ge 0$ and small enough $c>0$,

$$\mathbb{P}(Z>u)$$
$$\le \sum_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}\mathbb{P}\Bigg[\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\cdot\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\right)^2}{\left(n_{j,p+1}-(I^B_{j,p,\ell})_-\right)^{2d+1}}>u\Bigg]$$
$$\le \sum_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}e^{-cu}\prod_{j,p,\ell}\cdot\mathbb{E}\exp\Bigg[c\delta^B_{j,p,\ell}\max_{\tau\in I^B_{j,p,\ell}}\frac{\left(\sum_{i\in(\tau;(I^B_{j,p,\ell})_+]}(i-\tau)^d\varepsilon_i\right)^2}{\left(n_{j,p+1}-(I^B_{j,p,\ell})_-\right)^{2d+1}}\Bigg]$$
$$\lesssim \sum_{\{\delta^B_{j,p,\ell}\}\in\Delta^B}e^{-cu}\cdot\exp\Bigg(\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}C\delta^B_{j,p,\ell}\log\log\left(16(n_{j,p+1}-n_{j,p})\right)\Bigg)$$
$$\le \exp\left(\log|\Delta^B|-cu+Ck\log\log(16n/k)\right)$$
$$\le \exp(-cu+Ck\log\log(16n/k)).$$

Here, the second inequality follows from the independence of the partial sum processes over the partition $\{I^B_{j,p,\ell}\}$, the third inequality follows by choosing c to be sufficiently small and then applying Theorem 13 with $\psi(x)=\exp(cx^2)-1$ therein, and the fourth inequality follows from the fact that $n_{j,p+1}-n_{j,p}\le n/k$ and that $\sum_{j=0}^{k-1}\sum_{p=0}^{m_j-1}\sum_{\ell=1}^{t_{j,p}}\delta^B_{j,p,\ell}\le k$ for any $\{\delta^B_{j,p,\ell}\}\in\Delta^B$. The proof is now complete by integrating the tail estimate. □

Appendix A
Proof of Lower Bounds

A. Lower Bound in Section II

Proof of Proposition 2: We start with the first claim. In view of the fact that minimax rate over $\Theta(d,d_0,k)$ is non-decreasing in k and $\Theta(d,d-1,k)\subset\Theta(d,d_0,k)$ for any $d_0\in[-1;d-1]$, it suffices to show that

$$\inf_{\widetilde{\theta}}\sup_{\theta\in\Theta(d,d-1,2)}\mathbb{E}_\theta\|\widetilde{\theta}-\theta\|^2\ge c\log\log(16n).$$

For this, we will apply a standard reduction argument to multiple hypothesis testing (cf. Theorem 2.5 of [44]). Define the following series of splines. Let $M \equiv \lfloor \log_2(n/(d+1)) \rfloor$, and for each $\ell \in [1;M]$, $\tau_\ell \equiv \lfloor (1-2^{-\ell})n \rfloor$ and $f^\ell(x) \equiv \alpha_\ell (x-\tau_\ell/n)_+^d$ with $\alpha_\ell \equiv c(2^\ell)^{(2d+1)/2}\sqrt{\log\log(16n)/n}$ for some sufficiently small c. Further define $f^0(x) \equiv 0$ on $[0,1]$, and the induced vectors $\theta_i^\ell \equiv f^\ell(i/n)$ for $i \in [1;n]$ and $\ell \in [0;M]$. Denote the corresponding joint distribution of $\{Y_i\}_{i=1}^n$ under the experiment (1) with truth θ^ℓ as P_ℓ, $\ell \in [0;M]$. It can be readily verified that $\theta^\ell \in \Theta(d, d-1, 2)$, and the Kullback-Leibler divergence between P_0 and each P_ℓ, denoted as $\mathrm{KL}(P_0, P_\ell)$, satisfies

$$\mathrm{KL}(P_0, P_\ell) = \|\theta^0 - \theta^\ell\|^2/2 = \|\theta^\ell\|^2/2 \asymp \log\log(16n)$$

for every $\ell \in [1;M]$. Moreover, for any $1 \le j < k \le M$, it holds by direct calculation that

$$\begin{aligned} d(P_j, P_k) \equiv \|\theta^j - \theta^k\|^2 &\ge \sum_{i \in (\tau_j, \tau_k]} (\theta_i^j - \theta_i^k)^2 \\ &\asymp \alpha_j^2 \frac{(\tau_k - \tau_j)^{2d+1}}{n^{2d}} \asymp \alpha_j^2 \frac{(n-\tau_j)^{2d+1}}{n^{2d}} \\ &\asymp \alpha_j^2 \frac{2^{-j(2d+1)}}{n} \asymp \log\log(16n). \end{aligned}$$

Theorem 2.5 in [44] therefore entails the desired lower bound.

Next, we prove the second claim. By following the same reduction as in the previous claim, it suffices to show that for any $k \ge k_0+1$, there exists some nonzero $f \in \mathcal{F}_n(d, d_0, k_0+1)$ such that $f(x) = 0$ for $x \in [0,c] \cup [1-c, 1]$ with some universal c. Take $c = 1/3$. Let $\tau_0 \equiv 0$, $\tau_j \equiv 1/3 + (j-1)/(3(k_0-1))$ for $j \in [1;k_0]$, and $\tau_{k_0+1} \equiv 1$. Define

$$f(x) \equiv \Big(\sum_{j=1}^{k_0-1} \sum_{\ell=d_0+1}^{d} c_\ell^j (x-\tau_j)_+^\ell \Big) \cdot \mathbf{1}_{[1/3, 2/3]}(x), \quad x \in [0,1].$$

By definition, f vanishes on $[0,1/3] \cup [2/3, 1]$. Moreover, it can be readily checked that, for any real sequence $\{c_\ell^j\}_{j \in [1;k_0-1], \ell \in [d_0+1;d]}$, $f^{(\ell)}((\tau_j)_-) = f^{(\ell)}((\tau_j)_+)$ for $j \in [1;k_0-1]$ and $\ell \in [0;d_0]$. Therefore, in order to show that $f \in \mathcal{F}_n(d, d_0, k_0+1)$ and is non-zero, it suffices to show that there exists a non-zero realization of the sequence $\{c_\ell^j\}_{j \in [1;k_0-1], \ell \in [d_0+1;d]}$ such that $f^{(\ell)}((\tau_{k_0})_-) = f^{(\ell)}((\tau_{k_0})_+) = 0$ for all $\ell \in [0;d_0]$. This is equivalent to finding a non-zero solution for the homogeneous linear system $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{b}$, where $\boldsymbol{c} \equiv \{c_\ell^j\}_{j \in [1;k_0-1], \ell \in [d_0+1;d]} \in \mathbb{R}^{(k_0-1)(d-d_0)}$, $\boldsymbol{b} \equiv \boldsymbol{0}_{(k_0-1)(d-d_0)}$, and

$$\boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots \boldsymbol{A}_{k_0-1} \end{bmatrix}$$

with $\boldsymbol{A}_j$, shown at the bottom of the page, and $\tau_{j_1,j_2} \equiv \tau_{j_1} - \tau_{j_2}$. Note that the coefficient matrix $\boldsymbol{A}$ has d_0+1 rows and $(k_0-1)(d-d_0)$ columns, where, by definition of k_0,

$$(k_0-1)(d-d_0) \ge d_0+2 \iff \Big\lfloor \frac{d+1}{d-d_0} \Big\rfloor + 1 \ge \frac{d+2}{d-d_0}.$$

The above equivalence indeed holds since if $(d+1)/(d-d_0)$ is an integer, then

$$\Big\lfloor \frac{d+1}{d-d_0} \Big\rfloor + 1 = \frac{d+1+(d-d_0)}{d-d_0} \ge \frac{d+2}{d-d_0},$$

and if not

$$\Big\lfloor \frac{d+1}{d-d_0} \Big\rfloor + 1 \ge \Big\lceil \frac{d+1}{d-d_0} \Big\rceil \ge \frac{d+2}{d-d_0}.$$

This entails that the solution space of the linear system $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{b}$ is of dimension at least one and thus the system is guaranteed to have a non-trivial solution. The proof is thus complete. □

B. Lower Bound in Section III

Proof of Proposition 9: We will continue to adopt the standard reduction to multiple testing (cf. Theorem 2.5 of [44]) as in the proof of Proposition 2. We first introduce a set of basis functions. Let $\widetilde{k} \equiv k/3$ which we assume without loss of generality to be an integer, $\ell_0 \equiv \lfloor \log_2(n/(2\widetilde{k})) \rfloor$, and $\tau_\ell \equiv (1-2^{-(\ell-1)})/\widetilde{k}$ for $\ell \in [1;\ell_0+1]$. Next, for $x \in [0, 1/\widetilde{k}]$, let $\widetilde{f}_\ell(x) \equiv c(2^{\ell-1})^{3/2}\sqrt{\log\log(16n/k)/n}(x-\tau_\ell)_+$ for $\ell \in [1;\ell_0]$ and $f_{\mathrm{ref}}(x) \equiv c(2^{\ell_0})^{3/2}\sqrt{\log\log(16n/k)/n}(x-\tau_{\ell_0+1})_+$ (here the subscript "ref" stands for "reference" and f_{ref} will be pieced together later to be the true signal underlying the distribution P_0 in Theorem 2.5 of [44]). Then let $f_\ell(x) \equiv \widetilde{f}_\ell(x) \vee f_{\mathrm{ref}}(x)$, and it can be verified that $f_\ell(x) = \widetilde{f}_\ell(x)$ on $[0, \tau_{\ell_0+1}]$. The above set of functions resembles those constructed in the proof of Proposition 2, and satisfies the similar properties

$$\sum_{i:(i/n) \in (0, 1/\widetilde{k}]} \big(f_\ell(i/n) - f_{\ell'}(i/n)\big)^2 \ge c\log\log(16n/k) \tag{49}$$

for any $1 \le \ell \ne \ell' \le \ell_0$, and

$$\begin{aligned} &\sum_{i:(i/n) \in (0, 1/\widetilde{k}]} \big(f_\ell(i/n) - f_{\mathrm{ref}}(i/n)\big)^2 \\ &\le \sum_{i:(i/n) \in (0, 1/\widetilde{k}]} \big(f_\ell(i/n)\big)^2 \\ &\le 2\Big(\sum_{i:(i/n) \in (0, 1/\widetilde{k}]} \big(\widetilde{f}_\ell(i/n)\big)^2 + \sum_{i:(i/n) \in (0, 1/\widetilde{k}]} \big(f_{\mathrm{ref}}(i/n)\big)^2 \Big) \\ &\le C\log\log(16n/k). \end{aligned} \tag{50}$$

We now construct the hypotheses in the multiple testing framework. For $j \in [1;\widetilde{k}]$, let $f_\ell^j(\cdot), f_{\mathrm{ref}}^j(\cdot)$ be a set of functions defined on $[(j-1)/\widetilde{k}, j/\widetilde{k}]$ as follows. Let $f_\ell^1(x) \equiv f_\ell(x)$ and $f_{\mathrm{ref}}^1(x) \equiv f_{\mathrm{ref}}(x)$ as defined above. Next, for $j \in [2;k]$, we define inductively $f_{\mathrm{ref}}^j(x) \equiv f_{\mathrm{ref}}^{j-1}(x) + f_{\mathrm{ref}}(x-(j-1)/\widetilde{k})$,

$$\boldsymbol{A}_j \equiv \begin{bmatrix} \underline{\odot}(d_0+1;0)\tau_{k_0,j}^{d_0+1} & \underline{\odot}(d_0+2;0)\tau_{k_0,j}^{d_0+2} & \ldots & \underline{\odot}(d;0)\tau_{k_0,j}^{d} \\ \underline{\odot}(d_0+1;1)\tau_{k_0,j}^{d_0} & \underline{\odot}(d_0+2;1)\tau_{k_0,j}^{d_0+1} & \ldots & \underline{\odot}(d;1)\tau_{k_0,j}^{d-1} \\ & & \ldots & \\ \underline{\odot}(d_0+1;d_0)\tau_{k_0,j} & \underline{\odot}(d_0+2;d_0)\tau_{k_0,j}^{2} & \ldots & \underline{\odot}(d;d_0)\tau_{k_0,j}^{d-d_0} \end{bmatrix}$$

where $f_{\mathrm{ref}}^{j-1}(x)$ for $x \in [(j-1)/\widetilde{k}, j/\widetilde{k}]$ is to be understood as the extension from $[(j-2)/\widetilde{k}, (j-1)/\widetilde{k}]$. Also define $f_\ell^j(x) \equiv f_{\mathrm{ref}}^j(x) + f_\ell(x-(j-1)/\widetilde{k})$. Lastly, we piece them together as

$$f^0(x) \equiv \sum_{j=1}^{\widetilde{k}} f_{\mathrm{ref}}^j(x)\mathbf{1}_{((j-1)/\widetilde{k}, j/\widetilde{k}]}(x)$$

and

$$f^{\boldsymbol{\ell}}(x) \equiv \sum_{j=1}^{\widetilde{k}} f_{\ell_j}^j(x)\mathbf{1}_{((j-1)/\widetilde{k}, j/\widetilde{k}]}(x),$$

where $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_{\widetilde{k}})^\top \in [1;\ell_0]^{\widetilde{k}}$. One can readily verify that all of the f^0 and $f^{\boldsymbol{\ell}}$ belong to the class $\mathcal{F}_n^*(1,k)$. Indeed, continuity follows directly from the construction and since there are at most 3 pieces on each of $[(j-1)/\widetilde{k}, j/\widetilde{k}]$, there will be at most $3\widetilde{k} = k$ pieces in total. Therefore, the sequence counterparts $\theta^0 \equiv (f^0(i/n))_i$ and $\theta^{\boldsymbol{\ell}} \equiv (f^{\boldsymbol{\ell}}(i/n))_i$ belong to $\Theta^*(1,k)$.

