

# An improved global risk bound in concave regression

Sabyasachi Chatterjee

*Department of Statistics  
The University of Chicago  
e-mail: [sabyasachi@uchicago.galton.edu](mailto:sabyasachi@uchicago.galton.edu)*

**Abstract:** A new risk bound is presented for the problem of convex/concave function estimation, using the least squares estimator. The best known risk bound, as had appeared in Guntuboyina and Sen [8], scaled like  $\log(en) n^{-4/5}$  under the mean squared error loss, up to a constant factor. The authors in [8] had conjectured that the logarithmic term may be an artifact of their proof. We show that indeed the logarithmic term is unnecessary and prove a risk bound which scales like  $n^{-4/5}$  up to constant factors. Our proof technique has one extra peeling step than in a usual chaining type argument. Our risk bound holds in expectation as well as with high probability and also extends to the case of model misspecification, where the true function may not be concave.

**MSC 2010 subject classifications:** Primary 60K35, 60K35; secondary 60K35.

Received May 2016.

## Contents

1	Introduction . . . . .	1608
2	Main result . . . . .	1610
	2.1 Proof sketch for Theorem 2.1 . . . . .	1612
3	Proof of main result . . . . .	1613
	3.1 Proof of Theorem 2.1 . . . . .	1613
	3.2 Proof of Proposition 3.1 . . . . .	1616
	3.3 Comparison with [8] . . . . .	1624
4	Model misspecification . . . . .	1626
	Appendix . . . . .	1627
	Acknowledgements . . . . .	1628
	References . . . . .	1628

## 1. Introduction

In this paper we consider the problem of estimating a concave function in a standard additive noise regression model. We observe  $y = (y_1, \dots, y_n)$  which are noisy evaluations of an underlying concave function defined on the unit interval  $[0, 1]$  on sorted design points  $x = (x_1 < \dots < x_n) \in [0, 1]^n$  as follows;

$$y_i = f(x_i) + z_i \tag{1.1}$$

with  $z = (z_1, \dots, z_n)$  being independent  $N(0, \sigma^2)$  error for some unknown  $\sigma$ . The task of estimating the regression function  $f$  when knowing that the underlying  $f$  is concave is known as the concave regression problem. This problem has a relatively long history.

Interestingly, the problem of concave regression was discussed in the literature by Hildreth [10] even before the problem of monotone regression (see Robertson et al. [14], Ayer et al. [1]), which is perhaps the most well studied problem in shape constrained estimation. [10] discusses some applications in economics where concavity arises naturally, such as in production functions and utility functions. A very natural estimator of a concave regression function is the constrained least squares estimator (LSE) which is defined as follows:

$$\hat{f} = \arg \min_{f: [0,1] \rightarrow \mathbb{R}, f \text{ concave}} \sum_{i=1}^n (y_i - f(x_i))^2. \quad (1.2)$$

As is clear, the LSE  $\hat{f}$  is not unique but is uniquely defined at the design points. This estimator was proposed in [10] along with quadratic programming methods for solving (1.2). The first statistical analysis for the LSE was carried out by Hanson and Pledger [9] where the consistency of the LSE was established under the supremum loss over a compact interval. The problem of estimation of the concave function under a local loss function, that is the problem of estimating the regression function  $f$  at a given point  $x_0$ , has also received attention in the literature. Notable papers are Mammen [13], Groeneboom et al. [7] where the asymptotic distribution of the LSE at the point  $x_0$  was studied. The first work which studied rates of convergence of the LSE under a global loss appeared in [6]. The authors there showed that under the supremum loss over a compact interval containing the design points, the rates of convergence varies from  $(\frac{\log n}{n})^{1/3}$  to  $(\frac{\log n}{n})^{2/5}$  depending on the smoothness of the underlying concave function. For instance, if the regression function is also twice differentiable the rate of convergence under the supremum loss is  $(\frac{\log n}{n})^{2/5}$ .

In the most recent work on concave regression under a global loss function, Guntuboyina and Sen [8] considered the following sequence formulation of the concave regression problem. Assuming that the design points are equally spaced, note that the mean regression vector  $(f(x_1), \dots, f(x_n))$  lies in a polyhedral cone (defined by the concavity constraint on  $f$ )  $K_n \subset \mathbb{R}^n$ ,  $n \geq 3$  defined as follows:

$$K_n = \{\theta \in \mathbb{R}^n : \theta_i - \theta_{i-1} \geq \theta_{i+1} - \theta_i \text{ whenever } 2 \leq i \leq n-1\}.$$

Then it is clear that if  $x_1 < x_2 < \dots < x_n$  are equispaced design points in  $\mathbb{R}$  then one can also write

$$K_n = \{\theta \in \mathbb{R}^n : \theta = (f(x_1), \dots, f(x_n)) \text{ for some concave function } f : \mathbb{R} \rightarrow \mathbb{R}\}.$$

The problem of estimating the entire concave function in the sequence model then becomes the problem of estimating  $\theta^* \in K_n$  from observations

$$y_i = \theta_i^* + z_i \quad \forall 1 \leq i \leq n \quad (1.3)$$

where  $z_i$  are i.i.d  $N(0, \sigma^2)$  random variables.

In this paper, we study this sequence formulation of the concave regression problem, assuming equally spaced design points. Specifically, we study the risk properties of the Least Squares Estimator  $\hat{\theta}$  defined as

$$\hat{\theta} = \arg \min_{\theta \in K_n} \|y - \theta\|^2, \quad (1.4)$$

where  $\|\cdot\|$  refers to the usual Euclidean norm on  $\mathbb{R}^n$ . We are interested in studying the risk behaviour of  $\hat{\theta}$  under the natural mean squared error loss function defined as

$$R(\hat{\theta}, \theta^*) = \mathbb{E} \frac{1}{n} \|\hat{\theta} - \theta^*\|^2, \quad (1.5)$$

where the expectation is taken under the distribution of  $y$  as given in (1.3).

To describe the results of [8] under the risk function  $R(\hat{\theta}, \theta^*)$ , we first have to set up some notation. Let  $L$  be the subspace of  $\mathbb{R}^n$  spanned by the constant vector  $(1, \dots, 1)$  and the vector  $(1, 2, \dots, n)$ . In words,  $L$  is the subspace of  $n$  dimensional affine sequences. Let  $P_L$  denote the orthogonal projection matrix to the subspace  $L$ . Define  $G_{\theta^*} = \max\{\sigma^2, \frac{1}{n} \|(I - P_L)\theta^*\|^2\}$ . Then Theorem 2.2 in [8] shows that for any  $\theta^* \in K_n$  there exists a universal constant  $C$  such that whenever the sample size  $n \geq C \frac{\sigma^2}{(G_{\theta^*})^2} (\log \frac{en}{2})^{5/4}$ , we have

$$R(\hat{\theta}, \theta^*) \leq C \left(\log \frac{en}{2}\right) \left(\frac{\sigma^2 \sqrt{G_{\theta^*}}}{n}\right)^{4/5}. \quad (1.6)$$

To interpret the theorem, it is instructive to think of a regime where  $G(\theta^*)$  stays bounded as  $n$  grows. For example, this would be the case if  $\theta^*$  are evaluations on a grid of a concave function  $f: \mathbb{R} \rightarrow \mathbb{R}$  with  $\int_0^\infty f^2(x) dx < \infty$ . In this case, the bound in 1.6 yields  $R(\hat{\theta}, \theta^*) \leq (\log \frac{en}{2}) n^{-4/5}$  upto a multiplicative constant factor. [8] conjectured that there should be no logarithmic term in the upper bound and perhaps the logarithmic term is an artifact of their proof. The goal of this paper is to show that one can derive a risk bound in the same setting as in (1.6), but without the logarithmic factor.

The paper is organized as follows. In Section 2 we present our main result on the global risk bound, comment on its interpretations, and sketch the proof ideas. In Section 3, we present the proof of our main result. In Section 4, we state and prove an extension of our main result in the case of model misspecification.

## 2. Main result

We now describe the main result of this paper and discuss its consequences. For any vector  $\theta \in \mathbb{R}^n$  let us define  $V(\theta) = \max_{1 \leq i \leq n} \theta_i - \min_{1 \leq i \leq n} \theta_i$ .

**Theorem 2.1.** *Fix any positive integer  $n$ . Fix any  $\theta^* \in K_n$ . Fix any  $x > 0$ . The following upper bound on the risk holds for a universal constant  $C$  with probability greater than  $1 - 2 \exp(-x)$ :*

$$\begin{aligned} \frac{1}{n} \|\widehat{\theta} - \theta^*\|^2 &\leq \frac{2}{n} \sigma^2 (1 + \sqrt{2x} + x) \\ &\quad + \max\{C\sigma^{8/5} (V((I - P_L)\theta^*) + \sigma)^{2/5} n^{-4/5}, C\sigma^2 \frac{\log n}{n}\} \end{aligned} \quad (2.1)$$

Moreover, we also have a risk bound in expectation for a universal constant  $C$  as follows:

$$R(\widehat{\theta}, \theta^*) \leq \max\{C\sigma^{8/5} (V((I - P_L)\theta^*) + \sigma)^{2/5} n^{-4/5}, C\sigma^2 \frac{\log n}{n}\} + \frac{2\sigma^2}{n}. \quad (2.2)$$

Theorem 2.1 proves a high probability bound for  $\frac{1}{n} \|\widehat{\theta} - \theta^*\|^2$  which then immediately leads to a bound in expectation. We show in Section 4 that we can also extend our risk bound to the case of model misspecification. By model misspecification, we mean the case when the true underlying sequence  $\theta^*$  is not concave. In this case, we prove that the risk of the LSE  $R(\widehat{\theta}, \theta^*)$  is at most the squared Euclidean distance of  $\theta^*$  to  $K_n$  divided by  $n$ , plus a term which goes down to zero at the rate  $n^{-4/5}$ .

