Title
Shape-constrained regression in misspecified and multivariate settings

Author
Fang, Billy

Publication Date
2020

Peer reviewed|Thesis/dissertation

Shape-constrained regression in misspecified and multivariate settings

by

Billy Fang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Statistics

and the Designated Emphasis

in

Communication, Computation and Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Adityanand Guntuboyina, Co-chair
Professor Martin J. Wainwright, Co-chair
Professor Peter L. Bartlett
Professor Nikhil Srivastava

Spring 2020

Shape-constrained regression in misspecified and multivariate settings

Abstract

Shape-constrained regression in misspecified and multivariate settings

by

Billy Fang

Doctor of Philosophy in Statistics

Designated Emphasis in Communication, Computation and Statistics

University of California, Berkeley

Professor Adityanand Guntuboyina, Co-chair

Professor Martin J. Wainwright, Co-chair

In the context of nonparametric regression, shape-constrained estimators such as isotonic regression have a number of attractive properties. The shape constraints are typically mild and are often justified by the context of the estimation problem, allowing for more flexible fits than a more restrictive parametric model; yet at the same time such estimators can be computed efficiently and have reasonable risk properties. Additionally, these estimators are free of tuning parameters and often exhibit adaptation to certain types of hidden structure in the data (e.g., isotonic regression and piecewise constant functions). Properties of such estimators in the setting of univariate function estimation are well-studied. This thesis provides some new insights for shape-constrained regression on two fronts: misspecification and the multivariate setting.

In Chapter 2, we study least squares estimators under polyhedral convex constraints. Many estimators fall in this category, including shape constrained estimators like isotonic regression and convex regression, as well as other estimators like LASSO. We give an explicit geometric characterization of how the risk of such an estimator behaves when the truth it is trying to estimate lies outside of the constraint set, and show how this result generalizes what is known in the well-specified setting. This result leads to a better understanding of how isotonic regression behaves when applied in settings where the true function is not isotonic. This chapter is joint work with Adityanand Guntuboyina.

There has been recent interest in understanding shape-constrained estimation in the multivariate setting. It is known that multivariate isotonic regression suffers from the curse of dimensionality, making it unsuitable for most high-dimensional applications. In Chapter 3, we propose and analyze an alternate multivariate generalization of isotonic regression that uses a notion of monotonicity called entire monotonicity. It is restrictive enough to avoid the curse of dimensionality (the dependence on the dimension is only in the logarithmic terms), yet rich enough to include non-smooth functions like rectangular piecewise constant func-

tions. In parallel, we also propose and analyze a generalization of total variation denoising using a notion called Hardy-Krause variation, and show it has similar computational and statistical properties as the entirely monotonic estimator. This chapter is joint work with Adityanand Guntuboyina and Bodhisattva Sen.

Finally in Chapter 4, we show how entire monotonicity can be viewed as the introduction of "positive interactions" to the interaction-less additive monotonic model. In making this comparison between entire monotonicity and additive monotonicity, we introduce various intermediate models that have different combinations of interactions. We prove a risk rate for some of these intermediate models that generalizes the analogous risk rate for entire monotonicity established in the previous chapter, and also discuss hypothesis testing for interaction terms in these models. This chapter is joint work with Adityanand Guntuboyina and Hansheng Jiang.

To my parents

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I must thank my wonderful advisers Professor Adityanand Guntuboyina and Professor Martin Wainwright. I feel so privileged to work with two great thinkers who share so many positive qualities that I value highly. They not only are principled and methodical in their research and always searching for the right questions to ask, but also are terrific communicators who value clarity of writing in their papers, and excellent speakers who command the respect of their colleagues who attend their talks, and of their students who attend their classes. I am thankful for Aditya's patience from my very first semester at Berkeley when I was trying to get started with research, and his encouragement was instrumental in my development. He always has something to share from his endless fountain of ideas, holds me to high standards for quality of research, and was also willing to help me work out technical details. I am always astounded by both the depth and breadth of Martin's expertise. His sharp intuition lets him cut right to the heart of the issue and see what is interesting about a problem or what is worth exploring. At the same time, his familiarity and mastery of a wide array of techniques from statistics, mathematics, and related fields helps him craft and navigate complex arguments, while still being clear in presentation and writing. Thank you both for your guidance during my PhD and for always pushing me to improve the quality of my work.

I would also like to thank Professors Peter Bartlett and Nikhil Srivastava for being part of my committee. I am also thankful for the several conversations I had with Peter during my first year when I was trying to find my footing in the research world. I also appreciate the enlightening conversations I had with Professors Peter Bickel, Thomas Courtade, Deborah Nolan, Chris Paciorek, Jim Pitman, and Allan Sly.

Next, I must thank my great collaborators Hansheng Jiang, Bodhisattva Sen, and Yuting Wei. Hansheng was very quick in grasping the technical details of a project that Aditya, Bodhi, and I had worked on, and was able to immediately help us make progress on related research. Bodhi often provided perspectives and ideas that were complementary to Aditya's, and the many conversations the three of us had during Bodhi's stay at Berkeley were very fruitful. As Yuting's academic sibling, I tried my best to follow in her footsteps, but she left big shoes to fill. During my collaboration with her, I was amazed at her mathematical wizardry and tenacity, as well as her depth of understanding of statistics. At the same time she is so friendly and always gave me advice when I needed it. I am also thankful for the memorable week in Banff with Aditya, Bodhi, and Yuting for the workshop on shape-constrained methods. I'll remember to bring better footwear next time.

Additional thanks go to the journal editors and referees whose insightful comments improved the quality of my papers, as well as to Professors Dennis Amelunxen and Frank Gao for their helpful responses to my technical questions. I am also thankful for the support I received through the NSF Graduate Research Fellowship Program.

I would like to thank instructors Wenpin Tang, Oscar Hernan Madrid Padilla, Hank Ibser, and Gaston Sanchez, my fellow GSIs Sourav Sarkar, Milind Hegde, Yiming Shi, Jake Spertus, and Frank Qiu, as well as all my students for helping me be a better teacher. I

would also like to thank all the wonderful people in the department who keep things running smoothly and always are happy to help me with various problems at a moment's notice: Keyla Gomez, Ryan Lovett, Mary Melinn, Chris Paciorek, La Shana Porlaris, Majabeen Samadi, Laura Slakey, Luis Torres, Denise Yee (and Tuesday the dog), and many more. Thank you for making the department a wonderful place to be.

During my years at Berkeley, many of the more senior graduate students and postdocs were exemplary role models and gave me insightful career advice. These people include Yuansi Chen, Hye Soo Choi, Ryan Giordano, Peter Hintz, Nhat Ho, Johnny Hong, Lihua Lei, Ashwin Pananjady, Aaditya Ramdas, Yuting Wei, and Fanny Yang. I also thank my peers for their companionship in the graduate school experience, for contributing to a rich and diverse intellectual environment here at Berkeley, and more broadly for helping me through the ups and downs of life: Taejoo Ahn, Olivia Angiuli, Efe Aras, Zsolt Bartha, Joe Borja, Wilson Cai, Jianbo Chen, Zihao Chen, Xiang Cheng, Stephanie DeGraaf, Andrew Do, Raaz Dwivedi, Avishek Ghosh, Amanda Glazer, Geno Guerra, Vipul Gupta, Sang Min Han, Ella Hiesmayr, Steve Howard, Miyabi Ishihara, Hansheng Jiang, Cheng Ju, Koulik Khamaru, Kuan-Yun Lee, Xiao Li, Michael Lim, Tianyi Lin, Bryan Liu, Lydia Liu, Tyler Maltba, Jamie Murdoch, Kellie Ottoboni, Briton Park, Soham Phade, Yannik Pitcan, Frank Qiu, Jake Soloff, Dan Soriano, Tiffany Tang, Nilesh Tripuraneni, Alexander Tsigler, Maxim Rabinovich, Sujayam Saha, Sourava Sarkar, Vaishaal Shankar, Max Simchowitz, Jake Spertus, Mitchell Stern, Sara Stoudt, Simon Walter, Andre Waschka, Yu Wang, Ashia Wilson, Jason Wu, Elizabeth Yang, Puyudi Yang, Yuting Ye, Michelle Yu, Yun Zhou, Chelsea Zhang, and many more. Further thanks go to my friends outside of Berkeley who supported me during my PhD: Allen, Andrew, Bobby, Brian, Edward, Eric, Gabe, George, Ina, Ingrid, Jamie, Jeff, Kevin, Kimberly, Lisa, Mark, Michael, Min Joo, Spencer, Victor, Vivian, and Yeri.

I also thank Professor David Milnes and the UC Berkeley Symphony Orchestra for the whirlwind of music-making during my five years in the orchestra, and for trusting me to lead the orchestra as concertmaster during my last two years here. I am thankful for the many talented people I got to work with in the orchestra: Alex, Andre, Andrew C., Andrew R., Angel, Anna F., Anna O., Arjun, Aster, Bethany, Beverly, Cassandra, Chelsie, Cindy, Daniel G., Daniel W., Dannver, Doron, Drew, Ethan, Gabrielle, Gavin, Grace, Grady, Isaac, Jane, Jen, Jihoon, Jong, Josephine, Kana, Kane, Katherine, Kavya, Kyle, Leo, Liam, Lucia, Lucy, Melissa, Michael, Michelle, Mosa, Naoya, Nick, Noam, Pearl, Peter, Raina, Rebecca, Richard, Robert, Ryan, Sergio, Shalini, Simon, Sophie, Sneha, Thomas, Tony, Vicky, Yi-shen, Zhimin, and many more. Special thanks to Jolie and Emiel, the better two thirds of the Golden Bear Trio, for the opportunity to travel to Austin and Seattle for competitions and to play for hundreds of kids during our tour of Washington. Many thanks also go to my violin and piano teachers for helping me get my rusty playing back in shape: Antoine van Dongen, Eunbyol Ko, Jooyeon Kong, and Jeffrey Sykes.

Finally, this journey would not have been possible without my family. Thank you Mom, Dad, and Demi for sacrificing so much to raise me and for providing an environment for me to succeed. My accomplishments are as much yours as they are mine, if not more. Thank you for your love and unwavering support from day one.

Chapter 1

Introduction

In regression, we obtain noisy observations $y_i = f^*(\mathbf{x}_i) + \xi_i$ of an unknown function f^* from some function class $\mathcal{F}$ at design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and seek to estimate f^*. In the particular case of nonparametric shape-constrained regression, the function class $\mathcal{F}$ imposes certain constraints on its functions. One of the most well-studied examples of shape-constrained regression is isotonic regression [14, 5], where $\mathcal{F}$ consists of functions $f : [0, 1] \to \mathbb{R}$ that are nondecreasing. Another common example is convex regression (e.g., [48, 27]), where $\mathcal{F}$ consists of convex functions. The choice of function class is usually informed by the context of the regression problem, where the practitioner knows from domain knowledge that the unknown function f^* satisfies some shape constraint.

Shape-constrained regression has a number of attractive properties. The assumptions imposed on the functions are usually relatively mild and justified by the context of the problem, allowing much more flexibility when compared to more restrictive models like parametric models. One feature that distinguishes shape-constrained regression from other nonparametric regression problems is that one can use least squares or maximum likelihood without explicit regularization to obtain estimators that are free of tuning parameters. Certain shape-constrained least squares estimators also exhibit adaptation to certain types of hidden structure in the unknown function. For example, in isotonic regression, if the unknown function f^* is piecewise constant and nondecreasing, then the least squares estimator with respect to the class of nondecreasing functions can estimate the function nearly as well as an oracle estimator that knows the locations of the constant pieces [19]. The survey by Guntuboyina and Sen [45] provides more examples and discussion of shape-constrained regression. In this thesis, we provide some insight into shape-constrained regression on two fronts: misspecification and the multivariate setting.

1.1 Misspecification

In Chapter 2, we study the risk of constrained least squares estimators

$$\widehat{\theta}(Y) = \operatorname*{argmin}_{\theta \in \mathcal{C}} \|\theta - Y\|^2,$$

with respect to some closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Here, Y is a noisy vector with mean $\theta^* \in \mathcal{C}$ and variance $\sigma^2 I$. Many least squares estimators (including isotonic regression and convex regression) can be cast in this form by identifying θ^* with $(f^*(\mathbf{x}_1), \ldots, f^*(\mathbf{x}_n))$. Oymak and Hassibi [70] showed that the risk of such estimators is governed by the geometry of the constraint set $\mathcal{C}$ near θ^*. Specifically, they showed that the normalized risk $\sigma^{-2}\mathbb{E}\|\widehat{\theta}(Y) - \theta^*\|^2$ is upper bounded by the statistical dimension of the tangent cone of $\mathcal{C}$ at θ^*, and that this upper bound is tight in the limit as $\sigma \downarrow 0$. However, the above results are under the well-specified assumption $\theta^* \in \mathcal{C}$. In practice we cannot ensure that $\theta^* \in \mathcal{C}$ when using this estimator, so it is worth asking how the estimator behaves in the misspecified setting $\theta^* \notin \mathcal{C}$. Bellec [10] proved an analogous upper bound for the normalized risk, but we show that, unlike in the well-specified setting, this upper bound is not tight as $\sigma \downarrow 0$. In particular we provide an explicit formula for the low noise limit of the normalized risk in the case where $\mathcal{C}$ is a polyhedral cone, and observe that it can be strictly smaller than the upper bound. Essentially, the reason why the risk can be smaller in the in misspecified setting is due to certain directions of the noise in the observations being eliminated under the least squares projection to the boundary of $\mathcal{C}$. One application of this result is an explicit characterization of how properties of the target function govern the risk of isotonic regression when the target function is not nondecreasing. This chapter is based on joint work with Adityanand Guntuboyina [33].

1.2 Multivariate shape constraints

There has been recent interest in understanding shape-constrained estimators in the multivariate setting, such as regression with multivariate functions $f : [0,1]^d \to \mathbb{R}$, or estimation of multivariate densities. However, in many cases the estimators suffer from the curse of dimensionality, in that the risk is of the order $n^{-1/d}$, so that amount of observations needed to obtain a certain level of error grows exponentially in the dimension d (e.g., multivariate isotonic regression [47] and log-concave density estimation [53]). Thus it is meaningful to find classes of multivariate functions that are rich enough to be useful, but not so large that they suffer from the curse of dimensionality. In Chapter 3, we propose and analyze a multivariate generalization of isotonic regression using a notion called entire monotonicity. Entire monotonicity has appeared in other contexts before, but has not been studied in this nonparametric regression framework. We prove an upper bound and a minimax lower bound for this least squares estimator to show that it avoids the curse of dimensionality to some extent: the main term of the risk is $n^{-2/3}(\log n)^{\frac{2d-1}{3}}$, so the price of going from univariate isotonic regression to this estimator is only in additional logarithmic factors. Although this class is small enough to avoid the curse of dimensionality, it is still rich enough to contain non-smooth functions like rectangular piecewise constant functions. In fact, it adapts to rectangular piecewise constant functions in the same way that univariate isotonic regression adapts to piecewise constant functions. In this chapter we also propose a multivariate generalization of univariate total variation denoising using a notion called Hardy-Krause variation.

Compared to the entirely monotonic estimator, this Hardy-Krause variation denoising estimator drops the shape constraint at the cost of introducing a tuning parameter, but has similar statistical properties to the entirely monotonic estimator. Additionally, we show that these two estimators can be computed by solving a nonnegative least squares problem and a LASSO problem respectively. This chapter is joint work with Adityanand Guntuboyina and Bodhisattva Sen [34].

Finally in Chapter 4, we provide a different perspective on entire monotonicity by comparing it to the additive monotonic model [6], which considers the class of functions $f : [0,1]^d \to \mathbb{R}$ of the form $f(\mathbf{x}) = \sum_{j=1}^{d} f_j(x_j)$ where each f_j is nondecreasing . This latter model is relatively simple, since each function's behavior can be decomposed into individual univariate monotonic functions, but it excludes the possibility of interactions among its covariates: the effect of a change in one covariate when the other covariates are held fixed does not depend on the actual value of those other covariates. We show how entire monotonicity introduces interaction terms into the additive model, but also imposes a shape constraint on these interaction terms. In making this comparison between entire monotonicity and additive monotonicity, we introduce various intermediate models that have different combinations of interactions. We prove a risk rate for some of these intermediate models and show that it is almost the same as the analogous risk rate $n^{-2/3}(\log n)^{\frac{2d-1}{3}}$ for entire monotonicity established in the previous chapter, except the exponent of the logarithmic factor is reduced. In the context of linear regression, hypothesis testing is often used to decide whether to include certain interaction terms in a model. In this chapter we also describe how to set up a likelihood ratio test to test for the inclusion of interaction terms in these models, and we apply a result of Menéndez et al. [62] to show that it is dominated by another likelihood ratio test. This chapter is joint work with Adityanand Guntuboyina and Hansheng Jiang.

Chapter 2

On the risk of convex-constrained least squares estimators under misspecification

2.1 Introduction

In many statistical problems, it is common to model the observations $y_1, \dots, y_n \in \mathbb{R}$ as $y_i = \theta_i^* + \sigma z_i$ where $\theta_1^*, \dots, \theta_n^*$ are unknown parameters of interest, $z_1, \dots, z_n$ represent noise or error variables that have mean zero, and $\sigma > 0$ denotes a scale parameter. In vector notation, this is equivalent to writing

$$Y = \theta^* + \sigma Z,$$

where $Y \coloneqq (y_1, \dots, y_n)$, $\theta^* \coloneqq (\theta_1^*, \dots, \theta_n^*)$, and $Z \coloneqq (z_1, \dots, z_n)$. A common instance of this model is the Gaussian sequence model, where the $z_1, \dots, z_n$ are independent standard Gaussian random variables, in which case the model can be written as $Y \sim N(\theta^*, \sigma^2 I_n)$, where I_n is the $n \times n$ identity matrix.

A standard method of estimating θ^* from the observation vector Y is to fix a closed convex set $\mathcal{C}$ of $\mathbb{R}^n$ and use the least squares estimator under the constraint given by $\theta \in \mathcal{C}$. Specifically, the least squares projection is

$$\Pi_{\mathcal{C}}(x) \coloneqq \operatorname*{argmin}_{\theta \in \mathcal{C}} \|x - \theta\|^2,$$

(where $\|\cdot\|$ denotes the standard Euclidean norm in $\mathbb{R}^n$), and one estimates θ^* by

$$\hat{\theta}(Y) \coloneqq \Pi_{\mathcal{C}}(Y).$$

When $\mathcal{C}$ is taken to be $\{X\beta : \|\beta\|_1 \leq R\}$ for some deterministic $n \times p$ matrix X and $R > 0$, this estimator becomes LASSO in the constrained form as originally proposed by Tibshirani [80]. When $\mathcal{C}$ is taken to be $\{X\beta : \min_j \beta_j \geq 0\}$, this estimator becomes nonnegative least

squares. Note that shape restricted regression estimators are special cases of nonnegative least squares for appropriate choices of X (see, for example, Groeneboom and Jongbloed [41]). Also, note that both sets $\{X\beta : \|\beta\|_1 \leq R\}$ and $\{X\beta : \min_j \beta_j \geq 0\}$ are examples of polyhedral sets. Therefore in most applications, the constraint set $\mathcal{C}$ is polyhedral.

There exist many results in the literature studying the accuracy of $\hat{\theta}(Y)$ as an estimator for θ^*. Most of these results make the assumption that $\theta^* \in \mathcal{C}$. In this chapter, we shall refer to this assumption as the *well-specified* assumption. Essentially, the constraint set $\mathcal{C}$ can be taken to be a part of the model specification, and the assumption $\theta^* \in \mathcal{C}$ means that the true mean vector θ^* satisfies the model assumptions, i.e. the model is well-specified.

Under the well-specified assumption, it is reasonable and common to measure the accuracy of $\hat{\theta}(Y)$ via its risk under squared Euclidean distance. More precisely, the risk of $\hat{\theta}(Y)$ is defined by

$$R(\hat{\theta}, \theta^*) := \mathbb{E}_{\theta^*}\|\hat{\theta}(Y) - \theta^*\|^2$$

where $\mathbb{E}_{\theta^*}$ refers to expectation taken with respect to the noise Z in the model $Y = \theta^* + \sigma Z$.

Many results on $R(\hat{\theta}, \theta^*)$ in the well-specified setting are available in the literature. Of all the available results, let us isolate two results from Oymak and Hassibi [70] because of their generality. In the setting where $Z \sim N(0, I_n)$, Oymak and Hassibi [70] first proved the upper bound

$$\frac{1}{\sigma^2} R(\hat{\theta}, \theta^*) \leq \delta(T_{\mathcal{C}}(\theta^*)), \tag{2.1}$$

where $T_{\mathcal{C}}(\theta^*)$ denotes the *tangent cone* of $\mathcal{C}$ at θ^*, defined by

$$T_{\mathcal{C}}(\theta^*) = \mathrm{cl}\left\{\alpha(\theta - \theta^*) : \alpha \geq 0, \theta \in \mathcal{C}\right\}, \tag{2.2}$$

("cl" denotes closure), and where $\delta\left(T_{\mathcal{C}}(\theta^*)\right)$ denotes the *statistical dimension* of the cone $T_{\mathcal{C}}(\theta^*)$. In general, the statistical dimension of a closed cone $T \subseteq \mathbb{R}^n$ is defined as

$$\delta(T) := \mathbb{E}\|\Pi_T(Z)\|^2, \tag{2.3}$$

where the expectation is with respect to $Z \sim N(0, I_n)$. Many properties of the statistical dimension are covered by Amelunxen et al. [3].

In the case when the constraint set $\mathcal{C}$ is a subspace, the estimator $\hat{\theta}(Y)$ is linear and, in this case, it is easy to see that $\delta(T_{\mathcal{C}}(\theta^*))$ is simply the dimension of $\mathcal{C}$, so that inequality (2.1) becomes an equality. For general closed convex sets, it is therefore reasonable to ask how tight inequality (2.1) is. It is not hard to construct examples of $\mathcal{C}$ and $\theta^* \in \mathcal{C}$ where inequality (2.1) is loose for fixed $\sigma > 0$. However Oymak and Hassibi [70] proved remarkably that the upper bound in (2.1) is tight in the limit as $\sigma \downarrow 0$ (we shall refer to this in the sequel as the *low σ limit*); that is, when $Z \sim N(0, I_n)$,

$$\lim_{\sigma \downarrow 0} \frac{1}{\sigma^2} R(\hat{\theta}, \theta^*) = \delta(T_{\mathcal{C}}(\theta^*)). \tag{2.4}$$

In summary, Oymak and Hassibi [70] proved that $\sigma^2\delta(T_{\mathcal{C}}(\theta^*))$ is a nice formula for the risk of $\hat{\theta}(Y)$ that is, in general, an upper bound which is tight in the low σ limit.

We remark that although Oymak and Hassibi [70] state the results (2.1) and (2.4) for the specific case $Z \sim N(0, I_n)$, their proof automatically extends to the more general setting where Z is an arbitrary zero mean random vector with $\mathbb{E}\|Z\|^2 < \infty$ (the components $Z_1, \ldots, Z_n$ of Z can be arbitrarily dependent), provided we generalize the definition (2.3) of statistical dimension by taking the expectation with respect to Z, without assuming Z is standard Gaussian. We refer to this modification of the definition (2.3) as the *generalized statistical dimension* of the cone T. As a slight abuse of notation, we use the same notation $\delta(\cdot)$ for this more general concept, with the understanding that the expectation in the definition is with respect to the distribution of Z. By dropping the Gaussian assumption, the generalized statistical dimension loses much of the interpretability and nice geometric properties of the usual statistical dimension [3], but still serves as an abstract notion of the size of a cone T with respect to a distribution Z.

This chapter deals with the behavior of the estimator $\hat{\theta}(Y)$ when the assumption $\theta^* \in \mathcal{C}$ is violated. We shall refer to the situation when $\theta^* \notin \mathcal{C}$ as the *misspecified* setting. Note that, in practice, one can never know if the unknown θ^* truly lies in $\mathcal{C}$. It is therefore necessary to study the behavior of $\hat{\theta}(Y)$ under misspecification.

For the misspecified setting, one must first note that it is no longer reasonable to measure the performance of $\hat{\theta}(Y)$ by the risk $R(\hat{\theta}, \theta^*)$, simply because $\hat{\theta}(Y)$ is constrained to be in $\mathcal{C}$ and hence cannot be expected to be close to θ^* which is essentially unconstrained. There are two natural notions of accuracy of $\hat{\theta}(Y)$ in the misspecified setting, which we call the *misspecified risk* and the *excess risk*. The misspecified risk is defined as

$$M(\hat{\theta}, \theta^*) := \mathbb{E}_{\theta^*}\|\hat{\theta}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2, \tag{2.5}$$

and the excess risk is defined as

$$E(\hat{\theta}, \theta^*) := \mathbb{E}_{\theta^*}\|\hat{\theta}(Y) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2. \tag{2.6}$$

The misspecified risk, $M(\hat{\theta}, \theta^*)$, is motivated by the observation that, in the misspecifed case, the estimator $\hat{\theta}(Y)$ is really estimating $\Pi_{\mathcal{C}}(\theta^*)$ so it is natural to measure its squared distance from $\Pi_{\mathcal{C}}(\theta^*)$. On the other hand, the excess risk, $E(\hat{\theta}, \theta^*)$, measures the squared distance of the estimator from θ^* relative to the squared distance of $\Pi_{\mathcal{C}}(\theta^*)$ from θ^*. We refer the reader to Bellec [10] and Section 2.2 for some background and basic properties on these notions of accuracy under misspecification. For example, it can be shown that $M(\hat{\theta}, \theta^*)$ is always less than or equal to $E(\hat{\theta}, \theta^*)$ (see (2.12)). It is easy to see that both of these risk measures equal $R(\hat{\theta}, \theta^*)$ in the well-specifed case i.e.,

$$R(\hat{\theta}, \theta^*) = M(\hat{\theta}, \theta^*) = E(\hat{\theta}, \theta^*), \qquad \text{when } \theta^* \in \mathcal{C}.$$

An analogue to inequality (2.1) for the case of misspecification has been proved by Bellec [10, Corollary 2.2], who showed that

$$\frac{1}{\sigma^2} M(\hat{\theta}, \theta^*) \le \frac{1}{\sigma^2} E(\hat{\theta}, \theta^*) \le \delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))). \tag{2.7}$$

Again, although this was originally stated for $Z \sim N(0, I_n)$, it holds for arbitrary zero mean random vectors Z with $\mathbb{E}\|Z\|^2 < \infty$. Note the similarity between the right-hand sides of the inequalities (2.1) and (2.7). The only difference is that the tangent cone at θ^* is replaced by the tangent cone at $\Pi_{\mathcal{C}}(\theta^*)$ in the case of misspecification. Moreover, in the well-specified setting, the above inequality (2.7) reduces to (2.1).

It is now very natural to ask if the second inequality in (2.7) is tight in the low σ limit. One might guess that this should be the case given the result (2.4) for the well-specified setting. However, it turns out that (2.7) is not sharp in the low σ limit. The main contribution of this chapter is to provide an exact formula for the low σ limit of $M(\hat{\theta}, \theta^*)$ and $E(\hat{\theta}, \theta^*)$ when $\mathcal{C}$ is polyhedral. Specifically, in Theorem 2.3.1, we prove that if the noise Z is zero mean with $\mathbb{E}\|Z\|^2 < \infty$ and if $\mathcal{C}$ is polyhedral, then

$$\lim_{\sigma \downarrow 0} \frac{1}{\sigma^2} M(\hat{\theta}, \theta^*) = \lim_{\sigma \downarrow 0} \frac{1}{\sigma^2} E(\hat{\theta}, \theta^*) = \delta\big(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}\big), \tag{2.8}$$

where $v^{\perp} := \{u \in \mathbb{R}^n : \langle u, v \rangle = 0\}$ for vectors $v \in \mathbb{R}^n$. As we remarked earlier, in most applications, the constraint set $\mathcal{C}$ is polyhedral.

Because the set $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}$ is a subset of $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$, the right hand side of (2.8) is never larger than $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$. Under the assumption that the polyhedron $\mathcal{C}$ has a nonempty interior along with a mild condition on the noise Z, it can be proved that the right hand side of (2.8) is *strictly* smaller than $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$ when $\theta^* \notin \mathcal{C}$ (an even stronger statement is proved in Lemma 2.3.4), which then implies that $\lim_{\sigma \downarrow 0} \sigma^{-2} M(\hat{\theta}, \theta^*) < \lim_{\sigma \downarrow 0} \sigma^{-2} R(\hat{\theta}, \Pi_{\mathcal{C}}(\theta^*))$. This inequality is more interpretable in the following form:

$$\lim_{\sigma \downarrow 0} \frac{1}{\sigma^2} \mathbb{E}_{\theta^*} \|\hat{\theta}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2 < \lim_{\sigma \downarrow 0} \frac{1}{\sigma^2} \mathbb{E}_{\Pi_{\mathcal{C}}(\theta^*)} \|\hat{\theta}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2 \qquad \text{whenever } \theta^* \notin \mathcal{C}. \tag{2.9}$$

Inequality (2.9) can be qualitatively understood as follows. The left hand side above corresponds to misspecification where the data are generated from $\theta^* \notin \mathcal{C}$ while the right hand side corresponds to the well-specified setting where the data are generated from $\Pi_{\mathcal{C}}(\theta^*)$. Note that in both cases, the estimator $\hat{\theta}(Y)$ is really estimating $\Pi_{\mathcal{C}}(\theta^*)$ so it is natural to compare the squared expected distance to $\Pi_{\mathcal{C}}(\theta^*)$ in both situations. The interesting aspect is that (in the low σ limit) the expected squared distance is smaller in the misspecified setting compared to the well-specified setting. To the best of our knowledge, this fact has not been noted in the literature previously at this level of generality.

Our main result, Theorem 2.3.1, is stated and proved in Section 2.3 where some intuition is also provided for the exact form of the low σ limit in misspecification. The low σ limit can be explicitly computed in certain specific situations. In Section 2.4, we specialize to the Gaussian model $Z \sim N(0, I_n)$ and study in detail the examples when $\mathcal{C}$ is the nonnegative orthant and when $\mathcal{C}$ is the monotone cone (this latter case corresponds to isotonic regression).

In Section 2.5, we explore issues naturally related to Theorem 2.3.1. In Section 2.5.1, we consider the situation when $\mathcal{C}$ is not polyhedral. It seems hard to characterize the low σ misspecification limits in this case but it is possible to compute them when $\mathcal{C}$ is the unit

ball. It is interesting to note that the low σ limits of $M(\hat{\theta}, \theta^*)$ and $E(\hat{\theta}, \theta^*)$ are different in this case (in sharp contrast to the polyhedral situation). In Section 2.5.2, we deal with the risks when σ is large. Under some conditions, it is possible to write a formula for the large σ limits of $M(\hat{\theta}, \theta^*)$ and $E(\hat{\theta}, \theta^*)$; see Proposition 2.5.3. In Section 2.5.3, we deal with the maximum normalized risks:

$$\sup_{\sigma>0} \frac{1}{\sigma^2} M(\hat{\theta}, \theta^*) \quad \text{and} \quad \sup_{\sigma>0} \frac{1}{\sigma^2} E(\hat{\theta}, \theta^*). \tag{2.10}$$

In the well-specified setting, inequalities (2.1) and (2.4) together imply that the maximum normalized risk equals $\delta(T_{\mathcal{C}}(\theta^*))$. However in the misspecified setting, the quantities (2.10) lie between $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp})$ and $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$. It seems hard to write down an exact formula for the quantities (2.10) but we present some simulation evidence in Section 2.5.3 to argue that they can be strictly between $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp})$ and $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$.

We conclude with an appendix that contains technical lemmas and proofs of the various intermediate results throughout the chapter.

2.2 Background and Notation

In this short section, we shall set up some notation and also recollect standard results in convex analysis that will be used in the remainder of the chapter.

For $x \in \mathbb{R}^n$ and $r > 0$, we denote by $B_r(x) := \{u \in \mathbb{R}^n : \|u - x\| \leq r\}$ the closed ball of radius r centered at x. For $v \in \mathbb{R}^n$, let $v^{\perp} := \{u \in \mathbb{R}^n : \langle u, v\rangle = 0\}$ denote the hyperplane with normal vector v. For $\theta_0 \in \mathcal{C}$, let $F_{\mathcal{C}}(\theta_0) := \{\theta - \theta_0 : \theta \in \mathcal{C}\}$ be the result of re-centering the set $\mathcal{C}$ about θ_0. Also recall the definition of the tangent cone (2.2) and note that $T_{\mathcal{C}}(\theta_0) = \mathrm{cl}\{\alpha x : x \in F_{\mathcal{C}}(\theta_0), \alpha > 0\}$.

If A is an $m \times n$ matrix and $J \subseteq \{1, \ldots, m\}$, we let a_j denote the jth row of A, and let A_J denote the matrix obtained by combining the rows of A indexed by J.

A *polyhedron* refers to a set of the form $\{x \in \mathbb{R}^n : Ax \leq b\}$ for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$ where the inequality $\leq$ is interpreted coordinate-wise, i.e. $\langle a_j, x\rangle \leq b_j$ for $j = 1, \ldots, m$. We will assume that no two pairs (a_j, b_j) and (a_k, b_k) are scalar multiples of each other. A *polyhedral cone* is a set of the form $\{x \in \mathbb{R}^n : Ax \leq 0\}$ for some $A \in \mathbb{R}^{m \times n}$. Again, we will assume that no two rows of A are scalar multiples of each other. A *face* of a polyhedron refers to any subset obtained by setting some of the polyhedron's linear inequality constraints to equality instead.

In the remainder of this section, we shall collect some standard results above convex projections that will be used in the chapter. These results can be found in a standard reference such as [49]. Recall that $\Pi_{\mathcal{C}}(x)$ denotes the projection of a vector $x \in \mathbb{R}^n$ on a closed convex set $\mathcal{C}$. It is well known that $\Pi_{\mathcal{C}}(x)$ is the unique vector in $\mathcal{C}$ satisfying the optimality condition

$$\langle z - \Pi_{\mathcal{C}}(x), x - \Pi_{\mathcal{C}}(x)\rangle \leq 0, \qquad \forall z \in \mathcal{C}. \tag{2.11}$$

Consequently, we have the following Pythagorean inequality

$$\|z-x\|^2 = \|z-\Pi_{\mathcal{C}}(x)\|^2+\|\Pi_{\mathcal{C}}(x)-x\|^2+2\langle z-\Pi_{\mathcal{C}}(x), \Pi_{\mathcal{C}}(x)-x\rangle \geq \|z-\Pi_{\mathcal{C}}(x)\|^2+\|\Pi_{\mathcal{C}}(x)-x\|^2.$$

Plugging in $z = \Pi_{\mathcal{C}}(y)$ and $x = \theta^*$ shows that the misspecified error is upper bounded by the excess error, that is,

$$\|\Pi_{\mathcal{C}}(y) - \Pi_{\mathcal{C}}(\theta^*)\|^2 \leq \|\Pi_{\mathcal{C}}(y) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2, \qquad \forall y \in \mathbb{R}^n. \tag{2.12}$$

If instead we plug in $z = \Pi_{\mathcal{C}}(\theta^*)$ to (2.11), we have $\langle \Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(\theta^*), x - \Pi_{\mathcal{C}}(x)\rangle \geq 0$, which implies

$$\begin{aligned}\|\Pi_{\mathcal{C}}(x) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2 &= -\|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(\theta^*)\|^2 + 2\langle \Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(\theta^*), \Pi_{\mathcal{C}}(x) - \theta^*\rangle \\ &\leq -\|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(\theta^*)\|^2 + 2\langle \Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(\theta^*), x - \theta^*\rangle \\ &\leq \|x - \theta^*\|^2.\end{aligned}$$

Combining this with (2.12), we see that for $Y = \theta^* + \sigma Z$ we have

$$0 \leq \|\Pi_{\mathcal{C}}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2 \leq \|\Pi_{\mathcal{C}}(Y) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2 \leq \sigma^2\|Z\|^2. \tag{2.13}$$

In the special case where $\mathcal{C}$ is a cone, the optimality condition (2.11) implies that $\Pi_{\mathcal{C}}(x)$ is the unique vector in $\mathcal{C}$ satisfying

$$\langle \Pi_{\mathcal{C}}(x), x - \Pi_{\mathcal{C}}(x)\rangle = 0, \quad \text{and} \quad \langle z, x - \Pi_{\mathcal{C}}(x)\rangle \leq 0, \quad \forall z \in \mathcal{C}. \tag{2.14}$$

2.3 Main theorem: low noise limit for polyhedra

Our main result below provides a precise characterization of the low σ limits of the risks (2.5) and (2.6) (normalized by σ^2) in the misspecified setting (i.e., when $\theta^* \notin \mathcal{C}$) for polyhedral $\mathcal{C}$. An implication of this result is that the low σ limit can be much smaller than the upper bound (2.7) of Bellec [10].

Theorem 2.3.1 (Low noise limit of risk for polyhedra)**.** *Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a closed convex set, and let $Y = \theta^* + \sigma Z$ where $\theta^* \in \mathbb{R}^n$ is not necessarily in $\mathcal{C}$, and Z is zero mean with $\mathbb{E}\|Z\|^2 < \infty$. Suppose the following "locally polyhedral" condition holds.*

$$\begin{aligned}&T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \text{ is a polyhedral cone, and} \\ &T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap B_{r^*}(0) = F_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap B_{r^*}(0) \text{ for some } r^* > 0.\end{aligned} \tag{2.15}$$

Then,

$$\lim_{\sigma\downarrow 0} \frac{1}{\sigma^2} M(\hat{\theta}, \theta^*) = \lim_{\sigma\downarrow 0} \frac{1}{\sigma^2} E(\hat{\theta}, \theta^*) = \delta\big(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}\big). \tag{2.16}$$

Note again that $\delta(\cdot)$ denotes the generalized statistical dimension induced by the noise Z, and reduces to the usual statistical dimension [3] when $Z \sim N(0, I_n)$.

We remark that the "locally polyhedral" condition (2.15) essentially states that $\mathcal{C}$ looks like a polyhedron in a neighborhood around $\Pi_{\mathcal{C}}(\theta^*)$. As established in the following lemma, it automatically holds if $\mathcal{C}$ is a polyhedron, so one can replace any mention of condition (2.15) with "$\mathcal{C}$ is a polyhedron" for the sake of readability. We provide some remarks on the case when $\mathcal{C}$ is not polyhedral in Section 2.5.1.

Lemma 2.3.2. *Let $\mathcal{C}$ be a polyhedron. Then the locally polyhedral condition* (2.15) *holds for any $\theta^* \in \mathbb{R}^n$.*

Next, the following lemma establishes that the set $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ that appears in the limit (2.16) is a face of the tangent cone $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$.

Lemma 2.3.3. *Let $\theta^* \in \mathbb{R}^n$ and let $\mathcal{C} \subseteq \mathbb{R}^n$ be a closed convex set satisfying the locally polyhedral condition* (2.15). *Let $A \in \mathbb{R}^{m \times n}$ be such that $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) = \{u : Au \leq 0\}$. Then there exists some subset $J \subseteq \{1, \ldots, m\}$ such that*

$$T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp = \{u : A_J u = 0, A_{J^c} u \leq 0\}.$$

Thus, $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ is a face of $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$.

Both the above lemmas are proved in Section 2.6.

If $\theta^* \in \mathcal{C}$ then we have $\Pi_{\mathcal{C}}(\theta^*) = \theta^*$, and Theorem 2.3.1 reduces to the result (2.4) of Oymak and Hassibi [70]: the excess risk and the misspecified risk become the same, and the common limit is the statistical dimension of $T_{\mathcal{C}}(\theta^*)$. We must remark here that the result of Oymak and Hassibi [70] holds for non-polyhedral $\mathcal{C}$ as well. We discuss the non-polyhedral setting further in Section 2.5.1.

Theorem 2.3.1 states that in the misspecified case $\theta^* \notin \mathcal{C}$, the low sigma limit still involves the tangent cone $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$, but one needs to intersect it with the hyperplane $(\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ before taking the statistical dimension. Due to the optimality condition (2.11) characterizing $\Pi_{\mathcal{C}}$, the tangent cone lies entirely on one side of the hyperplane, so the hyperplane does not intersect the interior of the tangent cone. Therefore, the interior of the tangent cone $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$ does not contribute to the low σ limit of the risk under misspecification. This makes sense because when $\theta^* \notin \mathcal{C}$ and σ is small, the observation vector Y is outside $\mathcal{C}$ with high probability so that $\hat{\theta}(Y)$ lies on the boundary of $\mathcal{C}$.

In general, the intersection $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ can be anything from $\{0\}$ to the full tangent cone $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$ and so the low sigma limit can be anything between 0 and $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$. The case when the limit equals zero corresponds to the situation where θ^* lies in the interior of the preimage of $\Pi_{\mathcal{C}}(\theta^*)$ under the map $\Pi_{\mathcal{C}}$ so that every point in some neighborhood of θ^* is projected onto the same point $\Pi_{\mathcal{C}}(\theta^*)$ (see Figure 2.1c for an example).

The following lemma (proved in Section 2.6), provides mild conditions under which the intersection $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ has strictly smaller generalized statistical dimension than the full tangent cone $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$.

Lemma 2.3.4. *Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a polyhedron with nonempty interior. Then*

$$\sup_{\theta^* \notin \mathcal{C}: \Pi_{\mathcal{C}}(\theta^*) = \theta_0} \delta\big(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}\big) < \delta(T_{\mathcal{C}}(\theta_0)).$$

for every $\theta_0 \in \mathcal{C}$, provided the random vector Z has nonzero probability of lying in the interior of $T_{\mathcal{C}}(\theta_0)$.

As mentioned already, Lemma 2.3.4 combined with the main result Theorem 2.3.1 implies the risk gap (2.9). In summary, under the nonempty interior assumption, if we think of the low σ limit as a function of θ^*, we see that as θ^* approaches $\mathcal{C}$ from the outside there is a "jump" when θ^* enters $\mathcal{C}$. This "jump" phenomenon is not unique to the polyhedral case. In Section 2.5.1 we discuss a non-polyhedral example that also exhibits this jump phenomenon.

Theorem 2.3.1 suggests something that may seem nonintuitive: if $\theta^* \notin \mathcal{C}$ and we use the estimator $\hat{\theta}(Y) = \Pi_{\mathcal{C}}(Y)$, the risk when $Y = \theta^* + \sigma Z$ is smaller than the risk when $Y = \Pi_{\mathcal{C}}(\theta^*) + \sigma Z$. As mentioned already, in the case $Y = \theta^* + \sigma Z$ the estimator is actually estimating $\Pi_{\mathcal{C}}(\theta^*)$, not θ^*. Moreover, the risks (2.5) and (2.6) measure error relative to $\Pi_{\mathcal{C}}(\theta^*)$ rather than to θ^*. Furthermore, the intuition is that in the low σ limit, the estimator $\hat{\theta}(Y)$ in the misspecified setting is a projection onto a much smaller set than in the well-specified setting (essentially, a face of a tangent cone instead of the full tangent cone), so more of the original noise in Y is eliminated. This qualitatively explains why having Y generated from θ^* outside $\mathcal{C}$ allows the estimator to estimate $\Pi_{\mathcal{C}}(\theta^*)$ better than if Y were generated from $\Pi_{\mathcal{C}}(\theta^*)$ instead.

Finally, we observe that in the misspecified setting, there is a gap between Bellec's upper bound $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$ (2.7) and the low σ risk limit, unlike in the well-specified setting where the result (2.4) implies that the normalized risk increases to the upper bound in the low σ limit. The upper bound, which is constant in σ, can become very loose as $\sigma \downarrow 0$. However, in Section 2.5.3 we shown a few examples where the normalized risk is close to the upper bound for some σ, as well as examples where the normalized risk remains much smaller than the upper bound for all $\sigma > 0$.

2.3.1 Proof of Theorem 2.3.1

We establish one key lemma (proved in Section 2.6) before proving Theorem 2.3.1. It is a deterministic result that contains the core of the argument: roughly, if we have a polyhedral cone $\mathcal{T}$ and any $\theta^* \in \mathbb{R}^n$ satisfying $\Pi_{\mathcal{T}}(\theta^*) = 0$, then any point u sufficiently near θ^* will have its projection $\Pi_{\mathcal{T}}(u)$ lying in the hyperplane with normal direction θ^*.

Lemma 2.3.5 (Key lemma)**.** *Fix $\theta^* \in \mathbb{R}^n$, and let $\mathcal{T}$ be a closed convex set such that the re-centered set $\{\theta - \Pi_{\mathcal{T}}(\theta^*) : \theta \in \mathcal{T}\}$ is a polyhedral cone. Then there exists $r > 0$ such that*

$$\Pi_{\mathcal{T}}(u) - \Pi_{\mathcal{T}}(\theta^*) \in (\theta^* - \Pi_{\mathcal{T}}(\theta^*))^{\perp}, \qquad \forall u \in B_r(\theta^*). \tag{2.17}$$

With this lemma, along with some standard results collected in Section 2.2, we can proceed with proving Theorem 2.3.1.

Proof of Theorem 2.3.1. We first prove

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}M(\hat{\theta},\theta^*)=\delta\big(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))\cap(\theta^*-\Pi_{\mathcal{C}}(\theta^*))^{\perp}\big). \tag{2.18}$$

For any $r>0$ we can write

$$\frac{1}{\sigma^2}M(\hat{\theta},\theta^*)=\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\|\Pi_{\mathcal{C}}(Y)-\Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_{\{Y\in B_r(\theta^*)\}}\big]+\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\|\Pi_{\mathcal{C}}(Y)-\Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_{\{Y\notin B_r(\theta^*)\}}\big]. \tag{2.19}$$

We claim the second term on the right-hand side vanishes as $\sigma\downarrow 0$ (regardless of the value of $r>0$). Since the projection $\Pi_{\mathcal{C}}$ is non-expansive [49],

$$0\le\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\|\Pi_{\mathcal{C}}(Y)-\Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_{\{Y\notin B_r(\theta^*)\}}\big]\le\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\|Y-\theta^*\|^2\mathbf{1}_{\{Y\notin B_r(\theta^*)\}}\big]=\mathbb{E}_{\theta^*}\big[\|Z\|^2\mathbf{1}_{\{\sigma\|Z\|>r\}}\big].$$

Then, the dominated convergence theorem implies the right-hand side tends to zero as $\sigma\downarrow 0$, because $\mathbb{E}\|Z\|^2<\infty$ and the random variable $\|Z\|^2\mathbf{1}_{\{\sigma\|Z\|>r\}}$ converges to zero pointwise.

Thus, it remains to show

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\|\Pi_{\mathcal{C}}(Y)-\Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_{\{Y\in B_r(\theta^*)\}}\big]=\delta\big(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))\cap(\theta^*-\Pi_{\mathcal{C}}(\theta^*))^{\perp}\big) \tag{2.20}$$

for some $r>0$.

We define the re-centered tangent cone

$$\mathcal{T}\coloneqq\{\Pi_{\mathcal{C}}(\theta^*)+u:u\in T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))\}.$$

We claim there exists some $r>0$ such that

$$\Pi_{\mathcal{C}}(u)=\Pi_{\mathcal{T}}(u),\qquad\forall u\in B_r(\theta^*). \tag{2.21}$$

Indeed, note that the locally polyhedral condition (2.15) implies the existence of some $r^*>0$ such that

$$\mathcal{C}\cap B_{r^*}(\Pi_{\mathcal{C}}(\theta^*))=\mathcal{T}\cap B_{r^*}(\Pi_{\mathcal{C}}(\theta^*)) \tag{2.22}$$

Since both projections $\Pi_{\mathcal{C}}$ and $\Pi_{\mathcal{T}}$ are continuous [49] at θ^*, there exists some $r>0$ such that the image of $B_r(\theta^*)$ under both projections lies in $B_{r^*}(\Pi_{\mathcal{C}}(\theta^*))$. Thus the local equality (2.21) of the projections follows from the locally polyhedral condition (2.22).

By combining this argument with Lemma 2.3.5, we have shown there exists some $r>0$ that satisfies not only (2.21), but also (2.17). With this value of r, the equality (2.21) implies that replacing each instance of $\mathcal{C}$ with $\mathcal{T}$ in (2.20) does not change either side, since $\Pi_{\mathcal{C}}(Y)=\Pi_{\mathcal{T}}(Y)$, $\Pi_{\mathcal{C}}(\theta^*)=\Pi_{\mathcal{T}}(\theta^*)$, and

$$T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))=T_{\mathcal{C}\cap B_{r^*}(\Pi_{\mathcal{C}}(\theta^*))}(\Pi_{\mathcal{C}}(\theta^*))=T_{\mathcal{T}\cap B_{r^*}(\Pi_{\mathcal{C}}(\theta^*))}(\Pi_{\mathcal{C}}(\theta^*))=T_{\mathcal{T}}(\Pi_{\mathcal{T}}(\theta^*)),$$

by the equality (2.22) and the definition of the tangent cone. Thus it remains to prove

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\|\Pi_{\mathcal{T}}(Y)-\Pi_{\mathcal{T}}(\theta^*)\|^2\mathbf{1}_{\{Y\in B_r(\theta^*)\}}\big]=\delta(\mathcal{K}), \tag{2.23}$$

where $\mathcal{K}\coloneqq T_{\mathcal{T}}(\theta^*)\cap(\theta^*-\Pi_{\mathcal{T}}(\theta^*))^{\perp}$.

Since r satisfies (2.17), some re-centering yields

$$\Pi_{\mathcal{T}}(Y)-\Pi_{\mathcal{T}}(\theta^*)=\Pi_{T_{\mathcal{T}}(\theta^*)}(Y-\Pi_{\mathcal{T}}(\theta^*))=\Pi_{\mathcal{K}}(Y-\Pi_{\mathcal{T}}(\theta^*)) \tag{2.24}$$

in the event $\{Y\in B_r(\theta^*)\}$.

For $W\coloneqq(\theta^*-\Pi_{\mathcal{T}}(\theta^*))^{\perp}$, we claim

$$\Pi_{\mathcal{K}}=\Pi_{\mathcal{K}}\circ\Pi_W.$$

In fact this holds for any subspace W and closed convex $\mathcal{K}\subseteq W$, by the Pythagorean theorem:

$$\Pi_{\mathcal{K}}(x)=\operatorname*{argmin}_{u\in\mathcal{K}}\|x-u\|^2=\operatorname*{argmin}_{u\in\mathcal{K}}\big\{\|x-\Pi_W(x)\|^2+\|\Pi_W(x)-u\|^2\big\}=\Pi_{\mathcal{K}}(\Pi_W(x)).$$

Applying this to (2.24) yields

$$\begin{aligned}\Pi_{\mathcal{T}}(Y)-\Pi_{\mathcal{T}}(\theta^*)&=\Pi_{\mathcal{K}}(Y-\Pi_{\mathcal{T}}(\theta^*))\\&=\Pi_{\mathcal{K}}(\Pi_W(\theta^*+\sigma Z-\Pi_{\mathcal{T}}(\theta^*)))\\&=\Pi_{\mathcal{K}}(\Pi_W(\sigma Z)) && \Pi_W \text{ is linear, } \Pi_W(\theta^*-\Pi_{\mathcal{T}}(\theta^*))=0\\&=\Pi_{\mathcal{K}}(\sigma Z)=\sigma\Pi_{\mathcal{K}}(Z) && \mathcal{K} \text{ is a cone}\end{aligned}$$

in the event $\{Y\in B_r(\theta^*)\}$. By plugging this into the left-hand side of equation (2.23), we have

$$\lim_{\sigma\downarrow 0}\mathbb{E}_{\theta^*}\big[\|\Pi_{\mathcal{K}}(Z)\|^2\mathbf{1}_{\{Y\in B_r(\theta^*)\}}\big]=\mathbb{E}\|\Pi_{\mathcal{K}}(Z)\|^2=\delta(\mathcal{K}),$$

where the first equality follows by dominated convergence ($\|\Pi_{\mathcal{K}}(Z)\|^2\le\|Z\|^2$ and $\mathbb{E}\|Z\|^2<\infty$). This verifies the desired equality (2.23) and concludes the proof of the first low σ limit (2.18).

We now prove the other equality

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}M(\hat{\theta},\theta^*)=\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}E(\hat{\theta},\theta^*).$$

We claim

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\big(\|\Pi_{\mathcal{C}}(Y)-\theta^*\|^2-\|\Pi_{\mathcal{C}}(\theta^*)-\theta^*\|^2\big)\mathbf{1}_{\{Y\notin B_r(\theta^*)\}}\big]=0 \tag{2.25}$$

for any $r>0$. Applying some basic properties (2.13) of the projection $\Pi_{\mathcal{C}}$ yields

$$0\le\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\big[\big(\|\Pi_{\mathcal{C}}(Y)-\theta^*\|^2-\|\Pi_{\mathcal{C}}(\theta^*)-\theta^*\|^2\big)\mathbf{1}_{\{Y\notin B_r(\theta^*)\}}\big]\le\mathbb{E}\big[\|Z\|^2\mathbf{1}_{\{\sigma\|Z\|\ge r\}}\big],$$

so applying the dominated convergence theorem as before leads to the limit (2.25).

Thus, it suffices to prove

$$\begin{aligned}&\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\left[\|\Pi_{\mathcal{C}}(Y)-\Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_{\{Y\in B_r(\theta^*)\}}\right]\\&=\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\left[\left(\|\Pi_{\mathcal{C}}(Y)-\theta^*\|^2-\|\Pi_{\mathcal{C}}(\theta^*)-\theta^*\|^2\right)\mathbf{1}_{\{Y\in B_r(\theta^*)\}}\right]\end{aligned}\tag{2.26}$$

for some $r>0$. We choose r as before so that (2.17) and (2.21) both hold. By the same reasoning as before, we can replace each instance of $\mathcal{C}$ with $\mathcal{T}$ without changing anything. Furthermore, the condition (2.17) implies we have $\langle\Pi_{\mathcal{T}}(Y)-\Pi_{\mathcal{T}}(\theta^*),\theta^*-\Pi_{\mathcal{T}}(\theta^*)\rangle=0$ in the event $\{Y\in B_r(\theta^*)\}$, so the Pythagorean inequality (2.12) becomes equality:

$$\|\Pi_{\mathcal{T}}(Y)-\Pi_{\mathcal{T}}(\theta^*)\|^2\mathbf{1}_{\{Y\in B_r(\theta^*)\}}=\left(\|\Pi_{\mathcal{T}}(Y)-\theta^*\|^2-\|\Pi_{\mathcal{T}}(\theta^*)-\theta^*\|^2\right)\mathbf{1}_{\{Y\in B_r(\theta^*)\}}.$$

Therefore the equality (2.26) holds, which concludes the proof of Theorem 2.3.1. □

2.4 Examples

In this section, we assume the Gaussian noise model $Z\sim N(0,I_n)$, or equivalently $Y\sim N(\theta^*,\sigma^2I_n)$. Thus, $\delta(\cdot)$ denotes the usual statistical dimension [3], where Z in the definition (2.3) is a standard Gaussian vector.

2.4.1 Nonnegative orthant

We now apply Theorem 2.3.1 to the *nonnegative orthant* $\mathbb{R}^n_+\coloneqq\{u\in\mathbb{R}^n:u_i\geq 0,\forall i\}$. In Figure 2.1 we provide visualizations of the geometry of the main theorem when applied to this constraint set.

Corollary 2.4.1 (Nonnegative orthant)**.** *Let $Y\sim N(\theta^*,\sigma^2I)$ where $\theta^*\in\mathbb{R}^n$. Let $n_+\coloneqq\sum_{i=1}^n\mathbf{1}_{\{\theta_i^*>0\}}$ and $n_0\coloneqq\sum_{i=1}^n\mathbf{1}_{\{\theta_i^*=0\}}$ denote the number of positive components and number of zero components of θ^* respectively. Then the normalized excess risk* (2.6) *and normalized mispecified risk* (2.5) *of the least squares estimator $\hat{\theta}(Y)\coloneqq\Pi_{\mathbb{R}^n_+}(Y)$ with respect to $\mathbb{R}^n_+$ both tend to*

$$\frac{n_0}{2}+n_+$$

as $\sigma\downarrow 0$.

Proof. By Theorem 2.3.1, it suffices to prove that the statistical dimension term in (2.16) is $\frac{n_0}{2}+n_+$. Note that for $y\in\mathbb{R}^n$, $\Pi_{\mathbb{R}^n_+}(y)=\max\{y,0\}$ is obtained by taking the component-wise maximum of y with 0. Consequently,

$$T_{\mathbb{R}^n_+}(\Pi_{\mathbb{R}^n_+}(\theta^*))=\{u\in\mathbb{R}^n:u_i\geq 0\text{ if }(\Pi_{\mathbb{R}^n_+}(\theta^*))_i=0\}=\{u\in\mathbb{R}^n:u_i\geq 0\text{ if }\theta_i^*\leq 0\}.$$

Also,

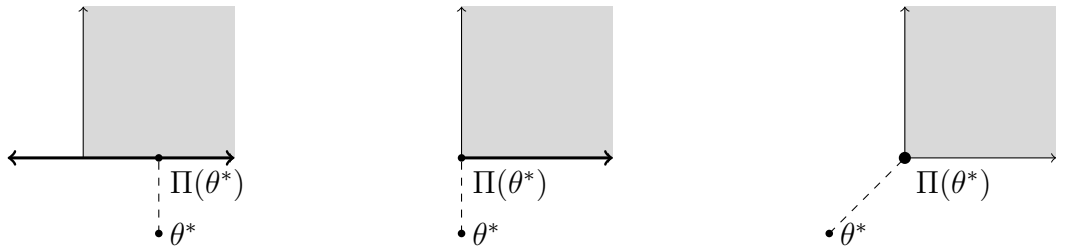
$$(\theta^* - \Pi_{\mathbb{R}^n_+}(\theta^*))^\perp = \left\{ u \in \mathbb{R}^n : \sum_{i:\theta_i^*<0} \theta_i^* u_i = 0 \right\}$$

The intersection is thus

$$T_{\mathbb{R}^n_+}(\Pi_{\mathbb{R}^n_+}(\theta^*)) \cap (\theta^* - \Pi_{\mathbb{R}^n_+}(\theta^*))^\perp = \left\{ u \in \mathbb{R}^n : \begin{array}{ll} u_i \geq 0 & \text{if } \theta_i^* = 0 \\ u_i = 0 & \text{if } \theta_i^* < 0 \end{array} \right\} \cong \mathbb{R}^{n_+} \times \mathbb{R}^{n_0}_+ \times \{0\}^{n-n_+-n_0}.$$

The result follows by noting $\delta(\mathbb{R}) = 1$ and $\delta(\mathbb{R}_+) = 1/2$ and by using the fact that $\delta(T_1 \times T_2) = \delta(T_1) + \delta(T_2)$ for any two cones T_1 and T_2 [3]. □

Remark 2.4.2. *For $\theta^* \in \mathbb{R}^n$ let n_+ and n_0 be as defined in Corollary 2.4.1. Then the low σ limit for the corresponding well-specified problem $Y \sim N(\Pi_{\mathbb{R}^n_+}(\theta^*), \sigma^2 I)$ is $\frac{n-n_+}{2} + n_+$ since all negative components of θ^* are sent to zero by $\Pi_{\mathbb{R}^n_+}$. This is larger than the low σ limit for the misspecified problem $Y \sim N(\theta^*, \sigma^2 I)$ because $n - n_+ \geq n_0$, with strict inequality if $\theta^* \notin \mathbb{R}^n_+$.*



(a) $\theta = (1, -1)$; $\delta = 1$ (b) $\theta = (0, -1)$; $\delta = 1/2$ (c) $\theta = (-1, -1)$; $\delta = 0$

Figure 2.1: $\mathbb{R}^2_+$ is marked by the gray area. The intersection $T_{\mathbb{R}^2_+}(\Pi_{\mathbb{R}^2_+}(\theta^*)) \cap (\theta^* - \Pi_{\mathbb{R}^2_+}(\theta^*))^\perp$ [translated to be centered at $\Pi_{\mathbb{R}^2_+}(\theta^*)$] is marked by the bold lines in the first two examples, and the bold point in the third example. Each sub-caption states the statistical dimension $\delta = \delta(T_{\mathbb{R}^2_+}(\Pi_{\mathbb{R}^2_+}(\theta^*)) \cap (\theta^* - \Pi_{\mathbb{R}^2_+}(\theta^*))^\perp)$.

2.4.2 Consequences for isotonic regression

This section details interesting consequences of Theorem 2.3.1 for isotonic regression under misspecification. Let

$$\mathcal{S}^n := \{u \in \mathbb{R}^n : u_1 \leq \cdots \leq u_n\}$$

be the *monotone cone*. We call elements of $\mathcal{S}^n$ *nondecreasing*.

By a *block*, we refer to a set of the form $\{k, k+1, \ldots, l\}$ for two nonnegative integers $k \le l$. Consider a partition of $\{1, \ldots, n\}$ into blocks $I_1, \ldots, I_m$ listed in increasing order (i.e., the maximum entry of I_i is strictly smaller than the minimum entry of I_j for $i < j$). Let $|I_j|$ denote the cardinality of I_j and note that $\sum_{j=1}^m |I_j| = n$ as $I_1, \ldots, I_m$ form a partition of $\{1, \ldots, n\}$. Let $\mathcal{S}_{|I_1|,\ldots,|I_m|}$ denote the induced *block monotone cone* defined as

$$\mathcal{S}_{|I_1|,\ldots,|I_m|} \coloneqq \{u \in \mathcal{S}^n : u \text{ is constant on each of the blocks } I_1, \ldots, I_m\} \tag{2.27}$$

For example,

$$\mathcal{S}_{2,3,2} = \{u \in \mathbb{R}^{2+3+2} : u_1 = u_2 \le u_3 = u_4 = u_5 \le u_6 = u_7\}.$$

Theorem 2.3.1 implies the following result, which we prove in Section 2.7.3.

Proposition 2.4.3 (Isotonic regression)**.** *Let $Y \sim N(\theta^*, \sigma^2 I)$ where $\theta^* \in \mathbb{R}^n$. Let $(J_1, \ldots, J_K)$ be the partition of $\{1, \ldots, n\}$ into blocks such that $\Pi_{\mathcal{S}^n}(\theta^*)$ is constant on each J_k with respective values $\mu_1 < \cdots < \mu_K$. For each $k \in \{1, \ldots, K\}$, there exists a unique finest partition $(I_1^k, \ldots, I_{m_k}^k)$ of J_k into blocks such that for all $j \in \{1, \ldots, m_k\}$, the mean of the components of θ^* on each I_j^k equals μ_k; that is,*

$$\frac{1}{|I_j^k|} \sum_{i \in I_j^k} \theta_i^* = \mu_k, \qquad 1 \le j \le m_k. \tag{2.28}$$

Then the common low σ limit of the normalized excess risk (2.6) *and normalized misspecified risk* (2.5) *of the isotonic least squares estimator $\hat{\theta}(Y) \coloneqq \Pi_{\mathcal{S}^n}(Y)$ equals*

$$\sum_{k=1}^K \delta\Big(\mathcal{S}_{|I_1^k|,\ldots,|I_{m_k}^k|}\Big). \tag{2.29}$$

It is clear from the above proposition that the low σ behavior of the isotonic estimator under misspecification crucially depends on the statistical dimension of the block monotone cone $\mathcal{S}_{|I_1^k|,\ldots,|I_{m_k}^k|}$. [We remark again that throughout this section we only deal with the usual statistical dimension, where the noise Z in the definition (2.3) is standard Gaussian.] Here, we provide two simple properties of the block monotone cone (2.27), each of which implies that when the block sizes are equal, the statistical dimension is simply that of $\mathcal{S}^{m_k}$. The first result provides a direct connection to weighted isotonic regression.

Lemma 2.4.4 (Weighted isotonic regression)**.** *Let $z \in \mathbb{R}^n$ and let $I_1, \ldots, I_m$ be a partition of $\{1, \ldots, n\}$ into blocks. Let $\bar{z}_{I_j} \coloneqq \frac{1}{|I_j|} \sum_{i \in I_j} z_i$. Then $\Pi_{\mathcal{S}_{|I_1|,\ldots,|I_m|}}(y)$ is the vector that is constant on the blocks $I_1, \ldots, I_m$ with constant values $x_1^*, \ldots, x_m^*$, where $x^* = (x_1^*, \ldots, x_m^*)$ is*

$$x^* = \operatorname*{argmin}_{x \in \mathcal{S}^m} \sum_{j=1}^m |I_j| (x_i - \bar{z}_{I_j})^2.$$

In other words, the values on the constant blocks of $\Pi_{\mathcal{S}_{|I_1|,\ldots,|I_m|}}(z)$ *can be found by weighted isotonic regression of* $(\bar{z}_{I_1}, \ldots, \bar{z}_{I_m}) \in \mathbb{R}^m$ *with weights* $|I_1|, \ldots, |I_m|$.

Consequently, when $|I_1| = \cdots = |I_m|$, *the statistical dimension of the block monotone cone is*

$$\delta(\mathcal{S}_{|I_1|,\ldots,|I_m|}) = \sum_{j=1}^{m} \frac{1}{j}.$$

The next lemma shows $\mathcal{S}_{|I_1|,\ldots,|I_m|}$ is isometric to a particular cone in the lower-dimensional space $\mathbb{R}^m$.

Lemma 2.4.5 (Block monotone cone isometry)**.** *The block monotone cone* $\mathcal{S}_{|I_1|,\ldots,|I_m|} \subseteq \mathbb{R}^n$ *is isometric to*

$$\left\{ v \in \mathbb{R}^m : \frac{v_1}{\sqrt{|I_1|}} \leq \cdots \leq \frac{v_m}{\sqrt{|I_m|}} \right\} \subseteq \mathbb{R}^m, \tag{2.30}$$

and thus both sets have the same statistical dimension. In particular, if $|I_1| = \cdots = |I_m|$, *then the statistical dimension of the block monotone cone is*

$$\delta(\mathcal{S}_{|I_1|,\ldots,|I_m|}) = \sum_{j=1}^{m} \frac{1}{j}.$$

Both lemmas are proved in Section 2.7.1. Note that for the case $|I_1| = \cdots = |I_m| = 1$, both lemmas reduce to the statement of the statistical dimension of the monotone cone $\mathcal{S}^n$ [3, Eq. D.12]. More generally, when the m blocks have equal size, the statistical dimension of the associated block monotone cone is the same as that of the monotone cone $\mathcal{S}^m$. In Section 2.7.2, we discuss what Lemma 2.4.5 suggests for the completely general case when the block sizes are arbitrary.

By combining either of these two lemmas with Proposition 2.4.3, we immediately obtain an explicit expression for the low σ limits in a special case. For $m \geq 1$, we denote the harmonic number $\sum_{j=1}^{m}(1/j)$ by H_m.

Corollary 2.4.6 (Isotonic regression with equal sub-block sizes)**.** *Consider the setting of Proposition 2.4.3. In the special case where*

$$|I_1^k| = \cdots = |I_{m_k}^k| \qquad \text{for each } k \in \{1, \ldots, K\}, \tag{2.31}$$

the common low σ *limit has the following explicit expression:*

$$\sum_{k=1}^{K} H_{m_k} = \sum_{k=1}^{K} \sum_{j=1}^{m_k} \frac{1}{j}.$$

See the examples to follow (as well as Section 2.7.2) for further discussion about how the statistical dimension of $\mathcal{S}_{|I_1^k|,\ldots,|I_{m_k}^k|}$ behaves in general, when the special condition (2.31) does not hold.

In Table 2.1, we demonstrate how to apply this theorem to various cases of θ^*. In the "partition of θ^*" column, we use square brackets to partition the components of θ^* into K blocks according to the constant pieces $\mu_1 < \cdots < \mu_K$ of $\Pi_{\mathcal{S}^n}(\theta^*)$, and then within the kth group use parentheses to further partition the components into m_k sub-blocks each with common mean μ_k.

θ^*	$\Pi_{\mathcal{S}^n}(\theta^*)$	partition of θ^*	$m_1, \ldots, m_K$	$\sum_{k=1}^K H_{m_k}$
$(0,0,0,0,0,0)$	$(0,0,0,0,0,0)$	$[(0),(0),(0),(0),(0),(0)]$	6	$H_6 = 2.45$
$(1,-1,1,-1,1,-1)$	$(0,0,0,0,0,0)$	$[(1,-1),(1,-1),(1,-1)]$	3	$H_3 = 1.8\bar{3}$
$(5,3,1,-1,-3,-5)$	$(0,0,0,0,0,0)$	$[(5,3,1,-1,-3,-5)]$	1	$H_1 = 1$
$(-1,-1,-1,-1,2,2)$	$(-1,-1,-1,-1,2,2)$	$[(-1),(-1),(-1),(-1)],[(2),(2)]$	$4,2$	$H_4 + H_2 = 3.58\bar{3}$
$(0,-2,1,-3,2,2)$	$(-1,-1,-1,-1,2,2)$	$[(0,-2),(1,-3)],[(2),(2)]$	$2,2$	$H_2 + H_2 = 3$
$(0,0,-2,-2,3,1)$	$(-1,-1,-1,-1,2,2)$	$[(0,0,-2,-2)],[(3,1)]$	$1,1$	$H_1 + H_1 = 2$

Table 2.1: Examples of how to compute the limit in Proposition 2.4.3 in the special case (2.31).

We now discuss in detail what Proposition 2.4.3 states for certain cases of θ^*.

1. In the well-specified case where $\theta^* \in \mathcal{S}^n$, we have $\theta^*_j = \mu_k$ for all $j \in J_k$ and $k \in \{1, \ldots, K\}$, so the finest partition of each J_k is the partition into singleton sets. Then $m_k = |J_k|$ for each k, and moreover $|I^k_j| = 1$ for all valid k and j. Thus, Proposition 2.4.3 implies that both low σ limits are

$$\sum_{k=1}^K H_{|J_k|} \coloneqq \sum_{k=1}^K \sum_{j=1}^{|J_k|} \frac{1}{j},$$

This is precisely the upper bound (2.7) for the monotone cone as computed by Bellec [10, Prop. 3.1], so we recover the low σ limit (2.4). Computations for the well-specified examples $\theta^* = (0,0,0,0,0,0)$ and $\theta^* = (-1,-1,-1,-1,2,2)$ appear in Table 2.1.

Now, consider the misspecified problem $Y \sim N(\theta^*, \sigma^2 I_n)$ with $\theta^* \notin \mathcal{S}^n$, and compare the statement of Proposition 2.4.3 with the corresponding statemetn for the well-specified problem $Y \sim N(\Pi_{\mathcal{S}^n}(\theta^*), \sigma^2 I)$. In both cases, the partition of $\{1, \ldots, n\}$ into $(J_1, \ldots, J_K)$ is the same. However, we showed above that in the well-specified problem, the sub-partition of each J_k consists of singletons, whereas for the misspecified problem we may get nontrivial partitions $(I^k_1, \ldots, I^k_{m_k})$. Noting the inclusion $\mathcal{S}_{|I^k_1|,\ldots,|I^k_{m_k}|} \subseteq \mathcal{S}^{|J_k|}$ for each k and comparing (2.29) for the two cases yields

$$\sum_{k=1}^K \delta(\mathcal{S}_{|I^k_1|,\ldots,|I^k_{m_k}|}) \leq \sum_{k=1}^K \delta(\mathcal{S}^{|J_k|}),$$

which shows that in general the misspecified low σ limit is smaller than the corresponding well-specified limit.

2. Suppose θ^* is nonincreasing and nonconstant i.e., $\theta^* \in (-\mathcal{S}^n) \setminus \mathcal{S}^n$. Then $\Pi_{\mathcal{S}^n}(\theta^*)$ is constant (see [73] for various properties of $\Pi_{\mathcal{S}^n}$), so $K = 1$ and $\mu_1 = \frac{1}{n}\sum_{i=1}^n \theta_i^*$. We also claim $m_1 = 1$. Indeed, if $m_1 > 1$ then there exists some $j < n$ such that $\mu = \frac{1}{j}\sum_{i=1}^j \theta_i^* = \frac{1}{n-j}\sum_{i=j+1}^n \theta_i^*$. However, the fact that θ^* is nonincreasing and nonconstant implies $\frac{1}{j}\sum_{i=1}^j \theta_i^* > \frac{1}{n-j}\sum_{i=j+1}^n \theta_i^*$, a contradiction. Thus, Proposition 2.4.3 implies that both low σ limits are 1. (In fact, by combining the above argument with the proof of Proposition 2.4.3, we have shown that the intersection $T_{\mathcal{S}^n}(\Pi_{\mathcal{S}^n}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{S}^n}(\theta^*))^\perp$ is simply the subspace of constant sequences.) On the other hand, since $\Pi_{\mathcal{S}^n}(\theta^*)$ is constant, the low σ limit in the well-specified setting $Y \sim N(\Pi_{\mathcal{S}^n}(\theta^*), \sigma^2 I_n)$ is $\sum_{j=1}^n \frac{1}{j} \asymp \log n$, which is much larger.

   The logarithmic term appears here in the well-specified case due to the well-known spiking effect of isotonic regression (documented, for example, by Pal [71], Wu et al. [93], Zhang [97]). Indeed, the isotonic estimator is inconsistent near the end points which leads to the logarithm term in the risk. However, in the misspecified case when θ^* is nonincreasing and nonconstant, a combination of the proof of Theorem 2.3.1 (in particular Lemma 2.3.5) with the fact that $T_{\mathcal{S}^n}(\Pi_{\mathcal{S}^n}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{S}^n}(\theta^*))^\perp$ is the subspace of all constant sequences implies $\hat{\theta}(Y)$ is a constant sequence with probability increasing to 1 as $\sigma \downarrow 0$, in which case the constant value must be the sample mean $\bar{Y} := \frac{1}{n}\sum_{i=1}^n Y_i$. Alternatively, one can rephrase the geometric argument in Lemma 2.3.5 more simply in this example; when σ is small, Y is near θ^* and thus is also nondecreasing with high probability, in which case $\hat{\theta}(Y)$ is constant, due to the properties of the projection $\Pi_{\mathcal{S}^n}$. Hence, in this situation the estimator does not suffer from any spiking at the endpoints, and consequently there are no logarithmic terms in the risk in the misspecified case in the low sigma limit.

   Computations for the specific example when $\theta^* = (5, 3, 1, -1, -3, -5)$ appear in Table 2.1.

3. In the first half of Table 2.1 we consider three choices for θ^* that project to $\Pi_{\mathcal{S}^n}(\theta^*) = (0, 0, 0, 0, 0, 0)$. Here $K = 1$ and the sub-block sizes $|I_1^1|, \ldots, |I_{m_1}^1|$ are equal in each case (namely, the common block size is 1, 2, and 6 respectively), so we are in the special case (2.31). Thus, the limit is $\sum_{j=1}^{m_1} \frac{1}{j}$ where m_1 is the number of sub-blocks. We see that for the misspecified θ^* the low σ limits are smaller.

   One can heuristically interpret Theorem 2.3.1 for the example $\theta^* = (1, -1, 1, -1, 1, -1)$ as follows. With probability increasing to 1 as $\sigma \downarrow 0$, the estimator $\hat{\theta}(Y)$ is nondecreasing and piecewise constant on three equally sized blocks, so the low σ limit is the same as if we were estimating $(0, 0, 0)$ in $\mathcal{S}^3$.

4. Similarly in the second half of Table 2.1 we consider three θ^* that project to $\Pi_{\mathcal{S}^n}(\theta^*) = (-1, -1, -1, -1, 2, 2)$. Here $K = 2$ but, since the low σ limit decomposes, we can simply consider each constant piece separately. Again, we see that the more sub-

blocks I_i^j, the higher the statistical dimension, with the well-specified case having the most sub-blocks (all singletons).

5. The concrete examples we have considered so far have been in the special case (2.31). In a few other cases we can still provide the low σ limit. (See also Section 2.7.2 for further discussion.)

   a) If $K = 1$ and $m_1 = 2$, then the low σ limit is $\delta(S_{|I_1^1|,|I_2^1|})$. By Lemma 2.4.5, this is the same as the statistical dimension of the half space $\{u \in \mathbb{R}^2 : u_1/\sqrt{|I_1|} \leq u_2/\sqrt{|I_2|}\}$, which is 1.5. However, when $m_1 > 2$, it is difficult to compute $\delta(S_{|I_1^1|,\dots|I_{m_1}^1|})$ unless we are in the special case $|I_1^1| = \cdots = |I_{m_1}^1|$.

   b) In some other extreme cases we can get an approximation. For example, if

$$\theta^* = (0, \underbrace{1, \dots, 1}_{(n-2)/2}, \underbrace{-1, \dots, -1}_{(n-2)/2}, 0),$$

   then $\Pi_{\mathcal{S}^n}(\theta^*) = (0, \dots, 0)$, so the low σ limit is $\delta(\mathcal{S}_{1,n-2,1})$. Lemma 2.4.5 shows that this is the same as the statistical dimension of $\{u \in \mathbb{R}^3 : u_1 \leq u_2/\sqrt{n-2} \leq u_3\}$. As $n \to \infty$ tends to this set tends to $\{u \in \mathbb{R}^3 : u_1 \leq 0 \leq u_3\}$ which has statistical dimension $1 + \frac{1}{2} + \frac{1}{2} = 2$. Thus $\delta(\mathcal{S}_{1,n-2,1}) \to 2$ as $n \to \infty$. We used simulations to verify that the low σ limit is indeed near 2 even for $n = 20$.

