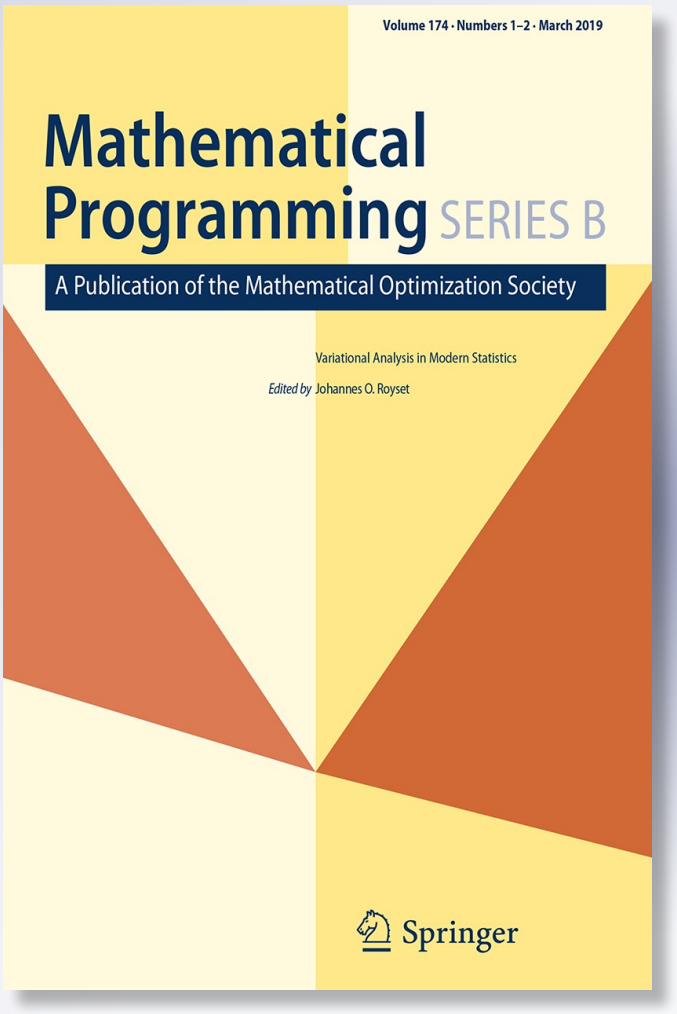
Concave regression: value-constrained estimation and likelihood ratio-based inference

Charles R. Doss

Mathematical Programming
A Publication of the Mathematical Optimization Society

ISSN 0025-5610
Volume 174
Combined 1-2

Math. Program. (2019) 174:5-39
DOI 10.1007/s10107-018-1338-5



Springer

Mathematical Programming (2019) 174:5–39
https://doi.org/10.1007/s10107-018-1338-5

FULL LENGTH PAPER

Series B

Concave regression: value-constrained estimation and likelihood ratio-based inference

Charles R. Doss[1]

Received: 12 April 2017 / Accepted: 25 September 2018 / Published online: 12 October 2018

Abstract

We propose a likelihood ratio statistic for forming hypothesis tests and confidence intervals for a nonparametrically estimated univariate regression function, based on the shape restriction of concavity (alternatively, convexity). Dealing with the likelihood ratio statistic requires studying an estimator satisfying a null hypothesis, that is, studying a concave least-squares estimator satisfying a further equality constraint. We study this null hypothesis least-squares estimator (NLSE) here, and use it to study our likelihood ratio statistic. The NLSE is the solution to a convex program, and we find a set of inequality and equality constraints that characterize the solution. We also study a corresponding limiting version of the convex program based on observing a Brownian motion with drift. The solution to the limit problem is a stochastic process. We study the optimality conditions for the solution to the limit problem and find that they match those we derived for the solution to the finite sample problem. This allows us to show the limit stochastic process yields the limit distribution of the (finite sample) NLSE. We conjecture that the likelihood ratio statistic is asymptotically pivotal, meaning that it has a limit distribution with no nuisance parameters to be estimated, which makes it a very effective tool for this difficult inference problem. We provide a partial proof of this conjecture, and we also provide simulation evidence strongly supporting this conjecture.

Mathematics Subject Classification Primary 62G08; Secondary 62G15 · 62G10 · 62G20

Supported in part by NSF Grant DMS-1712664.

✉ Charles R. Doss
cdoss@stat.umn.edu

[1] School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church Street SE, Minneapolis, MN 55455, USA

1 Introduction

In nonparametric density, regression, or other function estimation, forming hypothesis tests and confidence intervals is important but often challenging. For nonparametric estimators to be effective, they are generally tuned so as to balance their bias and variance (perhaps asymptotically). However, having non-negligible asymptotic bias is problematic for doing inference, since the bias must then be assessed to do honest and efficient inference. One approach is to ignore the bias (e.g., Chapter 5.7 of [47]), although this is clearly problematic. Often the bootstrap [18] can be used for inference in complicated problems, but it is frequently a poor estimate of bias and so requires corrections or modifications. Such corrections have been implemented in a large variety of cases. For instance in forming confidence intervals for a density function, one approach is to undersmooth a kernel density estimator and then use the bootstrap [25]. However the undersmoothed estimator used for the confidence interval is then different from that which would be optimal for pure estimation, and requires stronger smoothness assumptions than would be required for just estimation. Importantly, the inference is still dependent on a tuning parameter (the bandwidth), whose optimal selection can be challenging, can lead to different inferences for different users, and can add another layer of computational burden.

These issues motivate an alternative approach to nonparametric function estimation and inference, which relies on assumptions based on shape constraints and which often does not suffer from the above problems. Here we consider the regression setup,

$$Y_{n,i} = r_0(x_{n,i}) + \epsilon_{n,i}, \quad i = 1, \ldots, n, \tag{1}$$

where $Y_{n,i} \in \mathbb{R}$, we assume that the univariate predictor variables $x_{n,i}$ are fixed, and $\epsilon_{n,i}$ are independent and identically distributed (i.i.d.) with mean 0, and $Ee^{t\epsilon_{n,i}^2} < \infty$ for some $t > 0$. We assume that the target of estimation, $r_0 : \mathbb{R} \to \mathbb{R}$, is concave. (Concave regression is equivalent to convex regression by taking $-Y_{n,i}$ as our responses; we will sometimes use "concave/convex regression" to mean either concave regression or convex regression since they are equivalent.) As will be discussed in greater detail below, concave/convex regression estimators are solutions to convex programs, and so they have very different properties than many other nonparametric regression estimators such as kernel-based ones. Concave/convex regression estimation arises in a truly vast number of settings. It seems to have originally arisen in the econometrics literature [28]. As noted by Hildreth [28], in classical economic theory

> utility functions are usually assumed to be concave; marginal utility is often assumed to be convex; and functions representing productivity, supply, and demand curves are often assumed to be either concave or convex.

(The example worked through by Hildreth [28] is on production function estimation.) Related examples in finance also exhibit convexity restrictions (Aït-Sahalia and Duarte [1] study stock option pricing). Concavity/convexity also arises in operations research, where the concavity/convexity often arises theoretically, and then conveniently makes optimization of the estimated function very efficient [26,32,44,45]. See [37] and ref-

erences therein for further examples of uses of concavity restrictions. There have been a variety of works that have considered concave/convex regression, in the literatures of different fields, with most of the focus being on estimation [2,7,27,31,33,39,41]. Meyer [35,36] and Wang and Meyer [46] and consider spline-based approaches to concave regression estimation, and study testing the hypothesis of linearity against concavity/convexity, and testing the hypothesis of concavity/convexity against a general smooth alternative. Mammen [34] finds the rates of convergence of the univariate LSE, and [23,24] find its limit distribution. Algorithms for computing estimators in concave regression settings (sometimes combined with other constraints) have been studied by Dent [11], Dykstra [17], Fraser and Massam [20], Hudson [29], Meyer [35,36] and Wu [48]. In [10], the authors study upper and lower bounds for the lengths of confidence intervals (with a fixed coverage probability) for concave regression, but we do not know of any practical implementation for the intervals they study. In the Gaussian white noise model, Dümbgen [15] studies multiscale confidence bands (rather than pointwise intervals) for a concave function. (Confidence bands can of course be used for pointwise confidence intervals but will be unnecessarily long.)

The model for r_0, based only on the assumption that r_0 is concave, is nonparametric and infinite dimensional. However, it is still possible to estimate r_0 directly via least-squares, as in finite dimensional problems. We let

$$\widehat{r}_n := \operatorname*{argmin}_r \phi_n(r) := \operatorname*{argmin}_r \frac{1}{2}\sum_{i=1}^{n}\big(Y_{n,i} - r(x_{n,i})\big)^2 \tag{2}$$

where the argmin is taken over all concave functions $r\colon \mathbb{R} \to \mathbb{R}$. Perhaps surprisingly, minimizing the least-squares objective function over the class of all functions constrained only to be concave admits a solution that is uniquely specified at the data points. It is possible for the solution to be not uniquely specified at some other points, so we take $\widehat{r}_n$ to be piecewise linear between the $x_{n,i}$'s [28]. The limit distribution of the estimator at a fixed point x has been obtained (under a second derivative assumption and uniformity conditions on the design of the $x_{n,i}$'s) by Groeneboom et al. [24], who show that

$$d(r_0)n^{2/5}\,(\widehat{r}_n(x_0) - r_0(x_0)) \to_d U \tag{3}$$

where U is a universal limit distribution (meaning it does not depend on r_0), and $d(r_0) := (24/\sigma^4|r_0''(x_0)|)^{1/5}$, where $\sigma^2 = \mathrm{Var}(\epsilon_{n,i})$. (In fact, $U \equiv \widehat{r}_{1,1}(0)$ where $\widehat{r}_{1,1}$ is described below in Theorem 6.) We use $g' \equiv g^{(1)}$, $g'' \equiv g^{(2)}$, and $g^{(i)}$ to refer to the first, second, and ith derivatives of an appropriately differentiable function g.

One might attempt to directly use the limit result (3) as the basis for inference about $r_0(x_0)$. However, the limit distribution depends on $r_0''(x_0)$, and so using (3) requires somehow estimating $r_0''(x_0)$, which leads to many of the problems described in the first paragraph of this paper. We avoid this, rather pursuing a hypothesis test approach based on a likelihood ratio statistic (LRS), and using that to develop a confidence interval. (Here 'likelihood ratio' is a slight abuse of terminology, since it will be a likelihood ratio only if the $\epsilon_{n,i}$ are Gaussian, which we do not assume; the LRS could alternatively be referred to as a residual-sum-of-squares statistic.) We will consider the hypothesis test

$$H_0 : r_0(x_0) = y_0 \quad \text{against} \quad H_1 : r_0(x_0) \neq y_0 \tag{4}$$

for $(x_0, y_0) \in \mathbb{R}^2$ fixed. To form confidence intervals, we will invert the hypothesis test: assume we reject H_0 when $2 \log \lambda_n(y_0) > d_\alpha$ for a statistic $2 \log \lambda_n(y_0)$ (to be discussed shortly) and some critical value d_α, $\alpha \in (0, 1)$. Then the corresponding confidence interval is

$$\{y : 2 \log \lambda_n(y) \leq d_\alpha\}. \tag{5}$$

(Since y is univariate, the confidence interval can be computed by computing the test on a grid of y values.)

The statistic $2 \log \lambda_n(y_0)$ which we will study is based on a ratio statistic $\lambda_n(y_0)$ which depends on a 'null hypothesis statistic' and on an 'alternative hypothesis statistic.' The null hypothesis statistic will depend on a least-squares estimator (LSE) of a concave regression function, $\widehat{r}_n^0$, that is further constrained so that $\widehat{r}_n^0(x_0) = y_0$ where y_0 is fixed: $\widehat{r}_n^0 := \text{argmin}_r \sum_{i=1}^n (Y_{n,i} - r(x_{n,i}))^2$ where the argmin is over concave functions r satisfying $r(x_0) = y_0$. We refer to this estimator as the 'null hypothesis least-squares estimator' (NLSE). The 'alternative hypothesis statistic' depends on $\widehat{r}_n$, which we thus refer to as the 'alternative hypothesis least-squares estimator' (ALSE). With these two estimators in hand, we define our statistic by

$$2 \log \lambda_n \equiv 2 \log \lambda_n(y_0) := 2 \left(\phi_n \left(\widehat{r}_n^0 \right) - \phi_n \left(\widehat{r}_n \right) \right) \tag{6}$$

with ϕ_n defined in (2).