Comparing the bound in [8] with our risk bound, we have no logarithmic terms in our risk bound. However, we pay the price of a larger constant in order to remove the logarithmic term. The term  $G(\theta^*)$  is now replaced by  $V((I - P_L)\theta^*)$  which is generally larger. In the situation where  $\theta^*$  are evaluations of a fixed concave function  $f : [0, 1] \rightarrow \mathbb{R}$  on the grid  $\{1/n, \dots, (n-1)/n, 1\}$  the term  $V(\theta^*)$  does stay bounded by a number which is a property of  $f$  and does not change with  $n$ . In this situation, our risk bound gives the optimal rate  $n^{-4/5}$  upto constant factors. In general though, the risk bound proposed in this manuscript is not a uniform improvement over the bound (1.6) because  $G(\theta^*)$  and  $V(\theta^*)$  may have different rates of growth with  $n$  depending on the chosen sequence of concave sequences  $\theta^* \in K_n$ . A more detailed comparison between the risk bound (2.2) and the bound (1.6) is deferred after the completion of proof of our main Theorem 2.1 and is done in subsection 3.3.

Our analysis actually carries through for slightly more general design points than just equally spaced ones; see Remark 3.2 for more explanation. Coming to the issue of optimality of the LSE, the rate  $n^{-4/5}$  is known to be minimax rate optimal even if one considers the parameter space to be an appropriate local ball around a convex function with positive curvature everywhere; see Theorem 5.1 in [8] for the precise result. Apart from minimax rate optimality, a further motivation to study the LSE in concave regression is that it was shown to have automatic adaptation properties first in [8] and subsequently in [3] and [2]. Specifically, it was shown that the risk  $R(\widehat{\theta}, \theta^*)$  scaled like  $\frac{k}{n} \log n$  for piecewise linear concave sequences  $\theta^*$  with  $k$  pieces. In this paper though, our focus is not on the automatic adaptive properties of the LSE, but rather in proving an improved worst case performance.

It is worthwhile to point out that we do not assume any smoothness assumptions on the underlying concave function. Considering the result of Dümbgen et al. [6], their loss function is the supremum loss function over an appropriate compact interval. Notwithstanding some details, the mean squared error loss

function we consider is naturally bounded by the square of the supremum loss function that is considered in Dümbgen et al. [6]. Hence their risk bound directly applied to the mean squared error loss scales like  $(\frac{\log n}{n})^{4/5}$  only when the underlying concave function is twice differentiable. Our analysis shows that when the design points are equally spaced, we do not need any smoothness assumptions on the underlying concave function to obtain the  $n^{-4/5}$  rate in addition to showing that the logarithmic factor is unnecessary.

### 2.1. Proof sketch for Theorem 2.1

The goal of this subsection is to provide a high level overview of the method of proof of Theorem 2.1. We take the standard empirical process based approach in proving our risk bound. Specifically we take the recipe proposed by Chatterjee [4] which says that a key ingredient in proving risk bounds for the LSE is to control an expected Gaussian suprema term described as follows. Let  $\langle a, b \rangle$  denote the usual inner product between any two vectors  $a, b$ . For any  $\theta^* \in \mathbb{R}^n$  define the function  $f_{\theta^*} : \mathbb{R}_+ \rightarrow \mathbb{R}$  as follows:

$$f_{\theta^*}(t) = \mathbb{E} \sup_{\theta \in K_n : \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle \quad (2.3)$$

where  $z$  is a random Gaussian vector with each entry being independent, has mean zero and variance  $\sigma^2$ . [4] also shows that if one obtains  $s > 0$  such that  $f_{\theta^*}(s) \leq \frac{s^2}{2}$  then essentially one gets the risk bound  $R(\hat{\theta}, \theta^*) \leq s^2/n$  up to constant factors. Precise statements are given in later sections. Therefore, it suffices to tightly upper bound the function  $f_{\theta^*}$  and this paper basically gives a new way of upper bounding the function  $f_{\theta^*}(t)$  for all  $t \geq 0$ .

Since  $f_{\theta^*}(t)$  is an expected Gaussian maxima, a standard tool in empirical process theory to upper bound  $f_{\theta^*}(t)$  is to use the Dudley's entropy integral bound which requires good estimates of the covering number of the set  $K_n \cap B(\theta^*, t)$  where  $B(\theta^*, t)$  refers to the Euclidean ball of radius  $t$  centred at  $\theta^*$ . Tight estimates (without logarithmic factors) of the covering numbers for the space of bounded convex sequences are available in the literature, see Dryanov [5]. Specifically the result of [5] gives us tight upper bounds of the covering number for the space  $\{\theta \in K_n : \max_{1 \leq i \leq n} |\theta_i| \leq B\}$  for any constant  $B$ . But we require covering numbers for  $K_n$  intersected with a Euclidean ball of radius  $t$ . This was done in [8] by applying the basic result of [5] in appropriate subintervals. This approach gives rise to logarithmic factors in the risk bound as had appeared in [8].

Our approach is to do one more step of refining the function  $f_{\theta^*}$ . Essentially for any  $\theta \in K_n \cap B(\theta^*, t)$  we define its truncated version  $\theta' \in K'_n$  where the set  $K'_n \subset \mathbb{R}^n$  is not a much larger set than  $K_n$  and hence has comparable metric entropy. Note that one can write

$$\mathbb{E} \sup_{\theta \in K_n} \langle z, \theta - \theta^* \rangle \leq \mathbb{E} \sup_{\theta \in K_n} \langle z, \theta - \theta' \rangle + \mathbb{E} \sup_{\theta'} \langle z, \theta' - \theta^* \rangle. \quad (2.4)$$

We control the two terms on the right side of the above inequality separately. We show it is possible to define the truncation  $\theta'$  of  $\theta$  such that it satisfies two critical properties. The first property is that the first term on the right side of (2.4) is upper bounded by  $\frac{t^2}{4}$ . Since we finally have to compare  $f_{\theta^*}(t)$  with  $\frac{t^2}{2}$  an extra factor of  $\frac{t^2}{4}$  can only affect the risk bound upto constants. Now we are left with the task of upper bounding the second term in the right side of (2.4). The second critical property that  $\theta'$  satisfies is that  $\theta'$  is bounded entrywise by  $C(V(\theta^*) + \sigma)$  where  $C$  is a universal constant. This means that  $\theta'$  is bounded by a constant factor. Hence the second term in the right side of (2.4) can now be controlled by a direct application of Dudley's entropy integral bound. This only requires getting tight estimates of the covering number of bounded sequences in  $K'_n$ . We show that this covering number is very similar to the one given in [5] (and hence has no logarithmic factors) as  $K'_n$  is not much bigger than  $K_n$ . In this way our extra refinement step enables us to save a logarithmic factor of  $n$ .

The main crux of this paper lies in defining  $\theta'$  as described in the previous paragraph. This is done in Lemma 3.5 which is perhaps the most important step in our entire argument. In Lemma 3.5 we actually assume our underlying concave sequence  $\theta^*$  is also monotonic. Since a concave sequence always first increases and then decreases, it actually suffices to analyze the case when  $\theta^*$  is monotonic concave. We use the monotonicity of  $\theta^*$  crucially in coming up with a definition of the truncation  $\theta'$  of  $\theta$ , satisfying the critical properties as explained in the previous paragraph.

### 3. Proof of main result

#### 3.1. Proof of Theorem 2.1

The goal of this section is to state and prove Theorem 2.1. Before starting to prove the above theorem, we first go through some background results. By now it is known that a key ingredient in proving risk bounds for the LSE is to control the expected Gaussian suprema function  $f_{\theta^*}$  as defined in (2.3). For any  $\theta^* \in K_n$  it was actually shown in Theorem 1.1 in Chatterjee [4] that the loss term  $\|\widehat{\theta} - \theta^*\|$  concentrates around a deterministic value  $t_{\theta^*} = \arg \max_{t \geq 0} f_{\theta^*}(t) - \frac{t^2}{2}$ . Another result providing upper bounds on the loss term  $\|\widehat{\theta} - \theta^*\|$  in terms of the function  $f_{\theta^*}$  has been given in Theorem 12 in Bellec [2] which we use as our starting point. Before describing this result, we claim that since the projection onto the cone  $K_n$  is a sum of projection onto the subspace  $L$  and the projection onto the cone  $K_n \cap L^\perp$  it almost suffices to study the risk of the least squares estimator constrained to the cone  $K_n \cap L^\perp$  denoted by  $\widehat{\theta}_{K_n \cap L^\perp}$ . Specifically we mean

$$\widehat{\theta}_{K_n \cap L^\perp} = \arg \min_{\theta \in K_n \cap L^\perp} \|y - \theta\|^2.$$

Here  $L^\perp$  is the subspace of  $\mathbb{R}^n$  orthogonal to  $L$ . We make this claim clearer when we prove Theorem 2.1. With this viewpoint, we now state the following

lemma which is a direct consequence of Theorem 12 in [2], applied to the cone  $K_n \cap L^\perp$ .