2.5 Further discussion

2.5.1 Generalizing Theorem 2.3.1 to the non-polyhedral case

Note that Theorem 2.3.1 requires the condition (2.15) i.e., that $\mathcal{C}$ is locally a polyhedron near $\Pi_{\mathcal{C}}(\theta^*)$. Here we comment on the situation when $\mathcal{C}$ is non-polyhedral. Although non-polyhedral convex sets can be approximated by polyhedra, the low σ limit magnifies the local geometry of the set and ignores the goodness of such an approximation. As a stark counterexample, consider any closed convex $\mathcal{C} \subseteq \mathbb{R}^2$ with nonempty interior, and let $Z \sim N(0, I_n)$. For any polygon in $\mathbb{R}^2$, Theorem 2.3.1 implies that the low σ limits are either 0, 1/2, or 1 because in $\mathbb{R}^2$ the intersection of a convex cone with a line intersecting the origin is either the origin, a ray, or a line. Thus, for a sequence of polygons approximating $\mathcal{C}$ the sequence of corresponding low σ limits need not even have a limit, never mind the matter of two different sequences of polygonal approximations having a common limit. Therefore, the low σ limit for general $\mathcal{C}$ cannot be found using a polyhedral approximation.

In order to understand how the low σ limits behave for general $\mathcal{C}$, we consider the following specific example. Let $\mathcal{C} \coloneqq \{\theta \in \mathbb{R}^n : \|\theta\| \leq 1\}$ be the unit ball so that $\Pi_{\mathcal{C}}(x) = \frac{x}{\max\{\|x\|,1\}}$. Also let $\theta^* \coloneqq (r, 0, \dots, 0)$ for some $r > 1$ so that $\Pi_{\mathcal{C}}(\theta^*) = (1, 0, \dots, 0)$. By rotational symmetry of $\mathcal{C}$, the case of any general $\theta^* \notin \mathcal{C}$ can be reduced to this case.

In the corresponding well-specified case $Y \sim N(\Pi_{\mathcal{C}}(\theta^*), \sigma^2 I_n))$, the result (2.4) of Oymak and Hassibi [70] implies that the normalized misspecified risk (2.5) and the normalized excess risk (2.6) are equal in the low σ limit with common value

$$\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))) = n - \frac{1}{2},$$

since the tangent cone is the half space $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) = \{x \in \mathbb{R}^n : x_1 \leq 0\}$.

However, in the misspecified case, we observe some new phenomena that do not occur for polyhedra.

Proposition 2.5.1 (Low noise limits for the ball)**.** *Let $\mathcal{C} \coloneqq \{\theta \in \mathbb{R}^n : \|\theta\|_2 \leq 1\}$, $\theta^* \notin \mathcal{C}$, and $Y \sim N(\theta^*, \sigma^2 I_n)$. For the estimator $\hat{\theta}(Y) = \Pi_{\mathcal{C}}(Y)$, we have*

$$\lim_{\sigma \downarrow 0} \frac{1}{\sigma^2} M(\hat{\theta}, \theta^*) = \frac{n-1}{\|\theta^*\|^2}, \tag{2.32a}$$

$$\lim_{\sigma \downarrow 0} \frac{1}{\sigma^2} E(\hat{\theta}, \theta^*) = \frac{n-1}{\|\theta^*\|}. \tag{2.32b}$$

The proof involves direct computation and appears in Section 2.8.

We now highlight some of the interesting behavior. In the polyhedral case, both limits were equal; in the proof of Theorem 2.3.1 (in particular Lemma 2.3.5) we showed that with probability increasing to 1 (in the low σ limit), Y would be projected onto the hyperplane $(\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}$, producing the orthogonality required for the Pythagorean inequality (2.12) to become an equality. In the general case, the Pythagorean inequality is not tight, and we explicitly see from this example that even in the low noise limit the the excess risk can be strictly larger than the misspecified risk.

Note that in contrast to the corresponding well-specified case $Y \sim N(\Pi_{\mathcal{C}}(\theta^*), \sigma^2 I_n)$ which has limit $n - \frac{1}{2}$, the misspecified limits $\frac{n-1}{\|\theta^*\|^2}$ and $\frac{n-1}{\|\theta^*\|}$ both tend to $n-1$ as $\|\theta^*\| \downarrow 1$, so there is a "jump" in the limits between the misspecified and well-specified setting. This is also a feature of Theorem 2.3.1 when the polyhedron $\mathcal{C}$ has nonempty interior, as we discussed earlier (see Lemma 2.3.4).

This example shows that Theorem 2.3.1 does not hold for nonpolyhedral constraint sets $\mathcal{C}$, as the two normalized risks are not equal in this particular example of the unit ball, and moreover neither limit equals

$$\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}) = \delta(\{u : \langle u, \theta^* \rangle \leq 0\} \cap (\theta^*)^{\perp}) = \delta((\theta^*)^{\perp}) = n - 1.$$

The intuition for Theorem 2.3.1 is that, in the polyhedral case, the projections of Y largely end up in some face of the polyhedron $\mathcal{C}$, which can be approximated by a lower-dimensional cone, for which the statistical dimension is well defined. When $\mathcal{C}$ is not polyhedral, the generalization of this "face" is hard to conceptualize and is likely not well approximated by a cone, so a statistical dimension can not be even applied. Indeed, for general $\mathcal{C}$ such as the ball, tangent cones are extremely poor approximations for the set. Contrary to this

drawback, the result (2.4) of Oymak and Hassibi [70] shows that tangent cones are good enough for the well-specified setting. However for the misspecified setting, we expect that any general result for the low σ limits does not involve a statistical dimension of some cone, since the surface of $\mathcal{C}$ is the essential object of interest and cannot be approximated by some cone except in special settings like the polyhedral case.

As mentioned already, Theorem 2.3.1 shows that in the misspecified setting, the upper bound (2.7), which holds for all σ, is not tight in the low σ limit. One might ask whether a better upper bound for all σ can be achieved, but Figure 2.2 shows that for some values of σ the risks can be close to the upper bound, represented by the solid horizontal line. We observed this behavior in other examples (see also Figure 2.3): the risks can be close to the upper bound for some moderate values of σ, and then converge to the strictly smaller low σ limit. Replacing the upper bound (2.7), which is constant in σ, with a σ-dependent upper bound would be an interesting result, but it would have to be extremely dependent on the geometry of the set $\mathcal{C}$. In the following sections we further discuss the normalized risks as a function of σ.

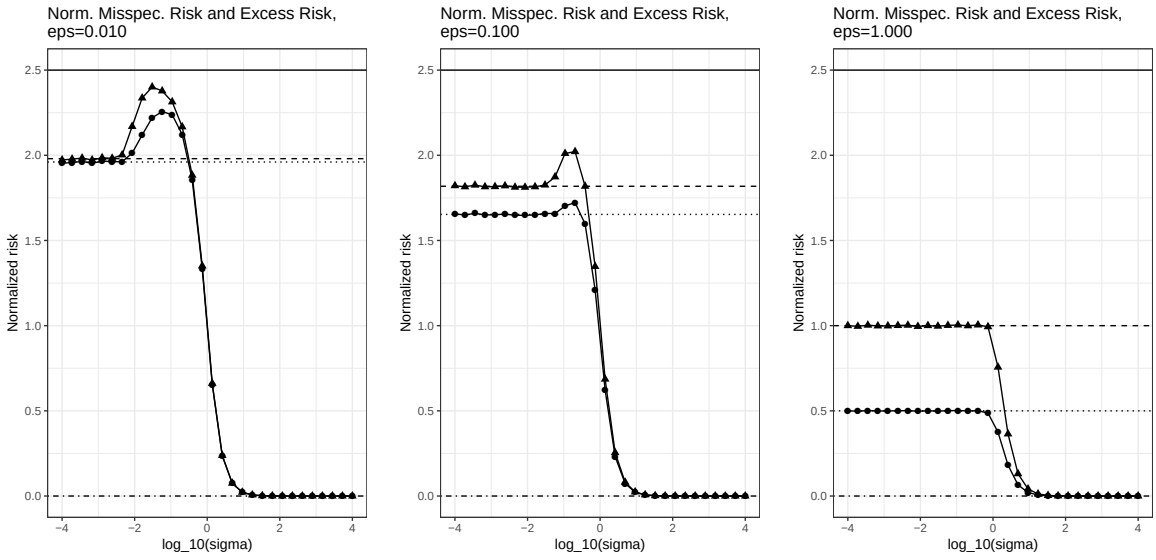


Figure 2.2: Empirical estimates of the normalized misspecified risk (•) and normalized excess risk (▲) plotted against $\log_{10}(\sigma)$, for the ball $\mathcal{C} = \{\theta \in \mathbb{R}^n : \|\theta\| \leq 1\}$ in the case $n = 3$ with $\theta^* = (1 + \epsilon, 0, 0)$ and $\epsilon \in \{0.01, 0.1, 1\}$. The solid horizontal line represents the upper bound $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))) = n - \frac{1}{2} = 2.5$ guaranteed by (2.7). The dotted lines and dashed lines are the predicted low σ limits $\frac{n-1}{(1+\epsilon)^2}$ and $\frac{n-1}{1+\epsilon}$ respectively. The dash-dot line is the high σ limit 0.

2.5.2 High noise limit

Although not interesting in its own right, the high noise limit of the normalized risks can help characterize the maximum risk as we discuss in the following section. Proofs for this section appear in Section 2.9.

For a closed convex set $\mathcal{C}$ we define the *core cone*

$$K_{\mathcal{C}} := \bigcap_{\theta \in \mathcal{C}} T_{\mathcal{C}}(\theta). \tag{2.33}$$

Recall the notation for the re-centered set $F_{\mathcal{C}}(\theta_0) = \{\theta - \theta_0 : \theta \in \mathcal{C}\}$ where $\theta_0 \in \mathcal{C}$. For a vector $v \in \mathbb{R}^n$ we let $\mathbb{R}_+ v := \{\alpha u : \alpha \geq 0\}$. We have the following equivalent characterizations of the core cone.

Lemma 2.5.2 (Characterizations of the core cone)**.** *Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a closed convex set. For any $\theta_0 \in \mathcal{C}$,*

$$K_{\mathcal{C}} \overset{(i)}{=} \{v : \mathbb{R}_+ v \subseteq F_{\mathcal{C}}(\theta_0)\} \overset{(ii)}{=} \bigcap_{\sigma > 0} \frac{F_{\mathcal{C}}(\theta_0)}{\sigma}.$$

Additionally, the inclusion $K_{\mathcal{C}} \subseteq T_{\mathcal{C}}(\theta)$ holds for any $\theta \in \mathcal{C}$. If furthermore $F_{\mathcal{C}}(\theta_0)$ is a cone, then the equality $K_{\mathcal{C}} = T_{\mathcal{C}}(\theta)$ holds if and only if $\theta_0 - (\theta - \theta_0) \in \mathcal{C}$; in particular, taking $\theta = \theta_0$ shows that $K_{\mathcal{C}} = T_{\mathcal{C}}(\theta_0) = F_{\mathcal{C}}(\theta_0)$.

Thus, up to a translation, the core cone can either be viewed as the result of shrinking $\mathcal{C}$ radially toward $\theta_0 \in \mathcal{C}$, or as the largest cone centered at $\theta_0 \in \mathcal{C}$ that is contained in $\mathcal{C}$. An interesting point is that $\theta_0 \in \mathcal{C}$ can be chosen arbitrarily.

Furthermore, in case when $\mathcal{C}$ is a cone, the core cone $K_{\mathcal{C}}$ is this cone $\mathcal{C}$, and we can characterize which tangent cones are the "smallest" in the sense that they equal the intersection (2.33) of all tangent cones.

The following result shows that under a boundedness condition, the core cone characterizes both high σ limits.

Proposition 2.5.3 (High noise limit)**.** *Let $\mathcal{C}$ be a closed convex set. Let $\theta^* \in \mathbb{R}^n$ and $Y := \theta^* + \sigma Z$ where Z is a zero mean random vector with $\mathbb{E}\|Z\|^2 < \infty$. If the condition*

$$\sup_{x \in \mathbb{R}^n} \left(\|\Pi_{F_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))}(x)\|^2 - \|\Pi_{K_{\mathcal{C}}}(x)\|^2 \right) < \infty. \tag{2.34}$$

holds, then

$$\lim_{\sigma \to \infty} \frac{1}{\sigma^2} M(\hat{\theta}, \theta^*) = \lim_{\sigma \to \infty} \frac{1}{\sigma^2} E(\hat{\theta}, \theta^*) = \delta(K_{\mathcal{C}}).$$

The main hurdle in applying Proposition 2.5.3 is verifying the condition (2.34). The following result covers two cases where it is easy to verify the condition.

Corollary 2.5.4 (Orthant and bounded sets)**.** *Let $\theta^* \in \mathbb{R}^n$ and $Y \sim N(\theta^*, \sigma^2 I_n)$.*

- *If $\mathcal{C} = \mathbb{R}^n_+$ is the nonnegative orthant, then the high σ limits are $\delta(\mathbb{R}^n_+) = n/2$.*
- *Let $\mathcal{C}$ be a closed convex set. $K_{\mathcal{C}} = \{0\}$ if and only if $\mathcal{C}$ is bounded, in which case both high σ limits are* 0.

Figure 2.2 and Figure 2.3 illustrate the result of this corollary.

Verifying (2.34) for more general $\mathcal{C}$ is more difficult. We believe it might hold for polyhedral cones with any θ^*, in which case Proposition 2.5.3 would imply that the high σ limits are $\delta(\mathcal{C})$. An interesting feature of the examples presented thus far is that the high σ limits (including the veracity of (2.34)) do not depend on θ^*.

Remark 2.5.5. *More generally, suppose $\mathcal{C}$ is a general cone. By applying Lemma 2.5.2 with $\theta_0 = 0$ and $\theta = \Pi_{\mathcal{C}}(\theta^*)$, we observe that the core cone $K_{\mathcal{C}}$ is $\mathcal{C}$, and moreover $\mathcal{C} \subseteq T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$, with equality if and only if $-\Pi_{\mathcal{C}}(\theta^*) \in \mathcal{C}$. Thus, if the condition* (2.34) *holds, then Proposition 2.5.3 implies the high σ limits are $\delta(\mathcal{C})$, and moreover Lemma 2.5.2 implies that these limits equal Bellec's upper bound* (2.7)*, $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$, if and only if θ^* satisfies $-\Pi_{\mathcal{C}}(\theta^*) \in \mathcal{C}$.*

However, the condition (2.34) does not hold for all $\mathcal{C}$. One can verify numerically that the epigraph $\mathcal{C} \coloneqq \{u \in \mathbb{R}^2 : u_2 \geq u_1^2\}$, whose core cone is $K_{\mathcal{C}} = \{(0, u_2) : u_2 \geq 0\}$, does not satisfy (2.34). Simulations also show that the high σ limits are larger than $\delta(K_{\mathcal{C}}) = 1/2$. In general, it is unclear exactly when the core cone does or does not characterize the high σ limits.

2.5.3 Maximum normalized risk

Our low and high σ limit results Theorem 2.3.1 and Proposition 2.5.3 provides an incomplete characterization of the maximum normalized risks (2.10). As mentioned already in (2.7), $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$ is an upper bound for both suprema.

In the well-specified case $\theta^* \in \mathcal{C}$, both suprema reduce to the usual normalized risk $\sigma^{-2}R(\hat{\theta}, \theta^*)$; moreover the upper bound becomes $\delta(T_{\mathcal{C}}(\theta^*))$, and is attained as $\sigma \downarrow 0$ by the result (2.4) of Oymak and Hassibi [70].

However, in the misspecified case we have shown in Theorem 2.3.1 that in general the low σ limit does not attain the upper bound (2.7). Moreover, simulations show that in some cases even the suprema do not attain the upper bound; see Figure 2.2 and Figure 2.3. We see that for some cases the suprema are close to the upper bound, but for others it is much smaller.

Of course, if one can show that either the low σ limit or the high σ limit is equal to the upper bound $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)))$, then we know the upper bound is attained either as $\sigma \downarrow 0$ or $\sigma \to \infty$ respectively. However, in the settings of Theorem 2.3.1 and Proposition 2.5.3, this seldom happens. As discussed already, if $\mathcal{C}$ is polyhedral with nonempty interior, then the low σ limit is strictly smaller than the upper bound. If Proposition 2.5.3 applies, then $K_{\mathcal{C}} = \bigcap_{\theta \in \mathcal{C}} T_{\mathcal{C}}(\theta) \subseteq T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$ shows that the high σ limit is typically strictly smaller than

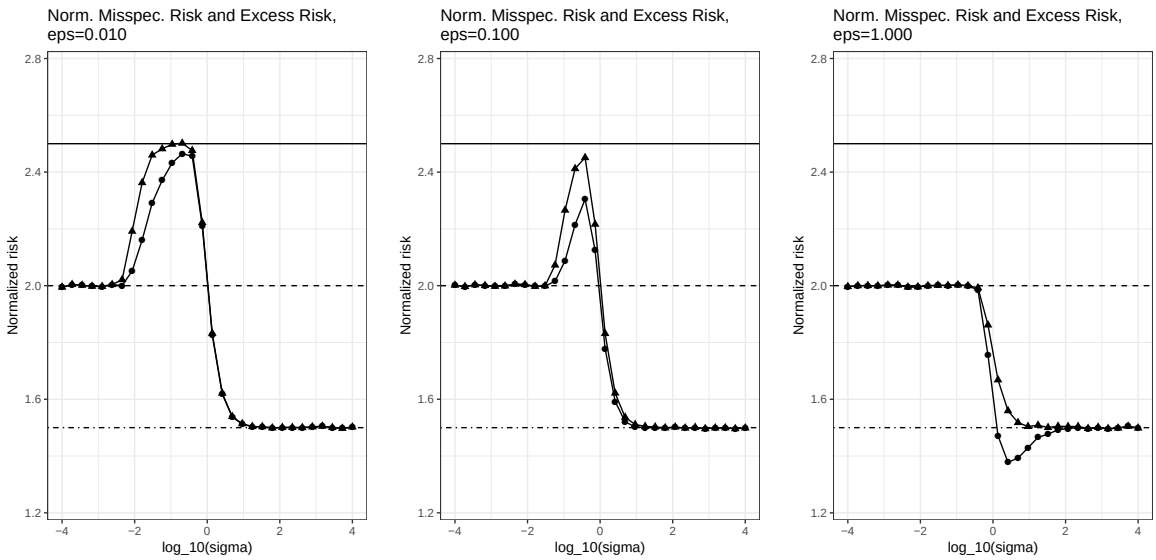


Figure 2.3: Empirical estimates of the normalized misspecified risk (•) and normalized excess risk (▲) plotted against $\log_{10}(\sigma)$, for the orthant $\mathcal{C} := \mathbb{R}^3_+$ and $\theta^* = (1, 1, -\epsilon)$ with $\epsilon \in \{0.01, 0.1, 1\}$. The solid horizontal line represents the upper bound $\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))) = n - \frac{1}{2}$ guaranteed by (2.7). The dashed line is the common low σ limit $n - 1$ (see Corollary 2.4.1). The dash-dot line is the high σ limit $\delta(\mathbb{R}^n_+) = 3/2$.

the upper bound; for the special case where $\mathcal{C}$ is a cone, see Remark 2.5.5 for a necessary and sufficient condition for the high σ limit to equal the upper bound.

Thus in most cases the suprema are attained at some moderate values of σ, but it is difficult to provide a characterization of these maximizing values σ, as well as the value of the suprema and whether they are close to the upper bound or not. The plots suggest that as θ^* gets closer to $\mathcal{C}$, the suprema get closer to the upper bound as well.

2.6 Proofs of lemmas in Section 2.3

The next lemma is a technical device for representing the largest face of a polyhedral cone that lies in a particular hyperplane. It is useful for proving Lemma 2.3.3 and Lemma 2.3.5.

Lemma 2.6.1 (Largest face in hyperplane)**.** *Let $\mathcal{K} = \{u : Au \leq 0\} \subseteq \mathbb{R}^n$ be a polyhedral cone, where $A \in \mathbb{R}^{m \times n}$ has distinct rows. For each $y \in \mathbb{R}^n$, consider the subsets $J \subseteq \{1, \ldots, m\}$ satisfying*

$$\{u : A_J u = 0\} \subseteq (y - \Pi_{\mathcal{K}}(y))^{\perp}. \tag{2.35}$$

We let J_y denote the smallest such subset.

This subset J_y characterizes a face of $\mathcal{K}$ in the following way.

$$\mathcal{K} \cap (y - \Pi_{\mathcal{K}}(y))^{\perp} = \{u : A_{J_y} u = 0, A_{J_y^c} u \leq 0\}. \tag{2.36}$$

Proof. The optimality condition for a projection onto a cone (2.14) implies $\langle y-\Pi_{\mathcal{K}}(y), u\rangle \leq 0$ for all $u \in \mathcal{K}$. If $\mathcal{K}$ contains both u and $-u$, then this implies $u \in (y - \Pi_{\mathcal{K}}(y))^{\perp}$. Thus for $J = \{1, \dots, m\}$, (2.35) holds because $\{u : A_J u = 0\} \subseteq \mathcal{K}$. This shows the existence of subsets J that satisfy (2.35).

Next, note that if J and J' both satisfy (2.35), then $J \cap J'$ does as well, because

$$\{u : A_{J\cap J'}u = 0\} = \{u + v : A_J u = A_{J'} v = 0\} \subseteq (y - \Pi_{\mathcal{K}}(y))^{\perp}.$$

So, letting J_y be the intersection of all J satisfying (2.35) yields the unique subset of minimal size.

The $\supseteq$ inclusion in (2.36) follows immediately from $\{u : A_{J_y}u = 0\} \subseteq (y - \Pi_{\mathcal{K}}(y))^{\perp}$. For the other inclusion, suppose $v \in \mathcal{K} \cap (y - \Pi_{\mathcal{K}}(y))^{\perp}$. Then $Av \leq 0$, so it remains to verify $A_{J_y}v = 0$. That is, if $J \subseteq \{1, \dots, m\}$ denotes the indices j for which $\langle a_j, v\rangle = 0$, we want to show $J_y \subseteq J$; furthermore, this reduces to showing J satisfies (2.35), by minimality of J_y.

Any u satisfying $A_J u = 0$ can be rewritten as $u = v + w$ for some w also satisfying $A_J w = 0$. There exists some $c > 0$ such that both $v + cw$ and $v - cw$ are in $\mathcal{K}$ because all the linear constraints outside of J are strict inequalities at v. Then, the optimality condition for the projection onto a cone, yields $\langle v + cw, y - \Pi_{\mathcal{K}}(y)\rangle \leq 0$ and $\langle v - cw, y - \Pi_{\mathcal{K}}(y)\rangle \leq 0$. Since $v \in (y - \Pi_{\mathcal{K}}(y))^{\perp}$, this yields $w \in (y - \Pi_{\mathcal{K}}(y))^{\perp}$ and thus $u \in (y - \Pi_{\mathcal{K}}(y))^{\perp}$, which verifies that J satisfies (2.35). □

Proof of Lemma 2.3.2. By definition there exist an integer m, matrix $A \in \mathbb{R}^{m\times n}$, and vector $b \in \mathbb{R}^m$ such that $\mathcal{C} \coloneqq \{u \in \mathbb{R}^n : Au \leq b\}$. Fix $\theta^* \in \mathbb{R}^n$ and let $\theta_0 \coloneqq \Pi_{\mathcal{C}}(\theta^*)$. We will show

$$T_{\mathcal{C}}(\theta_0) = \{u : A_J u \leq 0\},$$

where $J = \{j : \langle a_j, \theta_0\rangle = b_j\}$. Then $T_{\mathcal{C}}(\theta_0)$ is a polyhedral cone.

If $u \in T_{\mathcal{C}}(\theta_0)$ then for some $r^* > 0$ we have $\theta_0 + ru \in \mathcal{C}$. Thus, $b_J \geq A_J(\theta_0 + ru) = b_J + rA_J u$ which implies $A_J u \leq 0$.

Conversely, suppose u satisfies $A_J u \leq 0$. Choose $r^* > 0$ so that $r\langle a_j, u\rangle \leq b_j - \langle a_j, \theta_0\rangle$ for all $j \notin J$. This is possible because $b_j > \langle a_j, \theta_0\rangle$ for each $j \notin J$. Then $\theta_0 + r^* u \in \mathcal{C}$ so $u \in T_{\mathcal{C}}(\theta_0)$.

Finally, we need to prove the second part of the locally polyhedral condition (2.15), which will follow if we show $T_{\mathcal{C}}(\theta_0) \cap B_{r^*}(0) \subseteq F_{\mathcal{C}}(\theta_0)$ for some $r > 0$. If $u \in T_{\mathcal{C}}(\theta_0)$ then $A_J u \leq 0 = b_J - A_J\theta_0$, so it suffices to find some r such that $A_{J^c}u \leq b_{J^c} - A_{J^c}\theta_0$ for any $u \in B_{r^*}(0)$. For each $j \notin J$, we have $\langle a_j, \theta_0\rangle < b_j$ so there exists some $r^* > 0$ such that all $\theta \in B_{r^*}(\theta_0)$ satisfy $\langle a_j, \theta\rangle < b_j$ for all $j \notin J$. Taking $u = \theta - \theta_0$ concludes the proof. □

Proof of Lemma 2.3.3. Let $\mathcal{T} \coloneqq \{u + \Pi_{\mathcal{C}}(\theta^*) : u \in T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))\}$. Using the locally polyhedral condition (2.15) and continuity [49] of $\Pi_{\mathcal{C}}$ and $\Pi_{\mathcal{T}}$, we have $\Pi_{\mathcal{T}}(\theta^*) = \Pi_{\mathcal{C}}(\theta^*)$ (e.g., see the verification of (2.24)), and thus translating yields $\Pi_{T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))}(\theta^* - \Pi_{\mathcal{C}}(\theta^*)) = 0$. Applying Lemma 2.6.1 with $\mathcal{K} = T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$, $y = \theta^* - \Pi_{\mathcal{C}}(\theta^*)$, and $\Pi_{\mathcal{K}}(y) = 0$ concludes the proof. □

Proof of Lemma 2.3.4. Fix $\theta_0 \in \mathcal{C}$. For any $\theta^* \notin \mathcal{C}$ such that $\Pi_{\mathcal{C}}(\theta^*) = \theta_0$, Lemma 2.3.2 implies the locally polyhedral condition (2.15) holds, and thus Lemma 2.3.3 establishes that $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ is a face of the tangent cone $T_{\mathcal{C}}(\theta_0)$.

Since the tangent cone has finitely many faces, the supremum is actually a maximum over the statistical dimensions of finitely many such lower-dimensional faces. Thus it remains to show

$$\delta(T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp) < \delta(T_{\mathcal{C}}(\theta_0))$$

for each $\theta^* \notin \mathcal{C}$ such that $\Pi_{\mathcal{C}}(\theta^*) = \theta_0$.

The set $(\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ is a hyperplane (not all of $\mathbb{R}^n$) because $\theta^* \notin \mathcal{C}$. Using the fact that the tangent cone $T_{\mathcal{C}}(\theta_0)$ has nonempty interior (because it contains the translation $F_{\mathcal{C}}(\theta_0)$ of $\mathcal{C}$), we see that the intersection $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^\perp$ is a face that that lies in a strictly lower-dimensional subspace of $\mathbb{R}^n$, and is therefore strictly smaller than the full cone $T_{\mathcal{C}}(\theta_0)$. Thus, we just need to show $\delta(T') < \delta(T)$ for any polyhedral cone T with nonempty interior in $\mathbb{R}^n$, and any face T' of T that lies in a strictly lower-dimensional subspace of $\mathbb{R}^n$.

For a point $x \in \mathbb{R}^n$ and a set $S \subseteq \mathbb{R}^n$ let $d(x, S) \coloneqq \inf_{\theta \in S} \|x - \theta\|$. Note that the Moreau decomposition for cones [3, Sec. B] implies $\|\Pi_{\mathcal{K}}(x)\| = d(x, \mathcal{K}^\circ)$ for any $x \in \mathbb{R}^n$ and any cone $\mathcal{K}$, where $\mathcal{K}^\circ \coloneqq \{u \in \mathbb{R}^n : \langle u, \theta \rangle \leq 0, \forall \theta \in \mathcal{K}\}$ denotes the polar cone of $\mathcal{K}$. Since $T^\circ \subseteq (T')^\circ$, we have

$$d(x, (T')^\circ) \leq d(x, T^\circ), \qquad \forall x \in \mathbb{R}^n.$$

Thus, if we show the random vector Z has nonzero probability of being in the set

$$\mathcal{A} \coloneqq \{x \in \mathbb{R}^n : d(x, (T')^\circ) < d(x, T^\circ)\} = \{x \in \mathbb{R}^n : \|\Pi_{T'}(x)\| < \|\Pi_T(x)\|\},$$

then we immediately have the desired strict inequality

$$\delta(T') = \mathbb{E} d(Z, (T')^\circ) < \mathbb{E} d(Z, T^\circ) = \delta(T).$$

To prove the above claim that $\mathbb{P}(Z \in \mathcal{A}) > 0$, we show below that the interior of T is contained in $\mathcal{A}$; then our assumption on Z will conclude the proof.

Let x be in the interior of T. Then $x \in T \setminus T'$. Moreover, if we let U be the smallest linear subspace of $\mathbb{R}^n$ containing T', then $x \notin U$ as well. Note the the Pythagorean theorem implies

$$\|\Pi_T(x)\|^2 = \|x\|^2 = \|\Pi_U(x)\|^2 + \|x - \Pi_U(x)\|^2 > \|\Pi_U(x)\|^2. \tag{2.37}$$

We also have

$$\Pi_{T'}(x) = \operatorname*{argmin}_{\theta \in T'} \|\theta - x\|^2 = \operatorname*{argmin}_{\theta \in T'} \big\{ \|\theta - \Pi_U(x)\|^2 + \|\Pi_U(x) - x\|^2 \big\} = \Pi_{T'}(\Pi_U(x)),$$

so combining this with the above inequality (2.37) and the optimality condition (2.14) for the projection of $\Pi_U(x)$ onto the cone T', we have

$$\|\Pi_{T'}(x)\|^2 = \|\Pi_{T'}(\Pi_U(x))\|^2 = \|\Pi_U(x)\|^2 - \|\Pi_U(x) - \Pi_{T'}(\Pi_U(x))\|^2 \leq \|\Pi_U(x)\|^2 < \|\Pi_T(x)\|^2,$$

and thus $x \in \mathcal{A}$. □

Proof of Lemma 2.3.5. The lemma holds immediately if $\theta^* \in \mathcal{T}$, so we assume $\theta^* \notin \mathcal{T}$.

By translating, we may without loss of generality assume $\Pi_{\mathcal{T}}(\theta^*) = 0$ so that the cone is centered at 0 and can be written as $\mathcal{T} = \{u : Au \le 0\}$ for some number of constraints m and some matrix $A \in \mathbb{R}^{m\times n}$. The objective then reduces to

$$\Pi_{\mathcal{T}}(y) \in (\theta^*)^\perp, \qquad \text{for all } y \in B_r(\theta^*).$$

For any $y \in \mathbb{R}^n$ let $J_y \subseteq \{1, \dots, m\}$ be as defined in Lemma 2.6.1 for our polyhedral cone $\mathcal{T}$; it characterizes the largest face of $\mathcal{T}$ that lies in $(\theta^*)^\perp$. We claim there exists $r > 0$ such that

$$\{u : A_{J_y} u = 0\} \subseteq (\theta^*)^\perp, \quad \forall y \in B_r(\theta^*). \tag{2.38}$$

If not, then there exists a sequence of points $y_k \notin \mathcal{T}$ converging to θ^* such that $\{u : A_{J_{y_k}} u = 0\} \not\subseteq (\theta^*)^\perp$ for all k. Since there are finitely many distinct subsets J_{y_k}, we may take a subsequence and without loss of generality assume it is common subset $J = J_{y_k}$ for all k, and $\{u : A_J u = 0\} \not\subseteq (\theta^*)^\perp$. By the definition (2.35) of J_{y_k}, any u satisfying $A_J u = 0$ also satisfies $\langle y_k - \Pi_{\mathcal{T}}(y_k), u\rangle = 0$. By continuity of $\Pi_{\mathcal{T}}$ and taking $k \to \infty$, we have $\langle \theta^*, u\rangle = 0$ as well, a contradiction.

Finally, since the optimality condition (2.11) for $\Pi_{\mathcal{T}}$ implies $\langle \Pi_{\mathcal{T}}(y), y - \Pi_{\mathcal{T}}(y)\rangle = 0$ for any $y \in \mathbb{R}^n$, (2.36) implies $\Pi_{\mathcal{T}}(y) \in \{u : A_{J_y} u = 0\}$. Combining this with (2.38) concludes the proof. □

2.7 Proofs for Section 2.4.2 (isotonic regression)

2.7.1 Proofs of block monotone cone lemmas

Proof of Lemma 2.4.4. The first claim follows from decomposing the squared Euclidean distance into blocks.

$$\begin{aligned}
\min_{v\in\mathcal{S}_{|I_1|,\dots,|I_m|}} \|v - z\|^2 &= \min_{x\in\mathcal{S}^m} \sum_{j=1}^m \sum_{i\in I_j} (x_j - z_i)^2 \\
&= \min_{x\in\mathcal{S}^m} \sum_{j=1}^m \sum_{i\in I_j} ((x_j - \bar{z}_{I_j})^2 + (\bar{z}_{I_j} - y_i)^2) \\
&= \sum_{j=1}^m \sum_{i\in I_j} (z_i - \bar{z}_{I_j})^2 + \min_{x\in\mathcal{S}^m} \sum_{j=1}^m |I_j|(x_j - \bar{z}_{I_j})^2.
\end{aligned}$$

Let Z and Z' be standard Gaussian in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. If $|I_1| = \cdots = |I_m| = r$, then the first claim implies

$$\delta(\mathcal{S}_{|I_1|,\dots,|I_m|}) := \mathbb{E}\|\Pi_{\mathcal{S}_{|I_1|,\dots,|I_m|}}(Z)\|^2 \overset{(i)}{=} r\mathbb{E}\|\Pi_{\mathcal{S}^m}(Z'/\sqrt{r})\|^2 \overset{(ii)}{=} \mathbb{E}\|\Pi_{\mathcal{S}^m}(Z')\|^2 =: \delta(\mathcal{S}^m) = \sum_{j=1}^m \frac{1}{j},$$

where (i) is due to $Z'/\sqrt{r} \overset{d}{=} (\bar{Z}_{I_1}, \dots, \bar{Z}_{I_m})$, and (ii) is due to $\Pi_{\mathcal{C}}(cx) = c\Pi_{\mathcal{C}}(x)$ for a cone $\mathcal{C}$ and $c > 0$ (e.g., [10, Sec. 1.6]). The statistical dimension of $\mathcal{S}^m$ is proved by Amelunxen et al. [3, Sec. D.4]. □

Proof of Lemma 2.4.5. We use two useful properties of the statistical dimension of any cone $\mathcal{C}$ [3, Prop. 3.1].

- Rotational invariance: for any orthogonal transformation Q, we have $\delta(Q\mathcal{C}) = \delta(\mathcal{C})$.
- Invariance under embedding: $\delta(\mathcal{C} \times \{0\}^k) = \delta(\mathcal{C})$.

Thus it suffices to provide an orthogonal transformation Q such that $Q\mathcal{S}_{|I_1|,\dots,|I_m|}$ is an embedding of the cone (2.30) into $\mathbb{R}^n$.

Let e_i denote the ith standard basis vector in $\mathbb{R}^n$. Let the last element of each block be denoted $k_j \coloneqq \max I_j$ for $1 \le j \le m$, with $k_0 = 0$ for convenience. The block monotone cone $\mathcal{S}_{|I_1|,\dots,|I_m|}$ is defined by the following constraints for $u \in \mathbb{R}^n$.

$$\langle e_i - e_{i+1}, u\rangle \le 0, \qquad i \in \{k_1, \dots, k_m\} \tag{2.39a}$$

$$\langle e_i - e_{i+1}, u\rangle = 0, \qquad i \in \{1, \dots, n-1\} \setminus \{k_1, \dots, k_m\} \tag{2.39b}$$

Let us focus on an arbitrary block I_j. Consider the $|I_j| \times |I_j|$ matrix

$$\tilde{A}_j = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

Because $\tilde{A}_j$ is full rank, the QR decomposition implies there exists an $|I_j| \times |I_j|$ orthogonal matrix $\tilde{Q}_j$ such that $\tilde{R}_j \coloneqq \tilde{Q}_j \tilde{A}_j$ is upper triangular with positive diagonal entries, and this decomposition is unique.

The block diagonal matrix Q with blocks $\tilde{Q}_1, \dots, \tilde{Q}_m$ is an $n \times n$ orthogonal matrix. Let A and R also be block diagonal, each constructed similarly using the $\tilde{A}_j$ and the $\tilde{R}_j$ respectively, so that $U = QA$. We consider $Q\mathcal{S}_{|I_1|,\dots,|I_m|}$. We use the fact that if $v = Qu$ then $\langle b, u\rangle \le 0 \iff \langle Qb, v\rangle \le 0$ to rewrite the constraints (2.39a) and (2.39b). The following hold for each $j = 1, \dots, m$.

- Note that the ith column of A is $a_i = e_i - e_{i+1}$ when $k_{j-1} < i < k_j$. For these i, the equality constraints (2.39b) after the transformation become $0 = \langle Q(e_i - e_{i+1}), v\rangle = \langle r_i, v\rangle$ where r_i is the ith column of R. Since $\tilde{R}_j$ is upper triangular with nonzero diagonal entries (because $\tilde{A}_j$ is full rank), induction on $i = k_{j-1}+1, \dots, k_j - 1$ implies

$$v_i = 0, \qquad k_{j-1} < i < k_j.$$

- When $j < m$, we have $e_{k_j} = a_k$ and $e_{k_j+1} = a_{k_j} + a_{k_j+1} + \cdots + a_{k_{j+1}}$ Thus for $j < m$ the inequality constraint $\langle e_{k_j} - e_{k_j+1}, u\rangle \leq 0$ becomes

$$0 \geq \langle Q(e_{k_j} - e_{k_j+1}), v\rangle = \langle r_{k_j} - r_{k_j+1} - r_{k_j+2} - \cdots - r_{k_{j+1}}, v\rangle = \langle r_{k_j} - r_{k_{j+1}}, v\rangle,$$

  where the last equality is due to $\langle r_i, v\rangle = 0$ for $k_j < i < k_{j+1}$, by the previous point. Since $\tilde{R}_j$ and $\tilde{R}_{j+1}$ are each upper triangular, the inequality reduces to $r_{k_j,k_j} v_{k_j} \leq r_{k_{j+1},k_{j+1}} v_{k_{j+1}}$, where $r_{k,k}$ denotes the kth diagonal entry of R. Lemma 2.7.1 (proved below) computes these diagonal elements and yields

$$\frac{v_{k_j}}{\sqrt{|I_j|}} \leq \frac{v_{k_{j+1}}}{\sqrt{|I_{j+1}|}}.$$

Therefore we have shown that $Q\mathcal{S}_{|I_1|,\ldots,|I_m|}$ consists of all vectors satisfying

$$\frac{v_{k_1}}{\sqrt{|I_1|}} \leq \frac{v_{k_2}}{\sqrt{|I_2|}} \leq \cdots \leq \frac{v_{k_m}}{\sqrt{|I_m|}}, \text{ and } v_i = 0, \forall i \in \{1,\ldots,n\} \setminus \{k_1,\ldots,k_m\}.$$

We have thus verified the claim that $Q\mathcal{S}_{|I_1|,\ldots,|I_m|}$ is an embedding of (2.30) into $\mathbb{R}^n$.

When the blocks all have equal size r, the cone (2.30) becomes the monotone cone $\mathcal{S}^m$, whose statistical dimension is $\sum_{j=1}^m \frac{1}{j}$ [3, Sec. D.4]. □

Lemma 2.7.1. *Consider the $n \times n$ matrix*

$$A = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}.$$

There exists a unique orthogonal matrix Q and a unique upper triangular matrix R with positive diagonal entries such that $A = QR$. The bottom-right entry of R is $r_{n,n} = 1/\sqrt{n}$.

Proof. Let q_i be the ith column of Q. The last column q_n is orthogonal to the span of the first $n-1$ columns of A, so q_n is either $(1,\ldots,1)/\sqrt{n}$ or its negative. The positivity constraint on the diagonal entries of R implies the former, and thus $r_{n,n} = \langle q_n, e_n\rangle = 1/\sqrt{n}$. □

2.7.2 Statistical dimension of the block monotone cone in general

In Lemma 2.4.5 we provided an expression for the statistical dimension of the block monotone cone $\mathcal{S}_{|I_1|,\ldots,|I_m|}$ when the block sizes were equal. In general, the statistical dimension can be higher or lower than $\sum_{j=1}^m \frac{1}{j}$. Consider the following examples for $m = 3$.

Lemma 2.4.5 implies $\mathcal{S}_{n-2,1,1}$ has the same statistical dimension as $\{v \in \mathbb{R}^3 : v_1/\sqrt{n-2} \leq v_2 \leq v_3\}$. As $n \to \infty$ this latter cone approaches $\{v \in \mathbb{R}^3 : 0 \leq v_2 \leq v_3\}$ which has statistical dimension $1 + \left(\frac{1}{8} \cdot 2 + \frac{1}{2} \cdot 1\right) = \frac{7}{4} = 1.75$, which is smaller than $\sum_{j=1}^{3} \frac{1}{j} = \frac{11}{6} = 1.8\bar{3}$.

On the other hand, $\mathcal{S}_{1,n-2,1}$ has the same statistical dimension as $\{v \in \mathbb{R}^3 : v_1 \leq v_2/\sqrt{n-2} \leq v_3\}$. As $n \to \infty$ this latter cone approaches $\{v \in \mathbb{R}^3 : v_1 \leq 0, v_3 \geq 0\}$ which has statistical dimension $1 + \frac{1}{2} + \frac{1}{2} = 2$, which is larger than $1.8\bar{3}$.

We suspect that the approach used to prove the statistical dimension of $\mathcal{S}^n$ [3, Sec. D.4], which uses the theory of finite reflection groups, cannot be generalized for $\mathcal{S}_{|I_1|,\ldots,|I_m|}$, due to the asymmetry of (2.30). However, using a result of Klivans and Swartz [55], it is possible to show that the average statistical dimension among all block monotone cones with a given [unordered] set of m block sizes is H_m [2, Prop. 6.6].

2.7.3 Proof of Proposition 2.4.3

When applying Theorem 2.3.1, it is useful to characterize $\mathcal{S}^n$ and its tangent cones using conic generators. If $T \subseteq \mathbb{R}^n$ is a cone and there exist $x_1, \ldots, x_p \in T$ such that

$$T = \left\{ \sum_{i=1}^{p} \alpha_i x_i : \alpha_i \geq 0, \forall i \right\},$$

then we call $x_1, \ldots, x_p$ the *conic generators* of T, and write

$$T = \operatorname{cone}\{x_1, \ldots, x_p\}.$$

Lemma 2.7.2. *Let $\theta^* \in \mathbb{R}^n$ and let $\mathcal{C} \subseteq \mathbb{R}^n$ be closed and convex. If the tangent cone $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$ is generated by $x_1, \ldots, x_p \in \mathbb{R}^n$, i.e. $T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) = \operatorname{cone}\{x_1, \ldots, x_p\}$, then*

$$T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp} = \operatorname{cone}(\{x_1, \ldots, x_p\} \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}).$$

Proof of Lemma 2.7.2. The inclusion $\supset$ is immediate, so it remains to prove the inclusion $\subseteq$. Note that the optimality condition (2.11) implies $\langle \theta^* - \Pi_{\mathcal{C}}(\theta^*), x \rangle \leq 0$ for any $x \in T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$. In particular, if $v \in T_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}$, then v can be written as the conical combination $v = \sum_{i=1}^{p} \alpha_i x_i$ with $\alpha_i \geq 0$, and we have

$$0 = \langle \theta^* - \Pi_{\mathcal{C}}(\theta^*), v \rangle = \sum_{i=1}^{p} \alpha_i \underbrace{\langle \theta^* - \Pi_{\mathcal{C}}(\theta^*), x_i \rangle}_{\leq 0}.$$

Thus, if a generator x_i is not in the hyperplane $(\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}$, then $\alpha_i = 0$, so x_i does not contribute in the conical combination of v. Thus, v can be written as a conical combination of generators in $(\theta^* - \Pi_{\mathcal{C}}(\theta^*))^{\perp}$. □

We are now ready to prove Proposition 2.4.3.

Proof of Proposition 2.4.3. By Theorem 2.3.1, it suffices to prove that the statistical dimension term is $\sum_{k=1}^{K} \delta\left(\mathcal{S}_{|I_1^k|,\ldots,|I_{m_k}^k|}\right)$.

For $p \geq 1$ let

$$M_p \coloneqq \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & 1 & \cdots & 1 \\ & 1 & \cdots & 1 \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix} \in \mathbb{R}^{(p+1)\times p}.$$

The rows of M_p are the conic generators of $\mathcal{S}^p$.

Suppose first that $\Pi_{\mathcal{S}^n}(\theta^*)$ is constant, so that $K = 1$ and $J_1 = \{1, \ldots, n\}$. Then $\Pi_{\mathcal{S}^n}(\theta^*) = (\mu_1, \mu_1, \ldots, \mu_1)$ where $\mu_1 \coloneqq \frac{1}{n}\sum_{i=1}^{n} \theta_i^*$; this follows directly by minimizing $\sum_{i=1}^{n}(\theta_i^* - \mu_1)^2$ with respect to μ_1.

The finest partition $(I_1^1, \ldots, I_{m_1}^1)$ of J_1 into blocks satisfying (2.28) can be constructed greedily as follows. Begin populating I_1^1 with the elements of $\{1, \ldots, n\}$ in order, stopping as soon as the mean of the elements of I_1^1 is μ_1. Then begin populating I_2^1 with the remaining elements in order, again stopping when the mean of the elements in I_2^1 is μ_1. Continue in this manner until the last element n is placed in a subset $I_{m_1}^1$. The mean of the elements of this last subset $I_{m_1}^1$ is μ_1 as well, since the mean of all components of θ^* is μ_1. Thus this partition satisfies (2.28). To establish uniqueness, note that if some other partition of J_1 satisfies (2.28), then our partition $(I_1^1, \ldots, I_{m_1}^1)$ must be a refinement, due to the greedy construction.

Because $\Pi_{\mathcal{S}^n}(\theta^*)$ is constant, the tangent cone there is $T_{\mathcal{S}^n}(\Pi_{\mathcal{S}^n}(\theta^*)) = \mathcal{S}^n$ [10, Prop. 3.1], which is generated by the rows of M_n. In order to use Lemma 2.7.2, we need to determine which rows of M_n are in the hyperplane $(\theta^* - \Pi_{\mathcal{S}^n}(\theta^*))^\perp$. We already know the mean of the components of $\theta^* - \Pi_{\mathcal{S}^n}(\theta^*)$ is zero, so the first two rows are in the hyperplane.

We claim that exactly $m_1 - 1$ of the remaining $n-1$ rows of M_n also lie in the hyperplane. Explicitly, if $(I_1^1, \ldots, I_{m_1}^1)$ is without loss of generality assumed to be sorted in increasing order, then the remaining rows of M_n that lie in the hyperplane are the indicator vectors for

$$\bigcup_{j=u}^{m_k} I_j^1, \qquad 2 \leq u \leq m_k. \tag{2.40}$$

No other rows of M_n can be in the hyperplane, else there would exist a finer partition of J_1.

So, Lemma 2.7.2 implies $T_{\mathcal{S}^n}(\Pi_{\mathcal{S}^n}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{S}^n}(\theta^*))^\perp$ is the cone generated by $(-1, \ldots, -1)$, $(1, \ldots, 1)$, and the indicator vectors of the subsets (2.40), otherwise known as the cone of nondecreasing vectors that are piecewise constant on the blocks $I_1^1, \ldots, I_{m_1}^1$. Its statistical dimension is denoted by $\delta(\mathcal{S}_{|I_1^1|,\ldots,|I_{m_1}^1|})$. This concludes the proof in the case when $\Pi_{\mathcal{S}^n}(\theta^*)$ is constant.

We now turn to the general case where $\Pi_{\mathcal{S}^n}(\theta^*)$ is piecewise constant with values $\mu_1 < \cdots < \mu_K$ on $J_1, \ldots, J_K$ respectively. We claim

$$\mu_k = \frac{1}{|J_k|} \sum_{i \in J_k} \theta_i^*. \tag{2.41}$$

Since $\mathcal{S}^n$ is a cone, the projection satisfies $\langle \theta^* - \Pi_{\mathcal{S}^n}(\theta^*), x \rangle \leq 0$ for all $x \in \mathcal{S}^n$, with equality if $x = \Pi_{\mathcal{C}}(\theta^*)$ (e.g., [10, Sec. 1.6]). Letting $x_1, \ldots, x_{n+1}$ be the conic generators of $\mathcal{S}^n$ (the rows of M_n), we have $\Pi_{\mathcal{C}}(\theta^*) = \sum_{i=1}^{n+1} \alpha_i x_i$ for some coefficients $\alpha_i \geq 0$. Then,

$$0 = \langle \theta^* - \Pi_{\mathcal{S}^n}(\theta^*), \Pi_{\mathcal{C}}(\theta^*) \rangle = \sum_{i=1}^{n+1} \alpha_i \underbrace{\langle \theta^* - \Pi_{\mathcal{S}^n}(\theta^*), x_i \rangle}_{\leq 0},$$

which implies $\langle \theta^* - \Pi_{\mathcal{S}^n}(\theta^*), x_i \rangle = 0$ if $\alpha_i > 0$. Consequently, if $\Pi_{\mathcal{S}^n}(\theta^*)$ changes value from component $j-1$ to j, then $\sum_{i=j}^{n} [\theta_i^* - (\Pi_{\mathcal{S}^n}(\theta^*))_i] = 0$. Thus (2.41) holds.

By Proposition 3.1 of [10], the tangent cone is

$$T_{\mathcal{S}^n}(\Pi_{\mathcal{S}^n}(\theta^*)) = \mathcal{S}^{n_1} \times \cdots \times \mathcal{S}^{n_K},$$

which is generated by the rows of the block diagonal matrix

$$A \coloneqq \begin{bmatrix} M_{n_1} & & \\ & \ddots & \\ & & M_{n_K} \end{bmatrix}.$$

To find which rows of A are in the hyperplane $(\theta^* - \Pi_{\mathcal{S}^n}(\theta^*))^\perp$, we can treat each block M_{n_k} separately and repeat the above argument. Doing so shows that $T_{\mathcal{S}^n}(\Pi_{\mathcal{S}^n}(\theta^*)) \cap (\theta^* - \Pi_{\mathcal{S}^n}(\theta^*))^\perp$ is the cone of vectors that are piecewise constant on $(I_1^1, \ldots, I_{m_1}^1, \ldots, I_1^K, \ldots, I_{m_K}^K)$ and are increasing within each of the blocks $(J_1, \ldots, J_K)$. The statistical dimension of this cone is $\sum_{k=1}^{K} \delta(\mathcal{S}_{|I_1^k|, \ldots, |I_{m_k}^k|})$. □

2.8 Proof of Proposition 2.5.1

Let $r \coloneqq \|\theta^*\|$. By rotating the problem, we may without loss of generality assume $\theta^* = (r, 0, \ldots, 0)$.

Let $E \coloneqq \{Y \in B_{(r-1)/2}(\theta^*)\}$. Then we have $E \subseteq \{Y \notin \mathcal{C}\}$, so under the event E we have $\hat{\theta}(Y) = Y/\|Y\|$. Noting $\|Y\|^2 = \|\theta^* + \sigma Z\|^2 = r^2 + 2\sigma r Z_1 + \sigma^2 \|Z\|^2$, we have

$$\frac{1}{\sigma^2} \|\hat{\theta}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2 = \frac{1}{\sigma^2} \left(\frac{r + \sigma Z_1}{\sqrt{r^2 + 2\sigma r Z_1 + \sigma^2 \|Z\|^2}} - 1 \right)^2 + \frac{\sum_{i=2}^{n} Z_i^2}{r^2 + 2\sigma r Z_1 + \sigma^2 \|Z\|^2}.$$

The second term converges to $r^{-2}\sum_{i=2}^n Z_i^2$ as $\sigma \downarrow 0$. We show the first term vanishes as $\sigma \downarrow 0$. Defining $g(\sigma) := \|\theta^* + \sigma Z\|$, we have

$$\begin{aligned} g(\sigma) &= \sqrt{r^2 + 2\sigma r Z_1 + \sigma^2\|Z\|^2} \\ g'(\sigma) &= \frac{rZ_1 + \sigma\|Z\|^2}{g(\sigma)} \\ g''(\sigma) &= \frac{\|Z\|^2}{g(\sigma)} - \frac{(rZ_1 + \sigma\|Z\|^2)g'(\sigma)}{g(\sigma)^2} \end{aligned}$$

Moreover we have $g(0) = r$, $g'(0) = Z_1$, and $g''(0) = (\|Z\|^2 - Z_1^2)/r$. Then by L'Hôpital's rule,

$$\begin{aligned} &\lim_{\sigma\downarrow 0}\frac{1}{\sigma}\left(\frac{r+\sigma Z_1}{\sqrt{r^2+2\sigma r Z_1+\sigma^2\|Z\|^2}} - 1\right) \\ &= \lim_{\sigma\downarrow 0}\frac{r+\sigma Z_1 - g(\sigma)}{\sigma g(\sigma)} = \lim_{\sigma\downarrow 0}\frac{Z_1 - g'(\sigma)}{g(\sigma)+\sigma g'(\sigma)} = \frac{Z_1 - Z_1}{r+0} = 0. \end{aligned}$$

Note $\mathbf{1}_E \to 1$ almost surely as $\sigma \downarrow 0$. Thus, $\sigma^{-2}\|\hat{\theta}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_E \to r^{-2}\sum_{i=2}^n Z_i^2$ almost surely. By the upper bound (2.7) we may use the dominated convergence theorem to get

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\Big[\|\hat{\theta}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_E\Big] = \frac{1}{r^2}\sum_{i=2}^n \mathbb{E}Z_i^2 = \frac{n-1}{r^2}.$$

To conclude the proof of the first limit (2.32a), note that

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\Big[\|\hat{\theta}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2\mathbf{1}_{E^c}\Big] = 0,$$

which holds by the argument used in the proof of Theorem 2.3.1 (e.g., see the second term in (2.19)).

A similar proof holds for the second limit (2.32b). Let E and $g(\sigma)$ be the same as before. Then

$$\begin{aligned} &\frac{1}{\sigma^2}\Big(\|\hat{\theta}(Y) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2\Big) \\ &= \frac{1}{\sigma^2}\left(\frac{r+\sigma Z_1}{\sqrt{r^2+2\sigma r Z_1+\sigma^2\|Z\|^2}} - r\right)^2 + \frac{\sum_{i=2}^n Z_i^2}{r^2+2\sigma r Z_1+\sigma^2\|Z\|^2} - \frac{(r-1)^2}{\sigma^2} \\ &= \frac{1}{\sigma^2}\left[\left(\frac{r+\sigma Z_1}{g(\sigma)} - r\right)^2 - (r-1)^2\right] + \frac{\sum_{i=2}^n Z_i^2}{r^2+2\sigma r Z_1+\sigma^2\|Z\|^2}. \end{aligned}$$

Again, the second term tends to $r^{-2}\sum_{i=2}^n Z_i^2$ as $\sigma \downarrow 0$. To handle the first term we use L'Hôpital's rule again. Let

$$\begin{aligned}
h(\sigma) &:= \frac{r+\sigma Z_1}{g(\sigma)} - r \\
h'(\sigma) &= \frac{Z_1}{g(\sigma)} - \frac{(r+\sigma Z_1)g'(\sigma)}{g(\sigma)^2} \\
h''(\sigma) &= -\frac{Z_1 g'(\sigma)}{g(\sigma)^2} + 2\frac{(r+\sigma Z_1)g'(\sigma)^2}{g(\sigma)^3} - \frac{Z_1 g'(\sigma) + (r+\sigma Z_1)g''(\sigma)}{g(\sigma)^2}
\end{aligned}$$

Recalling the limits $g(0) = r$, $g'(0) = Z_1$, and $g''(0) = (\|Z\|^2 - Z_1^2)/r$, we have $h(\sigma) \to -(r-1)$, $h'(\sigma) \to 0$, and

$$h''(0) = -\frac{Z_1^2}{r^2} + 2\frac{rZ_1^2}{r^3} - \frac{Z_1^2 + \|Z\|^2 - Z_1^2}{r^2} = \frac{Z_1^2 - \|Z\|^2}{r^2}.$$

Then, L'Hôpital's rule allows us to compute the limit of the first term.

$$\begin{aligned}
&\lim_{\sigma\downarrow 0} \frac{1}{\sigma^2}\left[\left(\frac{r+\sigma Z_1}{g(\sigma)} - r\right)^2 - (r-1)^2\right] \\
&= \lim_{\sigma\downarrow 0} \frac{h(\sigma)^2 - (r-1)^2}{\sigma^2} = \lim_{\sigma\downarrow 0} \frac{h(\sigma)h'(\sigma)}{\sigma} = \lim_{\sigma\downarrow 0}(h'(\sigma)^2 + h(\sigma)h''(\sigma)) = \frac{(r-1)(\|Z\|^2 - Z_1^2)}{r^2}.
\end{aligned}$$

Combining terms yields

$$\frac{1}{\sigma^2}\Big(\|\hat\theta(Y) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2\Big)\mathbf{1}_E \to \frac{(r-1)(\|Z\|^2 - Z_1^2) + \sum_{i=2}^n Z_i^2}{r^2} = \frac{\sum_{i=2}^n Z_i^2}{r},$$

so again by dominated convergence with the upper bound (2.7), we have

$$\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\Big[\Big(\|\hat\theta(Y) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2\Big)\mathbf{1}_E\Big] \to \frac{n-1}{r}.$$

To conclude the proof of (2.32b), note that

$$\frac{1}{\sigma^2}\mathbb{E}_{\theta^*}\Big[\Big(\|\hat\theta(Y) - \theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*) - \theta^*\|^2\Big)\mathbf{1}_{E^c}\Big] \to 0,$$

which was proved in the proof of Theorem 2.3.1 (see (2.25)).

2.9 Proofs for Section 2.5.2

The following lemma shows that the left-hand side of (2.34) is nonnegative.

Lemma 2.9.1. *For any $\theta_0 \in \mathcal{C}$,*

$$\|\Pi_{F_{\mathcal{C}}(\theta_0)}(x)\|^2 \geq \|\Pi_{K_{\mathcal{C}}}(x)\|^2.$$

Proof of Lemma 2.9.1. Because $K_{\mathcal{C}}$ is a cone, we have $\langle x, \Pi_{K_{\mathcal{C}}}(x)\rangle = \|\Pi_{K_{\mathcal{C}}}(x)\|^2$. Since $K_{\mathcal{C}} \subseteq F_{\mathcal{C}}(\theta_0)$, the optimality condition for $\Pi_{F_{\mathcal{C}}(\theta_0)}(x)$ implies $\langle x - \Pi_{F_{\mathcal{C}}(\theta_0)}(x), \Pi_{K_{\mathcal{C}}}(x)\rangle \leq 0$ and thus

$$\|\Pi_{K_{\mathcal{C}}}(x)\|^2 \leq \langle \Pi_{F_{\mathcal{C}}(\theta_0)}(x), \Pi_{K_{\mathcal{C}}}(x)\rangle \leq \|\Pi_{F_{\mathcal{C}}(\theta_0)}(x)\|\|\Pi_{K_{\mathcal{C}}}(x)\|.$$

Thus $\|\Pi_{F_{\mathcal{C}}(\theta_0)}(x)\| \geq \|\Pi_{K_{\mathcal{C}}}(x)\|$ and $M_{\theta_0} \geq 0$. □

Proof of Lemma 2.5.2. We first prove the equalities (i) and (ii).

(i) Let $v \in \{u : \mathbb{R}_+ u \subseteq F_{\mathcal{C}}(\theta_0)\}$ and let $\theta \in \mathcal{C}$. For any $c > 0$ we have $\theta_0 + cv \in \mathcal{C}$, and convexity implies $\theta + \alpha(\theta_0 + cv - \theta) \in \mathcal{C}$ for all $\alpha \in [0, 1]$. For large c we have $\|\theta_0 + cv - \theta\| > 1$ and thus $\theta + \frac{\theta_0 + cv - \theta}{\|\theta_0 + cv - \theta\|} \in \mathcal{C}$. Taking $c \to \infty$ and using the fact that $\mathcal{C}$ is closed yields $\theta + \frac{v}{\|v\|} \in \mathcal{C}$ and thus $v \in T_{\mathcal{C}}(\theta)$. Since θ was arbitrary, we have $v \in K_{\mathcal{C}}$.

Conversely, suppose $v \in K_{\mathcal{C}}$. Let $c^* \coloneqq \sup\{c > 0 : \theta_0 + cv \in \mathcal{C}\}$. The supremum is over a nonempty set because $v \in T_{\mathcal{C}}(\theta_0)$. Suppose for sake of contradiction that $c^* < \infty$. Since $\mathcal{C}$ is closed, $\theta_0 + c^* v \in \mathcal{C}$. Thus $v \in T_{\mathcal{C}}(\theta_0 + c^* v)$ which implies $\theta_0 + (c^* + \alpha)v \in \mathcal{C}$ for some $\alpha > 0$, contradicting the definition of c^*. Thus $c^* = \infty$ and $\theta_0 + cv \in \mathcal{C}$ for all $c > 0$.

(ii) Both sides can be expressed as the set of $v \in \mathbb{R}^n$ satisfying $\theta_0 + \sigma v \in \mathcal{C}$ for all $\sigma > 0$.

We now prove the second part of the lemma. The definition (2.33) implies $K_{\mathcal{C}} \subseteq T_{\mathcal{C}}(\theta)$ for any $\theta \in \mathcal{C}$.

Now, assume $F_{\mathcal{C}}(\theta_0)$ is a cone. If the reverse inclusion $T_{\mathcal{C}}(\theta) \subseteq F_{\mathcal{C}}(\theta)$ holds, then $\theta_0 - \theta \in T_{\mathcal{C}}(\theta) = F_{\mathcal{C}}(\theta)$ so $\theta_0 - (\theta - \theta_0) \in \mathcal{C}$. Conversely, suppose $\theta_0 - (\theta - \theta_0) \in \mathcal{C}$. If $v \in T_{\mathcal{C}}(\theta)$, then $\theta + cv \in \mathcal{C}$ for some $c > 0$. By convexity, $\theta_0 + cv/2 \in \mathcal{C}$, so $v \in F_{\mathcal{C}}(\theta_0)$. Thus $T_{\mathcal{C}}(\theta) \subseteq F_{\mathcal{C}}(\theta)$. □

Proof of Proposition 2.5.3. We use Y instead of $\theta^* + \sigma Z$ throughout the proof, but note that Y depends on σ.

Without loss of generality we can translate the problem so that $\Pi_{\mathcal{C}}(\theta^*) = 0$.

In view of (2.13), we may use the dominated convergence theorem on $\sigma^{-2}\|\Pi_{\mathcal{C}}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2$, so

$$\begin{aligned}
&\lim_{\sigma\to\infty} \frac{1}{\sigma^2}\mathbb{E}\|\Pi_{\mathcal{C}}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2 \\
&= \mathbb{E} \lim_{\sigma\to\infty} \frac{1}{\sigma^2}\|\Pi_{\mathcal{C}}(Y) - \Pi_{\mathcal{C}}(\theta^*)\|^2 && \text{dom. conv. with } \mathbb{E}\|Z\|^2 \\
&= \mathbb{E} \lim_{\sigma\to\infty} \frac{1}{\sigma^2}\|\Pi_{\mathcal{C}}(Y)\|^2 \\
&\overset{(i)}{=} \mathbb{E}\|\Pi_{K_{\mathcal{C}}}(Z)\|^2 = \delta(K_{\mathcal{C}}),
\end{aligned}$$

where we verify the equality (i) below.

Similarly, (2.13) allows us to use the dominated convergence theorem again for the excess risk.

$$\begin{aligned}
&\lim_{\sigma\to\infty} \frac{1}{\sigma^2}\big(\mathbb{E}\|\Pi_{\mathcal{C}}(Y)-\theta^*\|^2 - \|\Pi_{\mathcal{C}}(\theta^*)-\theta^*\|^2\big) \\
&= \mathbb{E}\lim_{\sigma\to\infty} \frac{1}{\sigma^2}\big(\|\Pi_{\mathcal{C}}(Y)-\theta^*\|^2 - \|\theta^*\|\big) && \text{dom. conv. with } \mathbb{E}\|Z\|^2 \\
&= \mathbb{E}\lim_{\sigma\to\infty} \frac{1}{\sigma^2}\big(\|\Pi_{\mathcal{C}}(Y)\|^2 - 2\langle \Pi_{\mathcal{C}}(Y),\theta^*\rangle\big) \\
&\overset{(ii)}{=} \mathbb{E}\|\Pi_{K_{\mathcal{C}}}(Z)\|^2 = \delta(K_{\mathcal{C}}).
\end{aligned}$$

It remains to verify (i) and (ii).

(i)

$$\begin{aligned}
&\left|\frac{1}{\sigma^2}\|\Pi_{\mathcal{C}}(Y)\|^2 - \|\Pi_{K_{\mathcal{C}}}(Z)\|^2\right| \\
&\le \frac{1}{\sigma^2}\big|\|\Pi_{\mathcal{C}}(Y)\|^2 - \|\Pi_{K_{\mathcal{C}}}(Y)\|^2\big| \\
&\quad + \left|\frac{1}{\sigma^2}\|\Pi_{K_{\mathcal{C}}}(Y)\|^2 - \|\Pi_{K_{\mathcal{C}}}(Z)\|^2\right| \\
&\le \frac{c}{\sigma^2} + \big|\|\Pi_{K_{\mathcal{C}}}(\theta^*/\sigma+Z)\|^2 - \|\Pi_{K_{\mathcal{C}}}(Z)\|^2\big| && \text{Lemma 2.9.1; } K_{\mathcal{C}} \text{ is a cone} \\
&\overset{\sigma\to\infty}{\longrightarrow} 0. && x\mapsto \|\Pi_{K_{\mathcal{C}}}(x)\|^2 \text{ is continuous}
\end{aligned}$$

(ii) We already showed $\|\Pi_{\mathcal{C}}(Y)\|^2/\sigma^2 \to \|\Pi_{K_{\mathcal{C}}}(Z)\|^2$, so it suffices to show the cross term vanishes. Indeed, we have $\|\Pi_{\mathcal{C}}(Y)\|/\sigma \to \|\Pi_{K_{\mathcal{C}}}(Z)\|$, so

$$\frac{1}{\sigma^2}|\langle \Pi_{\mathcal{C}}(Y),\theta^*\rangle| \le \frac{1}{\sigma^2}\|\Pi_{\mathcal{C}}(Y)\|\|\theta^*\| \overset{\sigma\to\infty}{\longrightarrow} 0.$$

□

Proof of Corollary 2.5.4. We begin with the first claim. Since $\mathcal{C} = \mathbb{R}^n_+$ is a cone, we have $K_{\mathcal{C}} = \mathbb{R}^n_+$. Provided we verify (2.34), the result follows from Proposition 2.5.3. Let $\theta := \Pi_{\mathcal{C}}(\theta^*)$ and fix $x \in \mathbb{R}^n$. Then some casework yields

$$\|\Pi_{F_{\mathcal{C}}(\theta)}(x)\|^2 - \|\Pi_{K_{\mathcal{C}}}(x)\|^2 = \sum_{i=1}^n \max\{x_i,-\theta_i\}^2 - \sum_{i=1}^n \max\{x_i,0\}^2 \le \sum_{i=1}^n \theta_i^2 = \|\theta\|^2 =: c.$$

We now turn to the second claim. If $\mathcal{C}$ is bounded, then by Lemma 2.5.2, $K_{\mathcal{C}} = \{u : \mathbb{R}_+ u \subseteq F_{\mathcal{C}}(\theta_0)\} = \{0\}$ for any $\theta_0 \in \mathcal{C}$.

Conversely, suppose $\mathcal{C}$ is unbounded and fix $\theta_0 \in \mathcal{C}$. Let

$$U_r := \{v \in S^{n-1} : \theta_0 + cv \notin \mathcal{C} \text{ for some } c \in (0, r)\}.$$

This set is open: if (v_n) is a sequence in U_r^c converging to v, then the fact that $\mathcal{C}$ is closed implies $\theta_0 + rv_n \in \mathcal{C}$ for all n, and consequently $\theta_0 + rv \in \mathcal{C}$ and finally $v \in U_r^c$.

If $\bigcup_{r>0} U_r$ is an open cover of the compact set S^{n-1}, then $S^{n-1} \subseteq U_r$ for some $r > 0$, which implies $\mathcal{C} \subseteq B_r(\theta_0)$, a contradiction. Thus, some direction $v \in S^{n-1}$ does not lie in $\bigcup_{r>0} U_r$, i.e., $\theta_0 + cv \in \mathcal{C}$ for all $c \geq 0$. This implies $cv \in K_{\mathcal{C}}$ for all $c \geq 0$.

We now apply Proposition 2.5.3. If $\mathcal{C}$ is bounded, then so is $F_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$. Choosing c large enough so that $F_{\mathcal{C}}(\Pi_{\mathcal{C}}(\theta^*))$ lies in the ball of radius c suffices to satisfy (2.34). Then Proposition 2.5.3 implies that the high σ limits are $\delta(K_{\mathcal{C}}) = 0$. □

Chapter 3

Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy-Krause variation

3.1 Introduction

Consider the problem of nonparametric regression where the goal is to estimate an unknown regression function $f^* : [0,1]^d \to \mathbb{R}$ ($d \geq 1$) from noisy observations at fixed design points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in [0,1]^d$. Specifically, we observe responses $y_1, \ldots, y_n$ drawn according to the model

$$y_i = f^*(\mathbf{x}_i) + \xi_i, \qquad \text{where} \quad \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \quad \text{for } i = 1, \ldots, n, \tag{3.1}$$

$\sigma^2 > 0$ is unknown, and the purpose is to nonparametrically estimate f^* known to belong to a prespecified function class. In the univariate ($d = 1$) case, two such important function classes are: (i) the class of *monotone nondecreasing* functions in which case f^* is usually estimated by the isotonic least squares estimator (LSE) (see e.g., Robertson et al. [73], Groeneboom and Jongbloed [41], Barlow et al. [7], Brunk [13], Ayer et al. [5]); and (ii) the class of functions whose *total variation* is bounded by a specific constant in which case it is natural to estimate f^* by total variation denoising (see e.g., Rudin et al. [74], Mammen and van de Geer [60], Chambolle et al. [16], Condat [24]). Both these estimators—isotonic regression and total variation denoising—have a long history and are very well-studied. For example, it is known that both these estimators produce piecewise constant fits and have finite sample risk (under the squared error loss) bounded from above by a constant multiple of $n^{-2/3}$ (see e.g., Meyer and Woodroofe [63], Zhang [97], Mammen and van de Geer [60]). Moreover, it is well-known that both these estimators are especially useful in fitting piecewise constant functions where their risk is almost parametric (at most $1/n$ up to logarithmic factors); see e.g., Guntuboyina and Sen [45], Dalalyan et al. [26], and Guntuboyina et al.

[44] and the references therein.

In this chapter, we try to answer the following question: "What is a natural generalization of univariate isotonic regression and univariate total variation denoising to multiple dimensions?" To answer this question we introduce and study two (constrained) LSEs for estimating $f^* : [0,1]^d \to \mathbb{R}$ where $d \geq 1$. We show that both these LSEs yield rectangular piecewise constant fits and have finite sample risk that is bounded from above by $n^{-2/3}$ (modulo logarithmic factors depending on d), thereby avoiding the curse of dimensionality to some extent. Further, we study the characterization and computation of these two estimators: the LSEs are obtained as solutions to convex optimization problems—in fact, quadratic programs with linear constraints—and are thus easily computable. Moreover, as in the case $d = 1$, we illustrate that these LSEs are particularly useful in fitting rectangular piecewise constant functions and can have almost parametric risk (up to logarithmic factors). These results are directly analogous to the univariate results mentioned in the previous paragraph and thus justify our claim that our proposed estimators are natural multivariate generalizations of univariate isotonic regression and univariate total variation denoising.

Our first estimator is the LSE over $\mathcal{F}^d_{\mathrm{EM}}$, the class of *entirely monotone* functions on $[0,1]^d$:

$$\widehat{f}_{\mathrm{EM}} \in \operatorname*{argmin}_{f \in \mathcal{F}^d_{\mathrm{EM}}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2. \tag{3.2}$$

The class $\mathcal{F}^d_{\mathrm{EM}}$ of entirely monotone functions is formally defined in Section 3.2. Entire monotonicity is an existing generalization in multivariate analysis of the univariate notion of monotonicity (see e.g., [1, 56, 95, 51]). Indeed, in the univariate case when $d = 1$, the class $\mathcal{F}^1_{\mathrm{EM}}$ is precisely the class of nondecreasing functions on $[0,1]$ and thus, for $d = 1$, the estimator (3.2) reduces to the usual isotonic LSE. For $d = 2$, the class $\mathcal{F}^2_{\mathrm{EM}}$ consists of all functions $f : [0,1]^2 \to \mathbb{R}$ which satisfy both $f(a_1, a_2) \leq f(b_1, b_2)$ and

$$f(b_1, b_2) - f(a_1, b_2) - f(b_1, a_2) + f(a_1, a_2) \geq 0, \tag{3.3}$$

for every $0 \leq a_1 \leq b_1 \leq 1$ and $0 \leq a_2 \leq b_2 \leq 1$. The formal definition of $\mathcal{F}^d_{\mathrm{EM}}$ for general $d \geq 1$ is given in Section 3.2. We remark that in general, entire monotonicity is different from the usual notion of monotonicity in classical multivariate isotonic regression [73]; see Lemma 3.2.2 for a connection between these two notions. We also remark that $\mathcal{F}^d_{\mathrm{EM}}$ is closed under translation and nonnegative scaling; that is, if $f \in \mathcal{F}^d_{\mathrm{EM}}$, then $af + b \in \mathcal{F}^d_{\mathrm{EM}}$ for any $a \geq 0$ and $b \in \mathbb{R}$. Additionally, the collection of right-continuous functions in $\mathcal{F}^d_{\mathrm{EM}}$ is precisely the collection of cumulative distribution functions of nonnegative measures on $[0,1]^d$ (see Lemma 3.2.3).

Our terminology of entire monotonicity is taken from Young and Young [95]. As a word of caution, we note that some authors (e.g., Aistleitner and Dick [1]) use the term "completely monotone" in place of "entirely monotone." We use the latter terminology because "completely monotone" has been used in the literature for other notions (see e.g., [92, 37, 35]) which are unrelated to our definition of entire monotonicity. Entire monotonicity

has also been referred by other names in the literature (for example, it has been referred to as "quasi-monotone" in Hobson [51]).