One of the major benefits to using LRS's is that their limit distribution often does not depend on nuisance parameters. In regular parametric problems, two times the log of the LRS is asymptotically χ_k^2, where k is the reduction in parameter dimension in going from the alternative hypothesis to the null hypothesis. Notably, this chi-squared distribution is universal, meaning it is the same limit distribution regardless of what underlying parameter is the true one, so no nuisance parameters need to be estimated to perform inference, which can make inference more simple and more efficient. (In this case, one says that log likelihood ratios are (asymptotically) pivotal, or that they satisfy the Wilks phenomenon.)

In our shape-constrained setting, LRS's can be challenging to analyze theoretically. However, such analysis has been successful in some cases. Banerjee and Wellner [6] and Groeneboom and Jongbloed [22] study LRS's based on monotonicity shape constraints, and [13,14] consider an LRS based on the concavity shape constraint. The estimators underlying both of these tests are maximum likelihood estimators, and they do not require any tuning parameter selection. The LRS's were shown to have asymptotic distributions that are universal, not depending at all on the unknown true function, so do not require any additional procedures for their use for inference. Also, the assumptions needed for the LRS asymptotics to hold are the same as those for estimation, rather than stronger ones as in some other nonparametric settings.

These positive results motivate interest in using the statistic $2 \log \lambda_n$ of (6) for testing and forming confidence intervals for $r_0(x_0)$, and suggest that it may have a limit distribution that is universal and free of nuisance parameters. This would allow us to avoid the difficult estimation of $r_0''(x_0)$ and resulting tuning parameter

selection problem that would be required if we rely on (3) to do inference. We make the following conjecture. To state the conjecture we need some assumptions on the design variables (and on x_0); the assumptions (Assumptions 1 and 2) are stated and discussed in Sect. 4.1.

Conjecture 1 *Assume the regression model* (1) *holds where* $Ee^{t\epsilon_{n,i}^2} < \infty$ *for some* $t > 0$. *Assume* r_0 *is concave,* $r_0(x_0) = y_0$, r_0 *is twice continuously differentiable in a neighborhood of* x_0, *and* $r_0''(x_0) < 0$. *Let* Assumptions 1 *and* 2 *hold. Then, with* $2\log\lambda_n(y_0)$ *defined in* (6),

$$2\log\lambda_n(y_0) \to_d \sigma^2\mathbb{D}, \tag{7}$$

where $\mathbb{D}$ *is a universal random variable (not depending on* r_0 *or the distribution of* $\epsilon_{n,i}$*).*

A partial proof of the conjecture is given in Sect. 4.1. See Theorem 5 there. The form of the random variable $\mathbb{D}$ is given below in (84). Some discussion of the assumptions is given in remarks after Theorem 4.

A theorem analogous to Conjecture 1 was proved by Banerjee and Wellner [6] (see also [4]) in the context of the current status data model of survival analysis, by Banerjee [5] in the context of monotone response models, and by Groeneboom and Jongbloed [22] in the context of monotone density estimation. Those models are based on the shape restriction of monotonicity. In the context of a shape restriction based on concavity, Doss and Wellner [12–14] show a theorem analogous to Conjecture 1 for an LRS for the mode of a log-concave density. The likelihood ratio in the latter problem, based on a concavity assumption, involves remainder terms which are asymptotically negligible but are quite challenging to theoretically analyze. In the current status problem there are no such remainder terms, and in the monotone density problem they can be analyzed using the so-called min–max formula (see e.g., Lemma 3.2 of [22]), which does not have an analog for concavity-based problems. Thus it is quite difficult in general to analyze LRS's in concavity-based problems, and so proving Conjecture 1 in full is a large undertaking beyond the scope of the present paper. To study the asymptotics of $2\log\lambda_n$ and prove Conjecture 1, one needs to study the asymptotics of the constrained estimator $\widehat{r}_n^0$. Since $\widehat{r}_n^0$ is the solution to a strictly convex program, there are optimality conditions that characterize it (i.e., Karush–Kuhn–Tucker type conditions). One key component in developing the asymptotics of $\widehat{r}_n^0$ is to understand the conditions that characterize $\widehat{r}_n^0$, which we do in Theorem 1. We also study a corresponding limit version of the problem, which is to find the constrained concave least-squares estimator based on observing a Brownian motion with drift (i.e., observing the solution to a stochastic differential equation). We find conditions that characterize the solution to this limit problem (the limit LSE) in Theorem 2 (on a compact domain) and Theorem 3 (on all of $\mathbb{R}$) and we see that the conditions are analogous to those in the finite sample case. (Theorem 2 is used to prove Theorem 3.) Showing that the convex program optimality conditions are the same for the finite sample estimator and for the limit process is a crucial step in showing the limit process is indeed the limit distribution of the finite sample estimator. Finding the characterizing conditions, particularly in the limit problem, seems to be somewhat more challenging for the constrained problems than for the unconstrained ones. The process arising in

Theorem 3 is used in Theorem 4, which gives the limit distribution of $\widehat{r}_n^0(x_0)$. Finally, in Sect. 4.1, we use Theorem 4 to give a partial proof of of Conjecture 1. Specifically, in Theorem 5 we show that under an assumption on a certain remainder term, the conjectured limit statement (7) holds.

We further describe the structure of this paper, as follows. In Sect. 2 we consider the regression model and study some basic properties of the (finite sample) NLSE and ALSE, which includes presenting Theorem 1. In Sect. 3 we study the limiting version of the problem and present Theorems 2 and 3. In Sect. 4 we present Theorem 4. In Sect. 4.1 we present a partial proof of Conjecture 1. In Sect. 5, we provide simulations giving strong evidence in favor of Conjecture 1, and showing that the corresponding test and confidence interval have good finite sample performance. Section 6 has some concluding remarks and discussion of related problems. Appendix A has results we include for completeness and technical formulas.

2 Finite sample constrained concave regression

We begin with the regression setup $\widetilde{Y}_{n,i} = \tilde{r}_0(\widetilde{x}_{n,i}) + \epsilon_{n,i}$, $i = 1, \ldots, n$. We assume that $\epsilon_{n,i}$ are i.i.d. with mean 0, $Ee^{t\epsilon_{n,i}^2} < \infty$ for some $t > 0$, we assume $\{\widetilde{x}_{n,i}\}$ are fixed and without loss of generality we assume that $\widetilde{x}_{n,1} < \widetilde{x}_{n,2} < \cdots < \widetilde{x}_{n,n}$. Our model assumption is that $\tilde{r}_0 : \mathbb{R} \to \mathbb{R}$ is a concave function. Our interest is in using $2 \log \lambda_n$ from (6) to test for the value of $\tilde{r}_0$ at a fixed point x_0, and also in inverting those tests to form corresponding confidence intervals. Thus, we will study the constrained concave regression problem, where at a fixed point $x_0 \in \mathbb{R}$ we assume $\tilde{r}_0(x_0) = y_0$ for a fixed value y_0.

Let

$$\mathcal{C} := \{\varphi : \mathbb{R} \to [-\infty, \infty) \mid \varphi \text{ is concave, closed, and proper}\} \tag{8}$$

Here φ is proper if $\varphi(x) < \infty$ for all x and $\varphi(x) > -\infty$ for some x and φ is closed if it is upper semi-continuous (as in [40, pages 24 and 50]). We follow the convention that a concave function φ is defined on all of $\mathbb{R}$ by assigning φ the value $-\infty$ off its effective domain $\text{dom}(\varphi) := \{x : \varphi(x) > -\infty\}$ (as in [40, page 40]). For fixed $(x_0, y_0) \in \mathbb{R}^2$, let $\tilde{\mathcal{C}}^0 := \{r \in \mathcal{C} : r(x_0) = y_0\}$. We consider estimation of $\tilde{r}_0$ via minimization of the objective function $r \mapsto \frac{1}{2}\sum_{i=1}^n (\widetilde{Y}_{n,i} - r(\widetilde{x}_{n,i}))^2$. The constrained LSE is the minimum of the above objective function over $\tilde{\mathcal{C}}^0$; however, $\tilde{\mathcal{C}}^0$ is not a convex cone. Thus, to proceed further, we now introduce an augmented or auxiliary data set. We will (a) translate the original data set so that the corresponding set of possible regression functions forms a convex cone, and (b) potentially augment the $\widetilde{x}_{n,i}$ by x_0. In addition, in (a), without loss of generality, we will translate the data so that the true regression function may be assumed to satisfy $r_0'(x_0) = 0$. We define the auxiliary data set $\{(x_{n,i}, Y_{n,i})\}_{i=1}^{n_0}$, where n_0 will be either n or $n+1$, as follows.

1. If x_0 is equal to one of the data points, say $\widetilde{x}_{n,k^0} = x_0$ where $1 \le k^0 \le n$, then let $x_{n,i} := \widetilde{x}_{n,i}$ and let $n_0 := n$. Let $Y_{n,i} := \widetilde{Y}_{n,i} - y_0 - \tilde{r}_0'(x_0)(\widetilde{x}_{n,i} - x_0)$ for $i = 1, \ldots, n$.

2. If x_0 is not equal to any data point, then let $1 \le k^0 \le n+1$ be such that $\widetilde{x}_{n,k^0-1} < x_0 < \widetilde{x}_{n,k^0}$, where we let $\widetilde{x}_{n,0} = -\infty$ and $\widetilde{x}_{n,n+1} = \infty$ here. Then for $i = 1, \ldots, k^0 - 1$ let $x_{n,i} := \widetilde{x}_{n,i}$ and $Y_{n,i} := \widetilde{Y}_{n,i} - y_0 - \tilde{r}_0'(x_0)(\widetilde{x}_{n,i} - x_0)$, let $x_{n,k^0} := x_0$, and for $i = k^0 + 1, \ldots, n+1 =: n_0$, let $x_{n,i} := \widetilde{x}_{n,i-1}$ and $Y_{n,i} := \widetilde{Y}_{n,i-1} - y_0 - \tilde{r}_0'(x_0)(\widetilde{x}_{n,i-1} - x_0)$. Define $Y_{n,k^0} := 0$.

Thus the size of the augmented data set, n_0, is either n or $n+1$. In either case, define I to be the subset of $\{1, \ldots, n_0\}$ corresponding to data indices, so I has cardinality n and may or may not include k^0. Thus, with these definitions $Y_{n,i}$ and $x_{n,i}$ satisfy the regression relationship

$$Y_{n,i} = r_0(x_{n,i}) + \epsilon_{n,i} \quad \text{for} \quad i \in I, \tag{9}$$

where $r_0 \in \mathcal{C}^0 := \{r \in \mathcal{C} \mid r(x_0) = 0\}$. We thus consider the objective function[1]

$$\phi_n(r) = \frac{1}{2} \sum_{i \in I} \left(Y_{n,i} - r(x_{n,i})\right)^2. \tag{10}$$

A priori, $\operatorname{argmin}_{r \in \mathcal{C}} \phi_n(r)$ is uniquely specified only at the data points $x_{n,i}$ for $i \in I$, and $\operatorname{argmin}_{r \in \mathcal{C}^0} \phi_n(r)$ is only uniquely specified at the data points $x_{n,i}$ for $i = 1, \ldots, n_0$; thus we choose to restrict attention to solutions that are affine between the $x_{n,i}$. (The actual solutions will be uniquely specified on most of their domain in practice, because they will be piecewise linear with relatively few knot points.) Restricting attention to piecewise affine solutions is the standard approach, and the choice does not affect the asymptotic results, see e.g. [41]. For a concave function that is piecewise linear, we can identify the function with its values at its bend points, so we define corresponding subsets of $\mathbb{R}^n$ and $\mathbb{R}^{n_0}$ by

$$\mathcal{C}_n := \left\{\left(r\left(\widetilde{x}_{n,1}\right), \ldots, r\left(\widetilde{x}_{n,n}\right)\right) : r \in \mathcal{C}\right\} \text{ and } \mathcal{C}_n^0 := \left\{\left(r_0\left(x_{n,1}\right), \ldots, r_0\left(x_{n,n_0}\right)\right) : r_0 \in \mathcal{C}^0\right\}. \tag{11}$$

For a function r we let $\operatorname{eval}_n r := \left(r(\widetilde{x}_{n,1}), \ldots, r(\widetilde{x}_{n,n})\right)$ and $\operatorname{eval}_{n_0} r := (r(x_{n,1}), \ldots, r(x_{n,n_0}))$. For $r_n \in \mathcal{C}_n$, define the linear extrapolation $\operatorname{ext}(r_n) \in \mathbb{R}^{n_0}$ by $\operatorname{ext}(r_n) := \operatorname{eval}_{n_0}(r)$, where r is the function giving the linear interpolation of r_n. Then, slightly abusing notation (by giving ϕ_n a $\mathbb{R}^{n_0}$-vector argument rather than a function), we define the estimator vectors $\widehat{\underline{r}}_n$ and $\widehat{\underline{r}}_n^0$ by

$$\widehat{\underline{r}}_n \in \operatorname*{argmin}_{r \in \mathcal{C}_n} \phi_n(\operatorname{ext} r) \quad \text{and} \quad \widehat{\underline{r}}_n^0 \in \operatorname*{argmin}_{r \in \mathcal{C}_n^0} \phi_n(r), \tag{12}$$

and let $\widehat{r}_n$ and $\widehat{r}_n^0$ be the piecewise linear interpolation of $\widehat{\underline{r}}_n$ and $\widehat{\underline{r}}_n^0$ on $[x_{n,1}, x_{n,n_0}]$. We let $\widehat{r}_{n,i} := \widehat{r}_n(\widetilde{x}_{n,i})$, $i \in I$, and $\widehat{r}_{n,i}^0 := \widehat{r}_n^0(x_{n,i})$, $i = 1, \ldots, n_0$.