Let $\rho(\cdot,\cdot)$ denote the Hamming distance. Then, the Gilbert-Varshamov bound (cf. Theorems 5.1.7 and 5.1.9 in [45]) entails that with some small $c > 0$, there exists a subset $\mathcal{S} \subset [1;\ell_0]^{\widetilde{k}}$ with cardinality $|\mathcal{S}| \asymp \ell_0^{c\widetilde{k}}$ such that $\rho(\boldsymbol{\ell}, \boldsymbol{\ell}') \geq c\widetilde{k}$ for any $\boldsymbol{\ell} \neq \boldsymbol{\ell}' \in \mathcal{S}$. Adopting those in $\mathcal{S}$ as the truth in the experiment (1), we obtain a total of $M \equiv 1 + |\mathcal{S}| \asymp \ell_0^{c\widetilde{k}}$ hypotheses, which we denote as P^0 and $P^{\boldsymbol{\ell}}$, $\boldsymbol{\ell} \in \mathcal{S}$.

It remains to verify: (i) $\|\theta^{\boldsymbol{\ell}} - \theta^{\boldsymbol{\ell}'}\|^2 \geq ck\log\log(16n/k)$ for any $\boldsymbol{\ell} \neq \boldsymbol{\ell}' \in \mathcal{S}$; (ii) $\mathrm{KL}(P^0, P^{\boldsymbol{\ell}}) \leq C\log|\mathcal{S}|$ for any $\boldsymbol{\ell} \in \mathcal{S}$. We first verify (i). By definition of $\theta^{\boldsymbol{\ell}}$ and $\theta^{\boldsymbol{\ell}'}$, on each $[(j-1)/\widetilde{k}, j/\widetilde{k}]$ such that $\ell_j \neq \ell_j'$, we have by (49),

$$\begin{aligned}
&\sum_{i: \frac{i}{n} \in (\frac{j-1}{\widetilde{k}}, \frac{j}{\widetilde{k}}]} (\theta_i^{\boldsymbol{\ell}} - \theta_i^{\boldsymbol{\ell}'})^2 \\
&= \sum_{i: \frac{i}{n} \in (\frac{j-1}{\widetilde{k}}, \frac{j}{\widetilde{k}}]} \left[f_{\ell_j}\left(\frac{i}{n} - \frac{j-1}{\widetilde{k}}\right) - f_{\ell_j'}\left(\frac{i}{n} - \frac{j-1}{\widetilde{k}}\right)\right]^2 \\
&= \sum_{i: \frac{i}{n} \in (0, \frac{1}{\widetilde{k}}]} \left[f_{\ell_j}\left(\frac{i}{n}\right) - f_{\ell_j'}\left(\frac{i}{n}\right)\right]^2 \geq c\log\log(16n/k).
\end{aligned}$$

This entails that

$$\|\theta^{\boldsymbol{\ell}} - \theta^{\boldsymbol{\ell}'}\|^2 \geq \rho(\boldsymbol{\ell}, \boldsymbol{\ell}')c\log\log(16n/k) \geq ck\log\log(16n/k).$$

Similarly, for (ii), we have by (50)

$$\begin{aligned}
\mathrm{KL}(P^0, P^{\boldsymbol{\ell}}) &= \|\theta^0 - \theta^{\boldsymbol{\ell}}\|^2/2 \\
&\leq C\widetilde{k} \cdot \sum_{i: \frac{i}{n} \in (0, \frac{1}{\widetilde{k}}]} \left[f_{\ell_j}\left(\frac{i}{n}\right) - f_{\mathrm{ref}}\left(\frac{i}{n}\right)\right]^2 \\
&\leq C\widetilde{k}\log\log(16n/k) \asymp \log|\mathcal{S}|.
\end{aligned}$$

Application of Theorem 2.5 in [44] then completes the proof. $\square$

Proof of Theorem 11 (Lower Bound): This is immediate by realizing that $\Theta^*(d,2) \subset \Theta^*(d,k)$ for $k \geq 2$ and the lower bound construction in the first part of the proof of Proposition 2 can be directly applied to establish a lower bound for $\Theta^*(d,2)$. $\square$

APPENDIX B
PROOF OF THEOREM 13

Proof of Theorem 13: We first claim that there exists some $c = c(d)$ such that for any $t > 0$, the event

$$\begin{aligned}
\mathcal{E}_1 \equiv \Bigg\{ &\max_{1 \leq n_1 < n_2 \leq n} (n_2 - n_1)^{-d}(n_2 \wedge (n-n_1))^{-1/2} \\
&\bigg|\sum_{(n_1;n_2]} (i-n_1)^d \varepsilon_i\bigg| \geq t \Bigg\}
\end{aligned}$$

is contained in the event

$$\mathcal{E}_2 \equiv \left\{\max_{1 \leq n_1 < n_2 \leq n} (n_2 \wedge (n-n_1))^{-1/2}\bigg|\sum_{(n_1;n_2]} \varepsilon_i\bigg| \geq ct\right\}.$$

On $\mathcal{E}_2^c$, for any $1 \leq n_1 < n_2 \leq n$, it holds that $\left|\sum_{(n_1;n_2]} \varepsilon_i\right| \leq c(n_2 \wedge (n-n_1))^{1/2}t$. Then,

$$\begin{aligned}
&\bigg|\sum_{(n_1;n_2]} \varepsilon_i (i-n_1)^d\bigg| \\
&= \bigg|\sum_{i \in (n_1;n_2]} \varepsilon_i \sum_{j=1}^{i-n_1} \left(j^d - (j-1)^d\right)\bigg| \\
&= \bigg|\sum_{j=1}^{n_2-n_1} \left(j^d - (j-1)^d\right) \sum_{i \in [n_1+j;n_2]} \varepsilon_i\bigg| \\
&\leq \sum_{\ell=0}^{d-1} \binom{d}{\ell} \sum_{j=1}^{n_2-n_1} (j-1)^\ell \bigg|\sum_{i \in [n_1+j;n_2]} \varepsilon_i\bigg| \\
&\leq ct \cdot \sum_{\ell=0}^{d-1} \binom{d}{\ell} \sum_{j=1}^{n_2-n_1} (j-1)^\ell (\sqrt{n_2} \wedge \sqrt{n-n_1-(j-1)}) \\
&\leq 2ct \cdot \sum_{\ell=0}^{d-1} \binom{d}{\ell} \int_0^{n_2-n_1} x^\ell (\sqrt{n_2} \wedge \sqrt{n-n_1-x})\mathrm{d}x \quad (51) \\
&\leq 4ct \cdot \sum_{\ell=0}^{d-1} \binom{d}{\ell} (n_2-n_1)^{\ell+1}(n_2 \wedge (n-n_1))^{1/2} \quad (52) \\
&\leq c2^{d+2}t \cdot (n_2-n_1)^d (n_2 \wedge (n-n_1))^{1/2},
\end{aligned}$$

where the inequality (51) follows from the fact that the map $x \mapsto x^\ell(\sqrt{n_2} \wedge \sqrt{n-n_1-x})$ first increases and then decreases on $[0, n-n_1]$, and the inequality (52) follows from a separate discussion of $n_2 \leq n-n_1$ and $n_2 > n-n_1$ and the following two bounds: $\int_0^{n_2-n_1} x^\ell \,\mathrm{d}x = (\ell+1)^{-1}(n_2-n_1)^{\ell+1}$ and

$$\begin{aligned}
&\int_0^{n_2-n_1} x^\ell \sqrt{n-n_1-x}\,\mathrm{d}x \leq (n_2-n_1)^\ell \int_{n_1}^{n_2} \sqrt{n-x}\,\mathrm{d}x \\
&= (n_2-n_1)^\ell \int_{n-n_2}^{n-n_1} \sqrt{x}\,\mathrm{d}x \\
&= (n_2-n_1)^\ell \cdot \frac{2}{3}((n-n_1)^{3/2} - (n-n_2)^{3/2}) \\
&= (n_2-n_1)^\ell \\
&\cdot \frac{2}{3} \frac{(n_2-n_1)\left[(n-n_1)^2 + (n-n_1)(n-n_2) + (n-n_2)^2\right]}{(n-n_1)^{3/2} + (n-n_2)^{3/2}} \\
&\leq 2(n_2-n_1)^{\ell+1}(n-n_1)^{1/2}.
\end{aligned}$$

Therefore the claim holds by choosing $c = 2^{-(d+2)}$. This entails that, for any $t > 0$,

$$\begin{aligned}&\mathbb{P}(Z \geq t)\\&\leq \mathbb{P}\bigg(\max_{1\leq n_1<n_2\leq n}(n_2 \wedge (n-n_1))^{-1/2}\bigg|\sum_{(n_1;n_2]}\varepsilon_i\bigg| \geq ct\bigg)\\&\leq \mathbb{P}\bigg(\max_{n_1<n_2}\frac{\big|\sum_{(n_1;n_2]}\varepsilon_i\big|}{(n-n_1)^{1/2}} \geq ct\bigg)+\\&\quad \mathbb{P}\bigg(\max_{n_1<n_2}\frac{\big|\sum_{(n_1;n_2]}\varepsilon_i\big|}{n_2^{1/2}} \geq ct\bigg)\\&\equiv (I)\ + (II).\end{aligned}$$

Due to symmetry, we only bound (I). By the triangle inequality,

$$\begin{aligned}(I) &\leq \mathbb{P}\bigg(\sup_{n_1<n_2}\frac{\big|\sum_{i=n_1+1}^{n}\varepsilon_i\big|}{(n-n_1)^{1/2}} > ct/2\bigg)\\&\quad+\mathbb{P}\bigg(\sup_{n_1<n_2}\frac{\big|\sum_{i=n_2+1}^{n}\varepsilon_i\big|}{(n-n_1)^{1/2}} > ct/2\bigg).\end{aligned}$$

By Lévy's maximal inequality (cf. Theorem 1.1.5 of [46]), the first probability is bounded by

$$\begin{aligned}&\sum_{r=1}^{\lceil\log_2 n\rceil}\mathbb{P}\bigg(\sup_{2^{r-1}\leq(n-n_1)<2^r}2^{-(r-1)/2}\bigg|\sum_{i=n_1+1}^{n}\varepsilon_i\bigg| \geq ct/2\bigg)\\&\quad\leq 9\lceil\log_2 n\rceil e^{-c't^2}.\end{aligned}$$

Similarly, the second inequality is bounded by

$$\begin{aligned}&\sum_{r=1}^{\lceil\log_2 n\rceil}\mathbb{P}\bigg(\sup_{\substack{2^{r-1}\leq(n-n_1)<2^r\\1\leq n_1<n_2\leq n}}(n-n_1)^{-1/2}\bigg|\sum_{i=n_2+1}^{n}\varepsilon_i\bigg| \geq ct/2\bigg)\\&\leq \sum_{r=1}^{\lceil\log_2 n\rceil}\mathbb{P}\bigg(\sup_{n-2^r<n_2\leq n}2^{-(r-1)/2}\bigg|\sum_{i=n_2+1}^{n}\varepsilon_i\bigg| \geq ct/2\bigg)\\&\leq 9\lceil\log_2 n\rceil e^{-c't^2}.\end{aligned}$$

Putting together the pieces, it holds that $\mathbb{P}(Z \geq t) \leq 18\lceil\log_2 n\rceil e^{-c''t^2}$, where we take $c'' < c_0$ without loss of generality. Now, if $\psi(\cdot)$ is bounded on $[0,\infty)$ by some C, then the result holds trivially. Otherwise, $\psi(x)\uparrow\infty$ as $x\to\infty$, and integration by parts yields that for any $x_0 \geq 0$,

$$\begin{aligned}\mathbb{E}\psi(Z) &= \int_0^\infty \mathbb{P}(\psi(Z)\geq t)\ \mathrm{d}t = \int_0^\infty \mathbb{P}(Z\geq\psi^{-1}(t))\ \mathrm{d}t\\&\leq \int_0^\infty\Big\{1\wedge\big[C\log(16n)\cdot e^{-c''(\psi^{-1}(t))^2}\big]\Big\}\ \mathrm{d}t\\&\leq x_0 + C\cdot\int_{x_0}^\infty \log(16n)\cdot e^{-c''(\psi^{-1}(t))^2}\ \mathrm{d}t.\end{aligned}$$

By monotonicity of ψ^{-1}, for any $t \geq x_0$, $\psi^{-1}(t) \geq \psi^{-1}(t)/2+\psi^{-1}(x_0)/2$, so the integral above can be further bounded by

$$\begin{aligned}&\int_{x_0}^\infty\big[\log(16n)\cdot e^{-(c''/4)(\psi^{-1}(x_0))^2}\big]e^{-(c''/4)(\psi^{-1}(t))^2}\ \mathrm{d}t\\&\quad\leq\int_1^\infty e^{-(c''/4)(\psi^{-1}(t))^2}\ \mathrm{d}t,\end{aligned}$$

provided that $x_0 \geq 1$ and $\log(16n)\cdot e^{-(c''/4)(\psi^{-1}(x_0))^2} \leq 1$, or equivalently, $x_0 \geq 1\vee\psi\big(\sqrt{(4/c'')\log\log(16n)}\big)$. The claim now follows from the condition (22). $\square$