**Lemma 3.1** ([2]). *Let  $\theta^* \in K_n \cap L^\perp$ . Let  $s > 0$  be such that*

$$\mathbb{E} \sup_{\theta \in K_n \cap L^\perp: \|\theta - \theta^*\| \leq s} \langle z, \theta - \theta^* \rangle \leq \frac{s^2}{2}.$$

*Then for any  $x > 0$  the following inequality holds with probability greater than  $1 - \exp(-x)$ ,*

$$\|\widehat{\theta}_{K_n \cap L^\perp} - \theta^*\|^2 \leq 2 \max\{s^2, 8\sigma^2 x\}.$$

*Here  $\widehat{\theta}_{K_n \cap L^\perp}$  is the least squares estimator constrained to the cone  $K_n \cap L^\perp$ . Also  $z$  is a  $n$  dimensional gaussian random vector with mean zero and covariance matrix  $\sigma^2 I$ .*

**Remark 3.1.** Actually Theorem 12 in Bellec [2] implies a slightly stronger bound and is useful in the case of model misspecification. This issue is discussed in Section 4.

Recall the definition of the function  $f_{\theta^*}$  in (2.3). Also recall that for any vector  $\theta \in \mathbb{R}^n$  the range of  $\theta$  is denoted by  $V(\theta)$ . The key step in proving Theorem 2.1 is to obtain the following key upper bound on  $f_{\theta^*}$  which we state as a proposition:

**Proposition 3.1.** *There exists a universal  $C > 0$  such that for all  $n \geq 1$ , all  $\theta^* \in K_n$  and all  $t \geq 0$ , we have the following inequality:*

$$f_{\theta^*}(t) \leq C\sigma \left\{ (V(\theta^*) + \sigma)^{1/4} n^{1/8} t^{3/4} + \sqrt{\log nt} \right\} + \frac{t^2}{4}.$$

We prove the above proposition in the next subsection. We first prove Theorem 2.1 assuming the above proposition is true.

*Proof of Theorem 2.1.* Let  $K = K_n$  for simplicity. Fix  $\theta^* \in K$ . Recall that the subspace spanned by the  $n$  dimensional vectors  $(1, \dots, 1)$  and  $(1, \dots, n)$  is denoted as  $L$ . It is clear that  $L$  is the smallest subspace contained in  $K$ . Also recall that  $P_L$  is the orthogonal projection matrix to the subspace  $L$ . Let the projection of any  $\theta \in \mathbb{R}^n$  onto any closed convex cone  $C \subset \mathbb{R}^n$  be denoted by  $\Pi_C(\theta)$ . The projection exists and is unique because  $C$  is a closed convex set. Now by definition of the LSE and by orthogonal decomposition of projections we have

$$\widehat{\theta} = \Pi_K y = \Pi_K(\theta^* + z) = \Pi_{K \cap L^\perp}(\theta^* + z) + P_L(\theta^* + z).$$

Let  $P_{L^\perp} = I - P_L$  and  $\mu^* = P_{L^\perp} \theta^*$ . Now we can write

$$\begin{aligned} \|\widehat{\theta} - \theta^*\|^2 &= \|\Pi_{K \cap L^\perp}(\theta^* + z) - P_{L^\perp} \theta^*\|^2 + \|P_L(\theta^* + z) - P_L \theta^*\|^2 \\ &= \|\Pi_{K \cap L^\perp}(\mu^* + z) - \mu^*\|^2 + \|P_L z\|^2. \end{aligned} \tag{3.1}$$

The first equality is a sum of squares decomposition using orthogonality. The second equality is because  $\Pi_{K \cap L^\perp}(\theta^* + z) = \Pi_{K \cap L^\perp}(\mu^* + z)$  by definition of  $\mu^*$ . This can be checked by the usual KKT conditions for the projection of a point onto a closed convex cone. Now since  $L$  is a subspace with dimension 2 the term  $\frac{\|P_L z\|^2}{\sigma^2}$  is distributed as a  $\chi^2$  random variable with 2 degrees of freedom. Hence by a standard tail inequality of a  $\chi^2$  random variable we have the following inequality which holds for any  $x > 0$  with probability greater than  $1 - \exp(-x)$ ,

$$\|P_L z\|^2 \leq 2\sigma^2(1 + \sqrt{2x} + x). \quad (3.2)$$

The proof of this inequality can be found in [11]. Now since  $\mu^* \in K \cap L^\perp$  and  $\|\Pi_{K \cap L^\perp}(\mu^* + z) - \mu^*\|^2$  is exactly the squared error loss term for the least squares estimator constrained to the cone  $K \cap L^\perp$  we can use Lemma 3.1 to upper bound  $\|\Pi_{K \cap L^\perp}(\mu^* + z) - \mu^*\|^2$ . First we use Proposition 3.1 to get the following inequality for all  $t \geq 0$ ,

$$\begin{aligned} \mathbb{E} \sup_{\nu \in K \cap L^\perp: \|\nu - \mu^*\| \leq t} \langle z, \nu - \mu^* \rangle &\leq f_{\mu^*}(t) \leq \\ C\sigma \{ (V(\mu^*) + \sigma)^{1/4} n^{1/8} t^{3/4} + \sqrt{\log nt} \} &+ \frac{t^2}{4} \end{aligned}$$

where  $C$  is a universal constant. Now it is not too hard to check that by setting

$$s = \max\{ (C\sigma(V(\mu^*) + \sigma)^{1/4} n^{1/8})^{4/5}, C\sigma\sqrt{\log n} \} \quad (3.3)$$

for a large enough universal constant  $C$  we have  $f_{\mu^*}(s) < \frac{s^2}{2}$ . Setting this value of  $s$  in Lemma 3.1 alongwith the last display then gives us the following bound which holds for any  $x > 0$  with probability greater than  $1 - \exp(-x)$ ,

$$\|\Pi_{K \cap L^\perp}(\mu^* + z) - \mu^*\|^2 \leq 2 \max\{ C\sigma^{8/5} (V(\mu^*) + \sigma)^{2/5} n^{1/5}, C\sigma^2 \log n, 8\sigma^2 x \}. \quad (3.4)$$

Using the upperbounds (3.2) and (3.4) and combining them in (3.1) by a simple union bound argument finishes the proof for the high probability bound statement in (2.1). To prove the risk bound in expectation, let us denote  $W = \|\Pi_{K \cap L^\perp}(\mu^* + z) - \mu^*\|^2$ . Then we have the following inequality for any  $v \geq 0$  due to (3.4):

$$P(W \geq 16\sigma^2 v) = \begin{cases} 1 & \text{if } s^2 > 16\sigma^2 v \\ \leq \exp(-v) & \text{if } s^2 < 16\sigma^2 v \end{cases}$$

where  $s$  is as defined in (3.3).

Since  $\mathbb{E}W = \int_0^\infty P(W \geq v)dv$ , simple integral calculus now gives us

$$\mathbb{E}W \leq Cs^2$$

where  $C$  is a universal constant. Also we have

$$\mathbb{E}\|P_L z\|^2 \leq 2\sigma^2.$$

because  $\frac{\|P_L z\|^2}{\sigma^2}$  is a  $\chi^2$  random variable with degrees of freedom 2. The last two displays alongwith (3.1) finish the proof of the risk bound (4.1).  $\square$



**Remark 3.2.** We remark that we have defined  $K$  to be the space of concave sequences obtained by evaluations of a concave function  $f : [0, 1] \rightarrow \mathbb{R}$  on equally spaced design points  $(x_1, \dots, x_n) \in \mathbb{R}^n$ . Our risk bound calculations carry through if  $(x_1, \dots, x_n)$  are not equally spaced, the only difference being that the universal constant  $C$  would now be replaced by a constant which would only depend on the ratio  $\max_i(x_{i+1} - x_i) / \min_i(x_{i+1} - x_i)$ . This means that our risk bounds would continue to hold in the slightly more general situation where the gaps between consecutive design points lie between  $\frac{c_1}{n}$  and  $\frac{c_2}{n}$  for some constants  $c_1, c_2$ . This is in the same spirit as in the risk analysis given in [8].

### 3.2. Proof of Proposition 3.1

In order to prove Proposition 3.1, we will prove several lemmas along the way. We first set up some notations. For a metric space  $\mathbb{F}$  with metric  $D$  and  $\epsilon > 0$ , let  $N(\epsilon, \mathbb{F}, D)$  denote the  $\epsilon$ -covering number of  $\mathbb{F}$  under the metric  $D$ . That is,  $N(\epsilon, \mathbb{F}, D)$  is the minimum number of balls of radius  $\epsilon$  required to cover  $\mathbb{F}$ . If  $\mathbb{F} \subset \mathbb{R}^n$  and  $D$  is the usual Euclidean metric we simply denote the covering number by  $N(\epsilon, \mathbb{F})$ .