Proposition 1 *The estimators $\widehat{\underline{r}}_n$ and $\widehat{\underline{r}}_n^0$ exist and are unique (for any $n \ge 1$ or $n_0 \ge 1$, respectively).*

[1] Note that (9) and (10) are potentially different from (1) and (2) in the introduction, but only by a minor indexing modification.

Proof Both statements follow from writing the optimization as a quadratic program with linear inequality constraints giving the concavity restriction and one (linear) equality constraint corresponding to $r(x_0) = 0$ for the constrained estimator. □

The estimators $\widehat{r}_n$ and $\widehat{r}_n^0$ can be seen as projections of the data onto the convex cones $\mathcal{C}_n$ and $\mathcal{C}_n^0$, and so we begin by studying these cones. A convex subset $\mathcal{K}$ of a (possibly infinite dimensional) real vector space is a convex cone if $x \in \mathcal{K}$ implies $\lambda x \in \mathcal{K}$ if $\lambda \geq 0$. We say a convex cone $\mathcal{K}$ is (finitely) generated by or is spanned by a set of elements $k_1, \ldots, k_m \in \mathcal{K}$ if any $k \in \mathcal{K}$ can be written as $k = \sum_{i=1}^m \lambda_i k_i$ for some $\lambda_i \geq 0$. Define $(y)_- := \min(y, 0)$ for $y \in \mathbb{R}$.

Proposition 2 1. *A generating set for $\mathcal{C}_n$ is given by $\pm \operatorname{eval}_n 1, \pm \operatorname{eval}_n x$, and $\operatorname{eval}_n(\widetilde{x}_{n,i} - x)_-$ for $i = 2, \ldots, n-1$.*
2. *A generating set for $\mathcal{C}_n^0$ is given by $\pm \operatorname{eval}_{n_0}(x - x_0)$, $\operatorname{eval}_{n_0}(x - x_{n,i})_-$ for $i = 2, \ldots, k^0$, and $\operatorname{eval}_{n_0}(x_{n,i} - x)_-$ for $i = k^0 + 1, \ldots, n_0 - 1$.*

Proof First we show 1. Consider the subset of $\mathcal{C}$ that is piecewise affine with kinks only possible at $\widetilde{x}_{n,i}$, and where we restrict attention to $[\widetilde{x}_{n,1}, \widetilde{x}_{n,n}]$. Then for $x \in [\widetilde{x}_{n,1}, \widetilde{x}_{n,n}]$ we can write

$$r(x) = b + w_1 \left(x - \widetilde{x}_{n,1}\right) + w_2 \left(\widetilde{x}_{n,2} - x\right)_- + \cdots + w_{n-1} \left(\widetilde{x}_{n,n-1} - x\right)_-,$$

where $b, w_1 \in \mathbb{R}$, since r is piecewise affine, and $w_i \geq 0$ for $i = 2, \ldots, n-1$ since r is concave. Thus $\pm 1, \pm x$, and $(\widetilde{x}_{n,i} - x)_-$ for $i = 2, \ldots, n-1$ generate the cone of piecewise affine functions on $[\widetilde{x}_{n,1}, \widetilde{x}_{n,n}]$, and applying eval_n to these functions yields a generating set for $\mathcal{C}_n$.

Now we show 2. Any r that is piecewise affine with possible kinks at the $x_{n,i}$ can be written as

$$r(x) = b + w_1 \left(x - x_{n,1}\right) + w_2 \left(x_{n,2} - x\right)_- + \cdots + w_{n-1} \left(x_{n,n_0-1} - x\right)_-, \quad (13)$$

where $b, w_1 \in \mathbb{R}$, and $w_i \geq 0$ for $i = 2, \ldots, n_0 - 1$.

Since $(x_{n,i} - x)_- = (x - x_{n,i})_- - (x - x_{n,i})$, we can rewrite (13) as

$$\begin{aligned} &\tilde{b} + \tilde{w}_1(x - x_0) + \tilde{w}_2 \left(x - x_{n,2}\right)_- + \cdots + \tilde{w}_{k^0} \left(x - x_{n,k^0}\right)_- \\ &\quad + w_{k^0+1} \left(x_{n,k^0+1} - x\right)_- + \cdots + w_{n_0-1} \left(x_{n,n_0-1} - x\right)_- \end{aligned} \quad (14)$$

where $\tilde{b}, \tilde{w}_1 \in \mathbb{R}$, and $w_i, \tilde{w}_i \geq 0$. Thus enforcing $r(x_0) = 0$ amounts precisely to requiring $\tilde{b} = 0$ in (14). Thus $\pm(x - x_0)$, $(x - x_{n,i})_-$ for $i = 2, \ldots, k^0$, and $(x_{n,i} - x)_-$ for $i = k^0 + 1, \ldots, n_0 - 1$ span the cone of functions $r \in \mathcal{C}^0$ given by piecewise affine functions with kinks only possible at $x_{n,i}$ and $\operatorname{dom} r = [x_{n,1}, x_{n,n_0}]$. Correspondingly, applying $\operatorname{eval}_{n_0}$ to the above set of functions gives the span of $\mathcal{C}_n^0$. □

Next we study characterizations of the estimators. First, we state the result for the unconstrained estimator. The characterizations are derived from the previous proposition about the boundary elements of the cones we minimize over together with the

following optimality conditions, as given in Corollary 2.1 of [21]. We use $\langle \cdot, \cdot \rangle$ to denote the usual Euclidean inner product and, for a differentiable function $f : \mathbb{R}^m \to \mathbb{R}$, we use $\nabla f(x)$ to denote the gradient vector at x.

Proposition 3 ([21, Corollary 2.1]) *Let $\phi : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be differentiable and convex. Let $z_1, \ldots, z_k \in \mathbb{R}^m$ and let $\mathcal{K}$ be the convex cone generated by $z_1, \ldots, z_k$. Then $\hat{x} \in \mathcal{K}$ is the minimum of ϕ over $\mathcal{K}$ if and only if*

$$\left\langle z_i, \nabla\phi(\hat{x}) \right\rangle \geq 0 \quad \text{for} \quad 1 \leq i \leq k, \tag{15}$$

$$\left\langle z_i, \nabla\phi(\hat{x}) \right\rangle = 0 \quad \text{if} \quad \hat{\alpha}_i > 0, \tag{16}$$

where the nonnegative numbers $\hat{\alpha}_1, \ldots, \hat{\alpha}_k$ satisfy $\hat{x} = \sum_{i=1}^k \hat{\alpha}_i z_i$.

Proposition 4 ([24, Lemma 2.6]) *Define $\widehat{R}_{n,k} := \sum_{i=1}^k \widehat{r}_{n,i}$ and $S_{n,k} := \sum_{i=1}^k Y_{n,i}$ for $k \in I$. Then $\widehat{r}_n = \operatorname{argmin}_{r \in \mathcal{C}_n} \phi_n(r)$ if and only if $\widehat{R}_{n,n} = S_{n,n}$ and*

$$\sum_{k=1}^{j-1} \widehat{R}_{n,k} \left(\widetilde{x}_{n,k+1} - \widetilde{x}_{n,k} \right) \begin{cases} \leq \sum_{k=1}^{j-1} S_{n,k} \left(\widetilde{x}_{n,k+1} - \widetilde{x}_{n,k} \right) & j \in I,\ j \geq 2 \\ = \sum_{k=1}^{j-1} S_{n,k} \left(\widetilde{x}_{n,k+1} - \widetilde{x}_{n,k} \right) & \text{if } \widehat{r}_n \text{ has a kink at } \widetilde{x}_{n,j}. \end{cases}$$

We define $\widehat{r}_n$ to always have a kink at $\widetilde{x}_{n,n}$.

Proof This is proved in [24]. The proof follows from the first part of Proposition 2 together with Proposition 3. □

The inequality in the characterization is reversed from the original lemma in [24], since we are considering concave regression and [24] consider convex regression. Note that (15) and (16) are equivalent to saying

$$\left\langle \Delta, \nabla\phi(\hat{x}) \right\rangle \geq 0 \tag{17}$$

for all $\Delta \in \text{cone } \{\{z_i : 1 \leq i \leq k\} \cup \{-z_i : i \in I\}\}$ where $I := \left\{ i : \hat{\alpha}_i > 0 \right\}$ is the set of inactive constraints and

$$\text{cone } \{y_i : i \in \mathcal{I}\} := \left\{ \sum_{i \in \mathcal{I}} \alpha_i y_i : \alpha_i \geq 0 \right\}.$$

The cone we are now interested in is $\mathcal{C}_n^0$ which, by Proposition 2, is generated by

$$a_{\pm} := \pm \left(x_{n,j} - x_0 \right)_{j=1}^{n_0}, \tag{18}$$

$$a_i := \left(\left(x_{n,j} - x_{n,i} \right) \mathbb{1}_{\{i \geq j\}} \right)_{j=1}^{n_0} \quad \text{for } i = 2, \ldots, k^0, \text{ and} \tag{19}$$

$$a_i := \left(\left(x_{n,i} - x_{n,j} \right) \mathbb{1}_{\{i \leq j\}} \right)_{j=1}^{n_0} \quad \text{for } i = k^0 + 1, \ldots, n_0 - 1. \tag{20}$$

We now show an analog of the unconstrained characterization, Proposition 4, for the constrained case. For ease of presentation, we assume without any loss of generality that $x_0 = 0$.