Appendix C
Proofs for Technical Results in Section V

A. Proof of Proposition 16

Proof of Proposition 16: The basic inequality $\|Y-\widehat{\theta}\|^2 \leq \|Y-\theta_{\mathrm{oracle}}\|^2$ entails that

$$\|\widehat{\theta}-\theta_0\|^2 \leq \|\theta_{\mathrm{oracle}}-\theta_0\|^2 + 2\varepsilon\cdot\big(\widehat{\theta}-\theta_{\mathrm{oracle}}\big).$$

Then we have, for any $\eta > 0$,

$$\begin{aligned}&\varepsilon\cdot\big(\widehat{\theta}-\theta_{\mathrm{oracle}}\big)\\&=\sum_{j=0}^{k-1}(\varepsilon_{[j]}\cdot(\widehat{\theta}-\theta_{\mathrm{oracle}})_{[j]})\\&=\sum_{j=0}^{k-1}(\varepsilon_{[j]}\cdot v_j(\widehat{\theta}))\|(\widehat{\theta}-\theta_{\mathrm{oracle}})_{[j]}\|\\&\leq\eta^{-1}\cdot\sum_{j=0}^{k-1}\big(\varepsilon_{[j]}\cdot v_j(\widehat{\theta})\big)^2+\eta\cdot\sum_{j=0}^{k-1}\|(\widehat{\theta}-\theta_{\mathrm{oracle}})_{[j]}\|^2\\&=\eta^{-1}\cdot\sum_{j=0}^{k-1}\big(\varepsilon_{[j]}\cdot v_j(\widehat{\theta})\big)^2+\eta\cdot\|\widehat{\theta}-\theta_{\mathrm{oracle}}\|^2.\end{aligned}$$

Applying the inequality $\|\widehat{\theta}-\theta_{\mathrm{oracle}}\|^2 \leq 2\big(\|\widehat{\theta}-\theta_0\|^2 + \|\theta_{\mathrm{oracle}}-\theta_0\|^2\big)$ then yields that

$$\begin{aligned}\|\widehat{\theta}-\theta_0\|^2 &\leq \frac{1+2\eta}{1-2\eta}\|\theta_{\mathrm{oracle}}-\theta_0\|^2\\&\quad+\frac{1}{\eta(1-2\eta)}\sum_{j=0}^{k-1}(\varepsilon_{[j]}\cdot v_j(\widehat{\theta}))^2.\end{aligned}$$

For any given $\delta > 0$, choosing $\eta = \delta/(2\delta+4)$, upper bounding the right-hand side by the supremum over $\Theta(d,d_0,k)$, and then taking expectation on both sides yield the desired result. $\square$

B. Proof of Lemma 18

Proof of Lemma 18: On the pieces $(n_{i-1}/n, n_i/n]$ and $(n_i/n, n_{i+1}/n]$, the function f can be parametrized as

$$\begin{aligned}f_{i-1}(x) &\equiv \sum_{q=1}^{d+1}a_q^{i-1}\Big(x-\frac{n_{i-1}}{n}\Big)^{q-1},\\f_i(x) &\equiv \sum_{q=1}^{d+1}a_q^{i}\Big(x-\frac{n_i}{n}\Big)^{q-1}.\end{aligned}$$

By the fact that $0 \leq p-1 \leq d_0$ and thus the continuity of the $(p-1)$th derivative at knot n_i/n, it holds that $f_{i-1}^{(p-1)}(n_i/n) = f_i^{(p-1)}(n_i/n)$. But

$$\begin{aligned}f_{i-1}^{(p-1)}\Big(\frac{n_i}{n}\Big) &= \sum_{q=1}^{d+1}a_q^{i-1}\frac{\mathrm{d}^{p-1}}{\mathrm{d}x^{p-1}}\Big(x-\frac{n_{i-1}}{n}\Big)^{q-1}\Big|_{x=\frac{n_i}{n}}\\&=\sum_{q=p}^{d+1}a_q^{i-1}\underline{\odot}(q-1;p-1)n_{i;i-1}^{q-p},\end{aligned}$$

$$\begin{aligned} f_i^{(p-1)}\left(\frac{n_i}{n}\right) &= \sum_{q=1}^{d+1} a_q^i \frac{\mathrm{d}^{p-1}}{\mathrm{d}x^{p-1}}\left(x - \frac{n_i}{n}\right)^{q-1}\Bigg|_{x=\frac{n_i}{n}} \\ &= (p-1)!a_p^i. \end{aligned}$$

This entails that

$$\begin{aligned} (p-1)!a_p^i &= \sum_{q=p}^{d+1} \underline{\odot}(q-1;p-1)a_q^{i-1}n_{i;i-1}^{q-p} \\ &= \sum_{q=p}^{d+1} \frac{(q-1)!}{(q-p)!}a_q^{i-1}n_{i;i-1}^{q-p}. \end{aligned}$$

This implies that $\mathrm{Coef}[a_p^i; a_q^{i-1}] = (q-1)!/((q-p)!(p-1)!)n_{i;i-1}^{q-p} = \binom{q-1}{p-1}n_{i;i-1}^{q-p}$ if $q \geq p$; otherwise it is 0. $\square$

C. Proof of Lemma 19

Proof of Lemma 19: The baseline case $s = 0$ follows from the condition $\|\theta\| \leq 1$ and application of Lemma 27 to the piece $(n_{k_0-1}; n_{k_0}]$. The iteration from s to $s+1$ then follows from Lemma 31, which is to be stated and proved in Appendix E with its conditions satisfied since $n_{k_0;k_0-1} \geq \max\{n_{2;1}, n_{3;2}, \ldots, n_{k_0-1;k_0-2}\}$ by (33). $\square$

D. Proof of Lemma 21

Proof of Lemma 21: Fix $i \leq k_0 - 2$ as in the lemma statement. For simplicity, we again work under the condition $n_{k_0;k_0-1} = \max\{n_{2;1}, \ldots, n_{k_0;k_0-1}\}$. We will prove by induction: suppose the desired estimates hold for a_ℓ^i, $\ell \in [d_0+1; \ell_0]$ for some $\ell_0 \in [d_0+1; d]$ and we will prove that the estimate also holds for $a_{\ell_0+1}^i$. The condition of the lemma serves as the baseline $\ell_0 = d_0 + 1$. For the general induction from ℓ_0 to $\ell_0 + 1$, let $L \equiv 1 + (d - d_0)(k_0 - 1 - i)$. Then, Lemma 19 entails that

$$\begin{aligned} 1 \gtrsim \frac{(n - n_{k_0-1})^{2(\ell_0+1-L)+1}}{n^{2(\ell_0+1-L)}} \cdot \\ \left(\sum_{\ell=\ell_0+2-L}^{\ell_0+1} \overline{\beta}_{\ell_0+2-L,\ell-(\ell_0+2-L)}^{k_0-1-i} a_\ell^i\right)^2. \end{aligned}$$

On the other hand, we have

$$1 \gtrsim \sum_{\ell=\ell_0+2-L}^{\ell_0+1} n_{i+1;i}^{2(\ell-1)} M^2(n_{i+1}, n_i)(a_\ell^i)^2,$$

where the summands with $\ell \in [\ell_0+2-L; d_0+1]$ are from the condition of the lemma and those with $\ell \in [d_0+2; \ell_0+1]$ are from the induction assumption. Now, combining the above two estimates and applying Lemma 29 iteratively to cancel every a_ℓ^i, $\ell \in [\ell_0+2-L; \ell_0]$, we have

$$1 \gtrsim (a_{\ell_0+1}^i)^2((I) \wedge (II)),$$

where

$$\begin{aligned} (I) &\equiv \frac{(n - n_{k_0-1})^{2(\ell_0+1-L)+1}}{n^{2(\ell_0+1-L)}} (\overline{\beta}_{\ell_0+2-L,L-1}^{k_0-1-i})^2, \\ (II) &\equiv \bigwedge_{\ell=\ell_0+2-L}^{\ell_0} n_{i+1;i}^{2(\ell-1)} M(n_{i+1}, n_i) \frac{(\overline{\beta}_{\ell_0+2-L,L-1}^{k_0-1-i})^2}{(\overline{\beta}_{\ell_0+2-L,\ell-(\ell_0+2-L)}^{k_0-1-i})^2}. \end{aligned}$$

By Lemma 33 and the condition $n_{k_0;k_0-1} = \max\{n_{2;1}, \ldots, n_{k_0;k_0-1}\}$, we obtain that $(I) \gtrsim n_{i+1;i}^{2\ell_0} M(n_{i+1}, n_i)$. Similarly, by Lemma 33, as the factors $n_{\cdot;\cdot}$'s in the lower bound of $(\overline{\beta}_{\ell_0+2-L,L-1}^{k_0-1-i})^2 / (\overline{\beta}_{\ell_0+2-L,\ell-(\ell_0+2-L)}^{k_0-1-i})^2$ can all be further bounded below by $n_{i+1;i}$, we obtain by direct calculation that $(II) \gtrsim n_{i+1;i}^{2\ell_0} M(n_{i+1}, n_i)$. Putting together the lower bounds for (I), (II) completes the induction. $\square$

E. General Statement of Lemma 22

We restate here Lemma 22 for the case of general $d_0 \in [-1; d-1]$. Introduce the following notation:

$$\underline{\odot} n_{\cdot;j}(a,b,c) \equiv n_{a;j}^c \cdot n_{a-1;j}^c \ldots n_{a-\lfloor b/c \rfloor - 1;j}^{\mathrm{Mod}(b;c)}$$

for positive integers a, b, c. Fix $i \geq 2$. Recall the definition $M(a,b) = (a \wedge (n-b))^{1/2}$ for $a, b \in [1; n]$ and the condition (33).

Lemma 25: The following estimates hold for all locations $1 \leq j \leq i+1$:

$$\begin{aligned} 1 \gtrsim \max_{1 \leq \ell \leq d_0+1} \Bigg\{ & n_{i+2;j}^{2\ell \vee 2(d-d_0)(k_0-i-2)} \\ & \times \underline{\odot} n_{\cdot;j}^2 \Big(i+1, \{\ell - (d-d_0)(k_0-i-2) - 1\} \wedge \\ & \{(d-d_0)(i+1-j)\}, d-d_0 \Big) \\ & \times n_{j+1;j}^{2(\ell-1-(d-d_0)(k_0-j-1))_+} \Bigg\} \cdot M^2(n_{j+1}, n_j). \end{aligned}$$

In particular, for $j = i+1$:

$$1 \geq c \max_{1 \leq \ell \leq d_0+1} \left\{ (a_\ell^{i+1})^2 \cdot n_{i+2;i+1}^{2(\ell-1)} \cdot M^2(n_{i+2}, n_{i+1}) \right\}.$$

The proof for this general case is completely analogous to the one presented in Section V-E.

APPENDIX D
PROOFS FOR TECHNICAL RESULTS IN SECTION VI

A. Proof of Lemma 23

Proof of Lemma 23: For any $f \in \mathcal{F}_n^*(d,k)$, let $f_\circ \equiv f_\circ(f) \in \mathcal{F}_n^*(0,k)$ be such that $f = (I_{r_0,\ldots,r_{d-1};0}^d f_\circ)$ for some real sequence $\{r_\ell\}_{\ell=0}^{d-1}$, with corresponding knots $\{n_j\}_{j=1}^{k-1} = \{n_j(f_\circ)\}_{j=1}^{k-1}$ and magnitudes $\{\mu_j\}_{j=1}^k = \{\mu_j(f_\circ)\}_{j=1}^k$ between $(n_{j-1}/n, n_j/n]$, i.e., $f_\circ(x) = \sum_{j=1}^k \mu_j \mathbf{1}_{(n_{j-1}/n, n_j/n]}(x)$ for $x \in (0,1]$. Then $\mu_1 \leq \ldots \leq \mu_k$. Let

$$j^* \equiv j^*(f_\circ) \equiv \max\{1 \leq j \leq k : \mu_j \leq 0\}.$$

Define two sequences $\{\widetilde{a}_j\}_{j=1}^{j^*}$ and $\{\widetilde{b}_j\}_{j=j^*}^{k-1}$ as follows: $\widetilde{a}_{j^*} \equiv \mu_{j^*} \leq 0$ and $\widetilde{a}_j \equiv \mu_j - \mu_{j+1} \leq 0$ for $j \in [1; j^*-1]$, $\widetilde{b}_{j^*} \equiv \mu_{j^*+1} \geq 0$ and $\widetilde{b}_j \equiv \mu_{j+1} - \mu_j \geq 0$ for $j \in [j^*+1; k-1]$. Then, letting $\tau_j \equiv n_j/n$, $f_\circ$ can be re-parametrized as

$$f_\circ(x) = \sum_{j=1}^{j^*} \widetilde{a}_j \mathbf{1}_{(0,\tau_j]}(x) + \sum_{j=j^*}^{k-1} \widetilde{b}_j \mathbf{1}_{(\tau_j,1]}(x), \quad x \in (0,1].$$

Define the function $g_\ell^-(x;\tau) \equiv (\tau - x)_+^\ell$ with any parameter $\tau \in [0,1]$. Then, direct calculation shows that

$$\int_0^x g_\ell^-(u;\tau)\ \mathrm{d}u = \int_0^{x\wedge\tau} (\tau-u)^\ell\ \mathrm{d}u = \int_{\tau - x\wedge\tau}^{\tau} u^\ell\ \mathrm{d}u = \frac{\tau^{\ell+1}}{\ell+1} - \frac{(\tau-x)_+^{\ell+1}}{\ell+1} = \frac{\tau^{\ell+1}}{\ell+1} + \frac{(-1)}{\ell+1}\cdot g_{\ell+1}^-(x;\tau).$$

Similarly, with $g_\ell^+(x;\tau) \equiv (x-\tau)_+^\ell$, it holds that $\int_0^x g_\ell^+(u;\tau)\ \mathrm{d}u = \int_\tau^{x\vee\tau}(u-\tau)^\ell\ \mathrm{d}u = \int_0^{x\vee\tau-\tau} u^\ell\ \mathrm{d}u = g_{\ell+1}^+(x;\tau)/(\ell+1)$. This entails that

$$(I^d_{r_0,\ldots,r_{d-1};0} f_\circ)(x) = \sum_{j=1}^{j^*} (-1)^d \frac{\widetilde{a}_j}{d!}(\tau_j - x)_+^d + \sum_{j=j^*}^{k-1} \frac{\widetilde{b}_j}{d!}(x-\tau_j)_+^d + P_{d-1}(x),$$

where $P_{d-1}(x)$ is some polynomial of order $d-1$. The proof is then complete by noting that $\{(-1)^d\widetilde{a}_j/d!\}_{j=1}^{j^*}$ has sign $(-1)^{d+1}$ and $\{\widetilde{b}_j/d!\}_{j=j^*}^{k-1}$ is non-negative. $\square$

B. Proof of Lemma 24

We need the following simple fact that translates the ℓ_2 constraint on θ^* at the sequence level to an integral L_2 constraint on f^* at the underlying function level. Its proof can be found after the proof of Lemma 24.