We are required to upper bound the expected Gaussian maxima function  $f_{\theta^*}$ . To do this, we use a standard chaining bound (see [15]) as follows:

**Theorem 3.1** (Chaining). *Let  $\mathbb{F} \subset \mathbb{R}^n$  and fix any  $\theta^* \in \mathbb{F}$ . Let  $d = \sup_{\theta, \theta' \in \mathbb{F}} \|\theta - \theta'\|$  be the diameter of  $\mathbb{F}$ . Then we have*

$$\mathbb{E} \left[ \sup_{\theta \in \mathbb{F}} \langle z, \theta - \theta^* \rangle \right] \leq 12\sigma \int_0^d \sqrt{\log N(\epsilon, \mathbb{F})} \, d\epsilon$$

where  $z$  is again a  $n$  dimensional Gaussian random vector with mean zero and covariance matrix  $\sigma^2 \mathcal{I}$ .

We also use a standard Gaussian Concentration Inequality. The proof can be found in the argument after equation(2.35) in Ledoux [12]:

**Theorem 3.2** (Gaussian Concentration Inequality). *Let  $z$  be an  $n$  dimensional Gaussian random vector with covariance matrix  $\sigma^2 \mathcal{I}$ . Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a function that is  $L$  lipschitz, that is it satisfies  $|f(x) - f(y)| \leq L\|x - y\|$  for all  $x$  and  $y$ , where  $L$  is a positive constant. Then the following is true for any  $t \geq 0$ ,*

$$\mathbb{P}(f(z) \geq \mathbb{E}f(z) + \sigma t) \leq \exp\left(-\frac{t^2}{2L^2}\right).$$

We also need to use a log covering number bound for the space of bounded convex functions defined on the unit interval, as proved in Dryanov [5].

**Lemma 3.2** (Dryanov). *Let  $C[0, 1, B]$  be the space of real valued concave functions defined on the unit interval  $[0, 1]$  with absolute value bounded by  $B$  for some  $B > 0$ . Also let  $L_2(f, g) = (\int_0^1 (f(x) - g(x))^2 dx)^{1/2}$  for any  $f, g \in C[0, 1, B]$ .*

Then the following is true for a universal constant  $C$ ,

$$\log N(\epsilon, C[0, 1, B], L_2) \leq C\sqrt{\frac{B}{\epsilon}} \quad \forall \epsilon > 0.$$

Modifying the above result we can now prove a log covering number bound for the set of concave sequences bounded by a number  $B$ .

**Lemma 3.3.** Fix  $B > 0$ . Let  $K_{n,B} = \{\theta \in K_n : \max_{1 \leq i \leq n} |\theta_i| \leq B\}$ . The following is true for any  $n \geq 3$  and a universal constant  $C$ ,

$$\log N(\epsilon, K_{n,B}) \leq Cn^{1/4}\sqrt{\frac{B}{\epsilon}} \quad \forall \epsilon > 0.$$

*Proof.* Let  $\tau = \sqrt{\frac{24}{n}}\epsilon$ . Let  $\mathbb{F}$  be a finite subset of  $C[0, 1, B]$  such that for any  $f \in C[0, 1, B]$  there exists  $g \in \mathbb{F}$  such that  $L_2(f, g) \leq \tau$ . By Lemma 3.2  $\mathbb{F}$  can be chosen to have  $\log |\mathbb{F}| \leq C\sqrt{B}/\tau$  where  $C$  is a universal constant. Now take any  $\theta \in K_{n,B}$ . Define a convex function  $f_\theta \in C[0, 1, B]$  as follows: Set  $f_\theta(\frac{i-1}{n-1}) = \theta_i \quad \forall 1 \leq i \leq n$  and extend  $f_\theta$  to all other points in the unit interval by linear interpolation. Clearly  $f_\theta \in C[0, 1, B]$ . For each  $g \in \mathbb{F}$  check whether there exists  $\nu \in K_{n,B}$  such that  $L_2(g, f_\nu) \leq \tau$ . If there is, choose such a  $\nu$  arbitrarily and name it  $\nu_g$ . Let  $\mathbb{F}'$  be the set of such  $\nu_g$  obtained as we vary  $g \in \mathbb{F}$ . Clearly we then have for a universal constant  $C$ ,

$$|\mathbb{F}'| \leq |\mathbb{F}| \leq \exp(C\sqrt{B}/\tau). \tag{3.5}$$

Now we claim that a  $2\tau$  covering set for  $K_{n,B}$  has cardinality at most  $|\mathbb{F}'|$  which will suffice to prove the lemma.

Take any  $\theta \in K_{n,B}$ . By definition of  $\mathbb{F}$  there exists  $g \in \mathbb{F}$  such that  $L_2(f_\theta, g) \leq \tau$ . Therefore for this  $g$ , there exists  $\nu_g \in \mathbb{F}'$  such that  $L_2(g, f_{\nu_g}) \leq \tau$ . Hence by the triangle inequality, we have  $L_2(f_\theta, f_{\nu_g}) \leq 2\tau$ . Now a direct consequence of Lemma(A.4) in Guntuboyina and Sen [8] shows the following for a universal constant  $C$ :

$$\frac{1}{n} \|\theta - \nu_g\|^2 \leq C(L_2(f_\theta, f_{\nu_g}))^2 \leq 4C\tau^2.$$

Now set  $\epsilon = \sqrt{\frac{n}{4C}}\tau$  to conclude that  $\mathbb{F}'$  is a  $\epsilon$  cover for  $K_{n,B}$ . Setting the value of  $\epsilon$  in (3.5) now finishes the proof of the lemma.  $\square$

The space of  $n$  dimensional vectors with at most three concave blocks plays an important role in our analysis. We now define the space of sequences with at most three concave blocks as follows. For any vector  $\theta \in \mathbb{R}^n$  let  $\theta_{[a:b]} = (\theta_a, \dots, \theta_b)$  for any integers  $1 \leq a \leq b \leq n$ . Define

$$K3 = \{\theta \in \mathbb{R}^n : \theta_{[1:m_1]} \in K_{m_1}, \theta_{[m_1+1:m_2-1]} \in K_{m_2-m_1-1}, \theta_{[m_2:n]} \in K_{n-m_2+1} \text{ for some integers } 0 \leq m_1 < m_2 \leq n+1\}.$$

The interpretation is that if  $m_1 = 0$  then the first block is empty. Likewise if  $m_2 = n+1$  then the third block is empty. It is not too difficult to extend Lemma 3.3 to give a bound on the log covering number of all bounded sequences in  $K3$  as shown in our next lemma.

**Lemma 3.4.** *For every  $\epsilon > 0$  we have the following:*

$$\log N(\epsilon, \{\theta \in K_3 \subset \mathbb{R}^n : \max_{1 \leq i \leq n} |\theta_i| \leq B\}) \leq Cn^{1/4} \left(\frac{B}{\epsilon}\right)^{1/2} + 2 \log(n + 2).$$

*Proof.* Fix integers  $0 \leq m_1 < m_2 \leq n + 1$ . Define  $B_1 = \{1 \leq i \leq m_1\}, B_2 = \{m_1 + 1 \leq i \leq m_2 - 1\}, B_3 = \{m_2 \leq i \leq n\}$ . There are three concave pieces and it suffices to cover each of them separately at radius  $\epsilon/\sqrt{3}$  to get an  $\epsilon$  cover for sequences which are concave on  $B_1, B_2$  and  $B_3$  separately. Using Lemma 3.3 on each of these pieces one obtains a log covering number bound which is at most  $3Cn^{1/4}(\frac{B}{\epsilon})^{1/2}$ . Now there are exactly  $\binom{n+2}{2}$  ways of choosing  $m_1$  and  $m_2$ . We could cover each of the spaces defined by a fixed  $m_1, m_2$  at radius  $\epsilon$  separately and take the union of the covers. That would be a cover for  $K_3$  at radius  $\epsilon$ . The log cardinality of this cover is clearly upper bounded by  $Cn^{1/4}(\frac{B}{\epsilon})^{1/2} + \log \binom{n+2}{2}$ . This finishes the proof of the lemma.  $\square$

We now embark on proving a key result which gives an upper bound on the function  $f_{\theta^*}$  in case  $\theta^*$  is a concave monotonic sequence. As mentioned before, since any concave function first increases and then decreases, a critical step is to understand the behaviour of  $f_{\theta^*}$  when  $\theta^*$  is monotonic, in addition to being concave. Recalling the definition of  $f_{\theta^*}$  in (2.3), it is an expected supremum of Gaussian random variables where the supremum is over all  $\theta \in K_n$  which also lie within a Euclidean ball around  $\theta^*$ . As a first step we are only going to take the supremum over all  $\theta \in K_n$  lying within a Euclidean ball around  $\theta^*$  and having a maxima at a fixed index  $k$ . To explain further, let us define the set  $C_k$  of concave sequences with maxima at  $k$  as follows:

$$C_k = \{\theta \in K : \max_{1 \leq i \leq n} \theta_i \leq \theta_k\}.$$

As defined  $C_k$  is a closed convex cone in  $\mathbb{R}^n$ . Also it is clear that  $K_n = \cup_{k=1}^n C_k$ . Our next result is a key lemma controlling the expected Gaussian maxima term  $\mathbb{E}(\sup_{\theta \in C_k : \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle)$  where the supremum is taken over the restricted set  $C_k$  intersected with a Euclidean ball. Our bound would hold uniformly over  $k$ , as a function of  $t$  whenever the underlying  $\theta^*$  is concave and monotonic (non decreasing or non increasing).