Theorem 1 *Let $x_0 = 0$. For $\widehat{\underline{r}}_n^0 \in \mathcal{C}_n^0$, define*

$$\widehat{R}_{n,k}^L := \sum_{j=1}^{k} \widehat{r}_{n,j}^0, \quad S_{n,k}^L := \sum_{j=1}^{k} Y_{n,j}, \quad \text{for } k = 1, \ldots, k^0 - 1,$$
$$\widehat{R}_{n,k}^R := \sum_{j=k}^{n_0} \widehat{r}_{n,j}^0, \quad \text{and} \quad S_{n,k}^R := \sum_{j=k}^{n_0} Y_{n,j}, \quad \text{for } k = k^0 + 1, \ldots n_0.$$

Then $\widehat{\underline{r}}_n^0$ is the unique element of $\operatorname{argmin}_{r \in \mathcal{C}_n^0} \phi_n(r)$ *given by* Proposition 1 *if and only if*

$$\sum_{k=1}^{j-1} \widehat{R}_{n,k}^L \left(x_{n,k+1} - x_{n,k}\right) \le \sum_{k=1}^{j-1} S_{n,k}^L \left(x_{n,k+1} - x_{n,k}\right) \quad \text{for } 2 \le j \le k^0, \tag{21}$$

$$\sum_{k=j+1}^{n_0} \widehat{R}_{n,k}^R \left(x_{n,k} - x_{n,k-1}\right) \le \sum_{k=j+1}^{n_0} S_{n,k}^R \left(x_{n,k} - x_{n,k-1}\right) \quad \text{for } k^0 + 1 \le j \le n_0 - 1, \tag{22}$$

$$\sum_{k=1}^{k^0-1} \left(\widehat{R}_{n,k}^L - S_{n,k}^L\right) \left(x_{n,k+1} - x_{n,k}\right) = \sum_{k=k^0+1}^{n_0} \left(\widehat{R}_{n,k}^R - S_{n,k}^R\right) \left(x_{n,k} - x_{n,k-1}\right), \tag{23}$$

where the inequalities in (21) *and* (22) *are equalities if $x_{n,j}$ is a knot of $\widehat{r}_n^0$.*

Proof Let $a_\pm$ and a_i, $2 \le i \le n_0 - 1$, be defined as in (18), (19), and (20). Compute

$$\nabla_{n_0} \phi_n \left(\widehat{\underline{r}}_n^0\right) := \left(\widehat{r}_{n,1}^0 - Y_{n,1}, \ldots, \left(\widehat{r}_{n,k^0}^0 - Y_{n,k^0}\right) \mathbb{1}_{\{n_0 = n\}}, \ldots, \widehat{r}_{n,n_0}^0 - Y_{n,n_0}\right)' \tag{24}$$

or $\nabla_{n_0} \phi_n(\widehat{\underline{r}}_n^0) = \left((\widehat{r}_{n,j}^0 - Y_{n,j}) \mathbb{1}_{\{j \ne k^0 \text{ or } n_0 = n\}}\right)_{j=1}^{n_0}$. By Propositions 2 and 3 we see that $\widehat{\underline{r}}_n^0 = \operatorname{argmin}_{r \in \mathcal{C}_n^0} \phi_n(r)$ if and only if $\widehat{\underline{r}}_n^0 \in \mathcal{C}_n^0$ and

$$\left\langle (x_{n,j})_{j=1}^{n_0}, \nabla_{n_0} \phi_n \left(\widehat{\underline{r}}_n^0\right) \right\rangle = 0, \tag{25}$$

$$\left\langle \left((x_{n,j} - x_{n,i}) \mathbb{1}_{\{i > j\}}\right)_{j=1}^{n_0}, \nabla_{n_0} \phi_n \left(\widehat{\underline{r}}_n^0\right) \right\rangle \ge 0 \quad \text{for } i = 2, \ldots, k^0, \tag{26}$$

and

$$\left\langle \left((x_{n,i} - x_{n,j}) \mathbb{1}_{\{i < j\}}\right)_{j=1}^{n_0}, \nabla_{n_0} \phi_n \left(\widehat{\underline{r}}_n^0\right) \right\rangle \ge 0 \quad \text{for } i = k^0 + 1, \ldots, n_0 - 1, \tag{27}$$

with equalities in (26) and (27) if $x_{n,i} \in S(\widehat{r}_n^0)$. From (26), for $i = 2, \ldots, k^0 - 1, k^0$, we have

$$0 \le \sum_{j=1}^{n_0} \mathbb{1}_{\{j<i\}} \left(x_{n,j} - x_{n,i}\right) \left(\widehat{r}_{n,j}^0 - Y_{n,j}\right) = \sum_{j=1}^{i-1} \left(x_{n,j} - x_{n,i}\right) \left(\widehat{r}_{n,j}^0 - Y_{n,j}\right), \tag{28}$$

and from (27) for $i = k^0 + 1, \ldots, n_0 - 1$, we have

$$0 \leq \sum_{j=1}^{n_0} \mathbb{1}_{\{j>i\}}(x_{n,i} - x_{n,j})\left(\widehat{r}^0_{n,j} - Y_{n,j}\right) = \sum_{j=i+1}^{n_0} \left(x_{n,i} - x_{n,j}\right)\left(\widehat{r}^0_{n,j} - Y_{n,j}\right), \tag{29}$$

and from (25)

$$0 = \sum_{j=1, j\neq k^0}^{n_0} x_{n,j}\left(\widehat{r}^0_{n,j} - Y_{n,j}\right) \tag{30}$$

since $x_{n,k^0} = 0$. Summing by parts, we see for $i = 2, \ldots, k^0$ that (28) equals

$$\begin{aligned}\sum_{j=1}^{i-1}\sum_{k=j}^{i-1}\left(x_{n,k} - x_{n,k+1}\right)\left(\widehat{r}^0_{n,j} - Y_{n,j}\right) &= \sum_{j=1}^{i-1}\sum_{k=1}^{i-1}\mathbb{1}_{j\leq k}\left(x_{n,k} - x_{n,k+1}\right)\left(\widehat{r}^0_{n,j} - Y_{n,j}\right)\\ &= \sum_{k=1}^{i-1}\left(x_{n,k} - x_{n,k+1}\right)\sum_{j=1}^{i-1}\mathbb{1}_{j\leq k}\left(\widehat{r}^0_{n,j} - Y_{n,j}\right)\\ &= \sum_{k=1}^{i-1}\left(x_{n,k} - x_{n,k+1}\right)\left(\widehat{R}^L_{n,k} - S^L_{n,k}\right).\end{aligned} \tag{31}$$

Similarly, from (29), for $i = k^0 + 1, \ldots, n_0 - 1$, we see that

$$\begin{aligned}\sum_{j=i+1}^{n_0}\sum_{k=i+1}^{j}(x_{n,k-1} - x_{n,k})\left(\widehat{r}^0_{n,j} - Y_{n,j}\right) &= \sum_{j=i+1}^{n_0}\sum_{k=i+1}^{n_0}\mathbb{1}_{k\leq j}(x_{n,k-1} - x_{n,k})(\widehat{r}^0_{n,j} - Y_{n,j})\\ &= \sum_{k=i+1}^{n_0}(x_{n,k-1} - x_{n,k})\sum_{j=i+1}^{n_0}\mathbb{1}_{k\leq j}(\widehat{r}^0_{n,j} - Y_{n,j})\\ &= \sum_{k=i+1}^{n_0}(x_{n,k-1} - x_{n,k})(\widehat{R}^R_{n,k} - S^R_{n,k}).\end{aligned} \tag{32}$$

To finish, we use the same calculations once more. From (30), since $x_{n,k^0} = 0$, we see that

$$\sum_{j=1}^{k^0-1}\left(x_{n,j} - x_{n,k^0}\right)\left(\widehat{r}^0_{n,j} - Y_{n,j}\right) = \sum_{j=k^0+1}^{n_0}\left(x_{n,k^0} - x_{n,j}\right)\left(\widehat{r}^0_{n,j} - Y_{n,j}\right). \tag{33}$$

Identifying the left- and right-hand sides of (33) with the right-hand sides of (28) and (29) (with $i = k^0$ in both cases), and using (31) and (32), we see

$$\sum_{k=1}^{k^0-1}\left(x_{n,k} - x_{n,k+1}\right)\left(\widehat{R}^L_{n,k} - S^L_{n,k}\right) = \sum_{k=k^0+1}^{n_0}\left(x_{n,k-1} - x_{n,k}\right)\left(\widehat{R}^R_{n,k} - S^R_{n,k}\right).$$

This completes the proof. □

3 Limit process for constrained concave regression

Now we consider an asymptotic version of this problem. Let

$$dX(t) = -12t^2 dt + dW(t) \tag{34}$$

where W is a standard two-sided Brownian motion started from 0. This serves as a canonical/limiting/white noise version of a concave regression problem (with canonical regression function $r_0(t) = -12t^2$, where the constant 12 is not important). As has been seen in past work (e.g., Theorem 7 in Appendix A) and as will be seen below in Theorem 4, the white noise problem is important because it yields the limit distribution of (finite sample) estimators. On a compact interval $[-c, c]$ one can define a least-squares objective function

$$\phi_c(r) = \frac{1}{2}\int_{-c}^{c} r(u)^2\, du - \int_{-c}^{c} r(u) dX(u), \tag{35}$$

as in [23]. Note that, symbolically replacing dX with g, $(r-g)^2/2 = r^2/2 - rg + g^2/2$; we can drop the $g^2/2$ term (which is irrelevant when optimizing over r) which explains why (35) is a 'least-squares' objective function. We can now consider minimizing ϕ_c over concave functions r satisfying $r(0) = 0$. See the introduction (pages 1622–1623) of [23] for further explanation and derivation motivating the idea that (35) serves as a limit version of the objective function (10), and that (34) serves as an approximation to the (finite sample) observed data. For $c > 0$ and $k < 0$, let

$$\mathcal{C}_{c,k} := \{r : [-c, c] \to \mathbb{R};\ r \text{ concave}, r(0) = 0, r(\pm c) = k\}. \tag{36}$$

We add the extra constraints $r(\pm c) = k$ to compactify the problem. These constraints become irrelevant as $c \to \infty$. We start by showing existence and uniqueness of the minimizer of (35).

Proposition 5 *Let $k < 0$ and ϕ_c be given by (35). For Lebesgue-almost every $c > 0$, $\operatorname{argmin}_{r \in \mathcal{C}_{c,k}} \phi_c(r)$ exists and is unique with probability 1.*

Proof Let $r \in \mathcal{C}_{c,k}$. Note that if $M := \max r \to \infty$, then by concavity of r, $r > M/2$ on some interval of length at least $c/4$ for M large enough. Then the first term in (35) is of order M^2 whereas the second term is of order M, so the objective function value goes to ∞. We can thus almost surely restrict attention to only functions bounded above by some fixed value M.

Now consider the class

$$\mathcal{C}^{\circ}_{c,k,M} := \{r : [-c, c] \to \mathbb{R}, r \text{ concave}, r(0) = 0, M \geq r, r(\pm c) \geq k\}.$$

This class is closed under pointwise convergence because limits of concave functions are concave and the limit of a uniformly bounded function is uniformly bounded. It is thus a closed subset of a set of functions compact (by Tychonoff's theorem) under pointwise convergence, so is compact, and by the Lebesgue bounded convergence theorem, ϕ_c is continuous with respect to pointwise convergence. Thus ϕ_c attains a minimum on $\mathcal{C}^\circ_{c,k,M}$. We now show that the minimum satisfies the constraints $r(\pm c) = k$. Assume, to the contrary, that $r(c) > k$. Let $(y)_- := \min(y, 0)$ for $y \in \mathbb{R}$. Let $\Delta = \eta(c - \delta - \cdot)_-$ for some $\delta, \eta > 0$, where $\delta\eta = g := r(c) - k$, so that $r(c) + \Delta(c) = k$. We will show that $\phi_c(r + \Delta) < \phi_c(r)$. Let $\tilde{X}(u) := X(u) - X(c)$. Then

$$-\int \Delta dX = -\int \Delta d\tilde{X} = -(\tilde{X}\Delta)(c-\delta, c] + \int \tilde{X} d\Delta = \int \tilde{X} d\Delta$$

since $\Delta(c - \delta) = 0$ and $\tilde{X}(c) = 0$. Here, we let $g(a, b] := g(b) - g(a)$ for a function g and $a < b$. The previous display equals

$$-\eta \int_{c-\delta}^{c} (W(u) - W(c))du + \eta \int_{c-\delta}^{c} \left(4u^3 - 4c^3\right) du. \tag{37}$$

There exists a sequence of δ's converging to 0 such that the first term in (37) is 0 because integrated Brownian motion started from 0 crosses 0 an infinite number of times near 0, almost surely. The second term in (37) equals, to first order approximation (as $\delta \searrow 0$),

$$-\eta k \int_{c-\delta}^{c} (u - c)du = \eta\delta^2 k/2 = g\delta k/2. \tag{38}$$

On the other hand, the first order term in $\int_{c-\delta}^{c}((r + \Delta)^2 - r^2)d\lambda/2$ is $\int r\Delta d\lambda$ which equals, to first order,

$$r(c) \int \Delta d\lambda = -r(c) \int |\Delta| d\lambda = \frac{-r(c)}{2} g\delta < (-k/2)g\delta. \tag{39}$$

Thus we see that there exists δ such that (39) plus (38) is negative, i.e. such that $\phi_c(r + \Delta) - \phi_c(r) < 0$. Thus the minimum over $\mathcal{C}^\circ_{c,k,M}$ satisfies $r(\pm c) = k$, and so ϕ_c attains a minimum on $\mathcal{C}_{c,k}$.