The second main estimator that we study in this chapter involves $V_{\mathrm{HK}\mathbf{0}}(\cdot)$, the *variation in the sense of Hardy and Krause (anchored at* $\mathbf{0}$*)*, which we shorten to *Hardy-Krause variation* or *HK*$\mathbf{0}$ *variation*. The HK$\mathbf{0}$ variation of a univariate function $f : [0,1] \to \mathbb{R}$ is simply the total variation of the function, i.e.,

$$V_{\mathrm{HK}\mathbf{0}}(f) = \sup_{0=x_0<x_1<\cdots<x_k=1} \sum_{i=0}^{k-1} |f(x_{i+1}) - f(x_i)|, \tag{3.4}$$

where the supremum is over all $k \geq 1$ and all partitions $0 = x_0 < x_1 < \cdots < x_k = 1$ of $[0,1]$. Thus HK$\mathbf{0}$ variation is a generalization of one-dimensional total variation to multiple dimensions. For $d = 2$, HK$\mathbf{0}$ variation is defined in the following way: for $f : [0,1]^2 \to \mathbb{R}$,

$$\begin{aligned} V_{\mathrm{HK}\mathbf{0}}(f) \coloneqq\ & V_{\mathrm{HK}\mathbf{0}}(x \mapsto f(x,0)) + V_{\mathrm{HK}\mathbf{0}}(x \mapsto f(0,x)) \\ & + \sup \sum_{0 \leq l_1 < k_1, 0 \leq l_2 < k_2} \Big| f(x^{(1)}_{l_1+1}, x^{(2)}_{l_2+1}) - f(x^{(1)}_{l_1}, x^{(2)}_{l_2+1}) \\ & \qquad - f(x^{(1)}_{l_1+1}, x^{(2)}_{l_2}) + f(x^{(1)}_{l_1}, x^{(2)}_{l_2}) \Big| \end{aligned} \tag{3.5}$$

where the first two terms in the right hand side above are defined via the univariate definition (3.4) and the supremum in the third term above is over all pairs of partitions $0 = x^{(1)}_0 < x^{(1)}_1 < \cdots < x^{(1)}_{k_1} = 1$ and $0 = x^{(2)}_0 < x^{(2)}_1 < \cdots < x^{(2)}_{k_2} = 1$ of $[0,1]$. Note that a special role is played in the first two terms of the right hand side of (3.5) by the point $(0,0)$ and this is the reason for the phrase "anchored at $\mathbf{0}$". For smooth functions $f : [0,1]^2 \to \mathbb{R}$, it can be shown that

$$V_{\mathrm{HK}\mathbf{0}}(f) = \int_0^1 \int_0^1 \left| \frac{\partial^2 f}{\partial x_1 \partial x_2} \right| dx_1 dx_2 + \int_0^1 \left| \frac{\partial f(\cdot, 0)}{\partial x_1} \right| dx_1 + \int_0^1 \left| \frac{\partial f(0, \cdot)}{\partial x_2} \right| dx_2$$

and, from the first term in the right hand side above, it is clear that the HK$\mathbf{0}$ variation is related to the L^1 norm of the mixed derivative. The definition of HK$\mathbf{0}$ variation for general $d \geq 1$ is given in Section 3.2. HK$\mathbf{0}$ variation is quite different from the usual definition of multivariate total variation (see e.g., Ziemer [99, Chapter 5]) as explained briefly in Section 3.2.

Functions that are piecewise constant on axis-aligned rectangular pieces (see Definition 3.2.5) have finite HK$\mathbf{0}$ variation as explained in Section 3.2. More generally, the collection of right-continuous functions of finite HK$\mathbf{0}$ variation is precisely the same as the collection of cumulative distribution functions of finite signed measures (see Lemma 3.2.8). An example of a function with infinite HK$\mathbf{0}$ variation is the indicator function of an open d-dimensional ball contained in $[0,1]^d$ (see [69, Sec. 12]).

Our second estimator is the constrained LSE over functions with HK**0** variation bounded by some tuning parameter $V > 0$:

$$\widehat{f}_{\mathrm{HK}\mathbf{0},V} \in \operatorname*{argmin}_{f:V_{\mathrm{HK}\mathbf{0}}(f)\leq V} \frac{1}{n}\sum_{i=1}^{n}(y_i - f(\mathbf{x}_i))^2. \tag{3.6}$$

This estimator is a generalization of total variation denoising to $d \geq 2$ because in the case $d = 1$, HK**0** variation coincides with total variation and, thus, the above estimator performs univariate total variation denoising, sometimes also called trend filtering of first order [74, 60, 16, 24, 54, 81]. This generalization is different from the usual multivariate total variation denoising as in Rudin et al. [74] (see Section 3.5 for more discussion on how $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ is different from the multivariate total variation regularized estimator). It is also possible to define the HK**0** variation estimator in the following penalized form:

$$\widehat{f}_{\mathrm{HK}\mathbf{0},\lambda} \in \operatorname*{argmin}_{f} \frac{1}{n}\left\{\sum_{i=1}^{n}(y_i - f(\mathbf{x}_i))^2 + \lambda V_{\mathrm{HK}\mathbf{0}}(f)\right\} \tag{3.7}$$

for a tuning parameter $\lambda > 0$. In this chapter, we shall focus on the constrained form in (3.6) although analogues of our results for the penalized estimator (3.7) can also be proved.

Before proceeding further, let us note that entire monotonicity is related to HK**0** variation in much the same way as univariate monotonicity is related to univariate total variation. Indeed, for functions in one variable, the following two properties are well-known:

1. Every function $f : [0, 1] \to \mathbb{R}$ of bounded variation can be written as the difference of two monotone functions $f = f_+ - f_-$ and the total variation of f equals the sum of the variations of f_+ and f_-.

2. If $f : [0, 1] \to \mathbb{R}$ is nondecreasing, then its total variation on $[0, 1]$ is simply $f(1) - f(0)$.

These two facts generalize almost verbatim to entire monotonicity and HK**0** variation (see Lemma 3.2.7). Thus, in some sense, entire monotonicity is to Hardy-Krause variation as monotonicity is to total variation.

Although the terminology of "entire monotonicity" does not seem to have been used previously in the statistics literature, entirely monotone functions are closely related to cumulative distribution functions of nonnegative measures which appear routinely in statistics. HK**0** variation has appeared previously in statistics in the literature on quasi-Monte Carlo (see e.g., [69, 46]) as well as in the power analysis of certain sequential detection problems (see e.g., [72]). Additionally Benkeser and Van Der Laan [11] (see also [84, 83, 86, 85]) considered the class $\{f : V_{\mathrm{HK}\mathbf{0}}(f) \leq V\}$ in their "highly adaptive LASSO" estimator and exploited its connections to the LASSO in a setting that is different from our classical nonparametric regression framework. They also used the terminology of "sectional variation norm" to refer to the Hardy-Krause variation (see also [38, Section 2]). An estimator very similar to (3.6) was proposed by Mammen and van de Geer [60] for $d = 2$ when the design points take values

in a uniformly spaced grid (this estimator of [60] is described in Section 3.3.1). Also, Lin [58] proposed an estimator in the context of the Gaussian white noise model that bears some similarities to (3.6) (this connection is detailed in Section 2.5).

The goal of this chapter is to analyze the properties of the estimators (3.2) and (3.6). Here is a description of our main results. Section 3.3 concerns the computation of these estimators. Note that, as stated, the optimization problems defining our estimators (3.2) and (3.6) are convex (albeit infinite-dimensional). We show that, given arbitrary data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, the two estimators (3.2) and (3.6) can be computed by solving a nonnegative least squares (NNLS) problem and a LASSO problem respectively, with a suitable design matrix that only depends on the design-points $\mathbf{x}_1, \ldots, \mathbf{x}_n$. It is interesting to note that the design matrices in the two finite-dimensional problems for computing (3.2) and (3.6) are exactly the same. Our main results in this section (Proposition 3.3.1 and Proposition 3.3.3) imply that $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ can be taken to be of the form

$$\widehat{f}_{\mathrm{EM}} = \sum_{j=1}^{p} (\widehat{\beta}_{\mathrm{EM}})_j \cdot \mathbb{I}_{[\mathbf{z}_j, \mathbf{1}]} \quad \text{and} \quad \widehat{f}_{\mathrm{HK}\mathbf{0},V} = \sum_{j=1}^{p} (\widehat{\beta}_{\mathrm{HK}\mathbf{0},V})_j \cdot \mathbb{I}_{[\mathbf{z}_j, \mathbf{1}]} \tag{3.8}$$

for some $\mathbf{z}_1, \ldots, \mathbf{z}_p$ that only depend on the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and vectors $\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$ in $\mathbb{R}^p$ which are obtained by solving the NNLS problem (3.28) and the LASSO problem (3.30) respectively. Here $\mathbb{I}_{[\mathbf{z}_j, \mathbf{1}]}$ denotes the indicator of the rectangle $[\mathbf{z}_j, \mathbf{1}]$ (defined via (3.16)). Because NNLS and LASSO typically lead to sparse solutions, the vectors $\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$ will be sparse which clearly implies that $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ as given above (3.8) will be piecewise constant on axis-aligned rectangles. Therefore our estimators give rectangular piecewise constant fits to data and this generalizes the fact that univariate isotonic regression and total variation denoising yield piecewise constant fits. In the case when the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ form an equally spaced lattice in $[0, 1]^d$ (see the definition (3.34) for the precise formulation of this assumption), the points $\mathbf{z}_1, \ldots, \mathbf{z}_p$ can simply be taken to be $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and, in this case, more explicit expressions can be given for the estimators (see Section 3.3.1 for details). It should be noted that the lattice design is quite commonly used for theoretical studies in multidimensional nonparametric function estimation (see e.g., [64]) especially in connection with image analysis (see e.g., [16, 25]).

We also investigate the accuracy properties of $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ via the study of their risk behavior under the standard fixed design squared error loss function. Specifically, we define the risk of an estimator $\widehat{f}$ by

$$\mathcal{R}(\widehat{f}, f^*) \coloneqq \mathbb{E}\mathcal{L}(\widehat{f}, f^*) \qquad \text{where} \qquad \mathcal{L}(\widehat{f}, f^*) \coloneqq \frac{1}{n} \sum_{i=1}^{n} (\widehat{f}(\mathbf{x}_i) - f^*(\mathbf{x}_i))^2. \tag{3.9}$$

We prove results on the risk of $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ in the case of the aforementioned lattice design. In this setting, our main results are described below.

We analyze the risk of $\widehat{f}_{\mathrm{EM}}$ under the (well-specified) assumption that $f^* \in \mathcal{F}^d_{\mathrm{EM}}$. We prove in Theorem 3.4.1 that, for $n \geq 1$,

$$\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) \leq \frac{C(d, \sigma, V^*)}{n^{2/3}} (\log(en))^{\frac{2d-1}{3}} \tag{3.10}$$

where

$$V^* := f^*(1, \ldots, 1) - f^*(0, \ldots, 0)$$

and $C(d, \sigma, V^*)$ depends only on d, σ and V^* (see statement of Theorem 3.4.1 for the explicit form of $C(d, \sigma, V^*)$). Note that the dimension d appears in (3.10) only through the logarithmic term which means that we obtain "dimension independent rates" ignoring logarithmic factors. Some intuition for why the constraint of entire monotononicity is able to mitigate the usual curse of dimensionality is provided in Section 3.5. Other nonparametric estimators exhibiting such dimension independent rates can be found in [8, 58, 22, 65, 76, 89]. In Theorem 3.4.3, we prove a minimax lower bound which implies that the dependence on d through the logarithmic term in (3.10) cannot be avoided for any estimator.

We also prove in Theorem 3.4.5 that $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*)$ is smaller than the bound given by (3.10) when $f^* \in \mathcal{F}^d_{\mathrm{EM}}$ is *rectangular piecewise constant*. Loosely speaking, we say that $f : [0,1]^d \to \mathbb{R}$ is rectangular piecewise constant if it is constant on each set in a partition of $[0,1]^d$ into axis-aligned rectangles and the smallest cardinality of such a partition shall be denoted by $k(f)$ (see Definition 3.2.5 for the precise definitions). In Theorem 3.4.5, we prove that whenever $f^* \in \mathcal{F}^d_{\mathrm{EM}}$ is rectangular piecewise constant, we have

$$\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) \leq C_d \sigma^2 \frac{k(f^*)}{n} (\log(en))^{\frac{3d}{2}} (\log(e \log(en)))^{\frac{2d-1}{2}} \tag{3.11}$$

for a positive constant C_d which only depends on d. Note that when $k(f^*)$ is not too large, the right hand side of (3.11) converges to zero as $n \to \infty$ at a faster rate compared to the right hand side of (3.10). Thus rectangular piecewise constant functions which also satisfy the constraint of entire monotonicity are estimated at nearly the parametric rate (ignoring the logarithmic factor) by the LSE $\widehat{f}_{\mathrm{EM}}$.

Let us now describe our results for the other estimator $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$. In Theorem 3.4.6 we prove that when $V_{\mathrm{HK}\mathbf{0}}(f^*) \leq V$ (note that V is the tuning parameter in the definition of $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$), then

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) \leq \frac{C(d, \sigma, V)}{n^{2/3}} (\log(en))^{\frac{2d-1}{3}}. \tag{3.12}$$

Note that the right sides of the bounds (3.12) and (3.10) are the same and thus the estimator $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ also achieves dimension independent rates (ignoring logarithmic factors) (see Section 3.5 for an explanation of this phenomenon). We also prove a minimax lower bound in Theorem 3.4.9 which implies that the dependence on d in the logarithmic term in (3.12) cannot be completely removed for any estimator.

In univariate total variation denoising, it is known that one obtains faster rates than given by the bound (3.12) when $f^* : [0,1] \to \mathbb{R}$ is piecewise constant with not too many pieces. Indeed if f^* is piecewise constant for $d = 1$ with $k(f^*)$ pieces, then it has been proved that

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) \leq C(c)\sigma^2 \frac{k(f^*)}{n} \log(en) \tag{3.13}$$

provided $V = V_{\mathrm{HK}\mathbf{0}}(f^*)$ and f^* satisfies a minimum length condition in that each constant piece has length at least $c/k(f^*)$ (the multiplicative term $C(c)$ in (3.13) only depends on this c appearing in the minimum length condition). A proof of this result can be found in [44, Corollary 2.3] and, for other similar results, see [57, 26, 66, 98]. In light of this univariate result, it is plausible to expect a bound similar to (3.11) for $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ when f^* is an axis-aligned rectangular piecewise constant function provided that the tuning parameter V is taken to be equal to $V_{\mathrm{HK}\mathbf{0}}(f^*)$ and provided that f^* satisfies a minimum length condition. We prove such a result for a class of simple rectangular piecewise constant functions $f^* : [0,1]^d \to \mathbb{R}$ of the form

$$f^*(\cdot) = a_1 \mathbb{I}_{[\mathbf{x}^*, \mathbf{1}]}(\cdot) + a_0 \tag{3.14}$$

for some $a_1, a_0 \in \mathbb{R}$ and $\mathbf{x}^* \in [0,1]^d$ (here $\mathbb{I}$ stands for the indicator function). It is easy to see that (3.14) represents a rectangular piecewise constant function with $k(f^*) \leq 2^d$. In Theorem 3.4.10, we prove that when f^* is of the above form (3.14), then

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) \leq C(c,d) \frac{\sigma^2}{n} (\log(en))^{\frac{3d}{2}} (\log(e \log(en)))^{\frac{2d-1}{2}} \tag{3.15}$$

provided the tuning parameter V equals $V_{\mathrm{HK}\mathbf{0}}(f^*)$ and $\mathbf{x}^* \in [0,1]^d$ satisfies a *minimum size condition* (3.50). This latter condition, which is analogous to the minimum length condition in the univariate case, involves a positive constant c and the constant $C(c,d)$ appearing in (3.15) only depends on c and the dimension d. In the specific case when $d = 2$, the minimum length condition (3.50) can be weakened, as discussed in Section A.1.

We are unable to prove versions of (3.15) for more general rectangular piecewise constant functions. However, some results in that direction have been proved in a very recent paper by Ortelli and van de Geer [66]. Their results are of a different flavor as they work with a similar but different estimator and a smaller loss function. Their proof techniques are also completely different from ours.

The rest of the chapter is organized as follows. The notions of entire monotonicity and Hardy-Krause variation are formally defined for arbitrary $d \geq 1$ in Section 3.2 where we also collect some of their relevant properties. In Section 3.3, we discuss the computational aspects for solving the optimization problems in (3.2) and (3.6). The risk results for $\widehat{f}_{\mathrm{EM}}$ are given in Section 3.4.1 while the risk bounds for $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ are in Section 3.4.2. We discuss the connections of our contributions with other related work in Section 3.5. The proofs for our risk results are given in Section A.3 while the proofs of the results in Section 3.2 and Section 3.3 are given in Section A.4.Additional technical results used in the proofs of Section A.3 are proved in Section A.5.The section also contains another risk bound for $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$

(Section A.1), as well a section for simulations (Section A.2) that contains some examples and depictions of the two estimators, including an application to estimation in the bivariate current status model.

3.2 Entire monotonicity and Hardy-Krause variation

The aim of this section is to provide formal definitions of entire monotonicity and HK**0** variation for the convenience of the reader. We roughly follow the notation of Aistleitner and Dick [1] and Owen [69].

Let us first introduce some basic notation that will be used throughout the chapter. We let $\mathbf{0} = (0, \ldots, 0)$ and $\mathbf{1} = (1, \ldots, 1)$. Given an integer m, we take $[m] \coloneqq \{1, \ldots, m\}$. For two points $\mathbf{a} = (a_1, \ldots, a_d)$ and $\mathbf{b} = (b_1, \ldots, b_d) \in [0,1]^d$, we write

$$\mathbf{a} \prec \mathbf{b} \quad \text{if and only if} \quad a_j < b_j \text{ for every } j = 1, \ldots, d$$

and

$$\mathbf{a} \preceq \mathbf{b} \quad \text{if and only if} \quad a_j \leq b_j \text{ for every } j = 1, \ldots, d.$$

When $\mathbf{a} \preceq \mathbf{b}$, we write

$$[\mathbf{a}, \mathbf{b}] \coloneqq \{\mathbf{x} : \mathbf{a} \preceq \mathbf{x} \preceq \mathbf{b}\} \coloneqq \prod_{j=1}^{d} [a_j, b_j], \tag{3.16}$$

$$[\mathbf{a}, \mathbf{b}) \coloneqq \{\mathbf{x} : \mathbf{a} \preceq \mathbf{x} \prec \mathbf{b}\} \coloneqq \prod_{j=1}^{d} [a_j, b_j).$$

Note that $[\mathbf{a}, \mathbf{b}]$ is a closed axis-aligned rectangle and it has nonempty interior when $\mathbf{a} \prec \mathbf{b}$.

Given a function $f : [0,1]^d \to \mathbb{R}$ and two distinct points $\mathbf{a} = (a_1, \ldots, a_d), \mathbf{b} = (b_1, \ldots, b_d) \in [0,1]^d$ with $\mathbf{a} \preceq \mathbf{b}$, we define the *quasi-volume* $\Delta(f; [\mathbf{a}, \mathbf{b}])$ by

$$\sum_{j_1=0}^{J_1} \cdots \sum_{j_d=0}^{J_d} (-1)^{j_1 + \cdots + j_d} f(b_1 + j_1(a_1 - b_1), \ldots, b_d + j_d(a_d - b_d)), \tag{3.17}$$

where $J_i \coloneqq \mathbb{I}\{a_i \neq b_i\}$ for each i. For example, when $d = 2$, it is easy to see that $\Delta(f; [\mathbf{a}, \mathbf{b}])$ equals

$$\begin{aligned} f(b_1, b_2) - f(b_1, a_2) - f(a_1, b_2) + f(a_1, a_2) \ \text{ if } \mathbf{a} \prec \mathbf{b} \\ f(b_1, b_2) - f(b_1, a_2) \ \text{ if } a_1 = b_1 \text{ and } a_2 < b_2 \\ f(b_1, b_2) - f(a_1, b_2) \ \text{ if } a_2 = b_2 \text{ and } a_1 < b_1. \end{aligned} \tag{3.18}$$

We are now ready to define entire monotonicity.

Definition 3.2.1 (Entire monotonicity)**.** We say that a function $f : [0,1]^d \to \mathbb{R}$ is *entirely monotone* if

$$\Delta(f; [\mathbf{a}, \mathbf{b}]) \geq 0 \qquad \text{for every } \mathbf{a} \neq \mathbf{b} \in [0,1]^d \text{ with } \mathbf{a} \preceq \mathbf{b}. \tag{3.19}$$

In words, for a entirely monotone function f, every quasi-volume $\Delta(f; [\mathbf{a}, \mathbf{b}])$ is nonnegative. The class of such functions will be denoted by $\mathcal{F}^d_{\mathrm{EM}}$. By (3.18), note that entire monotonicity is equivalent to (3.3) for $d = 2$.

A more common generalization of monotonicity to multiple dimensions is the class $\mathcal{F}^d_{\mathrm{M}}$ consisting of all functions $f : [0,1]^d \to \mathbb{R}$ satisfying

$$f(a_1, \ldots, a_d) \leq f(b_1, \ldots, b_d), \qquad \text{for } 0 \leq a_i \leq b_i \leq 1, \quad i = 1, \ldots, d. \tag{3.20}$$

As the following result shows (see Section A.4.1 for a proof), $\mathcal{F}^d_{\mathrm{EM}}$ is a strict subset of $\mathcal{F}^d_{\mathrm{M}}$ when $d \geq 2$ (e.g., when $d = 2$, functions in $\mathcal{F}^d_{\mathrm{EM}}$ need to additionally satisfy the second constraint in (3.3)) and thus the estimator (3.2) is distinct from the LSE over $\mathcal{F}^d_{\mathrm{M}}$ for $d \geq 2$. This latter estimator is the classical multivariate isotonic regression estimator [73].

Lemma 3.2.2. *When $d = 1$, entire monotonicity coincides with monotonicity, i.e., $\mathcal{F}^1_{\mathrm{EM}} = \mathcal{F}^1_{\mathrm{M}}$. For $d \geq 2$, we have $\mathcal{F}^d_{\mathrm{EM}} \subsetneq \mathcal{F}^d_{\mathrm{M}}$.*

It is well-known that entirely monotone functions are closely related to cumulative distribution functions of nonnegative measures. The following result taken from Aistleitner and Dick [1, Theorem 3] makes this connection precise.

Lemma 3.2.3 ([1, Theorem 3])**.** *1. For every nonnegative Borel measure ν on $[0,1]^d$, the function $f(\mathbf{x}) := \nu([\mathbf{0}, \mathbf{x}])$ belongs to $\mathcal{F}^d_{\mathrm{EM}}$.*

2. If $f \in \mathcal{F}^d_{\mathrm{EM}}$ is right-continuous, then there exists a unique nonnegative Borel measure ν on $[0,1]^d$ such that $f(\mathbf{x}) - f(\mathbf{0}) = \nu([\mathbf{0}, \mathbf{x}])$.

We shall now define the notion of HK**0** variation. The HK**0** variation is defined through another variation called the *Vitali variation*. Let us first define the Vitali variation of a function $f : [0,1]^d \to \mathbb{R}$. To do so, we need some notation. By a partition of the univariate interval $[0,1]$, we mean a set of points $0 = x_0 < x_1 < \cdots < x_k = 1$ for some $k \geq 1$. Given d such univariate partitions:

$$0 = x^{(s)}_0 < x^{(s)}_1 < \cdots < x^{(s)}_{k_s} = 1, \qquad \text{for } s = 1, \ldots, d, \tag{3.21}$$

we can define a collection $\mathcal{P}$ of subsets of $[0,1]^d$ consisting of all sets of the form $A_1 \times \cdots \times A_d$ where for each $1 \leq s \leq d$, $A_s = [x^{(s)}_{l_s}, x^{(s)}_{l_s+1}]$ for some $0 \leq l_s \leq k_s - 1$. Note that each set in $\mathcal{P}$ is an axis-aligned closed rectangle and the cardinality of $\mathcal{P}$ equals $k_1 \ldots k_d$. The rectangles in $\mathcal{P}$ are not disjoint but they form a *split* of $[0,1]^d$ in the sense of Owen [69, Definition 3] and we shall refer to $\mathcal{P}$ as the split generated by the d univariate partitions (3.21).

Definition 3.2.4 (Vitali variation)**.** The Vitali variation of a function $f : [0,1]^d \to \mathbb{R}$ is defined as

$$V^{(d)}(f; [0,1]^d) := \sup_{\mathcal{P}} \sum_{A \in \mathcal{P}} |\Delta(f; A)| \tag{3.22}$$

where $\Delta(f; A)$ is the quasi-volume defined in (3.17) and the supremum above is taken over all splits $\mathcal{P}$ that are generated by d univariate partitions in the manner described above.

The following observations about the Vitali variation will be useful for us. Note first that when $d = 1$, Vitali variation is simply total variation (3.4) since the rectangles in this case are intervals. The second fact is that when f is smooth (in the sense that the partial derivatives appearing below exist and are continuous on $[0,1]^d$), we have

$$V^{(d)}(f;[0,1]^d) = \int_0^1 \cdots \int_0^1 \left| \frac{\partial^d f}{\partial x_1 \ldots \partial x_d} \right| dx_1 \ldots dx_d. \tag{3.23}$$

The third observation is that $V^{(d)}(f;[0,1]^d)$ can be written out explicitly when f is a rectangular piecewise constant function. In order to state this result, let us formally define the notion of a rectangular piecewise constant function on $[0,1]^d$. Given d univariate partitions as in (3.21), let $\mathcal{P}^*$ denote the collection of all sets of the form $A_1 \times \cdots \times A_d$ where for each $1 \le s \le d$, A_s is either equal to $[x_{l_s}^{(s)}, x_{l_s+1}^{(s)})$ for some $0 \le l_s \le k_s - 1$ or the singleton $\{1\}$. Note that, unlike $\mathcal{P}$, the sets in $\mathcal{P}^*$ are disjoint and hence $\mathcal{P}^*$ forms a partition of $[0,1]^d$. We shall refer to $\mathcal{P}^*$ as the partition generated by the d univariate partitions (3.21).

Definition 3.2.5 (Rectangular piecewise constant function)**.** We say that $f : [0,1]^d \to \mathbb{R}$ is rectangular piecewise constant if there exists a partition $\mathcal{P}^*$ generated by d univariate partitions as described above such that f is constant on each set in $\mathcal{P}^*$. We use $\mathfrak{R}^d$ to denote the class of all rectangular piecewise constant functions on $[0,1]^d$. For $f \in \mathfrak{R}^d$, we define $k(f)$ as the smallest value of $k_1 \ldots k_d$ for which there exist d univariate partitions of lengths $k_1, \ldots, k_d$ such that f is constant on each of the sets in $\mathcal{P}^*$ generated by these d univariate partitions.

The following lemma (proved in Section A.4.2) provides a formula for the Vitali variation of a rectangular piecewise constant function f on $[0,1]^d$. Note that this lemma implies, in particular, that the Vitali variation of every rectangular piecewise constant function is finite.

Lemma 3.2.6. *Suppose f is rectangular piecewise constant on $[0,1]^d$ with respect to a partition $\mathcal{P}^*$ generated by d univariate partitions and let $\mathcal{P}$ denote the split generated by these univariate partitions. Then*

$$V^{(d)}(f;[0,1]^d) = \sum_{A \in \mathcal{P}} |\Delta(f;A)|.$$

Despite these interesting properties, the Vitali variation is not directly suitable for our purposes because there exist many non-constant functions f on $[0,1]^d$ (such as $f(x,y) := x$) whose Vitali variation is zero. This weakness of the Vitali variation is well-known (see e.g., Owen [69] or Aistleitner and Dick [1]) and motivates the following definition of the HK**0** variation.

Given a nonempty subset of indices $S \subseteq [d] = \{1, \ldots, d\}$, let

$$U_S := \{(u_1, \ldots, u_d) \in [0,1]^d : u_j = 0, j \notin S\}. \tag{3.24}$$

Note that U_S is a face of $[0,1]^d$ adjacent to $\mathbf{0}$. By ignoring the components not in S, the restriction of the function f on $[0,1]^d$ to the set U_S can be viewed as a function $\widetilde{f} : [0,1]^{|S|} \to \mathbb{R}$. The Vitali variation of $\widetilde{f}$ viewed as a function of $[0,1]^{|S|}$ will be denoted by

$$V^{(|S|)}(f; S; [0,1]^d) \coloneqq V^{(|S|)}(\widetilde{f}; [0,1]^{|S|}).$$

The *Hardy-Krause variation (anchored at* $\mathbf{0}$*)* of $f : [0,1]^d \to \mathbb{R}$ is defined by

$$V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) := \sum_{\varnothing \neq S \subseteq [d]} V^{(|S|)}(f; S; [0,1]^d). \tag{3.25}$$

That is, the HK$\mathbf{0}$ variation is the sum of the Vitali variations of f restricted to each face of $[0,1]^d$ adjacent to $\mathbf{0}$. Note the special role played by the point $\mathbf{0}$ in this definition and this is the reason for the phrase "anchored at $\mathbf{0}$". It is also common to anchor the HK variation at $\mathbf{1}$ (see e.g., Aistleitner and Dick [1]) but we focus only on $\mathbf{0}$ as the anchor in this chapter. Because of the addition of the lower-dimensional Vitali variations, it is clear that the HK$\mathbf{0}$ variation equals zero only for constant functions and this property is the reason why the HK$\mathbf{0}$ variation is usually preferred to the Vitali variation.

Let us now remark that the HK$\mathbf{0}$ variation is quite different from the usual notion of multivariate total variation. Indeed, when f is smooth, the multivariate total variation of f only involves the first order partial derivatives of f. On the other hand, as can be seen from (3.23), the HK$\mathbf{0}$ variation is defined in terms of higher order mixed partial derivatives of f.

An important property of the HK$\mathbf{0}$ variation is that it is finite for rectangular piecewise constant functions. This is basically a consequence of Lemma 3.2.6 and the fact that the restriction of a rectangular piecewise constant function to each set U_S in (3.24) is also rectangular piecewise constant.

The following lemma formally establishes the connection between entire monotonicity and HK$\mathbf{0}$ variation, as mentioned earlier in the Introduction.

Lemma 3.2.7. *The following properties hold:*

(i) If $f : [0,1]^d \to \mathbb{R}$ has finite HK$\mathbf{0}$ variation, then there exist unique $f_+, f_- \in \mathcal{F}^d_{\mathrm{EM}}$ such that $f_+(\mathbf{0}) = f_-(\mathbf{0}) = 0$ and

$$f(\mathbf{x}) - f(\mathbf{0}) = f_+(\mathbf{x}) - f_-(\mathbf{x}), \qquad \mathbf{x} \in [0,1]^d$$

and

$$V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) = V_{\mathrm{HK}\mathbf{0}}(f_+; [0,1]^d) + V_{\mathrm{HK}\mathbf{0}}(f_-; [0,1]^d).$$

(ii) If $f \in \mathcal{F}^d_{\mathrm{EM}}$, then

$$V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) = f(\mathbf{1}) - f(\mathbf{0}).$$

The first fact in the above lemma is quite standard (see e.g., [1, Theorem 2]). We could not find an exact reference for the second fact so we included a proof in Section A.4.3).

Finally, let us mention that it is well-known that a result analogous to Lemma 3.2.3 holds for the connection between functions with finite HK**0** variation and cumulative distribution functions for signed measures. This result is stated next.

Lemma 3.2.8 ([1, Theorem 3])**.** *1. For every signed Borel measure ν on $[0,1]^d$, the function $f(\mathbf{x}) := \nu([\mathbf{0}, \mathbf{x}])$ has finite HK**0** variation.*

*2. If f has finite HK**0** variation and is right-continuous, then there exists a unique finite signed Borel measure ν on $[0,1]^d$ such that $f(\mathbf{x}) = \nu([\mathbf{0}, \mathbf{x}])$.*

3.3 Computational feasibility

The goal of this section is to describe procedures for computing the two estimators (3.2) and (3.6). We shall specifically show that the estimators (3.2) and (3.6) can be computed by solving a NNLS problem and a LASSO problem respectively, with a suitable design matrix that is the same for both the problems and that depends only on $\mathbf{x}_1, \ldots, \mathbf{x}_n$. This design matrix will be the matrix $\mathbf{A}$ whose columns are the distinct elements of the finite set

$$\mathcal{Q} \equiv \mathcal{Q}_{\mathbf{x}_1,\ldots,\mathbf{x}_n} := \{\mathbf{v}(\mathbf{z}) : \mathbf{z} \in [0,1]^d\} \subseteq \{0,1\}^n, \tag{3.26}$$

where

$$\mathbf{v}(\mathbf{z}) \equiv \mathbf{v}_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(\mathbf{z}) := (\mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{x}_1), \mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{x}_2), \ldots, \mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{x}_n)). \tag{3.27}$$

We assume without loss of generality that the first column of $\mathbf{A}$ is $\mathbf{v}(\mathbf{0}) = \mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^n$. Note that $\mathbf{A}$ has dimensions $n \times p$ where $p \equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n) := |\mathcal{Q}|$. By definition, there exist distinct points $\mathbf{z}_1, \ldots, \mathbf{z}_p \in [0,1]^d$ with $\mathbf{z}_1 = \mathbf{0}$ such that the jth column of $\mathbf{A}$ is $\mathbf{v}(\mathbf{z}_j)$ for each j.

Our first result below deals with problem (3.2). Given the design matrix $\mathbf{A}$, we can define the following NNLS problem

$$\widehat{\boldsymbol{\beta}}_{\mathrm{EM}} \in \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p : \beta_j \geq 0, \forall j \geq 2} \|\mathbf{y} - \mathbf{A}\boldsymbol{\beta}\|^2 \tag{3.28}$$

where $\mathbf{y}$ is the $n \times 1$ vector consisting of the observations $y_1, \ldots, y_n$ coming from model (3.1). (3.28) is clearly a finite dimensional convex optimization problem (in fact, a quadratic optimization problem with linear constraints). Its solution $\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$ is not necessarily unique but the vector $\mathbf{A}\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$ is the projection of the observation vector $\mathbf{y}$ onto the closed convex cone $\{\mathbf{A}\boldsymbol{\beta} : \min_{j \geq 2} \beta_j \geq 0\}$ and is thus unique. The next result (proved in Section A.4.6)shows how to obtain a solution to problem (3.2) using any solution $\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$ of (3.28).

Proposition 3.3.1. *One solution for the optimization problem* (3.2) *is*

$$\widehat{f}_{\mathrm{EM}} := \sum_{j=1}^{p} (\widehat{\beta}_{\mathrm{EM}})_j \cdot \mathbb{I}_{[\mathbf{z}_j,\mathbf{1}]}, \tag{3.29}$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{EM}} = ((\widehat{\beta}_{\mathrm{EM}})_1, \ldots, (\widehat{\beta}_{\mathrm{EM}})_p)$ is any solution to (3.28).

Thus, one way to compute the estimator (3.2) is to solve the NNLS problem (3.28) and use the resulting coefficients in the above manner (3.29). It is interesting to note that the solution (3.29) is a rectangular piecewise constant function and the quantity $k(\widehat{f}_{\mathrm{EM}})$ (see Definition 3.2.5) will be controlled by the sparsity of $\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$. The key to proving Proposition 3.3.1 is the following characterization of $\mathcal{F}^d_{\mathrm{EM}}$ (proved in Section A.4.5).

Proposition 3.3.2 (Discretization of entirely monotone functions)**.** *For every set of design points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in [0,1]^d$, we have*

$$\{\mathbf{A}\boldsymbol{\beta} : \beta_j \geq 0, \forall j \geq 2\} = \big\{(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) : f \in \mathcal{F}^d_{\mathrm{EM}}\big\}.$$

Note that Proposition 3.3.2 immediately implies that for every minimizer $\widehat{f}_{\mathrm{EM}}$ of (3.2), the vector $(\widehat{f}_{\mathrm{EM}}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{EM}}(\mathbf{x}_n))$ equals $\mathbf{A}\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$ and is thus unique.

We now turn to problem (3.6). Given the matrix $\mathbf{A}$ and a tuning parameter $V > 0$, we can define the following LASSO problem:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V} \in \operatorname*{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p : \sum_{j\geq 2}|\beta_j| \leq V} \|\mathbf{y} - \mathbf{A}\boldsymbol{\beta}\|^2. \tag{3.30}$$

Again $\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$ may not be unique but $A\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$ is unique as it is the projection of $\mathbf{y}$ onto the closed convex set

$$\mathcal{C}(V) \coloneqq \left\{\mathbf{A}\boldsymbol{\beta} : \sum_{j\geq 2}|\beta_j| \leq V\right\}. \tag{3.31}$$

The next result (proved in Section A.4.8)shows how to obtain a solution to (3.6) using any solution $\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$ of (3.30).

Proposition 3.3.3. *One solution for the optimization problem* (3.6) *is*

$$\widehat{f}_{\mathrm{HK}\mathbf{0},V} \coloneqq \sum_{j=1}^{p} (\widehat{\beta}_{\mathrm{HK}\mathbf{0},V})_j \cdot \mathbb{I}_{[\mathbf{z}_j,\mathbf{1}]}, \tag{3.32}$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V} = ((\widehat{\beta}_{\mathrm{HK}\mathbf{0},V})_1, \ldots, (\widehat{\beta}_{\mathrm{HK}\mathbf{0},V})_p)$ is the solution to the LASSO problem (3.30)*.*

Thus, one way to compute the estimator (3.6) is to solve the LASSO problem (3.30) and use the resulting coefficients to construct the rectangular piecewise constant function (3.6). Note the strong similarity between the two expressions (3.29) and (3.32). The following result (proved in Section A.4.7)is the key ingredient in proving the above.

Proposition 3.3.4. *For every set of design points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in [0,1]^d$, we have*

$$\mathcal{C}(V) = \{(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) : V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) \leq V\}.$$

Proposition 3.3.4 immediately implies that for every minimizer $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ of (3.6), the vector $(\widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_n))$ equals $\mathbf{A}\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$ and is thus unique.

We have thus shown that the LSEs defined by (3.2) and (3.6) can be computed via NNLS and LASSO estimators with respect to the design matrix $\mathbf{A}$ whose columns are the elements of the finite set $\mathcal{Q}$ defined in (3.26). Once the design matrix $\mathbf{A}$ is formed, we can use existing quadratic program solvers to solve the NNLS and LASSO problems. The key to forming $\mathbf{A}$ is to enumerate the elements of $\mathcal{Q}$ and we address this issue now. We first state the following result which provides a worst case upper bound on $p \equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, the cardinality of $\mathcal{Q}$.

Lemma 3.3.5. *The cardinality of $\mathcal{Q}$ satisfies*

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n) \leq \sum_{j=0}^{d} \binom{n}{j} \tag{3.33}$$

for every $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$.

Lemma 3.3.5 is a consequence of the Vapnik-Chervonenkis lemma [88] and is proved in Section A.4.9). Note that the upper bound (3.33) can be further bounded by $(en/d)^d$.

We emphasize here that Lemma 3.3.5 gives a worst case upper bound for $p(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ (here worst case is in terms of the design configurations $\mathbf{x}_1, \ldots, \mathbf{x}_n$). For specific choices of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the quantity $p(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ can be much smaller than the right hand side of (3.33). For example, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are an enumeration of the grid points $\big\{(i_1/n^{1/d}, \ldots, i_d/n^{1/d}) : i_1, \ldots, i_d \in \{1, \ldots, n^{1/d}\}\big\}$ (or form any other full grid) then $p(\mathbf{x}_1, \ldots, \mathbf{x}_n) = n$ whereas the upper bound in (3.33) is of order n^d. However, there exist design configurations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ where the upper bound can be tight. For instance, when $d = 2$, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ lie on the anti-diagonal (the line segment connecting $(0, 1)$ and $(1, 0)$), then $p(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{n(n+1)}{2}$, so the upper bound $\frac{n(n+1)}{2} + 1$ in (3.33) is nearly tight for $p(\mathbf{x}_1, \ldots, \mathbf{x}_n)$.

The task of enumerating $\mathcal{Q}$ in general can be simplified if we show that we only need to check the value of $\mathbb{I}_{[\mathbf{z},\mathbf{1}]}$ on the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ for all $\mathbf{z}$ in some finite set S, rather than all $\mathbf{z} \in (0, 1]^d$ as in definition (3.26). Then we can list all $|S|$ evaluation vectors (and remove duplicates if necessary) to form $\mathbf{A}$. The following two strategies can be used to construct the set S:

1. **Naïve gridding.** The simplest idea is to let S be the smallest grid that contains the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$. That is, let $S = S_1 \times \cdots \times S_d$ where $S_j \coloneqq \{(\mathbf{x}_1)_j, \ldots, (\mathbf{x}_n)_j\}$ is the set of unique jth component values among the design points. It is simple to check that for any $\mathbf{z} \in (0, 1]^d$, the value of $\mathbb{I}_{[\mathbf{z},\mathbf{1}]}$ on the design points is the same as $\mathbb{I}_{[\mathbf{z}',\mathbf{1}]}$, where $\mathbf{z}'$ is the smallest element of S such that $\mathbf{z} \preceq \mathbf{z}'$. In the worst case, $|S_j| = n$ for each j, so we would need to check at most $|S| = n^d$ vectors.

2. **Component-wise minimum.** A better approach is to let

$$S \coloneqq \{\min\{\mathbf{x}_i : i \in I\} : I \subseteq [n], |I| \leq d\},$$

where "min" denotes component-wise minimum of vectors. That is, for each subset of the design points of size $\leq d$, we take the component-wise minimum and include that vector in S. To see why this definition of S suffices, consider any $\mathbf{z} \in [0,1]^d$ and note the $\mathbb{I}_{[\mathbf{z},\mathbf{1}]}$ has the same values on the design points as $\mathbb{I}_{[\mathbf{z}',\mathbf{1}]}$, where $\mathbf{z}' \coloneqq \min\{\mathbf{x}_i : i \in J\}$ and $J \coloneqq \{i : \mathbf{z} \preceq \mathbf{x}_i\}$. Furthermore, by the same reasoning as in our VC dimension computation above, there must exist some subset $I \subseteq J$ of size $\leq d$ such that $\min\{\mathbf{x}_i : i \in J\} = \min\{\mathbf{x}_i : i \in I\}$, which proves $\mathbf{z}' \in S$. In the worst case, we would need to check $|S| = \sum_{j=0}^{d} \binom{n}{j}$ vectors, which is the VC upper bound (3.33).

3.3.1 Special Case: the equally-spaced lattice design

The results stated so far in the section hold for every configuration of design points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in [0,1]^d$. We now specialize to the setting where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ form an equally-spaced lattice (precisely defined below). Our theoretical results described in the next section work under this setting. Moreover, some of the estimators from the literature that are related to $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ are defined only under the lattice design so a discussion of the form of our estimators in this setting will make it easier for us to compare and contrast them with existing estimators (this comparison is the subject of Section 3.5).

Given positive integers $n_1, \ldots, n_d$ with $n = n_1 \ldots n_d$, by a lattice design of dimensions $n_1 \times \cdots \times n_d$, we mean that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ form an enumeration of the points in

$$\mathbb{L}_{n_1,\ldots,n_d} \coloneqq \{(i_1/n_1, \ldots, i_d/n_d) : 0 \leq i_j \leq n_j - 1, j = 1, \ldots, d\} \tag{3.34}$$

Note that, in this setting, the set $\mathcal{Q}$ (defined in (3.26)) can be enumerated by $\mathcal{Q} = \{\mathbf{v}(\mathbf{x}_1), \ldots, \mathbf{v}(\mathbf{x}_n), \mathbf{0}\}$. Without loss of generality, we may ignore the $\mathbf{0}$ element and assume the columns of $\mathbf{A}$ are $\mathbf{v}(\mathbf{x}_1), \ldots, \mathbf{v}(\mathbf{x}_n)$ so that the i,j entry of $\mathbf{A}$ is given by $\mathbf{A}(i,j) = \mathbb{I}_{[\mathbf{x}_j,\mathbf{1}]}(\mathbf{x}_i) = \mathbb{I}\{\mathbf{x}_j \preceq \mathbf{x}_i\}$. We also take $\mathbf{x}_1 := \mathbf{0}$ (corresponding to $i_1 = \cdots = i_d = 0$) so that the first column of $\mathbf{A}$ is the vector of ones. Therefore in the lattice design setting, the optimization problems (3.28) and (3.30) for computing the two estimators $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ can be rewritten as

$$\widehat{\boldsymbol{\beta}}_{\mathrm{EM}} = \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p : \beta_j \geq 0, \forall j \geq 2} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{n} \mathbb{I}\{\mathbf{x}_j \preceq \mathbf{x}_i\} \boldsymbol{\beta}_j \right)^2 \tag{3.35}$$

and

$$\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V} = \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p : \sum_{j \geq 2} |\beta_j| \leq V} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{n} \mathbb{I}\{\mathbf{x}_j \preceq \mathbf{x}_i\} \boldsymbol{\beta}_j \right)^2 \tag{3.36}$$

respectively. It also turns out that, in the lattice design setting, the matrix $\mathbf{A}$ is square and invertible (Lemma A.4.1). As a result, it is possible to write down the vectors $(\widehat{f}_{\mathrm{EM}}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{EM}}(\mathbf{x}_n))$ and $(\widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_n))$ as solutions to more explicit constrained quadratic optimization problems. This is the content of the next result which is

proved in Section A.4.10. Here, it will be convenient to represent vectors in $\mathbb{R}^n$ as tensors indexed by $\mathbf{i} := (i_1, \ldots, i_d) \in \mathcal{I}$ where

$$\mathcal{I} := \Big\{\mathbf{i} = (i_1, \ldots, i_d) : i_j \in \{0, 1, \ldots, n_j - 1\} \text{ for every } j = 1, \ldots, d\Big\}. \tag{3.37}$$

In other words, we write the components of a vector $\boldsymbol{\theta} \in \mathbb{R}^n$ by $\theta_{\mathbf{i}}$ for $\mathbf{i} = (i_1, \ldots, i_d) \in \mathcal{I}$. We will also denote the observation corresponding to the design point $(i_1/n_1, \ldots, i_d/n_d)$ by $y_{\mathbf{i}} = y_{i_1,\ldots,i_d}$.

Lemma 3.3.6. *Consider the setting of the lattice design of dimensions $n_1 \times \cdots \times n_d$. For each $\boldsymbol{\theta} \in \mathbb{R}^n$, associate the "differenced" vector $D\boldsymbol{\theta} \in \mathbb{R}^n$ whose $\mathbf{i}^{th}$ entry is given by*

$$\sum_{j_1=0}^{1} \cdots \sum_{j_d=0}^{1} I\{i_1 - j_1 \geq 0, \ldots, i_d - j_d \geq 0\}(-1)^{j_1+\cdots+j_d}\theta_{i_1-j_1,\ldots,i_d-j_d} \tag{3.38}$$

for every $\mathbf{i} = (i_1, \ldots, i_d) \in \mathcal{I}$. Then:

1. *The vector $\Big(\widehat{f}_{\mathrm{EM}}(i_1/n_1, \ldots, i_d/n_d) : \mathbf{i} = (i_1, \ldots, i_d) \in \mathcal{I}\Big)$ is the solution to the optimization problem*

$$\operatorname{argmin}\left\{\sum_{\mathbf{i}\in\mathcal{I}} (y_{\mathbf{i}} - \theta_{\mathbf{i}})^2 : (D\boldsymbol{\theta})_{\mathbf{i}} \geq 0 \text{ for all } \mathbf{i} \neq \mathbf{0}\right\}. \tag{3.39}$$

2. *The vector $\Big(\widehat{f}_{\mathrm{HK}\mathbf{0},V}(i_1/n_1, \ldots, i_d/n_d) : \mathbf{i} = (i_1, \ldots, i_d) \in \mathcal{I}\Big)$ is the solution to the optimization problem*

$$\operatorname{argmin}\left\{\sum_{\mathbf{i}} (y_{\mathbf{i}} - \theta_{\mathbf{i}})^2 : \sum_{\mathbf{i}\neq\mathbf{0}} |(D\boldsymbol{\theta})_{\mathbf{i}}| \leq V\right\}. \tag{3.40}$$

Remark 3.3.7 (The special case of $d = 2$)**.** *When $d = 2$, it is easy to see that the differenced vector $D\boldsymbol{\theta}$ is given by*

$$(D\boldsymbol{\theta})_{(i_1,i_2)} = \begin{cases} \theta_{i_1,i_2} - \theta_{i_1-1,i_2} - \theta_{i_1,i_2-1} + \theta_{i_1-1,i_2-1} & \text{if } i_1 > 0, i_2 > 0 \\ \theta_{i_1,0} - \theta_{i_1-1,0} & \text{if } i_1 > 0, i_2 = 0 \\ \theta_{0,i_2} - \theta_{0,i_2-1} & \text{if } i_1 = 0, i_2 > 0 \\ \theta_{0,0} & \text{if } i_1 = i_2 = 0. \end{cases}$$

Using this, it is easy to see that (3.40) *can be rewritten for* $d = 2$ *as*

$$\operatorname{argmin}\Bigg\{\sum_{i_1=0}^{n_1-1}\sum_{i_2=0}^{n_2-1}(y_{i_1,i_2}-\theta_{i_1,i_2})^2 : \sum_{i_1=1}^{n_1-1}\sum_{i_2=1}^{n_2-1}|\theta_{i_1,i_2}-\theta_{i_1-1,i_2}-\theta_{i_1,i_2-1}+\theta_{i_1-1,i_2-1}| + \sum_{i_1=1}^{n_1-1}|\theta_{i_1,0}-\theta_{i_1-1,0}| + \sum_{i_2=1}^{n_2-1}|\theta_{0,i_2}-\theta_{0,i_2-1}| \le V\Bigg\} \tag{3.41}$$

and a similar formula can be written for (3.39) *for* $d = 2$.

As mentioned in the Introduction, an estimator similar to $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ has been described by Mammen and van de Geer [60] for $d = 2$ under the lattice design setting. Specifically, the estimator of [60] for the vector $(f^*(i_1/n_1, i_2/n_2), 0 \le i_1 \le n_1 - 1, 0 \le i_2 \le n_2 - 1)$ is given by the solution to the optimization problem:

$$\operatorname{argmin}\Bigg\{\sum_{i_1,i_2}(y_{i_1,i_2}-\theta_{i_1,i_2})^2 + \lambda_1\sum_{i_1,i_2\ge 1}|\theta_{i_1,i_2}-\theta_{i_1-1,i_2}-\theta_{i_1,i_2-1}+\theta_{i_1-1,i_2-1}| + \lambda_2\sum_{i_1\ge 1}|\overline{\theta}^{(1)}_{i_1}-\overline{\theta}^{(1)}_{i_1-1}| + \lambda_2\sum_{i_2\ge 1}|\overline{\theta}^{(2)}_{i_2}-\overline{\theta}^{(2)}_{i_2-1}|\Bigg\} \tag{3.42}$$

where λ_1 and λ_2 are positive tuning parameters, $\overline{\theta}^{(1)}_{i_1} := \frac{1}{n_2}\sum_{i_2=0}^{n_2-1}\theta_{i_1,i_2}$ and $\overline{\theta}^{(2)}_{i_2} := \frac{1}{n_1}\sum_{i_1=0}^{n_1-1}\theta_{i_1,i_2}$. This optimization problem is similar to (3.41) in that the first term in the penalty is the same in both problems. However the remaining terms in the penalty above are different from the terms in (3.41) although they are of the same spirit in that both are penalizing lower dimensional variations. Moreover, our estimator (3.41) has one tuning parameter (in the constrained form) and (3.42) has two tuning parameters in the penalized form. It should also be noted that we defined our estimators for arbitrary design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ while Mammen and van de Geer [60] only considered the lattice design for $d = 2$.

3.4 Risk results

In this section, risk bounds for the estimators $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ are presented. We define risk under the standard fixed design squared error loss function (see (3.9)). Throughout this section, we assume that we are working with the lattice design of dimensions $n_1 \times \cdots \times n_d$ with $n = n_1 \times \cdots \times n_d$ and $n_j \ge 1$ for all $j = 1, \ldots, d$.

3.4.1 Risk results for $\widehat{f}_{\mathbf{EM}}$

In this subsection, we present bounds on the risk $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*)$ of $\widehat{f}_{\mathrm{EM}}$ under the well-specified assumption where we assume that $f^* \in \mathcal{F}^d_{\mathrm{EM}}$. The first result below (proved in Section A.3.2) bounds the risk in terms of the HK$\mathbf{0}$ variation of f^*. Note that from part (ii) of Lemma 3.2.7, $V_{\mathrm{HK}\mathbf{0}}(f^*; [0,1]^d) = f^*(\mathbf{1}) - f^*(\mathbf{0})$ as $f^* \in \mathcal{F}^d_{\mathrm{EM}}$.

Theorem 3.4.1. *Let $f^* \in \mathcal{F}^d_{\mathrm{EM}}$ and $V^* \coloneqq V_{\mathrm{HK}\mathbf{0}}(f^*; [0,1]^d)$. For the lattice design* (3.34)*, the estimator $\widehat{f}_{\mathrm{EM}}$ satisfies*

$$\begin{aligned}\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) \leq C_d \left(\frac{\sigma^2 V^*}{n}\right)^{\frac{2}{3}} \left(\log\left(2 + \frac{V^*\sqrt{n}}{\sigma}\right)\right)^{\frac{2d-1}{3}} \\ + C_d \frac{\sigma^2}{n} (\log(en))^{\frac{3d}{2}} (\log(e\log(en)))^{\frac{2d-1}{2}}.\end{aligned} \tag{3.43}$$

where C_d is a constant that depends only on the dimension d.

Note that the bound (3.10) in the Introduction is the dominant first term of this bound (3.43).

Remark 3.4.2 (Model misspecification)**.** *Theorem 3.4.1 is stated under the well-specified assumption $f^* \in \mathcal{F}^d_{\mathrm{EM}}$. In the misspecified setting where $f^* \notin \mathcal{F}^d_{\mathrm{EM}}$, our LSE $\widehat{f}_{\mathrm{EM}}$ will not be close to f^*, but rather to*

$$\widetilde{f} \in \operatorname*{argmin}_{f \in \mathcal{F}^d_{\mathrm{EM}}} \sum_{i=1}^n (f(\mathbf{x}_i) - f^*(\mathbf{x}_i))^2,$$

so it is reasonable to consider $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, \widetilde{f})$ rather than $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*)$. By the argument outlined in Remark A.3.3, one can show that $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, \widetilde{f})$ is upper bounded by the right hand side of (3.43) *after re-defining V^* as $V_{\mathrm{HK}\mathbf{0}}(\widetilde{f}; [0,1]^d)$.*

As mentioned in the Introduction, when $d = 1$, the estimator $\widehat{f}_{\mathrm{EM}}$ is simply the isotonic LSE for which Zhang [97] proved that

$$\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) \leq C \left(\frac{\sigma^2 V^*}{n}\right)^{\frac{2}{3}} + C \frac{\sigma^2}{n} \log(en) \tag{3.44}$$

for some constant $C > 0$. It is interesting to note that our risk bound (3.43) for general $d \geq 2$ has the same terms as the univariate bound (3.44) with additional logarithmic factors which depend on d. It is natural to ask therefore if these additional logarithmic factors are indeed necessary or merely artifacts of our analysis. The next result (a minimax lower bound) shows that every estimator pays a logarithmic multiplicative price of $\log n$ for $d = 2$ and $(\log n)^{2(d-2)/3}$ for $d \geq 3$ in the first $n^{-2/3}$ term. We do not, unfortunately, know if the $(\log n)^{3d/2}(\log\log n)^{(2d-1)/2}$ factor in the second term in (3.43) is necessary or artifactual,

although we can prove that it can be removed by a modification of the estimator $\widehat{f}_{\mathrm{EM}}$ (see Theorem 3.4.4 below).

The next result (proved in Section A.3.7) proves a lower bound for the minimax risk:

$$\mathfrak{M}_{\mathrm{EM},\sigma,V,d}(n) := \inf_{\widehat{f}_n} \sup_{f^* \in \mathcal{F}^d_{\mathrm{EM}}: V_{\mathrm{HK0}}(f^*) \leq V} \mathbb{E}_{f^*} \mathcal{L}(\widehat{f}_n, f^*), \tag{3.45}$$

where the expectation is with respect to model (3.1).

Theorem 3.4.3. *Let $d \geq 2$, $V > 0$, $\sigma > 0$ and let $n_j \geq c_s n^{1/d}$ for all $j = 1, \ldots, d$ for some $c_s \in (0, 1]$. Then there exists a positive constant C_d depending only on d and c_s, such that the minimax risk on the lattice design* (3.34) *satisfies*

$$\mathfrak{M}_{\mathrm{EM},\sigma,V,d}(n) \geq C_d \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} \left(\log\left(\frac{V\sqrt{n}}{\sigma}\right)\right)^{\frac{2(d-2)}{3}}$$

provided n is larger than a positive constant $c_{d,\sigma^2/V^2}$ depending only on d, σ^2/V^2, and c_s. In the case $d = 2$, this bound can be tightened to

$$\mathfrak{M}_{\mathrm{EM},\sigma,V,d}(n) \geq C \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} \log\left(\frac{V\sqrt{n}}{\sigma}\right). \tag{3.46}$$

Note that the assumption $n_j \geq c_s n^{1/d}$ for all j is reasonable, since if, for instance, $n_{d'+1} = n_{d'+2} \cdots = n_d = 1$ then we simply have a d'-dimensional problem where $d' < d$, which should have a smaller minimax risk.

As mentioned before, the above result shows that some dependence on dimension d in the logarithmic term cannot be avoided for any estimator. Note also, that for $d = 2$, the minimax lower bound (3.46) matches our upper bound in Theorem 3.4.1 implying minimaxity of $\widehat{f}_{\mathrm{EM}}$ for $d = 2$. For $d > 2$, there remains a gap of $\log n$ between our minimax lower bound and the upper bound in Theorem 3.4.1. This gap is due to a logarithmic gap between an upper bound and lower bound given by Blei et al. [12, Theorem 1.1] for the metric entropy of cumulative distribution functions of probability measures on $[0, 1]^d$, a gap that essentially reduces to improving estimates of a small ball probability of Brownian sheets (see discussion in [12] for more detail and references).

As mentioned earlier, the logarithmic factor $(\log n)^{3d/2}(\log\log n)^{(2d-1)/2}$ appearing in the second term of (3.43) can be removed by a modification of the estimator $\widehat{f}_{\mathrm{EM}}$. This is shown in the next result. For a tuning parameter $V \geq 0$, let

$$\widetilde{f}_{\mathrm{EM},V} \in \operatorname*{argmin}_{f \in \mathcal{F}^d_{\mathrm{EM}}: V_{\mathrm{HK0}}(f) \leq V} \frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2.$$

Note that this differs from the original estimator (3.2) only by the introduction of the additional constraint $V_{\mathrm{HK0}}(f) \leq V$.

Theorem 3.4.4. *Let $f^* \in \mathcal{F}^d_{\mathrm{EM}}$ and $V^* := V_{\mathrm{HK0}}(f^*; [0,1]^d)$. Assume the lattice design (3.34). If the tuning parameter V is such that $V \geq V^*$, then the estimator $\widetilde{f}_{\mathrm{EM},V}$ satisfies*

$$\mathcal{R}(\widetilde{f}_{\mathrm{EM},V}, f^*) \leq C_d \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} \left(\log\left(2 + \frac{V\sqrt{n}}{\sigma}\right)\right)^{\frac{2d-1}{3}} + C_d \frac{\sigma^2}{n}. \tag{3.47}$$

Note that the second term in (3.47) is just σ^2/n and smaller than the second term in (3.43) but this comes at the cost of introducing a tuning parameter V that needs to be at least V^*.

We will now prove near-parametric rates for $\widehat{f}_{\mathrm{EM}}$ when f^* is rectangular piecewise constant. To motivate these results, note first that when f^* is constant on $[0,1]^d$, we have $V^* = 0$ and thus the bound given by (3.43) is σ^2/n up to logarithmic factors. In the next result (proved in Section A.3.3), we generalize this fact and show that $\widehat{f}_{\mathrm{EM}}$ achieves nearly the parametric rate for rectangular piecewise constant functions $f^* \in \mathcal{F}^d_{\mathrm{EM}}$. Recall the definition of the class $\mathfrak{R}^d$ of all rectangular piecewise constant functions and the associated mapping $k(f), f \in \mathfrak{R}^d$, from Definition 3.2.5.

Theorem 3.4.5. *For every $f^* : [0,1]^d \to \mathbb{R}$, the LSE $\widehat{f}_{\mathrm{EM}}$ satisfies*

$$\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) \leq \inf_{f \in \mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{EM}}} \left\{ \mathcal{L}(f, f^*) + C_d \sigma^2 \frac{k(f)}{n} (\log(en))^{\frac{3d}{2}} (\log(e \log(en)))^{\frac{2d-1}{2}} \right\}.$$

Theorem 3.4.5 gives a sharp oracle inequality in the sense of [10] as it applies to every function f^* (even in the misspecified case when $f^* \notin \mathcal{F}^d_{\mathrm{EM}}$) and the constant in front of the first term inside the infimum equals 1. Even though the inequality holds for every f^*, the right hand side will be small only when f^* is close to some function f in $\mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{EM}}$. This implies that when $f^* \in \mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{EM}}$, we can take $f = f^*$ in the right hand side to obtain that the risk of $\widehat{f}_{\mathrm{EM}}$ decays as $\sigma^2 k(f^*)/n$ up to logarithmic factors. This rate will be faster than the rate given by Theorem 3.4.1 provided $k(f^*)$ is not too large. Note that one can combine the two bounds given by Theorem 3.4.1 and Theorem 3.4.5 by taking their minimum. In the case $d = 1$, Theorem 3.4.5 reduces to the adaptive rates for isotonic regression [19, 10] but with worse logarithmic factors.

We would also like to mention here that $\mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{EM}}$ is a smaller class compared to $\mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{M}}$ (recall that $\mathcal{F}^d_{\mathrm{M}}$ is defined via (3.20)). Risk results over the class $\mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{M}}$ for the LSE over $\mathcal{F}^d_{\mathrm{M}}$ and other related estimators have been proved in Han et al. [47] and Deng and Zhang [28].

Before closing this subsection, let us briefly describe the main ideas underlying the proofs of Theorems 3.4.1, 3.4.3, 3.4.4 and 3.4.5. For Theorem 3.4.1, we use standard results on the accuracy of LSEs on closed convex sets which related the risk of $\widehat{f}_{\mathrm{EM}}$ to covering numbers of local balls of the form $\left\{f \in \mathcal{F}^d_{\mathrm{EM}} : \mathcal{L}(f, f^*) \leq t^2\right\}$ for $t > 0$ sufficiently small in the pseudo-metric given by the square-root of the loss function $\mathcal{L}$. We calculated the covering numbers of these local balls by relating the functions in $\mathcal{F}^d_{\mathrm{EM}}$ to distribution functions of signed measures on $[0,1]^d$ and using existing covering number results for distribution functions of signed

measures from Blei et al. [12] and Gao [36]. The proof of Theorem 3.4.3 is also based on covering number arguments as we use general minimax lower bounds from Yang and Barron [94]. Finding lower bounds for the covering numbers under the pseudometric $\sqrt{\mathcal{L}}$ seems somewhat involved and we used a multiscale construction from Blei et al. [12, Section 4] for this purpose. The bound in Theorem 3.4.4 for $\widetilde{f}_{\mathrm{EM},V}$ is a quick consequence of the proof of the risk bound for $\widehat{f}_{\mathrm{HK0},V}$ (Theorem 3.4.6) which is stated in the next subsection. For Theorem 3.4.5, we used standard results relating $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*)$ to a certain size-related measure (statistical dimension) of the tangent cone to $\widehat{f}_{\mathrm{EM}}$ at f^*. When $f^* \in \mathfrak{R}^d$ (or when f^* is approximable by a function in $\mathfrak{R}^d$), this tangent cone is decomposable into tangent cones of certain lower-dimensional tangent cones. The statistical dimension of these lower-dimensional tangent cones is then bounded via an application of Theorem 3.4.1 in the case when $V^* = 0$.

3.4.2 Risk results for $\widehat{f}_{\mathbf{HK0},V}$

In this subsection, we present bounds on the risk $\mathcal{R}(\widehat{f}_{\mathrm{HK0},V}, f^*)$ of the estimator $\widehat{f}_{\mathrm{HK0},V}$. Note that the estimator $\widehat{f}_{\mathrm{HK0},V}$ involves a tuning parameter V and therefore these results will require some conditions on V. Our first result below assumes that $V \geq V^* := V_{\mathrm{HK0}}(f^*; [0,1]^d)$ and gives the $n^{-2/3}$ rate up to logarithmic factors. The proof of this result is given in Section A.3.4.

Theorem 3.4.6. *Assume the lattice design* (3.34)*. If the tuning parameter V is such that $V \geq V^* := V_{\mathrm{HK0}}(f^*; [0,1]^d)$, then the estimator $\widehat{f}_{\mathrm{HK0},V}$ satisfies*

$$\mathcal{R}(\widehat{f}_{\mathrm{HK0},V}, f^*) \leq C_d \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} \left(\log\left(2 + \frac{V\sqrt{n}}{\sigma}\right)\right)^{\frac{2d-1}{3}} + C_d \frac{\sigma^2}{n}. \tag{3.48}$$

Remark 3.4.7. *As mentioned earlier, Mammen and van de Geer [60] (see also the very recent paper Ortelli and van de Geer [67]) proposed the estimator* (3.42) *that is similar to $\widehat{f}_{\mathrm{HK0},V}$. Mammen and van de Geer [60] also proved a risk result for their estimator giving the rate $n^{-(1+d)/(1+2d)}$ which is strictly suboptimal compared to our rate in* (3.48) *for $d \geq 2$. This suboptimality is likely due to the use of suboptimal covering number bounds in [60].*

Remark 3.4.8 (Model misspecification)**.** *Theorem 3.4.6 is stated under the well-specified assumption $V_{\mathrm{HK0}}(f^*; [0,1]^d) \leq V$. In the misspecified setting where $V_{\mathrm{HK0}}(f^*; [0,1]^d) > V$, our LSE $\widehat{f}_{\mathrm{HK0},V}$ will not be close to f^*, but to $\widetilde{f} \in \operatorname{argmin}_{f: V_{\mathrm{HK0}}(f) \leq V} \sum_{i=1}^n (f(\mathbf{x}_i) - f^*(\mathbf{x}_i))^2$, so it is reasonable to consider $\mathcal{R}(\widehat{f}_{\mathrm{HK0},V}, \widetilde{f})$ rather than $\mathcal{R}(\widehat{f}_{\mathrm{HK0},V}, f^*)$. By the argument outlined in Remark A.3.3, $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, \widetilde{f})$ is upper bounded by the right hand side of* (3.48)*.*

In the next result, we prove a complementary minimax lower bound to Theorem 3.4.6 which proves that, for $d \geq 2$, the risk of every estimator over the class $\{f^* : V_{\mathrm{HK0}}(f^*) \leq V\}$ is bounded from below by $n^{-2/3}(\log n)^{2(d-1)/3}$ (ignoring terms depending on d, V and σ).

This implies that the logarithmic terms in (3.48) can perhaps be reduced slightly but cannot be removed altogether and must necessarily increase with the dimension d. Let

$$\mathfrak{M}_{\mathrm{HK},\sigma,V,d}(n) := \inf_{\widehat{f}_n} \sup_{f^*: V_{\mathrm{HK}\mathbf{0}}(f^*) \leq V} \mathbb{E}_{f^*} \mathcal{L}(\widehat{f}_n, f^*),$$

where the expectation is with respect to model (3.1). Note that $\{f^* \in \mathcal{F}^d_{\mathrm{EM}} : V_{\mathrm{HK}\mathbf{0}}(f^*) \leq V\} \subseteq \{f^* : V_{\mathrm{HK}\mathbf{0}}(f^*) \leq V\}$ which implies that

$$\mathfrak{M}_{\mathrm{HK},\sigma,V,d}(n) \geq \mathfrak{M}_{\mathrm{EM},\sigma,V,d}(n)$$

where $\mathfrak{M}_{\mathrm{EM},\sigma,V,d}(n)$ is defined in (3.45). This implies, in particular, that the lower bounds on $\mathfrak{M}_{\mathrm{EM},\sigma,V,d}(n)$ from Theorem 3.4.3 are also lower bounds on $\mathfrak{M}_{\mathrm{HK},\sigma,V,d}(n)$. However the next result (whose proof is in Section A.3.6) gives a strictly larger lower bound for $\mathfrak{M}_{\mathrm{HK},\sigma,V,d}(n)$ for $d > 2$ than that given by Theorem 3.4.3.

Theorem 3.4.9. *Let $d \geq 2$, $V > 0$, $\sigma > 0$ and let $n_j \geq c_s n^{1/d}$ for $j = 1, \ldots, d$, where $c_s \in (0, 1]$. Then there exists a positive constant C_d depending only on d and c_s, such that*

$$\mathfrak{M}_{\mathrm{HK},\sigma,V,d}(n) \geq C_d \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} \left(\log\left(\frac{V\sqrt{n}}{\sigma}\right)\right)^{\frac{2(d-1)}{3}}$$

provided n is larger than a positive constant $c_{d,\sigma^2/V^2}$ depending only on d, σ^2/V^2, and c_s. In the case $d = 2$, this bound can be tightened to

$$\mathfrak{M}_{\mathrm{HK},\sigma,V,d}(n) \geq C \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} \log\left(\frac{V\sqrt{n}}{\sigma}\right).$$

Theorems 3.4.6 and 3.4.9 together imply that $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ is minimax optimal over $\{f^* : V_{\mathrm{HK}\mathbf{0}}(f^*) \leq V\}$ for $d = 2$ and only possibly off by a factor of $(\log n)^{1/3}$ for $d > 2$.

We next explore the possibility of near parametric rates for $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ for rectangular piecewise constant functions. In the univariate case $d = 1$, it is known (see [44, Theorem 2.2]) that $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ satisfies the near-parametric risk bound (3.13) provided (a) the tuning parameter V is taken to be close to V^*, (b) f^* is piecewise constant, and (c) the length of each constant piece of f^* is bounded from below by $c/k(f^*)$ for some $c > 0$. The next result (proved in Section A.3.8) provides evidence that a similar story holds true for estimating certain rectangular piecewise constant functions.

For a given constant $0 < c \leq 1/2$, let $\mathfrak{R}^d_1(c)$ denote the collection of functions $f : [0, 1]^d \to \mathbb{R}$ of the form

$$f = a_1 \mathbb{I}_{[\mathbf{x}^*, \mathbf{1}]} + a_0 \tag{3.49}$$

for some $a_1, a_0 \in \mathbb{R}$ and $\mathbf{x}^* \in [0, 1]^d$ satisfying the minimum size condition

$$\min\{|\mathbb{L}_{n_1,\ldots,n_d} \cap [\mathbf{x}^*, \mathbf{1}]|, |\mathbb{L}_{n_1,\ldots,n_d} \cap [\mathbf{0}, \mathbf{x}^*)|\} \geq cn. \tag{3.50}$$

To gain more intuition about the above condition, note first that we are working with the lattice design so that $\mathbb{L}_{n_1,\ldots,n_d} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is the set containing all design points. Roughly speaking, (3.50) ensures that $\mathbf{x}^*$ is not too close to the boundary of $[0,1]^d$ so that each of the rectangles $[\mathbf{x}^*, \mathbf{1}]$ and $[\mathbf{0}, \mathbf{x}^*)$ contain at least some constant fraction of the n design points.

It is clear that $\mathfrak{R}_1^d(c)$ is a subset of $\mathfrak{R}^d$, i.e., every function of the form (3.49) is rectangular piecewise constant. Indeed, it is easy to see that $k(f) \leq 2^d$ for every $f \in \mathfrak{R}_1^d(c)$. The following result (proved in Section A.3.8 bounds the risk of $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ for $f^* \in \mathfrak{R}_1^d(c)$.

Theorem 3.4.10. *Consider the lattice design* (3.34) *with* $n > 1$. *Fix* $f^* : [0,1]^d \to \mathbb{R}$ *and consider the estimator* $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ *with a tuning parameter* V. *Then for every* $0 < c \leq 1/2$, *we have*

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) \leq \inf_{\substack{f \in \mathfrak{R}_1^d(c): \\ V_{\mathrm{HK}\mathbf{0}}(f) = V}} \left\{ \mathcal{L}(f, f^*) + C(c,d) \frac{\sigma^2}{n} (\log n)^{\frac{3d}{2}} (\log\log n)^{\frac{2d-1}{2}} \right\} \tag{3.51}$$

for a constant $C(c,d)$ *that depends only on* c *and* d.

Theorem 3.4.10 applies to every function f^* but the infimum on the right hand side of (3.51) is over all functions f in $\mathfrak{R}_1^d(c)$ with $V_{\mathrm{HK}\mathbf{0}}(f) = V$. Therefore, Theorem 3.4.10 implies that the risk of the estimator $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ with tuning parameter V at f^* is the near-parametric rate $\frac{\sigma^2}{n}(\log en)^{3d/2}(\log\log n)^{(2d-1)/2}$ provided f^* is close to some function f in $\mathfrak{R}_1^d(c)$ with $V = V_{\mathrm{HK}\mathbf{0}}(f)$. As an immediate consequence, we obtain that if $f^* \in \mathfrak{R}_1^d(c)$ and $V = V_{\mathrm{HK}\mathbf{0}}(f^*)$, then

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) \leq C(c,d) \frac{\sigma^2}{n} (\log(en))^{\frac{3d}{2}} (\log(e\log(en)))^{\frac{2d-1}{2}}.$$

Functions in $\mathfrak{R}_1^d(c)$ are constrained to satisfy the minimum size condition (3.50). A comparison of Theorem 3.4.10 with the corresponding univariate results shows that the near-parametric rate cannot be achieved without any minimum size condition (see e.g., [44, Remark 2.5] and [32, Section 4]). However, condition (3.50) might sometimes be too stringent for $d \geq 2$. For example, it rules out the case when $\mathbf{x}^* := (0.5, 0, \ldots, 0)$ which means that the function class $\mathfrak{R}_1^d(c)$ excludes simple functions such as $f(\mathbf{x}) := \mathbb{I}\{x_1 \geq 1/2\}$. In Theorem A.1.1 (deferred to Section A.1), we show that when $d = 2$, it is possible to obtain the same risk bound under a weaker minimum size condition which does not rule out functions such as $f(\mathbf{x}) := \mathbb{I}\{x_1 \geq 1/2\}$.

The implication of Theorems 3.4.10 and A.1.1 is that there exists a subclass of $\mathfrak{R}^d$ consisting of indicators of upper right rectangles in $[0,1]^d$ over which the estimator $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$, when ideally tuned, achieves the near-parametric rate with some logarithmic factors. Simulations (see Section A.2.3) indicate that this should also be true for a larger subclass of $\mathfrak{R}^d$ consisting of all functions in $\mathfrak{R}^d$ satisfying some minimum size condition, but our proof technique does not currently work in this generality. Ortelli and van de Geer [66] recently proved, for $d = 2$, near-parametric rates for the estimator (3.42) for a more general class of piecewise constant

functions, but for a smaller loss function. Their proof technique is completely different from our approach.

Let us now briefly discuss the key ideas behind the proofs of Theorems 3.4.6, 3.4.9 and 3.4.10. Theorem 3.4.6 is proved via covering number arguments which relate $\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*)$ to covering numbers of $\{f : V_{\mathrm{HK}\mathbf{0}}(f) \leq V\}$ and these covering numbers are controlled by invoking connections to distribution functions of signed measures. Theorem 3.4.9 is proved by Assouad's lemma with a multiscale construction of functions with bounded HK**0** variation. This multiscale construction is involved and taken from Blei et al. [12, Section 4].

The ideas for the proof of Theorem 3.4.10 (and also Theorem A.1.1) is borrowed from the proofs for the univariate case in Guntuboyina et al. [44] although the situation for $d \geq 2$ is much more complicated. At a high level, we use tangent cone connections where the goal is to control an appropriate size measure (Gaussian width) of the tangent cone of $\{f : V_{\mathrm{HK}\mathbf{0}}(f) \leq V^*\}$ at f^*. This tangent cone can be explicitly computed (see Lemma A.3.12). To bound its Gaussian width, our key observation is that for functions f^* in $\mathfrak{R}^d_1(c)$, every element of the tangent cone can be broken down into lower-dimensional elements each of which is either nearly entirely monotone or has low HK**0** variation. The Gaussian width of the tangent cone can then be bounded by a combination of (suitably strengthened) versions of Theorem 3.4.5 and Theorem 3.4.6. This method unfortunately does not seem to work for arbitrary functions $f^* \in \mathfrak{R}^d$ because of certain technical issues which are mentioned in Remark A.3.17.

3.5 On the "dimension-independent" rate $n^{-2/3}$ in Theorem 3.4.1 and Theorem 3.4.6

As mentioned previously, the dimension d appears in the bounds given by Theorem 3.4.1 and Theorem 3.4.6 only through the logarithmic term which means that $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ attain "dimension-independent rates" ignoring logarithmic factors. We shall provide some insight and put these results in proper historical context in this section. In nonparametric statistics, it is well-known that the rate of estimation of smooth functions based on n observations is $n^{-2m/(2m+d)}$ where d is the dimension and m is the order of smoothness [78]. The constraints of entire monotonicity and having finite HK**0** variation can be loosely viewed as smoothness constraints of order $m = d$. This is because, for smooth functions f, entire monotonicity is equivalent to

$$\frac{\partial^{|S|} f}{\prod_{j \in S} \partial x_j} \geq 0 \qquad \text{for every } \varnothing \neq S \subseteq \{1, \ldots, d\}$$

and the constraint of finite HK**0** variation is equivalent to

$$\frac{\partial^{|S|} f}{\prod_{j \in S} \partial x_j} \in L^1 \qquad \text{for every } \varnothing \neq S \subseteq \{1, \ldots, d\}. \tag{3.52}$$

Because derivatives of order d appear in these expressions, these constraints should be considered as smoothness constraints of order d. Note that taking $m = d$ in $n^{-2m/(2m+d)}$ gives $n^{-2/3}$.