Lemma 26: Let $f^* \in \mathcal{F}_n^*(d,k)$ and $(\theta^*)_i \equiv (f^*(i/n))_i$. Then, if $\|\theta^*\|^2 \le 1$, there exists some $c = c(d)$ such that $1 \ge c\cdot n\int_0^1 (f^*)^2(x)\ \mathrm{d}x$. Actually, this inequality holds for the larger unshaped spline space $\mathcal{F}_n(d,d_0,k)$.

Proof of Lemma 24: Fix any $\theta \in \Theta^*(d,k)$ and its generating spline $f \in \mathcal{F}_n^*(d,k)$. Then, under the condition $\|\theta\|^2 \le 1$, Lemma 26 entails that there exists some $K = K(d) > 0$ such that $\int_0^1 f^2(x)\ \mathrm{d}x \le K/n$. Due to scale invariance, it suffices to prove that $|c_\ell(f)| \le C$ for $\ell \in [0;d-1]$ for some $C = C(d)$ under the condition $\|f\|_2^2 = \int_0^1 f^2(x)\ \mathrm{d}x \le 1$.

For $f \in \mathcal{F}_n^*(d,k)$, let $\{n_j = n_j(f)\}_{j\in[0;k]}$ be its knots and $j^* = j^*(f)$ be as in its canonical form in Lemma 23. Let $\tau_j \equiv \tau_j(f) \equiv n_j(f)/n$ for $j \in [1;k]$ and $\tau^* \equiv \tau^*(f) \equiv \tau_{j^*(f)}(f)$. We will prove that for some $K = K(d) > 0$,

$$\int_0^1 f^2(x)\ \mathrm{d}x \ge K\cdot \max_{0\le\ell\le d-1} c_\ell^2(f), \quad \text{for any } f \in \mathcal{F}_n^*(d,k).$$

We focus on the case $\tau^*(f) \in [0,1/2]$ and prove that

$$\int_{\tau^*(f)}^1 f^2(x)\ \mathrm{d}x \ge K\cdot \max_{0\le\ell\le d-1} c_\ell^2(f), \quad \text{for any } f \in \mathcal{F}_n^*(d,k).$$

We present the proof for $c_{d-1}(f)$ whenever $c_{d-1}(f) \neq 0$; the bounds for $\{c_\ell(f)\}_{\ell\in[0;d-2]}$ follow from completely analogous arguments. Below we omit notational dependence on f if no confusion could arise. On $[\tau^*,1]$, f has the canonical form

$$f(x) = \sum_{j=j^*}^{k-1} b_j(x-\tau_j)_+^d + \sum_{\ell=0}^{d-1}\frac{c_\ell}{\ell!}x^\ell.$$

This can be alternatively parametrized as $f(x) = \sum_{\ell=0}^{d-1} c_\ell x^\ell/\ell! + (I^d_{0,\ldots,0;\tau^*(f)} f_\circ)(x)$, where $f_\circ(x) \equiv \sum_{j=j^*}^{k-1} (b_j\cdot d!)\mathbf{1}_{x>\tau_j} \in \mathcal{F}_n^*(0,k)$, and $\tau^*(f) = \tau^*(f_\circ)$. Therefore, we have

$$\begin{aligned}
1 &\ge \int_{\tau^*(f)}^1 \Big(\sum_{\ell=0}^{d-1} c_\ell x^\ell/\ell! + (I^d_{0,\ldots,0;\tau^*(f)} f_\circ)(x)\Big)^2\ \mathrm{d}x\\
&= c_{d-1}^2 \int_{\tau^*(f_\circ)}^1 \Big[\sum_{\ell=0}^{d-2}\frac{c_\ell}{|c_{d-1}|}\frac{x^\ell}{\ell!} + \mathrm{sgn}(c_{d-1})\frac{x^{d-1}}{(d-1)!} + \Big(I^d_{0,\ldots,0;\tau^*(f_\circ)}\frac{f_\circ}{|c_{d-1}|}\Big)(x)\Big]^2\ \mathrm{d}x\\
&\ge c_{d-1}^2 \inf_{\substack{c'_0,\ldots,c'_{d-2}\in\mathbb{R}, c'_{d-1}\in\{\pm1\}\\ \widetilde{f}_\circ\in\cup_n\mathcal{F}_n^*(0,k), \tau^*(\widetilde{f}_\circ)\le 1/2}} \int_{\tau^*(\widetilde{f}_\circ)}^1 \Big[\sum_{\ell=0}^{d-1}\frac{c'_\ell x^\ell}{\ell!} + (I^d_{0,\ldots,0;\tau^*(\widetilde{f}_\circ)}\widetilde{f}_\circ)(x)\Big]^2\ \mathrm{d}x\\
&= c_{d-1}^2 \inf_{\substack{c''_0,\ldots,c''_{d-2}\in\mathbb{R}, c''_{d-1}\in\{\pm1\}\\ \widetilde{f}_\circ\in\cup_n\mathcal{F}_n^*(0,k), \tau^*(\widetilde{f}_\circ)\le 1/2}} \int_{\tau^*(\widetilde{f}_\circ)}^1 \Big[\sum_{\ell=0}^{d-1}\frac{c''_\ell(x-\tau^*(\widetilde{f}_\circ))^\ell}{\ell!} + (I^d_{0,\ldots,0;\tau^*(\widetilde{f}_\circ)}\widetilde{f}_\circ)(x)\Big]^2\ \mathrm{d}x,
\end{aligned}$$

where in the third line we use the fact that $\widetilde{f}_\circ = f_\circ/|c_{d-1}| \in \mathcal{F}_n^*(0,k)$ and satisfies $\tau^*(\widetilde{f}_\circ) = \tau^*(f_\circ) \le 1/2$. Thus, to prove the desired result, it suffices to show that there exists some $K = K(d) > 0$ such that

$$\inf_{\substack{\widetilde{c}_0,\ldots,\widetilde{c}_{d-2}\in\mathbb{R}, \widetilde{c}_{d-1}\in\{\pm1\}\\ \widetilde{f}_\circ\in\cup_n\mathcal{F}_n^*(0,k), \tau^*(\widetilde{f}_\circ)\le 1/2, k\in\mathbb{Z}_+}} \int_{\tau^*(\widetilde{f}_\circ)}^1 \Big[\sum_{\ell=0}^{d-1}\frac{\widetilde{c}_\ell(x-\tau^*(\widetilde{f}_\circ))^\ell}{\ell!} + (I^d_{0,\ldots,0;\tau^*(\widetilde{f}_\circ)}\widetilde{f}_\circ)(x)\Big]^2\ \mathrm{d}x \ge K. \tag{53}$$

Suppose this is not true, then there exist a function sequence $\{\widetilde{f}_{n,\circ}\}_n \subset \cup_{n',k'}\mathcal{F}_{n'}^*(0,k')$ with $\tau_n^* \equiv \tau_n^*(\widetilde{f}_{n,\circ}) \subset [0,1/2]$ and real sequences $\{\widetilde{c}_{n,\ell}\}_{n,\ell}$ with $\widetilde{c}_{n,d-1} \in \{\pm1\}$, such that

$$\int_0^1 \mathbf{1}_{[\tau_n^*,1]}(x) \Big[\sum_{\ell=0}^{d-1}\frac{\widetilde{c}_{n,\ell}}{\ell!}(x-\tau_n^*)^\ell + (I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x)\Big]^2\ \mathrm{d}x \to 0.$$

Since L_2 convergence implies almost everywhere (a.e.) convergence, it follows that

$$\mathbf{1}_{[\tau_n^*,1]}(x)\cdot\Big[\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(x-\tau_n^*)^\ell/\ell! + (I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x)\Big] \to 0, \quad \text{a.e. on } [0,1].$$

Since the sequence $\{\tau_n^*\} \subset [0,1/2]$ is bounded, $\tau_n^* \to \tau^*$ along some subsequence for some $\tau^* \in [0,1/2]$, and we work with this subsequence below. As $\mathbf{1}_{[\tau_n^*,1]}(x) \to 1$ for any fixed $x \in (\tau^*,1]$, the sequence of functions in the brackets in the above display converges a.e. to 0 on $(\tau^*,1]$. In other words,

$$\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(x-\tau_n^*)^\ell/\ell! + (I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x) \to 0, \quad \text{a.e. on } (\tau^*,1]. \tag{54}$$

We first prove that under (54), $\{\widetilde{c}_{n,\ell}\}_n$ is necessarily bounded for each $\ell \in [0; d-1]$. Since $\{\widetilde{c}_{n,d-1}\}_n \subset \{-1,+1\}$ is already bounded, it suffices to prove the claim for $\ell \in [0; d-2]$. If this is not the case, then there exists some nonempty subset $\mathcal{L} \subset [0; d-2]$ such that for every $\ell \in \mathcal{L}$, $\{\widetilde{c}_{n,\ell}\}_n$ is divergent, i.e., $\limsup_n |\widetilde{c}_{n,\ell}| = +\infty$. As $\tau_n^* \to \tau^*$, we may find some slowly decaying $\varepsilon_n \downarrow 0$ such that (i) $\varepsilon_n > (\tau^* - \tau_n^*)_+$, (ii) $\{\widetilde{c}_{n,\ell}\varepsilon_n^\ell\}_n$ is still divergent for every $\ell \in \mathcal{L}$, and (54) holds with $x_n \equiv \tau_n^* + \varepsilon_n > \tau^*$. Now, by definition of $\widetilde{f}_{n,\circ}(\cdot)$, there exist some k_n, $j_n^* \in [1; k_n]$, $0 \equiv \tau_{n,0} \leq \ldots \leq \tau_{n,k_n} \equiv 1$, and non-negative sequence $\{\mu_{n,j}\}_{j=j_n^*}^{k_n-1}$ such that $\tau_n^* \equiv \tau_{n,j_n^*} \leq 1/2$ and for $x \in [\tau_n^*, 1]$, $\widetilde{f}_{n,\circ}(x) = \sum_{j=j_n^*}^{k_n-1} \mu_{n,j}\mathbf{1}_{x>\tau_{n,j}}$. Thus by a direct calculation, we have for $x \in [\tau_n^*, 1]$

$$(I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x) = \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!}\big(x-\tau_n^*\big)_+^d.$$

So by (54) and definition of $\{x_n\}$,

$$\sum_{\ell=0}^{d-1} \frac{\widetilde{c}_{n,\ell}}{\ell!}\varepsilon_n^\ell + \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!}\varepsilon_n^d \to 0.$$