Uniqueness of the minimum follows from the strict convexity of ϕ_c on the convex set $\mathcal{C}_{c,k}$: for any $w \in (0, 1)$, $r_1, r_2 \in \mathcal{C}_{c,k}$,

$$\phi_c(wr_1 + (1-w)r_2) = w\phi_c(r_1) + (1-w)\phi_c(r_2) - \frac{w(1-w)}{2} \int_{-c}^{c} (r_1(u) - r_2(u))^2 du,$$

where the right side is strictly less than $w\phi_c(r_1) + (1 - w)\phi_c(r_2)$ if $\int_{-c}^{c}(r_1(u) - r_2(u))^2 du > 0$, so that ϕ_c is strictly convex. This completes the proof. □

We now state and prove a characterization of the minimizer of (35). Unlike the unconstrained case, we must explicitly deal with the knot set $\widehat{S}^0$ (defined below) in the

statement and the proof of the theorem, because the correct definitions of the processes depends on knots τ_L and τ_R (also defined below). This complicates definitions, because $\widehat{S}^0$ is not necessarily a countable set. It is known to have Lebesgue measure zero [23,43]. For our next theorems, we let

Theorem 2 *Let X be given by (34). Fix $c > 0$ and $k < 0$ and let $\widehat{r}^0 \in \mathcal{C}_{c,k}$. Define $\widehat{S}^0$ by*

$$\left(\widehat{S}^0\right)^c := \left(\widehat{S}^0\left(\widehat{r}^0\right)\right)^c := \left\{t \in \mathbb{R} : \left(\widehat{r}^0\right)''(t) = 0\right\}.$$

For $t \in [-c, c]$, define

$$\tau_R := \inf\left(\widehat{S}^0 \cap [0, c)\right), \quad \tau_L := \sup\left(\widehat{S}^0 \cap (-c, 0]\right),$$

$$X_R(t) := \int_{\tau_R}^{t} dX, \quad X_L(t) := \int_{t}^{\tau_L} dX, \tag{40}$$

$$Y_R(t) := \int_{\tau_R}^{t} X_R(u)du, \quad Y_L(t) := \int_{t}^{\tau_L} X_L(u)du. \tag{41}$$

Let H_R be the primitive of the primitive of $\widehat{r}^0$ such that $H_R(c) = Y_R(c)$ and $H_R(\tau_R) = Y_R(\tau_R)$. Let H_L be the primitive of the primitive of $\widehat{r}^0$ such that $H_L(-c) = Y_L(-c)$ and $H_L(\tau_L) = Y_L(\tau_L)$. Then for Lebesgue-almost-every $c > 0$, $\widehat{r}^0 = \operatorname{argmin}_{r \in \mathcal{C}_{c,k}} \phi_c(r)$ if and only if the following three conditions hold:

1. $(H_R - Y_R)(0) = (H_L - Y_L)(0)$,
2. *for $c \geq t \geq 0$, $(H_R - Y_R)(t) \leq 0$ and for $-c \leq t \leq 0$, $(H_L - Y_L)(t) \leq 0$,*
3. *and*

$$\int_{[-c,0]} (H_L - Y_L)\, d\left(\widehat{r}^0\right)' = 0 = \int_{[0,c]} (H_R - Y_R)\, d\left(\widehat{r}^0\right)'. \tag{42}$$

For completeness we give, in Lemma 5 in the appendix, integration by parts formulas, which we will use in the proof without further reference. Recall also that, by Theorem 23.1 of [40], a finite, concave function on $\mathbb{R}$ has well-defined right and left derivatives on all of $\mathbb{R}$.

Proof Notice τ_L and τ_R are well defined and finite because $\widehat{r}^0(0) = 0$ and $\widehat{r}^0(\pm c) = k < 0$, so $\widehat{r}^0$ cannot be affine.

Sufficiency Assume Conditions 1, 2, and 3 hold for $\widehat{r}^0$. Since H_R is twice differentiable, if $H_R(c) = Y_R(c)$, ${H_R}'(c) = {Y_R}'(c)$, and $H_R \leq Y_R$ then there exists a 'one-sided parabolic tangent' [23] to Y_R at c. Because W is of infinite variation, for Lebesgue almost all $c > 0$, Y_R cannot have such a one-sided parabolic tangent, so we can thus assume that ${H_R}'(c) = F_R(c) > X_R(c) = {Y_R}'(c)$. Note we can rule out $F_R(c) < X_R(c)$ because then on an interval $[c - \delta, c]$, $\delta > 0$, we would have $Y_R < H_R$. Similarly, we can assume $-{H_L}'(-c) = F_L(-c) > X_L(-c) = -{Y_L}'(-c)$.

Note that for functions q and r,

$$q^2 - r^2 = (q - r)^2 + 2r(q - r) \geq 2r(q - r).$$

Thus for any $q \in \mathcal{C}_{c,k}$ which is not Lebesgue-a.e. identical to $\widehat{r}^0$,

$$\begin{aligned}\phi_c(q) - \phi_c\left(\widehat{r}^0\right) &> \int_{-c}^{c}\left(q-\widehat{r}^0\right)\left(\widehat{r}^0 d\lambda - dX\right)\\ &= -\int_{-c}^{0}\left(q-\widehat{r}^0\right)d(F_L - X_L) + \int_0^c\left(q-\widehat{r}^0\right)d(F_R - X_R),\end{aligned}$$

from (46). The previous display equals

$$\begin{aligned}&-\left[\left(\left(q-\widehat{r}^0\right)(F_L - X_L)\right)(-c,0] - \int_{-c}^{0}(F_L - X_L)\,d\left(q-\widehat{r}^0\right)\right]\\ &\quad+\left(\left(q-\widehat{r}^0\right)(F_R - X_R)\right)(0,c] - \int_0^c (F_R - X_R)\,d\left(q-\widehat{r}^0\right)\\ &= \int_{-c}^{0}(F_L - X_L)\,d\left(q-\widehat{r}^0\right) - \int_0^c (F_R - X_R)\,d\left(q-\widehat{r}^0\right)\end{aligned} \tag{43}$$

since $q, \widehat{r}^0 \in \mathcal{C}_{c,k}$, and, recalling $-(F_L - X_L) = (H_L - Y_L)'$, $(F_R - X_R) = (H_L - Y_L)'$, we see the previous display equals

$$\begin{aligned}&-\left((H_L - Y_L)\left(q-\widehat{r}^0\right)'\right)(-c,0] + \int_{(-c,0]}(H_L - Y_L)\,d\left(q-\widehat{r}^0\right)'\\ &\quad-\left((H_R - Y_R)\left(q-\widehat{r}^0\right)'\right)(0,c] + \int_{(0,c]}(H_R - Y_R)\,d\left(q-\widehat{r}^0\right)',\end{aligned}$$

and if both $q'(\pm c)$ and $(\widehat{r}^0)'(\pm c)$ are finite, then by Condition 1 and since $(H_L - Y_L)(-c) = 0$, $(H_R - Y_R)(c) = 0$, the previous display equals

$$\int_{(-c,0]}(H_L - Y_L)dq' + \int_{(0,c]}(H_R - Y_R)dq' \geq 0.$$

The final inequality follows by Condition 2 and because q' is nonincreasing (q is concave), so that q' defines a nonpositive measure.

We now show that we can take both $q'(\pm c)$ and $(\widehat{r}^0)'(\pm c)$ to be finite, which will complete the proof of sufficiency. Recall from the beginning of this sufficiency proof that we may assume that Y_R does not have a one-sided parabolic tangent at c, and thus that $F_R(c) > X_R(c)$. Now,

$$k\left(X_R(c) - F_R(c)\right) + \int_0^c\left(\widehat{r}^0\right)^2(u)du - \int_0^c \widehat{r}^0(u)dX(u) \tag{44}$$

equals $\int_0^c(\widehat{r}^0)'(u)(X_R(u) - F_R(u))du$. But if $(\widehat{r}^0)'(u) \to -\infty$ as $u \nearrow c$, then $\int_0^\eta(\widehat{r}^0)'(u)(X_R(u) - F_R(u))du \to \infty$ as $\eta \nearrow c$. A similar argument holds on $[-c, 0]$ to show that if $(\widehat{r}^0)'(u) \to \infty$ as $u \searrow -c$, then $\int_\eta^0(\widehat{r}^0)'(u)(X_L - F_L)(u)\,du \to \infty$ as $\eta \searrow -c$. Comparison with e.g. the triangle function linearly interpolating between

$\widehat{r}^0(\pm c) = k$ and $\widehat{r}^0(0) = 0$, shows that $\phi_c(\widehat{r}^0) < \infty$ and so (44) is also $< \infty$. Thus, by contradiction, we see that $(\widehat{r}^0)'$ (interpreted appropriately as the left or right derivative) is bounded above at $-c$ and below at c, and so by concavity is bounded on all of $[-c, c]$. Similarly, to see that q' can be assumed finite, notice that if $q'(u) \to -\infty$ as $u \nearrow c$ or $q'(u) \to \infty$ as $u \searrow -c$, then (43) would be infinite, so we would be finished. This completes the proof of sufficiency.

Necessity Assume $\widehat{r}^0 = \text{argmin}_{r \in \mathcal{C}_{c,k}} \phi_c(r)$. We argue by perturbations of $\widehat{r}^0$ to show the characterization holds. For a perturbation $\Delta : [-c, c] \to \mathbb{R}$, we will say that the perturbation is 'acceptable for small ϵ' if for all $\epsilon > 0$ small enough, $\widehat{r}^0 + \epsilon\Delta \in \mathcal{C}_{c,k}$. If for all $\epsilon > 0$ small enough, $\widehat{r}^0 + \epsilon\Delta$ is concave (but may not satisfy the constraints at $\pm c$), we say 'Δ preserves concavity for small ϵ.' If $\widehat{r}^0 + \Delta \in \mathcal{C}_{c,k}$ or $\widehat{r}^0 + \Delta$ is concave we will say that Δ is 'acceptable' or 'preserves concavity,' respectively (in which case the ϵ is generally explicitly given). We let $F_R := H_R{}'$ and $F_L := -H_L{}'$ (in analogy with $X_R = Y_R{}'$, $X_L = -Y_L{}'$). Note: this means $H_L(t) = \int_t^{\tau_L} F_L d\lambda$ and $H_R(t) = \int_{\tau_R}^t F_R d\lambda$. Recall that λ is Lebesgue measure. For a perturbation Δ that is acceptable for small ϵ,

$$0 \leq \lim_{\epsilon \searrow 0} \epsilon^{-1} \left(\phi_c \left(\widehat{r}^0 + \epsilon\Delta \right) - \phi_c \left(\widehat{r}^0 \right) \right) = \int_{-c}^{c} \Delta \left(\widehat{r}^0 d\lambda - dX \right) \tag{45}$$

$$= -\int_{-c}^{0} \Delta \, d(F_L - X_L) + \int_{0}^{c} \Delta \, d(F_R - X_R), \tag{46}$$

since $\widehat{r}^0$ minimizes ϕ_c. We now show a preliminary result. For $t \geq 0$ let $\Delta_t(u) := (t - u)_-$ and for $t < 0$ let $\Delta_t(u) := (u - t)_-$, where $(y)_- = \min(y, 0)$. Now fix $t \geq 0$, $\epsilon > 0$, and $\tau \in \widehat{S}^0$. Assume $\tau \geq 0$; the case $\tau \leq 0$ is analogous. Assume further that either $(\widehat{r}^0)'(\tau+) \neq (\widehat{r}^0)'(\tau-)$ or $(\widehat{r}^0)''_+(\tau) \neq 0$, where $(\widehat{r}^0)''_+(\tau) := \lim_{h \searrow 0} h^{-1} \left((\widehat{r}^0)'(\tau + h) - (\widehat{r}^0)'(\tau+) \right)$ is the second derivative from above. In the statement "$(\widehat{r}^0)''_+(\tau) \neq 0$," we allow the possibility $(\widehat{r}^0)''_+(\tau)$ is undefined. Notice that there exists a sequence of points $\{\tau_i\} \subset \widehat{S}^0$, $\tau_i \searrow \tau_R$, such that τ_i satisfies the conditions just described for τ, since $\widehat{r}^0$ is linear on $[\tau_L, \tau_R]$ (so has second derivative that is 0 from below). Thus either $\tau_R \in \widehat{S}^0$ or there are $\tau_i \in \widehat{S}^0$, $\tau_i \searrow \tau_R$, all either having discontinuous derivative or having nonzero second derivative from above. Now for $s \geq 0$, define the concave function $r_{\epsilon,s}$ by

$$\begin{aligned} r_{\epsilon,s}(u) &:= \min \left(\widehat{r}^0(s) + \left(\left(\widehat{r}^0 \right)' (s+) + \epsilon \right) (u - s), \ \widehat{r}^0(u) - \epsilon \Delta_s(u) \right) \\ &= \widehat{r}^0(u) \mathbb{1}_{(-\infty, s-\delta]}(u) + \left(\widehat{r}^0(s) + \left(\left(\widehat{r}^0 \right)' (s+) + \epsilon \right) (u - s) \right) \mathbb{1}_{(s-\delta, s)}(u) \\ &\quad + \left(\widehat{r}^0(u) - \epsilon \Delta_s(u) \right) \mathbb{1}_{[s, \infty)}(u) \end{aligned} \tag{47}$$

where the equality holds for some small $\delta \equiv \delta_{\widehat{r}^0,s,\epsilon} \geq 0$ which solves

$$\widehat{r}^0(s-\delta) = \widehat{r}^0(s) - \left(\left(\widehat{r}^0\right)'(s+) + \epsilon \right) \delta. \tag{48}$$