Some other papers which studied such higher order constraints to obtain estimators having nearly dimension-free rates include [8, 58, 22, 65, 76, 89]. In particular, Lin [58] studied estimation under the constraint:

$$\frac{\partial^{|S|} f}{\prod_{j \in S} \partial x_j} \in L^2 \qquad \text{for every } \varnothing \neq S \subseteq \{1, \ldots, d\}. \tag{3.53}$$

The difference between (3.52) and (3.53) is that L^1 in (3.52) is replaced by L^2 in (3.53). Lin [58] proved that the minimax rate of convergence under (3.53) is $n^{-2/3}(\log n)^{2(d-1)/3}$ and constructed a linear estimator which is optimal over the class (3.53). Let us remark here that the L^2 constraint makes the class smaller compared to (3.52) and also enables linear estimators to achieve the optimal rate. However, linear estimators will not be optimal over $\{f : V_{\mathrm{HK0}}(f) \leq V\}$ as is well-known in $d = 1$ (see Donoho and Johnstone [30]) and the estimator of Lin [58] will also not adapt to rectangular piecewise constant functions (note that it is not possible to extend (3.53) to nonsmooth functions in such a way that the constraint is satisfied by rectangular piecewise constant functions).

Let us also mention here that, in approximation theory, it is known that classes of smooth functions f on $[0, 1]^d$ satisfying mixed partial derivative constraints such as (3.52) or (3.53) allow one to overcome the curse of dimensionality to some extent from the perspective of metric entropy, approximation and interpolation (see e.g., [29, 79, 15]).

Another way to impose higher order smoothness is to impose the constraint:

$$\frac{\partial^d f}{\partial x_j^d} \in L^1 \qquad \text{for each } j = 1, \ldots, d \tag{3.54}$$

as in the Kronecker Trend filtering method of order $k+1 = d$ of Sadhanala et al. [76] who also proved that this leads to the dimension-free rate $n^{-2/3}$ up to logarithmic factors. There are some differences between the constraints (3.52) and (3.54). For example, product functions $f(x_1, \ldots, x_d) := f_1(x_1) \ldots f_d(x_d)$ satisfy (3.52) provided each f_j satisfies $f_j' \in L_1$ while they will satisfy (3.54) provided $f_j^{(d)} \in L_1$.

Finally, let us mention that, in the usual multivariate extensions of isotonic regression and total variation denoising, one uses partial derivatives only of the first order which leads to rates of convergence that are exponential in the dimension d. For example, the usual multivariate isotonic regression (see e.g., Robertson et al. [73, Section 1.3]) considers the class $\mathcal{F}_{\mathrm{M}}^d$ of multivariate monotone functions which only imposes first order constraints. The rate of convergence here is given by $n^{-1/d}$ as recently shown in Han et al. [47]. This rate is exponentially slow in the dimension d. One sees the same rate behavior for the multivariate total variation denoising estimator (which also imposes only first order constraints) originally proposed by Rudin et al. [74] and whose theoretical behavior is studied in Hütter and Rigollet [52], Sadhanala et al. [77], Chatterjee and Goswami [18], Ortelli and van de Geer [68], Ruiz et al. [75].

Chapter 4

Shape constraints and interactions

4.1 Introduction

Consider the nonparametric regression framework, where the goal is to estimate an unknown function $f^* : [0,1]^d \to \mathbb{R}$ from noisy observations

$$y_i = f^*(\mathbf{x}_i) + \xi_i, \qquad \text{where } \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \quad \text{for } i = 1, \ldots, n. \tag{4.1}$$

We continue to focus on least squares estimators (LSEs) of the form

$$\operatorname*{argmin}_f \frac{1}{n} \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2,$$

where the minimum is taken with respect to some function class. In the previous chapter, we proposed and analyzed the least squares estimator with respect to $\mathcal{F}^d_{\text{EM}}$, the class of entirely monotone functions (3.2) and noted it is one multivariate generalization of univariate isotonic regression. We discussed in Section 3.5 that this class of entirely monotone functions is a small enough subclass of multivariate monotonic functions to avoid the curse of dimensionality to some extent.

Yet another multivariate generalization of isotonic regression is additive monotonic regression [6], where the function class consists of multivariate functions that are sums of univariate isotonic functions in each component. From a practical standpoint, performing regression with respect to this class has a number of attractive properties. Computationally, one can efficiently compute the estimator using the cyclic pool-adjacent violators algorithm [6], a multivariate generalization of the well-known pool-adjacent violators algorithm [5] for univariate isotonic regression. From a modeling perspective, one only needs to establish the monotonic direction of the relationship between the response variable and each covariate, which is usually known from the context of the regression; this nonparametric modeling avoids making more stringent assumptions like linearity on the relationship between the response and the covariates. At the same time, the additivity of the model makes the fitted function fairly

interpretable, since the effect of a particular covariate on the response is entirely captured by the corresponding univariate nondecreasing fitted function.

In particular, the class of additive monotonic functions is a subclass of entirely monotone functions. Contrasting the two function classes allows us to view entirely monotonic regression from a new perspective. The additivity constraint in additive monotonic regression prevents any interaction in how the covariates affect the response; entire monotonicity can be viewed as a particular relaxation of this condition by allowing "positive interactions" of all orders among covariates. This perspective also naturally leads us to consider a variety of intermediate classes lying between the interaction-less additive monotonic class and the entirely monotonic class, by allowing only certain interactions.

In Section 4.2, we describe how relative to the additive monotonic function class, the entirely monotonic function class introduces positive interactions. We then define two types of intermediate function classes that each interpolate between additive monotonicity and entirely monotonicity. In Section 4.3 we show how the least squares estimators with respect to these function classes can be computed by solving corresponding nonnegative least squares problems. In Section 4.4 we prove a risk bound for one of these types of estimators, and show that it generalizes the risk bound for entirely monotonic functions (3.43). In Section 4.5, we describe how one can perform a hypothesis test for whether one should include certain interaction terms in a model. The test is a likelihood ratio test involving nested convex cones, and we use a result of Menéndez et al. [62] to prove that this test is dominated by a different likelihood ratio test for a set of reduced hypotheses.

4.2 Terminology and motivation

We define the class of *additive monotonic* functions $\mathcal{F}^d_{\mathrm{AM}}$ as functions $f : [0,1]^d \to \mathbb{R}$ of the form

$$f(\mathbf{x}) = \sum_{j=1}^{d} f_j(x_j), \tag{4.2}$$

for some nondecreasing univariate functions $f_j : [0,1] \to \mathbb{R}$, $j = 1, \ldots, d$. This generalization of univariate isotonic regression via additivity is analogous to how multiple linear regression generalizes simple linear regression by considering functions of the form (4.2) but with each f_j being a univariate linear function. One common feature of these additive models is that they exclude interactions of the covariates in the following sense. Given $\mathbf{x} = (x_1, \ldots, x_d)$ and $\mathbf{x}' = (x_1, \ldots, x'_j, \ldots, x_d)$ that differ only in the jth component, the effect of this change in the jth component is

$$f(\mathbf{x}') - f(\mathbf{x}) = f_j(x'_j) - f_j(x_j),$$

which does not depend on the value of the other components $\mathbf{x}_{\setminus j} \coloneqq (x_k : k \neq j)$. Equivalently, the univariate slice $x_j \mapsto f(\mathbf{x})$ for fixed $\mathbf{x}_{\setminus j}$ is $f_j(x_j) + c$ where the other covariates $\mathbf{x}_{\setminus j}$ only affect the constant shift c; see Figure 4.1 for a depiction of these parallel slices. Depending on the context, the class of additive monotonic functions may or may not be a good

modeling choice. Similar to how univariate isotonic regression is flexible in imposing only the mild shape constraint of monotonicity while producing piecewise constant fitted functions, additive monotonic regression retains the flexible piecewise constant fitted functions for each component. However, the imposition of having no covariate interaction may be too restrictive in certain applications where it is known that two or more covariates interact to affect the response. Contrasting the additive monotonic class with the entirely monotonic class of functions gives us perspectives on how to relax this no-interaction constraint.

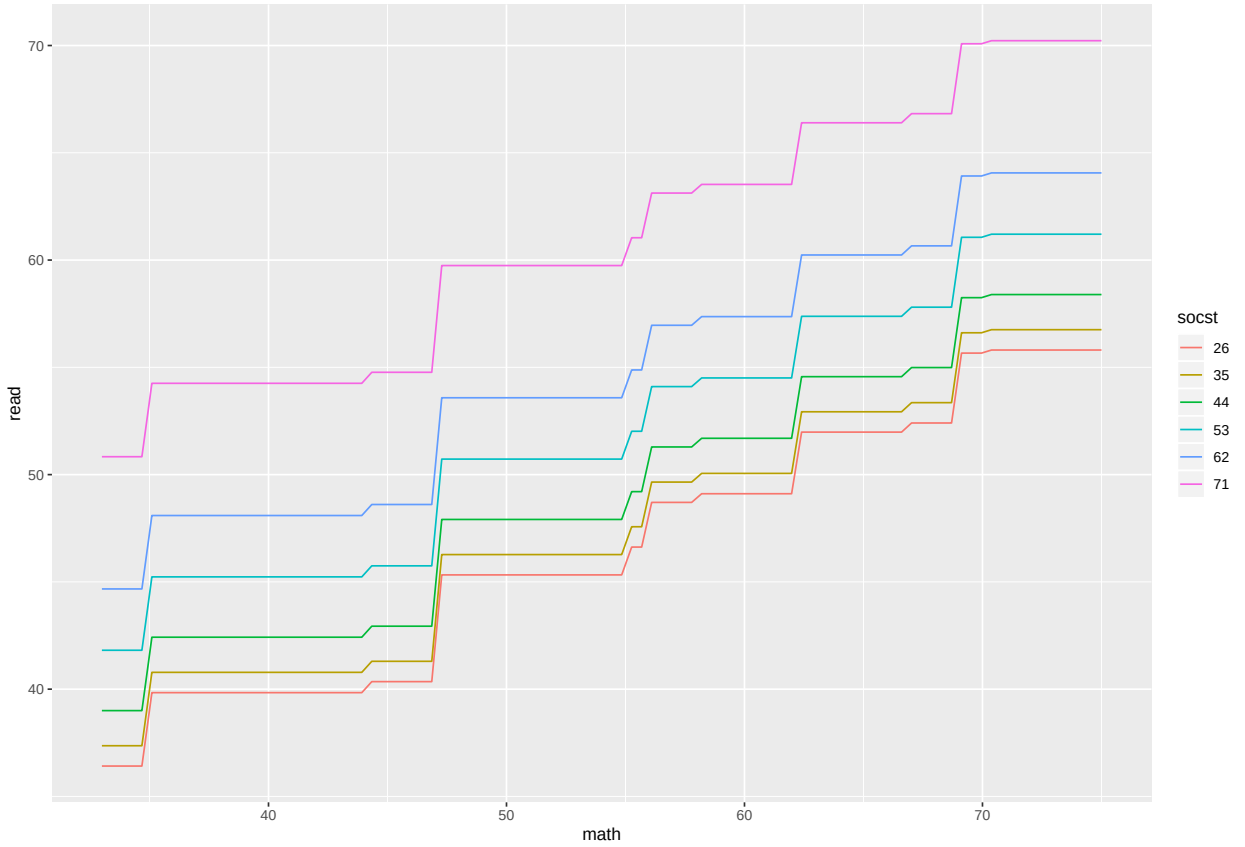


Figure 4.1: Additive monotonic regression of students' reading scores on their math and social studies scores: the slices of the fitted function (for fixed social studies scores) are parallel.

Recall from Section 3.2 the definition of the class of entire monotonic functions $\mathcal{F}^d_{\mathrm{EM}}$: a function $f : [0,1]^d \to \mathbb{R}$ is entirely monotonic if the quasi-volumes

$$\Delta(f; [\mathbf{a}, \mathbf{b}]) \coloneqq \sum_{j_1=0}^{J_1} \cdots \sum_{j_d=0}^{J_d} (-1)^{j_1+\cdots+j_d} f(b_1 + j_1(a_1 - b_1), \ldots, b_d + j_d(a_d - b_d)), \; J_k \coloneqq \mathbb{I}\{a_k \neq b_k\}, \tag{4.3}$$

are nonnegative for every rectangle $[\mathbf{a}, \mathbf{b}]$ in $[0,1]^d$ where $\mathbf{a} \preceq \mathbf{b}$ and $\mathbf{a} \neq \mathbf{b}$. The following lemma (proved in Section 4.6.4) shows exactly how the additive monotonic class is a subclass of entirely monotonic functions.

Lemma 4.2.1. *The class of additive monotonic functions can be characterized as*

$$\mathcal{F}^d_{\mathrm{AM}} = \{f \in \mathcal{F}^d_{\mathrm{EM}} : \Delta(f; [\mathbf{a}, \mathbf{b}]) = 0 \text{ for } \mathbf{a} \preceq \mathbf{b} \text{ where } |\{k : a_k \neq b_k\}| > 1\}.$$

Consequently, $\mathcal{F}^d_{\mathrm{AM}} \subseteq \mathcal{F}^d_{\mathrm{EM}}$*, with equality if and only if* $d = 1$*.*

In light of Lemma 4.2.1, we can see how the no-interaction constraint is relaxed in $\mathcal{F}^d_{\mathrm{EM}}$ as well as in $\mathcal{F}^d_{\mathrm{M}}$ (the class of multivariate monotonic functions (3.20)). Functions f in these three classes all satisfy

$$\Delta(f; [\mathbf{a}, \mathbf{b}]) \geq 0, \qquad \text{if } |\{k : a_k \neq b_k\}| = 1. \tag{4.4}$$

Stated more simply, any f in any of the three classes is nondecreasing in each component. How the three classes differ is how they constrain the quasi-volume $\Delta(f; [\mathbf{a}, \mathbf{b}])$ when $\mathbf{a}$ and $\mathbf{b}$ differ by more than one component.

- The additive monotonic class $\mathcal{F}^d_{\mathrm{AM}}$ is component-wise nondecreasing (4.4) and satisfies $\Delta(f; [\mathbf{a}, \mathbf{b}]) = 0$ when $|\{k : a_k \neq b_k\}| > 1$.
- The entirely monotonic class $\mathcal{F}^d_{\mathrm{EM}}$ is component-wise nondecreasing (4.4) and satisfies $\Delta(f; [\mathbf{a}, \mathbf{b}]) \geq 0$ when $|\{k : a_k \neq b_k\}| > 1$.
- The multivariate monotonic class $\mathcal{F}^d_{\mathrm{M}}$ is component-wise nondecreasing (4.4) and imposes no constraint on $\Delta(f; [\mathbf{a}, \mathbf{b}])$ when $|\{k : a_k \neq b_k\}| > 1$.

We see that entirely removing the no-interaction condition $\Delta(f; [\mathbf{a}, \mathbf{b}]) = 0$ leads to the very large class $\mathcal{F}^d_{\mathrm{M}}$, which suffers from the curse of dimensionality (as discussed in Section 3.5). Instead imposing a nonnegativity constraint on these quasi-volumes allows for "positive interactions." To understand the nature of a positive interaction, it is helpful to consider the simple case $d = 2$, where the quasi-volume constraint can be written as

$$f(a_1, b_2) - f(a_1, a_2) \leq f(b_1, b_2) - f(b_1, a_2), \qquad a_1 < b_1,\ a_2 < b_2.$$

Each side of the inequality represents the effect of a change in the second component from a_2 to b_2 while holding the first component fixed. The inequality implies that this effect is larger when the value of the first component is larger, i.e. the first component positively interacts with the second component to produce a larger effect. For higher d, there will also be nonnegativity constraints on higher-order differences. For instance, in the case $d = 3$, there will be constraints similar to the above, as well as the higher-order constraint

$$\begin{aligned} &f(a_1, b_2, b_3) - f(a_1, b_2, a_3) - f(a_1, a_2, b_3) + f(a_1, a_2, a_3) \\ &\leq f(b_1, b_2, b_3) - f(b_1, b_2, a_3) - f(b_1, a_2, b_3) + f(b_1, a_2, a_3), \qquad a_k < b_k, k = 1, 2, 3. \end{aligned}$$

Although these function classes are defined without any smoothness assumptions, it can be helpful to consider the special case of smooth functions and state these constraints in terms of derivatives for the sake of interpretation. All three classes impose the componentwise nondecreasing constraint

$$\frac{\partial f}{\partial x_j} \geq 0, \qquad j = 1, \ldots, d,$$

but differ in how they constraint the mixed derivatives

$$\frac{\partial^{|S|} f}{\prod_{j \in S} \partial x_j}, \qquad S \subseteq \{1, \ldots, d\}, |S| > 1.$$

The additive monotonic class $\mathcal{F}^d_{\mathrm{AM}}$ forces these mixed derivatives to be zero, the entirely monotonic class $\mathcal{F}^d_{\mathrm{EM}}$ imposes a nonnegativity constraint on these mixed derivatives, and the multivariate monotonic class $\mathcal{F}^d_{\mathrm{M}}$ does not impose any constraint on them.

If we revisit the linear regression analogy, the leap from the additive monotonic model (no interactions) to the entirely monotonic model (interactions of all orders) may seem a bit drastic. For instance when $d = 3$, this would be akin to leaping from the multiple linear regression model $f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ directly to

$$f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{1,2} x_1 x_2 + \beta_{1,3} x_1 x_3 + \beta_{2,3} x_2 x_3 + \beta_{1,2,3} x_1 x_2 x_3$$

by including second and third-order linear interactions. In practice, one might consider intermediate models, like only including some or all of the second-order interaction terms, without higher-order interaction terms. Moreover, one can help inform modeling decisions by performing an F-test to compare a simpler null model with an alternative model that includes certain interaction terms [23]. This begs the questions of a) whether we can analogously establish intermediate models between $\mathcal{F}^d_{\mathrm{AM}}$ and $\mathcal{F}^d_{\mathrm{EM}}$ in the monotonic setting, and b) whether we can establish hypothesis tests to compare two such models.

One way to define an intermediate class is to apply the additive structure on blocks of components rather than individual components. Let

$$\mathcal{S} = \{S_1, \ldots, S_M\} \tag{4.5}$$

be a partition of the covariates $\{1, \ldots, d\}$ into M disjoint subsets. Let $d_m \coloneqq |S_m|$. The class $\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}$ consists of functions of the form

$$f(\mathbf{x}) = \sum_{m=1}^{M} f_m(\mathbf{x}_{S_m}), \qquad f_m \in \mathcal{F}^{d_m}_{\mathrm{EM}}, \tag{4.6}$$

where $\mathbf{x}_{S_m}$ denotes the sub-vector of $\mathbf{x}$ indexed by S_m. That is, f is the sum of separate entirely monotone functions on each block S_m of covariates. For example, $\mathcal{F}^{\{\{1,2\},\{3,4,5,6\},\{7,8,9\}\}}_{\mathrm{EM}}$ consists of functions $f : [0,1]^9 \to \mathbb{R}$ of the form

$$f(x_1, \ldots, x_9) = f_1(x_1, x_2) + f_2(x_3, x_4, x_5, x_6) + f_3(x_7, x_8, x_9), \qquad f_1 \in \mathcal{F}^2_{\mathrm{EM}}, f_2 \in \mathcal{F}^4_{\mathrm{EM}}, f_3 \in \mathcal{F}^3_{\mathrm{EM}}.$$

To relate this new class to the earlier ones, note that when $M = 1$, we simply get the EM class $\mathcal{F}^d_{\mathrm{EM}}$, and in general

$$\mathcal{F}^d_{\mathrm{AM}} = \mathcal{F}^{\{\{1\},\ldots,\{d\}\}}_{\mathrm{EM}} \subseteq \mathcal{F}^{\mathcal{S}}_{\mathrm{EM}} \subseteq \mathcal{F}^{\{\{1,\ldots,d\}\}}_{\mathrm{EM}} = \mathcal{F}^d_{\mathrm{EM}}.$$

These classes are suitable in cases when the set of covariates can be partitioned into blocks that do not interact with each other across blocks, but might have positive interactions within each block. The following generalization of Lemma 4.2.1 also holds. (See Section 4.6.4 for the proof.)

Lemma 4.2.2. *The class $\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}$ can be expressed as*

$$\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}} = \{f \in \mathcal{F}^d_{\mathrm{EM}} : \Delta(f; [\mathbf{a}, \mathbf{b}]) = 0 \text{ for } \mathbf{a} \preceq \mathbf{b} \text{ where } \{k : a_k \neq b_k\} \not\subseteq S_m, \forall m\}. \tag{4.7}$$

That is, the quasi-volume over hyperrectangles $[\mathbf{a}, \mathbf{b}]$ is forced to be zero when the components in which $\mathbf{a}$ and $\mathbf{b}$ differ do not lie entirely within one block S_m. For smooth functions, this is equivalent to constraining the mixed derivatives $\frac{\partial^{|S|} f}{\prod_{j \in S} \partial x_j}$ to be nonnegative if $S \subseteq S_m$ for some m, and zero otherwise.

An alternative modeling choice could be to allow interactions across all covariates, but exclude interactions of order exceeding some threshold r. For $1 \leq r \leq d$, we define

$$\mathcal{F}^{d,\leq r}_{\mathrm{EM}} = \{f \in \mathcal{F}^d_{\mathrm{EM}} : \Delta(f; [\mathbf{a}, \mathbf{b}]) = 0 \text{ for } \mathbf{a} \preceq \mathbf{b} \text{ where } |\{k : a_k \neq b_k\}| > r\}.$$

In light of Lemma 4.2.1, we have

$$\mathcal{F}^d_{\mathrm{AM}} = \mathcal{F}^{d,\leq 1}_{\mathrm{EM}} \subseteq \cdots \subseteq \mathcal{F}^{d,\leq r}_{\mathrm{EM}} \subseteq \cdots \subseteq \mathcal{F}^{d,\leq d}_{\mathrm{EM}} = \mathcal{F}^d_{\mathrm{EM}}.$$

In general, this class imposes nonnegativity constraints on the quasi-volumes of order $\leq r$, and forces all higher-order quasi-volumes to be zero. For smooth functions, this is equivalent to constraining the mixed derivatives $\frac{\partial^{|S|} f}{\prod_{j \in S} \partial x_j}$ to be nonnegative when $|S| \leq r$, and zero when $|S| > r$.

4.3 Computational feasibility

In Proposition 3.3.1, we showed how the least squares estimator for the entirely monotonic class $\mathcal{F}^d_{\mathrm{EM}}$ could be computed by solving a nonnegative least squares (NNLS) problem (3.2) with respect to the matrix matrix $\mathbf{A}$ whose columns are the distinct elements of the finite set

$$\mathcal{Q} \equiv \mathcal{Q}_{\mathbf{x}_1,\ldots,\mathbf{x}_n} \coloneqq \{\mathbf{v}(\mathbf{z}) : \mathbf{z} \in [0,1]^d\} \subseteq \{0,1\}^n, \tag{4.8}$$

where

$$\mathbf{v}(\mathbf{z}) \equiv \mathbf{v}_{\mathbf{x}_1,\ldots,\mathbf{x}_n}(\mathbf{z}) \coloneqq (\mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{x}_1), \mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{x}_2), \ldots, \mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{x}_n)).$$

We show below that the LSEs of the intermediate classes $\mathcal{F}_{\mathrm{EM}}^{d,\leq r}$ and $\mathcal{F}_{\mathrm{EM}}^{\mathcal{S}}$, namely

$$\widehat{f}_{\mathrm{EM},\leq r} \in \operatorname*{argmin}_{f \in \mathcal{F}_{\mathrm{EM}}^{d,\leq r}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2 \tag{4.9a}$$

$$\widehat{f}_{\mathrm{EM},\mathcal{S}} \in \operatorname*{argmin}_{f \in \mathcal{F}_{\mathrm{EM}}^{\mathcal{S}}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2 \tag{4.9b}$$

can also be computed by solving a NNLS problem with the same matrix $\mathbf{A}$, but with some coefficients constrained to be zero. In other words, while the LSE with respect to the large class $\mathcal{F}_{\mathrm{EM}}^{d}$ can be computed from performing NNLS with all the columns of $\mathbf{A}$, the LSEs for the intermediate classes can be computed from performing NNLS with a subset of the columns of $\mathbf{A}$.

Let the columns of $\mathbf{A}_{\leq r} \in \mathbb{R}^{n \times p'}$ be the distinct elements of the finite set

$$\{\mathbf{v}(\mathbf{z}) : \mathbf{z} \in [0,1]^d, |\{j : z_j \neq 0\}| \leq r\}.$$

The restriction here forces $\mathbf{z}$ to have at most r nonzero elements. Let the columns of $\mathbf{A}_{\mathcal{S}} \in \mathbb{R}^{n \times p'}$ be the distinct elements of the finite set

$$\{\mathbf{v}(\mathbf{z}) : \mathbf{z} \in [0,1]^d, \{j : z_j \neq 0\} \in \{S_1, \ldots, S_M\}\}.$$

The restriction here forces the nonzero elements of $\mathbf{z}$ to lie entirely within one of the blocks. Without loss of generality assume the first columns of $\mathbf{A}_{\leq r}$ and $\mathbf{A}_{\mathcal{S}}$ are each $\mathbf{v}(\mathbf{0}) = (1, \ldots, 1)$. Note that the columns of $\mathbf{A}_{\leq r}$ and $\mathbf{A}_{\mathcal{S}}$ are all columns of $\mathbf{A}$. We can then consider the NNLS problems

$$\widehat{\boldsymbol{\beta}}_{\mathrm{EM},\leq r} \in \operatorname*{argmin}_{\beta \in \mathbb{R}^{p'} : \beta_j \geq 0, \forall j \geq 2} \|\mathbf{y} - \mathbf{A}_{\leq r}\boldsymbol{\beta}\|^2, \tag{4.10a}$$

$$\widehat{\boldsymbol{\beta}}_{\mathrm{EM},\mathcal{S}} \in \operatorname*{argmin}_{\beta \in \mathbb{R}^{p''} : \beta_j \geq 0, \forall j \geq 2} \|\mathbf{y} - \mathbf{A}_{\mathcal{S}}\boldsymbol{\beta}\|^2. \tag{4.10b}$$

The following result (proved in Section 4.6.6) shows how these NNLS problems can be used to compute the least squares estimators for $\mathcal{F}_{\mathrm{EM}}^{d,\leq r}$ and $\mathcal{F}_{\mathrm{EM}}^{\mathcal{S}}$.

Proposition 4.3.1. *The functions*

$$\widehat{f}_{\mathrm{EM},\leq r} \coloneqq \sum_{j=1}^{p'} (\widehat{\boldsymbol{\beta}}_{\mathrm{EM},\leq r})_j \cdot \mathbb{I}_{[\mathbf{z}_j, \mathbf{1}]}, \qquad \widehat{f}_{\mathrm{EM},\mathcal{S}} \coloneqq \sum_{j=1}^{p''} (\widehat{\boldsymbol{\beta}}_{\mathrm{EM},\mathcal{S}})_j \cdot \mathbb{I}_{[\mathbf{z}_j, \mathbf{1}]}, \tag{4.11}$$

are solutions to the respective optimization problems (4.9a) *and* (4.9b).

Since $\mathbf{A}_{\leq r}$ and $\mathbf{A}_{\mathcal{S}}$ have fewer columns than $\mathbf{A}$, the NNLS problems (4.10a) and (4.10b) are computationally less expensive than the analogous NNLS problem (3.28) for computing

the LSE with respect to $\mathcal{F}_{\text{EM}}^d$. This savings in computation can be quite large. For instance, if the design $\mathbf{x}_1, \dots, \mathbf{x}_n$ is a $n^{1/d} \times \cdots \times n^{1/d}$ lattice, then $\mathbf{A}_{\le r}$ has on the order of $p' \asymp n^{r/d}$ columns and $\mathbf{A}_{\mathcal{S}}$ has on the order of $p'' \asymp n^{d_{\max}/d}$ columns (where $d_{\max} = \max_m |S_m|$ is the size of the largest block). Both these numbers are much smaller than n (the number of columns of $\mathbf{A}$ for this lattice design) especially if r is small and if the blocks of $\mathcal{S}$ are small.

Similar to how the key to proving Proposition 3.3.1 was Proposition 3.3.2, the key to proving Proposition 4.3.1 is the following result that establishes the relationship between the matrices $\mathbf{A}_{\le r}$ and $\mathbf{A}_{\mathcal{S}}$ and the respective classes $\mathcal{F}_{\text{EM}}^{d,\le r}$ and $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$. We defer the proof to Section 4.6.5.

Proposition 4.3.2. *Given n design points $\mathbf{x}_1, \dots, \mathbf{x}_n \in [0,1]^d$, we have*

$$\{(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)) : f \in \mathcal{F}_{\text{EM}}^{d,\le r}\} = \{\mathbf{A}_{\le r}\boldsymbol{\beta} : \beta_j \ge 0, \forall j \ge 2\}, \tag{4.12a}$$

$$\{(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)) : f \in \mathcal{F}_{\text{EM}}^{\mathcal{S}}\} = \{\mathbf{A}_{d_1,\dots,d_M}\boldsymbol{\beta} : \beta_j \ge 0, \forall j \ge 2\}. \tag{4.12b}$$

4.4 Risk bound for $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$

In the rest of this chapter, we will assume the design $\mathbf{x}_1, \dots, \mathbf{x}_n$ is the equally-spaced lattice design

$$\mathbb{L}_{n_1,\dots,n_d} \coloneqq \{(i_1/n_1, \dots, i_d/n_d) : 0 \le i_j \le n_j - 1, j = 1, \dots, d\}. \tag{4.13}$$

Fix a partition $\mathcal{S}$ of $\{1, \dots, d\}$. Given the design $\mathbf{x}_1, \dots, \mathbf{x}_n$ from the lattice (4.13), let $y_1, \dots, y_n$ be obtained from the Gaussian model (4.1) with $f^* \in \mathcal{F}_{\text{EM}}^{\mathcal{S}}$. From these observations, we can the least squares estimator $\widehat{f}_{\text{EM},\mathcal{S}}$ (see Equation 4.9b) with respect to the class $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$. We define the risk by

$$\mathcal{R}(\widehat{f}_{\text{EM},\mathcal{S}}, f^*) \coloneqq \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}(\widehat{f}_{\text{EM},\mathcal{S}}(\mathbf{x}_i) - f^*(\mathbf{x}_i))^2,$$

where the expectation is taken with respect to the the noise in the Gaussian model (4.1). The following result (proved in Section 4.6.1) provides a bound on this risk.

Theorem 4.4.1. *Let $f^*(\mathbf{x}) = \sum_{m=1}^{M} f_m^*(\mathbf{x}_{S_m}) \in \mathcal{F}_{\text{EM}}^{\mathcal{S}}$. Under the lattice design (4.13) and Gaussian model (4.1), the risk of $\widehat{f}_{\text{EM},\mathcal{S}}$ with respect to $f^* \in \mathcal{F}_{\text{EM}}^{\mathcal{S}}$ is bounded as*

$$\mathcal{R}(\widehat{f}_{\text{EM},\mathcal{S}}, f^*)$$
$$\le C \max_{m \in \{1,\dots,M\}} \left[\left(\frac{\sigma^2 V_m^*}{n}\right)^{\frac{2}{3}} \left(\log(2 + \sqrt{n}V_m^*/\sigma)\right)^{\frac{2d_m - 1}{3}} + \frac{\sigma^2}{n}(\log \widetilde{n}_m)^{\frac{3d_m}{4}} (\log(e \log(\widetilde{n}_m)))^{\frac{2d_m - 4}{2}} \right],$$

where $d_m \coloneqq |S_m|$ is the size of the mth block, where $\widetilde{n}_m \coloneqq \prod_{j \in S_m} n_j$ is the size of the lattice for the block S_m, and where $V_m^* \coloneqq f_m^*(\mathbf{1}) - f_m^*(\mathbf{0})$ is the variation of the function f_m^* that governs the S_m block of f^*.

Note that this is a generalization of the entirely monotone risk bound Theorem 3.4.1: when $\mathcal{S} = \{\{1, \ldots, d\}\}$ we have $\mathcal{F}_{\text{EM}}^{\mathcal{S}} = \mathcal{F}_{\text{EM}}^{d}$, $M = 1$, $V_1^* = f(\mathbf{1}) - f(\mathbf{0})$, $\widetilde{n}_1 = n$, and $d_1 = d$, which then yields the EM risk bound (3.43).

The case $\mathcal{S} = \{\{1\}, \ldots, \{d\}\}$ corresponds to additive isotonic regression, in which case the main term of the risk bound is

$$\mathcal{R}(\widehat{f}_{\text{AM}}, f^*) \lesssim \left(\frac{\sigma^2 V_{\max}^*}{n}\right)^{\frac{2}{3}} (\log \sqrt{n} V_{\max}^*/\sigma))^{\frac{1}{3}},$$

where $V_{\max}^* = \max_j \{f_j^*(1) - f_j^*(0)\}$ is the largest variation among the component functions f_j^* of $f^*(\mathbf{x}) = \sum_{j=1}^d f_j^*(x_j)$. For the general case $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$, the dominant term in the bound is essentially $n^{-2/3}(\log n)^{\frac{2d_{\max}-1}{3}}$ where $d_{\max} = \max_m d_m$ is the size of the largest block. Thus the main term in the risk for all of the classes $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$ is $n^{-2/3}$ up to a logarithmic factor, and the only place the partition $\mathcal{S}$ affects this main term is in the exponent of the logarithmic factor, which is $\frac{1}{3}$ in the smallest case $\mathcal{F}_{\text{AM}}^{d}$ and $\frac{2d-1}{3}$ in the largest case $\mathcal{F}_{\text{EM}}^{d}$.

We conjecture that the risk of $\widehat{f}_{\text{EM},\leq r}$ has a main term of the form $n^{-2/3}(\log n)^{O(r)}$, and thus exhibits a similar behavior of having the same rate as the EM class (3.43) except with the exponent of the logarithmic term governed by r rather than d. That the risk takes this form is very plausible given that both $\mathcal{F}_{\text{EM}}^{d,\leq r}$ and $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$ interpolate between the additive monotonic class $\mathcal{F}_{\text{AM}}^{d}$ and the entirely monotonic class $\mathcal{F}_{\text{EM}}^{d}$, but the exact form of the logarithmic term's exponent is not obvious. The proof of Theorem 4.4.1 (see Section 4.6.1) took advantage of the fact that interactions in $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$ are over disjoint subsets of covariates, which allowed us to obtain the above rate despite bounding a Gaussian width rather crudely (see (4.19)). The disjointness was also useful in obtaining an explicit ANOVA decomposition (4.18) involving means of slices of functions in $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$. A proof for the risk of $\widehat{f}_{\text{EM},\leq r}$ would not be able to take advantage of this disjointness because each covariate is allowed to interact with any of the other covariates.

4.5 Hypothesis testing for interactions

In linear regression, the estimator can be interpreted as a projection onto the column space of the design matrix. If we are comparing two nested linear models, the column space of the smaller model $\mathcal{F}_0$ is a subspace of that of the larger model $\mathcal{F}_1$. If we were to test the null hypothesis $H_0 : f^* \in \mathcal{F}_0$ against the alternative $H_a - H_0$ where $H_a : f^* \in \mathcal{F}_1$, one can perform an F-test [23], which is equivalent to a likelihood ratio test for our Gaussian model (4.1).

In light of Proposition 4.3.1 and Proposition 4.3.2, the least squares estimators with respect to classes $\mathcal{F}_{\text{EM}}^{d,\leq r}$ and $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$ are least squares projections onto convex polyhedral cones

$$C_{\text{EM}}^{\leq r} \coloneqq \{\mathbf{A}_{\leq r}\boldsymbol{\beta} : \beta_j \geq 0, \forall j \geq 2\},$$
$$C_{\text{EM}}^{\mathcal{S}} \coloneqq \{\mathbf{A}_{\mathcal{S}}\boldsymbol{\beta} : \beta_j \geq 0, \forall j \geq 2\}.$$

Thus if we were to do a test comparing two nested classes, we would be comparing least squares estimators of two nested cones $C_0 \subseteq C_a$. Let $\Pi_C(\mathbf{v}) \coloneqq \operatorname{argmin}_{\mathbf{u}\in C}\|\mathbf{u}-\mathbf{v}\|^2$ denote the least squares projection onto a cone C. When the nested cones $C_0 \subseteq C_a$ satisfy the condition

$$\Pi_{C_0}(\boldsymbol{\theta}) = \Pi_{C_0}(\Pi_{C_a}(\boldsymbol{\theta})), \quad \forall \boldsymbol{\theta}, \tag{4.14}$$

they are referred to as *non-oblique*. The notion of non-obliqueness was first introduced by Warrack and Robertson [90], and when this condition holds, the corresponding likelihood ratio test (LRT) is well-understood (see [91] and references therein). Unfortunately, for the cones that we are considering, the non-obliqueness condition (4.14) does not necessarily hold. See Section 4.6.2 for an example where $C_{\mathrm{AM}} \subseteq C_{\mathrm{EM}}$ fail to satisfy the condition (4.14). However, we show below that a result of Menéndez et al. [62] allows us to obtain a test that dominates the LRT.

We continue to operate under the assumption that the design $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is the equally-spaced lattice (4.13). Let $\boldsymbol{\theta}^* = (f^*(\mathbf{x}_1), \ldots, f^*(\mathbf{x}_n))$ so that the observations are $\mathbf{y} = \boldsymbol{\theta}^* + \boldsymbol{\xi}$ where $\boldsymbol{\xi} \sim \mathcal{N}(0, \sigma^2)$.

If $C_0 \subseteq C_a$ are nested cones of the form $C_{\mathrm{EM}}^{\leq r}$ or $C_{\mathrm{EM}}^{\mathcal{S}}$, then there are matrices $\mathbf{A}_0$ and $\mathbf{A}_1$ such that the columns of $\mathbf{A}_0$ are columns of $\mathbf{A}_1$, and such that

$$\begin{aligned} C_0 &= \{\mathbf{A}_0\boldsymbol{\beta} : \beta_j \geq 0, \forall j \geq 2\} \\ C_a &= \{\mathbf{A}_1\boldsymbol{\beta} : \beta_j \geq 0, \forall j \geq 2\}. \end{aligned}$$

If we let L_{C_0} be the subspace closure of C_0 (that is, the column space of $\mathbf{A}_0$), then we have

$$C_0 = C_a \cap L_{C_0}.$$

The LRT for testing $H_0 : \boldsymbol{\theta}^* \in C_0$ against $H_a : \boldsymbol{\theta}^* \in C_a$ involves the statistic

$$T(\mathbf{y}) = \|\mathbf{y} - \Pi_{C_0}(\mathbf{y})\|^2 - \|\mathbf{y} - \Pi_{C_a}(\mathbf{y})\|^2.$$

Menéndez et al. [62] showed that if the condition

$$\Pi_{L_{C_0}}(\boldsymbol{\theta}) \in C_a, \quad \forall \boldsymbol{\theta} \in C_a \tag{4.15}$$

holds, then the LRT is dominated by the LRT for a different pair of hypotheses $H_0^* : \boldsymbol{\theta}^* \in L_{C_0}$ and $H_a^* \in C^* \coloneqq \mathrm{cl}(C_a + L_{C_0})$. The relevant statistic for the latter LRT is

$$T^*(\mathbf{y}) = \|\mathbf{y} - \Pi_{L_{C_0}}(\mathbf{y})\|^2 - \|\mathbf{y} - \Pi_{C^*}(\mathbf{y})\|^2.$$

The following result (proved in Section 4.6.3) states that this dominance occurs when the smaller cone C_0 is of the form $C_{\mathrm{EM}}^{\mathcal{S}}$.

Theorem 4.5.1. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the lattice design (4.13). Let $C_0 \subseteq C_a$ where $C_0 = C_{\mathrm{EM}}^{\mathcal{S}}$, and where C_a is either of the form $\mathcal{F}_{\mathrm{EM}}^{\mathcal{S}'}$ or $\mathcal{F}_{\mathrm{EM}}^{d,\leq r}$. The LRT for testing $H_0 : \boldsymbol{\theta} \in C_0$ against the alternative $H_a - H_0$ where $H_a : \boldsymbol{\theta} \in C_a$ is dominated by the LRT for testing $H_0^* : \boldsymbol{\theta}^* \in L_{C_0}$*

against $H_a^* - H_0^*$ *where* $H_a^* : \boldsymbol{\theta}^* \in C^* := \mathrm{cl}(C_a + L_{C_0})$*, in the sense that it has the same significance level*

$$\sup_{\boldsymbol{\theta}^* \in C_0} \mathbb{P}_{\boldsymbol{\theta}^*}(T(\mathbf{y}) \geq c) = \sup_{\boldsymbol{\theta}^* \in L_{C_0}} \mathbb{P}_{\boldsymbol{\theta}^*}(T^*(\mathbf{y}) \geq c)$$

and has at least as much power, i.e.

$$\mathbb{P}_{\boldsymbol{\theta}^*}(T(\mathbf{y}) \geq c) \leq \mathbb{P}_{\boldsymbol{\theta}^*}(T^*(\mathbf{y}) \geq c), \qquad \forall \boldsymbol{\theta}^*, \forall c.$$

One advantage of the new test with hypotheses H_0^* and H_a^* is that the statistic T^* is invariant under translations of $\boldsymbol{\theta}^*$ within the subspace L_{C_0}. In particular, the worst-case Type I error $\sup_{\boldsymbol{\theta}^* \in L_{C_0}} \mathbb{P}_{\boldsymbol{\theta}^*}(T^*(\mathbf{y}) \geq c)$ is attained at any $\boldsymbol{\theta}^* \in L_{C_0}$, such as $\boldsymbol{\theta}^* = \mathbf{0}$. This fact is useful in determining an appropriate cutoff c for the original LRT involving T.

Theorem 4.5.1 can be used to test for whether certain interactions should be included in a model. For instance, if one wanted to test whether the true function has no interactions ($H_0 : f^* \in \mathcal{F}^d_{\mathrm{AM}}$) against the alternative $H_a : f^* \in \mathcal{F}^d_{\mathrm{EM}}$, the LRT for this test would be dominated by the LRT for hypotheses

$$H_0^* : f^* \in \{f : \Delta(f; [\mathbf{a}, \mathbf{b}]) = 0 \text{ if } |\{k : a_k \neq b_k\}| > 1\} = \left\{ f : f(\mathbf{x}) = \sum_{j=1}^{d} f_j(x_j) \right\}$$

$$H_a^* : f^* \in \{f : \Delta(f; [\mathbf{a}, \mathbf{b}]) \geq 0 \text{ if } |\{k : a_k \neq b_k\}| > 1\}.$$

More generally, Theorem 4.5.1 can be used for any pair of nested models of the form

- $\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}} \subseteq \mathcal{F}^{\mathcal{S}'}_{\mathrm{EM}}$ where the sets of $\mathcal{S}'$ are unions of sets in $\mathcal{S}$, or
- $\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}} \subseteq \mathcal{F}^{d,\leq r}_{\mathrm{EM}}$ where $r \geq \max_m |S_m|$.

There are a few limitations of Theorem 4.5.1. It is unclear whether the same result holds when the smaller cone is of the other form $C^{\leq r}_{\mathrm{EM}}$. The proof of Theorem 4.5.1 again takes advantage of the disjointness of the interactions in $\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}$ to get an explicit expression for the linear projection $\Pi_{L_{C_0}}$ via an ANOVA decomposition. A separate drawback of Theorem 4.5.1 is its reliance on the lattice design (4.13). The condition (4.15) does not necessarily hold for non-lattice designs, so it is unclear whether this auxiliary LRT involving the statistic T^* has any bearing on the original LRT in the general case.

4.6 Proofs

4.6.1 Proof of Theorem 4.4.1

Without loss of generality, we may assume $\sigma = 1$. (The general case can be obtained by multiplying the estimation risk in the scaled model $y_i/\sigma = f^*(\mathbf{x}_i)/\sigma + \xi_i/\sigma$ and then multiplying by σ^2.)

Let $C_{\mathrm{EM}}^{\mathcal{S}} := \{(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) : f \in \mathcal{F}_{\mathrm{EM}}^{\mathcal{S}}\}$. Since the design is a lattice, the full EM matrix $\mathbf{A}$ (see (4.8)) is square with columns $(\mathbb{I}_{[\mathbf{x}_i,\mathbf{1}]}(\mathbf{x}_1), \ldots \mathbb{I}_{[\mathbf{x}_i,\mathbf{1}]}(\mathbf{x}_n))$ for $i = 1, \ldots, n$. By Proposition 4.3.2, $C_{\mathrm{EM}}^{\mathcal{S}}$ is exactly the set of vectors of the form $\mathbf{A}\boldsymbol{\beta}$ where $\boldsymbol{\beta} \in \mathbb{R}^n$ such that $\boldsymbol{\beta}_j \geq 0$ for $j \geq 2$ and $\beta_j = 0$ if $\{i : (\mathbf{x}_j)_i \neq 0\}$ is not contained within one block S_m. (The coefficients being forced to be zero correspond to the columns of $\mathbf{A}$ that are not columns of $\mathbf{A}_{\mathcal{S}}$.) Note that $\widetilde{n}_1 + \cdots + \widetilde{n}_M - (M-1)$ of the β_j are not constrained to be zero.

If we let $\boldsymbol{\theta}^* := (f^*(\mathbf{x}_1), \ldots, f^*(\mathbf{x}_n))$ and $\widehat{\boldsymbol{\theta}} := (\widehat{f}_{\mathrm{EM},\mathcal{S}}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{EM},\mathcal{S}}(\mathbf{x}_n))$ then

$$\widehat{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta} \in C_{\mathrm{EM}}^{\mathcal{S}}} \|\mathbf{y} - \boldsymbol{\theta}\|^2.$$

Thus, we have $\mathcal{R}(\widehat{f}_{\mathrm{EM},\mathcal{S}}, f^*) = \frac{1}{n}\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2$.

Because the design $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is the lattice defined earlier (4.13), it is convenient to treat elements $\boldsymbol{\theta} \in C_{\mathrm{EM}}^{\mathcal{S}}$ not as vectors in $\mathbb{R}^n$, but as arrays in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ and index their elements with integer vectors

$$\mathbf{i} = (i_1, \ldots, i_d) \in \mathcal{I} := \prod_{j=1}^{d} \{0, 1, \ldots, n_j - 1\}$$

so that, for instance $\theta_{\mathbf{i}}^* = f^*(i_1/n_1, \ldots, i_d/n_d)$. From the above representation of $\boldsymbol{\theta} \in C_{\mathrm{EM}}^{\mathcal{S}}$ as linear combinations of columns of $\mathbf{A}$, we can also similarly treat the nonnegative coefficients $\boldsymbol{\beta} \in \mathbb{R}^n$ as arrays $\mathbb{R}^{n_1 \times \cdots \times n_d}$, and we then obtain the direct relationship

$$\theta_{\mathbf{i}} = \sum_{\mathbf{k} \preceq \mathbf{i}} \beta_{\mathbf{k}}.$$

Again, recall from the definition of $C_{\mathrm{EM}}^{\mathcal{S}}$ that $\beta_{\mathbf{k}} = 0$ if $\mathbf{k}$ has nonzero components in more than one block. Thus if we let $(\mathbf{i}_{S_m}, \mathbf{0}_{S_m^c})$ denote the integer vector whose components in S_m equal $\mathbf{i}_{S_m}$, and whose all other components are zero, we have

$$\theta_{\mathbf{i}} = \sum_{\substack{\mathbf{k} \preceq \mathbf{i}, \\ \exists m : \{j : \mathbf{k}_j \neq 0\} \subseteq S_m}} \beta_{\mathbf{k}} = \beta_{\mathbf{0}} + \sum_{m=1}^{M} \sum_{\substack{\mathbf{k} \preceq \mathbf{i}, \mathbf{k} \neq \mathbf{0}, \\ \{j : \mathbf{k}_j \neq 0\} \subseteq S_m}} \beta_{\mathbf{k}} = \left(\sum_{m=1}^{M} \theta_{(\mathbf{i}_{S_m}, \mathbf{0}_{S_m^c})} \right) - (d-1)\theta_{\mathbf{0}}, \tag{4.16}$$

where the last step is due to $\beta_{\mathbf{0}} = \theta_{\mathbf{0}}$ and

$$\beta_{\mathbf{0}} + \sum_{\substack{\mathbf{k} \preceq \mathbf{i}, \mathbf{k} \neq \mathbf{0}, \\ \{j : \mathbf{k}_j \neq 0\} \subseteq S_m}} \beta_{\mathbf{k}} = \theta_{(\mathbf{i}_{S_m}, \mathbf{0}_{S_m^c})}.$$

Since $\widehat{\boldsymbol{\theta}}$ is a least squares estimator, we can use the result of Chatterjee [17] (see Theorem A.3.2) to bound the risk. Thus our goal is to bound

$$G(t) = \mathbb{E} \sup_{\boldsymbol{\theta} \in C_{\mathrm{EM}}^{\mathcal{S}} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq t} \langle Z, \boldsymbol{\theta} - \boldsymbol{\theta}^* \rangle, \tag{4.17}$$

where the expectation is with respect to the random array Z with i.i.d. standard Gaussian random variables.

Fix $\boldsymbol{\theta} \in C^{\mathcal{S}}_{\mathrm{EM}}$. We define $\mu \coloneqq \frac{1}{n}\sum_{\mathbf{i}} \theta_{\mathbf{i}}$ as well as, for $\mathbf{i}_{S_m} \in \mathcal{I}_m \coloneqq \prod_{j \in S_m} \{0, \dots, n_j - 1\}$,

$$\mu_{m,\mathbf{i}_{S_m}} \coloneqq \frac{1}{n/\widetilde{n}_m} \sum_{\mathbf{i}' : \mathbf{i}'_{S_m} = \mathbf{i}_{S_m}} \theta_{\mathbf{i}'}$$

$$\alpha_{m,\mathbf{i}_{S_m}} \coloneqq \mu_{m,\mathbf{i}_{S_m}} - \mu.$$

Then we claim

$$\theta_{\mathbf{i}} = \mu + \sum_{m=1}^{M} \alpha_{m,\mathbf{i}_{S_m}}, \quad \forall \mathbf{i} \in \mathcal{I}. \tag{4.18}$$

Indeed, combining the definitions of μ and $\mu_{m,\mathbf{i}_{S_m}}$ with the additive decomposition of $\theta_{\mathbf{i}}$ (4.16) yields

$$\begin{aligned}
\mu &= \frac{1}{n} \sum_{\mathbf{i} \in \mathcal{I}} \left[\left(\sum_{m=1}^{M} \theta_{(\mathbf{i}_{S_m}, \mathbf{0}_{S_m^c})} \right) - (d-1)\theta_{\mathbf{0}} \right] \\
&= -(d-1)\theta_{\mathbf{0}} + \sum_{m=1}^{M} \frac{1}{\widetilde{n}_m} \sum_{\mathbf{i}_{S_m}} \theta_{(\mathbf{i}_{S_m}, \mathbf{0}_{S_m^c})}
\end{aligned}$$

and

$$\begin{aligned}
\mu_{m,\mathbf{i}_{S_m}} &= \frac{1}{n/\widetilde{n}_m} \sum_{\mathbf{i}' : \mathbf{i}'_{S_m} = \mathbf{i}_{S_m}} \left[\left(\sum_{m=1}^{M} \theta_{(\mathbf{i}'_{S_m}, \mathbf{0}_{S_m^c})} \right) - (d-1)\theta_{\mathbf{0}} \right] \\
&= \theta_{(\mathbf{i}_{S_m}, \mathbf{0}_{S_m^c})} - (d-1)\theta_{\mathbf{0}} + \sum_{m' \neq m} \frac{1}{\widetilde{n}_{m'}} \sum_{\mathbf{i}_{S'_m} \in \mathcal{I}_m} \theta_{(\mathbf{i}_{S_{m'}}, \mathbf{0}_{S_{m'}^c})},
\end{aligned}$$

and thus

$$\begin{aligned}
\mu + \sum_{m=1}^{M} \alpha_{m,\mathbf{i}_{S_m}} &= -(M-1)\mu + \sum_{m=1}^{M} \mu_{m,\mathbf{i}_{S_m}} \\
&= -(M-1)\mu + \underbrace{\left(-(d-1)\theta_{\mathbf{0}} + \sum_{m=1}^{M} \theta_{(\mathbf{i}_{S_m}, \mathbf{0}_{S_m^c})} \right)}_{=\theta_{\mathbf{i}}} \\
&\qquad - (M-1)(d-1)\theta_{\mathbf{0}} + \sum_{m=1}^{M} \sum_{m' \neq m} \frac{1}{\widetilde{n}_{m'}} \sum_{\mathbf{i}_{S_{m'}} \in \mathcal{I}_m} \theta_{(\mathbf{i}_{S_{m'}}, \mathbf{0}_{S_{m'}^c})} \\
&= \theta_{\mathbf{i}} - (M-1)\mu + (M-1) \left(-(d-1)\theta_{\mathbf{0}} + \sum_{m'=1}^{M} \frac{1}{\widetilde{n}_{m'}} \sum_{\mathbf{i}_{S_{m'}} \in \mathcal{I}_m} \theta_{(\mathbf{i}_{S_{m'}}, \mathbf{0}_{S_{m'}^c})} \right) \\
&= \theta_{\mathbf{i}}.
\end{aligned}$$

Note that so far we have only used the fact that $\boldsymbol{\theta}$ is in the column space of $\mathbf{A}_{\mathcal{S}}$, and have not needed to use the fact that the coefficients need to be nonnegative.

Let the quantities μ^* and $\alpha^*_{m,\mathbf{i}_{S_m}}$ be analogous to μ and $\alpha_{m,\mathbf{i}_{S_m}}$, but for $\boldsymbol{\theta}^*$ instead of $\boldsymbol{\theta}$. By the above decomposition (4.18), the sum of squares in the Gaussian width (4.17) decomposes as

$$\begin{aligned}\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|^2 &= \sum_{\mathbf{i}\in\mathcal{I}}(\theta_{\mathbf{i}}-\theta^*_{\mathbf{i}})^2 \\ &= \sum_{\mathbf{i}\in\mathcal{I}}\left((\mu-\mu^*)+\sum_{m=1}^{M}(\alpha_{m,\mathbf{i}_{S_m}}-\alpha^*_{m,\mathbf{i}_{S_m}})\right)^2 \\ &= n(\mu-\mu^*)^2+\sum_{m=1}^{M}\frac{n}{\widetilde{n}_m}\sum_{\mathbf{i}_{S_m}\in\mathcal{I}_m}(\alpha_{m,\mathbf{i}_{S_m}}-\alpha^*_{m,\mathbf{i}_{S_m}})^2,\end{aligned}$$

where the cross terms cancel due to orthogonality. Similarly, the inner product in the Gaussian width (4.17) can be decomposed as

$$\begin{aligned}\langle Z,\boldsymbol{\theta}-\boldsymbol{\theta}^*\rangle &= \sum_{\mathbf{i}\in\mathcal{I}}Z_{\mathbf{i}}\left((\mu-\mu^*)+\sum_{m=1}^{M}(\alpha_{m,\mathbf{i}_{S_m}}-\alpha^*_{m,\mathbf{i}_{S_m}})\right) \\ &= \left(\sum_{\mathbf{i}\in\mathcal{I}}Z_{\mathbf{i}}\right)(\mu-\mu^*)+\sum_{m=1}^{M}\sum_{\mathbf{i}_{S_m}\in\mathcal{I}_m}\left(\sum_{\mathbf{k}:\mathbf{k}_{S_m}=\mathbf{i}_{S_m}}Z_{\mathbf{k}}\right)(\alpha_{m,\mathbf{i}_{S_m}}-\alpha^*_{m,\mathbf{i}_{S_m}}).\end{aligned}$$

Note that if we view $\boldsymbol{\alpha}_m=(\alpha_{m,\mathbf{i}_{S_m}})_{\mathbf{i}_{S_m}\in\mathcal{I}_m}$ as an array, it belongs to the entirely monotone class

$$C^{d_m}_{\mathrm{EM}}\equiv C^{\{1,\dots,d_m\}}_{\mathrm{EM}}=\{(f(\mathbf{x}_1),\dots,f(\mathbf{x}_n)):f\in\mathcal{F}^{d_m}_{\mathrm{EM}}\}$$

because it is obtained by averaging many S_m-slices of $f\in\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}$ (each of which is in $\mathcal{F}^{d_m}_{\mathrm{EM}}$ by definition of $\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}$).

Thus, we can bound the Gaussian width by

$$G(t)\le\mathbb{E}\sup_{\mu:|\mu-\mu^*|\le t/\sqrt{n}}\left(\sum_{\mathbf{i}\in\mathcal{I}}Z_{\mathbf{i}}\right)(\mu-\mu^*)+\sum_{m=1}^{M}G_m(t), \tag{4.19}$$

where

$$G_m(t):=\mathbb{E}\sup_{\substack{\boldsymbol{\alpha}_m\in C^{d_m}_{\mathrm{EM}}:\\ \|\boldsymbol{\alpha}_m-\boldsymbol{\alpha}^*_m\|\le t/\sqrt{n/\widetilde{n}_m}}}\sum_{\mathbf{i}_{S_m}\in\mathcal{I}_m}\left(\sum_{\mathbf{k}:\mathbf{k}_{S_m}=\mathbf{i}_{S_m}}Z_{\mathbf{k}}\right)(\alpha_{m,\mathbf{i}_{S_m}}-\alpha^*_{m,\mathbf{i}_{S_m}}).$$

The first term can be bounded easily. Since $\sum_{\mathbf{i}}Z_{\mathbf{i}}$ is equal in distribution to $\sqrt{n}U$ where U is a standard Gaussian random variables, we have

$$\mathbb{E}\sup_{\mu:|\mu-\mu^*|\le t/\sqrt{n}}\left(\sum_{\mathbf{i}\in\mathcal{I}}Z_{\mathbf{i}}\right)(\mu-\mu^*)=\sqrt{n}\mathbb{E}\sup_{\mu:|\mu-\mu^*|\le t/\sqrt{n}}U(\mu-\mu^*)t\mathbb{E}|U|=Ct.$$

We now turn to bounding the $G_m(t)$. Note that the difference between the largest and smallest values of $\boldsymbol{\alpha}_m^*$ is bounded by $V_m^* = f_m^*(\mathbf{1}) - f_m^*(\mathbf{0})$. Since $\sum_{\mathbf{k}:\mathbf{k}_{S_m}=\mathbf{i}_{S_m}} Z_{\mathbf{k}}$ is equal in distribution to $\sqrt{n/\widetilde{n}_m}U$ where $U = (U_{\mathbf{i}})_{\mathbf{i}\in\mathcal{I}_m}$ is an array with i.i.d. standard Gaussian entries, we are able to apply the Gaussian width bound for the EM class (A.18) to obtain

$$\begin{aligned}
G_m(t) &= \sqrt{\frac{n}{\widetilde{n}_m}}\mathbb{E} \sup_{\substack{\boldsymbol{\alpha}_m\in C_{\mathrm{EM}}^{d_m}:\\ \|\boldsymbol{\alpha}_m-\boldsymbol{\alpha}_m^*\|\le t/\sqrt{\frac{n}{\widetilde{n}_m}}}} \langle U, \boldsymbol{\alpha}_m - \boldsymbol{\alpha}_m^*\rangle \\
&\le C_{d_m}\sqrt{\frac{n}{\widetilde{n}_m}}\Big(tV_m^*/\sqrt{n/\widetilde{n}_m}\Big)^{1/2}\widetilde{n}_m^{1/4}\left(\log_+\frac{eV_m^*\sqrt{\widetilde{n}_m}}{t/\sqrt{n/\widetilde{n}_m}}\right)^{\frac{2d_m-1}{4}} \\
&\qquad + C_{d_m}\sqrt{\frac{n}{\widetilde{n}_m}}\cdot\frac{1}{\sqrt{n/\widetilde{n}_m}}t(\log\widetilde{n}_m)^{\frac{3d_m}{4}}(\log(e\log\widetilde{n}_m))^{\frac{2d_m-1}{4}} \\
&=: G_{m,1}(t) + G_{m,2}(t).
\end{aligned}$$

Let

$$t_{m,1} := \max\{1, C'_{d_m,M}\}(\sqrt{n}V_m^*)^{1/3}\big[\max\{1,\log_+(e(\sqrt{n}V_m^*)^{2/3}\}\big]^{\frac{2d_m-1}{6}}.$$

If $t \ge t_{m,1}$, then

$$\begin{aligned}
\frac{G_{m,1}(t)}{t^2} &= C_{d_m}(\sqrt{n}V_m^*)^{1/2}t^{-\frac{3}{2}}\left(\log_+\frac{eV_m^*\sqrt{n}}{t}\right)^{\frac{2d_m-1}{4}} \\
&\le C_{d_m}(\sqrt{n}V_m^*)^{1/2}t^{-\frac{3}{2}}\big(\log_+(e(\sqrt{n}V_m^*)^{2/3})\big)^{\frac{2d_m-1}{4}} \\
&\le \frac{1}{8M}.
\end{aligned}$$

Let

$$t_{m,2} := C''_{d_m,M}(\log\widetilde{n}_m)^{\frac{3d_m}{4}}(\log(e\log\widetilde{n}_m))^{\frac{2d_m-1}{4}}.$$

If $t \ge t_{m,2}$, then

$$\frac{G_{m,2}(t)}{t^2} = C_{d_m}t^{-1}(\log\widetilde{n}_m)^{\frac{3d_m}{4}}(\log(e\log\widetilde{n}_m))^{\frac{2d_m-1}{4}} \le \frac{1}{8M}.$$

Thus if

$$t_* = \max\{4C, \max_{1\le m\le M} t_{m,1}, \max_{1\le m\le M} t_{m,2}\}$$

we have

$$G(t_*) \le Ct_* + \sum_{m=1}^M G_m(t_*) \le \frac{t_*^2}{4} + \sum_{m=1}^M\left(\frac{t_*^2}{8M} + \frac{t_*^2}{8M}\right) \le \frac{t_*^2}{2}.$$

Applying Theorem A.3.2 implies

$$\frac{1}{n}\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leq \frac{t_*^2}{2n}$$
$$\leq C \max_{m\in\{1,\ldots,M\}} \left[\left(\frac{V_m^*}{n}\right)^{\frac{2}{3}} \left(\log(2+\sqrt{n}V_m^*)\right)^{\frac{2d_m-1}{3}} + \frac{1}{n}(\log \widetilde{n}_m)^{\frac{3d_m}{4}} (\log(e\log(\widetilde{n}_m)))^{\frac{2d_m-4}{2}} \right]$$

where the constant depends only on $d_1, \ldots, d_M$ and M.

4.6.2 Example of nested cones violating the non-obliqueness condition

Let $d = 2$ and $n_1 = n_2 = 2$, and consider the lattice design (4.13). Then

$$C_{\mathrm{AM}} = \{\boldsymbol{\theta} \in \mathbb{R}^{2\times 2} : \theta_{1,1} - \theta_{1,0} - \theta_{0,1} + \theta_{0,0} = 0, \theta_{1,0} \geq \theta_{0,0}, \theta_{0,1} \geq \theta_{0,0}\}$$
$$C_{\mathrm{EM}} = \{\boldsymbol{\theta} \in \mathbb{R}^{2\times 2} : \theta_{1,1} - \theta_{1,0} - \theta_{0,1} + \theta_{0,0} \geq 0, \theta_{1,0} \geq \theta_{0,0}, \theta_{0,1} \geq \theta_{0,0}\}.$$

However, if we let $\boldsymbol{\theta} = \begin{bmatrix} 0 & -6 \\ -6 & 0 \end{bmatrix}$, then

$$\Pi_{C_{\mathrm{AM}}}(\theta) = \begin{bmatrix} -3 & -3 \\ -3 & -3 \end{bmatrix}, \qquad \Pi_{C_{\mathrm{EM}}}(\theta) = \begin{bmatrix} -4 & -4 \\ -4 & 0 \end{bmatrix}, \qquad \Pi_{C_{\mathrm{AM}}}(\Pi_{C_{\mathrm{EM}}}(\theta)) = \begin{bmatrix} -5 & -3 \\ -3 & -1 \end{bmatrix},$$

which violates the non-obliqueness condition (4.14).

4.6.3 Proof of Theorem 4.5.1

The result follows directly from applying Theorem 2.1 of Menéndez et al. [62], provided we check the condition (4.15). To check the condition, we provide an explicit formula for the projection $\Pi_{L_{C_0}}$.

Let $\mathcal{I} \coloneqq \prod_{j=1}^d \{0, 1, \ldots, n_j - 1\}$ and $\mathcal{I}_m \coloneqq \prod_{j\in S_m} \{0, \ldots, n_j - 1\}$. Given $\boldsymbol{\theta} \in C_a$, we define $\mu \coloneqq \frac{1}{n}\sum_{\mathbf{i}} \theta_{\mathbf{i}}$ as well as, for $\mathbf{i}_{S_m} \in \mathcal{I}_m \coloneqq \prod_{j\in S_m}\{0, \ldots, n_j - 1\}$,

$$\mu_{m,\mathbf{i}_{S_m}} \coloneqq \frac{1}{n/\widetilde{n}_m} \sum_{\mathbf{i}' : \mathbf{i}'_{S_m} = \mathbf{i}_{S_m}} \theta_{\mathbf{i}'}$$
$$\alpha_{m,\mathbf{i}_{S_m}} \coloneqq \mu_{m,\mathbf{i}_{S_m}} - \mu.$$

Let $\boldsymbol{\theta}' \in L_{C_0}$, and let μ', $\mu'_{m,\mathbf{i}_{S_m}}$, and $\alpha'_{m,\mathbf{i}_{S_m}}$ be defined analogously. We showed earlier (4.18) that because $\boldsymbol{\theta}' \in L_{C_0}$, we have

$$\theta'_{\mathbf{i}} = \mu' + \sum_{m=1}^M \alpha'_{m,\mathbf{i}_{S_m}}, \quad \forall \mathbf{i} \in \mathcal{I}. \tag{4.20}$$

We know from the definition of $\mathbf{A}_{\mathcal{S}}$ (see Section 4.3) that there is a column in $\mathbf{A}_{\mathcal{S}}$ for every element of the lattice design with nonzero components entirely within a single block S_m. By counting the number of such design points, we have $\dim L_{C_0} = \widetilde{n}_1 + \cdots + \widetilde{n}_M - (M-1)$ where $\widetilde{n}_m \coloneqq \prod_{j \in S_m} n_j$ is the size of the lattice in block S_m. But the decomposition (4.20) also has the same number of degrees of freedom: there is 1 degree of freedom for parameter μ', and for each block there are $\widetilde{n} - 1$ more degrees of freedom (where the -1 is due to the constraint that $\sum_{\mathbf{i}_{S_m} \in \mathcal{I}_m} \alpha'_{m,\mathbf{i}_{S_m}} = 0$). Thus, we can parameterize the subspace L_{C_0} using these parameters (4.20) instead.

Using orthogonality, we obtain

$$\begin{aligned}
\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2 &= \sum_{\mathbf{i} \in \mathcal{I}} (\theta_{\mathbf{i}} - \theta'_{\mathbf{i}})^2 \\
&= \sum_{\mathbf{i} \in \mathcal{I}} \left[(\mu - \mu') + \sum_{m=1}^{M} (\alpha_{m,\mathbf{i}_{S_m}} - \alpha'_{m,\mathbf{i}_{S_m}}) + \left(\theta_{\mathbf{i}} - \mu - \sum_{m=1}^{M} \alpha_{m,\mathbf{i}_{S_m}} \right) \right]^2 \\
&= n(\mu - \mu')^2 + \sum_{m=1}^{M} \frac{n}{\widetilde{n}_m} (\alpha_{m,\mathbf{i}_{S_m}} - \alpha'_{m,\mathbf{i}_{S_m}})^2 + \sum_{\mathbf{i} \in \mathcal{I}} \left(\theta_{\mathbf{i}} - \mu - \sum_{m=1}^{M} \alpha_{m,\mathbf{i}_{S_m}} \right)^2 .
\end{aligned}$$

Thus, by focusing on the first two terms, we see that the choice of $\boldsymbol{\theta}' \in L_{C_0}$ that minimizes $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2$ is

$$\Pi_{L_{C_0}}(\boldsymbol{\theta}) = \mu + \sum_{m=1}^{M} \alpha_{m,\mathbf{i}_{S_m}}. \tag{4.21}$$

Now that we have an explicit formula for $\Pi_{L_{C_0}}$, we return to checking the condition (4.15) by arguing that $\Pi_{L_{C_0}}(\boldsymbol{\theta}) \in C_a$. Since $\Pi_{L_{C_0}}(\boldsymbol{\theta}) \in L_{C_0}$ by definition, this is equivalent to $\Pi_{L_{C_0}}(\boldsymbol{\theta}) \in L_{C_0} \cap C_a = C^{\mathcal{S}}_{\mathrm{EM}} = \{(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) : f \in \mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}\}$, so we only need to check the nonnegativity constraint of $\mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}$ within each block S_m. To see that the nonnegativity constraint within block S_m holds, note that from the formula (4.21) the only addend that varies within block S_m is $\alpha_{m,\mathbf{i}_{S_m}}$. But this term is obtained by taking an average of S_m-slices of $\boldsymbol{\theta}$, each of which satisfies the shape constraint within S_m, so $\boldsymbol{\theta}$ does as well.

4.6.4 Proof of Lemmas 4.2.1 and 4.2.2

Lemma 4.2.1 is a special case of Lemma 4.2.2, so it suffices to prove Lemma 4.2.2.

Suppose $f \in \mathcal{F}^{\mathcal{S}}_{\mathrm{EM}}$ and have the decomposition into entirely monotonic functions $f_1, \ldots, f_M$ given earlier (4.6). Suppose $K \coloneqq \{k : a_k \neq b_k\}$, the differing components of $\mathbf{a}$ and $\mathbf{b}$, lie entirely within a block, i.e. $K \subseteq S_m$ for some m. Recalling the notation $J_k \coloneqq \mathbb{I}\{a_k \neq b_k\}$ from the definition of quasi-volume (4.3), we see that the quasi-volume $\Delta(f; [\mathbf{a}, \mathbf{b}])$ reduces to a lower-dimensional quasi-volume $\Delta(\widetilde{f}; [\mathbf{a}_K, \mathbf{b}_K])$ where $\mathbf{a}_K$ and $\mathbf{b}_K$ are the sub-vectors of $\mathbf{a}$ and $\mathbf{b}$ indexed by K, and where $\widetilde{f} : [0,1]^{|K|} \to \mathbb{R}$ is obtained by plugging in $a_k = b_k$ for the kth covariate where $k \notin K$, and leaving the other covariates as

inputs. Since $K \subseteq S_m$, we have $\Delta(f; [\mathbf{a}, \mathbf{b}]) = \Delta(\widetilde{f}; [\mathbf{a}_K, \mathbf{b}_K]) = \Delta(f_m; [\mathbf{a}_{S_m}, \mathbf{b}_{S_m}]) \geq 0$ since $f_m \in \mathcal{F}^d_{\mathrm{EM}}$ by assumption.

Otherwise, suppose K intersects more than one of the blocks S_m. Without loss of generality, suppose $1 \in S_1$ and $2 \in S_2$ and $\{1, 2\} \subseteq K$. For any fixed values of $j_3, \ldots, j_d \in \{0, 1\}$, we have

$$\begin{aligned}
&\sum_{j_1=0}^{1} \sum_{j_2=0}^{1} (-1)^{j_1+\cdots+j_d} f(b_1 + j_1(a_1 - b_1), b_2 + j_2(a_2 - b_2), \ldots, b_d + j_d(a_d - b_d)) \\
&= f(b_1, b_2, b_3 + j_3(a_3 - b_3), \ldots, b_d + j_d(a_d - b_d)) \\
&\quad - f(b_1, a_2, b_3 + j_3(a_3 - b_3), \ldots, b_d + j_d(a_d - b_d)) \\
&\quad - f(a_1, b_2, b_3 + j_3(a_3 - b_3), \ldots, b_d + j_d(a_d - b_d)) \\
&\quad - f(a_1, a_2, b_3 + j_3(a_3 - b_3), \ldots, b_d + j_d(a_d - b_d)) \\
&= (\widetilde{f}_1(b_1) + \widetilde{f}_2(b_2)) - (\widetilde{f}_1(b_1) + \widetilde{f}_2(a_2)) - (\widetilde{f}_1(a_1) + \widetilde{f}_2(b_2)) + (\widetilde{f}_1(a_1) + \widetilde{f}_2(a_2)) \\
&= 0.
\end{aligned}$$

(Above, $\widetilde{f}_1$ is the univariate function obtained by plugging in $b_k + j_k(a_k - b_k)$ into f_1 for $k \in S_1 \setminus \{1\}$; $\widetilde{f}_2$ is defined similarly.) Summing over $j_3, \ldots, d_d$ yields $\Delta(f; [\mathbf{a}, \mathbf{b}]) = 0$ in this case.

We now show the reverse inclusion. Suppose f belongs to the right-hand side (4.7). We define $f_m : [0, 1]^{|S_m|} \to \mathbb{R}$ by

$$f_m(\mathbf{u}) \coloneqq f((\mathbf{u}, \mathbf{0}_{S_m^c})) - f(\mathbf{0}). \tag{4.22}$$

The first term on the right-hand side is obtained by plugging into f the inputs $\mathbf{u}$ for the covariates indexed by S_m, and zero for the other covariates. Because $f \in \mathcal{F}^{\mathcal{S}}_{\mathrm{EM}} \subseteq \mathcal{F}^d_{\mathrm{EM}}$, these f_m are each in $\mathcal{F}^{|S_m|}_{\mathrm{EM}}$. Thus it suffices to show

$$f(\mathbf{x}) = f(\mathbf{0}) + \sum_{m=1}^{M} f_m(\mathbf{x}_{S_m})$$

for $\mathbf{x} \in [0, 1]^d$. (The additive constant $f(\mathbf{0})$ can be absorbed by one of the f_m.) We can verify this claim (4.22) by strong induction on the number of nonzero components of $\mathbf{x}$. If the nonzero components of $\mathbf{x}$ lie within a single block S_m, then the claim holds immediately, since $f(\mathbf{0}) + \sum_{m=1}^{M} f_m(\mathbf{x}_m) = f(\mathbf{0}) + f_m(\mathbf{x}_m) = f(\mathbf{x})$. Now suppose the nonzero components $\{k : x_k \neq 0\}$ intersect with more than one block. By assumption (4.7), $\Delta(f; [\mathbf{0}, \mathbf{x}]) = 0$. From the definition (4.3), this quasi-volume is an alternating sum of f evaluated at the vertices of the hyperrectangle $[\mathbf{0}, \mathbf{x}]$, all of which have fewer nonzero components than $\mathbf{x}$. By strong induction, the claim (4.22) can be applied to each of these vertices of the hyperrectangle (excluding $\mathbf{x}$). After cancellations, we have

$$0 = \Delta(f; [\mathbf{0}, \mathbf{x}]) = f(\mathbf{x}) - \left(f(\mathbf{0}) + \sum_{m=1}^{M} f_m(\mathbf{x}_m) \right)$$

as desired.

4.6.5 Proof of Proposition 4.3.2

This proof builds upon the proof of Proposition 3.3.2 (see Section A.4.5). We may without loss of generality assume for the rest of the proof that the design $\mathbf{x}_1, \dots, \mathbf{x}_n$ forms a lattice including $\mathbf{0}$ (i.e. it is an enumeration of a set that is a Cartesian product $\prod_{j=1}^{d} \mathcal{U}_j$ of finite sets $\mathcal{U}_j = \{0, u_{j,1}, \dots, u_{j,m_j}\} \subseteq [0,1]$). (To handle the case of general design, one can use an argument similar to the proof in Section A.4.5, where one extends the design to a lattice.)