**Lemma 3.5.** *Fix a positive integer  $n$  and fix any  $1 \leq k \leq n$ . Also fix a non decreasing concave sequence  $\theta^* \in K_n$ . For all  $t \geq 0$  the following inequality is true for a universal constant  $C$ ,:*

$$\mathbb{E} \left( \sup_{\theta \in C_k : \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle \right) \leq C\sigma(V(\theta^*) + \sigma)^{1/4} n^{1/8} t^{3/4} + 2\sigma \sqrt{2 \log(n + 2)} t + \frac{t^2}{8}.$$

**Remark 3.3.** The conclusion for Lemma 3.5 works even when  $\theta^*$  is concave and non increasing by reasons of symmetry.

*Proof.* Let  $A = \{\theta \in C_k : \|\theta - \theta^*\| \leq t\}$ . Let us define  $A'$  as follows:

$$A' = \{\theta \in K_3 \subset \mathbb{R}^n : \max_{1 \leq i \leq n} \theta_i \leq \theta_n^* + L, \min_{1 \leq i \leq n} \theta_i \geq \theta_1^* - L, \|\theta - \theta^*\| \leq t\}$$

where  $L$  is a fixed positive number to be chosen later.

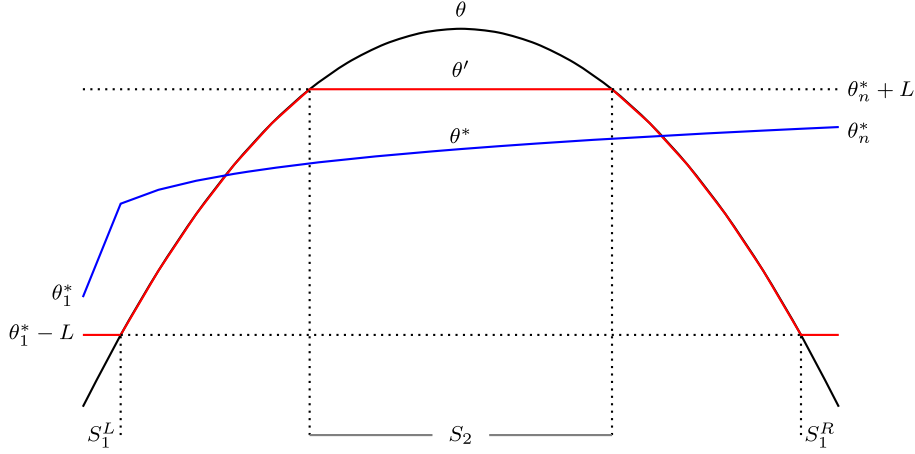


FIG 1. The proof of Lemma 3.5 is based on a truncation  $\theta'$  (in red) of an arbitrary concave sequence  $\theta \in C_k$  (in black) with respect to a fixed monotone increasing and concave sequence  $\theta^*$  (in blue). The tails of  $\theta$  are raised to  $\theta_1^* - L$  over  $S_1^L \cup S_1^R = S_1 = \{i : \theta_i < \theta_1^* - L\}$ . In the interval  $S_2 = \{i : \theta_i > \theta_n^* + L\}$  around the mode  $k$ , the sequence is lowered to the level  $\theta_n^* + L$ .

For any  $\theta \in A$  we will define a truncated version of  $\theta$  belonging in  $A'$  which will be denoted by  $\theta'(\theta)$ . Then we will have the inequality

$$\begin{aligned} \mathbb{E} \sup_{\theta \in A} \langle z, \theta - \theta^* \rangle &\leq \mathbb{E} \sup_{\theta \in A} \langle z, \theta - \theta'(\theta) \rangle + \mathbb{E} \sup_{\theta \in A} \langle z, \theta'(\theta) - \theta^* \rangle \leq \\ \mathbb{E} \sup_{\theta \in A} \langle z, \theta - \theta'(\theta) \rangle &+ \mathbb{E} \sup_{\tilde{\theta} \in A'} \langle z, \tilde{\theta} - \theta^* \rangle. \end{aligned} \quad (3.6)$$

To reduce notational clutter, we will now write  $\theta'(\theta)$  simply as  $\theta'$  while keeping in mind that  $\theta'$  is a function of  $\theta$ .

Fix an arbitrary  $\theta \in A$ . Let us denote  $S_1 = \{i : \theta_i < \theta_1^* - L\}$  and  $S_2 = \{i : \theta_i > \theta_n^* + L\}$ .

We now define  $\theta'$  as follows for each  $1 \leq i \leq n$ :

$$\theta'_i = \theta_i \mathcal{I}\{i \notin S_1 \cup S_2\} + (\theta_1^* - L) \mathcal{I}\{i \in S_1\} + (\theta_n^* + L) \mathcal{I}\{i \in S_2\}$$

where  $\mathcal{I}$  is the usual indicator function.

From the above definition, it is clear that  $\min_{1 \leq i \leq n} \theta'_i \geq \theta_1^* - L$ ,  $\max_{1 \leq i \leq n} \theta'_i \leq \theta_n^* + L$ . Also by construction of  $\theta'$ , we have the following contractive property for any  $1 \leq i \leq n$ :

$$|\theta_i - \theta'_i| \leq |\theta_i - \theta_i^*|.$$

Now by concavity of  $\theta$ , the set  $S_1$  is necessarily a union of at most two intervals of the form  $\{1 \leq i \leq m_1\} \cup \{m_2 \leq i \leq n\}$  for some integers  $0 \leq m_1 < m_2 \leq n+1$ . If  $m_1 = 0$  then the set  $\{1 \leq i \leq m_1\}$  is empty and similarly if  $m_2 = n+1$  then the set  $\{m_2 \leq i \leq n\}$  is empty. As defined,  $\theta'$  is a constant vector on the two

intervals in  $S_1$  and remains a concave vector on the complement of  $S_1$  which is an interval. Hence,  $\theta' \in K3$ . Combining these properties of  $\theta'$  imply that  $\theta' \in A'$ .

We now proceed to control the first term on the right side of the inequality in (3.6). We have the following inequality by definition of  $\theta'$ :

$$\begin{aligned} \sum_{i=1}^n z_i(\theta_i - \theta'_i) &= \sum_{i \in S_1} z_i(\theta_i - \theta'_i) + \sum_{i \in S_2} z_i(\theta_i - \theta'_i) \\ &\leq \sum_{i \in S_1} |z_i|(\theta'_i - \theta_i) + \sum_{i \in S_2} |z_i|(\theta_i - \theta'_i) \\ &= \sum_{i \in S_1} \sum_{j=0}^{\infty} |z_i|(\theta'_i - \theta_i) \mathcal{I}\{2^j L < \theta_1^* - \theta_i \leq 2^{j+1} L\} \\ &\quad + \sum_{i \in S_2} \sum_{j=0}^{\infty} |z_i|(\theta_i - \theta'_i) \mathcal{I}\{2^j L < \theta_i - \theta_n^* \leq 2^{j+1} L\} \\ &\leq \sum_{j=0}^{\infty} 2^{j+1} L \sum_{i \in S_1} |z_i| \mathcal{I}\{2^j L < \theta_1^* - \theta_i \leq 2^{j+1} L\} \\ &\quad + \sum_{j=0}^{\infty} 2^{j+1} L \sum_{i \in S_2} |z_i| \mathcal{I}\{2^j L < \theta_i - \theta_n^* \leq 2^{j+1} L\} \end{aligned} \tag{3.7}$$

where  $\mathcal{I}$  denotes the indicator function and the last inequality follows from the inequalities

$$\theta'_i - \theta_i \leq \theta_1^* - \theta_i \text{ for } i \in S_1 \tag{3.8}$$

$$\theta_i - \theta'_i \leq \theta_i - \theta_n^* \text{ for } i \in S_2. \tag{3.9}$$

Fix a non negative integer  $j$ . We now note that for any  $\theta \in A$  since  $\|\theta - \theta^*\| \leq t$  we have

$$|\{i : 2^j L < |\theta_i - \theta_i^*|\}| \leq \frac{t^2}{2^{2j} L^2} \equiv v_j. \tag{3.10}$$

We now make some further observations. Since  $\theta$  is concave, any set of the form  $\{i : \theta_i < a\}$  for some number  $a$  is necessarily at most a union of two intervals. One interval, if non empty, has to contain the index 1 and the other interval, if non empty has to contain the index  $n$ . Hence there exists integers  $0 \leq w_1 < w_2 \leq n + 1$  such that the following holds:

$$\begin{aligned} \mathcal{I}\{i \in S_1 : 2^j L < \theta_1^* - \theta_i \leq 2^{j+1} L\} &\leq \mathcal{I}\{i \in S_1 : \theta_i < \theta_i^* - 2^j L\} \leq \\ &\mathcal{I}\{1 \leq i \leq w_1\} + \mathcal{I}\{w_2 \leq i \leq n\}. \end{aligned}$$

Also we have

$$\sum_{i=1}^n \mathcal{I}\{i \in S_1 : 2^j L < \theta_1^* - \theta_i \leq 2^{j+1} L\} \leq \sum_{i=1}^n \mathcal{I}\{i \in S_1 : 2^j L < \theta_i^* - \theta_i\} \leq v_j$$