Let $\ell_0 \in \mathcal{L}$ be the index such that $\{\widetilde{c}_{n,\ell}\varepsilon_n^\ell\}_n$ has the fastest divergence rate, i.e., $\limsup_n |\widetilde{c}_{n,\ell_0}|\varepsilon_n^{\ell_0}/(|\widetilde{c}_{n,\ell}|\varepsilon_n^\ell) \geq \alpha$ for some positive α and every $\ell \in \mathcal{L}$. Without loss of generality, we further choose $\{\varepsilon_n\}$ such that the maximal divergence rate and the index that achieves this rate are unique, i.e., ℓ_0 is unique and satisfies $\limsup_n |\widetilde{c}_{n,\ell_0}|\varepsilon_n^{\ell_0}/(|\widetilde{c}_{n,\ell}|\varepsilon_n^\ell) = \infty$ for every $\ell \in \mathcal{L}\setminus\{\ell_0\}$. This then entails that

$$B_n \equiv \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!}\varepsilon_n^d \gtrsim |\widetilde{c}_{n,\ell_0}|\varepsilon_n^{\ell_0} \tag{55}$$

and is positive and divergent. Next, for the chosen sequence $\{\varepsilon_n\}$, choose $\{\eta_n\} \subset [1,\infty)$ as some slowly growing sequence such that (54) holds with the sequence $x_n' \equiv \tau_n^* + \varepsilon_n\eta_n \geq \tau_n^* + \varepsilon_n > \tau^*$, i.e.,

$$\sum_{\ell=0}^{d-1} \frac{\widetilde{c}_{n,\ell}}{\ell!}(\eta_n\varepsilon_n)^\ell + \sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!}(\eta_n\varepsilon_n)^d \to 0, \tag{56}$$

and that $\{\varepsilon_n\eta_n\} \downarrow 0$ and $\{\widetilde{c}_{n,\ell_0}(\eta_n\varepsilon_n)^{\ell_0}\}$ remains to be the fastest divergent sequence among $\mathcal{L}$, i.e., $\limsup_n |\widetilde{c}_{n,\ell_0}| |(\varepsilon_n\eta_n)^{\ell_0}/(|\widetilde{c}_{n,\ell}|(\varepsilon_n\eta_n)^\ell) = \infty$ for every $\ell \in \mathcal{L}\setminus\{\ell_0\}$. Similar to (55), we have $\sum_{j=j_n^*}^{k_n-1} \mu_{n,j}(\eta_n\varepsilon_n)^d/d! \gtrsim |\widetilde{c}_{n,\ell_0}|(\eta_n\varepsilon_n)^{\ell_0}$ and is positive and divergent. But this is impossible since

$$\begin{aligned}\sum_{j=j_n^*}^{k_n-1} \frac{\mu_{n,j}}{d!}(\eta_n\varepsilon_n)^d &= (\eta_n)^d B_n \gtrsim (\eta_n)^{d-\ell_0}(\widetilde{c}_{n,\ell_0}(\varepsilon_n\eta_n)^{\ell_0}) \\ &\asymp (\eta_n)^{d-\ell_0}\bigg|\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(\eta_n\varepsilon_n)^\ell/\ell!\bigg|,\end{aligned}$$

where the first inequality is by (55) and the last relation is by the maximal divergence rate of $\{\widetilde{c}_{n,\ell_0}(\eta_n\varepsilon_n)^{\ell_0}\}$, and thus

$$\begin{aligned}&\sum_{\ell=0}^{d-1} \frac{\widetilde{c}_{n,\ell}}{\ell!}(\eta_n\varepsilon_n)^\ell + \sum_{j=j^*}^{k_n-1} \frac{\mu_{n,j}}{d!}(\eta_n\varepsilon_n)^d \\ &\gtrsim \big[(\eta_n)^{d-\ell_0}-1\big]\bigg|\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(\eta_n\varepsilon_n)^\ell/\ell!\bigg| \\ &\geq \big[\eta_n-1\big]\bigg|\sum_{\ell=0}^{d-1}\widetilde{c}_{n,\ell}(\eta_n\varepsilon_n)^\ell/\ell!\bigg| \to \infty,\end{aligned}$$

a contradiction to (56). This concludes that $\{\widetilde{c}_{n,\ell}\}_n$ are necessarily bounded for every $\ell \in [0; d-1]$. Thus there exists a real sequence $\{c_\ell^*\}_{\ell=0}^{d-1}$ with $c_{d-1}^* \in \{\pm 1\}$ such that $\widetilde{c}_{n,\ell} \to c_\ell^*$ along some subsequence for each $\ell \in [0; d-1]$. Coming back to (54) and noting that $\tau_n^* \to \tau^*$ along some subsequence, we then conclude that

$$h_n(x) \equiv (I^d_{0,\ldots,0;\tau_n^*}\widetilde{f}_{n,\circ})(x) \to \sum_{\ell=0}^{d-1}\frac{-c_\ell^*}{\ell!}(x-\tau^*)^\ell \equiv h^*(x) \tag{57}$$

a.e. on $(\tau^*, 1]$ as $n \to \infty$. We will now prove that $\{c_\ell^*\}_{\ell=0}^{d-1}$ are necessarily non-positive. Fix some positive integer $m > d$ and define a regular grid on $(\tau^*, 1]$: $t_i \equiv \tau^* + i(1-\tau^*)/m$ for $i \in [0; m]$. Without loss of generality, assume that $\{t_i\}_{i=1}^m$ belongs to the set with full Lebesgue measure such that (57) holds. Define $(\xi_{n,i})_{i=1}^m \equiv (h_n(t_i))_{i=1}^m$ (resp. $(\xi_i^*)_{i=1}^m \equiv (h^*(t_i))_{i=1}^m$) to be the realization of $h_n(\cdot)$ (resp. $h^*(\cdot)$) on this grid. Define ∇ to be the finite difference operator that maps $(y_1,\ldots,y_m)^\top \in \mathbb{R}^m$ to $(y_2-y_1,\ldots,y_m-y_{m-1})^\top \in \mathbb{R}^{m-1}$. Then, since $\lim_n \min_{\ell\in[0;d]} h_n^{(\ell)}(x) \geq 0$ for $x \in (\tau^*, 1]$, it holds that for each fixed $m \geq d+1$, $\nabla^\ell\xi_n \in \mathbb{R}^{m-\ell}_{\geq 0}$ holds for all $\ell \in [0; d]$ for n large enough. On the other hand, for each $\ell \in [0; d-1]$ and $p \in [\ell; d-1]$, there exists some positive constant $L_{p,\ell} > 0$ for such that

$$\begin{aligned}\big(\nabla^\ell\xi^*\big)_1 &= \bigg(\nabla^\ell\bigg(\sum_{p=0}^{d-1}\frac{-c_p^*}{p!}(t_j-\tau^*)^p\bigg)_{j=1}^m\bigg)_1 \\ &= \sum_{p=\ell}^{d-1} -c_p^* L_{p,\ell}((1-\tau^*)/m)^p.\end{aligned}$$

Since for each fixed $m \geq d+1$, $\nabla^\ell\xi_n \to \nabla^\ell\xi^*$ as $n \to \infty$ by (57) and $\nabla^\ell\xi_n \in \mathbb{R}^{m-\ell}_{\geq 0}$ for n large enough, it holds that $\big(\nabla^\ell\xi^*\big)_1 \geq 0$ for each fixed $m \geq d+1$. Multiplying by m^ℓ on both sides of the above equation and letting $m \to \infty$ we conclude that $c_\ell^* \leq 0$ for $\ell \in [0; d-2]$ and $c_{d-1}^* = -1$.

With $\{c_\ell^*\}_{\ell=0}^{d-1} \in \mathbb{R}^d_{\leq 0}$, h_n, h^* have the property that their derivatives up to order $d-1$ are all convex functions, so on arbitrary compact interval contained in $(\tau^*, 1)$, $D^{(d-1)}h_n$ converges uniformly to $D^{(d-1)}h^* \equiv 1$ (cf. Theorem 25.7 of [47] and the remark after its proof). This cannot happen as $D^{(d-1)}h_n(\tau_n^*) = 0$, $\tau_n^* \to \tau^*$ and $D^{(d-1)}h_n$ is convex. We have therefore established the contradiction and proved (53). □

Proof of Lemma 26: By Lemma 23, any $f \in \mathcal{F}_n^*(d,k)$ has the canonical parametrization

$$f(x) = \sum_{j=1}^{j^*} a_j(\tau_j - x)_+^d + \sum_{j=j^*}^{k-1} b_j(x-\tau_j)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell,$$

where $\{\tau_j\}_{j=1}^{k-1} \equiv \{n_j/n\}_{j=1}^{k-1} \subset [0,1]$. Let $\tau^* \equiv \tau_{j^*}$. Then, it holds that $\int_0^1 f^2(x)\ \mathrm{d}x = (I)\ + (II)$, where

$$(I) \equiv \int_0^{\tau^*} \Big(\sum_{j=1}^{j^*} a_j(\tau_j - x)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x,$$

$$(II) \equiv \int_{\tau^*}^1 \Big(\sum_{j=j^*}^{k-1} b_j(x-\tau_j)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x.$$

We now upper bound (II) by its sequence counterpart; the bound for (I) is similar. Since

$$\begin{aligned}(II) &= \sum_{m=j^*}^{k-1} \int_{\tau_m}^{\tau_{m+1}} \Big(\sum_{j=j^*}^{k-1} b_j(x-\tau_j)_+^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x \\ &= \sum_{m=j^*}^{k-1} \int_{\tau_m}^{\tau_{m+1}} \Big(\sum_{j=j^*}^{m} b_j(x-\tau_j)^d + \sum_{\ell=0}^{d-1} c_\ell x^\ell\Big)^2\ \mathrm{d}x,\end{aligned}$$

we may bound the integral piece by piece. More generally, we show that there exists some $K = K(d) > 0$ such that for any $a, b \in [0; n]$ with $b - a \geq d+1$ and d-degree polynomial $P(x) \equiv \sum_{\ell=0}^d c_\ell x^\ell$,

$$\int_{a/n}^{b/n} P^2(x)\ \mathrm{d}x \leq K \cdot n^{-1} \sum_{i \in (a;b]} P^2(i/n). \tag{58}$$

The above display holds because

$$\begin{aligned}&\int_{a/n}^{b/n} P^2(x)\ \mathrm{d}x = \int_{a/n}^{b/n} \Big(\sum_{\ell=0}^{d} c_\ell x^\ell\Big)^2\ \mathrm{d}x \\ &\lesssim_d \sum_{\ell=0}^{d} c_\ell^2 \cdot \int_{a/n}^{b/n} x^{2\ell}\ \mathrm{d}x \leq \sum_{\ell=0}^{d} \frac{c_\ell^2}{n} \sum_{i\in(a;b]} \Big(\frac{i}{n}\Big)^{2\ell} \\ &= \frac{1}{n} \sum_{i\in(a;b]} \sum_{\ell=0}^{d} \bigg(c_\ell\Big(\frac{i}{n}\Big)^\ell\bigg)^2 \lesssim_d \frac{1}{n}\sum_{i\in(a;b]} \bigg(\sum_{\ell=0}^{d} c_\ell\Big(\frac{i}{n}\Big)^\ell\bigg)^2,\end{aligned}$$

where the last inequality is due to Lemma 27 and the condition $b-a \geq (d+1)$. Then for every $\theta \in \Theta(d, d_0, k)$ with unit norm constraint and the corresponding $f \in \mathcal{F}_n(d, d_0, k)$, by (58) we have

$$\begin{aligned}1 &\geq \|\theta\|^2 \geq \|\theta\|^2_{(n_{j^*};n]} \\ &= \sum_{m=j^*}^{k-1} \|\theta\|^2_{(n_m;n_{m+1}]} = \sum_{m=j^*}^{k-1} \sum_{i\in(n_m;n_{m+1}]} f^2(i/n) \\ &\gtrsim n \sum_{m=j^*}^{k-1} \int_{\tau_m}^{\tau_{m+1}} f^2(x)\ \mathrm{d}x = n\int_{\tau^*}^1 f^2(x)\ \mathrm{d}x.\end{aligned}$$

The bound for (II) is thus complete. □

APPENDIX E
AUXILIARY LEMMAS

Lemma 27: Fix any positive integer d. There exists some $c = c(d)$ such that for any integers $n \geq 0$, $m \geq d+1$, and real sequence $\{a_\ell\}_{\ell=1}^{d+1}$,

$$\sum_{i=1}^m \bigg[a_1 + a_2\Big(\frac{i}{n}\Big) + \ldots + a_{d+1}\Big(\frac{i}{n}\Big)^d\bigg]^2 \geq c\sum_{\ell=1}^{d+1} a_\ell^2 \frac{m^{2\ell-1}}{n^{2(\ell-1)}}.$$

Proof of Lemma 27: As the left hand side of the above inequality equals

$$\begin{aligned}&\sum_{i=1}^n \Big(\sum_{\ell=1}^{d+1} a_\ell (i/n)^{\ell-1}\Big)^2 = \sum_{1\leq \ell,\ell'\leq d+1} a_\ell a_{\ell'} \sum_{i=1}^m (i/n)^{\ell+\ell'-2} \\ &= \sum_{1\leq \ell,\ell'\leq d+1} a_\ell (m/n)^{\ell-1} m^{1/2} \cdot a_{\ell'}(m/n)^{\ell'-1} m^{1/2} \\ &\quad \cdot \bigg[m^{-(\ell+\ell'-1)} \sum_{i=1}^m i^{\ell+\ell'-2}\bigg],\end{aligned}$$

using matrix notation, it can be written as $x^\top Ax$, where $x \equiv (a_\ell(m/n)^{\ell-1}m^{1/2})_{\ell=1}^{d+1} \in \mathbb{R}^{d+1}$, and the matrix $(A)_{ij} \equiv (A(m,d))_{ij} \equiv (m^{-(i+j-1)}\sum_{k=1}^m k^{i+j-2})_{ij} \in \mathbb{R}^{(d+1)\times(d+1)}$.

We first show that A is strictly positive-definite for the fixed d and any $m \geq d+1$. Note that A is actually a moment matrix and can be written as $A_{ij} = \mathbb{E}(X^{i-1}\cdot X^{j-1})$, where X is uniformly distributed on the set $\{1/m, \ldots, m/m\}$. Therefore, for any $c \in \mathbb{S}^d$, writing, with a slight abuse of notation, $Z \equiv \sum_{i=1}^{d+1} c_i X^{i-1}$, it holds that

$$\begin{aligned}c^\top Ac &= \sum_{1\leq i,j\leq d+1} c_i c_j A_{ij} = \sum_{1\leq i,j\leq d+1} c_i c_j \mathbb{E}(X^{i-1}\cdot X^{j-1}) \\ &= \mathbb{E}\bigg(\sum_{i=1}^{d+1} c_i X^{i-1}\bigg)^2 = \mathbb{E}Z^2 = (\mathbb{E}Z)^2 + \mathrm{Var}(Z).\end{aligned}$$

If $\mathrm{Var}(Z) = 0$, then $Z \equiv \alpha$ almost surely for some constant α, which is equivalent to that the polynomial

$$T(x) \equiv (c_0 - \alpha) + c_1 x + \ldots + c_{d+1}x^d$$

having distinct roots $\{1/m, \ldots, m/m\}$. If $c_1 = \ldots = c_{d+1} = 0$, then $c_0 = \pm 1$ since $\|c\| = 1$, which implies that $Z = \pm 1$, and thus $c^\top Ac \geq (\mathbb{E}Z)^2 = 1$. Otherwise, we have $c_i \neq 0$ for some $i \in [1; d]$, and hence $T(x)$ is not a constant and thus has at most d roots, which contradicts the condition that $m \geq d+1$. So we conclude that $c^\top Ac > 0$ for any $c \in \mathbb{S}^d$ and thus A is strictly positive-definite.