Suppose $\beta \in \mathbb{R}^{p'}$ satisfies $\beta_j \geq 0$ for $j \geq 2$. Let $\mathbf{z}_1, \dots, \mathbf{z}_{p'}$ be the vectors with at most r nonzero entries such that $\mathbf{v}(\mathbf{z}_1), \dots, \mathbf{v}(\mathbf{z}_{p'})$ are the columns of $\mathbf{A}_{\leq r}$. Let $f \coloneqq \sum_{k=1}^{p'} \beta_j \mathbb{I}_{[\mathbf{z}_k, \mathbf{1}]}$ and note that $(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)) = \mathbf{A}_{\leq r}\boldsymbol{\beta}$. We claim that $f \in \mathcal{F}_{\mathrm{EM}}^{d,\leq r}$. By an inclusion-exclusion argument with the definition of quasi-volume, one can check that

$$\Delta(f; [\mathbf{a}, \mathbf{b}]) = \sum_{\substack{k: \mathbf{z}_k \preceq \mathbf{b}, \\ (\mathbf{z}_k)_j > a_j \text{ if } a_j < b_j}} \beta_k, \tag{4.23}$$

for any distinct $\mathbf{a}, \mathbf{b} \in [0,1]^d$ satisfying $\mathbf{a} \preceq \mathbf{b}$. When $\mathbf{a}$ and $\mathbf{b}$ differ by more than r components, the condition that $(\mathbf{z}_j) > a_j \geq 0$ is enforced for more than r components, which is impossible due to the $\mathbf{z}_k$ having at most r nonzero entries; thus in this case, the above sum is empty and the quasi-volume is zero. Otherwise, the sum is a sum of β_k which are all nonnegative (except β_1, which cannot appear in the sum anyway because $\mathbf{z}_1 = \mathbf{0}$ does not satisfy the condition $0 > a_j$ for the component j where $a_j < b_j$), so the quasi-volume is nonnegative. This proves the $\supseteq$ inclusion for the claim (4.12a).

For the reverse inclusion, suppose $f \in \mathcal{F}_{\mathrm{EM}}^{d,\leq r}$. Because our design is a lattice, the full EM matrix $\mathbf{A}$ can be taken to have columns $(\mathbb{I}_{[\mathbf{x},\mathbf{1}]}(\mathbf{x}_1), \dots, \mathbb{I}_{[\mathbf{x},\mathbf{1}]}(\mathbf{x}_n)$ for $\mathbf{x} \in \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. It is square and invertible (see Proposition 3.3.2) so there exists $\boldsymbol{\beta} \in \mathbb{R}^n$ such that $\mathbf{A}\boldsymbol{\beta} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$. One can check that the coefficient β_k corresponding to the indicator function $\mathbb{I}_{[\mathbf{x}_k,\mathbf{1}]}$ is equal to $\Delta(f; [\mathbf{x}'_k, \mathbf{x}_k])$ where $(\mathbf{x}'_k)_j$ is zero if $(\mathbf{x}_k)_j = 0$, and otherwise equals the next smallest value that appears in the jth component of the lattice. (This is essentially a special case of the above formula (4.23) where the right-hand side has exactly one term.) In light of this relationship between $\boldsymbol{\beta}$ and the quasi-volumes of f, we see that if $\mathbf{x}_k$ has more than r nonzero entries, then $\mathbf{x}_k$ and $\mathbf{x}'_k$ differ in more than r entries and the definition of $\mathcal{F}_{\mathrm{EM}}^{d,\leq r}$ forces the quasi-volume $\Delta(f; [\mathbf{x}'_k, \mathbf{x}_k])$ (which equals β_k) to be zero. Otherwise if $\mathbf{x}_k \neq \mathbf{0}$ has at most r nonzero entries, then the fact that $f \in \mathcal{F}_{\mathrm{EM}}^{d}$ leads us to $\beta_k \geq 0$. We see that in the representation $\mathbf{A}\boldsymbol{\beta} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$, we have nonnegativity constraints on all β_k where $k \geq 2$, and moreover coefficients corresponding to columns of $\mathbf{A}$ that do not appear in $\mathbf{A}_{\leq r}$ are forced to be zero. Thus it can be expressed as $\mathbf{A}_{\leq r}\boldsymbol{\beta}'$ for some $\boldsymbol{\beta}' \in \mathbb{R}^{p'}$ with $\beta'_k \geq 0$ for $k \geq 2$.

The proof of the other claim (4.12b) is very similar. Suppose $\beta \in \mathbb{R}^{p''}$ satisfies $\beta_j \geq 0$ for $j \geq 2$. Let $\mathbf{z}_1, \dots, \mathbf{z}_{p''}$ be the vectors with at most r nonzero entries such

that $\mathbf{v}(\mathbf{z}_1), \ldots, \mathbf{v}(\mathbf{z}_{p''})$ are the columns of $\mathbf{A}_{\mathcal{S}}$. Let $f \coloneqq \sum_{k=1}^{p''} \beta_j \mathbb{I}_{[\mathbf{z}_k, \mathbf{1}]}$ and note that $(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) = \mathbf{A}_{\leq r}\boldsymbol{\beta}$. We claim that $f \in \mathcal{F}_{\text{EM}}^{\mathcal{S}}$. Using the expression for the quasi-volume over $[\mathbf{a}, \mathbf{b}]$ above (4.23) , we see that when $a_j < b_j$, the vector $\mathbf{z}_k$ must satisfy $(\mathbf{z}_k)_j > a_j \geq 0$ to be included in the sum. Thus if $\{j : a_j < b_j\}$ intersects with more than one block S_m (see (4.5)), it is impossible for $\mathbf{z}_k$ to satisfy the condition to be included in the sum, since the definition of $\mathbf{A}_{\mathcal{S}}$ forces the nonzero elements of $\mathbf{z}_k$ to lie within a single block. Thus in this case the sum is empty, so the quasi-volume $\Delta(f; [\mathbf{a}, \mathbf{b}])$ is zero for such $\mathbf{a}$ and $\mathbf{b}$.

When $\mathbf{a}$ and $\mathbf{b}$ differ by more than r components, the condition that $(\mathbf{z}_j) > a_j \geq 0$ is enforced for more than r components, which is impossible due to the $\mathbf{z}_k$ having at most r nonzero entries; thus in this case, the above sum is empty and the quasi-volume is zero. Otherwise, the sum is a sum of β_k which are all nonnegative (except β_1, which cannot appear in the sum anyway because $\mathbf{z}_1 = \mathbf{0}$ does not satisfy the condition $0 > a_j$ for the component j where $a_j < b_j$), so the quasi-volume is nonnegative. In light of the characterization given by Lemma 4.2.2, this proves the $\supseteq$ inclusion for the claim (4.12b).

For the reverse inclusion, suppose $f \in \mathcal{F}_{\text{EM}}^{\mathcal{S}}$. Because our design is a lattice, the full EM matrix $\mathbf{A}$ can be taken to have columns $(\mathbb{I}_{[\mathbf{x},\mathbf{1}]}(\mathbf{x}_1), \ldots, \mathbb{I}_{[\mathbf{x},\mathbf{1}]}(\mathbf{x}_n)$ for $\mathbf{x} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. It is square and invertible (see Proposition 3.3.2) so there exists $\boldsymbol{\beta} \in \mathbb{R}^n$ such that $\mathbf{A}\boldsymbol{\beta} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$. One can check that the coefficient β_k corresponding to the indicator function $\mathbb{I}_{[\mathbf{x}_k,\mathbf{1}]}$ is equal to $\Delta(f; [\mathbf{x}'_k, \mathbf{x}_k])$ where $(\mathbf{x}'_k)_j$ is zero if $(\mathbf{x}_k)_j = 0$, and otherwise equals the next smallest value that appears in the jth component of the lattice. (This is essentially a special case of the above formula (4.23) where the right-hand side has exactly one term.) In light of this relationship between $\boldsymbol{\beta}$ and the quasi-volumes of f, we see that if $\mathbf{x}_k$ nonzero entries in more than one block, then $\{j : (\mathbf{x}_k)_j \neq (\mathbf{x}'_k)_j\}$ intersects more than one block, and the definition of $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$ forces the quasi-volume $\Delta(f; [\mathbf{x}'_k, \mathbf{x}_k])$ (which equals β_k) to be zero. Otherwise if the nonzero components of $\mathbf{x}_k \neq \mathbf{0}$ lie within a block, then the fact that $f \in \mathcal{F}_{\text{EM}}^{d}$ leads us to $\beta_k \geq 0$. We see that in the representation $\mathbf{A}\boldsymbol{\beta} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$, we have nonnegativity constraints on all β_k where $k \geq 2$, and moreover coefficients corresponding to columns of $\mathbf{A}$ that do not appear in $\mathbf{A}_{\mathcal{S}}$ are forced to be zero. Thus it can be expressed as $\mathbf{A}_{\mathcal{S}}\boldsymbol{\beta}''$ for some $\boldsymbol{\beta}' \in \mathbb{R}^{p'}$ with $\beta''_k \geq 0$ for $k \geq 2$.

4.6.6 Proof of Proposition 4.3.1

This proof is essentially the same as the proof of Proposition 3.3.1 (see Section A.4.6). The optimization problems (4.9a) and (4.9b) only involve the values of the function at the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$. By Proposition 4.3.2, we must have $(\widehat{f}_{\text{EM},\leq r}(\mathbf{x}_1), \ldots, \widehat{f}_{\text{EM},\leq r}(\mathbf{x}_n)) = \mathbf{A}_{\leq r}\widehat{\boldsymbol{\beta}}_{\text{EM},\leq r}$ and $(\widehat{f}_{\text{EM},\mathcal{S}}(\mathbf{x}_1), \ldots, \widehat{f}_{\text{EM},\mathcal{S}}(\mathbf{x}_n)) = \mathbf{A}_{\mathcal{S}}\widehat{\boldsymbol{\beta}}_{\text{EM},\mathcal{S}}$, so it remains to check that the functions (4.11) defined in the proposition satisfy this equality and belong to the respective classes $\mathcal{F}_{\text{EM}}^{d,\leq r}$ and $\mathcal{F}_{\text{EM}}^{\mathcal{S}}$. The equalities hold simply because the columns of the matrices $\mathbf{A}_{\leq r}$ and $\mathbf{A}_{\mathcal{S}}$ are defined using precisely the indicators that appear in the sums defining $\widehat{f}_{\text{EM},\leq r}$ and $\widehat{f}_{\text{EM},\mathcal{S}}$.

To show membership in the classes, we may assume without loss of generality that the design $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a lattice containing $\mathbf{0}$. (The general case can be handled as in Section A.4.6 by extending the design to such a lattice.) Let $\mathbf{z}_1, \ldots, \mathbf{z}_{p'}$ be such that the columns of $\mathbf{A}_{\leq r}$ are $\mathbf{v}(\mathbf{z}_k)$. The relationship (4.23) between quasi-volumes and the NNLS coefficients holds here as well.

$$\Delta(\widehat{f}_{\mathrm{EM},\leq r}; [\mathbf{a}, \mathbf{b}]) = \sum_{\substack{k:\mathbf{z}_k \preceq \mathbf{b}, \\ (\mathbf{z}_k)_j > a_j \text{ if } a_j < b_j}} (\beta_{\mathrm{EM},\leq r})_k,$$

and a by the same argument used in Section 4.6.5, this quantity is zero if $\mathbf{a}$ and $\mathbf{b}$ differ in more than r components, and nonnegative otherwise, which verifies $\widehat{f}_{\mathrm{EM},\leq r} \in \mathcal{F}_{\mathrm{EM}}^{d,\leq r}$. The proof of $\widehat{f}_{\mathrm{EM},\mathcal{S}} \in \mathcal{F}_{\mathrm{EM}}^{\mathcal{S}}$ is completely analogous.

Bibliography

[1] Aistleitner, C. and J. Dick (2015). Functions of bounded variation, signed measures, and a general Koksma-Hlawka inequality. *Acta Arith. 167*(2), 143–171.

[2] Amelunxen, D. and M. Lotz (2015). Intrinsic volumes of polyhedral cones: a combinatorial perspective. *arXiv preprint arXiv:1512.06033*.

[3] Amelunxen, D., M. Lotz, M. B. McCoy, and J. A. Tropp (2014). Living on the edge: phase transitions in convex programs with random data. *Inf. Inference 3*(3), 224–294.

[4] Assouad, P. (1983). Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique 296*, 1021–1024.

[5] Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist. 26*, 641–647.

[6] Bacchetti, P. (1989). Additive isotonic models. *J. Amer. Statist. Assoc. 84*(405), 289–294.

[7] Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression.* John Wiley & Sons, London-New York-Sydney. Wiley Series in Probability and Mathematical Statistics.

[8] Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory 39*(3), 930–945.

[9] Bellec, P. C. (2017). Optimistic lower bounds for convex regularized least-squares. *arXiv preprint arXiv:1703.01332*.

[10] Bellec, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist. 46*(2), 745–780.

[11] Benkeser, D. and M. Van Der Laan (2016). The highly adaptive lasso estimator. In *2016 IEEE international conference on data science and advanced analytics (DSAA)*, pp. 689–696. IEEE.

[12] Blei, R., F. Gao, and W. V. Li (2007). Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc. 135*(12), 4009–4018.

[13] Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist. 26*, 607–616.

[14] Brunk, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, pp. 177–197. London: Cambridge Univ. Press.

[15] Bungartz, H.-J. and M. Griebel (2004). Sparse grids. *Acta numerica 13*, 147–269.

[16] Chambolle, A., V. Caselles, D. Cremers, M. Novaga, and T. Pock (2010). An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery 9*(263-340), 227.

[17] Chatterjee, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist. 42*(6), 2340–2381.

[18] Chatterjee, S. and S. Goswami (2019). New risk bounds for 2d total variation denoising. *arXiv preprint arXiv:1902.01215*.

[19] Chatterjee, S., A. Guntuboyina, and B. Sen (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist. 43*(4), 1774–1800.

[20] Chatterjee, S., A. Guntuboyina, and B. Sen (2018). On matrix estimation under monotonicity constraints. *Bernoulli 24*(2), 1072–1100.

[21] Chen, X., A. Guntuboyina, and Y. Zhang (2017). A note on the approximate admissibility of regularized estimators in the gaussian sequence model. *arXiv preprint arXiv:1703.00542*.

[22] Chkifa, A., N. Dexter, H. Tran, and C. Webster (2018). Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. *Mathematics of Computation 87*(311), 1415–1450.

[23] Christensen, R. (2011). *Plane answers to complex questions: the theory of linear models*. Springer Science & Business Media.

[24] Condat, L. (2013). A direct algorithm for 1-d total variation denoising. *IEEE Signal Process. Lett. 20*(11), 1054–1057.

[25] Condat, L. (2017). Discrete total variation: New definition and minimization. *SIAM Journal on Imaging Sciences 10*(3), 1258–1290.

[26] Dalalyan, A., M. Hebiri, and J. Lederer (2017). On the prediction performance of the lasso. *Bernoulli 23*(1), 552–581.

[27] Demetriou, I. and P. Tzitziris (2017). Infant mortality and economic growth: modeling by increasing returns and least squares. In *Proceedings of the World Congress on Engineering*, Volume 2.

[28] Deng, H. and C.-H. Zhang (2018). Isotonic regression in multi-dimensional spaces and graphs. *arXiv preprint arXiv:1812.08944*.

[29] Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture 1*(32), 375.

[30] Donoho, D. L. and I. M. Johnstone (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist. 26*(3), 879–921.

[31] Dudley, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis 1*, 290–330.

[32] Fan, Z. and L. Guan (2018). Approximate ℓ_0-penalized estimation of piecewise-constant signals on graphs. *Ann. Statist. 46*(6B), 3217–3245.

[33] Fang, B. and A. Guntuboyina (2019). On the risk of convex-constrained least squares estimators under misspecification. *Bernoulli 25*(3), 2206–2244.

[34] Fang, B., A. Guntuboyina, and B. Sen (2019). Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy-Krause variation. *arXiv preprint arXiv:1903.01395*.

[35] Feller, W. (2015). Completely monotone functions and sequences. In *Selected Papers I*, pp. 497–510. Springer.

[36] Gao, F. (2013). Bracketing entropy of high dimensional distributions. In *High dimensional probability VI*, Volume 66 of *Progr. Probab.*, pp. 3–17. Birkhäuser/Springer, Basel.

[37] Gao, F., W. V. Li, and J. A. Wellner (2010). How many Laplace transforms of probability measures are there? *Proc. Amer. Math. Soc. 138*(12), 4331–4344.

[38] Gill, R. D., M. J. Laan, and J. A. Wellner (1995). Inefficient estimators of the bivariate survival function for three models. In *Annales de l'IHP Probabilités et statistiques*, Volume 31, pp. 545–597.

[39] Goddard, L. S. (1945). The accumulation of chance effects and the Gaussian frequency distribution. *Philos. Mag. (7) 36*, 428–433.

[40] Groeneboom, P. (2013). The bivariate current status model. *Electronic Journal of Statistics 7*, 1783–1805.

[41] Groeneboom, P. and G. Jongbloed (2014). *Nonparametric estimation under shape constraints*, Volume 38 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York. Estimators, algorithms and asymptotics.

[42] Groeneboom, P., T. Ketelaars, et al. (2011). Estimators for the interval censoring problem. *Electronic Journal of Statistics 5*, 1797–1845.

[43] Guntuboyina, A. (2011). Lower bounds for the minimax risk using f divergences, and applications. *IEEE Transactions on Information Theory 57*, 2386–2399.

[44] Guntuboyina, A., D. Lieu, S. Chatterjee, and B. Sen (2017). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist. (to appear)*; available at `https://arxiv.org/abs/1702.05113`.

[45] Guntuboyina, A. and B. Sen (2018). Nonparametric Shape-Restricted Regression. *Statist. Sci. 33*(4), 568–594.

[46] Guo, D. and X. Wang (2006). Quasi-monte carlo filtering in nonlinear dynamic systems. *IEEE transactions on signal processing 54*(6), 2087–2098.

[47] Han, Q., T. Wang, S. Chatterjee, and R. J. Samworth (2019). Isotonic regression in general dimensions. *Ann. Statist. 47*(5), 2440–2471.

[48] Hildreth, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc. 49*, 598–619.

[49] Hiriart-Urruty, J.-B. and C. Lemaréchal (2012). *Fundamentals of convex analysis*. Springer Science & Business Media.

[50] Hjort, N. L. and D. Pollard (1993). Asymptotics for minimisers of convex processes. Technical report. available at arXiv preprint arXiv:1107.3806.

[51] Hobson, E. W. (1950). *The theory of functions of a real variable and the theory of Fourier's series*, Volume 1. CUP Archive.

[52] Hütter, J.-C. and P. Rigollet (2016). Optimal rates for total variation denoising. In *Conference on Learning Theory*, pp. 1115–1146.

[53] Kim, A. K. and R. J. Samworth (2014). Global rates of convergence in log-concave density estimation. *arXiv preprint arXiv:1404.2298*.

[54] Kim, S.-J., K. Koh, S. Boyd, and D. Gorinevsky (2009). ℓ_1 trend filtering. *SIAM review 51*(2), 339–360.

[55] Klivans, C. J. and E. Swartz (2011). Projection volumes of hyperplane arrangements. *Discrete & Computational Geometry 46*(3), 417.

[56] Leonov, A. S. (1996). On the total variation for functions of several variables and a multidimensional analog of Helly's selection principle. *Mathematical Notes 63*(1), 61–71.

[57] Lin, K., J. L. Sharpnack, A. Rinaldo, and R. J. Tibshirani (2017). A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In *Advances in Neural Information Processing Systems*, pp. 6884–6893.

[58] Lin, Y. (2000). Tensor product space ANOVA models. *Ann. Statist. 28*(3), 734–755.

[59] Maathuis, M. H. (2005). Reduction algorithm for the npmle for the distribution function of bivariate interval-censored data. *Journal of Computational and Graphical Statistics 14*(2), 352–362.

[60] Mammen, E. and S. van de Geer (1997). Locally adaptive regression splines. *Ann. Statist. 25*(1), 387–413.

[61] Massart, P. (2007). *Concentration inequalities and model selection. Lecture notes in Mathematics*, Volume 1896. Berlin: Springer.

[62] Menéndez, J. A., C. Rueda, and B. Salvador (1992). Dominance of likelihood ratio tests under cone constraints. *Ann. Statist. 20*(4), 2087–2099.

[63] Meyer, M. and M. Woodroofe (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist. 28*(4), 1083–1104.

[64] Nemirovski, A. (2000). Topics in non-parametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, Volume 1738 of *Lecture Notes in Math.*, pp. 85–277. Springer, Berlin.

[65] Niyogi, P. and F. Girosi (1999). Generalization bounds for function approximation from scattered noisy data. *Advances in Computational Mathematics 10*(1), 51–80.

[66] Ortelli, F. and S. van de Geer (2018). On the total variation regularized estimator over the branched path graph. *arXiv preprint arXiv:1806.01009*.

[67] Ortelli, F. and S. van de Geer (2019a). Oracle inequalities for image denoising with total variation regularization. *arXiv preprint arXiv:1911.07231*.

[68] Ortelli, F. and S. van de Geer (2019b). Synthesis and analysis in total variation regularization. *arXiv preprint arXiv:1901.06418*.

[69] Owen, A. B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary multivariate analysis and design of experiments*, Volume 2 of *Ser. Biostat.*, pp. 49–74. World Sci. Publ., Hackensack, NJ.

[70] Oymak, S. and B. Hassibi (2013). Sharp mse bounds for proximal denoising. *Foundations of Computational Mathematics*, 1–65.

[71] Pal, J. K. (2008). Spiking problem in monotone regression: Penalized residual sum of squares. *Statistics & Probability Letters 78*(12), 1548–1556.

[72] Prause, A. and A. Steland (2017). Sequential detection of three-dimensional signals under dependent noise. *Sequential Analysis 36*(2), 151–178.

[73] Robertson, T., F. T. Wright, and R. L. Dykstra (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester.

[74] Rudin, L. I., S. Osher, and E. Fatemi (1992). Nonlinear total variation based noise removal algorithms. *Phys. D 60*(1-4), 259–268. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).

[75] Ruiz, M. d. Á., H. Li, and A. Munk (2018). Frame-constrained total variation regularization for white noise regression. *arXiv preprint arXiv:1807.02038*.

[76] Sadhanala, V., Y.-X. Wang, J. L. Sharpnack, and R. J. Tibshirani (2017). Higher-order total variation classes on grids: Minimax theory and trend filtering methods. In *Advances in Neural Information Processing Systems*, pp. 5800–5810.

[77] Sadhanala, V., Y.-X. Wang, and R. J. Tibshirani (2016). Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In *Advances in Neural Information Processing Systems*, pp. 3513–3521.

[78] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist. 10*(4), 1040–1053.

[79] Temlyakov, V. (2018). *Multivariate approximation*, Volume 32. Cambridge University Press.

[80] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

[81] Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist. 42*(1), 285–323.

[82] van de Geer, S. A. (2000). *Applications of empirical process theory*, Volume 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

[83] van der Laan, M. (2017a). Finite sample inference for targeted learning. *arXiv preprint arXiv:1708.09502*.

[84] van der Laan, M. (2017b). A generally efficient targeted minimum loss based estimator based on the highly adaptive Lasso. *Int. J. Biostat. 13*(2), 20150097, 35.

[85] van der Laan, M. J., D. Benkeser, and W. Cai (2019). Efficient estimation of pathwise differentiable target parameters with the undersmoothed highly adaptive lasso. *arXiv preprint arXiv:1908.05607*.

[86] van der Laan, M. J. and A. F. Bibaut (2017). Uniform consistency of the highly adaptive lasso estimator of infinite dimensional parameters. *arXiv preprint arXiv:1709.06256*.

[87] van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

[88] Vapnik, V. N. and A. Y. Chervonenkis (2015). On the uniform convergence of relative frequencies of events to their probabilities. pp. 11–30. Reprint of Theor. Probability Appl. 16 (1971), 264–280.

[89] Wahba, G., Y. Wang, C. Gu, R. Klein, and B. Klein (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. *Ann. Statist. 23*(6), 1865–1895.

[90] Warrack, G. and T. Robertson (1984). A likelihood ratio test regarding two nested but oblique order-restricted hypotheses. *J. Amer. Statist. Assoc. 79*(388), 881–886.

[91] Wei, Y., M. J. Wainwright, and A. Guntuboyina (2019). The geometry of hypothesis testing over convex cones: generalized likelihood ratio tests and minimax radii. *Ann. Statist. 47*(2), 994–1024.

[92] Widder, D. V. (1941). *The Laplace Transform*. Princeton Mathematical Series, v. 6. Princeton University Press, Princeton, N. J.

[93] Wu, J., M. C. Meyer, and J. D. Opsomer (2015). Penalized isotonic regression. *Journal of Statistical Planning and Inference 161*, 12–24.

[94] Yang, Y. and A. Barron (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist. 27*(5), 1564–1599.

[95] Young, W. and G. C. Young (1924). On the discontinuities of monotone functions of several variables. *Proceedings of the London Mathematical Society 2*(1), 124–142.

[96] Yu, B. (1997). Assouad, Fano, and Le Cam. In D. Pollard, E. Torgersen, and G. L. Yang (Eds.), *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pp. 423–435. New York: Springer-Verlag.

[97] Zhang, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist. 30*(2), 528–555.

[98] Zhang, T. (2019). Element-wise estimation error of a total variation regularized estimator for change point detection. *arXiv preprint arXiv:1901.00914*.

[99] Ziemer, W. P. (2012). *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*, Volume 120. Springer Science & Business Media.

Appendix A

Appendix for Chapter 3

The first section of this appendix contains a statement of another adaptation result for the HK Variation denoising estimator similar to Theorem 3.4.10. Section A.2 contains simulations containing examples and depictions of our EM and HK variation estimators, including an application to estimation in the bivariate current status model. The rest of the supplement contains proofs of all results from the main paper. The proofs for the risk results are given in Section A.3 while the proofs of the results in Section 3.2 and Section 3.3 are given in Section A.4. Additional technical results used in the proofs of Section A.3 are proved in Section A.5.

A.1 Another adaptation result for the Hardy-Krause variation denoising estimator

The goal of this section is to prove a result that is similar to but stronger than Theorem 3.4.10 for $d = 2$. Specifically, the minimum length condition appearing in (3.50) is relaxed for the next result. We take $d = 2$ in this section. For a given constant $0 < c \leq 1/2$, let $\widetilde{\mathfrak{R}}_1^2(c)$ denote the collection of functions $f : [0,1]^2 \to \mathbb{R}$ of the form (3.49) for some $a_1, a_0 \in \mathbb{R}$ and $\mathbf{x}^* = (x_1^*, x_2^*) \in [0,1]^2$ satisfying

$$\min\{|\mathbb{L}_{n_1,\ldots,n_d} \cap [\mathbf{x}^*, \mathbf{1}]|, |\mathbb{L}_{n_1,\ldots,n_d} \setminus [\mathbf{x}^*, \mathbf{1}]|\} \geq cn. \tag{A.1}$$

Note that the above condition is implied by the earlier minimum size condition (3.50) because $[\mathbf{0}, \mathbf{x}^*) \subseteq [\mathbf{x}^*, \mathbf{1}]^c$. Therefore we have $\mathfrak{R}_1^2(c) \subseteq \widetilde{\mathfrak{R}}_1^2(c)$. Note also that $\mathbf{x}^* := (0.5, 0, \ldots, 0)$ satisfies (A.1). The next result (proved in Section A.3.8) is the analogue of Theorem 3.4.10 for $d = 2$ which works under the weaker minimum size condition (A.1).

Theorem A.1.1. *Consider the lattice design* (3.34). *Fix* $f^* : [0,1]^2 \to \mathbb{R}$ *and consider the*

estimator $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ *with a tuning parameter* V*. Then for every* $0 < c \leq 1/2$*, we have*

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) \leq \inf_{\substack{f \in \widetilde{\mathfrak{R}}_1^2(c): \\ V_{\mathrm{HK}\mathbf{0}}(f) = V}} \left\{ \mathcal{L}(f, f^*) + C(c) \frac{\sigma^2}{n} (\log(en))^3 (\log(e \log(en)))^{\frac{3}{2}} \right\} \tag{A.2}$$

for a constant $C(c)$ *that depends only on* c*.*

When $f^* \in \widetilde{\mathfrak{R}}_1^2(c)$ and $V = V_{\mathrm{HK}\mathbf{0}}(f^*)$, inequality (A.2) readily implies

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) \leq C \frac{\sigma^2}{n} (\log(en))^3 (\log(e \log(en)))^{\frac{3}{2}}.$$

Note that previously we were only able to claim this result for functions f^* in the smaller class $\mathfrak{R}_1^2(c)$.

A.2 Simulation studies

Here we discuss some simulations we performed with the two estimators $\widehat{f}_{\mathrm{EM}}$ (3.2) and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ (3.6) for $d = 2$.

A.2.1 Examples of the estimators

We start by visual illustrations of our estimators for specific values of f^*. In Figure A.1 we depict an example of $\widehat{f}_{\mathrm{EM}}$ when fit on a 10×10 grid of observations (i.e., $n_1 = n_2 = 10$ and $n = 100$) from an EM function f^*. In Figure A.2, we consider a different example where f^* has $k(f^*) = 4$ and depict the estimate $\widehat{f}_{\mathrm{EM}}$ computed on a 10×10 grid of observations.

In Figure A.3 we consider a function $f^* \in \mathfrak{R}_1^d(1/4)$ (see equations (3.49) and (3.50)) and depict our estimate $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ computed from a 10×10 grid of observations for various values of the tuning parameter V.

We remark that in these examples, we have chosen the estimator to be rectangular piecewise constant, with values obtained by solving the finite-dimensional NNLS or LASSO problem as discussed in Section 3.3. Additionally, one can observe in Figures A.4 and A.3 that the performance of $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ improves as V approaches the optimal V^*. Note also that in Figures A.2 and A.3 (in the case $V = V^*$) where f^* is rectangular piecewise constant, the estimate is also rectangular piecewise constant with relatively few "jumps."

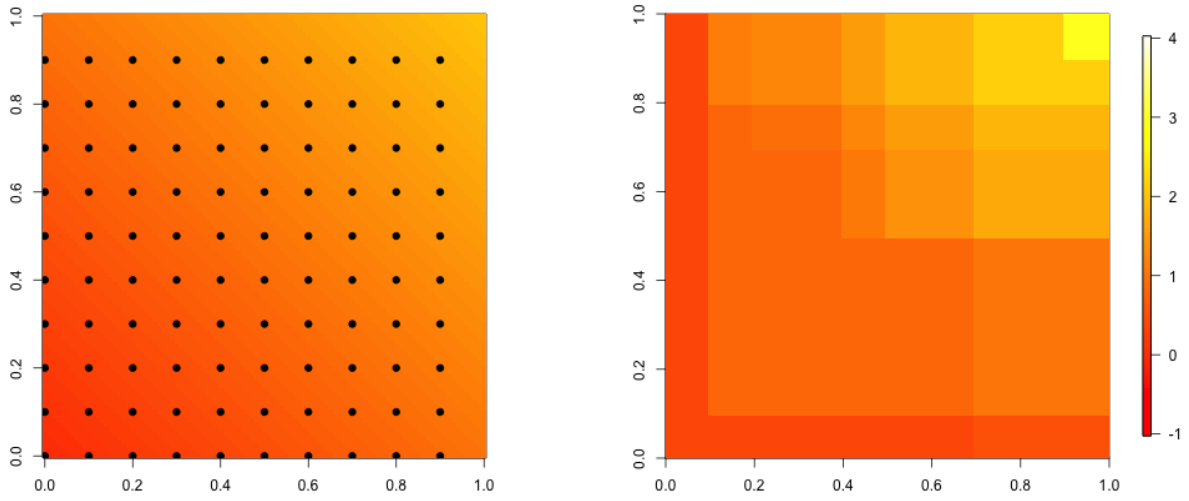


Figure A.1: The function $f^*(x_1, x_2) = x_1 + x_2$ (left), and the estimate $\widehat{f}_{\text{EM}}$ (right) performed on observations from f^* on the grid design ($n_1 = n_2 = 10$) with standard Gaussian noise ($\sigma^2 = 1$).

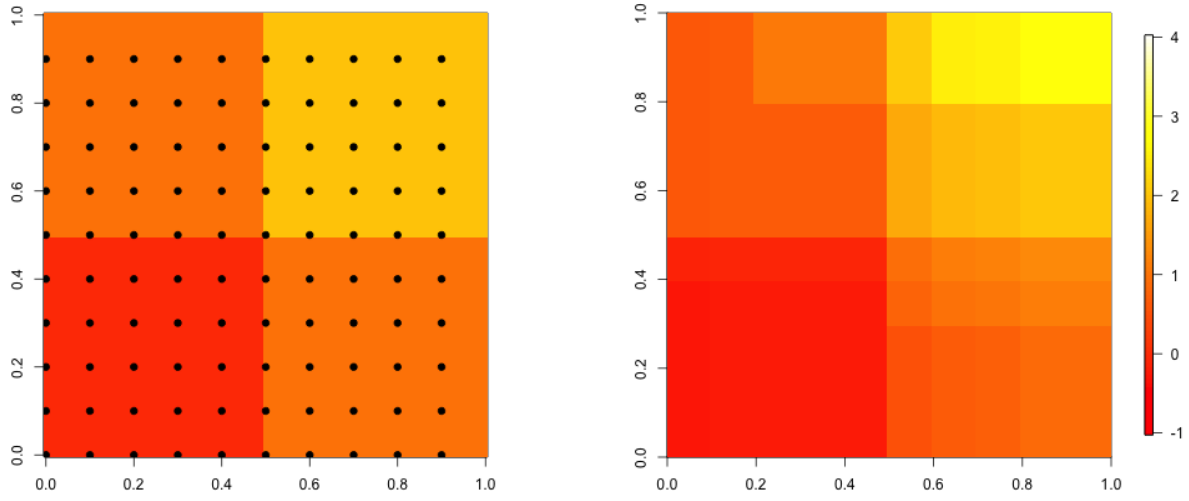


Figure A.2: The function $f^*(x_1, x_2) = \mathbb{I}\{x_1 \geq 0.5\} + \mathbb{I}\{x_2 \geq 0.5\}$ (left), and the estimate $\widehat{f}_{\text{EM}}$ (right) performed on observations from f^* with the grid design ($n_1 = n_2 = 10$) and standard Gaussian noise ($\sigma^2 = 1$).

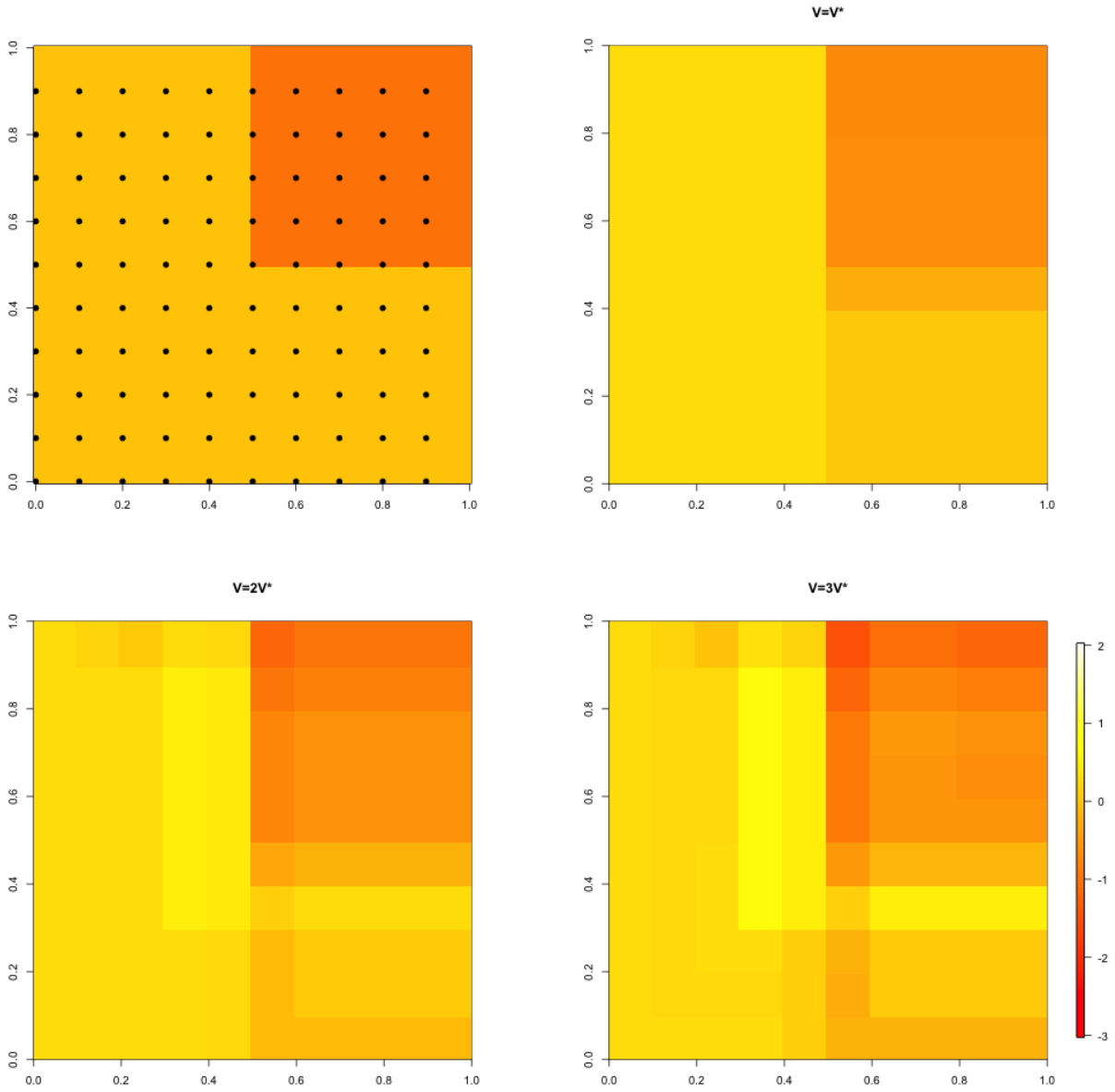


Figure A.3: The function $f^*(x_1, x_2) = -\mathbb{I}\{x_1 \geq 0.5, x_2 \geq 0.5\}$ (upper left), and the estimate $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ for $V = V^*, 2V^*, 3V^*$, performed on observations from f^* on the grid design ($n_1 = n_2 = 10$) with standard Gaussian noise ($\sigma^2 = 1$).

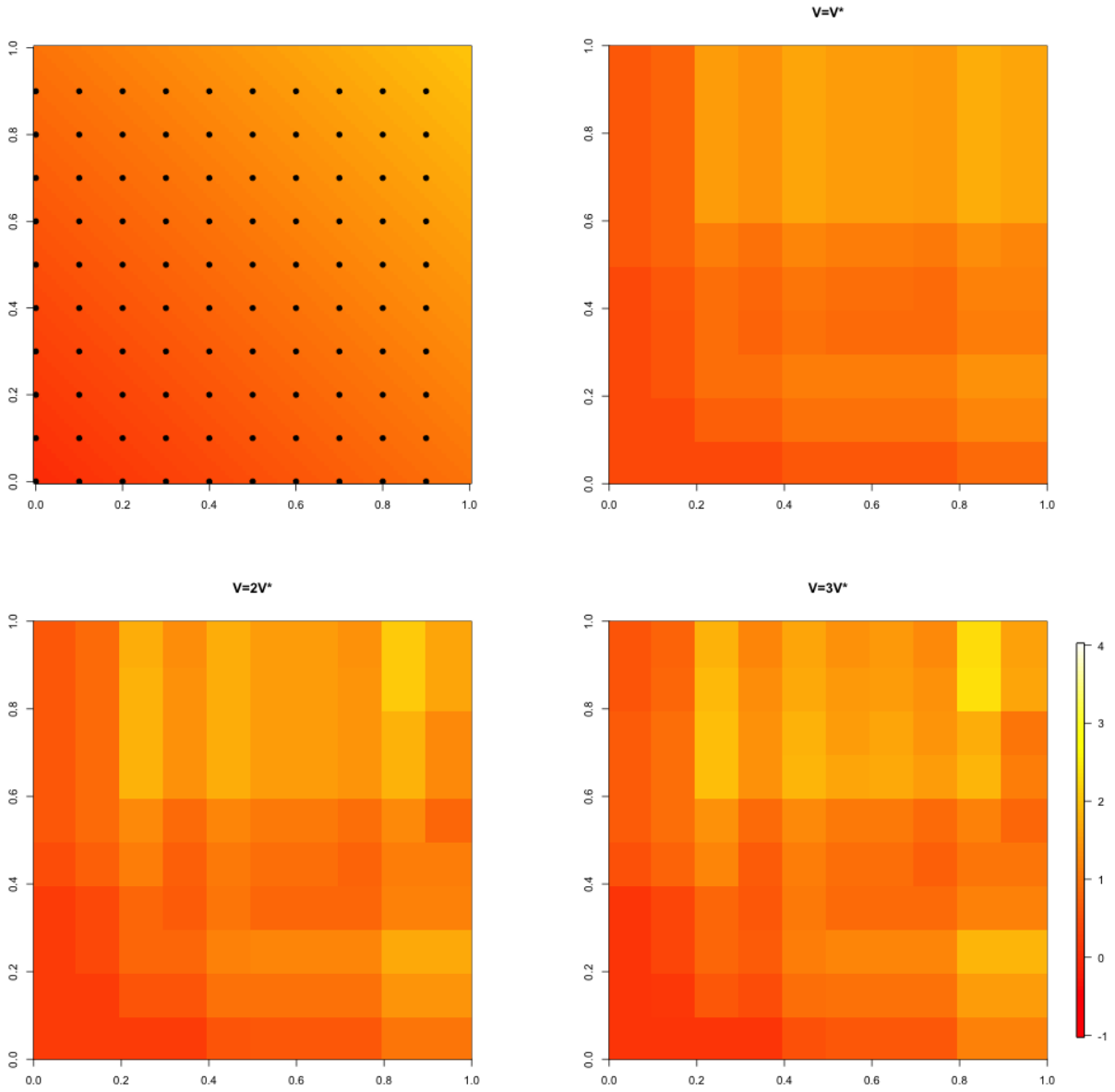


Figure A.4: The function $f^*(x_1, x_2) = x_1 + x_2$ (upper left), and the estimate $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ for $V = V^*, 2V^*, 3V^*$, performed on observations from f^* on the grid design ($n_1 = n_2 = 10$) with standard Gaussian noise ($\sigma^2 = 1$).

Although our theorems in Section 3.4.1 and Section 3.4.2 only apply in case of lattice design, we can still compute the estimator for arbitrary design. In Figure A.5, we used the "naïve gridding" approach described in Section 3.3 to compute the design matrix for the NNLS optimization problem. Note that the "jumps" in our estimate are located at design points.

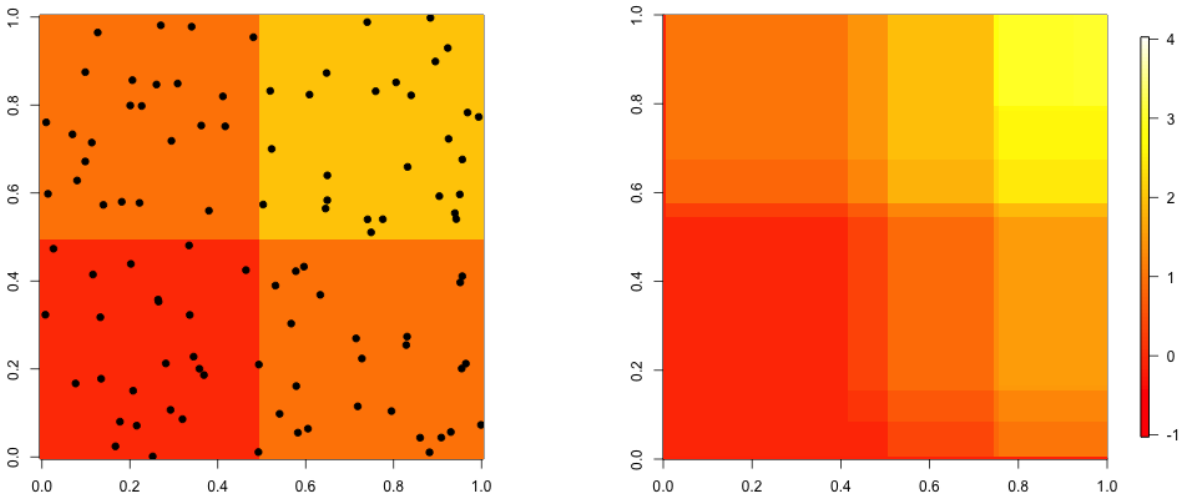


Figure A.5: The function $f^*(x_1, x_2) = \mathbb{I}\{x_1 \geq 0.5\} + \mathbb{I}\{x_2 \geq 0.5\}$ (left), and the estimate $\widehat{f}_{\text{EM}}$ (right) performed on observations from f^* on a uniformly drawn random design ($n = 100$) and standard Gaussian noise ($\sigma^2 = 1$).

A.2.2 Bivariate current status model

One practical setting where our estimator may be useful is in the bivariate current status model, which is a particular variant of the interval censoring problem [40, 42, 59]. In this setting we observe $(\mathbf{x}_i, y_i)$ where the y_i are independent Bernoulli random variables y_i with success parameter $F_0(\mathbf{x}_i)$, for some bivariate CDF F_0. Since F_0 is an entirely monotone function of $\mathbf{x}$, it is plausible to use our EM estimator (3.2) on these observations to estimate F_0. In Figure A.6, we simulated $n = 500$ observations in the case where $F_0(\mathbf{x}) = \frac{1}{2}(x_1^2 x_2 + x_1 x_2^2)$ on $[0,1]^2$ (the CDF of the density $f_0(\mathbf{x}) = x_1 + x_2$), and where $\mathbf{x}_i$ are drawn uniformly from $[0,1]^2$, and where $y_i \mid \mathbf{x}_i \sim \text{Bern}(F_0(\mathbf{x}_i))$. We get a fairly reasonable estimate of the original CDF on the interior of the square $[0,1]^2$. The estimated function is not a proper CDF, as it can take values outside of $[0,1]$, which happens often along the boundaries of the square $[0,1]^2$. One could avoid this by modifying the estimator $\widehat{f}_{\text{EM}}$ by restricting the

least squares optimization to functions in $\widehat{f}_{\mathrm{EM}}$ that take values in $[0,1]$, which would amount to adding two more linear constraints on the corresponding NNLS problem (3.28). This issue of obtaining an estimate that is not a proper CDF also occurs with a plug-in estimator studied by Groeneboom [40], which they address by proposing a truncation procedure on the boundaries of the square.

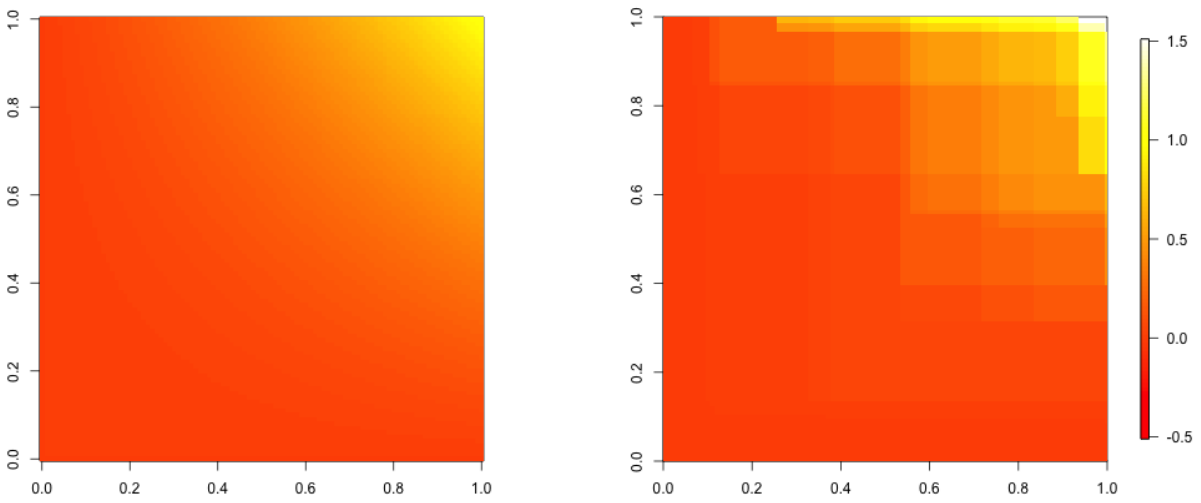


Figure A.6: The CDF $F_0(x_1, x_2) = \frac{1}{2}(x_1 x_2^2 + x_1^2 x_2)$ (left), and the estimate $\widehat{f}_{\mathrm{EM}}$ (right) applied on $n = 500$ observations of the form $(\mathbf{x}_i, y_i)$ where $y_i \sim \mathrm{Bern}(F_0(\mathbf{x}_i))$.

A.2.3 Adaptation to more general rectangular piecewise constant functions

One severe limitation of Theorems 3.4.10 and A.1.1 is that they only consider functions of the form (3.49), which only has one "jump" and two continguous constant pieces.

The following simulation study suggests that the upper bound of $n^{-1}(\log n)^{\gamma}$ that we proved in Theorems 3.4.10 and A.1.1 may also hold for a larger subclass of rectangular piecewise constant functions $\mathfrak{R}^d$.

The function $f^* : [0,1]^2 \mapsto \mathbb{R}$ we consider is

$$f^*(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \left([\frac{1}{3}, \frac{2}{3}) \times ([0, \frac{1}{3}) \cup [\frac{2}{3}, 1)\right) \cup \left(([0, \frac{1}{3}) \cup [\frac{1}{3}, 1]) \times [\frac{1}{3}, \frac{2}{3})\right) \\ 0 & \text{otherwise.} \end{cases}$$

One can check that $V^* = 12$. Visually, it has a checkered pattern (see Figure A.7).

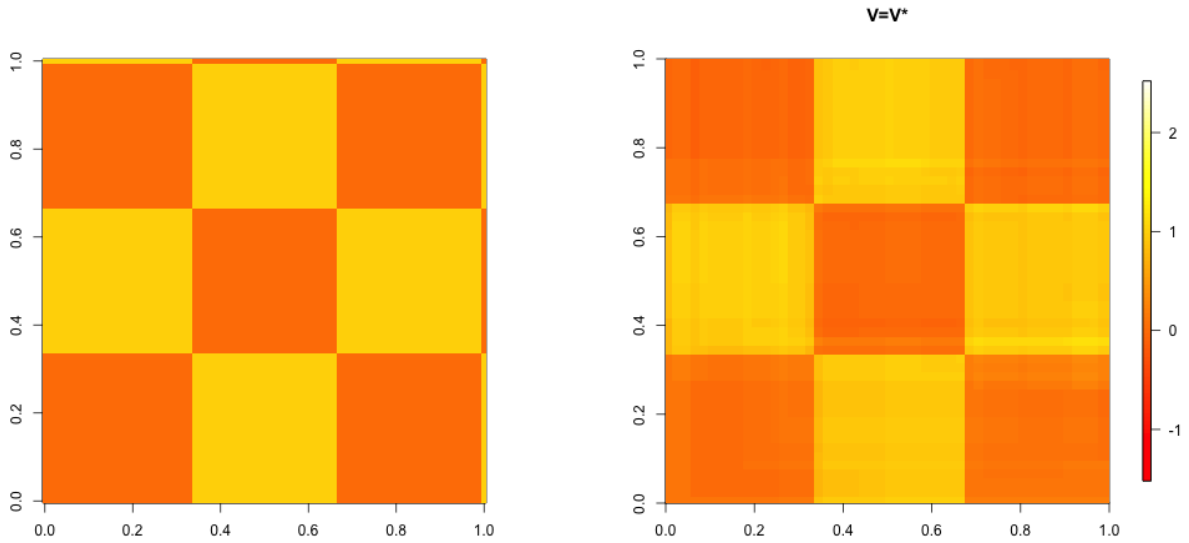


Figure A.7: Depiction of f^* (left), and an example of $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ (right) when given noisy measurements ($\sigma = 0.5$) from f^* on the grid design ($n_1 = n_2 = 50$).

We considered the lattice design $\mathbb{L}_{n_1,\ldots,n_d}$ with $n_1 = n_2 \in \{50, 60, 80, 95, 110\}$ (note that consequently $n = n_1 n_2$ ranges between 2500 and 12100). For each value of n, we performed 20 trials of generating observations $y_1, \ldots, y_n$ with noise $\sigma = 0.5$, computed $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ with $V = V^*$, and computed the error $\frac{1}{n}\sum_{i=1}^{n}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_i) - f^*(\mathbf{x}_i))^2$. Averaging over the 20 trials gives us an estimate r_n of $\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*)$ for that value of n.

As shown in Figure A.8 A linear regression of $\log r_n$ over $\log n$ yielded a slope of -0.85 which indicates that the estimator is performing better than the worst-case rate of $n^{-2/3}$ given in Theorem 3.4.6. A linear regression of $\log r_n$ over $\log \frac{n}{\log n}$ yielded a slope of -0.96, while a regression of $\log r_n$ over $\log \frac{n}{(\log n)^2}$ yielded a slope of -1.11. Thus these simulations suggest that the estimator $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ has risk on the order of $n^{-1}(\log n)^{\gamma}$ (possibly for $\gamma \leq 2$) for rectangular piecewise constant functions beyond the ones considered in Theorems 3.4.10 and A.1.1.

A.3 Proofs of Risk Results

A.3.1 Preliminaries

Note that the risks $\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*)$ and $\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*)$ both only depend on the values of the estimators $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ at the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Also by the results from Sec-

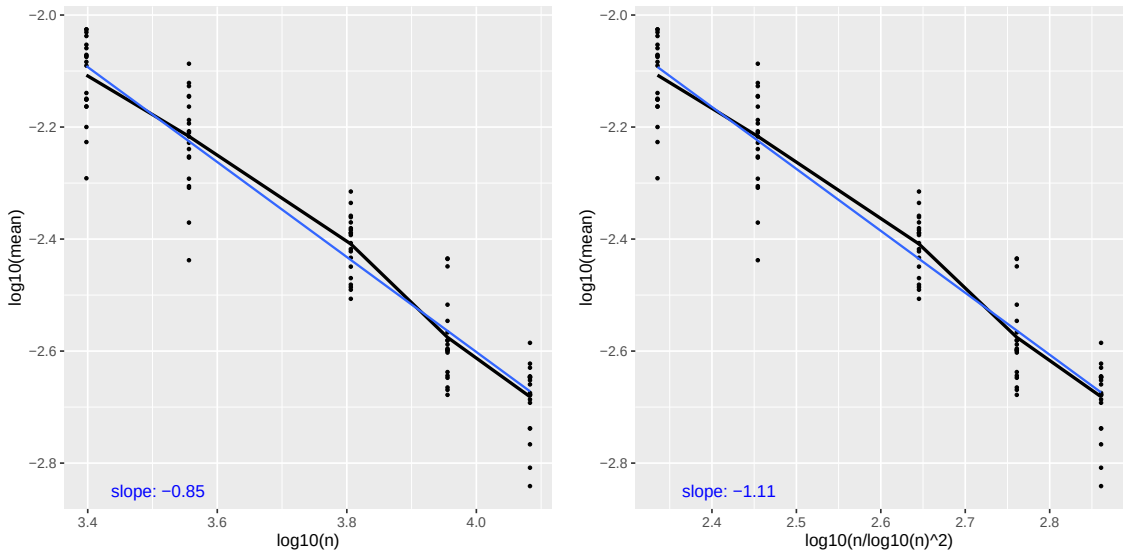


Figure A.8: Plot of estimate of $\log \mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*)$ vs. $\log n$ (left) and vs. $\log \frac{n}{(\log n)^2}$ (right).

tion 3.3, it is clear that the vectors $(\widehat{f}_{\mathrm{EM}}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{EM}}(\mathbf{x}_n))$ and $(\widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_n))$ are Euclidean projections of the data vector $\mathbf{y} = (y_1, \ldots, y_n)$ on the closed convex sets

$$\left\{\mathbf{A}\boldsymbol{\beta} : \min_{j\geq 2} \beta_j \geq 0\right\} \quad \text{and} \quad \left\{\mathbf{A}\boldsymbol{\beta} : \sum_{j\geq 2} |\beta_j| \leq V\right\}$$

respectively. Consequently, we can apply general results from the theory of convex-constrained LSEs to prove the risk results for $\widehat{f}_{\mathrm{EM}}$ and $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$. This theory is, by now, well established (see e.g., van de Geer [82], van der Vaart and Wellner [87], Hjort and Pollard [50], Chatterjee [17]). The following result from Chatterjee [17] provides upper bounds for the risk of general convex-constrained LSEs. This result will be used in the proofs of Theorem 3.4.1 and Theorem 3.4.6.

Theorem A.3.1 (Chatterjee [17])**.** *Let $\mathcal{K}$ be a closed convex set in $\mathbb{R}^n$ and let*

$$\widehat{\boldsymbol{\theta}} := \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathcal{K}} \|\mathbf{y} - \boldsymbol{\theta}\|^2, \tag{A.3}$$

where $\mathbf{y} \sim \mathcal{N}_n(\boldsymbol{\theta}^*, \mathbf{I}_n)$ for some $\boldsymbol{\theta}^* \in \mathbb{R}^n$ (not necessarily in $\mathcal{K}$). Then there exists a universal positive constant C such that

$$\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leq C \max(t_*^2, 1)$$

for every $t_* > 0$ which satisfies

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{K}:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\leq t_*} \langle \boldsymbol{\xi}, \boldsymbol{\theta} - \boldsymbol{\theta}^*\rangle\right] \leq \frac{t_*^2}{2} \qquad \text{where } \boldsymbol{\xi} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n). \tag{A.4}$$

Theorem A.3.1 is sufficient to prove Theorem 3.4.1 and Theorem 3.4.6. However, in order to handle the misspecified setting discussed in Remark 3.4.2 and Remark 3.4.8, one needs the following generalization of Theorem A.3.1. Below,

$$\Pi_{\mathcal{K}}(\mathbf{v}) \coloneqq \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathcal{K}} \|\mathbf{v}-\boldsymbol{\theta}\|^2$$

denotes the projection of $\mathbf{v}$ onto the closed convex set $\mathcal{K}$. The following result generalizes Theorem A.3.1 to the case of model misspecification. It is similar to related generalizations of Theorem A.3.1 from Chen et al. [21] and Bellec [9]. We omit the proof of this result as it can be proved by a straightforward generalization of the proof of the original result, Theorem A.3.1, from Chatterjee [17].

Theorem A.3.2. *Let $\mathcal{K}$ be a closed convex set in $\mathbb{R}^n$, and let $\widehat{\boldsymbol{\theta}} \coloneqq \Pi_{\mathcal{K}}(\mathbf{y})$ be as defined above* (A.3)*, with $\mathbf{y} \sim \mathcal{N}_n(\boldsymbol{\theta}^*, \mathbf{I}_n)$ and $\boldsymbol{\theta}^* \in \mathbb{R}^n$. Then there exists a universal positive constant C such that*

$$\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \Pi_{\mathcal{K}}(\boldsymbol{\theta}^*)\|^2 \le C\max(t_*^2, 1),$$

for every $t_* > 0$ which satisfies

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{K}:\|\boldsymbol{\theta}-\Pi_{\mathcal{K}}(\boldsymbol{\theta}^*)\|\le t_*} \langle \boldsymbol{\xi}, \boldsymbol{\theta} - \Pi_{\mathcal{K}}(\boldsymbol{\theta}^*)\rangle\right] \le \frac{t_*^2}{2} \qquad \text{where } \boldsymbol{\xi} \sim \mathcal{N}_n(0, \mathbf{I}_n). \tag{A.5}$$

Note that in the well-specified setting $\boldsymbol{\theta}^* \in \mathcal{K}$, we have $\Pi_{\mathcal{K}}(\boldsymbol{\theta}^*) = \boldsymbol{\theta}^*$, and thus Theorem A.3.1 and Theorem A.3.2 are identical. On the other hand, in the misspecified setting $\boldsymbol{\theta}^* \notin \mathcal{K}$, the two results differ in the risk quantity they control: $\mathbb{E}\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|^2$ and $\mathbb{E}\|\widehat{\boldsymbol{\theta}}-\Pi_{\mathcal{K}}(\boldsymbol{\theta}^*)\|^2$ respectively and the fact that $\boldsymbol{\theta}^*$ appearing in (A.4) is replaced by $\Pi_{\mathcal{K}}(\boldsymbol{\theta}^*)$ in (A.5).

Remark A.3.3 (Risk bounds under misspecification)**.** *Theorem 3.4.1 and Theorem 3.4.6 are proved via Theorem A.3.1 by establishing* (A.4) *for an appropriate t_*. If we replace $\boldsymbol{\theta}^*$ in these proofs by $\Pi_{\mathcal{K}}(\boldsymbol{\theta}^*)$ and replace the use of Theorem A.3.1 with that of Theorem A.3.2, we obtain the risk bounds under misspecification described in Remark 3.4.2 and Remark 3.4.8.*

The risk of the estimator $\widehat{\boldsymbol{\theta}}$ in (A.3) can also be related to the tangent cones of the closed convex set K at $\boldsymbol{\theta}^*$. To describe these results, we need some notation and terminology. The tangent cone of $\mathcal{K}$ at $\boldsymbol{\theta} \in \mathcal{K}$ is defined as

$$\mathcal{T}_{\mathcal{K}}(\boldsymbol{\theta})\{t(\boldsymbol{\eta}-\boldsymbol{\theta}) : t \ge 0, \boldsymbol{\eta} \in \mathcal{K}\}.$$

Informally, $\mathcal{T}_{\mathcal{K}}(\boldsymbol{\theta})$ represents all directions in which one can move from $\boldsymbol{\theta}$ and still remain in K. Note that $\mathcal{T}_{\mathcal{K}}(\boldsymbol{\theta})$ is a cone which means that $a\boldsymbol{\alpha} \in \mathcal{T}_{\mathcal{K}}(\boldsymbol{\theta})$ for every $\boldsymbol{\alpha} \in \mathcal{T}_{\mathcal{K}}(\boldsymbol{\theta})$ and $a \ge 0$. It is also easy to see that $\mathcal{T}_{\mathcal{K}}(\boldsymbol{\theta})$ closed and convex.

The statistical dimension of a closed convex cone $\mathcal{T} \subseteq \mathbb{R}^n$ is defined as

$$\delta(\mathcal{T}) \coloneqq \mathbb{E}\|\Pi_{\mathcal{T}}(Z)\|^2, \qquad \text{where } Z \sim \mathcal{N}_n(0, \mathbf{I}_n)$$

and $\Pi_{\mathcal{T}}(Z) \coloneqq \operatorname{argmin}_{\mathbf{u}\in\mathcal{T}} \|Z-\mathbf{u}\|^2$ is the projection of Z onto T. The terminology of statistical dimension is due to Amelunxen et al. [3] and we refer the reader to this paper for many properties of the statistical dimension.

The relevance of these notions to the estimator $\widehat{\boldsymbol{\theta}}$ (defined in (A.3)) is that the risk of $\widehat{\boldsymbol{\theta}}$ can be related to the statistical dimension of tangent cones of K. This is the content of the following result due to Bellec [10, Corollary 2.2].

Theorem A.3.4. *Suppose $Y \sim \mathcal{N}_n(\boldsymbol{\theta}^*, \sigma^2\mathbf{I}_n)$ for some $\boldsymbol{\theta}^* \in \mathbb{R}^n$ and $\sigma^2 > 0$ and consider the estimator $\widehat{\boldsymbol{\theta}}$ defined in* (A.3) *for a closed convex set K. Then*

$$\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \le \inf_{\boldsymbol{\theta}\in\mathcal{K}} \left[\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|^2 + \sigma^2\delta(\mathcal{T}_{\mathcal{K}}(\boldsymbol{\theta}))\right]. \tag{A.6}$$

The statistical dimension $\delta(\mathcal{T})$ of a closed convex cone $\mathcal{T}$ is closely related to the Gaussian width of $\mathcal{T}$ which is defined as

$$w(\mathcal{T}) \coloneqq \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{T}:\|\boldsymbol{\theta}\|\le 1} \langle Z, \boldsymbol{\theta}\rangle\right] \qquad \text{where } Z \sim \mathcal{N}_n(0, \mathbf{I}_n). \tag{A.7}$$

Indeed, it has been shown in Amelunxen et al. [3, Proposition 10.2] that

$$w^2(\mathcal{T}) \le \delta(\mathcal{T}) \le w^2(\mathcal{T}) + 1$$

for every closed convex cone T. Using this relation in conjunction with (A.6), we obtain the following bound on the risk of the estimator $\widehat{\boldsymbol{\theta}}$ defined in (A.3) when $Y \sim \mathcal{N}_n(\boldsymbol{\theta}^*, \sigma^2\mathbf{I}_n)$:

$$\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \le \inf_{\boldsymbol{\theta}\in K} \left[\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|^2 + \sigma^2 + \sigma^2 w^2(T_K(\boldsymbol{\theta}))\right]. \tag{A.8}$$

A.3.2 Proof of Theorem 3.4.1

Let

$$\widehat{\boldsymbol{\theta}} \coloneqq (\widehat{f}_{\mathrm{EM}}(\mathbf{x}_1), \dots, \widehat{f}_{\mathrm{EM}}(\mathbf{x}_n)) \quad \text{and} \quad \boldsymbol{\theta}^* \coloneqq (f^*(\mathbf{x}_1), \dots, f^*(\mathbf{x}_n)) \tag{A.9}$$

and note that

$$\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) = \mathbb{E}\frac{1}{n}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2$$

where $\|\cdot\|$ denotes the usual Euclidean norm in $\mathbb{R}^n$.

Observe that by Proposition 3.3.1, it follows that $\widehat{\boldsymbol{\theta}} = \mathbf{A}\widehat{\boldsymbol{\beta}}_{\mathrm{EM}}$ is the projection of the data vector $\mathbf{y}$ on the closed convex cone

$$\mathcal{D}_{n_1,\dots,n_d} \coloneqq \{\mathbf{A}\boldsymbol{\beta} : \beta_j \ge 0, \forall j \ge 2\} = \{(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)) : f \in \mathcal{F}^d_{\mathrm{EM}}\}. \tag{A.10}$$

where $\mathbf{A}$ is the design matrix introduced in Section 3.3. Note that, under the lattice design (3.34), the set $\mathcal{D}_{n_1,\dots,n_d}$ is completely determined by the values of $n_1, \dots, n_d$. We can therefore employ Theorem A.3.1 to bound the risk $\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2/n$.

First, we claim that it suffices to prove the theorem under the assumption $n_j \geq 2$ for all $j = 1, \ldots, d$. To see this, note first that when $n = n_1 \cdots n_d = 1$, we have $\widehat{\boldsymbol{\theta}} = \mathbf{y}$ so that $\mathcal{R}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = \sigma^2/n$ and the result holds which means that we can assume that $\max_j n_j \geq 2$ for some j. Now if $n_j = 1$ for some values of j, we can simply ignore these components and focus on the equivalent problem with a lattice design (3.34) in a lower-dimensional space that has at least two grid points in each component. We can apply the bound (3.43) to this lower-dimensional problem (for instance, the dimension would be $d' = \#\{j : n_j \geq 2\}$ instead of d) and then remark that the bound (3.43) for the original problem is even larger.

Next, we claim that it suffices to prove the theorem under the assumption $\sigma^2 = 1$. Indeed in general we may consider the rescaled problem with $\widetilde{f} \coloneqq f^*/\sigma$, $\widetilde{V}^* \coloneqq V^*/\sigma$, and $\widetilde{y}_i \sim \mathcal{N}(\widetilde{f}(\mathbf{x}_i), 1)$, apply the bound (3.43), and then multiply the risk bound by σ^2 to account for rescaling the fitted function by σ. This is possible because $\mathcal{F}^d_{\mathrm{EM}}$ is a cone.

So, we assume $n_j \geq 2$ for all $j = 1, \ldots, d$ and $\sigma^2 = 1$. As mentioned above, we want to bound $\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2/n$ using Theorem A.3.1. For this, we need to obtain upper bounds for

$$G(t) \coloneqq \mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d} \cap \mathcal{B}_2(\boldsymbol{\theta}^*, t)} \langle \boldsymbol{\xi}, \boldsymbol{\theta} - \boldsymbol{\theta}^* \rangle$$

where $\boldsymbol{\xi} \sim \mathcal{N}_n(0, \mathbf{I}_n)$ and $\mathcal{B}_2(\boldsymbol{\theta}^*, t) \coloneqq \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < t\}$ denotes the ball of radius t centered at $\boldsymbol{\theta}^*$.

In what follows, we sometimes treat vectors in $\mathbb{R}^n$ as arrays in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ indexed by $\mathbf{i} = (i_1, \ldots, i_d)$ for $0 \leq i_j \leq n_j - 1$ and $j = 1, \ldots, d$.

For each $j \in 1, \ldots, d$, let

$$S_j^{(0)} \coloneqq \{i_j : 0 \leq i_j \leq \frac{n_j}{2} - 1\}, \qquad S_j^{(1)} \coloneqq \{i_j : \frac{n_j}{2} - 1 < i_j \leq n_j - 1\},$$

so that

$$\langle \boldsymbol{\xi}, \boldsymbol{\theta} - \boldsymbol{\theta}^* \rangle = \sum_{i_1=0}^{n_1-1} \cdots \sum_{i_d=0}^{n_d-1} \xi_{\mathbf{i}}(\theta_{\mathbf{i}} - \theta^*_{\mathbf{i}}) = \sum_{\mathbf{z} \in \{0,1\}^d} \sum_{\mathbf{i} \in S_1^{(z_1)} \times \cdots \times S_d^{(z_d)}} \xi_{\mathbf{i}}(\theta_{\mathbf{i}} - \theta^*_{\mathbf{i}}).$$

We then obtain the bound

$$G(t) \leq \sum_{\mathbf{z} \in \{0,1\}^d} \underbrace{\mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d} \cap \mathcal{B}_2(\boldsymbol{\theta}^*, t)} \sum_{\mathbf{i} \in S_1^{(z_1)} \times \cdots \times S_d^{(z_d)}} \xi_{\mathbf{i}}(\theta_{\mathbf{i}} - \theta^*_{\mathbf{i}})}_{=: H_{\mathbf{z}}(t)}. \tag{A.11}$$

We now bound $H_{\mathbf{z}}(t)$ for fixed $\mathbf{z} \in \{0,1\}^d$. For each $j = 1, \ldots, d$ let K_j denote the largest positive integer k_j for which

$$\left\{ i_j \in S_j^{(z_j)} n_j 2^{-(k_j+1)} - 1 + z_j n_j/2 < i_j \leq n_j 2^{-k_j} - 1 + z_j n_j/2 \right\}$$

is nonempty. Let $\mathcal{K} \coloneqq \bigtimes_{j=1}^{d}\{1,\dots,K_j\}$ and note that $|\mathcal{K}| = K_1K_2\cdots K_d$. For $\mathbf{k} \coloneqq (k_1,\dots,k_d) \in \mathcal{K}$ and $\boldsymbol{\theta} \in \mathbb{R}^{n_1\times\cdots\times n_d}$, let

$$\boldsymbol{\theta}^{(\mathbf{k})} \coloneqq \Big\{\boldsymbol{\theta}_{\mathbf{i}} : n_j 2^{-(k_j+1)} - 1 + z_j n_j/2 < i_j \le n_j 2^{-k_j} - 1 + z_j n_j/2, \\ i_j = 0,\dots,n_j-1,\, j=1,\dots,d\Big\}.$$

Let $\mathcal{M} \coloneqq \{(m_{\mathbf{k}})_{\mathbf{k}\in\mathcal{K}} : 1 \le m_{\mathbf{k}} \le |\mathcal{K}|, \sum_{\mathbf{k}\in\mathcal{K}} m_{\mathbf{k}} \le 2|\mathcal{K}|\}$. For $\mathbf{m} \in \mathcal{M}$, we define

$$T_{\mathbf{m}}(t) \coloneqq \left\{\boldsymbol{\theta} \in \mathcal{D}_{n_1,\dots,n_d} \cap \mathcal{B}_2(\boldsymbol{\theta}^*, t) : \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \le t, \|\boldsymbol{\theta}^{(\mathbf{k})} - (\boldsymbol{\theta}^*)^{(\mathbf{k})}\|^2 \le \frac{m_{\mathbf{k}}t^2}{|\mathcal{K}|}, \forall \mathbf{k}\in\mathcal{K}\right\}.$$

We claim

$$\mathcal{D}_{n_1,\dots,n_d} \cap \mathcal{B}_2(\boldsymbol{\theta}^*, t) \subseteq \bigcup_{\mathbf{m}\in\mathcal{M}} T_{\mathbf{m}}(t). \tag{A.12}$$

Indeed suppose $\boldsymbol{\theta} \in \mathcal{D}_{n_1,\dots,n_d} \cap \mathcal{B}_2(\boldsymbol{\theta}^*, t)$; then we have $t^2 \ge \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|^2 \ge \|\boldsymbol{\theta}^{(\mathbf{k})} - (\boldsymbol{\theta}^*)^{(\mathbf{k})}\|^2$ for each $\mathbf{k}$, and thus there exists $1 \le m_{\mathbf{k}} \le |\mathcal{K}|$ such that

$$m_{\mathbf{k}} - 1 \le |\mathcal{K}| \frac{\|\boldsymbol{\theta}^{(\mathbf{k})} - (\boldsymbol{\theta}^*)^{(\mathbf{k})}\|^2}{t^2} \le m_{\mathbf{k}}.$$

This implies

$$1 \ge t^{-2}\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|^2 \ge t^{-2}\sum_{\mathbf{k}\in\mathcal{K}}\|\boldsymbol{\theta}^{(\mathbf{k})} - (\boldsymbol{\theta}^*)^{(\mathbf{k})}\|^2 \ge |\mathcal{K}|^{-1}\sum_{\mathbf{k}\in\mathcal{K}}(m_{\mathbf{k}}-1)$$

and thus $\sum_{\mathbf{k}\in\mathcal{K}} m_{\mathbf{k}} \le 2|\mathcal{K}|$, so $\mathbf{m} \in \mathcal{M}$ and $\boldsymbol{\theta} \in T_{\mathbf{m}}(t)$, which verifies the claim (A.12).