By our assumptions on τ, we can check that there exist sequences $s_i \searrow \tau$ and $\epsilon_i \searrow 0$ such that $\delta_i := s_i - \tau \geq 0$ satisfy $\delta_i = O(\epsilon_i)$ as $i \to \infty$. If $(\widehat{r}^0)'(\tau+) \neq (\widehat{r}^0)'(\tau-)$ then $-\Delta_\tau$ preserves concavity for small ϵ and we may take $s_i = \tau$ and $\delta_i = 0$ for any $\epsilon_i > 0$ small enough, so clearly $\delta_i = O(\epsilon_i)$. Similarly, if there exists a sequence $s_i \searrow \tau$ such that $(\widehat{r}^0)'(s_i+) \neq (\widehat{r}^0)'(s_i-)$ then we may take $\delta_i = 0$. If $(\widehat{r}^0)'$ exists but $(\widehat{r}^0)''_+(\tau) \neq 0$ then there exists a sequence of points $s_i \searrow \tau$, at which we assume without loss of generality at that $\widehat{r}^0$ is differentiable, such that $\widehat{r}^0(\tau) - \widehat{r}^0(s_i) - (\widehat{r}^0)'(s_i)(\tau - s_i) \leq \eta(\tau - s_i)^2$ for some $\eta < 0$ (by concavity). (Note: we can also assume without loss of generality, since $\widehat{r}^0$ is differentiable at τ and $(\widehat{r}^0)'$ is monotonic, that $(\widehat{r}^0)'(s_i) \to (\widehat{r}^0)'(\tau)$.) Thus, for a sequence $\gamma_i \geq 1$, by differentiability at s_i,

$$\eta(\tau - s_i)^2\gamma_i = \widehat{r}^0(\tau) - \widehat{r}^0(s_i) - (\tau - s_i)\left(\widehat{r}^0\right)'(s_i) = o(\tau - s_i) \quad \text{as } s_i \searrow \tau. \tag{49}$$

Thus, set $\epsilon_i := (-\eta)(\gamma_i)(s_i - \tau)$. By (49), $\epsilon_i \searrow 0$, and with $\delta_i = s_i - \tau$, we see $\delta_i = O(\epsilon_i)$ as $i \to \infty$, since $\gamma_i \geq 1$ and $\eta < 0$ is fixed.

Our first goal is to show

$$\lim_{i\to\infty} \frac{\phi_c\left(r_{\epsilon_i}\right) - \phi_c\left(\widehat{r}^0\right)}{\epsilon_i} = \lim_{\epsilon \searrow 0} \frac{\phi_c\left(\widehat{r}^0 - \epsilon\Delta_\tau\right) - \phi_c\left(\widehat{r}^0\right)}{\epsilon}. \tag{50}$$

Note that if τ is an isolated knot then $r_{\epsilon,\tau} = \widehat{r}^0 - \epsilon\Delta_\tau$ for ϵ small enough, and also that for ϵ small enough, $\widehat{r}^0 - \epsilon\Delta_\tau$ is concave (thus for small ϵ, $-\Delta_\tau$ preserves concavity although it is not an acceptable perturbation). If τ is not an isolated knot then $\widehat{r}^0 - \epsilon\Delta_\tau$ is not concave. However, r_{ϵ_i,s_i} is indeed concave and by (50), we can use $\tilde{\Delta}_i := \widehat{r}^0 - r_{\epsilon_i,s_i}$ in place of $\epsilon\Delta_\tau$. Now, notice that

$$\begin{aligned}\lim_{i\to\infty} \frac{\phi_c\left(r_{\epsilon_i}\mathbb{1}_{[s_i,\infty)}\right) - \phi_c\left(\widehat{r}^0\mathbb{1}_{[s_i,\infty)}\right)}{\epsilon_i} &= \lim_{i\to\infty} \frac{\phi_c\left(\widehat{r}^0 - \epsilon_i\Delta_{s_i}\right) - \phi_c\left(\widehat{r}^0\right)}{\epsilon_i} \\ &= -\int \Delta_\tau\left(\widehat{r}^0 d\lambda - dX\right),\end{aligned} \tag{51}$$

by the same calculation as in (45), since $s_i \to \tau$. Here, for a set A, $\mathbb{1}_A(u)$ is 1 if $u \in A$ and 0 otherwise. Thus, we will show

$$\lim_{i\to\infty} \frac{\phi_c\left(r_{\epsilon_i}\mathbb{1}_{(-\infty,s_i)}\right) - \phi_c\left(\widehat{r}^0\mathbb{1}_{(-\infty,s_i)}\right)}{\epsilon_i} = 0, \tag{52}$$

and then conclude that (50) holds. We assume without loss of generality that $(\widehat{r}^0)'(s_i-) = (\widehat{r}^0)'(s_i+)$ (since if this does not hold for an infinite subsequence of

$\{s_i\}$, then we can take the subsequence as our sequence, and then $-\Delta_{s_i}$ preserves concavity, $\delta_i = 0$, and (52) is immediate). Now $\phi_c(r_{\epsilon_i}\mathbb{1}_{(-\infty,s_i)}) - \phi_c(\widehat{r}^0\mathbb{1}_{(-\infty,s_i)})$ equals

$$\begin{aligned}&\frac{1}{2}\int_\tau^{s_i}\left(\left(\widehat{r}^0(s_i)+\left(\left(\widehat{r}^0\right)'(s_i+)+\epsilon_i\right)(u-s_i)\right)^2-\left(\widehat{r}^0\right)^2(u)\right)du\\&-\int_\tau^{s_i}\left(\widehat{r}^0(s_i)-\widehat{r}^0(u)+\left(\left(\widehat{r}^0\right)'(s_i+)+\epsilon_i\right)(u-s_i)\right)dX(u).\end{aligned}\tag{53}$$

The first term in (53) equals

$$\begin{aligned}&\frac{1}{2}\int_\tau^{s_i}\Bigg[\left(\widehat{r}^0(s_i)+\left(\widehat{r}^0\right)'(s_i-)(u-s_i)\right)^2+2\epsilon_i\left(\widehat{r}^0(s_i)\right.\\&\left.+\left(\widehat{r}^0\right)'(s_i-)(u-s_i)\right)(u-s_i)+\epsilon_i^2(u-s_i)^2\\&-\Bigg(\left(\widehat{r}^0(s_i)+\widehat{r}^0(s_i-)(u-s_i)\right)^2\\&+o(u-s_i)\left(\widehat{r}^0(s_i)+\widehat{r}^0(s_i-)(u-s_i)\right)+o(u-s_i)^2\Bigg)\Bigg]du\end{aligned}$$

and since $\delta_i = O(\epsilon_i)$, the previous display is $O(\epsilon_i^2)$. Recalling $\tilde{\Delta}_i = \widehat{r}^0 - r_{\epsilon_i,s_i}$ is, on $[\tau, s_i]$, the integrand of the second term in (53), we see that the negative of the second term in (53) equals

$$\left(X\tilde{\Delta}_i\right)(\tau,s_i]-\int_\tau^{s_i}X(u)\left(\left(\widehat{r}^0\right)'(s_i-)du-d\left(\widehat{r}^0\right)(u)\right).\tag{54}$$

Recall: for a function g and $a < b$, we let $g(a,b] := g(b) - g(a)$. Since $\tilde{\Delta}_i(s_i) = 0$, $\delta_i = O(\epsilon_i)$ as $\epsilon_i \searrow 0$, $\tilde{\Delta}_i(s_i - \delta_i) = o(\delta_i) = o(\epsilon_i)$ (recall $(\widehat{r}^0)'(s_i+) = (\widehat{r}^0)'(s_i-)$), and X is continuous, (54) is $o(\epsilon_i)$. Thus both terms in (53) are $o(\epsilon_i)$ and so we have shown (52), so by (51), we have shown (50).

From now on we take $\tau = \tau_R$. In the case where $\tau_R \notin \widehat{S}^0$ but $\tau_R = \lim_{i\to\infty}\tau_i$ with $\tau_i \in \widehat{S}^0$, the below arguments go through with $\tau = \tau_i$ and taking the limit of τ_i.

We now show Condition 1 holds. Recall $\Delta_t(u) := (t-u)_-$ for $t \geq 0$ and $\Delta_t(u) := (u-t)_-$ for $t < 0$. Let $\Delta_-(u) := -u$. Now let

$$\Delta_{1,i}(u) := -\frac{c\epsilon_i\Delta_{\tau_L}}{\Delta_{\tau_L}(-c)}+\epsilon_i\Delta_-(u)+\tilde{\Delta}_i(u)\frac{c}{-\Delta_{s_i}(c)}.\tag{55}$$

Note that $\Delta_{1,i}$ is an acceptable perturbation with $\Delta_{1,i}(\pm c) = 0$ and $\Delta_{1,i}(0) = 0$. Furthermore, as in the proof of (50), we can show that

$$0 \le \lim_{i\to\infty} \frac{\phi_c\left(\widehat{r}^0 + \Delta_{1,i}\right) - \phi_c\left(\widehat{r}^0\right)}{\epsilon_i} = \int_{-c}^{c} \left(-\frac{c}{\Delta_{\tau_L}(-c)}\Delta_{\tau_L} + \Delta_- + \Delta_{\tau_R}\frac{c}{\Delta_{\tau_R}(c)} \right) \left(\widehat{r}^0 d\lambda - dX \right). \tag{56}$$

Let Δ_1 denote the integrand on the right side of (56). By (46), the right side of (56) equals $-\int_{-c}^{0} \Delta_1\, d(F_L - X_L) + \int_0^c \Delta_1\, d(F_R - X_R)$ which equals

$$-((F_L - X_L)\Delta_1)(-c, 0] + \int_{-c}^{0} (F_L - X_L)\, d\Delta_1 + ((F_R - X_R)\Delta_1)(0, c] - \int_0^c (F_R - X_R)\, d\Delta_1$$

and, because $\Delta_1(\pm c) = 0$, $\Delta_1(0) = 0$, the previous display equals

$$\int_{-c}^{0} (F_L - X_L)\, d\Delta_1 - \int_0^c (F_R - X_R)\, d\Delta_1 = (H_L - Y_L)(0) - (H_R - Y_R)(0)$$

since by definition $(H_L - Y_L)(-c)$, $(H_R - Y_R)(c)$, $(H_R - Y_R)(\tau_R)$, and $(H_L - Y_L)(\tau_L)$ are all 0. This shows that $(H_L - Y_L)(0) - (H_R - Y_R)(0) \ge 0$.