Next, we show that for any $i \in [1; d+1]$, the $(-i,-i)$-minor of A (i.e. A minus the ith row and column) is also strictly positive-definite. For this, define Q_i as the permutation matrix that switches row i with row $i+1$, and define $P_i \equiv Q_iQ_{i+1}\ldots Q_d$ for $i \leq d$ and $P_{d+1} \equiv I_{d+1}$, the $(d+1)$-dimensional identity matrix. Further define $B \equiv P_i^\top AP_i$. Then, the $(-i,-i)$-minor of A is the $(-(d+1),-(d+1))$-minor of B. By Sylvester's criterion, it suffices to show that B is strictly positive-definite, but for any $c \in \mathbb{S}^d$, it holds that

$$c^\top Bc = c^\top P_i^\top AP_i c \equiv \widetilde{c}^\top A\widetilde{c} > 0,$$

where in the last inequality we have used the fact that

$$\widetilde{c}^\top\widetilde{c} = c^\top P_i^\top P_i c = c^\top Q_d \ldots Q_i Q_i \ldots Q_d c = c^\top c = 1.$$

Next, we show that $x^\top Ax \geq ca_d^2 m^{2d+1}/n^{2d}$ for some $c = c(d)$; bounds involving $a_0, \ldots, a_{d-1}$ can be similarly obtained. For this, write A in the block form $[A_{11}, A_{12}; A_{21}, A_{22}]$, where $A_{12} \in \mathbb{R}^{d\times 1}$. Writing y as the first d components of x, i.e. $y \equiv (a_0 m^{1/2}, a_1 m^{3/2}/n, \ldots, a_{d-1} m^{(2d-1)/2}/n^{d-1})^\top$, we have

$$\begin{aligned} &x^\top A x \\ &= (y, a_d m^{(2d+1)/2}/n^d)^\top \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} (y, a_d m^{(2d+1)/2}/n^d) \\ &= y^\top A_{11} y + 2\frac{y^\top A_{21} a_d m^{(2d+1)/2}}{n^d} + A_{22}\Big(\frac{a_d m^{(2d+1)/2}}{n^d}\Big)^2. \end{aligned}$$

This is a quadratic form in y and achieves its minimum at $y^* = -A_{11}^{-1} A_{12} a_d m^{(2d+1)/2}/n^d$ (note that A_{11}, the $(-(d+1), -(d+1))$-minor of A, is indeed invertible as proved before), which implies that

$$x^\top A x \geq a_d^2 \frac{m^{2d+1}}{n^{2d}} (A_{22} - A_{21} A_{11}^{-1} A_{12}).$$

Therefore if we can show that $A_{22} \geq (1+\varepsilon) A_{21} A_{11}^{-1} A_{12}$ for some positive $\varepsilon = \varepsilon(d)$, then we have

$$\begin{aligned} x^\top A x &\geq \frac{\varepsilon}{1+\varepsilon} a_d^2 \frac{m^{2d+1}}{n^{2d}} A_{22} \\ &= \frac{\varepsilon}{1+\varepsilon} a_d^2 \frac{m^{2d+1}}{n^{2d}} m^{-(2d+1)} \sum_{k=1}^m k^{2d} \\ &\geq \frac{\varepsilon}{1+\varepsilon} a_d^2 \frac{m^{2d+1}}{n^{2d}} m^{-(2d+1)} \int_0^m x^{2d}\, \mathrm{d}x \\ &= \frac{\varepsilon}{(2d+1)(1+\varepsilon)} a_d^2 \frac{m^{2d+1}}{n^{2d}}. \end{aligned}$$

Using the block matrix inverse formula $(A^{-1})_{d+1,d+1} = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}$ and the fact that $(A^{-1})_{d+1,d+1} \leq \|A^{-1}\|_2 = \lambda_{\min}^{-1}(A)$ ($\lambda_{\min}$ takes the smallest eigenvalue), we have

$$\begin{aligned} & A_{22} \geq (1+\varepsilon) A_{21} A_{11}^{-1} A_{12} \\ \iff & (1+\varepsilon)(A_{22} - A_{21} A_{11}^{-1} A_{12}) \geq \varepsilon A_{22} \\ \iff & (A^{-1})_{d+1,d+1} \leq \frac{1+\varepsilon}{\varepsilon} A_{22}^{-1} \\ \Leftarrow\ & \lambda_{\min}^{-1}(A) \leq \frac{1+\varepsilon}{\varepsilon} \min_{1\leq j\leq d+1} A_{jj}^{-1}, \end{aligned}$$

which is further implied by

$$\lambda_{\min}(A) \geq \frac{\varepsilon}{1+\varepsilon} \max_{1\leq j\leq d+1} A_{jj}. \tag{59}$$

For this, we have, for every $j \in [1; d+1]$,

$$\begin{aligned} A_{jj} &= m^{-(2j-1)} \sum_{k=1}^m k^{2j-2} \leq m^{-(2j-1)} \int_1^{m+1} x^{2j-2}\, \mathrm{d}x \\ &\leq \frac{1}{2j-1}\Big(1 + \frac{1}{m}\Big)^{2j-1} \leq 2^{2d+1}. \end{aligned}$$

It remains to show that there exists some sufficiently small $c^* = c^*(d)$ such that $\lambda_{\min}(A) \geq c^* > 0$, then we can take $\varepsilon = c^*/(2^{2d+1} - c^*)$ in (59). For this, let U be a random variable uniformly distributed on $[0,1]$ and define matrix $\bar{A}$ as $\bar{A}_{i,j} \equiv \mathbb{E}(U^{i-1} \cdot U^{j-1})$. Then, since d is fixed, it holds by the definition of $A, \bar{A}$, and the Portmanteau theorem that $A \to \bar{A}$ in the matrix spectral norm as $m \to \infty$. By Weyl's inequality, there exists some positive integer $N = N(d)$ such that for $m \geq N$, $\lambda_{\min}(A) \geq \lambda_{\min}(\bar{A})/2$. On the other hand, a similar argument that establishes the positive definiteness of A yields that $\lambda_{\min}(\bar{A}) \geq c > 0$ for some $c = c(d)$. Therefore we can take $c^* = c^*(d) = \min_{d+1\leq m\leq N} \lambda_{\min}(A(m,d)) \wedge (c/2)$. This completes the proof. □

Lemma 28: Let $\{a_i\}_{i=1}^m, \{b_i\}_{i=1}^m$ be two non-negative sequences. Then, it holds that $(\bigwedge_{i=1}^m a_i) \cdot (\bigvee_{i=1}^m b_i) \geq \bigwedge_{i=1}^m a_i b_i$.

Proof of Lemma 28: Without loss of generality, let a_1 be the smallest value among $\{a_i\}_{i=1}^m$. Then, it holds that $\big(\bigwedge_{i=1}^m a_i\big) \cdot \big(\bigvee_{i=1}^m b_i\big) = a_1 \cdot \big(\bigvee_{i=1}^m b_i\big) \geq a_1 b_1 \geq \big(\bigwedge_{i=1}^m a_i b_i\big)$. □

Lemma 29: Let $\alpha_1, \alpha_2 > 0$ and β_1, β_2 be real numbers. Then, for any $x \in \mathbb{R}$, it holds that

$$\alpha_1(x+\beta_1)^2 + \alpha_2(x+\beta_2)^2 \geq (\alpha_1 \wedge \alpha_2)(\beta_1 - \beta_2)^2/2.$$

Proof of Lemma 29: At $x^* \equiv -(\alpha_1/(\alpha_1+\alpha_2) \cdot \beta_1 + \alpha_2/(\alpha_1+\alpha_2) \cdot \beta_2)$, the quadratic form achieves it minimum value $\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}(\beta_1 - \beta_1)^2$, which is further lower bounded by $(\alpha_1 \wedge \alpha_2)(\beta_1 - \beta_2)^2/2$. □

Lemma 30: Let n be any positive integer. Then, for any polynomial $P(\cdot)$ of degree strictly smaller than n, it holds that

$$\sum_{j=0}^n \binom{n}{j} P(j)(-1)^j = 0.$$

Proof of Lemma 30: We prove by induction. The claim clearly holds for $n = 1$. Suppose the claim holds for some n, we will prove that it also holds for $n+1$. Let d be the degree of $P(\cdot)$. We will prove that the claim holds for all monomials $P(x) \equiv x^d$ where $0 \leq d \leq n = (n+1) - 1$. The case $d = 0$ follows from the binomial identity:

$$\sum_{j=0}^{n+1} \binom{n+1}{j} (-1)^j = (1 + (-1))^{n+1} = 0.$$

Next, for any $1 \leq d \leq n$, it holds that

$$\begin{aligned} \sum_{j=0}^{n+1} \binom{n+1}{j} j^d (-1)^j &= \sum_{j=1}^{n+1} \binom{n+1}{j} j^d (-1)^j \\ &= (n+1) \sum_{j=1}^{n+1} \binom{n}{j-1} j^{d-1} (-1)^j \\ &= (n+1) \sum_{j=0}^{n} \binom{n}{j} (j+1)^{d-1} (-1)^j = 0, \end{aligned}$$

where the last identity follows from the claim for n and the fact that $0 \leq d-1 \leq n-1 < n$. □

For the following lemma, recall the definition of the sequence $\{\overline{\beta}_{\cdot,\cdot}\}$ defined before Lemma 19.

Lemma 31: Fix d, d_0, k_0 as defined in (11), and any $s \in [0; \lfloor (d_0+1)/(d-d_0) \rfloor - 1]$. Suppose there exists some $c_1 = c_1(d)$ such that

$$1 \geq c_1 \cdot \sum_{k=1}^{(s+1)d_0 - sd+1} \frac{(n - n_{k_0-1})^{2k-1}}{n^{2(k-1)}} \Big(\sum_{\ell=0}^{s(d-d_0)} \overline{\beta}_{k,\ell}^s a_{k+\ell}^{k_0-1-s} \Big)^2. \tag{60}$$

Furthermore, assume that $n_{k_0;k_0-1} \ge n_{k_0-1-s;k_0-2-s}$. Then, there exists some positive constant $c_2 = c_2(d)$ such that

$$1 \ge c_2 \cdot \sum_{k=1}^{(s+1)d_0-sd+1} \frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}} \left(\sum_{\ell=0}^{(s+1)(d-d_0)} \overline{\beta}^{s+1}_{k,\ell} a^{k_0-2-s}_{k+\ell} \right)^2.$$

Note that in the above lemma the hypothesis involves only quadratic forms with 'shared coefficients' $\{a^{\cdot}_{\ell}\}_{\ell\in[1;d_0+1]}$, while the conclusion involves the ones with both 'shared coefficients' $\{a^{\cdot}_{\ell}\}_{\ell\in[1;d_0+1]}$ and 'nuisance coefficients' $\{a^{\cdot}_{\ell}\}_{\ell\in[d_0+2;d+1]}$.

Before the proof of Lemma 31, we need one further result. For this, some extra notation is needed:

$$\overline{v}^s_{i,j} \equiv \frac{\overline{\odot}(i+d-d_0-j); s(d-d_0))}{j!\,\overline{\odot}(i+d-d_0; s(d-d_0))}(-1)^j n^j_{k_0-s;k_0-1-s} \times \prod_{m=1}^{j} \Big(d-\big(i+(d-1-d_0)\big)-s(d-d_0)+m\Big),$$

$$T_k \equiv \sum_{\ell=0}^{s(d-d_0)} \overline{\beta}^s_{k,\ell} a^{k_0-1-s}_{k+\ell}.$$

Lemma 32: Fix d, d_0, and s. It holds for $i \in [1;(s+1)d_0-sd+1]$ that

$$M \equiv \sum_{k=i}^{(s+1)d_0-sd+1} \overline{v}^{s+1}_{\overline{i},k-i} \cdot T_k = \sum_{k=0}^{(d-d_0)(s+1)} \overline{\beta}^{s+1}_{i,k} a^{k_0-2-s}_{i+k}.$$

Proof: In order to prove the desired result, we need to show the following two claims:

- The coefficient of $a^{k_0-2-s}_{i+j}$ in M equals 0 for $(s+1)(d-d_0)+1 \le j \le d-i+1$;
- The coefficient of $a^{k_0-2-s}_{i+j}$ in M equals $\overline{\beta}^{s+1}_{i,j}$ for $0 \le j \le (s+1)(d-d_0)$.