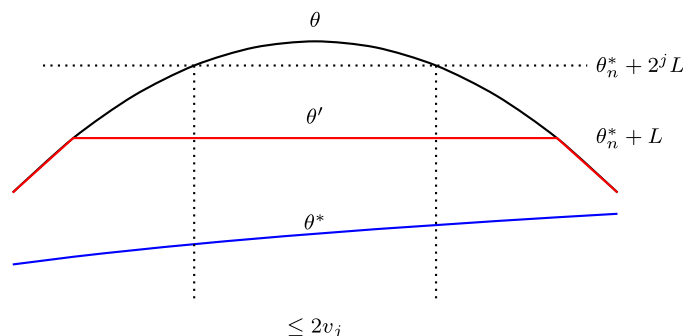


FIG 2. The set  $\{i \in S_2 : 2^j L < \theta_i - \theta_n^*\}$  if it is nonempty, is an interval of length no greater than  $2v_j$ . The figure above indicates this set. Similarly, the set  $\{i \in S_1 : 2^j L < \theta_1^* - \theta_i\}$  is the union of at most two intervals. Each has size no larger than  $v_j$ .

where the first inequality is because  $\theta^*$  is non decreasing and the second inequality is due to (3.10). The last two displays imply the following inequality:

$$\mathcal{I}\{i \in S_1 : 2^j L < \theta_1^* - \theta_i \leq 2^{j+1} L\} \leq \mathcal{I}\{1 \leq i \leq v_j\} + \mathcal{I}\{n - v_j < i \leq n\}. \quad (3.11)$$

Similarly, by concavity of  $\theta$  and since  $\theta \in C_k$  any set of the form  $\{i : \theta_i > a\}$  for some number  $a$  is necessarily an interval containing  $k$  if it is non empty. Hence there exists integers  $1 \leq w_3 \leq k \leq w_4 \leq n$  such that the following holds:

$$\begin{aligned} \mathcal{I}\{i \in S_2 : 2^j L < \theta_i - \theta_n^* \leq 2^{j+1} L\} &\leq \mathcal{I}\{i \in S_2 : \theta_i > \theta_n^* + 2^j L\} \leq \\ &\mathcal{I}\{w_3 \leq i \leq w_4\}. \end{aligned}$$

Also we have

$$\sum_{i=1}^n \mathcal{I}\{i \in S_2 : 2^j L < \theta_i - \theta_n^* \leq 2^{j+1} L\} \leq \sum_{i=1}^n \mathcal{I}\{i \in S_2 : 2^j L < \theta_i - \theta_i^*\} \leq v_j$$

where the first inequality is because  $\theta^*$  is non decreasing and the second inequality is due to (3.10). The last two displays imply the following inequality:

$$\mathcal{I}\{i \in S_2 : 2^j L \leq \theta_i - \theta_n^* \leq 2^{j+1} L\} \leq \mathcal{I}\{k - v_j < i < k + v_j\}. \quad (3.12)$$

Now using the inequalities (3.11) and (3.12) and applying them in (3.7) we obtain

$$\sum_{i=1}^n (\theta_i - \theta'_i) z_i \leq \sum_{j=0}^{\infty} 2^{j+1} L \left( \sum_{i=1}^{v_j} |z_i| + \sum_{i=k-v_j+1}^{k+v_j-1} |z_i| + \sum_{i=n-v_j+1}^n |z_i| \right).$$

Note that in our argument  $\theta$  is an arbitrary element in  $A$  and the upper bound in the previous inequality does not depend on the choice of  $\theta$ . Therefore, using

the fact  $\mathbb{E}|z_i| = \sigma\sqrt{2/\pi} \leq \sigma$  alongwith (3.10), the last display gives us

$$\mathbb{E} \sup_{\theta \in A} \sum_{i=1}^n (\theta_i - \theta'_i) z_i \leq \sum_{j=0}^{\infty} 2^{j+1} L \sigma (4v_j) \leq \frac{4t^2 \sigma}{L} \sum_{j=0}^{\infty} 2^{j+1-2j} = \frac{16t^2 \sigma}{L}.$$

We now set  $L = 128\sigma$  to finally get

$$\mathbb{E} \sup_{\theta \in A} \langle z, \theta - \theta' \rangle \leq \frac{t^2}{8} \tag{3.13}$$

Now we come to controlling the second term in the right side of (3.6). Setting  $\mathbb{F} = A'$  in the chaining result in Theorem 3.1 we get the upper bound

$$\mathbb{E} \sup_{\tilde{\theta} \in A'} \langle z, \tilde{\theta} - \theta^* \rangle \leq 12\sigma \int_0^{2t} \sqrt{\log N(\epsilon, A')} d\epsilon. \tag{3.14}$$

The upper limit of the integral is the diameter of  $A'$  which is at most  $2t$ . By definition of  $A'$  we can now apply Lemma 3.4 with  $B = \frac{V(\theta^*)}{2} + L$  to obtain

$$\log N(\epsilon, A') \leq Cn^{1/4} \left(\frac{B}{\epsilon}\right)^{1/2} + 2 \log(n+2).$$

Using (3.14) and integrating the above expression gives us

$$\mathbb{E} \sup_{\theta' \in A'} \langle z, \theta' - \theta^* \rangle \leq C\sigma \left(\frac{V(\theta^*)}{2} + L\right)^{1/4} n^{1/8} t^{3/4} + 2\sigma \sqrt{2 \log(n+2)} t$$

where we have also used the elementary inequality  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  for any two positive numbers  $a, b$ . Combining the last display with (3.13) and (3.6) finishes the proof of the lemma.  $\square$

Our next step is to extend Lemma 3.5 to the case when the supremum of the Gaussian inner products are taken over all of  $K_n$  (not just  $C_k$ ) intersected with a Euclidean ball. This result is presented in our next lemma.

**Lemma 3.6.** *Fix a positive integer  $n$ . Also fix a concave sequence  $\theta^*$  which is non decreasing sequence or non increasing. For all  $t \geq 0$  the following inequality is true for a universal constant  $C$ ,*

$$\mathbb{E} \left( \sup_{\theta \in K_n: \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle \right) \leq C\sigma \{ (V(\theta^*) + \sigma)^{1/4} n^{1/8} t^{3/4} + \sigma \sqrt{\log nt} \} + \frac{t^2}{8}.$$

*Proof.* We prove the lemma when  $\theta^*$  is a concave non decreasing sequence. The proof when  $\theta^*$  is concave non increasing is analogous. For each  $1 \leq k \leq n$  and  $t > 0$  define the random variables

$$X_k(t) = \sup_{\theta \in C_k: \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle.$$

We first note that

$$\sup_{\theta \in K_n: \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle = \max_{1 \leq k \leq n} X_k(t).$$

Applying Lemma A.3 (see appendix) we see that the random variables  $X_k(t)$  are lipschitz functions of  $z$  with lipschitz constant  $t$ . Hence using the Gaussian Concentration Theorem 3.2 for lipschitz functions we get for all  $x > 0$  and all  $1 \leq k \leq n$ ,

$$P(X_k(t) \leq \mathbb{E}X_k(t) + t\sigma x) \leq \exp\left(-\frac{x^2}{2}\right).$$

A standard argument involving maxima of random variables with Gaussian like tails is given in Lemma A.2 for the sake of completeness. Using this lemma and the last display we finally get for a universal constant  $C$ ,

$$\mathbb{E} \max_{1 \leq k \leq n} X_k(t) \leq \max_{1 \leq k \leq n} \mathbb{E}X_k(t) + C\sigma t\sqrt{\log n}.$$

Now Lemma 3.5 gives us an upper bound on the term  $\max_{1 \leq k \leq n} \mathbb{E}X_k(t)$  because the upper bound in Lemma 3.5 does not depend on  $k$ . Using this upper bound alongwith the last display finishes the proof of the lemma.  $\square$

We are now finally ready to prove Proposition 3.1. The main idea is to use the fact that any concave sequence first increases and then decreases and use Lemma 3.6.

*Proof of Proposition 3.1.* Let  $1 \leq i^* \leq n$  be such that  $\theta^* \in C_{i^*}$ . Let  $\theta^* = (\theta_1^*, \theta_2^*)$  where  $\theta_1^*$  is an  $i^*$  dimensional vector and  $\theta_2^*$  is an  $n - i^*$  dimensional vector. Similarly, let  $z = (z_1, z_2)$ . We can write

$$\begin{aligned} \mathbb{E}\left(\sup_{\theta \in K_n: \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle\right) &\leq \mathbb{E}\left(\sup_{\theta \in K_{i^*}: \|\theta - \theta_1^*\| \leq t} \langle z_1, \theta - \theta_1^* \rangle\right) \\ &\quad + \mathbb{E}\left(\sup_{\theta \in K_{n-i^*}: \|\theta - \theta_2^*\| \leq t} \langle z_2, \theta - \theta_2^* \rangle\right). \end{aligned}$$

There are two terms on the right side of the above inequality. Let us first bound the first term on the right side. The second term can be bounded exactly in the same way. Since  $\theta_1^*$  is non decreasing, we can use Lemma 3.6 to obtain for all  $t \geq 0$ :

$$\mathbb{E}\left(\sup_{\theta \in K_{i^*}: \|\theta - \theta_1^*\| \leq t} \langle z_1, \theta - \theta_1^* \rangle\right) \leq C\sigma\{(V(\theta^*) + \sigma)^{1/4} n^{1/8} t^{3/4} + \sqrt{\log nt}\} + \frac{t^2}{8}.$$

where we also use the fact that  $(i^*)^{1/8} \leq n^{1/8}$ . We upper bound the second term by using Lemma 3.6 exactly similarly. We then obtain for an appropriate universal constant  $C$ ,

$$\begin{aligned} f_{\theta^*}(t) &= \mathbb{E}\left(\sup_{\theta \in K_n: \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle\right) \\ &\leq C\sigma\{(V(\theta^*) + \sigma)^{1/4} n^{1/8} t^{3/4} + \sqrt{\log nt}\} + \frac{t^2}{4}. \end{aligned}$$