Using this claim (A.12) we obtain

$$H_{\mathbf{z}}(t) \le \mathbb{E}\max_{\mathbf{m}\in\mathcal{M}}\sup_{\boldsymbol{\theta}\in T_{\mathbf{m}}} \sum_{\mathbf{i}\in S_1^{(z_1)}\times\cdots\times S_d^{(z_d)}} \xi_{\mathbf{i}}(\theta_{\mathbf{i}} - \theta^*_{\mathbf{i}}).$$

Lemma D.1 from [44] then implies

$$H_{\mathbf{z}}(t) \le \max_{\mathbf{m}\in\mathcal{M}} \mathbb{E}\sup_{\boldsymbol{\theta}\in T_{\mathbf{m}}(t)} \sum_{\mathbf{i}\in S_1^{(z_1)}\times\cdots\times S_d^{(z_d)}} \xi_{\mathbf{i}}(\theta_{\mathbf{i}} - \theta^*_{\mathbf{i}}) + t\sqrt{2\log|\mathcal{M}|} + t\sqrt{\pi/2}.$$

Because the number of $|\mathcal{K}|$-tuples of positive integers summing to p is $\binom{p-1}{|\mathcal{K}|-1} = \binom{p-1}{p-|\mathcal{K}|}$, we can bound the cardinality of $\mathcal{M}$ by

$$|\mathcal{M}| \le \sum_{p=|\mathcal{K}|}^{2|\mathcal{K}|}\binom{p-1}{p-|\mathcal{K}|} \le \sum_{p=|\mathcal{K}|}^{2|\mathcal{K}|}\binom{2|\mathcal{K}|-1}{p-|\mathcal{K}|} \le 2^{2|\mathcal{K}|-1}.$$

Thus,

$$H_{\mathbf{z}}(t) \le \max_{\mathbf{m}\in\mathcal{M}} \underbrace{\mathbb{E} \sup_{\boldsymbol{\theta}\in T_{\mathbf{m}}(t)} \sum_{\mathbf{i}\in S_1^{(z_1)}\times\cdots\times S_d^{(z_d)}} \xi_{\mathbf{i}}(\theta_{\mathbf{i}} - \theta^*_{\mathbf{i}})}_{=:U_{\mathbf{z},\mathbf{m}}(t)} + 2t\sqrt{|\mathcal{K}|} + t\sqrt{\pi/2}. \tag{A.13}$$

Since $\sum_{\mathbf{i}\in S_1^{(z_1)}\times\cdots\times S_d^{(z_d)}} \xi_{\mathbf{i}}(\theta_{\mathbf{i}} - \theta^*_{\mathbf{i}}) = \sum_{\mathbf{k}\in\mathcal{K}} \left\langle \boldsymbol{\xi}^{(\mathbf{k})}, \boldsymbol{\theta}^{(\mathbf{k})} - (\boldsymbol{\theta}^*)^{(\mathbf{k})} \right\rangle$, we have

$$U_{\mathbf{z},\mathbf{m}}(t) \le \sum_{\mathbf{k}\in\mathcal{K}} \underbrace{\mathbb{E} \sup_{\substack{\boldsymbol{\theta}\in\mathcal{D}_{n_1,\ldots,n_d}\cap\mathcal{B}_2(\boldsymbol{\theta}^*,t):\\ \|\boldsymbol{\theta}^{(\mathbf{k})} - (\boldsymbol{\theta}^*)^{(\mathbf{k})}\|^2 \le t^2 m_{\mathbf{k}}/|\mathcal{K}|}} \left\langle \boldsymbol{\xi}^{(\mathbf{k})}, \boldsymbol{\theta}^{(\mathbf{k})} - (\boldsymbol{\theta}^*)^{(\mathbf{k})} \right\rangle}_{=:U_{\mathbf{z},\mathbf{m},\mathbf{k}}(t)}. \tag{A.14}$$

We claim that for any $\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d} \cap \mathcal{B}_2(\boldsymbol{\theta}^*, t)$ and any $\mathbf{i} \in \bigtimes_{j=1}^d \{0, 1, \ldots, n_j - 1\}$ and $\mathbf{k} \in \mathcal{K}$ satisfying

$$n_j 2^{-(k_j+1)} - 1 + z_j n_j/2 < i_j \le n_j 2^{-k_j} - 1 + z_j n_j/2, \tag{A.15}$$

then $\theta_{\mathbf{i}}$ can be bounded as

$$\theta^*_{\mathbf{0}} - t(2^{d+k_+}/n)^{1/2} \le \theta_{\mathbf{i}} \le \theta^*_{\mathbf{n-1}} + t(2^{d+k_+}/n)^{1/2}. \tag{A.16}$$

where $k_+ \coloneqq k_1 + \cdots + k_d$. We prove each bound by contradiction. If the upper bound of (A.16) does not hold, then

$$\theta_{\boldsymbol{\ell}} \ge \theta_{\mathbf{i}} > \theta^*_{\mathbf{n-1}} + t(2^{d+k_+}/n)^{1/2} \ge \theta^*_{\boldsymbol{\ell}} + t(2^{d+k_+}/n)^{1/2}$$

as long as $\boldsymbol{\ell} \succeq \mathbf{i}$, which yields

$$t^2 \ge \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2 \ge \sum_{\boldsymbol{\ell}\succeq\mathbf{i}} (\theta_{\boldsymbol{\ell}} - \theta^*_{\boldsymbol{\ell}})^2 > t^2 2^{d+k_+} n^{-1} \cdot \prod_{j=1}^d (n_j - i_j).$$

Noting that our condition on i_j (A.15) implies $n_j - i_j \ge n_j(1 - z_j/2 - 2^{-k_j}) \ge n_j 2^{-(k_j+1)}$, we obtain $\prod_{j=1}^d (n_j - i_j + 1) \ge n 2^{-(d+k_+)}$ which yields the contradiction $t^2 > t^2$.

Similarly if the lower bound of (A.16) does not hold, then

$$\theta_{\boldsymbol{\ell}} \le \theta_{\mathbf{i}} < \theta^*_{\mathbf{n-1}} - t(2^{d+k_+}/n)^{1/2} \le \theta^*_{\boldsymbol{\ell}} - t(2^{d+k_+}/n)^{1/2}$$

as long as $\boldsymbol{\ell} \preceq \mathbf{i}$, which yields

$$t^2 \ge \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2 \ge \sum_{\boldsymbol{\ell}\preceq\mathbf{i}} (\theta_{\boldsymbol{\ell}} - \theta^*_{\boldsymbol{\ell}})^2 > t^2 2^{d+k_+} n^{-1} \cdot \prod_{j=1}^d (i_j + 1).$$

Noting that our condition on i_j (A.15) implies $i_j + 1 > n_j 2^{-(k_j+1)}$ we obtain $\prod_{j=1}^d (i_j + 1) \ge n 2^{-(d+k_+)}$ which yields the contradiction $t^2 > t^2$.

Thus, the bounds (A.16) hold. So, for each $\boldsymbol{\theta} \in \mathcal{D}_{n_1,\dots,n_d} \cap \mathcal{B}_2(\boldsymbol{\theta}^*, t)$ and $\mathbf{k} \in \mathcal{K}$, the number of entries in $\boldsymbol{\theta}^{(\mathbf{k})}$ is at most

$$\prod_{j=1}^{d} (n_j 2^{-k_j} - n_j 2^{-(k_j+1)}) \le n 2^{-(d+k_+)},$$

and each entry lies in the interval

$$[a, b] \coloneqq \left[\theta^*_{\mathbf{0}} - t(2^{d+k_+}/n)^{1/2},\ \theta^*_{\mathbf{n-1}} + t(2^{d+k_+}/n)^{1/2}\right].$$

Moreover, $\boldsymbol{\theta}^{(\mathbf{k})}$ lies in some $\widetilde{\mathcal{D}} \coloneqq \mathcal{D}_{\widetilde{n}_1,\dots,\widetilde{n}_d}$ where $\widetilde{n}_1, \dots, \widetilde{n}_d$ are the dimensions of $\boldsymbol{\theta}^{(\mathbf{k})}$ as a sub-array.

We make use of the following metric entropy result, proved in Section A.5.1

Lemma A.3.5. *For $a < b$, we have*

$$\log N_2(\epsilon, \mathcal{D}_{n_1,\dots,n_d} \cap [a,b]^n) \le C_d \frac{(b-a)\sqrt{n}}{\epsilon} \left(\log \frac{(b-a)\sqrt{n}}{\epsilon}\right)^{d-\frac{1}{2}} \mathbb{I}\{\epsilon \le (b-a)\sqrt{n}\}.$$

Combining this metric entropy bound with Dudley's entropy bound [31] (for instance see [20, Thm. 3.2]) yields

$$U_{\mathbf{z},\mathbf{m},\mathbf{k}}(t) \le c \int_0^{t\sqrt{m_{\mathbf{k}}/|\mathcal{K}|}} \sqrt{\frac{B}{\epsilon}\left(\log \frac{B}{\epsilon}\right)^{d-\frac{1}{2}}}\, d\epsilon,$$

where

$$\begin{aligned} B &\coloneqq (n2^{-(d+k_+)})^{1/2}(V^* + 2t(n2^{-(d+k_+)})^{-1/2}) \\ &= (n2^{-(d+k_+)})^{1/2} V^* + 2t \end{aligned} \tag{A.17}$$

and $V^* = f^*(\mathbf{1}) - f^*(\mathbf{0}) \ge \theta^*_{\mathbf{n-1}} - \theta^*_{\mathbf{0}}$. Note that $\epsilon \le t\sqrt{m_{\mathbf{k}}/|\mathcal{K}|} \le t < B$, so $\log(B/\epsilon) > 0$.

The following lemma (proved in Section A.5.2) allows us to bound the above integral.

Lemma A.3.6. *For every $d \ge 1$ there exists a positive constant C_d such that for every $s \in (0, B]$, the following inequality holds.*

$$\int_0^s \sqrt{\frac{B}{\epsilon}\left(\log \frac{B}{\epsilon}\right)^{d-\frac{1}{2}}}\, d\epsilon \le C_d \sqrt{sB} \left(\log \frac{B}{s}\right)^{\frac{2d-1}{4}}$$

Applying Lemma A.3.6 with $s \coloneqq t\sqrt{m_{\mathbf{k}}/|\mathcal{K}|} \ge t/\sqrt{|\mathcal{K}|}$ yields

$$U_{\mathbf{z},\mathbf{m},\mathbf{k}}(t) \le C_d \sqrt{Bt} (m_{\mathbf{k}}/|\mathcal{K}|)^{1/4} \left(\log \frac{eB\sqrt{|\mathcal{K}|}}{t}\right)^{\frac{2d-1}{4}}.$$

We bound this with two terms depending on which of the two terms in the definition (A.17) of B is larger. In the case $V^*(n2^{-(d+k_+)})^{1/2} > 2t$, we have $B \leq 2V^*(n2^{-(d+k_+)})^{1/2} \leq 2V^*\sqrt{n}$ and

$$U_{\mathbf{z},\mathbf{m},\mathbf{k}}(t) \leq C_d\sqrt{tV^*}(n2^{-k_+})^{1/4}\left(\log\frac{2eV^*\sqrt{n|\mathcal{K}|}}{t}\right)^{\frac{2d-1}{4}}.$$

In the other case where $V^*(n2^{-(d+k_+)})^{1/2} \leq 2t$, we have $B \leq 3t$, which yields

$$U_{\mathbf{z},\mathbf{m},\mathbf{k}}(t) \leq C_d t(m_{\mathbf{k}}/|\mathcal{K}|)^{1/4}(\log(2e\sqrt{|\mathcal{K}|}))^{\frac{2d-1}{4}}.$$

Combining the two cases and using the indicator bounds $\mathbb{I}\{V^*(n2^{-(d+k_+)})^{1/2} > 2t\} \leq \mathbb{I}\{V^*\sqrt{n} > t\}$ and $\mathbb{I}\{V^*(n2^{-(d+k_+)})^{1/2} \leq 2t\} \leq 1$, we obtain

$$\begin{aligned}U_{\mathbf{z},\mathbf{m},\mathbf{k}}(t) \leq{}& C_d\sqrt{tV^*}(n2^{-k_+})^{1/4}\left(\log\frac{2eV^*\sqrt{n|\mathcal{K}|}}{t}\right)^{\frac{2d-1}{4}}\mathbb{I}\{V^*\sqrt{n} > t\}\\ &+ C_d t(m_{\mathbf{k}}/|\mathcal{K}|)^{1/4}(\log(2e\sqrt{|\mathcal{K}|}))^{\frac{2d-1}{4}}.\end{aligned}$$

Applying this observation to the earlier bound $U_{\mathbf{z},\mathbf{m}}(t) \leq \sum_{\mathbf{k}\in\mathcal{K}} U_{\mathbf{z},\mathbf{m},\mathbf{k}}(t)$ from (A.14) yields

$$\begin{aligned}U_{\mathbf{z},\mathbf{m}}(t) \leq{}& C_d\sqrt{tV^*}n^{1/4}\left(\log\frac{2eV^*\sqrt{n|\mathcal{K}|}}{t}\right)^{\frac{2d-1}{4}}\mathbb{I}\{V^*\sqrt{n} > t\}\sum_{\mathbf{k}\in\mathcal{K}}2^{-k_+/4}\\ &+ C_d t(\log(2e\sqrt{|\mathcal{K}|}))^{\frac{2d-1}{4}}\sum_{\mathbf{k}\in\mathcal{K}}(m_{\mathbf{k}}/|\mathcal{K}|)^{1/4}.\end{aligned}$$

The first sum can be bounded as

$$\sum_{\mathbf{k}\in\mathcal{K}}2^{-k_+/4} \leq \prod_{j=1}^{d}\sum_{k_j=1}^{\infty}2^{-k_j/4} \leq C_d.$$

For the second sum, note that Hölder's inequality combined with the fact that $\sum_{\mathbf{k}\in\mathcal{K}} m_{\mathbf{k}} \leq 2|\mathcal{K}|$ yields

$$\sum_{\mathbf{k}\in\mathcal{K}} m_{\mathbf{k}}^{1/4} \leq \left(\sum_{\mathbf{k}\in\mathcal{K}} m_{\mathbf{k}}\right)^{1/4}|\mathcal{K}|^{3/4} \leq 2^{1/4}|\mathcal{K}|.$$

Additionally, note that $K_j \leq C\log n_j$ for each j, so $\log|\mathcal{K}| \leq \sum_{j=1}^{d}\log(C\log n_j) \leq C_d\log n$, which allows us to bound the logarithmic term as

$$\log\frac{2eV^*\sqrt{n|\mathcal{K}|}}{t} \leq \log\frac{2eV^*\sqrt{n}}{t} + \frac{1}{2}\log|\mathcal{K}| \leq C_d\log\frac{eV^*\sqrt{n}}{t}.$$

Finally, note that

$$\left(\log \frac{eV^*\sqrt{n}}{t}\right)^{\frac{2d-1}{4}} \mathbb{I}\{V^*\sqrt{n} > t\} = \left(\log_+ \frac{eV^*\sqrt{n}}{t}\right)^{\frac{2d-1}{4}},$$

where $\log_+(x) \coloneqq \max(0, \log x)$.

Combining these four observations yields

$$U_{\mathbf{z},\mathbf{m}}(t) \le C_d\sqrt{tV^*}n^{1/4}\left(\log_+ \frac{eV^*\sqrt{n}}{t}\right)^{\frac{2d-1}{4}} + C_d t|\mathcal{K}|^{3/4}(\log(2e\sqrt{|\mathcal{K}|}))^{\frac{2d-1}{4}}.$$

Combining this bound with the earlier bound (A.13) on $H_{\mathbf{z}}(t)$ yields

$$\begin{aligned} H_{\mathbf{z}}(t) &\le \max_{\mathbf{m}\in\mathcal{M}} U_{\mathbf{z},\mathbf{m}}(t) + 2t\sqrt{|\mathcal{K}|} + t\sqrt{\pi/2} \\ &\le \underbrace{C_d\sqrt{tV^*}n^{1/4}\left(\log_+ \frac{eV^*\sqrt{n}}{t}\right)^{\frac{2d-1}{4}}}_{=:G_1(t)} + \underbrace{C_d t|\mathcal{K}|^{3/4}(\log(2e\sqrt{|\mathcal{K}|}))^{\frac{2d-1}{4}}}_{=:G_2(t)}. \end{aligned}$$

By observing the earlier bound (A.11), we see that the above upper bound for $H_{\mathbf{z}}(t)$ also holds for $G(t)$ (after multiplying the constants by 2^d). That is,

$$G(t) \le G_1(t) + G_2(t). \tag{A.18}$$

where G_1 and G_2 are the two terms of the previous inequality. Let

$$t_1 \coloneqq \max\{1, (4C_d)^{2/3}\}(\sqrt{n}V^*)^{1/3}\left[\max\{1, \log_+(e(\sqrt{n}V^*)^{2/3})\}\right]^{\frac{2d-1}{6}}.$$

Then $t_1 \ge (\sqrt{n}V^*)^{1/3}$, so for $t \ge t_1$ we have

$$\begin{aligned} \frac{G_1(t)}{t^2} &= C_d\frac{\sqrt{V^*}n^{1/4}}{t^{3/2}}\left(\log_+ \frac{eV^*\sqrt{n}}{t}\right)^{\frac{2d-1}{4}} \\ &\le C_d\frac{\sqrt{V^*}n^{1/4}}{t^{3/2}}\left(\log_+(e(V^*\sqrt{n})^{2/3})\right)^{\frac{2d-1}{4}} \le \frac{1}{4}. \end{aligned}$$

Next, with the definition

$$t_2 \coloneqq 4C_d|\mathcal{K}|^{3/4}(\log(2e\sqrt{|\mathcal{K}|}))^{\frac{2d-1}{4}},$$

for $t \ge t_2$ we have

$$\frac{G_2(t)}{t^2} = \frac{C_d|\mathcal{K}|^{3/4}}{t}(\log(2e\sqrt{|\mathcal{K}|}))^{\frac{2d-1}{4}} \le \frac{1}{4}.$$

Combining the two inequalities, we obtain $G(t) \leq t^2/2$ for $t \geq \max\{t_1, t_2\}$. By Theorem A.3.1 and the bound $K_j \leq c \log n_j$, we obtain

$$\begin{aligned}
\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) = \mathbb{E}\frac{1}{n}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 &\leq \frac{t_1^2 + t_2^2}{n} \\
&\leq C_d\left(\frac{V^*}{n}\right)^{\frac{2}{3}}\left[\max\{1, \log_+(e(\sqrt{n}V^*)^{2/3})\}\right]^{\frac{2d-1}{3}} \\
&\quad + \frac{C_d}{n}\left(\prod_{j=1}^{d}\log n_j\right)^{\frac{3}{2}}\left(\sum_{j=1}^{d}\log(e\log n_j)\right)^{\frac{2d-1}{2}} \\
&\leq C_d\left(\frac{V^*}{n}\right)^{\frac{2}{3}}\left[\log(2 + \sqrt{n}V^*)\right]^{\frac{2d-1}{3}} \\
&\quad + \frac{C_d}{n}(\log n)^{\frac{3d}{2}}(\log(e\log n))^{\frac{2d-1}{2}}.
\end{aligned}$$

A.3.3 Proof of Theorem 3.4.5

We use the earlier notation (A.9). As observed in the proof of Theorem 3.4.1, it follows from Proposition 3.3.1 and Proposition 3.3.2 that $\widehat{\boldsymbol{\theta}}$ is the projection of the data vector $\mathbf{y}$ onto the closed convex cone (A.10). We then apply Theorem A.3.4 to obtain

$$\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) = \mathbb{E}\frac{1}{n}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leq \inf_{\boldsymbol{\theta}\in K}\left\{\frac{1}{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2 + \frac{\sigma^2}{n}\delta(T_K(\boldsymbol{\theta}))\right\}$$

where $K = \mathcal{D}_{n_1,\ldots,n_d}$ is the set (A.10). Using the notation $\boldsymbol{\theta}_f := (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$ for $f \in \mathcal{F}^d_{\mathrm{EM}}$, we can rewrite the above inequality as

$$\begin{aligned}
\mathcal{R}(\widehat{f}_{\mathrm{EM}}, f^*) &\leq \inf_{f\in\mathcal{F}^d_{\mathrm{EM}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(f(\mathbf{x}_i) - f^*(\mathbf{x}_i)\right)^2 + \frac{\sigma^2}{n}\delta(T_K(\boldsymbol{\theta}_f))\right\} \\
&\leq \inf_{f\in\mathfrak{R}^d\cap\mathcal{F}^d_{\mathrm{EM}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(f(\mathbf{x}_i) - f^*(\mathbf{x}_i)\right)^2 + \frac{\sigma^2}{n}\delta(T_K(\boldsymbol{\theta}_f))\right\}.
\end{aligned}$$

Therefore to complete the proof of Theorem 3.4.5, it is enough to show that

$$\delta(T_K(\boldsymbol{\theta}_f)) \leq C_d k(f)(\log(en))^{\frac{3d}{2}}(\log\log n)^{\frac{2d-1}{2}} \qquad \text{for every } f \in \mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{EM}}. \tag{A.19}$$

Fix $f \in \mathfrak{R}^d \cap \mathcal{F}^d_{\mathrm{EM}}$ with $k(f) = k$. By the definition of $\mathfrak{R}^d$, there exist d univariate partitions as in (3.21) such that f is constant on each of the k rectangles

$$R_{l_1,\ldots,l_d} := \prod_{s=1}^{d}[x^{(s)}_{l_s}, x^{(s)}_{l_s+1}) \qquad l_s = 0, 1, \ldots, k_s - 1 \text{ and } s = 1, \ldots, d. \tag{A.20}$$

For every $s = 1, \ldots, d$ and $l_s = 0, 1, \ldots, k_s - 1$, let $n_s(l_s)$ be the number of indices $i_s = 0, 1, \ldots, n_s - 1$ such that $i_s/n_s \in [x_{l_s}^{(s)}, x_{l_s+1}^{(s)})$. It will be convenient in the sequel to, as in Section 3.3.1, index vectors in $\mathbb{R}^n$ by $(i_1, \ldots, i_d) \in \mathcal{I}$ (recall that $\mathcal{I}$ is defined as in (3.37)). Specifically the components of $\boldsymbol{\theta} \in R^n$ will be denoted by $\theta_{i_1,\ldots,i_d}, (i_1, \ldots, i_d) \in \mathcal{I}$. Also, for $\boldsymbol{\theta} \in \mathbb{R}^n$ and the rectangle (A.20), let $\boldsymbol{\theta}(R_{l_1,\ldots,l_d})$ denote the vector in $\mathbb{R}^{n_1(l_1)} \times \cdots \times \mathbb{R}^{n_d(l_d)}$ with components given by $\theta_{i_1,\ldots,i_d}$ as each i_s varies over the indices in $0, 1, \ldots, n_s - 1$ such that $i_s/n_s \in [x_{l_s}^{(s)}, x_{l_s+1}^{(s)})$. We now make the key observation that for every $\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d}$ and rectangle $R_{l_1,\ldots,l_d}$ in (A.20), we have

$$\boldsymbol{\theta}(R_{l_1,\ldots,l_d}) \in \mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}. \tag{A.21}$$

To see this, fix $\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d}$ and let $f \in \mathcal{F}_{\mathrm{EM}}^d$ be such that $\theta_{i_1,\ldots,i_d} = f(i_1/n_1, \ldots, i_d/n_d)$ for every $i_1, \ldots, i_d$. Then

$$\begin{aligned}\boldsymbol{\theta}(R_{l_1,\ldots,l_d}) &= \left\{ \left(f(\frac{i_1}{n_1}), \ldots, f(\frac{i_d}{n_d}) \right) : \frac{i_s}{n_s} \in [x_{l_s}^{(s)}, x_{l_s+1}^{(s)}), s = 1, \ldots, d \right\} \\ &= \left\{ \left(g(\frac{j_1}{n_1(l_1)}, \ldots, \frac{j_d}{n_d(l_d)}) \right) : j_s = 0, 1, \ldots, n_s(l_s) - 1, s = 1, \ldots, d \right\}\end{aligned}$$

where $g : [0,1]^d \to \mathbb{R}$ is defined as

$$g(x_1, \ldots, x_d) := f\left((1 - x_1)x_{l_1}^{(1)} + x_1 x_{l_1+1}^{(1)}, \ldots, (1 - x_d)x_{l_d}^{(d)} + x_d x_{l_1+d}^{(d)} \right)$$

It is easy to see that $g \in \mathcal{F}_{\mathrm{EM}}^d$ which proves (A.21). The fact (A.21) will be used to prove (A.19) in the following way. We first observe that

$$T_K(\boldsymbol{\theta}_f) \subseteq \left\{ v \in \mathbb{R}^n : v(R_{l_1,\ldots,l_d}) \in \mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}, \forall l_s = 0, 1, \ldots, k_s - 1, \forall s = 1, \ldots, d \right\} \tag{A.22}$$

To prove (A.22), note first that, by the definition of the tangent cone, we have

$$T_K(\boldsymbol{\theta}_f) = \text{Closure}\left\{ \alpha(\boldsymbol{\theta} - \boldsymbol{\theta}_f) : \theta \in K, \alpha \geq 0 \right\}.$$

Since the right hand side of (A.22) is a closed set, we only need to show that $v = \alpha(\boldsymbol{\theta} - \boldsymbol{\theta}_f)$ belongs to the right hand side of (A.22) for every $\boldsymbol{\theta} \in K$ and $\alpha \geq 0$. Fix $l_1, \ldots, l_d$. By (A.21), we have that $\boldsymbol{\theta}(R_{l_1,\ldots,l_d}) \in \mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}$. On the other hand, $\boldsymbol{\theta}_f(R_{l_1,\ldots,l_d})$ is a constant vector, because f is constant on $R_{l_1,\ldots,l_d}$. As a result, with $R = R_{l_1,\ldots,l_d}$, we obtain that $v(R) = \alpha(\boldsymbol{\theta}(R) - \boldsymbol{\theta}_f(R)) \in \mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}$ as $\mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}$ is a cone that is invariant under translation by constant vectors. This proves (A.22).

The observation (A.22) implies (using the monotonicity of statistical dimension; see Amelunxen et al. [3, Proposition 3.1]) that $\delta(T_K(\boldsymbol{\theta}_f)) \leq \delta(T)$ where T denotes the right hand side of (A.22). It is now easy to see that

$$\delta(T) = \sum_{l_1=0}^{k_1-1} \cdots \sum_{l_d=0}^{k_d-1} \mathbb{E} \| \Pi_{\mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}}(Z(R_{l_1,\ldots,l_d})) \|^2$$

where $Z \sim \mathcal{N}_n(0, \mathbf{I}_n)$ and $\Pi_{\mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}}$ is the projection operator on the closed convex set $\mathcal{D}_{n_1(l_1),\ldots,n_d(l_d)}$ Each addend on the right-hand side is simply the risk of the NNLS estimator $\widehat{f}_{\mathrm{EM}}$ when the design points are $(j_1/n_1(l_1), \ldots, j_d/n_d(l_d)), j_s = 0, 1, \ldots, n_s(l_s)-1, s = 1, \ldots, d$ and when the true function f^* is constantly equal to zero. Thus, by the second term in (3.43), and noting that the number of design points here is $\prod_{s=1}^d n_s(l_s) \le n$, we obtain

$$\delta(T_K(\boldsymbol{\theta}_f)) \le \delta(T) \le C_d k (\log(en))^{\frac{3d}{2}} (\log(e\log(en)))^{\frac{2d-1}{2}},$$

which proves (A.19) and completes the proof of Theorem 3.4.5.

A.3.4 Proof of Theorem 3.4.6

Let

$$\widehat{\boldsymbol{\theta}} \coloneqq (\widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_n)) = \mathbf{A}\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V} \quad \text{and} \quad \boldsymbol{\theta}^* \coloneqq (f^*(\mathbf{x}_1), \ldots, f^*(\mathbf{x}_n)) \tag{A.23}$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$ is defined by the LASSO problem (3.30). Note that $\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) = \frac{1}{n}\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2$.

Similar to the proof of Theorem 3.4.1, we take $\sigma = 1$ without loss of generality. To see this, note that we can consider the scaled problem $y_i/\sigma = f^*(\mathbf{x}_i)/\sigma + \xi_i/\sigma$ so that noise is scaled to have variance 1 and the variation is now $V_{\mathrm{HK}\mathbf{0}}(f^*/\sigma, [0,1]^d) \le V/\sigma$. Note also that the estimator for the scaled problem is $\widehat{f}_{\mathrm{HK}\mathbf{0},V}/\sigma$ where $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ is the estimator in the original problem. We may apply the bound (3.48) to the scaled problem, and convert this into a bound on the risk of the original problem by multiplying the bound by σ^2 and replacing the variation term V/σ with V. Thus, for the rest of the proof we assume $\sigma = 1$.

Observe first that $\widehat{\boldsymbol{\theta}}$ is the projection of $\mathbf{y}$ on the closed convex set $\mathcal{C}(V)$ defined in (3.31). We use Theorem A.3.1 to bound $\mathbb{E}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2$ and the key is to bound the quantity

$$G(t) \coloneqq \mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{C}(V): \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 \le t} \langle \boldsymbol{\xi}, \boldsymbol{\theta} - \boldsymbol{\theta}^* \rangle \qquad \text{for } t > 0 \tag{A.24}$$

where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ in order to find $t_* > 0$ such that $G(t_*) \le t_*^2/2$.

Throughout, $\mathbf{A}$ is the design matrix from Section 3.3. If $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta}$ and $\boldsymbol{\theta}^* = \mathbf{A}\boldsymbol{\beta}^*$ both belong to $\mathcal{C}(V)$ then $\sum_{j=2}^n |\beta_j - \beta_j^*| \le \sum_{j=2}^n |\beta_j| + \sum_{j=2}^n |\beta_j^*| \le 2V$, so we have

$$G(t) \le H(t) \coloneqq \mathbb{E} \sup_{\boldsymbol{\alpha} \in \mathcal{C}(2V): \|\boldsymbol{\alpha}\|_2 \le t} \langle \boldsymbol{\xi}, \boldsymbol{\alpha} \rangle.$$

Let $\mathcal{C}(V, t) \coloneqq \mathcal{C}(V) \cap \mathcal{B}_2(\mathbf{0}, t)$. We now use Dudley's entropy bound (see Chatterjee et al. [20, Thm. 3.2]) to control the right hand side above:

$$H(t) \le c \int_0^t \sqrt{\log N(\epsilon, \mathcal{C}(2V, t))}\, d\epsilon.$$

The covering numbers above are bounded in the following lemma whose proof is deferred to Section A.5.3.

Lemma A.3.7. *For every $V > 0$ and $t > 0$, we have*

$$\log N(\epsilon, \mathcal{C}(V,t)) \le C_d\left(\frac{V\sqrt{n}}{\epsilon}+1\right)\left(\log\left(\frac{2V\sqrt{n}}{\epsilon}+1\right)\right)^{d-\frac{1}{2}} + \log\left(2+2\frac{t+V\sqrt{n}}{\epsilon}\right).$$

Lemma A.3.7 and the inequality $\sqrt{a^2+b^2} \le a+b$ for $a, b \ge 0$ together give

$$\begin{aligned}\sqrt{\log N(\epsilon, \mathcal{C}(2V,t))} \le{}& C_d\sqrt{\left(\frac{V\sqrt{n}}{\epsilon}+1\right)\left(\log\left(\frac{2V\sqrt{n}}{\epsilon}+1\right)\right)^{d-\frac{1}{2}}} \\ &+ C_d\sqrt{\log\left(2+2\frac{t+V\sqrt{n}}{\epsilon}\right)}\end{aligned}$$

and thus

$$\begin{aligned}H(t) \le{}& C_d\int_0^t \sqrt{\left(\frac{V\sqrt{n}}{\epsilon}+1\right)\left(\log\left(\frac{2V\sqrt{n}}{\epsilon}+1\right)\right)^{d-\frac{1}{2}}}\, d\epsilon \\ &+ C_d\int_0^t \sqrt{\log\left(2+2\frac{t+V\sqrt{n}}{\epsilon}\right)}\, d\epsilon\end{aligned}$$

We can upper bound the second integral as follows.

Let $B \coloneqq 4t + 2V\sqrt{n}$. Using the fact that $\epsilon \le t$ in the integral, and peforming some substitutions and integration by parts, we obtain

$$\begin{aligned}&\int_0^t \sqrt{\log\left(2+2\frac{t+V\sqrt{n}}{\epsilon}\right)}\, d\epsilon \\ &\le \int_0^t \sqrt{\log\frac{4t+2V\sqrt{n}}{\epsilon}}\, d\epsilon \\ &= \int_0^t \sqrt{\log\frac{B}{\epsilon}}\, d\epsilon \\ &= B\int_\alpha^\infty u^{1/2}e^{-u}\, du && u = \log\frac{B}{\epsilon}, \alpha \coloneqq \log\frac{B}{t} \\ &= B\sqrt{\alpha}e^{-\alpha} + B\int_\alpha^\infty \frac{e^{-u}}{2\sqrt{u}}\, du\end{aligned}$$

where the last step is due to integration by parts. The last integral can be bounded by

$$\int_\alpha^\infty \frac{e^{-u}}{2\sqrt{u}}\, du \le \frac{1}{2\sqrt{\alpha}}\int_\alpha^\infty e^{-u}\, du \le \frac{1}{2\sqrt{\alpha}}e^{-\alpha}.$$

Noting that $\alpha = \log(B/t) \geq \log(4)$ and $Be^{-\alpha} = t$, we obtain

$$\begin{aligned}\int_0^t \sqrt{\log\left(2 + 2\frac{t + V\sqrt{n}}{\epsilon}\right)}\, d\epsilon &\leq Be^{-\alpha}\left(\sqrt{\alpha} + \frac{1}{2\sqrt{\alpha}}\right) \\ &\leq Ct\sqrt{1 + \log(B/t)} \\ &\leq Ct\sqrt{\log(4 + 2V\sqrt{n}/t)}.\end{aligned}$$

We now return to the first integral.

$$\begin{aligned}&C_d \int_0^t \sqrt{\left(\frac{V\sqrt{n}}{\epsilon} + 1\right)\left(\log\left(\frac{2V\sqrt{n}}{\epsilon} + 1\right)\right)^{d-\frac{1}{2}}}\, d\epsilon \\ &\leq C_d \int_0^t \sqrt{\frac{V\sqrt{n} + t}{\epsilon}\left(\log\frac{2V\sqrt{n} + t}{\epsilon}\right)^{d-\frac{1}{2}}}\, d\epsilon \\ &\leq C_d\sqrt{t(2V\sqrt{n} + t)}\left(\log\frac{e(2V\sqrt{n} + t)}{t}\right)^{\frac{2d-1}{4}} \\ &\leq C_d\left(t + \sqrt{2tV\sqrt{n}}\right)\left(\log(1 + 2eV\sqrt{n}/t)\right)^{\frac{2d-1}{4}},\end{aligned}$$

where we have used Lemma A.3.6 to bound the integral.

Combining these two terms yields

$$\begin{aligned}G(t) \leq C_d\left(t + \sqrt{2tV\sqrt{n}}\right)\left(\log(1 + 2eV\sqrt{n}/t)\right)^{\frac{2d-1}{4}} \\ + C_d t\sqrt{\log(4 + 2V\sqrt{n}/t)}.\end{aligned} \tag{A.25}$$

As always, the constants C_d that appear below vary from line to line. We have

$$C_d t\left(\log(1 + 2eV\sqrt{n}/t)\right)^{\frac{2d-1}{4}} \leq \frac{t^2}{6}$$

whenever $t \geq C_d \max\left\{1, \left(\log(1 + 2eV\sqrt{n})\right)^{\frac{2d-1}{4}}\right\}$. We have

$$C_d\sqrt{2tV\sqrt{n}}\left(\log(1 + 2eV\sqrt{n}/t)\right)^{\frac{2d-1}{4}} \leq \frac{t^2}{6}$$

whenever $t \geq c_d \max\left\{1, (V\sqrt{n})^{1/3}(\log(1 + 2eV\sqrt{n}))^{\frac{2d-1}{6}}\right\}$. Finally, we have

$$2C_d t\sqrt{\log(4 + 2V\sqrt{n}/t)} \leq \frac{t^2}{8}$$

whenever $t \geq C_d \max\Big\{1, \sqrt{\log(4+2V\sqrt{n})}\Big\}$. So, with

$$t = C_d \max\Big\{ (V\sqrt{n})^{1/3} \big(\log\big(1+2eV\sqrt{n}\big)\big)^{\frac{2d-1}{6}}, \sqrt{\log(4+2V\sqrt{n})}, \big(\log\big(1+2eV\sqrt{n}\big)\big)^{\frac{2d-1}{4}}, 1 \Big\}$$

the above three inequalities hold, and we obtain $G(t) \leq t^2/2$, and we may then use Theorem A.3.1 to obtain

$$\mathcal{R}(\widehat{\boldsymbol{\theta}}_{\text{LASSO}}, f^*) \leq \frac{t^2}{n} \leq C_d \max\Bigg\{ \bigg(\frac{V}{n}\bigg)^{\frac{2}{3}} \big(\log(1+2eV\sqrt{n})\big)^{\frac{2d-1}{3}}, \frac{1}{n}\log(4+2V\sqrt{n}), \\ \frac{1}{n}\big(\log(1+2eV\sqrt{n})\big)^{\frac{2d-1}{2}}, \frac{1}{n} \Bigg\}.$$

We claim we can remove the log terms in the second and third terms as well. Note that $\log(4+x) \leq x^{2/3}$ for $x \geq 3$. Thus, we may bound the second term by

$$\frac{1}{n}\log(4+2V\sqrt{n}) \leq \bigg(\frac{2V}{n}\bigg)^{\frac{2}{3}} \mathbb{I}\{2V\sqrt{n} \geq 3\} + \frac{\log(7)}{n}\mathbb{I}\{2V\sqrt{n} < 3\}$$

Similarly, $\log(1+x)^{\frac{2d-1}{2}} \leq x^{2/3}$ for $x \geq C_d$, so we may bound the third term by

$$\begin{aligned} &\frac{1}{n}\big(\log(1+2eV\sqrt{n})\big)^{\frac{2d-1}{2}} \\ &\leq \bigg(\frac{2eV}{n}\bigg)^{\frac{2}{3}} \mathbb{I}\{2eV\sqrt{n} \geq C_d\} + \frac{(\log(1+C_d))^{\frac{2d-1}{2}}}{n}\mathbb{I}\{2eV\sqrt{n} < C_d\}. \end{aligned}$$

This allows us to rewrite our risk bound as

$$\mathcal{R}(\widehat{\boldsymbol{\theta}}_{\text{LASSO}}, f^*) \leq C_d \bigg(\frac{V}{n}\bigg)^{\frac{2}{3}} \big(\log(1+2eV\sqrt{n})\big)^{\frac{2d-1}{3}} + C_d \frac{1}{n}$$

which is the desired bound in the case $\sigma^2 = 1$. The general result can be obtained by rescaling as discussed earlier.

A.3.5 Proof of Theorem 3.4.4

Let

$$\widetilde{\boldsymbol{\theta}} \coloneqq (\widetilde{f}_{\text{EM},V}(\mathbf{x}_1), \ldots, \widetilde{f}_{\text{EM},V}(\mathbf{x}_n)) \text{ and } \boldsymbol{\theta}^* \coloneqq (f^*(\mathbf{x}_1), \ldots, f^*(\mathbf{x}_n))$$

As discussed in Section 3.3, $\mathcal{D}_{n_1,\ldots,n_d} \cap (\theta_n - \theta_1) = \mathcal{D}_{n_1,\ldots,n_d} \cap \mathcal{C}(V)$ (since if $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta} \in \mathcal{D}_{n_1,\ldots,n_d}$ then $\theta_n - \theta_1 = \sum_{j\geq 2} \beta_j = \sum_{j\geq 2} |\beta_j|$), and we have

$$\widetilde{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d} \cap \mathcal{C}(V)} \|\mathbf{y} - \boldsymbol{\theta}\|^2$$

As in Section A.3.4, we may without loss of generality assume $\sigma^2 = 1$, and then rescale to handle the general case.

We again appeal to Theorem A.3.1. We need to bound

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{D}_{n_1,\dots,n_d} \cap \mathcal{C}(V) : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \le t} \langle \xi, \boldsymbol{\theta} - \boldsymbol{\theta}^* \rangle$$

for $t > 0$ where $\xi \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$. But by removing the $\mathcal{D}_{n_1,\dots,n_d}$ constraint in the supremum, we immediately see that this quantity is bounded from above by $G(t)$ as defined above (A.24). Thus we may exactly follow the argument that bounds $G(t)$ in Section A.3.4, and ultimately end up with the same bound (3.48) in Theorem 3.4.6.

A.3.6 Proof of Theorem 3.4.9

See the end of Section A.3.7 for the proof of the tighter bound in the case $d = 2$.

We use Assouad's lemma [4] (see also [96] for more discussion) in the following form:

Lemma A.3.8 (Assouad's lemma [96, Lemma 2])**.** *Let q be a positive integer, and assume that for every $\boldsymbol{\eta} \in \{-1, 1\}^q$ there is an associated function $f_{\boldsymbol{\eta}}$ satisfying $V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}) \le V$. Then*

$$\mathfrak{M}_{\sigma,V,d}(n) \ge \frac{q}{2} \min_{\boldsymbol{\eta} \ne \boldsymbol{\eta}'} \frac{\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'})}{d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}')} \min_{d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}') = 1} \Big(1 - \|\mathbb{P}_{f_{\boldsymbol{\eta}}} - \mathbb{P}_{f_{\boldsymbol{\eta}'}}\|_{\mathrm{TV}}\Big),$$

where $\mathcal{L}(f, g) \coloneqq \frac{1}{n} \sum_{i=1}^n (f(\mathbf{x}_i) - g(\mathbf{x}_i))^2$, where $\mathbb{P}_f$ denotes the probability measure of $y_1, \dots, y_n$ drawn from the model (3.1) where $f^* = f$, and where $d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}') \coloneqq \sum_{j=1}^q \mathbb{I}\{\eta_j \ne \eta'_j\}$ denotes the Hamming distance.

Below we construct a collection of functions $\{f_{\boldsymbol{\eta}}, \boldsymbol{\eta} \in \{-1, 1\}^q\}$ such that the right-hand side of Assouad's bound above is the resulting bound $C_d(\sigma^2 V/n)^{2/3}(\log(n(V/\sigma)^2))^{2(d-1)/3}$ of Theorem 3.4.9, but under the assumption that $n_1 = \cdots = n_d$ and that n_1 is a power of 2.

Our construction of the functions $\{f_{\boldsymbol{\eta}}, \boldsymbol{\eta} \in \{-1, 1\}^q\}$ closely roughly mirrors that of Blei et al. [12, Section 4]. First let

$$\ell \coloneqq \left\lceil \frac{1}{3 \log 2} \big(\log(C_d n V^2/\sigma^2) - (d-1) \log\log(C_d n V^2/\sigma^2) \big) \right\rceil. \tag{A.26}$$

The particular choice of this integer ℓ will be relevant later. We define the index set

$$\mathcal{M}_\ell \coloneqq \left\{ (m_1, \dots, m_d) \in \mathbb{N}^d : \sum_{j=1}^d m_j = \ell,\ \max_{j \in [d]} m_j \le 2\ell/d \right\},$$

and for each $\mathbf{m} \in \mathcal{M}_\ell$ we define

$$\mathcal{I}_{\mathbf{m}} \coloneqq \big\{ (i_1, \dots, i_d) \in \mathbb{N}^d : i_j \in [2^{m_j}] \text{ for each } j \in [d] \big\}.$$

One can check that $|\mathcal{I}_{\mathbf{m}}| = \prod_{j=1}^d 2^{m_j} = 2^\ell$ for each $\mathbf{m} \in \mathcal{M}_\ell$. We also have the following lower bound which is proved in Section A.5.4.

Lemma A.3.9. *There exist positive constants a_d and $c'_{d,\sigma^2/V^2}$ such that*

$$|\mathcal{M}_\ell| \geq a_d \ell^{d-1} \qquad \text{for all } n \geq c'_{d,\sigma^2/V^2}.$$

Finally, let

$$q \coloneqq |\mathcal{M}_\ell| \cdot 2^\ell \tag{A.27}$$

be the cardinality of the set $\{(\mathbf{m}, \mathbf{i}) : \mathbf{m} \in \mathcal{M}_\ell, \mathbf{i} \in \mathcal{I}_\mathbf{m}\}$. We index the components of $\boldsymbol{\eta} \in \{-1, 1\}^q$ by $\eta_{\mathbf{m},\mathbf{i}}$ for $\mathbf{m} \in \mathcal{M}_\ell, \mathbf{i} \in \mathcal{I}_\mathbf{m}$.

We now define a function $f_{\boldsymbol{\eta}}$ for each $\boldsymbol{\eta} \in \{-1, 1\}^q$. For natural numbers m and natural number $i \in [2^{m_j}]$ we define the function $\phi_{m,i} : [0, 1] \to \mathbb{R}$ by

$$\phi_{m,i}(x) \coloneqq \begin{cases} 0 & x \notin [(i-1)2^{-m}, i2^{-m}], \\ 2^{-m-2} & x = (i - \frac{3}{4})2^{-m}, \\ -2^{-m-2} & x = (i - \frac{1}{4})2^{-m}, \\ \text{linear} & \text{otherwise.} \end{cases} \tag{A.28}$$

Note that consequently

$$\phi'_{m,i}(x) = \begin{cases} 1 & x \in ((i-1)2^{-m}, (i - \frac{3}{4})2^{-m}) \cup ((i - \frac{1}{4})2^{-m}, i2^{-m}), \\ -1 & x \in ((i - \frac{3}{4})2^{-m}, (i - \frac{1}{4})2^{-m}). \end{cases}$$

We define the function $f_{\boldsymbol{\eta}} : [0, 1]^d \to \mathbb{R}$ as

$$f_{\boldsymbol{\eta}} \coloneqq \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \sum_{\mathbf{m} \in \mathcal{M}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_\mathbf{m}} \eta_{\mathbf{m},\mathbf{i}} \bigotimes_{j=1}^{d} \phi_{m_j, i_j},$$

that is,

$$f_{\boldsymbol{\eta}}(\mathbf{x}) \coloneqq \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \sum_{\mathbf{m} \in \mathcal{M}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_\mathbf{m}} \eta_{\mathbf{m},\mathbf{i}} \prod_{j=1}^{d} \phi_{m_j, i_j}(x_j).$$

The following lemma (proved in Section A.5.5) contains the key ingredients for the application of Lemma A.3.8.

Lemma A.3.10. *For the functions $f_{\boldsymbol{\eta}}$ defined above, the following three inequalities hold.*

$$V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}; [0, 1]^d) \leq V, \tag{A.29}$$

$$\max_{d_\mathrm{H}(\boldsymbol{\eta}, \boldsymbol{\eta}') = 1} \|\mathbb{P}_{f_{\boldsymbol{\eta}}} - \mathbb{P}_{f_{\boldsymbol{\eta}'}}\|_\mathrm{TV} \leq \sqrt{\frac{n}{\sigma^2} \frac{V^2}{|\mathcal{M}_\ell|} 2^{-3\ell-4d}}, \tag{A.30}$$

and

$$\min_{\boldsymbol{\eta} \neq \boldsymbol{\eta}'} \frac{\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'})}{d_\mathrm{H}(\boldsymbol{\eta}, \boldsymbol{\eta}')} \geq \frac{4V^2}{|\mathcal{M}_\ell|} 2^{-3\ell-6d}. \tag{A.31}$$

The three inequalities in the above lemma, together with Lemma A.3.8, Lemma A.3.9 and equation (A.27) imply

$$\begin{aligned}\mathfrak{M}_{\sigma,V,d}(n) &\geq \frac{q}{2}\cdot\frac{4V^2}{|\mathcal{M}_\ell|}2^{-3\ell-6d}\left[1-\sqrt{\frac{n}{\sigma^2}\frac{V^2}{|\mathcal{M}_\ell|}2^{-3\ell-4d}}\right]\\ &\geq 2^{\ell+1}V^2 2^{-3\ell-6d}\left[1-\sqrt{\frac{n}{\sigma^2}\frac{V^2}{a_d\ell^{d-1}}2^{-3\ell-4d}}\right]\\ &\geq V^2 2^{-2\ell-6d+1}\left[1-\sqrt{C_d\frac{n}{\sigma^2}\frac{V^2}{\ell^{d-1}}2^{-3\ell}}\right] \qquad \text{(A.32)}\end{aligned}$$

where $C_d \coloneqq 2^{-4d}/a_d$.

Note that our choice (A.26) of ℓ implies

$$2^{-\ell} = \left(\frac{\sigma^2}{C_d n V^2}\right)^{\frac{1}{3}}\left(\log(C_d n V^2/\sigma^2)\right)^{\frac{d-1}{3}}.$$

Then

$$\begin{aligned}C_d\frac{n}{\sigma^2}\frac{V^2}{\ell^{d-1}}2^{-3\ell} &= \left(\frac{\log(C_d n V^2/\sigma^2)\cdot\log 2}{\frac{1}{3}\log(C_d n V^2/\sigma^2)-\frac{d-1}{3}\log\log(C_d n V^2/\sigma^2)}\right)^{d-1}\\ &= \left(\frac{3}{\log 2}\left(1-(d-1)\frac{\log\log(C_d n V^2/\sigma^2)}{\log(C_d n V^2/\sigma^2)}\right)\right)^{-(d-1)}. \qquad \text{(A.33)}\end{aligned}$$

For all $x>1$ we have $\frac{\log\log x}{\log x}\leq(\log x)^{-1/2}$. Thus if we have

$$nV^2/\sigma^2 \geq e^{d^2/4}/C_d \qquad \text{(A.34)}$$

then we obtain

$$\frac{\log\log(C_d n V^2/\sigma^2)}{\log(C_d n V^2/\sigma^2)} \leq (\log(C_d n V^2/\sigma^2))^{-1/2} \leq \frac{2}{d}.$$

Applying this bound to the earlier equality (A.33) yields

$$C_d\frac{n}{\sigma^2}\frac{V^2}{\ell^{d-1}}2^{-3\ell} \leq \left(\frac{2\log 2}{3}\right)^{d-1} \leq \frac{1}{2}.$$

Thus continuing from the earlier lower bound (A.32), we obtain

$$\begin{aligned}\mathfrak{M}_{\sigma,V,d}(n) &\geq \widetilde{c}_d V^2\left(\frac{\sigma^2}{C_d n V^2}\right)^{\frac{2}{3}}\left(\log(C_d n V^2/\sigma^2)\right)^{\frac{2(d-1)}{3}}\\ &= c'_d\left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}}\left(\log(C_d n V^2/\sigma^2)\right)^{\frac{2(d-1)}{3}},\end{aligned}$$

where $\widetilde{c}_d \coloneqq 2^{-6d+1}(1-2^{-1/2})$ and $c'_d \coloneqq C_d^{-2/3}\widetilde{c}_d$, provided the sample size condition (A.34) holds.

We claim we may replace $\log(C_d n V^2/\sigma^2)$ with $\log(n(V/\sigma)^2)$ in the above lower bound for sufficiently large n. Indeed as long as $n(V/\sigma)^2 \geq C_d^{-2}$ we have $\log(C_d n V^2/\sigma^2) \geq \frac{1}{2}\log(n(V/\sigma)^2)$, so we obtain

$$\mathfrak{M}_{\sigma,V,d}(n) \geq c''_d \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} (\log(n(V/\sigma)^2))^{\frac{2(d-1)}{3}}$$

for all n larger than a constant depending only on d and σ^2/V^2.

Relaxing assumptions We have proved the theorem under the assumption $n_1 = \cdots = n_d$ with n_1 a power of 2. We now argue that this suffices to handle the general case. First, suppose $n_1 = \cdots = n_d$, but n_1 is not a power of 2. Let n'_1 be the largest power of 2 less than n_1, and let $n' = (n'_1)^d$. Then we may apply the argument on the $n'_1 \times \cdots \times n'_1$ and obtain a collection $\{f_{\boldsymbol{\eta}}, \boldsymbol{\eta} \in \{-1,1\}^q\}$ such that the right-hand side of Assouad's bound is $C_d(\sigma^2 V/n')^{2/3}(\log(n'(V/\sigma)^2))^{2(d-1)/3}$. We now adapt this collection for our original $n_1 \times \cdots \times n_d$ grid. Since $\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'})$ and $\|\mathbb{P}_{f_{\boldsymbol{\eta}}} - \mathbb{P}_{f_{\boldsymbol{\eta}'}}\|_{\mathrm{TV}}$ depend only the values of the functions at the design points $\mathbf{x}_i$, we may assume without loss of generality that the functions are piecewise constant with respect to the $n'_1 \times \cdots \times n'_1$ grid, since keeping the values of $f_{\boldsymbol{\eta}}(\mathbf{x}_i)$ intact for all $\boldsymbol{\eta}$ and $\mathbf{x}_i$ while making the function piecewise constant elsewhere can only decrease the HK-variation, and thus not violate the $V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}) \leq V$ condition. Note that $n_1 - n'_1 < n'_1$. To move from the $n'_1 \times \cdots \times n'_1$ grid $\bigtimes_{j=1}^d \{0, \frac{1}{n'_1}, \ldots, \frac{n'_1-1}{n'_1}\}$ to a $n_1 \times \cdots \times n_d$ grid, we simply include the $n_1 - n'_1$ extra points $\frac{1}{2n'_1}, \frac{3}{2n'_1}, \ldots, \frac{2(n_1-n'_1)-1}{2n'_1}$ to the set $\{0, \frac{1}{n'_1}, \ldots, \frac{n'_1-1}{n'_1}\}$ before taking the Cartesian product d times. This is not an evenly spaced grid, but we may consider an isotonic function g that maps these n_1 points

$$0, \frac{1}{2n'_1}, \frac{2}{2n'_1}, \ldots, \frac{2(n_1-n'_1)-1}{2n'_1}, \frac{n_1-n'_1}{n'_1}, \frac{n_1-n'_1+1}{n'_1}, \ldots \frac{n'_1-1}{n'_1}$$

to the evenly spaced grid $0, \frac{1}{n_1}, \ldots, \frac{n_1-1}{n_1}$, and let $\widetilde{f}_{\boldsymbol{\eta}} = f_{\boldsymbol{\eta}} \circ G$ where $G = \bigotimes_{i=1}^d g$.

We now account for how the right-hand side of Assouad's bound (Lemma A.3.8) changes when using $\{\widetilde{f}_{\boldsymbol{\eta}}\}$ on the full $n_1 \times \cdots \times n_d$ grid instead of $\{f_{\boldsymbol{\eta}}\}$ on the smaller grid. Since HK variation is invariant under "stretching" of the domain, $V_{\mathrm{HK}\mathbf{0}}(\widetilde{f}_{\boldsymbol{\eta}}) = V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}) \leq V$. Furthermore, since the $f_{\boldsymbol{\eta}}$ are piecewise constant, the addition of the extra points simply means that certain values of $f_{\boldsymbol{\eta}}$ on the smaller grid appear up to 2^d times as values of $\widetilde{f}_{\boldsymbol{\eta}}$ on the larger grid (since $n_j < 2n'_1$ for each j, and $n < 2^d n'$). Thus, using the fact that $n' < n < 2^d n'$, the loss $\widetilde{\mathcal{L}}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'})$ with respect to the larger grid satisfies

$$2^{-d}\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'}) \leq \widetilde{\mathcal{L}}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'}) \leq \mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'})$$

where $\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'})$ is with respect to the smaller grid. In particular, we still have the bound in (A.30) for $\|\mathbb{P}_{\widetilde{f}_{\boldsymbol{\eta}}} - \mathbb{P}_{\widetilde{f}_{\boldsymbol{\eta}'}}\|_{\mathrm{TV}}$, since in the proof of (A.3.10) we show $\|\mathbb{P}_{\widetilde{f}_{\boldsymbol{\eta}}} - \mathbb{P}_{\widetilde{f}_{\boldsymbol{\eta}'}}\|_{\mathrm{TV}} \leq$

$\sqrt{\frac{n}{4\sigma^2}\mathcal{L}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'})}$. For (A.31), we need to multiply the right-hand side by a factor of 2^{-d}, which amounts to changing a few constants that depend on d. Thus, up to this d-dependent factor, the result of Lemma A.3.10 hold, and we can apply Assouad's bound as before, with the only changes being an adjustment in the constants that depend on d. Thus, we obtain a final lower bound of the form

$$\mathfrak{M}_{\sigma,V,d}(n) \geq C_d \left(\frac{\sigma^2 V}{n'}\right)^{\frac{2}{3}} (\log(n'(V/\sigma)^2))^{\frac{2(d-1)}{3}}.$$

To conclude, note that $2^{-d}n \leq n' \leq n$, so we have

$$\mathfrak{M}_{\sigma,V,d}(n) \geq C'_d \left(\frac{\sigma^2 V}{n}\right)^{\frac{2}{3}} (\log(n(V/\sigma)^2))^{\frac{2(d-1)}{3}}.$$

for n larger than a [now slightly larger] constant depending only on $(\sigma/V)^2$ and d.

We have now proven the theorem under the assumption $n_1 = \cdots = n_d$ where n_1 is any sufficiently large positive integer. The argument for relaxing this assumption to $n_j \geq cn^{1/d}$ is similar. We can consider a smaller square grid $n'_1 \times \cdots \times n'_1$ where $n'_1 = c_s n^{1/d}$, and use the above argument to obtain a collection of $f_{\boldsymbol{\eta}}$ (which may be assumed to be rectangular piecewise constant on the small grid) for which Assouad's bound yields $C_d(\sigma^2 V/n')^{2/3}(\log(n'(V/\sigma)^2))^{2(d-1)/3}$ where $n' = c_s^d n$. To move to the larger grid, we need to add $n_j - c_s n'_1$ points to each dimension of the grid in the same fashion as above, by distributing them evenly among the gaps between the points of the smaller grid. We can again make this larger grid evenly spaced by stretching the domain as before to obtain a new collection of functions $\widetilde{f}_{\boldsymbol{\eta}}$. Since we have enlarged the grid by a factor of c_s^{-d}, each value of $f_{\boldsymbol{\eta}}$ on the small grid appears at most c_s^{-d} times as values of $\widetilde{f}_{\boldsymbol{\eta}}$ on the larger grid. Thus,

$$c_s^d \mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'}) \leq \widetilde{\mathcal{L}}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'}) \leq \mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'})$$

We may then use the bounds in Lemma A.3.10 (with the bound (A.31) having an extra factor of c_s^d that will later be absorbed into constants) and apply Assouad's bound to obtain the same bound $C_d(\sigma^2 V/n')^{2/3}(\log(n'(V/\sigma)^2))^{2(d-1)/3}$. Substituting $n' = c_s^d n$ and absorbing c_s^d into the constant and taking n larger than a constant depending only on c_s, $(\sigma/V)^2$, and d yields the desired bound.

A.3.7 Proof of Theorem 3.4.3

Let us first consider the case $\sigma^2 = 1$. Let $\mathcal{F}^d_{\mathrm{EM}}(V) \coloneqq \{f \in \mathcal{F}^d_{\mathrm{EM}} : V_{\mathrm{HK}\mathbf{0}}(f) \leq V\}$.

Let $\mathcal{F}^d_{\mathrm{DF}}$ denote the class of cumulative distribution functions of probability distributions on $[0,1]^d$. We immediately have $V\mathcal{F}^d_{\mathrm{DF}} \subseteq \mathcal{F}^d_{\mathrm{EM}}(V)$, which implies

$$\inf_{\widehat{f}_n} \sup_{f^* \in \mathcal{F}^d_{\mathrm{EM}}(V)} \mathbb{E}_{f^*} \mathcal{L}(\widehat{f}_n, f^*) \geq \inf_{\widehat{f}_n} \sup_{f^* \in V\mathcal{F}^d_{\mathrm{DF}}} \mathbb{E}_{f^*} \mathcal{L}(\widehat{f}_n, f^*).$$

Thus it suffices to prove a minimax lower bound for $V\mathcal{F}^d_{\mathrm{DF}}$. To do so, we employ the Yang and Barron bound [94], roughly in the form appearing in [43, Thm. IV.1] (after specializing the Kullback-Leibler divergence to our Gaussian model):

$$\inf_{\widehat{f}_n} \sup_{f^* \in \mathcal{F}^d_{\mathrm{DF}}} \mathbb{E}_{f^*} \mathcal{L}(\widehat{f}_n, f^*) \geq \frac{\eta^2}{4}\left(1 - \frac{\log 2 + \log N(\epsilon/V; \mathcal{F}^d_{\mathrm{DF}}) + n\epsilon^2}{\log M(\eta/V; \mathcal{F}^d_{\mathrm{DF}})}\right) \tag{A.35}$$

for any positive η and ϵ. Here, $N(\epsilon; \mathcal{F}^d_{\mathrm{DF}})$ is the covering number of $\mathcal{F}^d_{\mathrm{DF}}$ (cardinality N of the smallest set $g^1, \ldots, g^N$ satisfying $\min_j \mathcal{L}(f^j, g) \leq \epsilon^2$ for any $g \in \mathcal{F}^d_{\mathrm{DF}}$) and $M(\eta; \mathcal{F}^d_{\mathrm{DF}})$ is the packing number of $\mathcal{F}^d_{\mathrm{DF}}$ (cardinality M of the largest set $g^1, \ldots, g^M$ satisfying $\mathcal{L}(f^j, f^k) > \eta^2$ for all $j \neq k$).

The case $d \geq 2$. We first prove the general minimax bound for cases $d \geq 2$, before specializing to the case $d = 2$. We claim

$$\log N(\epsilon'; \mathcal{F}^d_{\mathrm{DF}}) \leq C_d \frac{1}{\epsilon'}\left(\log \frac{1}{\epsilon'}\right)^{d-\frac{1}{2}}, \qquad \epsilon' < e^{-1} \tag{A.36a}$$

$$\log M(\eta'; \mathcal{F}^d_{\mathrm{DF}}) \geq C_d \frac{1}{\eta'}\left(\log \frac{1}{\eta'}\right)^{d-1}. \tag{A.36b}$$

Assuming these two equations are true, then applying the Yang-Barron bound (A.35) with $\epsilon = a_d(V/n)^{\frac{1}{3}}(\log(nV^2))^{\frac{2d-1}{6}}$ and $\eta = b_d(V/n)^{\frac{1}{3}}(\log(nV^2))^{\frac{d-2}{3}}$, for certain constants a_d and b_d, allows us to conclude the proof. Specifically, we then have $n\epsilon^2 = a_d^2(nV^2)^{\frac{1}{3}}(\log(nV^2))^{\frac{2d-1}{3}}$ as well as

$$\begin{aligned}
&\log N(\epsilon/V; \mathcal{F}^d_{\mathrm{DF}}) \\
&= C_d a_d (nV^2)^{\frac{1}{3}}(\log(nV^2))^{-\frac{2d-1}{6}}\left[\frac{1}{3}\log(nV^2/a_d^3) - \frac{2d-1}{6}\log\log(nV^2)\right]^{d-\frac{1}{2}} \\
&\lesssim (nV^2)^{\frac{1}{3}}(\log(nV^2))^{\frac{2d-1}{3}}
\end{aligned}$$

and

$$\begin{aligned}
&\log M(\eta/V; \mathcal{F}^d_{\mathrm{DF}}) \\
&= C_d b_d (nV^2)^{\frac{1}{3}}(\log(nV^2))^{-\frac{d-2}{3}}\left[\frac{1}{3}\log(nV^2/b_d) - \frac{d-2}{3}\log\log(nV^2)\right]^{d-1} \\
&\gtrsim (nV^2)^{\frac{1}{3}}(\log(nV^2))^{\frac{2d-1}{3}}.
\end{aligned}$$

In particular, the quantities $n\epsilon^2$, $\log N(\epsilon; \mathcal{F}^d_{\mathrm{DF}})$, and $\log M(\eta; \mathcal{F}^d_{\mathrm{DF}})$ are of the same order, so a judicious choice of constants a_d and b_d will make the Yang-Barron bound (A.35) be on the order of

$$\eta^2 \asymp \left(\frac{V}{n}\right)^{\frac{2}{3}}(\log(nV^2))^{\frac{2(d-2)}{3}}, \tag{A.37}$$

which yields the desired minimax bound in the case $\sigma^2 = 1$. Note that n must be sufficiently large (larger than a constant depending on d and V) in order for $\epsilon/V < e^{-1}$ in order to use the covering number bound (A.36a). For general σ^2 and V, we may rescale the problem to have noise level $(\sigma')^2 = 1$ and variation $V' = V/\sigma$, apply the above bound (A.37), and multiply by σ^2 to obtain the final minimax bound that appears in Theorem 3.4.3.

It now remains to verify the above two claims. The first claim (A.36a) is due to Blei et al. [12]; see (A.67) with $R = 1$ and note that our notion of distance in the present proof is normalized by n.

We now turn to the other claim (A.36b). Let ℓ, $\mathcal{M}_\ell$, $q \coloneqq |\mathcal{M}_\ell| 2^\ell$, and $\{f_{\boldsymbol{\eta}} : \boldsymbol{\eta} \in \{-1,1\}^d\}$ be as defined in Section A.3.6 (see (A.26), (A.27), etc.), and let with $V = 1$. Note that the $f_{\boldsymbol{\eta}}$ are continuous functions with $V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}) \le 1$, so they belong to $\mathcal{F}^d_{\mathrm{DF}} - \mathcal{F}^d_{\mathrm{DF}}$.

The Gilbert-Varshamov lemma (see [61, Lemma 4.7]) guarantees a subset $T \subseteq \{-1,1\}^q$ satisfying $\log|T| \gtrsim q$ and $d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}') \gtrsim q/2$ for all distinct $\boldsymbol{\eta}, \boldsymbol{\eta}' \in T$. Recalling from Lemma A.3.10 that

$$\min_{\boldsymbol{\eta} \neq \boldsymbol{\eta}'} \frac{\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'})}{d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}')} \ge \frac{2^{-3\ell-6d+2}}{|\mathcal{M}_\ell|}, \tag{A.38}$$

we obtain a packing set $\{f_{\boldsymbol{\eta}} : \boldsymbol{\eta} \in T\}$ of $\mathcal{F}^d_{\mathrm{DF}} - \mathcal{F}^d_{\mathrm{DF}}$ satisfying $\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'}) \ge \frac{q \cdot 2^{-3\ell-6d+1}}{|\mathcal{M}_\ell|} = 2^{-2\ell-6d+1} =: (\eta')^2$. Note that $\ell = c \log \frac{1}{\eta'}$. Recalling $|\mathcal{M}_\ell| \gtrsim \ell^{d-1}$ from Lemma A.3.9, the log cardinality of this packing set $\{f_{\boldsymbol{\eta}} : \boldsymbol{\eta} \in T\}$ with radius η' is

$$\log M(\eta'; \mathcal{F}^d_{\mathrm{DF}} - \mathcal{F}^d_{\mathrm{DF}}) = |\mathcal{M}_\ell| 2^\ell \gtrsim 2^\ell \ell^{d-1} \asymp \frac{1}{\eta'} \left(\log \frac{1}{\eta'} \right)^{d-1}.$$

Using basic relationships between covering numbers and packing numbers, we have

$$\begin{aligned}
\frac{1}{\eta'} \left(\log \frac{1}{\eta'} \right)^{d-1} &\lesssim \log M(\eta'; \mathcal{F}^d_{\mathrm{DF}} - \mathcal{F}^d_{\mathrm{DF}}) \\
&\le \log N(\eta'/2; \mathcal{F}^d_{\mathrm{DF}} - \mathcal{F}^d_{\mathrm{DF}}) \\
&\overset{(*)}{\le} 2 \log N(\eta'/4; \mathcal{F}^d_{\mathrm{DF}}) \\
&\le 2 \log M(\eta'/4; \mathcal{F}^d_{\mathrm{DF}}),
\end{aligned}$$

where the starred inequality is due to the fact that one can obtain a covering set for $\mathcal{F}^d_{\mathrm{DF}} - \mathcal{F}^d_{\mathrm{DF}}$ by taking a covering set for $\mathcal{F}^d_{\mathrm{DF}}$ with half the radius, and taking the differences between all pairs drawn from the covering set.

The only place we used the assumption that $n_1 = \cdots = n_d$ with n_1 a power of 2 is in our appeal to the construction of $\{f_{\boldsymbol{\eta}}\}$ in proving the packing bound (A.36b). We may follow the same argument as in the end of Section A.3.6 to relax these assumptions to the setting of the theorem and obtain the same risk lower bound, since the argument there only results in changing the right-hand side of the lower bound (A.38) by a factor that depends on d and c_s.

The case $d = 2$. We now prove the tighter bound in the case $d = 2$, which will follow from tightening the packing number bound (A.36b).

We again refer to notation in Section A.3.6. Let

$$\widetilde{\mathcal{M}}_\ell \coloneqq \{(m_1, m_2) \in \mathbb{N}^d : m_1 + m_2 = \ell,\ m_1 \text{ and } m_2 \text{ both even}\}$$

and let $\widetilde{q} \coloneqq |\widetilde{\mathcal{M}}_\ell| 2^\ell$. Let $\phi_{m,i}$ be as before (A.28). For $\boldsymbol{\eta} \in \{-1, 1\}^q$, we define

$$F_{\boldsymbol{\eta},\mathbf{m}}(t_1, t_2) \coloneqq \sum_{\mathbf{i} \in \mathcal{I}_\mathbf{m}} \eta_{\mathbf{m},\mathbf{i}} \phi'_{m_1,i_1}(t_1) \phi'_{m_2,i_2}(t_2)$$

and

$$\widetilde{f}_{\boldsymbol{\eta}}(\mathbf{x}) \coloneqq \int_0^{x_1} \int_0^{x_2} \prod_{\mathbf{m} \in \widetilde{\mathcal{M}}_\ell} (1 + F_{\boldsymbol{\eta},\mathbf{m}}(t_1, t_2))\, dt_1\, dt_2$$

Note that we can rewrite this function as

$$\widetilde{f}_{\boldsymbol{\eta}}(\mathbf{x}) = x_1 x_2 + \sum_{\mathbf{m} \in \widetilde{\mathcal{M}}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_\mathbf{m}} \eta_{\mathbf{m},\mathbf{i}} \phi_{m_1,i_1}(x_1) \phi_{m_2,i_2}(x_2) + Q_{\boldsymbol{\eta}}(\mathbf{x}),$$

where

$$Q_{\boldsymbol{\eta}}(\mathbf{x}) \coloneqq \sum_{P \geq 2} \sum_{k_1, \ldots, k_P} \int_0^{x_1} \int_0^{x_2} \prod_{p=1}^{P} F_{\boldsymbol{\eta},(k_p, \ell - k_p)}(t_1, t_2)\, dt_1\, dt_2,$$

and the inner sum above is over even integers $0 \leq k_1 < k_2 < \cdots < k_P \leq \ell$.

These functions satisfying the following properties (proved in Section A.5.6).

Lemma A.3.11. *The functions $\widetilde{f}_{\boldsymbol{\eta}}$ belong to $\mathcal{F}^2_{\mathrm{DF}}$ and satisfy*

$$\min_{\boldsymbol{\eta} \neq \boldsymbol{\eta}'} \frac{\mathcal{L}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'})}{d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}')} \geq 2^{-3\ell - 10}.$$

From here, we apply the Gilbert-Varshamov lemma again to obtain a subset $T \subseteq \{-1, 1\}^{\widetilde{q}}$ satisfying $\log|T| \gtrsim \widetilde{q}$ and $d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}') \geq \widetilde{q}/2$ for all distinct $\boldsymbol{\eta}, \boldsymbol{\eta}' \in T$. From the above inequality, we can obtain a packing set $\{\widetilde{f}_{\boldsymbol{\eta}} : \boldsymbol{\eta} \in T\}$ of $\mathcal{F}^2_{\mathrm{DF}}$ satisfying $\mathcal{L}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'}) \geq \widetilde{q} \cdot 2^{-3\ell - 11} \gtrsim \ell \cdot 2^{-2\ell} =: (\eta')^2$ where we have used $\widetilde{q} = |\widetilde{\mathcal{M}}_\ell| 2^\ell \gtrsim \ell \cdot 2^\ell$. Note that then we have

$$\frac{1}{\eta'} \left(\log \frac{1}{\eta'} \right)^{3/2} = 2^\ell \ell^{-1/2} (c\ell - \frac{1}{2} \log \ell)^{3/2} \lesssim \ell \cdot 2^\ell \lesssim \widetilde{q} \leq \log M(\eta'; \mathcal{F}^2_{\mathrm{DF}})$$

since $\widetilde{q} \lesssim \log|T|$. Note that this packing number bound is of the same order as the earlier covering number bound (A.36a).

We now return to the Yang-Barron bound (A.35) with $\epsilon = a(V/n)^{\frac{1}{3}} (\log(nV^2))^{\frac{1}{2}}$ and $\eta = b(V/n)^{\frac{1}{3}} (\log(nV^2))^{\frac{1}{2}}$. We have $n\epsilon^2 \asymp (nV^2)^{\frac{1}{3}} \log(nV^2)$ as well as

$$\log N(\epsilon/V; \mathcal{F}^2_{\mathrm{DF}}) \lesssim (nV^2)^{\frac{1}{3}} \log(nV^2) \lesssim \log M(\eta/V; \mathcal{F}^2_{\mathrm{DF}}).$$

Thus with appropriate choices of constants, obtain a lower bound on the minimax risk on the order of

$$\eta^2 \asymp \left(\frac{V}{n}\right)^{\frac{2}{3}} \log(nV^2),$$

in the case $\sigma^2 = 1$. Repeating the rescaling argument produces the bound for general σ^2.

We can relax the assumption that $n_1 = n_2$ with n_1 a power of 2 in the same manner as before, and again, the result of applying the same argument amounts to an additional factor depending only on c_s for the lower bound in Lemma A.3.11.

Having proved the tighter minimax lower bound of Theorem 3.4.3 in the case $d = 2$, we note that the analogous bound of Theorem 3.4.9 follows immediately, since

$$\{f^* \in \mathcal{F}^2_{\mathrm{EM}} : V_{\mathrm{HK}\mathbf{0}}(f^*) \leq V\} \subseteq \{f^* : V_{\mathrm{HK}\mathbf{0}}(f^*) \leq V\}.$$

A.3.8 Proofs of Theorem 3.4.10 and Theorem A.1.1

We shall first introduce some notation and state some auxiliary results which will hold for every $d \geq 1$ and which will used in the proofs of both Theorem 3.4.10 and Theorem A.1.1. After that we shall give the proofs of Theorem 3.4.10 and Theorem A.1.1 separately in two subsections.

Throughout, $\mathbf{A}$ is the design matrix from Section 3.3. As observed in Section 3.3.1, $\mathbf{A}$ is square and invertible (note that we are working under the assumption that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ come from the lattice design (3.34)). This means that every $\boldsymbol{\theta} \in \mathbb{R}^n$ can be expressed as $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta}$ for a unique $\boldsymbol{\beta} \in \mathbb{R}^n$. By an abuse of notation, we define

$$V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta}) := \sum_{j=2}^{n} |\beta_j|$$

where $(\beta_1, \ldots, \beta_n)$ are the components of $\boldsymbol{\beta}$. This abuse of notation is justified by noting that if $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta}$, then $\boldsymbol{\theta} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$ for $f := \sum_{i=1}^{m} \beta_i \mathbb{I}_{[\mathbf{x}_i, 1]}$. For this function f, it is easy to see that $V_{\mathrm{HK}\mathbf{0}}(f) = \sum_{j=2}^{n} |\beta_j|$. In other words, we are defining $V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta})$ to be equal to $V_{\mathrm{HK}\mathbf{0}}(f)$ for a specific canonical function on $[0,1]^d$ which satisfies $f(\mathbf{x}_i) = \boldsymbol{\theta}_i$ for each $i = 1, \ldots, n$.

We shall say that a vector $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta} \in \mathbb{R}^n$ is entirely monotone if $\min_{j \geq 2} \beta_j \geq 0$. This can be justified by noting that the function $f := \sum_{i=1}^{m} \beta_i \mathbb{I}_{[\mathbf{x}_i, 1]}$ belongs to $\mathcal{F}^d_{\mathrm{EM}}$ if and only if $\min_{j \geq 2} \beta_j \geq 0$. We also say that $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta}$ is nearly entirely monotone if

$$\sum_{j=2}^{n} (|\beta_j| - \beta_j) \leq \delta \tag{A.39}$$

for a small $\delta > 0$. Note that, by the definition of $V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta})$, this is equivalent to the inequality: $V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta}) \leq \theta_n - \theta_1 + \delta$. Note that if $\boldsymbol{\theta}$ is entirely monotone, then (A.39) is true with $\delta = 0$ and this justifies the terminology of nearly entirely monotone.

We also use the notation in (A.23). Because $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ is the LSE over the class $\mathcal{C}(V)$, inequality (A.8) with $K = \mathcal{C}(V)$ gives

$$\mathcal{R}(\widehat{f}_{\mathrm{HK}\mathbf{0},V}, f^*) = \mathbb{E}\frac{1}{n}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leq \frac{1}{n}\inf_{\widetilde{\boldsymbol{\theta}}\in K}\Big\{\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 + \sigma^2 w^2(\mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}})) + \sigma^2\Big\}. \tag{A.40}$$

To further bound the right hand side above, it is important to understand the structure of the tangent cone $\mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}})$. The following result (proved in Section A.5.7) provides an explicit characterization of this tangent cone at $\widetilde{\boldsymbol{\theta}} = A\widetilde{\boldsymbol{\beta}}$.

Lemma A.3.12. *Suppose $\widetilde{\boldsymbol{\beta}}$ is such that $\mathbf{A}\widetilde{\boldsymbol{\beta}} \in \mathcal{C}(V)$. Then the tangent cone of $\mathcal{C}(V)$ at $\mathbf{A}\widetilde{\boldsymbol{\beta}}$ is*

$$\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}}) = \left\{\mathbf{A}\boldsymbol{\beta} : \sum_{j\geq 2:\widetilde{\beta}_j = 0} |\beta_j| \leq -\sum_{j\geq 2:\widetilde{\beta}_j \neq 0} \beta_j \operatorname{sign}(\widetilde{\beta}_j)\right\}, \tag{A.41}$$

if $\sum_{j=2}^n |\widetilde{\beta}_j| = V$; otherwise, $\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}}) = \mathbb{R}^n$.

The structure of the tangent cone given above (in the case $\sum_{j=2}^n |\widetilde{\beta}_j| = V$) has the implication that, when $\widetilde{\boldsymbol{\beta}}$ corresponds to a function of the form (3.49), every vector in $\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}})$ can be broken down into lower-dimensional elements each of which is either nearly entirely monotone or has low HK**0** variation. This is the content of the next result. For this result, it will be necessary, as in Section 3.3.1, to view vectors in $\mathbb{R}^n$ as arrays in $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$. Indeed, we shall denote the elements $\boldsymbol{\theta} \in \mathbb{R}^n$ by $\boldsymbol{\theta}_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}$ (where $\mathcal{I}$ is as defined in (3.37). Note that the columns of the design matrix $\mathbf{A}$ can also be indexed in this way so that the $\mathbf{i}^{th}$ column (where $\mathbf{i} = (i_1, \ldots, i_d)$) of $\mathbf{A}$ corresponds to the vector (3.27) with $\mathbf{z} = (i_1/n_1, \ldots, i_d/n_d)$. Note that this implies that the $\mathbf{0}^{th}$ column is the column of ones.