Let

$$\begin{aligned} i_0 &\equiv i_0(d,d_0,s,i) \equiv (s+1)d_0-sd+1-i,\\ \overline{i} &\equiv \overline{i}(d,d_0,s,i) \equiv (s+2)d_0-(s+1)d+2-i\\ &= i_0-(d-1-d_0),\\ \Delta n &\equiv n_{k_0-1-s;k_0-2-s}. \end{aligned}$$

By definition of M and Lemma 18, we have

$$\begin{aligned} &\mathrm{Coef}[M; a^{k_0-2-s}_{i+j}] = \sum_{k=i}^{(s+1)d_0-sd+1} \overline{v}^{s+1}_{\overline{i},k-i}\mathrm{Coef}[T_k; a^{k_0-2-s}_{i+j}]\\ &= \sum_{k=i}^{(s+1)d_0-sd+1} \frac{\overline{\odot}(\overline{i}+d-d_0-(k-i)); (s+1)(d-d_0))(-1)^{k-i}(\Delta n)^{k-i}}{(k-i)!\,\overline{\odot}(\overline{i}+d-d_0; (s+1)(d-d_0))}\\ &\quad\times \prod_{m=1}^{k-i}(d-i_0-(s+1)(d-d_0)+m)\\ &\quad\cdot \mathrm{Coef}\left[\sum_{\ell=0}^{s(d-d_0)} \overline{\beta}^s_{k,\ell} a^{k_0-1-s}_{k+\ell}; a^{k_0-2-s}_{i+j}\right]\\ &= \sum_{k=0}^{i_0} \frac{\overline{\odot}(i_0+1-k;(s+1)(d-d_0))(-1)^k(\Delta n)^k}{k!\,\overline{\odot}(i_0+1;(s+1)(d-d_0))}\overline{\odot}(i;k)\\ &\quad\times\left(\sum_{\ell=0}^{s(d-d_0)} \overline{\beta}^s_{i+k,\ell}\binom{i+j-1}{i+k+\ell-1}(\Delta n)^{j-k-\ell}\right)\\ &\equiv \sum_{\ell=0}^{s(d-d_0)} (\Delta n)^{j-\ell}\overline{\beta}^s_{\ell}\cdot A_\ell, \end{aligned}$$

where

$$A_\ell \equiv \sum_{k=0}^{i_0}(-1)^k \frac{\overline{\odot}(i_0+1-k;(s+1)(d-d_0))}{k!\,\overline{\odot}(i_0+1;(s+1)(d-d_0))}\overline{\odot}(i;k) \times \binom{i+j-1}{i+k+\ell-1}\frac{\overline{\odot}(i+k;\ell)}{\underline{\odot}(d+1-i-k;\ell)},$$

and we used $\overline{\beta}^s_{i+k,\ell} = D(i+k,\ell)\overline{\beta}^s_{\ell} = \frac{\overline{\odot}(i+k;\ell)}{\underline{\odot}(d+1-i-k;\ell)}\overline{\beta}^s_{\ell}$, with $\overline{\beta}^s_{\ell}$ defined in (32). Let $C(i,j,\ell) \equiv \binom{i+j-1}{i+\ell-1}\cdot\overline{\odot}(i;\ell)$. Then $C(i,j,\ell)\binom{j-\ell}{k} = \overline{\odot}(i;k)\binom{i+j-1}{i+k+\ell-1}\overline{\odot}(i+k;\ell)/k!$. So A_ℓ equals

$$\begin{aligned} &C(i,j,\ell)\sum_{k=0}^{i_0}\binom{j-\ell}{k}\\ &\frac{\overline{\odot}(i_0+1-k;(s+1)(d-d_0))}{\overline{\odot}(i_0+1;(s+1)(d-d_0))}(-1)^k\frac{1}{\underline{\odot}(d-i-k+1;\ell)}\\ &= C(i,j,\ell)\sum_{k=0}^{i_0+(s+1)(d-d_0)}\binom{j-\ell}{k}\\ &\frac{\overline{\odot}(i_0+1-k;(s+1)(d-d_0))}{\overline{\odot}(i_0+1;(s+1)(d-d_0))}(-1)^k\frac{1}{\underline{\odot}(d-i-k+1;\ell)}\\ &= C(i,j,\ell)\sum_{k=0}^{j-\ell}\binom{j-\ell}{k}\\ &\frac{\overline{\odot}(i_0+1-k;(s+1)(d-d_0))}{\overline{\odot}(i_0+1;(s+1)(d-d_0))}(-1)^k\frac{1}{\underline{\odot}(d-i-k+1;\ell)}\\ &= \frac{C(i,j,\ell)}{\overline{\odot}(i_0+1;(s+1)(d-d_0))}\\ &\cdot\sum_{k=0}^{j-\ell}\binom{j-\ell}{k}(-1)^k\underline{\odot}(d-i-k+1-\ell;(s+1)(d-d_0)-\ell), \end{aligned}$$

where the first identity follows from the fact that $\overline{\odot}(i_0+1-k;(s+1)(d-d_0)) = 0$ for any $i_0+1 \le k \le i_0+(s+1)(d-d_0)$, the second identity follows from the fact that $i_0+(s+1)(d-d_0) = d-i+1 \ge j \ge j-\ell$, the third identity follows from the fact that $\ell \le s(d-d_0) < (s+1)(d-d_0)$.

For the first claim, as $\underline{\odot}(d-i-k+1-\ell;(s+1)(d-d_0)-\ell)$ is a polynomial of degree at most $(s+1)(d-d_0)-\ell < j-\ell$, Lemma 30 entails that $A_\ell = 0$ for all $0 \le \ell \le s(d-d_0)$, thus proving the first claim. We now prove the second claim under the condition $j \le (s+1)(d-d_0)$. By definition of the $\{\overline{\beta}_{\cdot,\cdot}\}$ sequence, we have

$$\begin{aligned} \overline{\beta}^{s+1}_{i,j} &= D(i,j)\overline{\beta}^{s+1}_j\\ &= D(i,j)\left\{\sum_{\ell=0}^{j}\binom{(s+1)(d-d_0)-\ell}{j-\ell}(\Delta n)^{j-\ell}\overline{\beta}^s_\ell\right\}. \end{aligned}$$

Therefore, to prove the claim, it suffices to match the coefficients of $\overline{\beta}^s_\ell$ for $0 \le \ell \le s(d-d_0)$, as $\overline{\beta}^s_\ell = 0$ for $\ell > s(d-d_0)$ from the definition of $\overline{\beta}^s_{\cdot}$, and $A_\ell = 0$ for $\ell \ge j$. In other words, we only need to show $A_\ell = D(i,j)\binom{(s+1)(d-d_0)-\ell}{j-\ell}$. By using iteratively the identity $\binom{n}{k} = \binom{n}{k-1} + \binom{n-1}{k-1}$, one has

$$
\begin{aligned}
&\sum_{k=0}^{j-\ell}\binom{j-\ell}{k}(-1)^k\underline{\odot}(d-i-k+1-\ell;(s+1)(d-d_0)-\ell)\\
&=\underline{\odot}((s+1)(d-d_0)-\ell;1)\\
&\times\sum_{k=0}^{j-\ell-1}\binom{j-\ell-1}{k}(-1)^k\\
&\quad\underline{\odot}(d-i-k-\ell;(s+1)(d-d_0)-1-\ell)\\
&\qquad\cdots\\
&=\underline{\odot}((s+1)(d-d_0)-\ell;j-\ell-1)\\
&\times\sum_{k=0}^{1}\binom{1}{k}(-1)^k\underline{\odot}(d-i-k+2-j;(s+1)(d-d_0)+1-j)\\
&=\underline{\odot}((s+1)(d-d_0)-\ell;j-\ell)\\
&\quad\underline{\odot}(d-i+1-j;(s+1)(d-d_0)-j).
\end{aligned}
$$

Lastly, by direct calculation, we have

$$
\begin{aligned}
A_\ell &= \frac{C(i,j,\ell)}{\overline{\odot}(i_0+1;(s+1)(d-d_0))}\\
&\quad\times\underline{\odot}((s+1)(d-d_0)-\ell;j-\ell)\\
&\qquad\times\underline{\odot}(d-i+1-j;(s+1)(d-d_0)-j)\\
&= D(i,j)\binom{(s+1)(d-d_0)-\ell}{j-\ell}.
\end{aligned}
$$

The proof is complete. □

Proof of Lemma 31: Define for $i\in[1;(s+1)d_0-sd+1]$

$$
M_i \equiv \sum_{k=i}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}}T_k^2.
$$

Inequality (60) entails that $1\gtrsim c\sum_{i=1}^{(s+1)d_0-sd+1}M_i$ for some $c=c(d)$. We have for $i\in[1;(s+1)d_0-sd+1]$,

$$
\begin{aligned}
M_i &= \frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}T_i^2+\\
&\sum_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}}\frac{(\overline{v}^{s+1}_{\bar{i},k-i}\cdot T_k)^2}{(\overline{v}^{s+1}_{\bar{i},k-i})^2}\\
&\geq\left(\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\wedge\bigwedge_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}^{s+1}_{\bar{i},k-i})^2}\right)\\
&\quad\left(T_i^2+\sum_{k=i+1}^{(s+1)d_0-sd+1}(\overline{v}^{s+1}_{\bar{i},k-i}\cdot T_k)^2\right)\\
&\gtrsim\left(\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\wedge\bigwedge_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}^{s+1}_{\bar{i},k-i})^2}\right)\\
&\quad\left(T_i+\sum_{k=i+1}^{(s+1)d_0-sd+1}\overline{v}^{s+1}_{\bar{i},k-i}\cdot T_k\right)^2\\
&=\left(\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\wedge\bigwedge_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}^{s+1}_{\bar{i},k-i})^2}\right)\\
&\quad\left(\sum_{k=0}^{(d-d_0)(s+1)}\overline{\beta}^{s+1}_{i,k}a^{k_0-2-s}_{i+k}\right)^2\\
&\gtrsim\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\cdot\left(\sum_{k=0}^{(d-d_0)(s+1)}\overline{\beta}^{s+1}_{i,k}a^{k_0-2-s}_{i+k}\right)^2.
\end{aligned}
$$

Here, the second identity follows from Lemma 32, and the last inequality follows, by definition of $\{\overline{v}_{\cdot,\cdot}\}$ and the condition $n_{k_0;k_0-1}\geq n_{k_0-1-s;k_0-2-s}$, from the calculation:

$$
\begin{aligned}
&\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}\wedge\bigwedge_{k=i+1}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}}{n^{2(k-1)}(\overline{v}^{s+1}_{\bar{i},k-i})^2}\\
&\asymp\bigwedge_{k=i}^{(s+1)d_0-sd+1}\frac{(n-n_{k_0-1})^{2k-1}(n_{k_0-1-s}-n_{k_0-2-s})^{-2(k-i)}}{n^{2(k-1)}}\\
&\asymp\frac{(n-n_{k_0-1})^{2i-1}}{n^{2(i-1)}}.
\end{aligned}
$$

Putting together the lower bounds for M_i, $i\in[1;(s+1)d_0-sd+1]$ yields the result. □

Lemma 33: Fix any $1\leq s\leq\lfloor(d_0+1)/(d-d_0)\rfloor$ and $1\leq i\leq sd_0-(s-1)d+1$. For any $0\leq j_1\leq j_2\leq s(d-d_0)$, define the following two quantities:

$$
\begin{aligned}
\underline{S}(j_1)&\equiv\underline{S}(j_1;d,d_0,s)\\
&\equiv\prod_{\ell=1}^{\lfloor j_1/(d-d_0)\rfloor}n^{d-d_0}_{k_0-\ell;k_0-1-s}\times n^{\mathrm{Mod}(j_1;d-d_0)}_{k_0-1-\lfloor j_1/(d-d_0)\rfloor;k_0-1-s},\\
\overline{S}(j_2)&\equiv\overline{S}(j_2;d,d_0,s)\\
&\equiv\prod_{\ell=-\lfloor-j_2/(d-d_0)\rfloor+1}^{s}n^{d-d_0}_{k_0-\ell;k_0-1-s}\\
&\quad\times n^{\mathrm{Mod}(-j_2;d-d_0)}_{k_0-(-\lfloor-j_2/(d-d_0)\rfloor);k_0-1-s}.
\end{aligned}
$$

Then, there exists some positive constant $c=c(d)$ such that

$$
\frac{\overline{\beta}^s_{i,j_2}}{\overline{\beta}^s_{i,j_1}}\geq c\frac{\prod_{\ell=1}^s n^{d-d_0}_{k_0-\ell;k_0-1-s}}{\underline{S}(j_1)\overline{S}(j_2)}.
$$

When $j_1=j_2$, the product on the right hand side is to be understood as 1.

Proof: We only prove the special case $d_0=d-1$ (the proof for the general case is completely analogous). Then $k_0=d+2$, and

$$
\underline{S}(j_1)=\prod_{\ell=1}^{j_1}n_{d+2-\ell;d+1-s},\quad\overline{S}(j_2)=\prod_{\ell=j_2+1}^{s}n_{d+2-\ell;d+1-s},
$$

so we only need to prove for $s\in[1;d]$, $i\in[1;d+1-s]$, and $0\leq j_1\leq j_2\leq s$,

$$
\frac{\overline{\beta}^s_{i,j_2}}{\overline{\beta}^s_{i,j_1}}\geq c\prod_{k=j_1+1}^{j_2}n_{d+2-k;d+1-s}.
$$

We prove this by induction on s.

First consider $s=1$. Then $\overline{\beta}^1_j=n^j_{d+1;d}$, and $\overline{\beta}^1_{i,j}=D(i,j)\overline{\beta}^1_j\asymp n^j_{d+1;d}$. The only non-trivial case is $j_1=0$, $j_2=1$, so the claim follows.