The perturbation $\Delta_{1,i}$ is based about Δ_-. Since Δ_- does not satisfy the side constraints at $\pm c$, we modified Δ_- by adding two further perturbations, (constant multiples of) Δ_{τ_L} and Δ_i, to yield $\Delta_{1,i}$. The perturbation Δ_i is approximately equal to $-\Delta_{s_i}$, but modified so as to preserve concavity, and $-\Delta_{s_i}$ is approximately equal to $-\Delta_{\tau_R}$. A totally symmetric argument allows us to use a perturbation based around $-\Delta_-(u) = u$ that is modified by adding (constant multiples of) Δ_{τ_R} and a perturbation that approximates $-\Delta_{\tau_L}$. This shows that $(H_L - Y_L)(0) - (H_R - Y_R)(0) \le 0$, and allows us to conclude that Condition 1 holds.

Now let

$$\Delta_{2,i}(u) := \epsilon_i \Delta_t(u) + \tilde{\Delta}_i(u)\frac{\Delta_t(c)}{\Delta_{s_i}(c)}. \tag{57}$$

Then $\Delta_{2,i}(c) = 0$ and $\Delta_{2,i}(t \wedge \tau_R) = 0$ where $a \wedge b = \min(a, b)$. Thus $\Delta_{2,i}$ is an acceptable perturbation for all i, and, as above one can check that

$$0 \le \lim_{i\to\infty} \epsilon_i^{-1}\left(\phi_c\left(\widehat{r}^0 + \Delta_{2,i}\right) - \phi_c\left(\widehat{r}^0\right)\right) = \lim_{\epsilon \searrow 0} \epsilon^{-1}\left(\phi_c\left(\widehat{r}^0 + \epsilon\Delta_2\right) - \phi_c\left(\widehat{r}^0\right)\right)$$

where $\Delta_2 := \Delta_t - \Delta_{\tau_R}\Delta_t(c)/\Delta_{\tau_R}(c)$. Thus, from (45),

$$0 \le \int_0^c \Delta_2\, d(F_R - dX_R) = ((F_R - X_R)\Delta_2)(t \wedge \tau_R, c] - \int_0^c (F_R - X_R)\, d\Delta_2. \tag{58}$$

The term $((F_R - X_R)\Delta_2)(t \wedge \tau_R, c]$ equals 0 because Δ_2 is 0 at both $t \wedge \tau_R$ and at c. Since $\int (F_R - X_R) d\Delta_{\tau_R} = (H_R - Y_R)(\tau_R, c] = 0$, we see (58) equals

$$(H_R - Y_R)(t, c] = -(H_R - Y_R)(t).$$

This shows Condition 2 holds for $t \geq 0$. The argument for $t \leq 0$ is analogous.

Now let

$$\Delta_{3+,i} := \epsilon_i \widehat{r}^0_{\tau_R} + \tilde{\Delta}_i \frac{\widehat{r}^0_{\tau_R}(c)}{\Delta_{s_i}(c)}$$

where $\widehat{r}^0_{\tau_R} := \mathbb{1}_{[\tau_R,\infty)}(\widehat{r}^0 - \widehat{r}^0(\tau_R))$ is continuous, concave, and satisfies $\widehat{r}^0_{\tau_R}(\tau_R) = 0$. As above, we can check that

$$0 \leq \lim_{i\to\infty} \epsilon_i^{-1} \left(\phi_c\left(\widehat{r}^0 + \Delta_{3+,i}\right) - \phi_c\left(\widehat{r}^0\right)\right) = \int_0^c \Delta_{3+}\, d(F_R - X_R) \tag{59}$$

where $\Delta_{3+} := \widehat{r}^0_{\tau_R} - \Delta_{\tau_R}\widehat{r}^0_{\tau_R}(c)/\Delta_{\tau_R}(c)$. Then (59) equals

$$((F_R - X_R)\Delta_{3+})(\tau_R, c] - \int_{\tau_R}^c (F_R - X_R)\, d\Delta_{3+} = -\int_{\tau_R}^c (F_R - X_R)\, d\Delta_{3+} \tag{60}$$

since Δ_{3+} is 0 at τ_R and c. Then, since $(H_R - Y_R)$ is 0 at τ_R and at c, (60) equals

$$\begin{aligned} -\int_{\tau_R}^c (F_R - X_R)\, d\widehat{r}^0 &= -\left((H_R - Y_R)\left(\widehat{r}^0\right)'\right)(\tau_R, c] + \int_{(\tau_R,c]} (H_R - Y_R) d\left(\widehat{r}^0\right)' \\ &= \int_{(\tau_R,c]} (H_R - Y_R) d\left(\widehat{r}^0\right)' \end{aligned} \tag{61}$$

Here we used that $(\widehat{r}^0)'(c)$ is finite, which follows from the same argument used in the proof of sufficiency since we have already shown that Condition 2 holds. Thus, we have shown that $\int_{(\tau_R,c]}(H_R - Y_R)d(\widehat{r}^0)' \geq 0$. To show the reverse inequality, let

$$\Delta_{3-,i} := -\epsilon_i \widehat{r}^0_{\tau_R} + \epsilon_i \Delta_{\tau_R} \frac{\widehat{r}^0_{\tau_R}(c)}{\Delta_{\tau_R}(c)}.$$

Notice that $\widehat{r}^0 + \Delta_{3-,i}$ is concave by checking its right and left derivatives at τ_R:

$$\begin{aligned} &\left(\widehat{r}^0\right)'(\tau_R+) - \epsilon_i \left(\widehat{r}^0\right)'(\tau_R+) - \epsilon_i \frac{\widehat{r}^0_{\tau_R}(c)}{\Delta_{\tau_R}(c)} \leq \left(\widehat{r}^0\right)'(\tau_R+) \leq \left(\widehat{r}^0\right)'(\tau_R-) \\ &= \left(\widehat{r}^0 + \Delta_{3-,i}\right)'(\tau_R-), \end{aligned} \tag{62}$$

since $(\widehat{r}^0)'(\tau_R+) \geq -\widehat{r}^0_{\tau_R}(c)/\Delta_{\tau_R}(c)$ by concavity. By (62), we see that $(\widehat{r}^0 + \Delta_{3-,i})'$ is monotonic in a neighborhood of τ_R and thus is monotonic everywhere. Thus $\Delta_{3-,i}$ is

an acceptable perturbation for all i. Then $\Delta_{3-,i}$ is approximately equal to $-\Delta_{3+,i}$, and replicating the arguments in displays (59)–(61) shows that $\int_{(\tau_R,c]}(H_R-Y_R)d(\widehat{r}^0)' \leq 0$. Thus, we can conclude $\int_{(\tau_R,c]}(H_R-Y_R)d(\widehat{r}^0)' = 0$, and an analogous argument shows $\int_{[-c,\tau_L)}(H_L - Y_L)\,d(\widehat{r}^0)' = 0$. We can extend the domain of integration to include $[0, \tau_R]$ or $[\tau_L, 0]$, respectively, since $(H_R - Y_R)(\tau_R) = 0$ and $(H_L - Y_L)(\tau_L) = 0$ and $(\widehat{r}^0)'' \equiv 0$ on (τ_L, τ_R) by definition. This shows Condition 3 holds and completes the proof of the necessity of Conditions 1, 2, and 3, and thus completes the proof. □

For $c^- < 0 < c^+$ let

$$\phi_c(r) := \frac{1}{2}\int_{c^-}^{c^+} r(u)^2 du - \int_{c^-}^{c^+} r(u)dX(u) \tag{63}$$

[slightly modifying the definition given in (35)].

Corollary 1 *Let $c^- < 0 < c^+$, $k^\pm < 0$, and $M > 0$ be random variables, and let ϕ_c be given by (63). Let $\mathcal{C}^\circ_{c,k} := \left\{r \in \mathcal{C}^0 : M \geq r, r(c^\pm) \geq k^\pm\right\}$. Let $\widehat{r}^0 \in \mathcal{C}^\circ_{c,k}$ and let Y_R, Y_L, H_R, H_L be as in* Theorem 2*, with $\pm c$ replaced by $c^\pm$. Assume further that $|(\widehat{r}^0)'(c^\pm)| < \infty$, where $(\widehat{r}^0)'$ refers to the right or left derivative, and assume that*

$$(H_R - Y_R)'\left(c^+\right) = 0 = (H_L - Y_L)'\left(c^-\right). \tag{64}$$

Then for M large enough, almost surely $\widehat{r}^0$ is a minimizer of ϕ_c and is thus unique in $\mathcal{C}^\circ_{c,k}$ on $[c^-, c^+]$.

Proof The proof of Proposition 5 shows that, for M large enough, there is a minimizer of ϕ_c over $\mathcal{C}^\circ_{c,k}$ and the minimizer is unique on $[c^-, c^+]$. (The minimizer does not necessarily satisfy $\widehat{r}^0(c^\pm) = k^\pm$, since $c^\pm$ are random.) The proof that $\widehat{r}^0$ is indeed such a minimizer follows by a slightly modified version of the sufficiency part of the proof of Theorem 2. The equality (43) follows from (64) (rather than from $\widehat{r}^0(c^\pm) = k^\pm$). Note also that we now assume directly that $|(\widehat{r}^0)'(c^+)| < \infty$ (since $c^\pm$ are random, we do not know that Y_L, Y_R do not have so-called 'one-sided parabolic tangents' at $c^\pm$, respectively). □

The previous theorem and corollary are used to prove the next theorem, which gives characterizing conditions for a so-called "value-constrained invelope process" on all of $\mathbb{R}$. (The term "invelope process" originates in [23].) The process on $\mathbb{R}$ governs the limit distribution of $\widehat{r}^0_n(x_0)$.

Theorem 3 *Let $\widehat{r}^0 \in \mathcal{C}^0$. Define*

$$\left(\widehat{S}^0\right)^c := \left(\widehat{S}^0\left(\widehat{r}^0\right)\right)^c := \left\{t \in \mathbb{R} : \left(\widehat{r}^0_c\right)''(t) = 0\right\}.$$

Then define

$$\tau_R := \inf\left(\widehat{S}^0 \cap [0,\infty)\right), \qquad \tau_L := \sup\left(\widehat{S}^0 \cap (-\infty, 0]\right),$$

and define X, X_L, X_R, Y_L, and Y_R as in (34), (40), and (41). For $t \in \mathbb{R}$, let $H_R(t) := \int_{\tau_R}^{t} \int_{\tau_R}^{u} \widehat{r}^0(v)dvdu$, and $H_L(t) := \int_{t}^{\tau_L} \int_{u}^{\tau_L} \widehat{r}^0(v)dvdu$. Assume

1. $(H_R - Y_R)(0) = (H_L - Y_L)(0)$,
2. *for $t \geq 0$, $(H_R - Y_R)(t) \leq 0$ and for $t \leq 0$, $(H_L - Y_L)(t) \leq 0$,*
3. *and*

$$\int_{(-\infty,0]} (H_L - Y_L)\, d\left(\widehat{r}^0\right)' = 0 = \int_{[0,\infty)} (H_R - Y_R)\, d\left(\widehat{r}^0\right)'. \tag{65}$$

Then $\widehat{r}^0$ is unique.

Note that H_L, H_R have different definitions in Theorem 2 and in Theorem 3.

Proof We need several lemmas for the proof. The following lemma connects the H-processes to the Gaussian processes about which we can make explicit statements and computations.

Lemma 1 *Let $\tau_1, \tau_2 \in \widehat{S}^0$ with $0 < \tau_1 < \tau_2$ be such that $\widehat{r}^0$ is affine on $[\tau_1, \tau_2]$ and let $t \in [\tau_1, \tau_2]$. Define, for any function g, $\nabla g = g(\tau_2) - g(\tau_1)$, $\bar{g} = (g(\tau_1) + g(\tau_2))/2$, $\nabla\tau = \tau_2 - \tau_1$, and $\bar{\tau} = (\tau_1 + \tau_2)/2$. Then*

$$H_R(t) = \frac{Y_R(\tau_2)(t - \tau_1) + Y_R(\tau_1)(\tau_2 - t)}{\nabla\tau} - \frac{1}{2}\left(\frac{\nabla X_R}{\nabla\tau} + \frac{4}{(\nabla\tau)^3}(\bar{X}_R\nabla\tau - \nabla Y_R\right)(t - \tau_1)(\tau_2 - t),$$

and so $H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\nabla X_R \nabla\tau$. Analogous formulas can be stated for the left-side processes.