This finishes the proof.  $\square$



### 3.3. Comparison with [8]

Let us now compare Theorem 2.1 with the risk bound obtained in [8] as stated in (1.6). The first issue is the absence of the  $\log(en)$  term in (2.2). We now outline the differences in both the techniques which lead to a removal of the log factor. It will be simpler to discuss the differences if we consider the case when  $\theta^* = (I - P_L)\theta^*$ . The main ingredient to prove a risk bound at a concave sequence  $\theta^*$  is to obtain a good upper bound to the function  $f_{\theta^*}(t)$  as defined in (2.3). The authors in [8] directly use Dudley's chaining inequality which requires upper bounding  $\log N(\epsilon, B(\theta^*, t) \cap K_n)$  where  $B(\theta^*, t)$  is the Euclidean ball in  $\mathbb{R}^n$  around  $\theta^*$  of radius  $t > 0$ . The key step in their proof is to obtain an upper bound to the log covering number when  $\theta^*$  is essentially the zero vector. For a general  $\theta^*$  they then use the triangle inequality to say that any vector in a  $t$  ball around  $\theta^*$  is within a  $t + \|\theta^*\|$  ball around the origin and then they use their covering number bound for this larger ball which is now centred at the origin. The result we are referring to is Theorem 3.1 in [8]. Their log covering number upper bound is as follows:

$$\log N(\epsilon, B(\theta^*, t) \cap K_n) \leq C \log(en)^{5/4} \frac{(t^2 + \|\theta^*\|^2)^{1/4}}{\epsilon^{1/2}}. \quad (3.15)$$

The use of the chaining inequality to the above leads to the worst case risk bound obtained in [8] of the form given in equation 1.6 in the current manuscript. This argument of bounding the covering numbers will not be able to remove the log covering number due to the presence of the logarithmic term on the right side of the last display. Note again that the authors in [8] were directly bounding the covering number of the *Euclidean* ball around  $\theta^*$  intersected with the set of concave sequences. To get around this hurdle, the current manuscript does an extra peeling step to upper bound the function  $f_{\theta^*}(t)$  and hence avoids computing the log covering number of the *Euclidean* ball around  $\theta^*$ . Instead, after the peeling step, we essentially show that we only need to compute a log covering number for a  $l_\infty$  ball in  $K_n$  around  $\theta^*$  of constant radius. Covering numbers for a  $l_\infty$  ball in  $K_n$  around  $\theta^*$  are available and does not suffer from logarithmic factors.

The second issue is that we no longer get the factor  $G(\theta^*)$  in our risk bound and instead we get the factor  $V(\theta^*)$ . Again assume  $\theta^* = (I - P_L)\theta^*$ . For example, any  $\theta^*$  which is mean zero and is symmetric in the sense  $\theta_i^* = \theta_{n+1-i}^*$  satisfies  $\theta^* = (I - P_L)\theta^*$ . As a simple illustration, let us take  $\theta^* = (-(n-2), 2, \dots, 2, (n-2))$ . The previous bound available in [8] is of the form  $C\sigma^{8/5}G(\theta^*)^{1/5}(\log(n)/n)^{-4/5}$  for an absolute constant  $C > 0$  whereas our bound is of the form  $C\sigma^{8/5}\max\{\sigma, V(\theta^*)\}^{2/5}n^{-4/5}$ . In general, for any  $\theta^*$  satisfying  $\theta^* = (I - P_L)\theta^*$  we have the relation

$$G(\theta^*) \leq \max\{\sigma, V(\theta^*)\}^2.$$

In our chosen example,  $G(\theta^*) = O(1)$  and  $V(\theta^*) = O(n)$  and hence the bound given in Theorem 2.1 scales like  $n^{-2/5}$  which is worse than the rate  $(\log(n)/n)^{-4/5}$  given by the risk bound of [8].

However, since there are only two "big" terms in  $\theta^*$ , it is possible to prove a risk bound scaling like  $n^{-4/5}$  without log factors by using the techniques developed in this manuscript. To illustrate this fact, again we control the expected supremum term  $f_{\theta^*}(t) = \mathbb{E} \sup_{\theta \in K_n: \|\theta - \theta^*\| \leq t} \sum_{i=1}^n \langle Z_i, \theta - \theta^* \rangle$ . The inner product can be broken up as

$$\langle Z, \theta - \theta^* \rangle = Z_1 \theta_1 + \langle Z[2 : (n-1)], (\theta - \theta^*)[2 : (n-1)] \rangle + Z_n \theta_n$$

and now taking supremum over  $\theta \in K_n$  gives us

$$\begin{aligned} f_{\theta^*}(t) &\leq \mathbb{E} \sup_{\theta_1: |\theta_1 - \theta_1^*| \leq t} Z_1(\theta_1 - \theta_1^*) \\ &\quad + \mathbb{E} \sup_{\theta \in K_{n-2}: \|\theta - v\| \leq t} \sum_{i=2}^{n-1} Z_i(\theta_i - \theta_i^*) \\ &\quad + \mathbb{E} \sup_{\theta_n: |\theta_n - \theta_n^*| \leq t} Z_n(\theta_n - \theta_n^*) \end{aligned} \tag{3.16}$$

where  $v$  denotes  $\theta^*[2 : (n-1)]$ . The first and the third term on the right side of the above inequality can now be bounded by  $Ct\sigma$  for some constant  $C$ . Proposition 3.1 applies to the middle term where the same proof as given in this manuscript goes through. The rest of the proof can now be carried through just as in the proof of Theorem 2.1 in the manuscript to obtain a very similar risk bound which we do not carry out in full detail here. The only important change would be that the  $V(\theta^*)$  term in the risk bound would be replaced by  $V(\theta^*[2 : (n-1)])$  which is actually 0 for this particular example. This simple way of controlling the expected Gaussian suprema can actually be carried through for the first and last stretch of  $m$  entries where  $m = O(n^{1/5})$  in order to get a risk bound scaling like  $O(n^{-4/5})$ . Again if we break the expected Gaussian suprema into three terms, the first and the last term can be bounded by  $Cn^{1/10}t$  by an application of the Cauchy Schwarz inequality. The middle term can be handled in exactly the same way as in proof of Theorem 2.1. Again the only important change would be that the  $V(\theta^*)$  term in the risk bound would now be replaced by  $V(\theta^*[m+1 : (n-m)])$ . This is one easy way to prove risk bounds for  $\theta^*$  where the  $V(\theta^*)$  is of a higher order than  $G(\theta^*)$  but only so because of some large values in the last stretch of entries where this last stretch can be tolerated to have length of the order  $O(n^{1/5})$  and the  $V$  and  $G$  function are both  $O(1)$  for the middle stretch of  $\theta^*$ .

A natural question which now arises is whether one can prove a risk bound without the log factor and with the better constant  $G(\theta^*)$  for all  $\theta^* \in K_n$ ? It seems hard to answer this question in general using the techniques of this manuscript. Infact this question can also be asked for the LSE  $\hat{\theta}$  in Isotonic regression where a bound for  $R(\hat{\theta}, \theta^*)$  of the form  $V(\theta^*)/n^{2/3}$  upto a universal constant constant is available when  $V(\theta^*) > 0$ ; see [16]. As far as I am aware, it is not known whether one can prove the risk bound  $G(\theta^*)/n^{2/3}$  upto a universal constant in Isotonic regression too without suffering log factors.

#### 4. Model misspecification

Our risk analysis actually extends to the case of model misspecification, where  $\theta^*$  is now not necessarily in  $K_n$ . A natural quantity in the misspecified case for measuring the performance of an estimator  $\hat{\theta}$  is to control what is called the regret, defined as  $\|\hat{\theta} - \theta^*\|^2 - \min_{\theta \in K_n} \|\theta - \theta^*\|^2$ . The goal of this section is to prove the next theorem upper bounding the regret of the LSE. This theorem gives an oracle risk bound generalizing Theorem 2.1 to the case when the true underlying sequence  $\theta^*$  is not necessarily concave. Recall that  $\Pi_C$  denotes the projection operator to any closed convex cone  $C$  and for any vector  $\theta \in \mathbb{R}^n$  the range of the vector is denoted by  $V(\theta) = \max_{1 \leq i \leq n} \theta_i - \min_{1 \leq i \leq n} \theta_i$ . For simplicity of exposition, we state the theorem below for  $\theta^*$  satisfying  $(I - P_L)\theta^* = \theta^*$ . The general case can be dealt with easily.