Lemma A.3.13. *Let $\widetilde{\boldsymbol{\beta}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ satisfy $\widetilde{\beta}_{\mathbf{i}} = 0$ for all $\mathbf{i} \notin \{\mathbf{0}, \mathbf{i}^*\}$ for some $\mathbf{i}^*$. Let $\mathbf{i}^u$ and $\mathbf{i}^\ell$ be two indices such that $\mathbf{i}^* \preceq \mathbf{i}^u$ and $\mathbf{i}^\ell \not\succeq \mathbf{i}^*$, and let $L_u := \{\mathbf{i} : \mathbf{i} \preceq \mathbf{i}^u\}$ and $L_\ell := \{\mathbf{i} : \mathbf{i} \preceq \mathbf{i}^\ell\}$. Then for every $\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\beta} \in \mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}})$ where $V = V_{\mathrm{HK}\mathbf{0}}(\mathbf{A}\widetilde{\boldsymbol{\beta}}) = \sum_{\mathbf{i}\neq\mathbf{0}} |\widetilde{\beta}_{\mathbf{i}}|$, we have*

$$\sum_{\mathbf{i}\notin\{\mathbf{0},\mathbf{i}^*\}} (|\beta_{\mathbf{i}}| - \mathfrak{s}(\mathbf{i})\beta_{\mathbf{i}}) \leq -\operatorname{sign}(\widetilde{\beta}_{\mathbf{i}^*})(\alpha_{\mathbf{i}^u} - \alpha_{\mathbf{i}^\ell}), \tag{A.42}$$

where

$$\mathfrak{s}(\mathbf{i}) := \begin{cases} 1 & \mathbf{i} \in L_u \cap L_\ell^c \setminus \{\mathbf{i}^*\} \\ -1 & \mathbf{i} \in L_u^c \cap L_\ell \\ 0 & \mathbf{i} \in (L_u \cap L_\ell) \cup (L_u^c \cap L_\ell^c) \setminus \{\mathbf{0}\} \end{cases} \tag{A.43}$$

Lemma A.3.13 will be used to bound the Gaussian width $w(\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}}))$ for every $\widetilde{\boldsymbol{\beta}}$ as in the statement of Lemma A.3.13 in the following way. Assume first that $\mathbf{i}^u$ and $\mathbf{i}^\ell$ are chosen

so that the right hand side of (A.42) is small. Specifically, for $\widetilde{\boldsymbol{\beta}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and indices $\mathbf{i}^*, \mathbf{i}^u, \mathbf{i}^\ell$ as in the statement of Lemma A.3.13 and a fixed $\delta \geq 0$, let

$$T(\mathbf{i}^u, \mathbf{i}^\ell, \delta) \coloneqq \left\{ \boldsymbol{\alpha} \in \mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}}) : |\alpha_{\mathbf{i}^u} - \alpha_{\mathbf{i}^\ell}| \leq \delta \right\} \cap \mathcal{B}_2(\mathbf{0}, 1), \tag{A.44}$$

where $\mathcal{B}_2(\mathbf{0}, 1) \coloneqq \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| < 1\}$. The intersection with the unit ball here arises because of the presence of the unit norm restriction in the definition of the Gaussian width (see (A.7)). For every $\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\beta} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)$, it is clear that:

$$\sum_{\mathbf{i} \notin \{\mathbf{0}, \mathbf{i}^*\}} (|\beta_{\mathbf{i}}| - \mathfrak{s}(\mathbf{i})\beta_{\mathbf{i}}) \leq |\alpha_{\mathbf{i}^u} - \alpha_{\mathbf{i}^\ell}| \leq \delta. \tag{A.45}$$

Suppose now that δ is small. Then, if we restrict the indices $\mathbf{i}$ to the set $L_u \cap L_\ell^c \setminus \{\mathbf{i}^*\}$, we would have $\mathfrak{s}(\mathbf{i}) = 1$ according to (A.43) and, consequently,

The inequality (A.45) implies that

$$\sum_{\mathbf{i} \notin L_u \cap L_\ell^c \setminus \{\mathbf{i}^*\}} (|\beta_{\mathbf{i}}| - \beta_{\mathbf{i}}) \leq \delta, \tag{A.46}$$

which resembles the definition of nearly entire monotonicity (A.39). This might suggest that the restriction of $\boldsymbol{\alpha}$ to its components indexed by $L_u \cap L_\ell^c \setminus \{\mathbf{i}^*\}$ is nearly entirely monotone, but there are a few issues, one of which is that the definition of nearly entire monotonicity for a sub-array $\boldsymbol{\alpha}_Q$ of $\boldsymbol{\alpha}$ is not quite the same as taking the condition (A.46) and taking the sum only over indices $\mathbf{i}$ in the subset Q (specifically, the $\beta_{\mathbf{i}}$ terms should also be replaced with the analogous quantities for $\boldsymbol{\alpha}_Q$, which are different than the original $\beta_{\mathbf{i}}$ terms derived from the full array $\boldsymbol{\alpha}$). Similarly we also have $\sum_{\mathbf{i} \notin L_u^c \cap L_\ell} (|\beta_{\mathbf{i}}| + \beta_{\mathbf{i}}) \leq \delta$ and $\sum_{\mathbf{i} \notin (L_u \cap L_\ell) \cup (L_u^c \cap L_\ell^c) \setminus \{\mathbf{0}\}} |\beta_{\mathbf{i}}| \leq \delta$, which also might suggest nearly entire monotonicity of $-\boldsymbol{\alpha}$ on $L_u^c \cap L_\ell$ and low HK$\mathbf{0}$ variation on $(L_u \cap L_\ell) \cup (L_u^c \cap L_\ell^c) \setminus \{\mathbf{0}\}$ respectively, but for similar reasons is not immediately true.

Another complication is that the sets $L_u \cap L_\ell^c \setminus \{\mathbf{i}^*\}$ and $(L_u \cap L_\ell) \cup (L_u^c \cap L_\ell^c) \setminus \{\mathbf{0}\}$ are not necessarily rectangular. To deal with these above issues, we show that we can further partition these sets into rectangles such that $\boldsymbol{\alpha}$ restricted to each rectangle is indeed either nearly entirely monotone or has small HK$\mathbf{0}$ variation. This observation would allow us to bound $w(\mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}))$ based on bounds for the Gaussian width of nearly entirely monotone vectors and vectors with small HK$\mathbf{0}$ variation.

The following result gives conditions on a rectangle Q such that the above holds. To state this result, it will be convenient to introduce the following notation. For each $\boldsymbol{\theta} \in \mathbb{R}^n$, let $D\boldsymbol{\theta}$ denote the differenced vector defined as in (3.38). It is easy to check that

$$(D\boldsymbol{\theta})_{\mathbf{0}} \coloneqq \theta_{\mathbf{0}} \quad \text{and} \quad \theta_{\mathbf{i}} \coloneqq \sum_{\mathbf{i}' : \mathbf{i}' \preceq \mathbf{i}} (D\boldsymbol{\theta})_{\mathbf{i}} \text{ for } \mathbf{i} \neq \mathbf{0}. \tag{A.47}$$

As a result, it follows that $D\boldsymbol{\theta} = \mathbf{A}^{-1}\boldsymbol{\theta}$ or, equivalently, $\boldsymbol{\theta} = \mathbf{A}(D\boldsymbol{\theta})$.

Every two indices $\mathbf{q}^\ell$ and $\mathbf{q}^u$ in $\mathcal{I}$ with $\mathbf{q}^\ell \preceq \mathbf{q}^u$ define the following rectangle in $\mathcal{I}$:

$$Q \coloneqq [\mathbf{q}^\ell, \mathbf{q}^u] := \left\{\mathbf{i} \in \mathcal{I} : \mathbf{q}^\ell \preceq \mathbf{i} \preceq \mathbf{q}^u\right\} \tag{A.48}$$

For this rectangle Q and an arbitrary $\boldsymbol{\theta} \in \mathbb{R}^n$, we let $\boldsymbol{\theta}_Q$ be the vector in $\mathbb{R}^{|Q|}$ given by the elements $\theta_{\mathbf{i}}, \mathbf{i} \in Q$. For convenience, we shall index elements of $\boldsymbol{\theta}_Q$ by the entries of Q i.e., for every $q \in Q$, we have $(\boldsymbol{\theta}_Q)_q := \boldsymbol{\theta}_q$. We also define $D\boldsymbol{\theta}_Q := D(\boldsymbol{\theta}_Q)$ to be the differencing operator applied to $\boldsymbol{\theta}_Q$ in a manner analogous to (3.38). Specifically, we take

$$(D\boldsymbol{\theta}_Q)_{\mathbf{i}} = \sum_{\mathbf{z}\in\{0,1\}^d} \mathbb{I}\{\mathbf{i} - \mathbf{z} \succeq \mathbf{q}^\ell\}(-1)^{z_1+\cdots+z_d}\theta_{\mathbf{i}-\mathbf{z}} \qquad \text{for } \mathbf{i} \in Q \tag{A.49}$$

Note that the elements of $D\boldsymbol{\theta}_Q$ are also indexed by the indices in Q. It is important to observe here that $D\boldsymbol{\theta}_Q = D(\boldsymbol{\theta}_Q)$ is different from $(D\boldsymbol{\theta})_Q$. A formula for $D\boldsymbol{\theta}_Q$ in terms of $(D\boldsymbol{\theta})_Q$ is given in Lemma A.5.1.
For the rectangle Q in (A.48) and every $\mathbf{i} = (i_1, \ldots, i_d) \in Q$, we let

$$J(\mathbf{i}) := \left\{1 \le j \le d : i_j > q_j^\ell\right\} \qquad \text{where } \mathbf{q}^\ell := (q_1^\ell, \ldots, q_d^\ell) \tag{A.50}$$

Also for $\mathbf{i} \in Q$ and $\mathbf{i}' \preceq \mathbf{i}$, let

$$t(\mathbf{i}', \mathbf{i}) := \mathbb{I}\left\{\mathbf{i}'_{J(\mathbf{i})} = \mathbf{i}_{J(\mathbf{i})}\right\} \tag{A.51}$$

where we are using the notation $\mathbf{k}_J := (k_j : j \in J)$ for $\mathbf{k} = (k_1, \ldots, k_d) \in \mathcal{I}$ and $J \subseteq \{1, \ldots, d\}$.

Lemma A.3.14. *Consider the same notation and setting as Lemma A.3.13 (in particular, the signs $\mathfrak{s}(\mathbf{i})$ below come from* (A.43)*). Suppose $Q = [\mathbf{q}^\ell, \mathbf{q}^u]$ is a rectangle satisfying the following.*

(a) If $\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}$ and $\mathbf{i} \succeq \mathbf{i}^*$, then $t(\mathbf{i}^*, \mathbf{i}) = 0$ and $t(\mathbf{0}, \mathbf{i}) = 0$.

(b) Given $\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}$, the quantity $\mathfrak{s}(\mathbf{i}')$ is constant over all $\mathbf{i}'$ satisfying $\mathbf{i}' \preceq \mathbf{i}$, $t(\mathbf{i}', \mathbf{i}) \neq 0$, and $\mathfrak{s}(\mathbf{i}') \neq 0$.

(c) Q is a subset of one of $L_u \cap L_\ell$, $L_u^c \cap L_\ell$, $L_u \cap L_\ell^c$, or $L_u^c \cap L_\ell^c$.

Then for any $\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)$,

$$\sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}} (|(D\boldsymbol{\alpha}_Q)_{\mathbf{i}}| - \widetilde{\mathfrak{s}}(\mathbf{i})(D\boldsymbol{\alpha}_Q)_{\mathbf{i}}) \le 2\delta, \tag{A.52}$$

where $\widetilde{\mathfrak{s}}(\mathbf{i}) \coloneqq \mathfrak{s}(\mathbf{i})$ for $\mathbf{i} \succ \mathbf{q}^\ell$, and otherwise for $\mathbf{i} \not\succ \mathbf{q}^\ell$ we have $\widetilde{\mathfrak{s}}(\mathbf{i}) \coloneqq \mathfrak{s}(\mathbf{i}')$ for any $\mathbf{i}'$ satisfying $\mathbf{i}' \preceq \mathbf{i}$, $t(\mathbf{i}', \mathbf{i}) \neq 0$, and $\mathfrak{s}(\mathbf{i}') \neq 0$.

As mentioned earlier, our idea will be to partition $\mathcal{I}$ into a finite number of rectangles Q each satisfying the conditions of Lemma A.3.14. This will enable us to employ bounds for the Gaussian width of nearly entirely monotone vectors and vectors with small HK$\mathbf{0}$ variation to bound $w(\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}}))$. The next result (proved in Section A.5.11) bounds the Gaussian width of nearly entirely monotone vectors.

Lemma A.3.15. *For every $n \geq 1$, $\delta \geq 0$ and $t > 0$, we have*

$$\mathbb{E} \sup_{\substack{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|\leq t,\\ V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta})\leq \theta_n-\theta_1+\delta}} \langle Z, \boldsymbol{\theta}\rangle \leq C_d(t+\delta\sqrt{n})(\log(en))^{\frac{3d}{4}}(\log(e\log(en)))^{\frac{2d-1}{4}}$$

where $Z \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$.

For bounding the Gaussian width of vectors with small HK$\mathbf{0}$ variation, we use the bound derived in (A.25) in the proof of Theorem 3.4.6. This bound gives (here $Z \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$)

$$\mathbb{E} \sup_{\substack{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|\leq 1,\\ V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta})\leq 2V}} \langle Z, \boldsymbol{\theta}\rangle \leq C_d(1+\sqrt{2V\sqrt{n}})\big(\log(1+2eV\sqrt{n})\big)^{\frac{2d-1}{4}} + C_d\sqrt{\log(4+2V\sqrt{n})} \tag{A.53}$$

for every $V \geq 0$.

In addition to the above two Gaussian width bounds, we also need the following result (proved in Section A.5.12) for the proof of Theorem A.1.1. This result is stated for $d = 2$ as Theorem A.1.1 only applies to $d = 2$.

Lemma A.3.16. *Let $d = 2$ and $Z \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. For every $\delta \geq 0$ and $s_1, s_2 \in \{-1, 0, 1\}$, we have*

$$\begin{aligned}\mathbb{E} \sup_{\substack{\boldsymbol{\theta}=\mathbf{A}\boldsymbol{\beta}:\|\boldsymbol{\theta}\|\leq 1\\ V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta})\leq s_1(\theta_{n_1,1}-\theta_{1,1})+s_2(\theta_{1,n_2}-\theta_{1,1})+\delta\\ \beta_{\mathbf{i}}=0,\,\forall \mathbf{i}\succ\mathbf{0}}} \langle Z, \boldsymbol{\theta}\rangle \\ \leq C\Big\{(1+\delta\sqrt{n})\sqrt{\log(en)}\mathbb{I}_{\{s_1\neq 0\}\cup\{s_2\neq 0\}} \\ +\Big[(\delta\sqrt{n})^{\frac{1}{2}}+\sqrt{\log(en)}\Big]\mathbb{I}_{\{s_1=0\}\cup\{s_2=0\}}\Big\}+\sqrt{2/\pi}.\end{aligned}$$

Before proceeding to the proofs of Theorem 3.4.10 and Theorem A.1.1, let us add a brief remark below on why our proof technique does not seem to work for more general functions f^* in $\mathfrak{R}^d$.

Remark A.3.17. *The main technical reason why our adaptive results Theorem 3.4.10 and Theorem A.1.1 deal only with functions of the form* (3.49) *and not more general functions in $\mathfrak{R}^d$ is that our proof technique seems to break down for these general functions. In particular, for more complicated functions $f^* \in \mathfrak{R}^d$, it seems that it may not be possible to obtain a partition of $\mathcal{I}$ into a constant (depending only on d) number of rectangles Q satisfying the conditions in Lemma A.3.14.*

Proof of Theorem 3.4.10

We shall use (A.40). Note that the right hand side of (A.40) consists of infimum over all

$$\widetilde{\boldsymbol{\theta}} \in K = \mathcal{C}(V) = \{(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) : V_{\mathrm{HK}\mathbf{0}}(f) \leq V\}.$$

It is clear then that (A.40) will still be true if we restrict the infimum to $\widetilde{\boldsymbol{\theta}}$ belonging to any subset of K. We shall consider the subset

$$\left\{(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) : f \in \mathfrak{R}_1^d(c) \text{ and } V_{\mathrm{HK}\mathbf{0}}(f) = V\right\}.$$

We shall therefore fix a function $f \in \mathfrak{R}_1^d(c)$ with $V_{\mathrm{HK}\mathbf{0}}(f) = V$ and bound the Gaussian width

$$\mathbb{E} \sup_{\boldsymbol{\alpha} \in \mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}) \cap \mathcal{B}_2(\mathbf{0},1)} \langle Z, \boldsymbol{\alpha} \rangle$$

where $\widetilde{\boldsymbol{\theta}} = \mathbf{A}\widetilde{\boldsymbol{\beta}} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$. Due to the structure of f, there exists $\mathbf{i}^*$ such that $\widetilde{\beta}_{\mathbf{i}} = 0$ for all $\mathbf{i} \notin \{\mathbf{0}, \mathbf{i}^*\}$. Explicitly, if $f = \mathbb{I}_{[\mathbf{x}^*, \mathbf{1}]}$, then $\mathbf{i}^*$ is the index corresponding to the smallest design point $\mathbf{x} = (i_1/n_1, \ldots, i_d/n_d)$ satisfying $\mathbf{x} \succeq \mathbf{x}^*$.

The minimum length assumption (3.50) implies that the sets $\{\mathbf{i} : \mathbf{i} \succeq \mathbf{i}^*\}$ and $\{\mathbf{i} : \mathbf{i} \prec \mathbf{i}^*\}$ each have $\geq cn$ elements. Therefore if $\boldsymbol{\alpha} \in \mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}) \cap \mathcal{B}_2(\mathbf{0}, 1)$, the pigeonhole principle and fact that $\|\boldsymbol{\alpha}\| \leq 1$ together imply that there exist $\mathbf{i}^u \succeq \mathbf{i}^*$ and $\mathbf{i}^\ell \prec \mathbf{i}^*$ such that $|\alpha_{\mathbf{i}^u}| \leq (cn)^{-1/2}$ and $|\alpha_{\mathbf{i}^\ell}| \leq (cn)^{-1/2}$. This implies that

$$\mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}) \subseteq \bigcup_{\mathbf{i}^u, \mathbf{i}^\ell : \mathbf{i}^\ell \prec \mathbf{i}^* \preceq \mathbf{i}^u} T(\mathbf{i}^u, \mathbf{i}^\ell, 2(cn)^{-\frac{1}{2}}).$$

where $T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)$ is defined in (A.44). By Lemma D.1 of [44] and noting that the above union is over $\leq n^2$ indices, we obtain

$$\begin{aligned}
&\mathbb{E} \sup_{\boldsymbol{\alpha} \in \mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}) \cap \mathcal{B}_2(\mathbf{0},1)} \langle Z, \boldsymbol{\alpha} \rangle \\
&\quad \leq \max_{\mathbf{i}^u, \mathbf{i}^\ell : \mathbf{i}^\ell \prec \mathbf{i}^* \preceq \mathbf{i}^u} \mathbb{E} \sup_{\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, 2(cn)^{-\frac{1}{2}})} \langle Z, \boldsymbol{\alpha} \rangle + \sqrt{4 \log n} + \sqrt{\pi/2}.
\end{aligned}$$

The following lemma bounds the expectations appearing on the right-hand side above and is proved below.

Lemma A.3.18. *Let $\mathbf{i}^\ell$ and $\mathbf{i}^u$ satisfy $\mathbf{i}^\ell \prec \mathbf{i}^* \preceq \mathbf{i}^u$. For $\delta \geq 0$,*

$$\mathbb{E} \sup_{\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)} \langle Z, \boldsymbol{\alpha} \rangle \leq C_d(1 + 2\delta\sqrt{n})(\log(en))^{\frac{3d}{4}} (\log(e \log(en)))^{\frac{2d-1}{4}}.$$

Plugging $\delta = 2(cn)^{-1/2}$ into Lemma A.3.18 yields

$$\mathbb{E} \sup_{\boldsymbol{\alpha} \in \mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}) \cap \mathcal{B}_2(\mathbf{0},1)} \langle Z, \boldsymbol{\alpha} \rangle \leq C_d (\log(en))^{\frac{3d}{4}} (\log(e \log(en)))^{\frac{2d-1}{4}} + \sqrt{4 \log n} + \sqrt{\pi/2}$$
$$\leq C_d (\log(en))^{\frac{3d}{4}} (\log(e \log(en)))^{\frac{2d-1}{4}}.$$

Plugging this bound into the oracle inequality (A.40) concludes the proof of Theorem 3.4.10.

It therefore suffices to prove Lemma A.3.18. For every partition $Q_1, \ldots, Q_R$ of $\mathcal{I}$ into rectangles of the form (A.48), we have

$$\mathbb{E} \sup_{\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)} \langle Z, \boldsymbol{\alpha} \rangle \leq \sum_{r=1}^{R} \mathbb{E} \sup_{\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)} \langle Z_{Q_r}, \boldsymbol{\alpha}_{Q_r} \rangle. \tag{A.54}$$

Our idea is to choose the partition such that each Q_r satisfies the conditions of Lemma A.3.14 so that then each $\boldsymbol{\alpha}_{Q_r}$ for $\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)$ satisfies (A.52) which would allow us to bound each expectation appearing in the right hand side above.

Here is how we construct the partition. For each $j \in \{1, \ldots, d\}$ we partition the interval $\{0, \ldots, n_j - 1\}$ into at most 4 intervals by splitting at $i^u_j + 0.5$, $i^\ell_j + 0.5$, and $i^*_j - 0.5$. We then take the Cartesian product of these partitions over $j = 1, \ldots, d$ to obtain a partition $Q_1, \ldots, Q_R$ of $\mathcal{I}$ into at most $R \leq 4^d$ rectangles.

We now check that the rectangles each satisfy the three conditions of Lemma A.3.14. Let $Q = [\mathbf{q}^\ell, \mathbf{q}^u]$ be one of the rectangles of the above partition. Here the auxiliary technical Lemma A.5.1 will be used. By the second part of Lemma A.5.1, the quantity $t(\mathbf{0}, \mathbf{i})$ is zero for all $\mathbf{i} \in Q$ except when $\mathbf{i} = \mathbf{q}^\ell$. Now suppose $\mathbf{i}^* \preceq \mathbf{q}^u$. Due to the splits at $i^*_j - 0.5$ for all j, we have $\max\{q^\ell_j, i^*_j\} = q^\ell_j$ for all j, so the second part of Lemma A.5.1 implies $t(\mathbf{i}^*, \mathbf{i}) = 0$ for all $\mathbf{i} \in Q$ except $\mathbf{i} = \mathbf{q}^\ell$. Thus condition (a) is satisfied.

Recall that by assumption $\mathbf{i}^\ell \prec \mathbf{i}^* \preceq \mathbf{i}^u$, so $L^c_u \cap L_\ell$ is empty. Thus by definition (A.43), $\mathfrak{s}(\mathbf{i}) \in \{0, 1\}$ for all $\mathbf{i}$. Thus condition (b) holds automatically. Finally, note that Q is contained in either L_u or L^c_u due to the splits at $i^u_j + 0.5$ for all $j \in [d]$. Similarly Q is contained in either L_ℓ or L^c_u. Thus condition (c) holds.

We have thus proved that for each rectangle $Q_r, r = 1, \ldots, R$, the inequality (A.52) holds. We now fix such a rectangle $Q \in \{Q_1, \ldots, Q_R\}$ and bound the expected supremum term appearing on the right hand side of (A.54). By condition (c) of Lemma A.3.14, there exists $s \in \{-1, 0, 1\}$ such that $\mathfrak{s}(\mathbf{i}) = s$ for all $\mathbf{i} \in Q$. We separate the two cases where $s = 0$ and $s \neq 0$.

Case 1: $s = 0$**.** Because $L^c_u \cap L_\ell$ is empty, we must have $\widetilde{\mathfrak{s}}(\mathbf{i}) \in \{0, 1\}$ for all $\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}$. For $\mathbf{i} \in Q$ such that $\mathbf{i} \succ \mathbf{q}^\ell$, we further have $\widetilde{\mathfrak{s}}(\mathbf{i}) = s = 0$. Thus (A.52) can be rewritten as

$$\sum_{\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\} : \mathbf{i} \succ \mathbf{q}^\ell} |(D\boldsymbol{\alpha}_Q)_\mathbf{i}| + \sum_{\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\} : \mathbf{i} \not\succ \mathbf{q}^\ell} (|(D\boldsymbol{\alpha}_Q)_\mathbf{i}| - \widetilde{\mathfrak{s}}(\mathbf{i})(D\boldsymbol{\alpha}_Q)_\mathbf{i}) \leq 2\delta.$$

Using the fact that $-(D\boldsymbol{\alpha}_Q)_{\mathbf{i}} \le |(D\boldsymbol{\alpha}_Q)_{\mathbf{i}}|$ and

$$(|(D\boldsymbol{\alpha}_Q)_{\mathbf{i}}| - (D\boldsymbol{\alpha}_Q)_{\mathbf{i}}) \le 2(|(D\boldsymbol{\alpha}_Q)_{\mathbf{i}}| - \widetilde{\mathfrak{s}}(\mathbf{i})(D\boldsymbol{\alpha}_Q)_{\mathbf{i}})$$

for every $\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}$, we deduce

$$\sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}} (|(D\boldsymbol{\alpha}_Q)_{\mathbf{i}}| - (D\boldsymbol{\alpha}_Q)_{\mathbf{i}}) \le 4\delta.$$

Thus, Lemma A.3.15 (with 4δ in place of δ, as well as $t = 1$ and $\sigma = 1$) implies

$$\begin{aligned}&\mathbb{E} \sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)} \langle Z_Q, \boldsymbol{\alpha}_Q\rangle \\ &\le C_d(1 + 4\delta\sqrt{n})(\log(en))^{\frac{3d}{4}}(\log(e\log(en)))^{\frac{2d-1}{4}} \end{aligned} \tag{A.55}$$

Case 2. $s \neq 0$**.** Then the fact that $L_u^c \cap L_\ell$ is empty implies $s = 1$. Thus $\widetilde{\mathfrak{s}}(\mathbf{i}) = 1$ for all $\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}$. Therefore, the shape constraint (A.52) can be rewritten as

$$\sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}} (|(D\boldsymbol{\alpha}_Q)_{\mathbf{i}}| - (D\boldsymbol{\alpha}_Q)_{\mathbf{i}}) \le 2\delta.$$

Thus the above bound (A.55) holds as well.

Returning to the earlier inequality (A.54) and recalling the sum is over $R \le 4^d$ rectangles, we obtain

$$\mathbb{E} \sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)} \langle Z, \boldsymbol{\alpha}\rangle \le C_d(1 + 2\delta\sqrt{n})(\log(en))^{\frac{3d}{4}}(\log(e\log(en)))^{\frac{2d-1}{4}}.$$

We have thus proved Lemma A.3.18 which completes the proof of Theorem 3.4.10.

Proof of Theorem A.1.1

In this proof we take $d = 2$. This proof is similar to but longer than the proof of Theorem 3.4.10. We upper bound the oracle inequality (A.40) by taking the infimum only over $\boldsymbol{\theta}$ of the form $\boldsymbol{\theta} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$ where $f \in \widetilde{\mathfrak{R}}_1^2(c)$ and $V_{\mathrm{HK}\mathbf{0}}(f) = V$. It then suffices to control the Gaussian width $\mathbb{E}\sup_{\boldsymbol{\alpha}\in\mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}):\|\boldsymbol{\alpha}\|\le 1}\langle Z, \boldsymbol{\alpha}\rangle$ for such $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta}$.

Let $\mathbf{i}^* := (i_1^*, i_2^*) \neq (0, 0)$ be the unique index such that $\widetilde{\beta}_{\mathbf{i}^*} \neq 0$, which is guaranteed by the form (3.49) of functions in $\widetilde{\mathfrak{R}}_1^2$. Specifically, if $f \in \widetilde{\mathfrak{R}}_1^2$, it is of the form $a_1\mathbb{I}_{[\mathbf{x}^*,\mathbf{1}]} + a_0$, and $\mathbf{i}^*$ is the index corresponding to the smallest design point $\mathbf{x}$ satisfying $\mathbf{x} \succeq \mathbf{x}^*$.

The minimum size assumption (A.1) implies that the set $\{\mathbf{i} : \mathbf{i} \succeq \mathbf{i}^*\}$ and its complement have cardinality $\ge cn$. By the pigeonhole principle, for any $\boldsymbol{\alpha}$ satisfying $\|\boldsymbol{\alpha}\| \le 1$, there exists some $\mathbf{i}^u \succeq \mathbf{i}^*$ such that $|\alpha_{\mathbf{i}^u}| \le (cn)^{-1/2}$ and some $\mathbf{i}^\ell \not\succeq \mathbf{i}^*$ such that $|\alpha_{\mathbf{i}^\ell}| \le (cn)^{-1/2}$. Then we have

$$|\alpha_{\mathbf{i}^u} - \alpha_{\mathbf{i}^\ell}| \le 2(cn)^{-\frac{1}{2}}.$$

Thus,

$$\mathcal{T}_{\mathcal{C}(V)}(\widetilde{\boldsymbol{\theta}}) \cap \mathcal{B}_2(\mathbf{0},1) \subseteq \bigcup_{\mathbf{i}^u,\mathbf{i}^\ell:\mathbf{i}^u\succeq\mathbf{i}^*,\mathbf{i}^\ell\not\succeq\mathbf{i}^*} T(\mathbf{i}^u,\mathbf{i}^\ell,2(cn)^{-\frac{1}{2}}),$$

where $T(\mathbf{i}^u,\mathbf{i}^\ell,2(cn)^{-\frac{1}{2}})$ is defined in (A.44).

Using Lemma D.1 of [44] and noting the above union is over $\leq n^2$ sets, we then have

$$\begin{aligned}&\mathbb{E}\sup_{\boldsymbol{\alpha}\in\mathcal{T}_{\mathcal{C}(V)}:\|\boldsymbol{\alpha}\|\leq 1}\langle Z,\boldsymbol{\alpha}\rangle\\ &\quad\leq \max_{\mathbf{i}^u,\mathbf{i}^\ell:\mathbf{i}^u\succeq\mathbf{i}^*,\mathbf{i}^\ell\not\succeq\mathbf{i}^*}\mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,2(cn)^{-\frac{1}{2}})}\langle Z,\boldsymbol{\alpha}\rangle+\sqrt{4\log n}+\sqrt{\pi/2}.\end{aligned}$$

Therefore it remains to bound the expectation on the right-hand side for each set $T(\mathbf{i}^u,\mathbf{i}^\ell,2(cn)^{-\frac{1}{2}})$. This is the content of the following lemma.

Lemma A.3.19. *For $d=2$, $\delta\geq 0$ and every $\mathbf{i}^u\succeq\mathbf{i}^*$ and $\mathbf{i}^\ell\not\succeq\mathbf{i}^*$,*

$$\begin{aligned}&\mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)}\langle Z,\boldsymbol{\alpha}\rangle\\ &\leq c(1+(\delta\sqrt{n})^{\frac{1}{2}})(\log(\delta\sqrt{n}+1))^{\frac{3}{4}}\\ &\qquad+(1+\delta\sqrt{n})\left[(\log(en))^{\frac{3}{2}}(\log(e\log(en)))^{\frac{3}{4}}+\sqrt{\log(4+2\delta\sqrt{n})}\right].\end{aligned}$$

The proof of this result is quite involved and given below. Note that Lemma A.3.19 only deals with $d=2$ while Lemma A.3.18 is true for arbitrary d. On the other hand, for $d=2$, Lemma A.3.19 is stronger than Lemma A.3.18 because it applies to a more general set of indices $\mathbf{i}^\ell$ (the condition $\mathbf{i}^\ell\not\succeq\mathbf{i}^*$ is weaker than $\mathbf{i}^\ell\prec\mathbf{i}^*$).

Before proving Lemma A.3.19, let us quickly note that plugging in $\delta=2(cn)^{-\frac{1}{2}}$ in Lemma A.3.19 yields

$$\mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,2(cn)^{-\frac{1}{2}})}\langle Z,\boldsymbol{\alpha}\rangle\leq C(\log(en))^{\frac{3}{2}}(\log(e\log(en)))^{\frac{3}{4}},$$

which concludes the proof of Theorem A.1.1.

Let $Q_1,\ldots,Q_R$ be the partition constructed in the proof of Theorem 3.4.10. We shall first prove that each rectangle $Q=[\mathbf{q}^\ell,\mathbf{q}^u]$ in $\{Q_1,\ldots,Q_R\}$ satisfies the three conditions of Lemma A.3.14. Note that this was proved in the proof of Theorem 3.4.10 under the stronger condition $\mathbf{i}^\ell\prec\mathbf{i}^*$ but now we are working under the weaker condition $\mathbf{i}^\ell\not\succeq\mathbf{i}^*$. Conditions (a) and (c) hold by exactly the same argument as in proof of Theorem 3.4.10. To show condition (b), we need to crucially use $d=2$. If $\mathbf{i}\in Q$ satisfies $\mathbf{i}\succ\mathbf{q}^\ell$, then $t(\mathbf{i}',\mathbf{i})=0$ for all $\mathbf{i}'\preceq\mathbf{i}$ except $\mathbf{i}'=\mathbf{i}$, so condition (b) holds automatically. We now consider $\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}$ such that $\mathbf{i}\not\succ\mathbf{q}^\ell$. Suppose without loss of generality that $\mathbf{i}=(q_1^\ell,i_2)$ for $i_2>q_2^\ell$; the other case $\mathbf{i}=(i_1,q_2^\ell)$ for $i_1>q_1^\ell$ can be handled similarly. Then $t(\mathbf{i}',\mathbf{i})=1$ only when $\mathbf{i}'$ satisfies

$i'_2 = i_2$ and $i'_1 \le q_1^\ell$. Therefore, to verify condition (b) for such $\mathbf{i}$, it suffices to show the stronger claim that $\mathfrak{s}(\mathbf{i}')$ is constant over all $\mathbf{i}'$ in the set

$$\{\mathbf{i}' : i'_1 \le q_1^\ell,\, i'_2 \in [q_2^\ell + 1, q_2^u]\} \tag{A.56}$$

satisfying $\mathfrak{s}(\mathbf{i}') \neq 0$. Suppose for sake of contradiction that $\mathbf{i}'$ and $\mathbf{i}''$ belong to this set and satisfy $\mathfrak{s}(\mathbf{i}') = 1$ and $\mathfrak{s}(\mathbf{i}') = -1$. Then $\mathbf{i}' \in L_u \cap L_\ell^c$ and $\mathbf{i}'' \in L_u^c \cap L_\ell$. We then must have $i'_j \le i_j^u < i''_j$ for some $j \in \{1,2\}$, and $i''_j \le i_j^\ell < i'_j$ for some $j \in \{1,2\}$. From here, we deduce that either i_2^u or i_2^ℓ lies in $[\min\{i'_2, i''_2\}, \max\{i'_2, i''_2\}) \subseteq [q_2^\ell, q_2^u)$. But due to the splits at $i_2^u + 0.5$ and $i_2^\ell + 0.5$ in the construction of the partition, this is a contradiction.

A similar argument shows that $\mathfrak{s}(\mathbf{i}')$ is constant over all $\mathbf{i}'$ in the set

$$\{\mathbf{i}' : i'_2 \le q_2^\ell,\, i'_1 \in [q_1^\ell + 1, q_1^u]\} \tag{A.57}$$

satisfying $\mathfrak{s}(\mathbf{i}') \neq 0$. Let this constant value be denoted by s_1, and let the constant value for the earlier set (A.56) be denoted by s_2. Thus condition (b) holds as well, and we have the inequality (A.52) by Lemma A.3.14.

We shall now bound the Gaussian width

$$\mathbb{E} \sup_{\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)} \langle Z_Q, \boldsymbol{\alpha}_Q \rangle$$

by splitting into the two cases $s \neq 0$ and $s = 0$ where s is the common value of $\mathfrak{s}(\mathbf{i})$ for $\mathbf{i} \in Q$ (the fact that $\mathfrak{s}(\mathbf{i})$ is the same for every $\mathbf{i} \in Q$ is guaranteed by condition (c) of Lemma A.3.14).

Case 1: $s \neq 0$ By definition $\mathfrak{s}(\mathbf{i}) = s$ for all $\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}$, so (A.52) can be written as

$$\sum_{\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}} (|(D(s\boldsymbol{\alpha}_Q))_\mathbf{i}| - (D(s\boldsymbol{\alpha}_Q))_\mathbf{i}) = \sum_{\mathbf{i} \in Q \setminus \{\mathbf{q}^\ell\}} (|(D\boldsymbol{\alpha}_Q)_\mathbf{i}| - s(D\boldsymbol{\alpha}_Q)_\mathbf{i}) \le 2\delta.$$

Since the sets $T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)$ and $-T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)$ have the same Gaussian width, we may apply Lemma A.3.15 to obtain

$$\begin{aligned} &\mathbb{E} \sup_{\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)} \langle Z_Q, \boldsymbol{\alpha}_Q \rangle \\ &\le c(1 + 2\delta\sqrt{n})(\log(en))^{\frac{3}{2}} (\log(e\log(en)))^{\frac{3}{4}} \end{aligned} \tag{A.58}$$

Case 2: $s = 0$ In this case $\widetilde{\mathfrak{s}}(\mathbf{i}) = 0$ for all $\mathbf{i} \succ \mathbf{q}^\ell$, and is otherwise equal to s_1 (if $i_2 = q_2^\ell$) or s_2 (if $i_1 = q_1^\ell$) because we showed that $\mathfrak{s}(\mathbf{i}')$ is constant over the sets (A.56) and (A.57). So, inequality (A.52) can be rewritten as

$$\begin{aligned} \sum_{\mathbf{i} \in Q : \mathbf{i} \succ \mathbf{q}^\ell} |(D(\boldsymbol{\alpha}_Q))_\mathbf{i}| + \sum_{\substack{\mathbf{i} = (i_1, q_2^\ell): \\ i_1 \in [q_1^\ell + 1, q_1^u]}} (|(D(\boldsymbol{\alpha}_Q))_\mathbf{i}| - s_1(D(\boldsymbol{\alpha}_Q))_\mathbf{i}) \\ + \sum_{\substack{\mathbf{i} = (q_1^\ell, i_2): \\ i_2 \in [q_2^\ell + 1, q_2^u]}} (|(D(\boldsymbol{\alpha}_Q))_\mathbf{i}| - s_2(D(\boldsymbol{\alpha}_Q))_\mathbf{i}) \le 2\delta. \end{aligned} \tag{A.59}$$

Let us define $\boldsymbol{\alpha}_Q^{(0)} := \sum_{\mathbf{i}\in Q:\mathbf{i}\succ\mathbf{q}^\ell}(D(\boldsymbol{\alpha}_Q))_\mathbf{i}$ and $\boldsymbol{\alpha}_Q^{(1)} := \sum_{\mathbf{i}\in Q:\mathbf{i}\not\succ\mathbf{q}^\ell}(D(\boldsymbol{\alpha}_Q))_\mathbf{i}$. Since $\boldsymbol{\alpha}_Q = \boldsymbol{\alpha}_Q^{(0)} + \boldsymbol{\alpha}_Q^{(1)}$, we obtain

$$\mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)}\langle Z_Q,\boldsymbol{\alpha}_Q\rangle \le \mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)}\langle Z_Q,\boldsymbol{\alpha}_Q^{(0)}\rangle + \mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)}\langle Z_Q,\boldsymbol{\alpha}_Q^{(1)}\rangle. \tag{A.60}$$

We now bound the first term in the right hand side above. Because $\widetilde{\mathfrak{s}}(\mathbf{i}) = 0$ for $\mathbf{i}\succ\mathbf{q}^\ell$, inequality (A.59) implies

$$V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\alpha}_Q^{(0)}) = \sum_{\mathbf{i}\in Q:\mathbf{i}\succ\mathbf{q}^\ell}|(D(\boldsymbol{\alpha}_Q))_\mathbf{i}| \le 2\delta,$$

so applying (A.53) yields

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)}\langle Z_Q,\boldsymbol{\alpha}_Q^{(0)}\rangle &\le \mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbb{R}^{|Q|}:\|\boldsymbol{\theta}\|\le 1, V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta})\le 2\delta}\langle Z_Q,\boldsymbol{\theta}\rangle \\ &\le C_d(1+\sqrt{2\delta\sqrt{n}})\big(\log(1+2e\delta\sqrt{n})\big)^{\frac{3}{4}} \\ &\quad + C_d\sqrt{\log(4+2\delta\sqrt{n})}.\end{aligned} \tag{A.61}$$

We turn to the second term in (A.60). Inequality (A.59) implies

$$\begin{aligned}V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\alpha}_Q^{(1)}) &= \sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}:\mathbf{i}\not\succ\mathbf{q}^\ell}|(D\boldsymbol{\alpha}_Q)_\mathbf{i}| \\ &\le s_1\sum_{\substack{\mathbf{i}=(i_1,q_2^\ell):\\ i_1\in[q_1^\ell+1,q_1^u]}}(D\boldsymbol{\alpha}_Q)_\mathbf{i} + s_2\sum_{\substack{\mathbf{i}=(q_1^\ell,i_2):\\ i_2\in[q_2^\ell+1,q_2^u]}}(D\boldsymbol{\alpha}_Q)_\mathbf{i} + 2\delta \\ &= s_1[(\boldsymbol{\alpha}_Q^{(1)})_{q_1^u,q_2^\ell} - (\boldsymbol{\alpha}_Q^{(1)})_{\mathbf{q}^\ell}] + s_2[(\boldsymbol{\alpha}_Q^{(1)})_{q_1^\ell,q_2^u} - (\boldsymbol{\alpha}_Q^{(1)})_{\mathbf{q}^\ell}] + 2\delta.\end{aligned}$$

Lemma A.3.16 then implies

$$\begin{aligned}&\mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)}\langle Z_L,\boldsymbol{\alpha}_L^{(1)}\rangle \\ &\le c\Big\{(1+\delta\sqrt{n})\sqrt{\log(en)}\mathbb{I}_{\{s_1\neq 0\}\cup\{s_2\neq 0\}} \\ &\qquad + \Big[(\delta\sqrt{n})^{\frac{1}{2}} + \sqrt{\log(en)}\Big]\mathbb{I}_{\{s_1=0\}\cup\{s_2=0\}}\Big\} + \sqrt{2/\pi}.\end{aligned} \tag{A.62}$$

Summing the bounds (A.61) and (A.62) yields

$$\begin{aligned}&\mathbb{E}\sup_{\boldsymbol{\alpha}\in T(\mathbf{i}^u,\mathbf{i}^\ell,\delta)}\langle Z_L,\boldsymbol{\alpha}_L\rangle \\ &\le c(1+(\delta\sqrt{n})^{\frac{1}{2}})(\log(\delta\sqrt{n}+1))^{\frac{3}{4}} + c(1+\delta\sqrt{n})\Big[\sqrt{\log(en)} + \sqrt{\log(4+2\delta\sqrt{n})}\Big].\end{aligned} \tag{A.63}$$

Having handled the two cases $s = 0$ and $s \neq 0$, we take the maximum of (A.58) and (A.63) to obtain

$$\begin{aligned}
&\mathbb{E} \sup_{\boldsymbol{\alpha} \in T(\mathbf{i}^u, \mathbf{i}^\ell, \delta)} \langle Z_L, \boldsymbol{\alpha}_L \rangle \\
&\leq c(1 + (\delta\sqrt{n})^{\frac{1}{2}})(\log(\delta\sqrt{n} + 1))^{\frac{3}{4}} \\
&\quad + (1 + \delta\sqrt{n}) \left[(\log(en))^{\frac{3}{2}} (\log(e \log(en)))^{\frac{3}{4}} + \sqrt{\log(4 + 2\delta\sqrt{n})} \right].
\end{aligned}$$

Finally, in view of the inequality (A.54), multiplying this bound by $4^2 = 16$ (the maximum number of rectangles in the partition constructed at the beginning of this proof) produces the final bound given by Lemma A.3.19 thereby completing the proof of Theorem A.1.1.

A.4 Proofs of results from Section 3.2 and Section 3.3

This section contains the proofs of all the results from Section 3.2 and Section 3.3. Specifically, we prove Lemma 3.2.2, Lemma 3.2.6, part (ii) of Lemma 3.2.7, Proposition 3.3.2, Proposition 3.3.1, Proposition 3.3.4, Proposition 3.3.3 and Lemma 3.3.5. In addition, we also state and prove a result in Section A.4.4 which asserts that the columns of the design matrix $\mathbf{A}$ span $\mathbb{R}^n$ provided the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct.

A.4.1 Proof of Lemma 3.2.2

When $d = 1$, the only rectangles are intervals $[a, b]$, so the definition of entire monotonicity (3.19) reduces to $0 \leq \Delta(f, [a, b]) = f(b) - f(a)$ for all $0 \leq a \leq b \leq 1$, which is precisely the definition of $\mathcal{F}^d_{\mathrm{M}}$ (3.20).

More generally for $d \geq 1$, suppose $\mathbf{a}, \mathbf{b} \in [0, 1]^d$ agree in all but one component, that is, $|\{i : a_i \neq b_i\}| = 1$. Then entire monotonicity implies $0 \leq \Delta(f, [\mathbf{a}, \mathbf{b}]) = f(\mathbf{b}) - f(\mathbf{a})$. To see how this inequality implies monotonicity (3.20) note that for $\mathbf{a} \preceq \mathbf{b}$ we can apply the above inequality repeatedly to obtain

$$f(\mathbf{a}) \leq f(b_1, a_2, \ldots, a_d) \leq f(b_1, b_2, a_3, \ldots, a_d) \leq \cdots \leq f(\mathbf{b}).$$

Thus $\mathcal{F}^d_{\mathrm{EM}} \subseteq \mathcal{F}^d_{\mathrm{M}}$ for $d \geq 1$.

Finally, for $d \geq 2$ consider the function $f : [0, 1]^d \to \mathbb{R}$ defined by

$$f(\mathbf{u}) \coloneqq \begin{cases} 0 & \max\{u_1, u_2\} < 1/2 \\ 3 & \min\{u_1, u_2\} \geq 1/2 \\ 2 & \text{otherwise} \end{cases}$$

Note that f is constant in all components except the first two. One can directly check that $f \in \mathcal{F}^d_{\mathrm{M}}$. However, for $\mathbf{a} = (\frac{1}{4}, \frac{1}{4}, 0, \ldots, 0)$ and $\mathbf{b} = (\frac{3}{4}, \frac{3}{4}, 0, \ldots, 0)$, we have

$$\Delta(f; [\mathbf{a}, \mathbf{b}]) = 3 - 2 - 2 + 0 = -1 < 0,$$

so $f \notin \mathcal{F}^d_{\mathrm{EM}}$.

A.4.2 Proof of Lemma 3.2.6

Let $\mathcal{P}^*$ be given by the d univariate partitions (3.21) and let $\mathcal{P}$ be the split of $[0,1]^d$ formed from these univariate partitions (as described after (3.21)). Because $\mathcal{P}$ forms a split of $[0,1]^d$, it follows from Owen [69, Lemma 1] that

$$V^{(d)}(f;[0,1]^d) = \sum_{A\in\mathcal{P}} V^{(d)}(f;A)$$

where $V^{(d)}(f;A)$ is the Vitali variation of f on the rectangle A (which is defined analogously to $V^{(d)}(f;[0,1]^d)$). Let us now fix a rectangle $A = [\mathbf{a},\mathbf{b}] \in \mathcal{P}$ where $\mathbf{a} = (a_1,\dots,a_d)$ and $\mathbf{b} = (b_1,\dots,b_d)$. Because f is rectangular piecewise constant with respect to $\mathcal{P}^*$, it follows that f is constant on each of the sets $B_1 \times \cdots \times B_d$ where each B_i is either $\{b_i\}$ or $[a_i,b_i)$. Using this, it is easy to observe that

$$V^{(d)}(f;A) = |\Delta(f;A)|$$

which completes the proof of Lemma 3.2.6.

A.4.3 Proof of part (ii) of Lemma 3.2.7

If $f \in \mathcal{F}^d_{\mathrm{EM}}$ is entirely monotone, then one can check that for each S,

$$V^{(|S|)}(f;S;[0,1]^d) = \Delta(f;U_S),$$

where U_S is the face adjacent to $\mathbf{0}$ defined earlier (3.24). Thus the HK variation of f is the sum of quasi-volumes of all faces adjacent to $\mathbf{0}$. From the definition of quasi-volume (3.17), this sum involves only the value of f at vertices of $[0,1]^d$ (possibly multiplied by -1), and one can check that all terms cancel except for $f(\mathbf{1}) - f(\mathbf{0})$.

A.4.4 Statement and proof of a fact about the design matrix A

Recall the definition of $\mathbf{A}$ as the matrix whose columns are the elements of the finite set $\mathcal{Q} \coloneqq \{\mathbf{v}(\mathbf{z}) : \mathbf{z} \in [0,1]^d\}$.

Lemma A.4.1. *Suppose $\mathbf{x}_1,\dots,\mathbf{x}_n$ are unique. Then the columns of $\mathbf{A}$ span $\mathbb{R}^n$.*

Proof. It suffices to show the standard basis vector $\mathbf{e}_i$ lies in the column space of $\mathbf{A}$, for each $i = 1,\dots,n$.

Fix i. If $\mathbf{x}_i = \mathbf{1}$, then $\mathbf{e}_i = \mathbf{v}(\mathbf{1}) \in \mathcal{Q}$, which concludes the proof.

Otherwise we assume $\mathbf{x}_i \neq \mathbf{1}$. Let $\mathbf{u}^\delta$ be defined by $u^\delta_j \coloneqq \min\{1, (\mathbf{x}_i)_j + \delta\}$ for $j = 1,\dots,d$. There exists $\delta > 0$ such that the hyperrectangle $[\mathbf{x}_i, \mathbf{u}^\delta]$ contains no design point except $\mathbf{x}_i$. Let $S \coloneqq \{j : u^\delta_j \neq (\mathbf{x}_i)_j\}$, and note that the rectangle $[\mathbf{x}_i, \mathbf{u}^\delta]$ is $|S|$-dimensional.

For a subset $S' \subseteq [d]$ let $\mathbf{e}_{S'}$ denote the indicator vector of S'; that is, $(\mathbf{e}_{S'})_j$ is 1 if $j \in S'$ and is zero otherwise. We claim

$$\mathbf{e}_i = \sum_{S' \subseteq S} (-1)^{|S'|} \mathbf{v}(\mathbf{x}_i + \delta \mathbf{e}_{S'}).$$

To verify this, note that an inclusion-exclusion argument shows that the right-hand side is $(\mathbb{I}_{[\mathbf{x}_i, \mathbf{u}^\delta)}(\mathbf{x}_1), \ldots, \mathbb{I}_{[\mathbf{x}_i, \mathbf{u}^\delta)}(\mathbf{x}_n))$, and this is $\mathbf{e}_i$ due to the fact that $[\mathbf{x}_i, \mathbf{u}^\delta)$ contains no design point except $\mathbf{x}_i$.

□

A.4.5 Proof of Proposition 3.3.2

If Proposition 3.3.2 holds for a given design $\mathbf{x}_1, \ldots, \mathbf{x}_n$, then adding an additional design point $\mathbf{x}_{n+1} \coloneqq \mathbf{x}_i$ that is a copy of one of the original design points simply gives $\mathbf{A}$ a new row that is a copy of its ith row, and one can observe that the equality in the proposition still holds even after adding this extra design point. Thus without loss of generality we may assume the design points are distinct.

Suppose we replace the original design $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ with $\mathcal{U} \coloneqq \prod_{j=1}^d \mathcal{U}_j$ where $\mathcal{U}_j = \{0, (\mathbf{x}_1)_j, \ldots, (\mathbf{x}_n)_j\}$ for each $j = 1, \ldots, d$. This is a lattice that contains the original design. Using this new design, we define a square matrix $\mathbf{A}'$ whose kth column is $(\mathbb{I}_{[\mathbf{u}_k, \mathbf{1}]}(\mathbf{u}_1), \ldots, \mathbb{I}_{[\mathbf{u}_k, \mathbf{1}]}(\mathbf{u}_m))$. Let $\mathbf{u}_1 = \mathbf{0}$ so that the first column of $\mathbf{A}'$ is $\mathbf{1}$. If we let $K \coloneqq (k_1, \ldots, k_n)$ be such that $\mathbf{u}_{k_i} = \mathbf{x}_i$ so that it indexes the elements of the new design that are also in the old design, then we claim $\{(\mathbf{A}'\boldsymbol{\beta}')_K : \beta'_k \geq 0, \forall k \geq 2\} = \{\mathbf{A}\boldsymbol{\beta} : \beta_j \geq 0, \forall j \geq 2\}$. Indeed, this holds simply because each column of $(\mathbf{A}')_K$ is also a column in $\mathbf{A}$, so both sets are linear combinations of the same columns with the same nonnegativity constraints.

Thus it remains to show

$$\{(\mathbf{A}'\boldsymbol{\beta}')_K : \beta'_k \geq 0, \forall k \geq 2\} = \{(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n) : f \in \mathcal{F}^d_{\mathrm{EM}}\}.$$

We first show the forward inclusion $\subseteq$. Suppose $\boldsymbol{\beta}'$ satisfies $\beta'_k \geq 0$ for all $k \geq 2$. If $f \coloneqq \sum_{k=1}^m \beta'_k \cdot \mathbb{I}_{[\mathbf{u}_k, \mathbf{1}]}$, then $(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)) = (\mathbf{A}'\boldsymbol{\beta}')_K$. We now show $f \in \mathcal{F}^d_{\mathrm{EM}}$. For each pair of distinct points $\mathbf{a} \preceq \mathbf{b}$ in $[0,1]^d$, we want to show $\Delta(f; [\mathbf{a}, \mathbf{b}]) \geq 0$. Then there exist a pair $\mathbf{u}_k \preceq \mathbf{u}_{k'}$ in $\mathcal{U}$ such that $f(\mathbf{a}) = f(\mathbf{u}_k)$, $f(\mathbf{b}) = f(\mathbf{u}_{k'})$, and $\{j : \mathbf{a}_j \neq \mathbf{b}_j\} = \{j : (\mathbf{u}_k)_j \neq (\mathbf{u}_{k'})_j\}$, so that $\Delta(f; [\mathbf{a}, \mathbf{b}]) = \Delta(f; [\mathbf{u}_k, \mathbf{u}_{k'}])$.

Recall that $\Delta(f; [\mathbf{u}_k, \mathbf{u}_{k'}])$ by definition is the sum of terms of the form $f(\mathbf{u}_\ell)$ for some $\mathbf{u}_\ell \in \mathcal{U}$ (possibly with sign changes), since $\mathcal{U}$ is a lattice. Note that $f(\mathbf{u}_\ell) = \sum_{i : \mathbf{u}_i \preceq \mathbf{u}_\ell} \beta'_i$ for each ℓ. Putting the pieces together with an inclusion-exclusion argument yields

$$\Delta(f; [\mathbf{u}_k, \mathbf{u}_{k'}]) = \sum_{\substack{i : \mathbf{u}_i \preceq \mathbf{u}_{k'}, \\ (\mathbf{u}_i)_j > (\mathbf{u}_k)_j \text{ if } (\mathbf{u}_k)_j < (\mathbf{u}_{k'})_j}} \beta'_i \geq 0.$$

We now show the reverse inclusion $\supseteq$. The matrix $\mathbf{A}'$ is square and has spanning columns (Lemma A.4.1), so it is invertible. Thus there exists $\boldsymbol{\beta}'$ such that $\mathbf{A}'\boldsymbol{\beta}' = (f(\mathbf{u}_1), \ldots, f(\mathbf{u}_m))$. Sub-indexing by K yields $(\mathbf{A}'\boldsymbol{\beta}')_K = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$.

A.4.6 Proof of Proposition 3.3.1

The optimization problem (3.2) only involves the values of the function at $\mathbf{x}_1, \dots, \mathbf{x}_n$. Thus by Proposition 3.3.2, the solution $\widehat{f}_{\text{EM}}$ to the optimization problem (3.2) must satisfy $(\widehat{f}_{\text{EM}}(\mathbf{x}_1), \dots, \widehat{f}_{\text{EM}}(\mathbf{x}_n)) = \mathbf{A}\widehat{\boldsymbol{\beta}}_{\text{EM}}$. It remains to show that the function $\widehat{f}_{\text{EM}}$ defined in the result (3.29) satisfies this equality and also lies in $\mathcal{F}^d_{\text{EM}}$.

The equality holds by definition, since $\widehat{f}_{\text{EM}}$ satisfies

$$\widehat{f}_{\text{EM}}(\mathbf{x}_i) = \sum_{j=1}^{p} (\widehat{\beta}_{\text{EM}})_j \cdot \mathbb{I}_{[\mathbf{z}_j, \mathbf{1}]}(\mathbf{x}_i) = (\mathbf{A}\widehat{\boldsymbol{\beta}}_{\text{EM}})_i, \qquad i = 1, \dots, n.$$

To check $\widehat{f}_{\text{EM}}$ as defined in the result (3.29) lies in $\mathcal{F}^d_{\text{EM}}$, we need to show $\Delta(\widehat{f}_{\text{EM}}; [\mathbf{a}, \mathbf{b}]) \geq 0$ for any rectangle $[\mathbf{a}, \mathbf{b}] \subseteq [0,1]^d$, $\mathbf{a} \neq \mathbf{b}$. Similar to the proof in Section A.4.5, we consider the augmented design $\mathcal{U} \coloneqq \prod_{j=1}^d \mathcal{U}_j$ where $\mathcal{U}_j = \{0, (\mathbf{x}_1)_j, \dots, (\mathbf{x}_n)_j\}$ for each $j = 1, \dots, d$. This is a lattice that contains the original design. Moreover, for each $\mathbf{z}_j$ there exists some $\mathbf{u} \in \mathcal{U}$ such that $\mathbb{I}_{[\mathbf{z}_j, \mathbf{1}]}(\mathbf{x}_i) = \mathbb{I}_{[\mathbf{u}, \mathbf{1}]}(\mathbf{x}_i)$ holds for all $\mathbf{x}_i$. Thus the function defined in the result (3.29) can be written as $\widehat{f}_{\text{EM}} = \sum_{\mathbf{u} \in \mathcal{U}} \widetilde{\beta}_{\mathbf{u}} \mathbb{I}_{[\mathbf{u}, \mathbf{1}]}$ for some coefficients $\{\widetilde{\beta}_{\mathbf{u}} : \mathbf{u} \in \mathcal{U}\}$ that are either zero or equal to $(\widehat{\beta}_{\text{EM}})_j$ for some j. Then, as in Section A.4.5, there exist a pair $\mathbf{u}_k \preceq \mathbf{u}_{k'}$ in $\mathcal{U}$ such that $f(\mathbf{a}) = f(\mathbf{u}_k)$, $f(\mathbf{b}) = f(\mathbf{u}_{k'})$, and $\{j : \mathbf{a}_j \neq \mathbf{b}_j\} = \{j : (\mathbf{u}_k)_j \neq (\mathbf{u}_{k'})_j\}$, so that $\Delta(\widehat{f}_{\text{EM}}; [\mathbf{a}, \mathbf{b}]) = \Delta(\widehat{f}_{\text{EM}}; [\mathbf{u}_k, \mathbf{u}_{k'}])$, and by the same reasoning as in the earlier section, $\Delta(\widehat{f}_{\text{EM}}; [\mathbf{u}_k, \mathbf{u}_{k'}]) = \sum_{\mathbf{u} \in \mathcal{U} : \mathbf{u}_k \prec \mathbf{u} \preceq \mathbf{u}_{k'}} \widetilde{\beta}_{\mathbf{u}} \geq 0$.

A.4.7 Proof of Proposition 3.3.4

If Proposition 3.3.4 holds for a given design $\mathbf{x}_1, \dots, \mathbf{x}_n$, then adding an additional design point $\mathbf{x}_{n+1} \coloneqq \mathbf{x}_i$ that is a copy of one of the original design points simply gives $\mathbf{A}$ a new row that is a copy of its ith row, and one can observe that the equality in the proposition still holds even after adding this extra design point. Thus without loss of generality we may assume the design points are distinct.

We claim that the feasible set $\mathcal{C}(V)$ (3.31) does not change if we append additional columns to $\mathbf{A}$ (and append corresponding components to $\boldsymbol{\beta}$) that are copies of columns already in $\mathbf{A}$. Concretely, if $\mathbf{A}'$ is the augmented matrix (without loss of generality assume the new columns are appended on the right) and $\mathcal{C}'(V) \coloneqq \{\mathbf{A}'\boldsymbol{\beta}' : \sum_{j \geq 2} |\beta'_j| \leq V\}$ is the analogue of $\mathcal{C}(V)$, then the inclusion $\mathcal{C}(V) \subseteq \mathcal{C}'(V)$ holds immediately by noting $\mathbf{A}\boldsymbol{\beta} = \mathbf{A}'\boldsymbol{\beta}'$ and $\sum_{j \geq 2} |\beta_j| = \sum_{j \geq 2} |\beta'_j|$ where $\boldsymbol{\beta}'$ is the result of taking $\boldsymbol{\beta}$ and having coefficients 0 for the added components. For the reverse inclusion, suppose we are given $\mathbf{A}'\boldsymbol{\beta}'$ such that $\sum_{j \geq 2} |\beta'_j| \leq V$. Then $\mathbf{A}'\boldsymbol{\beta}' = \mathbf{A}\boldsymbol{\beta}$ where $\beta_j \coloneqq \sum_{k : \mathbf{A}'_{\cdot,k} = \mathbf{A}_{\cdot,j}} \beta'_j$ so the triangle inequality implies

$$\sum_{j \geq 2} |\beta_j| \leq \sum_{j \geq 2} \left| \sum_{k : \mathbf{A}'_{\cdot,k} = \mathbf{A}_{\cdot,j}} \beta'_j \right| \leq \sum_{j \geq 2} |\beta'_j| \leq V.$$

Above, $\mathbf{A}_{\cdot,j}$ denotes the jth column of $\mathbf{A}$, and $\mathbf{A}'_{\cdot,k}$ denotes the kth column of $\mathbf{A}'$.

Thus, similar to Section A.4.5, we may assume without loss of generality that the columns of $\mathbf{A}$ are $\mathbf{v}(\mathbf{u}_1), \dots, \mathbf{v}(\mathbf{u}_m)$ where $\mathbf{u}_1, \dots, \mathbf{u}_m$ are the elements of the lattice $\prod_{j=1}^d \mathcal{U}_j$ and $\mathcal{U}_j \coloneqq \{0, (\mathbf{x}_1)_j, \dots, (\mathbf{x}_n)_j, 1\}$ for $j = 1, \dots, d$. Note the inclusion of 0 and 1 in each $\mathcal{U}_j$, so that the lattice spans the entire hypercube $[0,1]^d$. Without loss of generality we assume the $\mathbf{u}_j$ are ordered such that $\mathbf{u}_{j'} \preceq \mathbf{u}_j$ implies $j' \le j$. Note that as a result, $\mathbf{u}_1 = \mathbf{0}$.

Fix $\boldsymbol{\beta}$ and let $\mathbf{A}\boldsymbol{\beta}$. Let $f \coloneqq \sum_{j=1}^m \beta_j \mathbb{I}_{[\mathbf{u}_j, \mathbf{1}]}$. By construction we have $f(\mathbf{x}_i) = (\mathbf{A}\boldsymbol{\beta})_i$ for all $i = 1, \dots, n$. It remains to compute the HK$\mathbf{0}$ variation of f. One can check that a maximizing partition in the definition of the Vitali variation (3.22) is the partition induced by the lattice $\prod_{j=1}^d \mathcal{U}_j$ (that is, the unique partition $\mathcal{P}^*$ whose rectangles each intersect the lattice only at its vertices). That is,

$$V^{(d)}(f; [\mathbf{0}, \mathbf{x}_n]) = \sum_{R \in \mathcal{P}^*} |\Delta(f; R)|. \tag{A.64}$$

Similarly, the maximizing partitions for the Vitali variations over each face U_S adjacent to $\mathbf{0}$ (3.24) can also be shown to be induced by the corresponding face of the lattice. By construction, the quasi-volume for the rectangle whose largest vertex is $\mathbf{u}_j$ will turn out to be β_j, so by the definition of HK$\mathbf{0}$ variation (3.25), $V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) = \sum_{j \ge 2} |\beta_j| \le V$.

Conversely, suppose we are given $f : [0,1]^d \to \mathbb{R}$ with $V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) \le V$. Suppose first that the original design $\mathbf{x}_1, \dots, \mathbf{x}_n$ is already a lattice spanning $[0,1]^d$, i.e. $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} = \prod_{j=1}^d \mathcal{U}_j$ and $n = m$. We remove this assumption at the end of the proof.

Because $\mathbf{A}$ has full column rank (Lemma A.4.1), there exists some $\boldsymbol{\beta}$ such that $(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)) = \mathbf{A}\boldsymbol{\beta}$. By the above argument, the function $\widetilde{f} \coloneqq \sum_{j=1}^n \widetilde{\beta}_j \mathbb{I}_{[\mathbf{u}_j, 1]}$ agrees with f at all the $\mathbf{x}_i$ (i.e. all the lattice points $\mathbf{u}_j$) and satisfies $V_{\mathrm{HK}\mathbf{0}}(\widetilde{f}; [0,1]^d) = \sum_{j \ge 2} |\beta_j|$. It then suffices to show $V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) \le V_{\mathrm{HK}\mathbf{0}}(\widetilde{f}; [0,1]^d)$.

Let $\mathcal{P}^*$ be the partition of $[0,1]^d$ induced by the lattice $\prod_{j=1}^d \mathcal{U}_j$. As noted already (A.64), this partition is maximal for the definition of the Vitali variation of $\widetilde{f}$ on $[0,1]^d$. Therefore, since f and $\widetilde{f}$ agree on all the lattice points $\mathbf{u}_j$, their quasi-volumes on all the rectangles of $\mathcal{P}^*$ are the same, so we have

$$V^{(d)}(\widetilde{f}; [0,1]^d) = \sum_{R \in \mathcal{P}^*} \left| \Delta(\widetilde{f}; R) \right| = \sum_{R \in \mathcal{P}^*} |\Delta(f; R)| \le V^{(d)}(f; [0,1]^d) \le V.$$

A similar argument on the lower-dimensional faces adjacent to $\mathbf{0}$ shows that $V^{(|S|)}(\widetilde{f}; S; [0,1]^d) \le V^{(|S|)}(f; S; [0,1]^d)$ for all $S \subseteq [d]$. Summing these inequalities over all Vitali variations in the definition of HK$\mathbf{0}$ variation (3.25) leads to

$$\sum_{j \ge 2} |\beta_j| = V_{\mathrm{HK}\mathbf{0}}(\widetilde{f}; [0,1]^d) \le V_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) \le V$$

as desired.

We now consider the case when the design $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is not a lattice. Recall we have assumed the columns of $\mathbf{A}$ are $\mathbf{v}(\mathbf{u}_1), \ldots, \mathbf{v}(\mathbf{u}_m)$. We can augment $\mathbf{A}$ further by redefining $\mathbf{v}(\mathbf{z})$ as $(\mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{u}_1), \ldots, \mathbb{I}_{[\mathbf{z},\mathbf{1}]}(\mathbf{u}_m))$ which amounts to adding new rows to $\mathbf{A}$. This new matrix, call it $\mathbf{A}''$, is precisely the matrix that would have resulted if our original design $\mathbf{x}_1, \ldots, \mathbf{x}_n$ were the full lattice $\mathbf{u}_1, \ldots, \mathbf{u}_m$. By the above argument, there exists $\boldsymbol{\beta}$ such that $\mathbf{A}''\boldsymbol{\beta} = (f(\mathbf{u}_1), \ldots, f(\mathbf{u}_m))$ and $\sum_{j\geq 2}|\beta_j| \leq V$. Discarding the rows of $\mathbf{A}''$ that correspond to lattice points $\mathbf{u}_j$ that do not belong to the original design $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, we obtain $\mathbf{A}\boldsymbol{\beta} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$.

A.4.8 Proof of Proposition 3.3.3

The optimization problem (3.6) only involves the values of the function at $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Thus by Proposition 3.3.4, $\widehat{f}_{\mathrm{HK}\mathbf{0},V}$ must satisfy $(\widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_1), \ldots, \widehat{f}_{\mathrm{HK}\mathbf{0},V}(\mathbf{x}_n)) = \mathbf{A}\widehat{\boldsymbol{\beta}}_{\mathrm{HK}\mathbf{0},V}$. Furthermore, in Section A.4.7 we construct precisely the function in the result (3.32) and shows that it has HK$\mathbf{0}$ variation equal to $\sum_{j=2}^{p}|(\widehat{\beta}_{\mathrm{HK}\mathbf{0},V})_j|$.

A.4.9 Proof of Lemma 3.3.5

We will argue that the Vapnik-Chervonenkis (VC) dimension of "upper-right rectangles": $\{(\mathbf{z}, \mathbf{1}] : \mathbf{z} \in [0,1]^d\}$ is d. A direct application of the Vapnik-Chervonenkis lemma [88] would then yield Lemma 3.3.5.

To show that the VC dimension is d, one can first check that the set $\{\mathbf{1} - \frac{1}{2}\mathbf{e}_1, \ldots, \mathbf{1} - \frac{1}{2}\mathbf{e}_d\}$ can be shattered by these rectangles, so the VC dimension is $\geq d$. To show that no set $\{\mathbf{a}_1, \ldots, \mathbf{a}_{d+1}\}$ of size $d+1$ can be shattered (so that the VC dimension is $\leq d$), note that there must exist some point $\mathbf{a}_i$ such that the component-wise minimum of the $d+1$ points does not change after removing $\mathbf{a}_i$; thus the rectangles cannot select the other d points without also selecting $\mathbf{a}_i$.

A.4.10 Proof of Lemma 3.3.6

Let us first start by describing some basic notation. Since we are working in the lattice design setting, we shall write the components of a vector $\boldsymbol{\theta} \in \mathbb{R}^n$ by $\boldsymbol{\theta}_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}$ (note that $\mathcal{I}$ is defined in (3.37)). We shall also write the design points as $\mathbf{x}_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}$ where

$$\mathbf{x}_{\mathbf{i}} = \left(\frac{i_1}{n_1}, \ldots, \frac{i_d}{n_d}\right) \qquad \text{for } \mathbf{i} = (i_1, \ldots, i_d).$$

The design matrix $\mathbf{A}$ is $n \times n$. We shall index the rows and columns of $\mathbf{A}$ by $\mathcal{I}$ so that

$$\mathbf{A}(\mathbf{i}, \mathbf{j}) = \mathbb{I}\{\mathbf{x}_{\mathbf{j}} \preceq \mathbf{x}_{\mathbf{i}}\} = \mathbb{I}\{\mathbf{j} \preceq \mathbf{i}\}$$

where $\mathbf{j} \preceq \mathbf{i}$ simply refers to $j_1 \leq i_1, \ldots, j_d \leq i_d$. The key to proving Lemma 3.3.6 is the observation that for every $\boldsymbol{\theta} \in \mathbb{R}^n$, we have

$$\mathbf{A}(D\boldsymbol{\theta}) = \boldsymbol{\theta}. \tag{A.65}$$

In other words, the differencing operator D is simply equal to the inverse of $\mathbf{A}$. From (A.65), it should be clear that (3.39) and (3.40) follow directly from immediately (3.35) and (3.36) respectively. To prove (A.65), we need to show that the $\mathbf{i}^{th}$ component of $\mathbf{A}(D\boldsymbol{\theta})$ equals the $\mathbf{i}^{th}$ component of $\boldsymbol{\theta}$ for every $\mathbf{i} \in \mathcal{I}$. For this, we write

$$\begin{aligned}
(\mathbf{A}(D\boldsymbol{\theta}))_{\mathbf{i}} &= \sum_{\mathbf{j}\in\mathcal{I}} \mathbf{A}(\mathbf{i},\mathbf{j})(D\boldsymbol{\theta})_{\mathbf{j}} \\
&= \sum_{\mathbf{j}\in\mathcal{I}} \mathbb{I}\{\mathbf{j} \preceq \mathbf{i}\}(D\boldsymbol{\theta})_{\mathbf{j}} \\
&= \sum_{\mathbf{j}} \mathbb{I}\{\mathbf{j} \preceq \mathbf{i}\} \sum_{\boldsymbol{\ell}\in\{0,1\}^d} \mathbb{I}\{\boldsymbol{\ell} \preceq \mathbf{j}\} (-1)^{l_1+\cdots+l_d} \theta_{\mathbf{j}-\boldsymbol{\ell}} \\
&= \sum_{\mathbf{k}\in\mathcal{I}} \theta_{\mathbf{k}} \left(\sum_{\boldsymbol{\ell}\in\{0,1\}^d} \mathbb{I}\{\mathbf{0} \preceq \mathbf{k} \preceq \mathbf{i}-\boldsymbol{\ell}\}(-1)^{l_1+\cdots+l_d} \right) \\
&= \sum_{\mathbf{k}\in\mathcal{I}} \theta_{\mathbf{k}} \left(\prod_{u=1}^{d} \sum_{l_u=0}^{1} \mathbb{I}\{0 \le k_u \le i_u - l_u\}(-1)^{l_u} \right) \\
&= \sum_{\mathbf{k}\in\mathcal{I}} \theta_{\mathbf{k}} \prod_{u=1}^{d} \mathbb{I}\{k_u = i_u\} = \theta_{\mathbf{i}}.
\end{aligned}$$

This proves (A.65) and completes the proof of Lemma 3.3.6.

A.5 Proofs of technical lemmas from section A.3

In this section, we prove the all the lemmas stated in Section A.3. Specifically, we provide proofs of Lemma A.3.5, Lemma A.3.6, Lemma A.3.7, Lemma A.3.9, Lemma A.3.10, Lemma A.3.12, Lemma A.3.13, Lemma A.3.14, Lemma A.3.15 and Lemma A.3.16. In addition, we also state and prove Lemma A.5.1 which was used in the proof of Theorem 3.4.10 and which is also needed for the proof of Lemma A.3.14.

A.5.1 Proof of Lemma A.3.5

Let $\Omega \coloneqq [0,1]^d$, and let $\Omega_0 \coloneqq \Omega \backslash \{\mathbf{x}_1\}$ be the result of removing the first design point $\mathbf{x}_1 \coloneqq \mathbf{0}$. Recall that by definition (A.10), the elements of $\mathcal{D}_{n_1,\ldots,n_d}$ are of the form $\mathbf{A}\boldsymbol{\beta}$ where $\beta_j \ge 0$ for $j \ge 2$. Recall also that the jth column of $\mathbf{A}$ is $\mathbf{v}(\mathbf{x}_j)$ due to the lattice design (3.34) so $(\mathbf{A}\boldsymbol{\beta})_i = \sum_{i':\mathbf{x}_{i'} \preceq \mathbf{x}_i} \beta_{i'}$ for $i = 1, \ldots, n$. This suggests we can express $\mathcal{D}_{n_1,\ldots,n_d}$ in terms of distribution functions.

Given such a $\boldsymbol{\beta}$, we define a measure μ supported on Ω_0 by $\mu\{\mathbf{x}_j\} = \beta_j$ for $j \ge 2$. We also let $b \coloneqq \beta_1$. If we consider the distribution function $F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}) \coloneqq (\mu + b\delta_{\mathbf{x}_1})([\mathbf{0}, \mathbf{x}])$ of the signed measure $\mu + b\delta_{\mathbf{x}_1}$, then $F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_i) = \sum_{i':\mathbf{x}'_i \preceq \mathbf{x}_i} \beta_{i'} = (\mathbf{A}\boldsymbol{\beta})_i$ for all $i = 1, \ldots, n$

Conversely, given any measure μ supported on Ω_0 and real number b, we may define $\beta_j \coloneqq (\mu + b\delta_{\mathbf{x}_1})\{\mathbf{x}_j\}$ for all $j = 1, \dots, n$ and note that it satisfies $\beta_j \geq 0$ for $j \geq 2$ and $F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_i) = \sum_{i':\mathbf{x}'_i \preceq \mathbf{x}_i} \beta_{i'} = (\mathbf{A}\boldsymbol{\beta})_i$ for all $i = 1, \dots, n$.