Suppose the claim holds up to $s-1$. Fix any $1\leq j_1\leq j_2\leq s$. The claim clearly holds for $j_1=j_2=s$. If $j_2=s$ and

$j_1 \le s-1$, then it holds by the recursion formula of $\{\overline{\beta}_j^s\}_{j=0}^s$ in (32) that $\overline{\beta}_{i,s}^s/\overline{\beta}_{i,j_1}^s \asymp n_{d+2-s;d+1-s}\overline{\beta}_{s-1}^s/\overline{\beta}_{j_1}^s$, and we can reduce to the following case with $1 \le j_1 \le j_2 \le s-1$. For this case, note that

$$\begin{aligned}\prod_{k=j_1+1}^{j_2} n_{d+2-k;d+1-s} &= \prod_{k=j_1+1}^{j_2} (n_{d+2-k;d+2-s}+n_{d+2-s;d+1-s})\\ &= \sum_{k=0}^{j_2-j_1} n_{d+2-s;d+1-s}^k \\ &\qquad \sum_{j_1+1\le m_1\neq\ldots\neq m_{j_2-j_1-k}\le j_2} n_{d+2-m_1;d+2-s}\cdots n_{d+2-m_k;d+2-s}\\ &\asymp \bigvee_{k=0}^{j_2-j_1}\Big\{ n_{d+2-s;d+1-s}^k \prod_{m=1}^{j_2-j_1-k} n_{d+2-j_1+m;d+2-s}\Big\}.\end{aligned}$$

Treating the above display as a polynomial of $n_{d+2-s;d+1-s}$, it suffices to match the corresponding coefficients of $n_{d+2-s;d+1-s}^k$ for $k \in [0; j_2-j_1]$ in $\overline{\beta}_{i,j_2}^s/\overline{\beta}_{i,j_1}^s$. To this end, we have

$$\begin{aligned}\frac{\overline{\beta}_{i,j_2}^s}{\overline{\beta}_{i,j_1}^s} &\asymp \frac{\bigvee_{\ell=0}^{j_2} n_{d+2-s;d+1-s}^{j_2-\ell}\overline{\beta}_\ell^{s-1}}{\bigvee_{\ell=0}^{j_1} n_{d+2-s;d+1-s}^{j_1}\overline{\beta}_\ell^{s-1}}\\ &\asymp \bigvee_{k=0}^{j_2-j_1} \frac{\bigvee_{\ell=0}^{j_1} n_{d+2-s;d+1-s}^{j_1+k-\ell}\overline{\beta}_{j_2-j_1-k+\ell}^{s-1}}{\bigvee_{\ell=0}^{j_1} n_{d+2-s;d+1-s}^{j_1-\ell}\overline{\beta}_\ell^{s-1}}\\ &\ge \bigvee_{k=0}^{j_2-j_1}\Big\{ n_{d+2-s;d+1-s}^k \bigwedge_{\ell=0}^{j_1} \frac{\overline{\beta}_{j_2-j_1-k+\ell}^{s-1}}{\overline{\beta}_\ell^{s-1}}\Big\}\\ &\quad \text{(by Lemma 28)}\\ &\gtrsim \bigvee_{k=0}^{j_2-j_1}\Big\{ n_{d+2-s;d+1-s}^k \bigwedge_{\ell=0}^{j_1}\prod_{m=\ell+1}^{j_2-j_1-k+\ell} n_{d+2-m;d+2-s}\Big\}\\ &\quad \text{(by induction)}\\ &= \bigvee_{k=0}^{j_2-j_1}\Big\{ n_{d+2-s;d+1-s}^k \prod_{m=j_1+1}^{j_2-k} n_{d+2-m;d+2-s}\Big\}\\ &\quad \text{(minimum at } \ell=j_1),\end{aligned}$$

matching the calculation in the previous display, completing the proof. □

Acknowledgment

The authors would like to thank the associate editor and the anonymous referees for helpful comments.

References

[1] C. de Boor, *A Practical Guide to Splines* (Applied Mathematical Sciences), vol. 27. New York, NY, USA: Springer-Verlag, 1978.

[2] P. J. Green and B. W. Silverman, *Nonparametric Regression Generalized Linear Models* (Monographs on Statistics and Applied Probability), vol. 58. London, U.K.: Chapman & Hall, 1994, doi: 10.1007/978-1-4899-4473-3.

[3] G. Wahba, *Spline Models for Observational Data* (CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics), vol. 59. Philadelphia, PA, USA: SIAM, 1990, doi: 10.1137/1.9781611970128.

[4] E. Mammen and S. van de Geer, "Locally adaptive regression splines," *Ann. Statist.*, vol. 25, no. 1, pp. 387–413, Feb. 1997.

[5] R. J. Tibshirani, "Adaptive piecewise polynomial estimation via trend filtering," *Ann. Statist.*, vol. 42, no. 1, pp. 285–323, 2014, doi: 10.1214/13-AOS1189.

[6] R. Baranowski, Y. Chen, and P. Fryzlewicz, "Narrowest-over-threshold detection of multiple change points and change-point-like features," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 81, no. 3, pp. 649–672, Jul. 2019.

[7] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992, doi: 10.1016/0167-2789(92)90242-F.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 43, no. 1, pp. 129–159, 1999, doi: 10.1137/S003614450037906X.

[9] P. L. Davies and A. Kovac, "Local extremes, runs, strings and multiresolution," *Ann. Statist.*, vol. 29, no. 1, pp. 1–65, 2001, doi: 10.1214/aos/996986500.

[10] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. Roy. Statist. Soc. B, Statist. Methodol.*, vol. 67, no. 1, pp. 91–108, 2005, doi: 10.1111/j.1467-9868.2005.0%0490.x.

[11] G. Steidl, S. Didas, and J. Neumann, "Splines in higher order TV regularization," *Int. J. Comput. Vis.*, vol. 70, no. 3, pp. 241–255, Dec. 2006.

[12] A. Rinaldo, "Properties and refinements of the fused lasso," *Ann. Statist.*, vol. 37, no. 5B, pp. 2922–2952, 2009, doi: 10.1214/08-AOS665.

[13] Z. Harchaoui and C. Lévy-Leduc, "Multiple change-point estimation with a total variation penalty," *J. Amer. Stat. Assoc.*, vol. 105, no. 492, pp. 1480–1493, 2012, doi: 10.1198/jasa.2010.tm09181.

[14] J.-C. Hütter and P. Rigollet, "Optimal rates for total variation denoising," in *Proc. Conf. Learn. Theory*, vol. 2016, pp. 1115–1146.

[15] A. S. Dalalyan, M. Hebiri, and J. Lederer, "On the prediction performance of the Lasso," *Bernoulli*, vol. 23, no. 1, pp. 552–581, 2017, doi: 10.3150/15-BEJ756.

[16] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "ℓ_1 trend filtering," *SIAM Rev.*, vol. 51, no. 2, pp. 339–360, 2009, doi: 10.1137/070690274.

[17] Y.-X. Wang, A. Smola, and R. Tibshirani, "The falling factorial basis and its statistical applications," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 730–738.

[18] A. Guntuboyina, D. Lieu, S. Chatterjee, and B. Sen, "Adaptive risk bounds in univariate total variation denoising and trend filtering," *Ann. Statist.*, vol. 48, no. 1, pp. 205–229, 2020, doi: 10.1214/18-AOS1799.

[19] S. van de Geer, "On tight bounds for the Lasso," *J. Mach. Learn. Res.*, vol. 19, p. 46 and 48, 2018.

[20] Z. Fan and L. Guan, "Approximate ℓ_0-penalized estimation of piecewise-constant signals on graphs," *Ann. Statist.*, vol. 46, no. 6B, pp. 3217–3245, 2018.

[21] I. Auger and C. Lawrence, "Algorithms for the optimal identification of segment neighborhoods," *Bull. Math. Biol.*, vol. 51, no. 1, pp. 39–54, 1989, doi: 10.1016/S0092-8240(89)8004% 7-3.

[22] G. Winkler and V. Liebscher, "Smoothers for discontinuous signals," *J. Nonparametric Statist.*, vol. 14, nos. 1–2, pp. 203–222, 2002, doi: 10.1080/10485250211388.

[23] B. Jackson *et al.*, "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Process. Lett.*, vol. 12, no. 2, pp. 105–108, Feb. 2005.

[24] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler, "Complexity penalized M-estimation: Fast computation," *J. Comput. Graph. Statist.*, vol. 17, no. 1, pp. 201–224, 2008, doi: 10.1198/106186008X285591.

[25] P. Fearnhead, R. Maidstone, and A. Letchford, "Detecting changes in slope with an L_0 penalty," *J. Comput. Graph. Statist.*, vol. 28, no. 2, pp. 265–275, 2019, doi: 10.1080/10618600.2018.1512868.

[26] C. Gao, F. Han, and C.-H. Zhang, "On estimation of isotonic piecewise constant signals," *Ann. Statist.*, vol. 48, no. 2, pp. 629–654, 2020, doi: 10.1214/18-AOS1792.

[27] F. Balabdaoui and J. A. Wellner, "Estimation of a K-monotone density: Limit distribution theory and the spline connection," *Ann. Statist.*, vol. 35, no. 6, pp. 2536–2564, 2007, doi: 10.1214/009053607000000262.

[28] S. Chatterjee, A. Guntuboyina, and B. Sen, "On risk bounds in isotonic and other shape restricted regression problems," *Ann. Statist.*, vol. 43, no. 4, pp. 1774–1800, 2015, doi: 10.1214/15-AOS1324.

[29] C.-H. Zhang, "Risk bounds in isotonic regression," *Ann. Statist.*, vol. 30, no. 2, pp. 528–555, 2002, doi: 10.1214/aos/1021379864.

[30] P. C. Bellec, "Sharp oracle inequalities for least squares estimators in shape restricted regression," *Ann. Statist.*, vol. 46, no. 2, pp. 745–780, Apr. 2018.

[31] A. Guntuboyina and B. Sen, "Global risk bounds and adaptation in univariate convex regression," *Probab. Theory Rel. Fields*, vol. 163, nos. 1–2, pp. 379–411, Oct. 2015, doi: 10.1007/s00440-014-0595-3.
[32] Y. Yu, S. Chatterjee, and H. Xu, "Localising change points in piecewise polynomials of general degrees," 2020, *arXiv:2007.09910*.
[33] D. L. Donoho and I. M. Johnstone, "Minimax risk over l_p-balls for l_q-error," *Probab. Theory Rel. Fields*, vol. 99, no. 2, pp. 277–303, 1994, doi: 10.1007/BF01199026.
[34] L. Birgé and P. Massart, "Gaussian model selection," *J. Eur. Math. Soc.*, vol. 3, no. 3, pp. 203–268, 2001, doi: 10.1007/s100970100031.
[35] L. Birgé and P. Massart, "Rates of convergence for minimum contrast estimators," *Probab. Theory Rel. Fields*, vol. 97, nos. 1–2, pp. 113–150, Mar. 1993, doi: 10.1007/BF01199316.
[36] A. Barron, L. Birgé, and P. Massart, "Risk bounds for model selection via penalization," *Prob. Theory Rel. Fields*, vol. 113, no. 3, pp. 301–413, 1999, doi: 10.1007/s004400050210.
[37] P. Massart, *Concentration Inequalities Model Selection* (Lecture Notes in Mathematics), vol. 1896. Berlin, Germany: Springer, 2007.
[38] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *J. Amer. Statist. Assoc.*, vol. 107, no. 500, pp. 1590–1598, 2012, doi: 10.1080/01621459.2012.737745.
[39] R. Maidstone, T. Hocking, G. Rigaill, and P. Fearnhead, "On optimal multiple changepoint algorithms for large data," *Statist. Comput.*, vol. 27, no. 2, pp. 519–533, 2017, doi: 10.1007/s11222-016-9636-3.
[40] G. Balázs, A. György, and C. Szepesvári, "Near-optimal max-affine estimators for convex regression," in *Proc. 18th Int. Conf. Artif. Intell. Statist.*, 2015, pp. 56–64.
[41] F. Ye and C.-H. Zhang, "Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls," *J. Mach. Learn. Res.*, vol. 11, pp. 3519–3540, 2010.
[42] G. Raskutti, M. J. Wainwright, and B. Yu, "Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6976–6994, Oct. 2011.
[43] R. Durrett, *Probability—Theory Examples* (Cambridge Series in Statistical and Probabilistic Mathematics), vol. 49. Cambridge, U.K.: Cambridge Univ. Press, Cambridge, 2019, doi: 10.1017/9781108591034.
[44] A. B. Tsybakov, *Introduction to Nonparametric Estimation* (Springer Series in Statistics). New York, NY, USA: Springer, 2009, doi: 10.1007/b13794.
[45] J. H. van Lint, *Introduction to Coding Theory* (Graduate Texts in Mathematics), vol. 86, 3rd ed. Berlin, Germany: Springer-Verlag, doi: 10.1007/978-3-642-58575-3.
[46] V. H. de la Peña and E. Giné, *Decoupling* (Probability and its Applications). New York, NY, USA: Springer-Verlag, 1999, doi: 10.1007/978-1-4612-0537-1.
[47] R. T. Rockafellar, *Convex Analysis* (Princeton Landmarks in Mathematics). Princeton, NJ, USA: Princeton Univ. Press, 1997.

Yandi Shen received the bachelor's degree in mathematics from Zhejiang University, China, in 2016, and the Ph.D. degree in statistics from the University of Washington, Seattle, WA, USA, in 2021. He is currently a William H. Kruscal Instructor with the Department of Statistics, The University of Chicago. He is broadly interested in nonparametric and semi-parametric statistics, high dimensional inference, and applied probability.

Qiyang Han received the bachelor's degree in mathematics from Fudan University, China, in 2013, and the Ph.D. degree from the University of Washington, Seattle, WA, USA, in 2018. He is currently an Assistant Professor of statistics at Rutgers, the State University of New Jersey. He is broadly interested in mathematical statistics and high-dimensional probability, with a particular focus on empirical process theory and its applications to nonparametric and high dimensional statistics.

Fang Han received the Ph.D. degree from the Department of Biostatistics, Johns Hopkins University. He is currently an Associate Professor of statistics and economics at the University of Washington. His main research interests are in rank- and graph-based statistics, statistical optimal transport, mixture models, nonparametric and semi-parametric regressions, time series analysis, and random matrix theory.