**Theorem 4.1.** *Fix any positive integer  $n$ . Fix any  $\theta^* \in \mathbb{R}^n \cap L^\perp$ . Let  $K = K_n$  for simplicity. Let  $H(\theta^*) = V(\Pi_K(\theta^*))$ . Fix any  $x > 0$ . The following upper bound on the risk holds for a universal constant  $C$  with probability greater than  $1 - \exp(-x)$ :*

$$\frac{1}{n} \|\hat{\theta} - \theta^*\|^2 \leq \frac{1}{n} \|\Pi_K(\theta^*) - \theta^*\|^2 + \max\left\{C\sigma^{8/5}(H(\theta^*) + \sigma)^{2/5}n^{-4/5}, C\sigma^2 \frac{\log n}{n}, \frac{8\sigma^2 x}{n}\right\}$$

Moreover, the above high probability risk bound also immediately implies a risk bound in expectation for an appropriate universal constant  $C$ :

$$R(\hat{\theta}, \theta^*) \leq \frac{1}{n} \|\Pi_K(\theta^*) - \theta^*\|^2 + C\sigma^{8/5}(H(\theta^*) + \sigma)^{2/5}n^{-4/5} + C\sigma^2 \frac{\log n}{n}.$$

**Remark 4.1.** The above theorem implies that in case the true underlying sequence  $\theta^*$  is non concave then the regret of the LSE converges at the rate  $n^{-4/5}$  upto constant factors. Infact by the usual Pythagoras inequality for projection onto convex sets, we have the inequality

$$\|\hat{\theta} - \Pi_K(\theta^*)\|^2 \leq \|\hat{\theta} - \theta^*\|^2 - \|\Pi_K(\theta^*) - \theta^*\|^2.$$

The above equation states that  $\|\hat{\theta} - \Pi_K(\theta^*)\|^2$  is upper bounded by the regret. Therefore, Theorem 4.1 implies a risk bound for  $\|\hat{\theta} - \Pi_K(\theta^*)\|^2$  as well.

The main tool in proving Theorem 4.1 is actually a slightly stronger version of Lemma 3.1 and is a direct consequence of Theorem 12 in Bellec [2] when applied to the cone  $K_n$ .

**Lemma 4.1** (Bellec). *Let  $\theta^* \in \mathbb{R}^n$  and let  $K = K_n$ . Let  $s > 0$  be such that  $f_{\Pi_K(\theta^*)}(s) \leq \frac{s^2}{2}$ . Then for any  $x > 0$  the following inequality holds with probability greater than  $1 - \exp(-x)$ ,*

$$\|\hat{\theta} - \theta^*\|^2 \leq \|\theta^* - \Pi_K(\theta^*)\|^2 + 2 \max\{s^2, 8\sigma^2 x\}.$$

Here  $z$  is a  $n$  dimensional gaussian random vector with mean zero and covariance matrix  $\sigma^2 I$ .

Using the above Lemma, the proof of Theorem 4.1 goes through verbatim as in the proof of Theorem 2.1.

### Appendix

**Lemma A.2.** *Let  $X_1, X_2, \dots, X_n$  be random variables such that the following holds for every  $1 \leq i \leq n$  and  $a > 0$ ,*

$$P(X_i \geq \mathbb{E}X_i + ax) \leq \exp\left(-\frac{x^2}{2}\right) \quad \forall x \geq 0. \tag{A.1}$$

Then the following is true:

$$\mathbb{E} \max_{1 \leq i \leq n} X_i \leq \max_{1 \leq i \leq n} \mathbb{E}X_i + a \left( \sqrt{2 \log n} + \sqrt{2\pi} \right).$$

*Proof.* Let  $m = \max_{1 \leq i \leq n} \mathbb{E}X_i$ . Define  $Y_i = X_i - m$ . Then by (A.1) we have sub gaussian tail behaviour for every  $1 \leq i \leq n$ ,

$$P(Y_i \geq x) \leq \exp\left(-\frac{x^2}{2a^2}\right) \quad \forall x \geq 0. \tag{A.2}$$

Now by defining  $Z_i = \max\{Y_i, 0\}$  we certainly have  $\mathbb{E} \max Y_i \leq \mathbb{E} \max Z_i$ . Also for any  $x \geq 0$  we have  $P(\max Z_i \geq x) = P(\max Y_i \geq x)$ . Combining these two and using the tail integral formula for the expectation of a non negative random variable we obtain

$$\mathbb{E} \max Y_i \leq \int_0^\infty P(\max Y_i \geq x) dx.$$

Therefore, now we can write

$$\begin{aligned} \mathbb{E} \max Y_i &\leq \int_0^\infty P(\max Y_i \geq x) dx \leq \int_0^\infty \min\{n \exp(-\frac{x^2}{2a^2}), 1\} dx = \\ &\int_0^{a\sqrt{2 \log n}} 1 dx + \int_{a\sqrt{2 \log n}}^\infty n \exp(-\frac{x^2}{2a^2}) dx \end{aligned}$$

where the inequalities follow from the previous display alongwith (A.2). The rest follows from simple integral calculus. Now a standard fact about Gaussian tails give us the following inequality

$$\int_{a\sqrt{2 \log n}}^\infty \exp(-\frac{x^2}{2a^2}) \leq \sqrt{2\pi} \frac{a}{n}.$$

The last two displays finish the proof of the lemma. □

**Lemma A.3.** *Let  $A \subset \mathbb{R}^n$  be a closed set. Fix any  $\theta^* \in \mathbb{R}^n$  and  $t > 0$ . Define the function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  as follows:*

$$f(z) = \sup_{\theta \in A: \|\theta - \theta^*\| \leq t} \langle z, \theta - \theta^* \rangle.$$

Then  $f$  is a Lipschitz function of  $z$  with Lipschitz constant  $t$ .

*Proof.* Since the function  $\langle z, \theta - \theta^* \rangle$  is a continuous function of  $z$  and the set  $\{\theta \in A : \|\theta - \theta^*\| \leq t\}$  is a compact set, its supremum is attained at  $\tilde{\theta}$  say. Then we have

$$\begin{aligned} f(z) - f(z') &= \langle z, \tilde{\theta} - \theta^* \rangle - \langle z', \tilde{\theta} - \theta^* \rangle \\ &= \langle z - z', \tilde{\theta} - \theta^* \rangle \leq \|z - z'\| \|\tilde{\theta} - \theta^*\| \leq t \|z - z'\|. \end{aligned}$$

The first inequality is because we set  $\theta = \tilde{\theta}$  instead of taking supremum over  $\theta$ , the second inequality is just the Cauchy Schwarz inequality and the last inequality follows from the fact that  $\|\tilde{\theta} - \theta^*\| \leq t$  by the choice of  $\tilde{\theta}$ . This finishes the proof of the lemma.  $\square$

### Acknowledgements

The author thanks Adityanand Guntuboyina, John Lafferty, Bodhisattva Sen and Yuancheng Zhu for helpful discussions. The author also thanks the anonymous referees for their helpful suggestions.

### References

- [1] M. Ayer, H. D. Brunk, G. M. Ewing, W.T. Reid, E. Silverman, et al. An empirical distribution function for sampling with incomplete information. *The annals of mathematical statistics*, 26(4):641–647, 1955. [MR0073895](#)
- [2] P. C. Bellec. Sharp oracle inequalities for least squares estimators in shape restricted regression. *arXiv preprint arXiv:1510.08029*, 2015.
- [3] S. Chatterjee, A. Guntuboyina, and B. Sen, et al. On risk bounds in isotonic and other shape restricted regression problems. *The Annals of Statistics*, 43(4):1774–1800, 2015. [MR3357878](#)
- [4] S. Chatterjee. A new perspective on least squares under convex constraint. *The Annals of Statistics*, 42(6):2340–2381, 2014. [MR3269982](#)
- [5] D. Dryanov. Kolmogorov entropy for classes of convex functions. *Constructive Approximation*, 30:137–153, 2009. [MR2519658](#)
- [6] L. Dümbgen, S. Freitag, and G. Jongbloed. Consistency of concave regression with an application to current-status data. *Math. Methods Statist.*, 13(1):69–81, 2004. ISSN 1066-5307. [MR2078313](#)
- [7] P. Groeneboom, G. Jongbloed, and J. A. Wellner. A canonical process for estimation of convex functions: The “invelope” of integrated brownian motion  $+t^4$ . *Annals of Statistics*, 29:1620–1652, 2001a. [MR1891741](#)
- [8] A. Guntuboyina and B. Sen. Global risk bounds and adaptation in univariate convex regression. *Probability Theory and Related Fields*, 163(1-2):379–411, 2015. [MR3405621](#)
- [9] D.L. Hanson and G. Pledger. Consistency in concave regression. *Ann. Statist.*, 4(6):1038–1050, 1976. ISSN 0090-5364. [MR0426273](#)
- [10] C. Hildreth. Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.*, 49:598–619, 1954. ISSN 0162-1459. [MR0065093](#)

- [11] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, 1302–1338, 2000. [MR1805785](#)
- [12] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001. [MR1849347](#)
- [13] E. Mammen. Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.*, 19(2):741–759, 1991. ISSN 0090-5364. [doi:10.1214/aos/1176348118](#). [MR1105842](#)
- [14] T. Robertson, F.T. Wright, and R.L. Dykstra. *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 1988. ISBN 0-471-91787-7. [MR0961262](#)
- [15] S. Van de Geer. *Applications of Empirical Process Theory*. Cambridge University Press, 2000. [MR1739079](#)
- [16] C.-H. Zhang. Risk bounds in isotonic regression. *Ann. Statist.*, 30(2):528–555, 2002. ISSN 0090-5364. [doi:10.1214/aos/1021379864](#). [MR1902898](#)