Therefore,

$$\mathcal{D}_{n_1,\dots,n_d} = \big\{(F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_1), \dots, F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_n)) : b \in \mathbb{R},\ \text{finite measure } \mu \text{ on } \Omega_0\big\},$$

Recall that the total variation of a signed measure ν on Ω is defined by $\|\nu\|_{\mathrm{TV}} \coloneqq \nu_+(\Omega) + \nu_-(\Omega)$ where $\nu = \nu_+ - \nu_-$ is the Jordan decomposition of the signed measure. We define the more restricted set

$$\begin{aligned}\mathcal{D}_{n_1,\dots,n_d}(R) \coloneqq \big\{(F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_1), \dots, F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_n)) : b \in \mathbb{R},& \\ \text{finite measure } \mu \text{ on } \Omega_0, \|\mu + b\delta_{\mathbf{x}_1}\|_{\mathrm{TV}} \leq R\big\},&\end{aligned} \tag{A.66}$$

which will be useful in our goal of bounding the metric entropy of $\mathcal{D}_{n_1,\dots,n_d} \cap \mathcal{B}_2(\mathbf{0}, r)$. Note that the total variation term can be written as

$$\|\mu + b\delta_{\mathbf{x}_1}\|_{\mathrm{TV}} = \mu(\Omega_0) + |b|.$$

Let $\boldsymbol{\theta} \coloneqq (F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_1), \dots, F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_n))$ and $\boldsymbol{\theta}' \coloneqq (F_{\mu'+b'\delta_{\mathbf{x}_1}}(\mathbf{x}_1), \dots, F_{\mu'+b'\delta_{\mathbf{x}_1}}(\mathbf{x}_n))$; recall these distribution functions belong to the function class $\mathcal{F}^d_{\mathrm{EM}}$ (Proposition 3.3.2). The Euclidean distance on $\mathcal{D}_{n_1,\dots,n_d}$ is related to the L^2 distance on $\mathcal{F}^d_{\mathrm{EM}}$, as

$$\begin{aligned}&n \int_{[0,1]^d} (F_{\mu+b\delta_{\mathbf{x}_1}} - F_{\mu'+b'\delta_{\mathbf{x}_1}})^2 \, d\lambda \\ &\quad = n \sum_{i=1}^{n} (\theta_i - \theta'_i)^2 \lambda([\mathbf{x}_i, \mathbf{x}_i + (n_1^{-1}, \dots, n_d^{-1})]) = \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2,\end{aligned}$$

where the integral is respect to the Lebesgue measure λ. Note that this equality holds even when $n_j = 1$ for some of the j.

Thus, the ϵ-metric entropy of $\mathcal{D}_{n_1,\dots,n_d}(R)$ (in the Euclidean norm) is bounded by the $\epsilon/\sqrt{n}$-metric entropy of distribution functions of signed measures with total variation norm $\leq R$ (in the L^2 norm). As explained in Blei et al. [12, Sec. 3] (see also Gao [36]), we have:

$$\log N_2(\epsilon, \mathcal{D}_{n_1,\dots,n_d}(R)) \leq C_d \frac{R\sqrt{n}}{\epsilon} \left(\log \frac{R\sqrt{n}}{\epsilon}\right)^{d-\frac{1}{2}} \qquad \text{whenever } \frac{\epsilon}{R\sqrt{n}} < e^{-1} \tag{A.67}$$

for $d > 1$. We remark again that this inequality holds even when $n_j = 1$ for some of the j.

The following inclusions show that $\mathcal{D}_{n_1,\dots,n_d}(R)$ is essentially the same as $\mathcal{D}_{n_1,\dots,n_d} \cap [-R, R]^n$ up to a constant scaling factor.

$$\mathcal{D}_{n_1,\dots,n_d}(R) \subseteq \mathcal{D}_{n_1,\dots,n_d} \cap [-R, R]^n \subseteq \mathcal{D}_{n_1,\dots,n_d}(3R). \tag{A.68}$$

To verify these inclusions, it is useful to recall that for $\boldsymbol{\theta} \coloneqq (F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_1), \ldots, F_{\mu+b\delta_{\mathbf{x}_1}}(\mathbf{x}_n))$ we have $\max_i \theta_i = \theta_n$ and $\min_i \theta_i = \theta_1$, as well as the fact that if $\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d}$ is associated with the pair (μ, b), then $\|\mu + b\delta_{\mathbf{x}_1}\|_{\mathrm{TV}} = \mu(\Omega_0) + |b| = (\theta_n - \theta_1) + |\theta_1|$. The first inclusion follows from the fact that $(\theta_n - \theta_1) + |\theta_1| \le R$ implies $\theta_n \le R$ and $\theta_1 \ge -R$. For the second inclusion, note that $-R \le \theta_1 \le \theta_n \le R$ implies $(\theta_n - \theta_1) + |\theta_1| \le 3R$.

The second inclusion (A.68) immediately yields

$$\log N_2(\epsilon, \mathcal{D}_{n_1,\ldots,n_d} \cap [-R, R]^n) \le C_d \frac{3R\sqrt{n}}{\epsilon} \left(\log \frac{3R\sqrt{n}}{\epsilon} \right)^{d-\frac{1}{2}}, \qquad \forall \epsilon < 3R\sqrt{n}/e.$$

Because $\mathcal{D}_{n_1,\ldots,n_d}$ is translation invariant, we may translate a hyperrectangle of the form $[a, b]^n$ to $[-R, R]^d$ for $R \coloneqq \frac{b-a}{2}$, and obtain

$$\log N_2(\epsilon, \mathcal{D}_{n_1,\ldots,n_d} \cap [a, b]^n) \le C_d \frac{(b-a)\sqrt{n}}{\epsilon} \left(\log \frac{(b-a)\sqrt{n}}{\epsilon} \right)^{d-\frac{1}{2}}$$

for $\epsilon < \frac{3}{2e}(b-a)\sqrt{n}$, where we have absorbed some constants into C_d.

To show that this bound holds under the more general condition $\epsilon \le \sqrt{n}(b-a)$, simply observe that if $\epsilon \ge \sqrt{n}(b-a)/2$, then a single point whose entries are each $(a+b)/2$ covers $\mathcal{D}_{n_1,\ldots,n_d} \cap [a, b]^n$, and so the log covering number is 0, which is bounded by the right-hand side as long as $\epsilon \le (b-a)\sqrt{n}$.

A.5.2 Proof of Lemma A.3.6

The substitution $u = \frac{1}{2} \log \frac{B}{\epsilon}$ and $du = -\frac{1}{2\epsilon}\, d\epsilon$ allows us to rewrite the integral as

$$2^{\frac{2d+3}{4}} B \int_a^\infty e^{-u} u^{\frac{2d-1}{4}}\, du, \qquad \text{with } a \coloneqq \frac{1}{2} \log \frac{B}{s}.$$

It thus suffices to show

$$I(a) \coloneqq \int_a^\infty e^{-u} u^{\frac{2d-1}{4}}\, du \le C_d e^{-a} (a + 1/2)^{\frac{2d-1}{4}}, \qquad \forall a \ge 0. \tag{A.69}$$

If $a \le 1$, then

$$I(a) \le \int_0^\infty e^{-u} u^{\frac{2d-1}{4}}\, du \le C_d e^{-a} 2^{-\frac{2d-1}{4}}$$

for $C_d \ge e 2^{\frac{2d-1}{4}} \int_0^\infty e^{-u} u^{\frac{2d-1}{4}}\, du$, proving the claim (A.69).

Now suppose $a > 1$. Let v be the smallest positive integer strictly larger than $\frac{2d-1}{4}$. Performing integration by parts v times yields

$$\begin{aligned} I(a) &\le C_d e^{-a} \sum_{r=1}^{v} a^{\frac{2d-1}{4} - r + 1} + C_d \int_a^\infty e^{-u} u^{\frac{2d-1}{4} - v}\, du \\ &\le C_d e^{-a} a^{\frac{2d-1}{4}} + C_d e^{-a} \\ &\le (C_d + C_d 2^{\frac{2d-1}{4}}) e^{-a} (a + 1/2)^{\frac{2d-1}{4}}, \end{aligned}$$

which proves the claim (A.69).

A.5.3 Proof of Lemma A.3.7

Suppose $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta} \in \mathcal{C}(V,t)$, where $\mathbf{A}$ is the usual design matrix defined in Section 3.3. Note that for any i we have

$$|\theta_i - \theta_1| = \left| \sum_{j:\mathbf{x}_j \preceq \mathbf{x}_i} \beta_j - \beta_1 \right| \leq \sum_{j=2}^{n} |\beta_j| \leq V.$$

Thus, using the simple inequality $(a+b)^2 \geq \frac{1}{2}a^2 - b^2$ along with the fact that $\|\boldsymbol{\theta}\|^2 \leq t$ we obtain

$$\theta_i^2 = (\theta_1 + \theta_i - \theta_1)^2 \geq \frac{1}{2}\theta_1^2 - (\theta_i - \theta_1)^2 \geq \frac{1}{2}\theta_1^2 - V^2$$

for each i, and thus

$$t^2 \geq \sum_{i=1}^{n} \theta_i^2 \geq \theta_1^2 + (n-1)\left(\frac{1}{2}\theta_1^2 - V^2\right).$$

Rearranging this and applying the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for nonnegative a, b yields

$$|\theta_1| \leq \sqrt{\frac{2}{n+1}(t^2 + (n-1)V^2)} \leq t\sqrt{\frac{2}{n}} + V\sqrt{2} =: \widetilde{t}.$$

We fix $\delta > 0$, whose value will be chosen later. If for an integer k we define

$$\widetilde{\mathcal{C}}_k(V,t) := \{\boldsymbol{\theta} \in \mathcal{C}(V,t) : k\delta \leq \theta_1 \leq (k+1)\delta\},$$

then we have

$$\mathcal{C}(V,t) \subseteq \bigcup_{-K-1 \leq k \leq K} \widetilde{\mathcal{C}}_k(V,t)$$

where $K = \lfloor \widetilde{t}/\delta \rfloor$. Then,

$$\log N(\epsilon, \mathcal{C}(V,t)) \leq \log\left(2 + \frac{\widetilde{t}}{\delta}\right) + \max_{-K-1 \leq k \leq K} \log N(\epsilon, \widetilde{\mathcal{C}}_k(V,t)). \tag{A.70}$$

Since $\widetilde{\mathcal{C}}_{-k-1}(V,t) = -\widetilde{\mathcal{C}}_k(V,t)$ for $k \geq 0$, we may restrict the maximum on the right-hand side to $0 \leq k \leq K$.

Fix $k \geq 0$. If $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta} \in \widetilde{\mathcal{C}}_k(V,t)$, we let $\boldsymbol{\pi}(\boldsymbol{\theta}) := \mathbf{A}\boldsymbol{\beta}^+$ and $\boldsymbol{\nu}(\boldsymbol{\theta}) := \mathbf{A}\boldsymbol{\beta}^-$, where $\beta_j^+ := \max\{\beta_j, 0\}$ and $\beta_j^- := \max\{-\beta_j, 0\}$ so that $\boldsymbol{\theta} = \boldsymbol{\pi}(\boldsymbol{\theta}) - \boldsymbol{\nu}(\boldsymbol{\theta})$. Defining

$$\mathcal{C}_{\boldsymbol{\pi}}(V,t) := \{\boldsymbol{\pi}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \widetilde{\mathcal{C}}_k(V,t)\},$$
$$\mathcal{C}_{\boldsymbol{\nu}}(V,t) := \{\boldsymbol{\nu}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \widetilde{\mathcal{C}}_k(V,t)\},$$

we therefore have

$$\log N(\epsilon, \widetilde{\mathcal{C}}_k(V,t)) \leq \log N(\epsilon/2, \mathcal{C}_{\boldsymbol{\pi}}(V,t)) + \log N(\epsilon/2, \mathcal{C}_{\boldsymbol{\nu}}(V,t)). \tag{A.71}$$

We bound the second term first. Because $\beta_1 = \theta_1 \geq k\delta \geq 0$ (recall the first column of $\mathbf{A}$ is the all-ones vector) for $\boldsymbol{\theta} \in \widetilde{\mathcal{C}}_k(V,t)$, we have $(\boldsymbol{\pi}(\boldsymbol{\theta}))_1 = \beta_1$ and $(\boldsymbol{\nu}(\boldsymbol{\theta}))_1 = 0$. Also,

$$(\boldsymbol{\pi}(\boldsymbol{\theta}))_n - (\boldsymbol{\pi}(\boldsymbol{\theta}))_1 + (\boldsymbol{\nu}(\boldsymbol{\theta}))_n - (\boldsymbol{\nu}(\boldsymbol{\theta}))_1 = \sum_{j=2}^{n} \beta_j \mathbb{I}_{\beta_j \geq 0} - \sum_{j=2}^{n} \beta_j \mathbb{I}_{\beta_j \leq 0} = \sum_{j=2}^{n} |\beta_j| \leq V,$$

which implies $\boldsymbol{\nu}(\boldsymbol{\theta})_n \leq V$ for $\boldsymbol{\theta} \in \mathcal{C}(V,t)$. Since elements of $\mathcal{C}_{\boldsymbol{\nu}}(V,t)$ are of the form $\mathbf{A}\boldsymbol{\beta}$ with $\beta_j \geq 0$ for $j \geq 2$, we use Proposition 3.3.2 and recall a definition (A.66) to obtain the inclusion $\mathcal{C}_{\boldsymbol{\nu}}(V,t) \subseteq \mathcal{D}_{n_1,\ldots,n_d} \cap [0,V]^n$. Thus, using the bound (A.68) along with Lemma A.3.5 we obtain

$$\log N(\epsilon/2, \mathcal{C}_{\boldsymbol{\nu}}(V,t)) \leq C_d \frac{V\sqrt{n}}{\epsilon} \left(\log \frac{2V\sqrt{n}}{\epsilon} \right)^{d-\frac{1}{2}} \mathbb{I}\{\epsilon \leq 2V\sqrt{n}\}$$

where we have absorbed constants into C_d.

We now bound the first term from earlier (A.71). We claim

$$\mathcal{C}_{\boldsymbol{\pi}}(V,t) \subseteq \mathcal{D}_{n_1,\ldots,n_d} \cap [0, V+\delta]^n + \{k\delta\}.$$

To see the last inclusion, note that if $\boldsymbol{\eta} \in \mathcal{C}_{\boldsymbol{\pi}}(V,t)$ satisfies $k\delta \leq \eta_1 \leq (k+1)\delta$ then $\boldsymbol{\eta} - k\delta\mathbf{1}$ lies in $\mathcal{D}_{n_1,\ldots,n_d}$ and has all entries lying in the interval $[0, V+\delta]$ (since $k\delta \leq \eta_1 \leq \eta_i$ and $\eta_i - k\delta \leq \eta_i - (\eta_1 - \delta) \leq V + \delta$). Noting that $\mathcal{D}_{n_1,\ldots,n_d}$ is invariant under translation, we need only compute the metric entropy of $\mathcal{D}_{n_1,\ldots,n_d} \cap [0, V+\delta]^n$. Applying Lemma A.3.5 again yields

$$\log N(\epsilon/2, \mathcal{C}_{\boldsymbol{\pi}}(V,t)) \leq C_d \frac{(V+\delta)\sqrt{n}}{\epsilon} \left(\log \frac{2(V+\delta)\sqrt{n}}{\epsilon} \right)^{d-\frac{1}{2}} \mathbb{I}\{\epsilon \leq 2(V+\delta)\sqrt{n}\}.$$

Choosing $\delta = \epsilon/\sqrt{n}$ yields

$$\log N(\epsilon/2, \mathcal{C}_{\boldsymbol{\pi}}(V,t)) \leq C_d \left(\frac{V\sqrt{n}}{\epsilon} + 1 \right) \left(\log \left(\frac{2V\sqrt{n}}{\epsilon} + 1 \right) \right)^{d-\frac{1}{2}}.$$

Returning to (A.71) we obtain

$$\log N(\epsilon, \widetilde{\mathcal{C}}_k(V,t)) \leq C_d \left(\frac{V\sqrt{n}}{\epsilon} + 1 \right) \left(\log \left(\frac{2V\sqrt{n}}{\epsilon} + 1 \right) \right)^{d-\frac{1}{2}}.$$

Going further back to (A.70) and plugging our definitions of δ and $\widetilde{t}$ yields

$$\log N(\epsilon, \mathcal{C}(V,t)) \leq C_d \left(\frac{V\sqrt{n}}{\epsilon} + 1 \right) \left(\log \left(\frac{2V\sqrt{n}}{\epsilon} + 1 \right) \right)^{d-\frac{1}{2}} + \log \left(2 + 2\frac{t + V\sqrt{n}}{\epsilon} \right).$$

A.5.4 Proof of Lemma A.3.9

Let $r \coloneqq \lfloor 2\ell/d \rfloor$. By an inclusion-exclusion argument, we have the following exact formula for the cardinality.

$$|\mathcal{M}_\ell| = \sum_{k=0}^{d} (-1)^k \binom{d}{k} \binom{\ell - kr - 1}{d-1},$$

with the convention that $\binom{a}{b} = 0$ if $a < b$.

If $k \geq d/2$ then we have $\ell - kr \leq \ell - \frac{d}{2}\left(\frac{2\ell}{d} - 1\right) = \frac{d}{2} < d$ which implies $\binom{\ell-kr-1}{d-1} = 0$.

Otherwise, for $k < d/2$ we have

$$\ell^{-(d-1)} \binom{\ell - kr - 1}{d-1} = \frac{1}{(d-1)!} \prod_{i=1}^{d-1} \frac{\ell - kr - i}{\ell}.$$

Noting that $\lim_{\ell\to\infty} \frac{\ell-kr-i}{\ell} = 1 - k \lim_{\ell\to\infty} \frac{r}{\ell} = 1 - \frac{2k}{d}$, we obtain

$$\lim_{\ell\to\infty} \ell^{-(d-1)} \binom{\ell - kr - 1}{d-1} = \frac{\left(1 - \frac{2k}{d}\right)^{d-1}}{(d-1)!}.$$

for $k < d/2$. Combining these observations for all k yields

$$\lim_{\ell\to\infty} \frac{|\mathcal{M}_\ell|}{\ell^{d-1}} = \sum_{k=0}^{d} (-1)^k \binom{d}{k} \frac{\left(1 - \frac{2k}{d}\right)_+^{d-1}}{(d-1)!} = \frac{d^{d-1}}{(d-1)!} \sum_{k=0}^{d} (-1)^k \binom{d}{k} (d - 2k)_+^{d-1},$$

where $(x)_+ \coloneqq \max\{x, 0\}$. It then suffices to check

$$b_d \coloneqq \sum_{k=0}^{d} (-1)^k \binom{d}{k} (d - 2k)_+^{d-1} > 0$$

for each fixed $d \geq 2$. Indeed, Goddard [39] showed

$$\frac{b_d}{2^d (d-1)!} = \frac{1}{\pi} \int_0^\infty \left(\frac{\sin x}{x}\right)^d dx.$$

When d is even, this clearly positive. When d is odd, we have

$$\int_0^\infty \left(\frac{\sin x}{x}\right)^d dx = \sum_{k=0}^{\infty} \int_{k\pi}^{(k+1)\pi} \left(\frac{\sin x}{x}\right)^d dx = \sum_{k=0}^{\infty} (-1)^k \int_0^\pi \left(\frac{\sin x}{x + k\pi}\right)^d dx,$$

which is positive because the last expression is an alternating sum whose addends' magnitudes $\int_0^\pi \left(\frac{\sin x}{x+k\pi}\right)^d dx$ form a positive decreasing sequence in k.

A.5.5 Proof of Lemma A.3.10

We prove the three inequalities (A.29), (A.30) and (A.31) separately.

Proof of (A.29). For functions $f, g : [0,1]^d \to \mathbb{R}$ we let $\|f\|_2 \coloneqq \left(\int_{[0,1]^d} |f(x)|^2\,dx\right)^{1/2}$ and $\|f\|_1 \coloneqq \int_{[0,1]^d} |f(x)|\,dx$ denote the L^2 and L^1 norms on $[0,1]^d$ with respect to the Lebesgue measure, and $\langle f, g\rangle \coloneqq \int_{[0,1]^d} f(x)g(x)\,dx$ denote the L^2 inner product.

Recall the definition of HK$\mathbf{0}$ variation (3.25) as the sum of Vitali variations over faces adjacent to $\mathbf{0}$. Because $f_{\boldsymbol{\eta}}(\mathbf{x})$ is zero whenever $x_j = 0$ for some j, all these Vitali variations are zero except for the Vitali variation over the entire space $[0,1]^d$. Thus, recalling that the Vitali variation can be written as the integral of the magnitude of a mixed partial derivative (3.23), we have

$$\begin{aligned} V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}; [0,1]^d) = V^{(d)}(f; [0,1]^d) &= \left\| \frac{\partial^d f_{\boldsymbol{\eta}}}{\partial x_1 \cdots \partial x_d} \right\|_1 \\ &\le \left\| \frac{\partial^d f_{\boldsymbol{\eta}}}{\partial x_1 \cdots \partial x_d} \right\|_2 = \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \left\| \sum_{\mathbf{m} \in \mathcal{M}_\ell} g_{\boldsymbol{\eta},\mathbf{m}} \right\|_2, \end{aligned}$$

where

$$g_{\boldsymbol{\eta},\mathbf{m}} \coloneqq \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} \eta_{\mathbf{m},\mathbf{i}} \bigotimes_{j=1}^{d} \phi'_{m_j,i_j}.$$

For natural numbers $m < m'$ and natural numbers $i \le 2^m$ and $i' \le 2^{m'}$, the functions $\phi'_{m,i}$ and $\phi'_{m',i'}$ are orthogonal. Thus for distinct $\mathbf{m}, \mathbf{m}' \in \mathcal{M}_\ell$, the functions $g_{\boldsymbol{\eta},\mathbf{m}}$ and $g_{\boldsymbol{\eta},\mathbf{m}'}$ are orthogonal as well. Thus from above we have

$$V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}; [0,1]^d) \le \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \sqrt{\sum_{\mathbf{m} \in \mathcal{M}_\ell} \|g_{\boldsymbol{\eta},\mathbf{m}}\|_2^2}.$$

For a fixed natural number m and distinct natural numbers $i, i' \le 2^m$, the functions $\phi'_{m,i}$ and $\phi'_{m,i'}$ are also orthogonal because they have different supports. Thus for fixed $\mathbf{m} \in \mathcal{M}$ and distinct $\mathbf{i}, \mathbf{i}' \in \mathcal{I}_{\mathbf{m}}$, the functions $\bigotimes_{j=1}^d \phi'_{m_j,i_j}$ and $\bigotimes_{j=1}^d \phi'_{m_j,i'_j}$ are orthogonal. Continuing from above, we obtain

$$\begin{aligned} V_{\mathrm{HK}\mathbf{0}}(f_{\boldsymbol{\eta}}; [0,1]^d) &\le \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \sqrt{\sum_{\mathbf{m} \in \mathcal{M}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} \left\| \bigotimes_{j=1}^{d} \phi'_{m_j,i_j} \right\|_2^2} \\ &= \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \sqrt{\sum_{\mathbf{m} \in \mathcal{M}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} 2^{-\ell}} = V, \end{aligned}$$

where we used the fact that $|\mathcal{I}_{\mathbf{m}}| = 2^\ell$ and

$$\left\| \bigotimes_{j=1}^{d} \phi'_{m_j,i_j} \right\|_2^2 = \prod_{j=1}^{d} \left\| \phi'_{m_j,i_j} \right\|_2^2 = \prod_{j=1}^{d} 2^{-m_j} = 2^{-\ell}.$$

□

Proof of (A.30). By Pinsker's inequality, we can bound the total variation distance between $\mathbb{P}_{f_{\boldsymbol{\eta}}}$ and $\mathbb{P}_{f_{\boldsymbol{\eta}'}}$ by their Kullback-Leibler divergence.

$$\|\mathbb{P}_{f_{\boldsymbol{\eta}}} - \mathbb{P}_{f_{\boldsymbol{\eta}'}}\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(\mathbb{P}_{f_{\boldsymbol{\eta}}} \| \mathbb{P}_{f_{\boldsymbol{\eta}'}})}.$$

The KL divergence can be computed as

$$D_{\mathrm{KL}}(\mathbb{P}_{f_{\boldsymbol{\eta}}} \| \mathbb{P}_{f_{\boldsymbol{\eta}'}}) = \frac{1}{2\sigma^2} \sum_{i=1}^{n} (f_{\boldsymbol{\eta}}(\mathbf{x}_i) - f_{\boldsymbol{\eta}'}(\mathbf{x}_i))^2 = \frac{n}{2\sigma^2} \mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'}),$$

where $\mathcal{L}$ denotes the discrete loss as defined earlier (3.9).

Note that

$$f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}'} = \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \sum_{\mathbf{m} \in \mathcal{M}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} (\eta_{\mathbf{m},\mathbf{i}} - \eta'_{\mathbf{m},\mathbf{i}}) \bigotimes_{j=1}^{d} \phi_{m_j,i_j}.$$

If $d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}') = 1$, then there exists a unique pair $\mathbf{m} \in \mathcal{M}_\ell$ and $\mathbf{i} \in \mathcal{I}_{\mathbf{m}}$ such that $\eta_{\mathbf{m},\mathbf{i}} \neq \eta'_{\mathbf{m},\mathbf{i}}$. Then

$$f_{\boldsymbol{\eta}} - f_{\boldsymbol{\eta}'} = \frac{V}{\sqrt{|\mathcal{M}_\ell|}} (\eta_{\mathbf{m},\mathbf{i}} - \eta'_{\mathbf{m},\mathbf{i}}) \bigotimes_{j=1}^{d} \phi_{m_j,i_j}.$$

Thus, recalling that the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ come from the lattice $\mathbb{L}_{n_1,\ldots,n_d}$ (see (3.34))

$$\begin{aligned} \mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'}) &= \frac{4V^2}{n|\mathcal{M}_\ell|} \sum_{k_1=0}^{n_1-1} \cdots \sum_{k_d=0}^{n_d-1} \prod_{j=1}^{d} (\phi_{m_j,i_j}(k_j/n_j))^2 \\ &= \frac{4V^2}{|\mathcal{M}_\ell|} \prod_{j=1}^{d} \left(\frac{1}{n_j} \sum_{k_j=0}^{n_j-1} (\phi_{m_j,i_j}(k_j/n_j))^2 \right). \end{aligned}$$

Note that for each j, the number of nonzero addends $(\phi_{m_j,i_j}(k_j/n_j))^2$ (of the above inner sum) is bounded by $n_j 2^{-m_j}$, so we obtain

$$\frac{1}{n_j} \sum_{k_j=0}^{n_j-1} (\phi_{m_j,i_j}(k_j/n_j))^2 \le \frac{1}{n_j} \cdot n_j 2^{-m_j} \cdot 2^{-2m_j-4} = 2^{-3m_j-4}.$$

Multiplying over all j yields

$$\prod_{j=1}^{d}\left(\frac{1}{n_j}\sum_{k_j=0}^{n_j-1}(\phi_{m_j,i_j}(k_j/n_j))^2\right)=\prod_{j=1}^{d}2^{-3m_j-4}=2^{-3\ell-4d}.$$

By combining our work above, we obtain

$$\max_{d_{\mathrm{H}}(\boldsymbol{\eta},\boldsymbol{\eta}')=1}\|\mathbb{P}_{f_{\boldsymbol{\eta}}}-\mathbb{P}_{f_{\boldsymbol{\eta}'}}\|_{\mathrm{TV}}\le\sqrt{\frac{n}{4\sigma^2}\mathcal{L}(f_{\boldsymbol{\eta}},f_{\boldsymbol{\eta}'})}\le\sqrt{\frac{n}{\sigma^2}\frac{V^2}{|\mathcal{M}_\ell|}2^{-3\ell-4d}}.$$

□

Proof of (A.31). To compute the loss $\mathcal{L}(f_{\boldsymbol{\eta}},f_{\boldsymbol{\eta}'})$ for some $\boldsymbol{\eta},\boldsymbol{\eta}'\in\{-1,1\}^q$, only the values of ϕ_{m_j,i_j} at points $\{k/n_j:k\in\{0,\dots,n_j-1\}\}$ matter. In particular, for each fixed $j\in[d]$ and $m_j\in\mathbb{N}$ and $i_j\in[2^{m_j}]$ we define the step function $\widetilde{\phi}:[0,1]\to\mathbb{R}$ by

$$\widetilde{\phi}_{j,m_j,i_j}(x)\coloneqq\phi_{m_j,i_j}(\lfloor xn_j\rfloor/n_j). \tag{A.72}$$

Recall our assumption that n_j is a power of 2. Thus function $\widetilde{\phi}_{j,m_j,i_j}$ is a step function supported on $[(i_j-1)2^{-m_j},i_j2^{-m_j}]$ that is constant on intervals $[k/n_j,(k+1)/n_j)$ for $k=0,\dots,n_j-1$, and agrees with the value of ϕ_{m_j,i_j} at points k/n_j for $k=0,\dots,n_j-1$.

If for $\boldsymbol{\eta},\boldsymbol{\eta}'\in\{-1,1\}^q$ we define

$$g_{\boldsymbol{\eta},\boldsymbol{\eta}'}\coloneqq\frac{V}{\sqrt{|\mathcal{M}_\ell|}}\sum_{\mathbf{m}\in\mathcal{M}_\ell}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{m}}}(\eta_{\mathbf{m},\mathbf{i}}-\eta'_{\mathbf{m},\mathbf{i}})\bigotimes_{j=1}^{d}\widetilde{\phi}_{m_j,i_j},$$

then $g_{\boldsymbol{\eta},\boldsymbol{\eta}'}$ agrees with $f_{\boldsymbol{\eta}}-f_{\boldsymbol{\eta}'}$ on points of the form $(k_1/n_1,\dots,k_d/n_d)$ for $k_j=0,\dots,n_j$ and all $j\in[d]$, and is constant on rectangles of the form $\bigtimes_{j=1}^{d}[k_j/n_j,(k_j+1)/n_j)$. Therefore,

$$\mathcal{L}(f_{\boldsymbol{\eta}},f_{\boldsymbol{\eta}'})\coloneqq\frac{1}{n}\sum_{i=1}^{n}(f_{\boldsymbol{\eta}}(\mathbf{x}_i)-f_{\boldsymbol{\eta}'}(\mathbf{x}_i))^2=\int_{[0,1]^d}(g_{\boldsymbol{\eta},\boldsymbol{\eta}'}(\mathbf{x}))^2\,d\mathbf{x}=\|g_{\boldsymbol{\eta},\boldsymbol{\eta}'}\|_{L^2}^2.$$

For natural number m and $i\in[2^m]$, we define the function $h_{m,i}:[0,1]\to\mathbb{R}$ by

$$h_{m,i}(x)=\begin{cases}2^{m/2} & x\in[(i-1)2^{-m},(i-\frac{1}{2})2^{-m}],\\ -2^{m/2} & x\in[(i-\frac{1}{2})2^{-m},i2^{-m}],\\ 0 & \text{otherwise.}\end{cases} \tag{A.73}$$

One can check $\{h_{m,i}:m\in\mathbb{N},i\in[2^m]\}$ is an orthonormal set. If we define $H_{\mathbf{m},\mathbf{i}}\coloneqq\bigotimes_{j=1}^{d}h_{m_j,i_j}$, then $\{H_{\mathbf{m},\mathbf{i}}:\mathbf{m}\in\mathcal{M}_\ell,\mathbf{i}\in\mathcal{I}_{\mathbf{m}}\}$ is an orthonormal set of functions on $[0,1]^d$.

Thus by Bessel's inequality,

$$\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'}) = \|g_{\boldsymbol{\eta},\boldsymbol{\eta}'}\|_2^2 \geq \sum_{\mathbf{m}'\in\mathcal{M}_\ell} \sum_{\mathbf{i}'\in\mathcal{I}_{\mathbf{m}}} \left\langle g_{\boldsymbol{\eta},\boldsymbol{\eta}'}, H_{\mathbf{m}',\mathbf{i}'}\right\rangle^2.$$

Fix $\mathbf{m}' \in \mathcal{M}_\ell$ and $\mathbf{i}' \in \mathcal{I}_{\mathbf{m}}$. We have

$$\left\langle g_{\boldsymbol{\eta},\boldsymbol{\eta}'}, H_{\mathbf{m}',\mathbf{i}'}\right\rangle = \frac{V}{\sqrt{|\mathcal{M}_\ell|}} \sum_{\mathbf{m}\in\mathcal{M}_\ell} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{m}}} (\eta_{\mathbf{m},\mathbf{i}} - \eta'_{\mathbf{m},\mathbf{i}}) \prod_{j=1}^d \langle \widetilde{\phi}_{j,m_j,i_j}, h_{m'_j,i'_j}\rangle.$$

We claim the inner products satisfy

$$\langle \widetilde{\phi}_{j,m_j,i_j}, h_{m'_j,i'_j}\rangle = \begin{cases} 0 & (m_j,i_j) \neq (m'_j,i'_j), m_j \leq m'_j \\ 0 & \log_2(n_j) \leq m_j + 1 \\ 2^{-3m_j/2-3} & (m_j,i_j) = (m'_j,i'_j), \log_2(n_j) \geq m_j + 2 \end{cases}$$

For the first case, if $m_j = m'_j$ and $i_j \neq i'_j$, then the supports of $\widetilde{\phi}_{j,m_j,i_j}$ and $h_{m'_j,i'_j}$ are disjoint so their inner product is zero. If instead $m_j < m'_j$, then recall $\int_0^1 \widetilde{\phi}_{j,m_j,i_j}(x)\,dx = 0$ and note that $h_{m'_j,i'_j}$ is constant on the support $[(i_j-1)2^{-m_j}, i_j 2^{-m_j}]$ of $\widetilde{\phi}_{j,m_j,i_j}$. The second case is due to the fact that n_j is a power of 2 and consequently $\widetilde{\phi}_{j,m_j,i_j} \equiv 0$ when $\log_2 n_j \leq m_j + 1$. For the third case where $(m_j,i_j) = (m'_j,i'_j)$ and $\log_2(n_j) \geq m_j + 2$, we have

$$\begin{aligned} \langle \widetilde{\phi}_{j,m_j,i_j}, h_{m_j,i_j}\rangle &= 2\cdot 2^{m_j/2} \int_{(i_j-1)2^{-m_j}}^{(i_j-\frac{1}{2})2^{-m_j}} \widetilde{\phi}_{j,m_j,i_j}(x)\,dx \\ &= 2\cdot 2^{m_j/2} \int_{(i_j-1)2^{-m_j}}^{(i_j-\frac{1}{2})2^{-m_j}} \phi_{m_j,i_j}(x)\,dx \\ &= 2^{m_j/2}\cdot 2^{-m_j-1}2^{-m_j-2} = 2^{-3m_j/2-3}, \end{aligned}$$

where the equality of integrals is a consequence of $\log_2(n_j) \geq m_j + 2$ and the fact that n_j is a power of 2.

If $\mathbf{m}$ and $\mathbf{m}'$ both belong to $\mathcal{M}$, they satisfy $\sum_{i=1}^d m_i = \sum_{i=1}^d m'_i = \ell$. Thus if $\mathbf{m} \neq \mathbf{m}'$, then because $d \geq 2$ there is some j for which $m_j < m'_j$, and we obtain

$$\prod_{j=1}^d \langle \widetilde{\phi}_{m_j,i_j}, h_{m'_j,i'_j}\rangle = \begin{cases} 0 & (\mathbf{m},\mathbf{i}) \neq (\mathbf{m}',\mathbf{i}') \\ 2^{-3\ell/2-3d} & (\mathbf{m},\mathbf{i}) = (\mathbf{m}',\mathbf{i}'),\ m_j + 2 \leq \log_2 n_j \forall j \end{cases} \tag{A.74}$$

Thus,

$$\left\langle g_{\boldsymbol{\eta},\boldsymbol{\eta}'}, H_{\mathbf{m}',\mathbf{i}'}\right\rangle = \frac{V}{\sqrt{|\mathcal{M}_\ell|}} (\eta_{\mathbf{m}',\mathbf{i}'} - \eta'_{\mathbf{m}',\mathbf{i}'}) 2^{-3\ell/2-3d} \prod_{j=1}^d \mathbb{I}\{m'_j + 2 \leq \log_2 n_j\}.$$

If we show that the above product of indicators is always equal to 1, then plugging this into Bessel's inequality above yields

$$\mathcal{L}(f_{\boldsymbol{\eta}}, f_{\boldsymbol{\eta}'}) \geq \frac{4V^2}{|\mathcal{M}_\ell|} 2^{-3\ell-6d} d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}')$$

which would complete the proof of the desired claim (A.31).

It remains to show this last unverified claim about the product of indicators. Equivalently, if $\mathbf{m} \in \mathcal{M}_\ell$, then we want to show $m_j + 2 \leq \log_2 n_j$ for all $j \in [d]$, provided n is large enough. Since $n_j = n^{1/d}$ and since $\max_{j\in[d]} m_j \leq 2\ell/d$, it suffices to show

$$\frac{2\ell}{d} + 2 \leq \frac{1}{d} \log n \tag{A.75}$$

Plugging in the definition (A.26) of ℓ yields

$$\frac{2}{3d \log 2} \log(C_d n V^2/\sigma^2) - \frac{2(d-1)}{3d \log 2} \log\log(C_d n V^2/\sigma^2) + 2 \leq \frac{1}{d} \log n.$$

For fixed d and σ^2/V^2, we have

$$\begin{aligned} &\lim_{n\to\infty} \frac{d}{\log n} \left[\frac{2}{3d \log 2} \log(C_d n V^2/\sigma^2) - \frac{2(d-1)}{3d \log 2} \log\log(C_d n V^2/\sigma^2) + 2 \right] \\ &= \frac{2}{3 \log 2} < 1, \end{aligned}$$

so there exists a constant $c_{d,\sigma^2/V^2}$ such that the bound (A.75) holds if $n \geq c_{d,\sigma^2/V^2}$. □

A.5.6 Proof of Lemma A.3.11

We claim the functions $F_{\boldsymbol{\eta},\mathbf{m}}$ and $F_{\boldsymbol{\eta},\mathbf{m}'}$ are orthogonal for distinct $\mathbf{m}, \mathbf{m}' \in \widetilde{\mathcal{M}}_\ell$. We have

$$\begin{aligned} &\int_0^1 \int_0^1 F_{\boldsymbol{\eta},\mathbf{m}}(t_1, t_2) F_{\boldsymbol{\eta},\mathbf{m}'}(t_1, t_2)\, dt_1\, dt_2 \\ &= \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{m}}} \sum_{\mathbf{i}'\in\mathcal{I}_{\mathbf{m}'}} \boldsymbol{\eta}_{\mathbf{m},\mathbf{i}} \boldsymbol{\eta}_{\mathbf{m}',\mathbf{i}'} \int_0^1 \phi'_{m_1,i_1}(t_1) \phi'_{m'_1,i'_1}(t_1)\, dt_1 \int_0^1 \phi'_{m_2,i_2}(t_2) \phi'_{m'_2,i'_2}(t_2)\, dt_2 \end{aligned}$$

Fix $(\mathbf{m}, \mathbf{i})$ and $(\mathbf{m}', \mathbf{i}')$. Since $\mathbf{m}$ and $\mathbf{m}'$ are distinct, we must have $m_1 \neq m'_1$ (since $m_1 + m_2 = m'_1 + m'_2$). Without loss of generality suppose $m_1 < m'_1$. Then ϕ'_{m_1,i_1} is constant on the support of $\phi'_{m'_1,i'_1}$ for any $i_1 \in [2^{m_1}]$ and $i'_1 \in [2^{m'_1}]$, and thus $\int_0^1 \phi'_{m_1,i_1}(t_1) \phi'_{m'_1,i'_1}(t_1)\, dt_1 = 0$. The other case $m_1 > m'_1$ can be handled similarly. In the end all terms in the above double sum are zero.

A similar argument shows that the integral of the product of $F_{\boldsymbol{\eta},\mathbf{m}^{(1)}}, \ldots, F_{\boldsymbol{\eta},\mathbf{m}^{(k)}}$ for distinct $\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(k)}$ is zero, since $m_1^{(1)}, \ldots, m_1^{(k)}$ are distinct in this case where $d = 2$.

Note also that $1+F_{\boldsymbol{\eta},\mathbf{m}} \geq 0$ for all $\mathbf{m} \in \widetilde{\mathcal{M}}_\ell$, and thus $\partial^2 \widetilde{f}_{\boldsymbol{\eta}}/(\partial x_1 \partial x_2) \geq 1$. Consequently,

$$\begin{aligned} V_{\mathrm{HK}\mathbf{0}}(\widetilde{f}_{\boldsymbol{\eta}}) &= \left\| \frac{\partial^2 \widetilde{f}_{\boldsymbol{\eta}}}{\partial x_1 \partial x_2} \right\|_1 = \left\| \prod_{\mathbf{m} \in \widetilde{\mathcal{M}}_\ell} (1 + F_{\boldsymbol{\eta},\mathbf{m}}(x_1, x_2)) \right\|_1 \\ &= \int_0^1 \int_0^1 \prod_{\mathbf{m} \in \widetilde{\mathcal{M}}_\ell} (1 + F_{\boldsymbol{\eta},\mathbf{m}}(t_1, t_2))\, dt_1\, dt_2 \\ &= 1 + \sum_{\mathbf{m} \in \widetilde{\mathcal{M}}_\ell} \int_0^1 \int_0^1 F_{\boldsymbol{\eta},\mathbf{m}}(t_1, t_2)\, dt_1\, dt_2 + 0 \\ &= 1. \end{aligned}$$

Combined with the fact that $\widetilde{f}_{\boldsymbol{\eta}}$ is continuous, we have $\widetilde{f}_{\boldsymbol{\eta}} \in \mathcal{F}^2_{\mathrm{DF}}$.

We now define $\widetilde{g}_{\boldsymbol{\eta}}(x_1, x_2) \coloneqq \widetilde{f}_{\boldsymbol{\eta}}(\lfloor x_1 n_1 \rfloor / n_1, \lfloor x_2 n_2 \rfloor / n_2)$. This function agrees with $\widetilde{f}_{\boldsymbol{\eta}}$ at the design points $(i/n_1, j/n_2)$, and is piecewise constant on rectangles of the grid. Thus for $\boldsymbol{\eta} \neq \boldsymbol{\eta}'$ we have

$$\mathcal{L}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'}) = \|\widetilde{g}_{\boldsymbol{\eta}} - \widetilde{g}_{\boldsymbol{\eta}'}\|^2_{L^2}.$$

Let us similarly define $\widetilde{Q}_{\boldsymbol{\eta}}(x_1, x_2) \coloneqq \widetilde{Q}_{\boldsymbol{\eta}}(\lfloor x_1 n_1 \rfloor / n_1, \lfloor x_2 n_2 \rfloor / n_2)$. Let $h_{m,r}$ be as defined above (A.73). We now show $\langle \widetilde{Q}_{\boldsymbol{\eta}}, h_{m_1,r_1} \otimes h_{m_2,r_2} \rangle$ for all $\mathbf{m} \in \widetilde{\mathcal{M}}_\ell$ and $\mathbf{r} \in \mathcal{I}_{\mathbf{m}}$. Note

$$\widetilde{Q}_{\boldsymbol{\eta}}(\mathbf{x}) = \sum_{P \geq 2} \sum \int_0^{\lfloor x_1 n_1 \rfloor / n_1} \prod_{p=1}^P \phi'_{k_p,i_p}(t_1)\, dt_1 \int_0^{\lfloor x_2 n_2 \rfloor / n_2} \prod_{p=1}^P \phi'_{\ell - k_p, j_p}(t_2)\, dt_2$$

where the inner sum above is over even integers $0 \leq k_1 < k_2 < \cdots < k_P \leq \ell$, and all $1 \leq i_p \leq 2^{k_p}$, $1 \leq j_p \leq 2^{\ell - k_p}$, $1 \leq p \leq P$.

Because $\phi'_{k_1,i_1}, \ldots, \phi'_{k_{P-1},i_{P-1}}$ are constant on the support of ϕ'_{k_P,i_P} we have for some constant c_1

$$\int_0^{\lfloor x_1 n_1 \rfloor / n_1} \prod_{p=1}^P \phi'_{k_p,i_p}(t_1)\, dt_1 = c_1 \phi_{k_P,i_P}(\lfloor x_1 n_1 \rfloor / n_1) =: c_1 \widetilde{\phi}_{1,k_P,i_P}(x_1),$$

where the last equality is due to the earlier definition (A.72). Similarly,

$$\int_0^{\lfloor x_2 n_2 \rfloor / n_2} \prod_{p=1}^P \phi'_{\ell - k_p, j_p}(t_2)\, dt_2 = c_2 \phi_{\ell - k_1, j_1}(\lfloor x_2 n_2 \rfloor / n_2) =: c_2 \widetilde{\phi}_{2,k_1,j_1}(x_2),$$

Because $k_P + (\ell - k_1) > \ell = m_1 + m_2$, we must have either $k_P > m_1$ or $\ell - k_1 > m_2$. If $k_P > m_1$, then for any $1 \leq r_1 \leq 2^{m_1}$, h_{m_1,r_1} is constant on the support of $\widetilde{\phi}_{k_P,i_P}$, and thus

$$\int_0^1 \widetilde{\phi}_{k_P,i_P}(x_1) h_{m_1,r_1}(x_1)\, dx_1 = c' \int_0^1 \widetilde{\phi}_{k_P,i_P}(x_1)\, dx_1 = 0.$$

Otherwise, if $\ell - k_1 > m_2$, then $\int_0^1 \widetilde{\phi}_{\ell-k_1,j_1}(x_2) h_{m_2,r_2}(x_2)\,dx_2 = 0$ for all $1 \le r_2 \le 2^{m_2}$. In either case we have $\langle \widetilde{\phi}_{k_P,i_P} \otimes \widetilde{\phi}_{\ell-k_1,j_1}, h_{m_1,r_1} \otimes h_{m_2,r_2} \rangle = 0$, and thus $\langle \widetilde{Q}_{\boldsymbol{\eta}}, h_{m_1,r_1} \otimes h_{m_2,r_2} \rangle = 0$ for all $\mathbf{m} \in \widetilde{\mathcal{M}}_\ell$ and $\mathbf{r} \in \mathcal{I}_{\mathbf{m}}$. Therefore, using the earlier observation (A.74) concerning inner products between $\widetilde{\phi}_{m,i}$ and $h_{m',i'}$, we obtain

$$\begin{aligned}
&\langle \widetilde{g}_{\boldsymbol{\eta}} - \widetilde{g}_{\boldsymbol{\eta}'}, h_{m_1',i_1'} \otimes h_{m_2',i_2'} \rangle \\
&= \sum_{\mathbf{m} \in \widetilde{\mathcal{M}}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} (\eta_{\mathbf{m},\mathbf{i}} - \eta'_{\mathbf{m},\mathbf{i}}) \langle \widetilde{\phi}_{m_1,i_1} \otimes \widetilde{\phi}_{m_2,i_2}, h_{m_1',i_1'} \otimes h_{m_2',i_2'} \rangle \\
&= (\eta_{\mathbf{m}',\mathbf{i}'} - \eta'_{\mathbf{m}',\mathbf{i}'}) 2^{-3\ell/2-6} \mathbb{I}\{m_1' + 2 \le \log_2 n_1, m_2' + 2 \le \log_2 n_2\}.
\end{aligned}$$

As argued before (A.75), the event in the indicator function holds for sufficiently large n, so we may ignore it. Applying Bessel's inequality yields

$$\mathcal{L}(\widetilde{f}_{\boldsymbol{\eta}}, \widetilde{f}_{\boldsymbol{\eta}'}) = \|\widetilde{g}_{\boldsymbol{\eta}} - \widetilde{g}_{\boldsymbol{\eta}'}\|_{L^2}^2 \ge \sum_{\mathbf{m} \in \widetilde{\mathcal{M}}_\ell} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} (\eta_{\mathbf{m},\mathbf{i}} - \eta'_{\mathbf{m},\mathbf{i}})^2 2^{-3\ell-12} = d_{\mathrm{H}}(\boldsymbol{\eta}, \boldsymbol{\eta}') 2^{-3\ell-10}.$$

A.5.7 Proof of Lemma A.3.12

If $\sum_{j=2}^n |\widetilde{\beta}_j| < R$, then $\mathbf{A}\boldsymbol{\beta}$ lies in the interior of $\mathcal{C}(V)$, so the tangent cone there is $\mathbb{R}^n$. Thus it remains to consider the case $\sum_{j=2}^n |\widetilde{\beta}_j| = R$.

Let $\mathcal{T}$ denote the right-hand side of the equality (A.41). We first show $\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}}) \subseteq \mathcal{T}$. Since $\mathcal{T}$ is a closed convex cone, it suffices to show that $\mathbf{A}\boldsymbol{\beta} := \mathbf{A}(\boldsymbol{\beta}' - \widetilde{\boldsymbol{\beta}})$ lies in $\mathcal{T}$ for any $\mathbf{A}\boldsymbol{\beta}' \in \mathcal{C}(V)$. Indeed, using the fact that $\beta_j = \beta_j'$ whenever $\widetilde{\beta}_j = 0$, we have

$$\begin{aligned}
\sum_{\substack{j \ge 2: \\ \widetilde{\beta}_j = 0}} |\beta_j| + \sum_{\substack{j \ge 2: \\ \widetilde{\beta}_j \neq 0}} \beta_j' \operatorname{sign}(\widetilde{\beta}_j) &= \sum_{\substack{j \ge 2: \\ \beta_j = 0}} |\beta_j'| + \sum_{\substack{j \ge 2: \\ \beta_j \neq 0}} \beta_j' \operatorname{sign}(\beta_j) \\
&\le \sum_{j=2}^n |\beta_j'| \le V = \sum_{j=2}^n \widetilde{\beta}_j \operatorname{sign}(\widetilde{\beta}_j).
\end{aligned}$$

Some rearrangement leads to $\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}}) \subseteq \mathcal{T}$.

For the reverse inclusion, suppose $\mathbf{A}\boldsymbol{\beta} \in \mathcal{T}$. We claim that there exists some $c > 0$ such that $\mathbf{A}\widetilde{\boldsymbol{\beta}} + c\mathbf{A}\boldsymbol{\beta} \in \mathcal{C}(V)$. Indeed, there exists a sufficiently small $c > 0$ such that

$\operatorname{sign}(\widetilde{\beta}_j + c\beta_j) = \operatorname{sign}(\widetilde{\beta}_j)$ for all j satisfying $\widetilde{\beta}_j \neq 0$, for which we have

$$\begin{aligned}\sum_{j=2}^{n}|\widetilde{\beta}_j + c\beta_j| &= c\sum_{\substack{j\geq 2:\\ \widetilde{\beta}_j=0}}|\beta_j| + \sum_{\substack{j\geq 2:\\ \widetilde{\beta}_j=0}}(\widetilde{\beta}_j + c\beta_j)\operatorname{sign}(\widetilde{\beta}_j)\\ &= \sum_{j=2}^{n}|\widetilde{\beta}_j| + c\underbrace{\left(\sum_{\substack{j\geq 2:\\ \widetilde{\beta}_j=0}}|\beta_j| + \sum_{\substack{j\geq 2:\\ \widetilde{\beta}_j\neq 0}}\beta_j\operatorname{sign}(\widetilde{\beta}_j)\right)}_{\leq 0}\\ &\leq V,\end{aligned}$$

where the quantity in parentheses is nonpositive due to the definition of $\mathbf{A}\boldsymbol{\beta} \in \mathcal{T}$. The above implies $\mathbf{A}\widetilde{\boldsymbol{\beta}} + c\mathbf{A}\boldsymbol{\beta} \in \mathcal{C}(V)$, concluding the proof.

A.5.8 Proof of Lemma A.3.13

Using the fact that $\sum_{\mathbf{i}':\mathbf{i}'\preceq\mathbf{i}}\beta_{\mathbf{i}'} = \alpha_{\mathbf{i}}$ we have

$$\begin{aligned}\operatorname{sign}(\widetilde{\beta}_{\mathbf{i}^*})(\alpha_{\mathbf{i}^u} - \alpha_{\mathbf{i}^\ell}) &= \operatorname{sign}(\widetilde{\beta}_{\mathbf{i}^*})\left(\sum_{\mathbf{i}\in L_u}\beta_{\mathbf{i}} - \sum_{\mathbf{i}\in L_\ell}\beta_{\mathbf{i}}\right)\\ &= \operatorname{sign}(\widetilde{\beta}_{\mathbf{i}^*})\left(\sum_{\mathbf{i}\in L_u\cap L_\ell^c}\beta_{\mathbf{i}} - \sum_{\mathbf{i}\in L_u^c\cap L_\ell}\beta_{\mathbf{i}}\right)\\ &= \operatorname{sign}(\widetilde{\beta}_{\mathbf{i}^*})\left(\beta_{\mathbf{i}^*} + \sum_{\mathbf{i}\in(L_u\cap L_\ell^c)\setminus\{\mathbf{i}^*\}}\beta_{\mathbf{i}} - \sum_{\mathbf{i}\in L_u^c\cap L_\ell}\beta_{\mathbf{i}}\right) \qquad \text{(A.76)}\\ &\leq \beta_{\mathbf{i}^*}\operatorname{sign}(\widetilde{\beta}_{\mathbf{i}^*}) + \sum_{\mathbf{i}\notin\{\mathbf{0},\mathbf{i}^*\}}|\beta_{\mathbf{i}}| \qquad \text{(A.77)}\\ &\leq 0,\end{aligned}$$

where the last inequality is due to the characterization (A.41) of $\mathcal{T}_{\mathcal{C}(V)}(\mathbf{A}\widetilde{\boldsymbol{\beta}})$. The above chain of inequalities implies that the difference between the expressions (A.77) and (A.76) is bounded by $-\operatorname{sign}(\widetilde{\beta}_{\mathbf{i}^*})(\alpha_{\mathbf{i}^u} - \alpha_{\mathbf{i}^\ell})$. This is precisely the desired inequality (A.42).

A.5.9 Statement and proof of a result connecting $D(\boldsymbol{\theta}_Q)$ and $D\boldsymbol{\theta}$

Lemma A.5.1. *Consider Q as in* (A.48) *for two indices $\mathbf{q}^\ell$ and $\mathbf{q}^u$ in $\mathcal{I}$ with $\mathbf{q}^\ell \preceq \mathbf{q}^u$. Recall the notation* (A.50) *and* (A.51)*. For every $\boldsymbol{\theta} \in \mathbb{R}^n$, we have*

$$(D\boldsymbol{\theta}_Q)_{\mathbf{i}} = \sum_{\mathbf{i}'\preceq\mathbf{i}} t(\mathbf{i}', \mathbf{i})(D\boldsymbol{\theta})_{\mathbf{i}}, \qquad \textit{for every } \mathbf{i} \in Q \qquad \text{(A.78)}$$

Furthermore, for every $\mathbf{i}' \preceq \mathbf{q}^u$*, there is a unique* $\mathbf{i} \in Q$ *such that* $\mathbf{i} \succeq \mathbf{i}'$ *and* $\mathbf{i}'_{J(\mathbf{i})} = \mathbf{i}_{J(\mathbf{i})}$*; this* $\mathbf{i}$ *is given by* $i_j := \max\{q_j^\ell, i'_j\}, j = 1, \ldots, d$.

Proof. For $\mathbf{i} \in Q$, the identities (A.49) and (A.47) together yield

$$\begin{aligned}(D\boldsymbol{\theta}_Q)_{\mathbf{i}} &= \sum_{\mathbf{z}\in\{0,1\}^d} \mathbb{I}\{\mathbf{i}-\mathbf{z} \succeq \mathbf{q}^\ell\}(-1)^{z_1+\cdots+z_d}\theta_{\mathbf{i}-\mathbf{z}} \\ &= \sum_{\mathbf{z}\in\{0,1\}^d} \mathbb{I}\{\mathbf{i}-\mathbf{z} \succeq \mathbf{q}^\ell\}(-1)^{z_1+\cdots+z_d} \sum_{\mathbf{i}'\preceq\mathbf{i}-\mathbf{z}} (D\boldsymbol{\theta})_{\mathbf{i}'} \\ &= \sum_{\mathbf{i}'\preceq\mathbf{i}} (D\boldsymbol{\theta})_{\mathbf{i}'} \sum_{\mathbf{z}\in\{0,1\}^d} \mathbb{I}\{\mathbf{i}-\mathbf{z} \succeq \mathbf{q}^\ell\}(-1)^{z_1+\cdots+z_d}\mathbb{I}\{\mathbf{i}'\preceq\mathbf{i}-\mathbf{z}\}.\end{aligned}$$

It then remains to show that the last inner sum equals $t(\mathbf{i}', \mathbf{i})$. We have

$$\begin{aligned}&\sum_{\mathbf{z}\in\{0,1\}^d} \mathbb{I}\{\mathbf{i}-\mathbf{z} \succeq \mathbf{q}^\ell\}(-1)^{z_1+\cdots+z_d}\mathbb{I}\{\mathbf{i}'\preceq\mathbf{i}-\mathbf{z}\} \\ &= \prod_{j=1}^d \sum_{z_j=0}^1 (-1)^{z_j}\mathbb{I}\{i_j - z_j \ge q_j^\ell\}\mathbb{I}\{i'_j \le i_j - z_j\} \\ &= \prod_{j=1}^d \big(\mathbb{I}\{q_j^\ell \le i_j; i'_j \le i_j\} - \mathbb{I}\{q_j^\ell \le i_j - 1; i'_j \le i_j - 1\}\big).\end{aligned}$$

For $j \in J(\mathbf{i})$ we have $i_j > q_j^\ell$, so the quantity in parentheses is $\mathbb{I}\{i'_j \le i_j\} - \mathbb{I}\{i'_j \le i_j - 1\} = \mathbb{I}\{i'_j = i_j\}$. For $j \notin J(\mathbf{i})$ we have $i_j = q_j^\ell$, so the quantity in parentheses is 1. Thus the above product is $\mathbb{I}\{\mathbf{i}'_{J(\mathbf{i})} = \mathbf{i}_{J(\mathbf{i})}\}$, and we obtain (A.78).

We now prove the second claim of the lemma. Fix $\mathbf{i}' \preceq \mathbf{q}^u$. We would like to produce $\mathbf{i} \in Q$ such that $\mathbf{i} \succeq \mathbf{i}'$ and $i_j = i_j$ for $j \in J(\mathbf{i}) = \{j : i_j > q_j^\ell\}$. If $i'_j > q_j^\ell$, we have no choice but to let $i_j = i'_j$. If $i'_j \le q_j^\ell$, we must let $i_j = q_j^\ell$ in order to have $\mathbf{i} \in Q$. This defines the unique $\mathbf{i}$ satisfying the conditions. □

A.5.10 Proof of Lemma A.3.14

Fix $\mathbf{i}' \preceq \mathbf{i}$ such that $t(\mathbf{i}', \mathbf{i}) \neq 0$. If $s(\mathbf{i}') \neq 0$, then $\widetilde{\mathfrak{s}}(\mathbf{i}) := \mathfrak{s}(\mathbf{i}')$ so we have

$$|\beta_{\mathbf{i}'}| - \widetilde{\mathfrak{s}}(\mathbf{i})t(\mathbf{i}', \mathbf{i})\beta_{\mathbf{i}'} = |\beta_{\mathbf{i}'}| - \mathfrak{s}(\mathbf{i}')\beta_{\mathbf{i}'}.$$

Otherwise if $s(\mathbf{i}') = 0$, we have

$$|\beta_{\mathbf{i}'}| - \widetilde{\mathfrak{s}}(\mathbf{i})t(\mathbf{i}', \mathbf{i})\beta_{\mathbf{i}'} \le 2|\beta_{\mathbf{i}'}| = 2(|\beta_{\mathbf{i}'}| - \mathfrak{s}(\mathbf{i}')\beta_{\mathbf{i}'}).$$

Using these two observations along with the relation (A.78), we have

$$\begin{aligned}
&\sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}:\mathbf{i}\not\succ\mathbf{q}^\ell} (|(D\boldsymbol{\alpha}_Q)_\mathbf{i}| - \widetilde{\mathfrak{s}}(\mathbf{i})(D\boldsymbol{\alpha}_Q)_\mathbf{i}) \\
&= \sum_{\mathbf{i}\in Q:\mathbf{i}\succ\mathbf{q}^\ell} \left\{ \left| \sum_{\mathbf{i}'\preceq\mathbf{i}} t(\mathbf{i}',\mathbf{i})\beta_{\mathbf{i}'} \right| - \widetilde{\mathfrak{s}}(\mathbf{i}) \sum_{\mathbf{i}'\preceq\mathbf{i}} t(\mathbf{i}',\mathbf{i})\beta_{\mathbf{i}'} \right\} \\
&= \sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}:\mathbf{i}\not\succ\mathbf{q}^\ell} \sum_{\mathbf{i}'\preceq\mathbf{i}} t(\mathbf{i}',\mathbf{i})\{|\beta_{\mathbf{i}'}| - \widetilde{\mathfrak{s}}(\mathbf{i})t(\mathbf{i}',\mathbf{i})\beta_{\mathbf{i}'}\} \\
&\le 2 \sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}:\mathbf{i}\not\succ\mathbf{q}^\ell} \sum_{\mathbf{i}'\preceq\mathbf{i}} t(\mathbf{i}',\mathbf{i})\{|\beta_{\mathbf{i}'}| - \mathfrak{s}(\mathbf{i}')\beta_{\mathbf{i}'}\} \\
&= 2 \sum_{\mathbf{i}'\preceq\mathbf{q}^u} \{|\beta_{\mathbf{i}'}| - \mathfrak{s}(\mathbf{i}')\beta_{\mathbf{i}'}\} \sum_{\mathbf{i}\in Q} t(\mathbf{i}',\mathbf{i})\mathbb{I}\{\mathbf{i}'\preceq\mathbf{i}, \mathbf{i}\not\succ\mathbf{q}^\ell, \mathbf{i}\neq\mathbf{q}^\ell\} \\
&\le 2 \sum_{\mathbf{i}'\preceq\mathbf{q}^u:\mathbf{i}'\not\succ\mathbf{q}^\ell,\mathbf{i}'\notin\{\mathbf{0},\mathbf{i}^*\}} \{|\beta_{\mathbf{i}'}| - \mathfrak{s}(\mathbf{i}')\beta_{\mathbf{i}'}\}.
\end{aligned}$$

In the last step we noted that for a given $\mathbf{i}' \preceq \mathbf{q}^u$, there is a unique $\mathbf{i}$ such that $t(\mathbf{i}',\mathbf{i})$ is nonzero (second part of Lemma A.5.1), so the inner sum only has one nonzero addend. Then we used assumption (a) to note that $\mathbb{I}\{\mathbf{i}'\preceq\mathbf{i}, \mathbf{i}\not\succ\mathbf{q}^\ell, \mathbf{i}\neq\mathbf{q}^\ell\} \le \mathbb{I}\{\mathbf{i}'\not\succ\mathbf{q}^\ell, \mathbf{i}'\notin\{\mathbf{0},\mathbf{i}^*\}\}$ for any $\mathbf{i}\in Q$ such that $t(\mathbf{i}',\mathbf{i}) = 1$.

Finally, for $\mathbf{i}\in Q$ such that $\mathbf{i}\succ\mathbf{q}^\ell$, let $\widetilde{\mathfrak{s}}(\mathbf{i}) \coloneqq \mathfrak{s}(\mathbf{i})$. Combining the above work with the fact that (A.78) implies that $(D\boldsymbol{\alpha}_Q)_\mathbf{i} = (D\boldsymbol{\alpha})_\mathbf{i}$ for $\mathbf{i}\succ\mathbf{q}^\ell$, we obtain

$$\begin{aligned}
&\sum_{\mathbf{i}\in Q\setminus\{\mathbf{q}^\ell\}} (|(D\boldsymbol{\alpha}_Q)_\mathbf{i}| - \widetilde{\mathfrak{s}}(\mathbf{i})(D\boldsymbol{\alpha}_Q)_\mathbf{i}) \\
&\le \sum_{\mathbf{i}\in Q:\mathbf{i}\succ\mathbf{q}^\ell} \{|\beta_\mathbf{i}| - \mathfrak{s}(\mathbf{i})\beta_\mathbf{i}\} + 2 \sum_{\mathbf{i}\preceq\mathbf{q}^u:\mathbf{i}\not\succ\mathbf{q}^\ell,\mathbf{i}\notin\{\mathbf{0},\mathbf{i}^*\}} \{|\beta_\mathbf{i}| - \mathfrak{s}(\mathbf{i})\beta_\mathbf{i}\} \\
&\le 2 \sum_{\mathbf{i}\notin\{\mathbf{0},\mathbf{i}^*\}} \{|\beta_{\mathbf{i}'}| - \mathfrak{s}(\mathbf{i})\beta_\mathbf{i}\}.
\end{aligned}$$

The last inequality follows by noting that the two sums indexed by $\mathbf{i}$ are over disjoint sets. Finally, the right-hand side can be bounded by 2δ due to (A.45).

A.5.11 Proof of Lemma A.3.15

Without loss of generality we assume $t = 1$ (the general result can then be obtained by scaling and replacing δ by δ/t).

Let $\boldsymbol{\beta}$ be such that $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta}$. Let $\boldsymbol{\pi}(\boldsymbol{\theta}) \coloneqq \mathbf{A}\boldsymbol{\beta}^+$ and $\boldsymbol{\nu}(\boldsymbol{\theta}) \coloneqq \mathbf{A}\boldsymbol{\beta}^-$, where $\beta_1^+ = \beta_1$ and $\beta_i^+ \coloneqq \max\{\beta_i, 0\}$ for $i \ge 2$, and where $\boldsymbol{\beta}^- \coloneqq \boldsymbol{\beta}^+ - \boldsymbol{\beta}$. Then $\boldsymbol{\theta} = \boldsymbol{\pi}(\boldsymbol{\theta}) - \boldsymbol{\nu}(\boldsymbol{\theta})$, and both $\boldsymbol{\pi}(\boldsymbol{\theta})$ and $\boldsymbol{\nu}(\boldsymbol{\theta})$ are entirely monotone.

We have the following two equalities.

$$\theta_n - \theta_1 = [(\boldsymbol{\pi}(\boldsymbol{\theta}))_n - (\boldsymbol{\pi}(\boldsymbol{\theta}))_1] - (\boldsymbol{\nu}(\boldsymbol{\theta}))_n,$$
$$V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta}) = \sum_{i=2}^{n} |\beta_i| = [(\boldsymbol{\pi}(\boldsymbol{\theta}))_n - (\boldsymbol{\pi}(\boldsymbol{\theta}))_1] + (\boldsymbol{\nu}(\boldsymbol{\theta}))_n.$$

Combining these two equalities shows that the constraint $V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta}) \le \theta_n - \theta_1 + \delta$ is equivalent to

$$\sum_{i\ge 2}^{n} \beta_i^- = (\boldsymbol{\nu}(\boldsymbol{\theta}))_n \le \frac{\delta}{2}.$$

Then

$$\|\boldsymbol{\nu}(\boldsymbol{\theta})\|^2 \le n(\boldsymbol{\nu}(\boldsymbol{\theta}))_n^2 \le \frac{\delta^2}{4} n.$$

By the triangle inequality,

$$\|\boldsymbol{\pi}(\boldsymbol{\theta})\| \le \|\boldsymbol{\theta}\| + \|\boldsymbol{\nu}(\boldsymbol{\theta})\| \le 1 + \frac{\delta}{2}\sqrt{n}.$$

Thus,

$$\mathbb{E} \sup_{\substack{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|\le 1,\\ V_{\mathrm{HK}\mathbf{0}}(\boldsymbol{\theta}) \le \theta_n - \theta_1 + \delta}} \langle \boldsymbol{\theta}, \xi \rangle \qquad \le \mathbb{E} \sup_{\substack{\boldsymbol{\pi} \in \mathcal{D}_{n_1,\ldots,n_d}:\\ \|\boldsymbol{\theta}\| \le 1 + \delta\sqrt{n}/2}} \langle \boldsymbol{\pi}, \xi \rangle + \mathbb{E} \sup_{\substack{\boldsymbol{\nu} \in \mathcal{D}_{n_1,\ldots,n_d}:\\ \|\boldsymbol{\nu}\| \le \delta\sqrt{n}/2}} \langle -\boldsymbol{\nu}, \xi \rangle$$

Since $\mathcal{D}$ is a cone and since $\xi \overset{d}{=} -\xi$, the right-hand side can be written as

$$\sigma(1 + \delta\sqrt{n})\ \mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d}: \|\boldsymbol{\theta}\| \le 1} \langle \boldsymbol{\theta}, \mathbf{z} \rangle,$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. From the earlier Gaussian width bound (A.18) (with $\boldsymbol{\theta}^* = 0$, $V^* = 0$, $t = 1$, and using the bounds $|\mathcal{K}| \le C(\log n)^d$ and $\log(2e\sqrt{|\mathcal{K}|}) \le C_d \log(e \log n))$ we have

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \mathcal{D}_{n_1,\ldots,n_d}: \|\boldsymbol{\theta}\| \le 1} \langle \boldsymbol{\theta}, \mathbf{z} \rangle \le C_d (\log(en))^{\frac{3d}{4}} (\log(e \log(en)))^{\frac{2d-1}{4}}.$$

A.5.12 Proof of Lemma A.3.16

Because $\beta_{\mathbf{i}} = 0$ for all $\mathbf{i} \succ \mathbf{0}$, we have the following equality for all $\mathbf{i} \in \{0, \ldots, n_1 - 1\} \times \{0, \ldots, n_2 - 1\}$.

$$\theta_{\mathbf{i}} = \sum_{\mathbf{i}' \preceq \mathbf{i}} \beta_{\mathbf{i}'} = \sum_{i_1'=0}^{i_1} \beta_{i_1',0} + \sum_{i_2'=0}^{i_2} \beta_{0,i_2'} - \beta_{0,0} = \theta_{i_1,0} + \theta_{0,i_2} - \theta_{0,0}, \qquad \forall \mathbf{i}. \tag{A.79}$$

Let $\overline{\theta}_1 \coloneqq \frac{1}{n_1} \sum_{i_1=1}^{n_1} \theta_{i_1,1}$ and $\overline{\theta}_2 \coloneqq \frac{1}{n_2} \sum_{i_2=1}^{n_2} \theta_{1,i_2}$

Note that the identity (A.79) implies

$$\begin{aligned}1 &\geq \|\boldsymbol{\theta}\|^2 \\ &= \sum_{i_1=0}^{n_1-1}\sum_{i_2=0}^{n_2-1}[(\theta_{i_1,0}-\overline{\theta}_1)+(\theta_{0,i_2}-\overline{\theta}_2)-(\theta_{0,0}-\overline{\theta}_1-\overline{\theta}_2)]^2 \\ &= n_2\sum_{i_1=0}^{n_1-1}(\theta_{i_1,0}-\overline{\theta}_1)^2+n_1\sum_{i_2=0}^{n_2 01}(\theta_{0,i_2}-\overline{\theta}_2)^2+n_1n_2(\theta_{0,0}-\overline{\theta}_1-\overline{\theta}_2)^2,\end{aligned}$$

where the cross terms vanish in the last step due to $\sum_{i_1=0}^{n_1-1}(\theta_{i_1,0}-\overline{\theta}_1)=0$ and $\sum_{i_2=0}^{n_2-1}(\theta_{0,i_2}-\overline{\theta}_2)=0$. Thus the vectors $\sqrt{n_2}(\theta_{i_1,0}-\overline{\theta}_1)_{i_1=0}^{n_1-1}$, $\sqrt{n_2}(\theta_{0,i_2}-\overline{\theta}_1)_{i_2=0}^{n_2-1}$, and $\sqrt{n_1n_2}(\theta_{0,0}-\overline{\theta}_1-\overline{\theta}_2)$ each have norm bounded by 1.

Let us view Z as a $n_1\times n_2$ matrix, and define $Z_{\cdot,i_2}\coloneqq\sum_{i_1=0}^{n_1-1}Z_{i_1,i_2}$, $Z_{i_1,\cdot}\coloneqq\sum_{i_2=0}^{n_2-1}Z_{i_1,i_2}$, and $Z_{\cdot,\cdot}\coloneqq\sum_{i_1=0}^{n_1-1}\sum_{i_2=0}^{n_2-1}Z_{i_1,i_2}$. Then, using the identity (A.79) we can decompose the inner product as

$$\begin{aligned}\langle Z,\boldsymbol{\theta}\rangle=\sum_{\mathbf{i}}Z_{\mathbf{i}}\theta_{\mathbf{i}} &= \sum_{i_1=0}^{n_1-1}\sum_{i_2=0}^{n_2-1}Z_{i_1,i_2}(\theta_{i_1,0}+\theta_{0,i_2}-\theta_{0,0}) \\ &= \sum_{i_1=0}^{n_1-1}Z_{i_1,\cdot}\theta_{i_1,0}+\sum_{i_2=0}^{n_2-1}Z_{\cdot,i_2}\theta_{0,i_2}-Z_{\cdot,\cdot}\theta_{1,1} \\ &= \sum_{i_1=0}^{n_1-1}Z_{i_1,\cdot}(\theta_{i_1,0}-\overline{\theta}_1)+\sum_{i_2=0}^{n_2-1}Z_{\cdot,i_2}(\theta_{0,i_2}-\overline{\theta}_2)-Z_{\cdot,\cdot}(\theta_{0,0}-\overline{\theta}_1-\overline{\theta}_2) \\ &= \sum_{i_1=0}^{n_1-1}\frac{Z_{i_1,\cdot}}{\sqrt{n_2}}\sqrt{n_2}(\theta_{i_1,0}-\overline{\theta}_1)+\sum_{i_2=0}^{n_2-1}\frac{Z_{\cdot,i_2}}{\sqrt{n_1}}\sqrt{n_1}(\theta_{0,i_2}-\overline{\theta}_2) \\ &\quad -\frac{Z_{\cdot,\cdot}}{\sqrt{n_1n_2}}\sqrt{n_1n_2}(\theta_{0,0}-\overline{\theta}_1-\overline{\theta}_2).\end{aligned}$$

Note that $(Z_{\cdot,i_2}/\sqrt{n_1})_{i_1=1}^{n_1}$, $(Z_{i_1,\cdot}/\sqrt{n_2})_{i_2=1}^{n_2}$, and $Z_{\cdot,\cdot}/\sqrt{n_1,n_2}$ are each standard Gaussian vectors.

Finally, note that because $\beta_{\mathbf{i}}=0$ for $\mathbf{i}\succ\mathbf{1}$, the HK variation condition on $\boldsymbol{\theta}$ can be written as

$$\sum_{i_1=1}^{n_1-1}(|\beta_{i_1,0}|-s_1\beta_{i_1,0})+\sum_{i_2=1}^{n_2-1}(|\beta_{0,i_2}|-s_2\beta_{0,i_2})\leq\delta,$$

and thus each of these two sums is bounded by δ

Thus, we can bound the expectation in the lemma by

$$\mathbb{E}\sup_{\substack{\widetilde{\boldsymbol{\theta}}\in\mathbb{R}^{n_1}:\|\widetilde{\boldsymbol{\theta}}\|\leq 1\\ V_{\mathrm{HK}\mathbf{0}}(\widetilde{\boldsymbol{\theta}})\leq s_1(\widetilde{\theta}_{n_1}-\widetilde{\theta}_1)+\delta}}\langle Z_{n_1},\widetilde{\boldsymbol{\theta}}\rangle+\mathbb{E}\sup_{\substack{\widetilde{\boldsymbol{\theta}}\in\mathbb{R}^{n_2}:\|\widetilde{\boldsymbol{\theta}}\|\leq 1\\ V_{\mathrm{HK}\mathbf{0}}(\widetilde{\boldsymbol{\theta}})\leq s_2(\widetilde{\theta}_{n_2}-\widetilde{\theta}_1)+\delta}}\langle Z_{n_2},\widetilde{\boldsymbol{\theta}}\rangle+\mathbb{E}\sup_{\widetilde{\theta}\in\mathbb{R}:|\widetilde{\theta}|\leq 1}Z_1\widetilde{\theta},$$

where Z_{n_1}, Z_{n_2}, and Z_1 are standard Gaussian vectors of the appropriate dimension. The third term is readily computed to be $\mathbb{E}|Z_1| = \sqrt{2/\pi}$.

We now focus on the first term; the second term can be bounded analogously. If $s_1 \in \{-1, 1\}$, then Lemma C.8 of Guntuboyina et al. [44] implies a bound of

$$c(1 + \delta\sqrt{n_1})\sqrt{\log(en_1)}$$

Otherwise if $s_1 = 0$, then Lemma B.1 of the same paper [44] yields a bound of

$$c(\delta\sqrt{n_1})^{\frac{1}{2}} + c\sqrt{\log(en_1)}.$$

Handling the second term in the same fashion concludes the proof.