Proof The proof follows from the proofs of Lemma 2.3 of [23], and Lemma 8.9 of [14]. □

The previous lemma is used to prove the next lemma, about the "knot" behavior of $\widehat{r}^0$.

Lemma 2 *Fix $t > 0$. Let $\tau_+(t)$ be the infimum of the points of touch of Y_R and H_R in $[t, \infty)$. Then for all $\epsilon > 0$, there exists M, independent of t, such that $P(\tau_+(t) - t > M) < \epsilon$. An analogous statement can be made for the left-side processes, $t < 0$, and the supremum of the points of touch of Y_L and H_L in $(-\infty, t]$.*

Proof The result follows from Lemma 1, via the analysis used in the the proof of Lemma 8.10 of [14] (see also Lemma 2.7 of [23]). □

The uniqueness of $\widehat{r}^0$ follows from showing that if two different processes both satisfy the characterizing conditions of the theorem then they are equal. One considers the cases where the two processes share (sequences of) knots (converging to infinity) or they do not. The following lemma handles the former case.

Lemma 3 *Suppose $G_{R,1}$ and $G_{R,2}$ both satisfy the conditions of* Theorem 3 *on H_R and $G_{L,1}$ and $G_{L,2}$ satisfy the theorem conditions for H_L (we do not assume a priori*

that they have the same value for τ_R or τ_L, respectively). Let $r_1 := G''_{R,1} \equiv G''_{L,1}$ and $r_2 := G''_{R,2} \equiv G''_{L,2}$. If $(G_{R,i} - Y_{R,i})(s_R) = 0$ and $(G_{L,i} - Y_{L,i})(s_L) = 0$, $i = 1, 2$, where $Y_{R,i}$, $Y_{L,i}$ are defined by (41) based on the knots of r_i, $i = 1, 2$, then $r_1 = r_2$ on $[s_L, s_R]$.

Proof This follows from Corollary 1. We let Y_L, Y_R, H_L and H_R be as defined in Theorem 3, we assume $H_R(s_R) - Y_L(s_R) = 0$ and $H_L(s_L) - Y_L(s_L) = 0$, and we will show that the conditions of Corollary 1 are satisfied by H_L and H_R. This will then show the statement of the lemma. Since $H_R(\tau_R) - Y_R(\tau_R) = 0$ by definition, we see that H_R is the primitive of the primitive of H''_R satisfying the constant conditions (at $c^+ = s_R$ and at τ_R) used to define H_R in Corollary 1. Furthermore, by Condition 2 and because $(H_R - Y_R)(\tau_R) = 0$, we see that $(H_R - Y_R)'(\tau_R) = 0$. A similar argument can be made for the left-side processes. Since $H_R^{(3)}$ is finite on $\mathbb{R}$, the condition $|H_R^{(3)}(c^{\pm})| < \infty$ is automatically satisfied (for either the left or right third derivative). Therefore we have shown that the conditions of Corollary 1 are satisfied. We apply this to $G_{R,1}$ and $G_{R,2}$. Let

$$M := \sup_{x \in [s_L, s_R]} \{r_1(x), r_2(x)\}, \; k^+ := \min(r_1(s_R), r_2(s_R)),$$
$$\text{and } k^- := \min(r_1(s_L), r_2(s_L)).$$

Then $r_i \in \mathcal{C}^{\circ}_{c,k}$ for $i = 1, 2$, and both the $i = 1$ and $i = 2$ processes satisfy the conditions of Corollary 1 by the argument in the previous paragraph, so $r_1 = r_2$ on $[s_L, s_R]$ as desired. □

For the remainder of the proof of Theorem 3, one considers cases where on either the left side, the right side, or both sides, there is no sequence of shared touch points converging to infinity, and deriving a contradiction. The argument follows as in the proof of Theorem 5.2 of [14].

This completes the proof of Theorem 3. □

Remark 1 If X is replaced by $X_{a,\sigma}(t) := \sigma W(t) - 4at^3$ for constants $a, \sigma > 0$, then the conclusion of Theorem 3 still holds; in this case, we denote the process $\widehat{r}^0$ of the theorem by $\widehat{r}^0_{a,\sigma}$.

Remark 2 The knot definitions in Theorem 3 differ from those in Theorem 5.2 of [14], in the context of a mode constraint. Condition (iii) of Theorem 5.2 of [14] (which is analogous to Condition 3 of Theorem 2) is based on knots τ^0_+ and τ^0_-, one (but almost surely not both) of which may be 0. These knots are potentially distinct from the knots τ_L and τ_R in that setup, where τ_L, τ_R can never be 0. In the height-constrained problem we consider in this paper, there is only one pair of knots, τ_L, τ_R, and they may be 0; if one is 0 then both are 0.

4 Asymptotics

We can now study the asymptotic behavior of $\widehat{r}^0_n$. To do so, we will make the following assumptions on the design.

Assumption 1 The design points $x_{n,i} \in [0,1]$ satisfy $c/n \leq x_{n,i+1} - x_{n,i} \leq C/n$, $i = 1, \ldots, n-1$ for some $0 < c < C < \infty$.

Assumption 2 For $0 \leq x \leq 1$, let $F_n(x) := n^{-1} \sum_{i=1}^n \mathbb{1}_{[0,x]}(x_{n,i})$. There exists $\delta > 0$ such that $\sup_{x:|x-x_0|\leq\delta} |F_n(x) - x| = o(n^{-1/5})$.

Theorem 4 *Suppose that the regression model* (9) *holds where r_0 is concave with $r_0(x_0) = y_0$, suppose that $\epsilon_{n,1}, \ldots, \epsilon_{n,n}$ are i.i.d. with $E^{\epsilon_{n,1}^2 t} < \infty$ for some $t > 0$, that r_0 is twice continuously differentiable in a neighborhood of x_0, and that $r_0''(x_0) < 0$. Let* Assumptions 1 *and* 2 *hold. Let $a = |r_0''(x_0)|/24$, and $\sigma^2 = \operatorname{Var} \epsilon_{n,i}$. Let $\widehat{r}_n^0 = \operatorname{argmin}_r \phi_n(r)$ where the argmin is over concave functions r such that $r(x_0) = y_0$. Then*

$$n^{2/5}\left(\widehat{r}_n^0\left(x_0 + tn^{-1/5}\right) - r_0(x_0) - r_0'(x_0)tn^{-1/5}\right) \to_d \widehat{r}_{a,\sigma}^0(t)$$

in $L^p[-K, K]$ for all $K > 0$, where $\widehat{r}_{a,\sigma}^0$ is given in Theorem 3 *(and* Remark 1*).*

Remark 3 We suspect asymptotic distributions and the Wilks phenomenon for $2\log\lambda_n(y_0)$ can be derived under more general conditions than Assumption 2, but this assumption is used by Groeneboom et al. [24] (it is their Assumption 6.1) to derive the limit distribution of $\widehat{r}_n(x_0)$, so we rely on it here too and leave generalizations for future research.

Remark 4 We require a sub-Gaussian tail assumption on $\epsilon_{n,i}$ in Theorem 4. In [9], the asymptotic distribution for a monotone regression function estimator is derived under only second moment assumptions for the error variables. However, for deriving the rates of convergence for concave regression least-squares estimators, Mammen [34, Theorem 4] (and then [24]) assume sub-Gaussian tails on the error variables. Mammen [34, page 749] states, "We do not believe that this strong condition is really necessary." However, in the present paper we have not attempted to weaken this assumption.

Proof of Theorem 4 We take $x_0 = 0$ for simplicity and take $r_0(0) = 0$ and $r_0'(0) = 0$ by the translation discussed in Sect. 2. Let $x_n(t) := tn^{-1/5} = tn^{-1/5} + x_0$ be the "global" parameter corresponding to the "local" parameter $t \in \mathbb{R}$. Then let $\tau_{n,R}$ be the smallest nonnegative bend point of $\widehat{r}_n^0$.

Recall $F_n(x) := n^{-1} \sum_{i=1}^n \mathbb{1}_{[0,x]}(x_{n,i})$. Then define

$$\mathbb{S}_{n,R}(v) := \int_{[\tau_{n,R},v]} Y(u)\, dF_n(u) \qquad Y_{n,R}(t) := n^{4/5} \int_{\tau_{n,R}}^{x_n(t)} \mathbb{S}_{n,R}(u)du$$

$$\mathbb{R}_{n,R}(v) := \int_{[\tau_{n,R},v]} \widehat{r}_n^0(u)\, dF_n(u) \qquad H_{n,R}(t) := n^{4/5} \int_{\tau_{n,R}}^{x_n(t)} \mathbb{R}_{n,R}(u)du + A_{n,R}\left(t - n^{1/5}\tau_{n,R}\right),$$

where Y is the function such that $Y(x_{n,i}) = Y_{n,i}$ (and whose value is 0 elsewhere), and where

$$A_{n,R} := n^{3/5} \int_{[\tau_{n,R},\infty)} Y(u) - \widehat{r}_n^0(u)\, dF_n(u).$$

Define also

$$\tilde{\mathbb{R}}_{n,R}(v) := \int_{[\tau_{n,R},v]} \widehat{r}_n^0(u)du, \quad \tilde{H}_{n,R}(t) := n^{4/5} \int_{[\tau_{n,R},x_n(t)]} \tilde{\mathbb{R}}_{n,R}(u)du + A_{n,R}\left(t - n^{1/5}\tau_{n,R}\right)$$

which we will show to be equivalent to $\mathbb{R}_{n,R}$ and to $H_{n,R}$, respectively. For brevity, we will make definitions and arguments only for the right-side processes. Analogous definitions and arguments can be made for the left-side processes.

By Theorem 1, one can check that

$$H_{n,R}(t) - Y_{n,R}(t) \leq 0$$

for all $t \geq 0$, with equality if t is a knot point of $\widehat{r}_n^0$ (see Lemma 8.18 of [14] for similar calculations). Additionally, defining $Y_{n,L}$ and $H_{n,L}$ in an analogous fashion as $Y_{n,R}$ and $H_{n,R}$, we can check that

$$H_{n,L}(0) - Y_{n,L}(0) = H_{n,R}(0) - Y_{n,R}(0). \tag{66}$$

Next, we can check that

$$\sup_{|t| \leq c} \left| H_{n,R}(t) - \tilde{H}_{n,R}(t) \right| = o_p(1) \tag{67}$$

for any $c > 0$, by Assumption 2 [24, see page 1696]. One can define $\tilde{H}_{n,L}$ and make an analogous statement for $\tilde{H}_{n,L}$ and $H_{n,L}$.

We can then conclude that

$$\tilde{H}_{n,L}(0) - Y_{n,L}(0) = \tilde{H}_{n,R}(0) - Y_{n,R}(0) + o_p(1), \tag{68}$$

$$\tilde{H}_{n,R}(t) - Y_{n,R}(t) + o_p(1) \leq 0, \tag{69}$$

$$\tilde{H}_{n,L}(t) - Y_{n,L}(t) + o_p(1) \leq 0 \tag{70}$$

where the inequalities are equalities for knot points of $\widehat{r}_n^0$.

Let $\mathbb{S}_n(t) := n^{-1} \sum_{i=1}^n Y_{n,i} \mathbb{1}_{\{x_{n,i} \leq t\}}$, and let $Y_n(t) := \int_{x_0}^{x_n(t)} (\mathbb{S}_n(v) - \mathbb{S}_n(x_0)) dv$. Then, for any $c > 0$, we can then check that Y_n converges weakly to $\sigma \int_0^t W(s) ds - at^4 = Y_{a,\sigma}(t)$ in the space of continuous functions on $[-c, c]$ with the uniform metric ([24, (6.12), page 1694]). A similarly structured argument shows that along certain subsequences of $\{n\}_1^\infty$, $Y_{n,R}$ converges to a process $Y_{R,a,\sigma} \equiv Y_R$ (which may a priori depend on the subsequence, but eventually is seen not to depend on the subsequence). The convergence argument for $Y_{n,R}$ requires more care than that for Y_n because the definition of the former depends on the knot $\tau_{n,R}$. Nonetheless it can be rigorously carried out, in a fashion similar to that of the proofs of Lemmas 8.16 and 8.17 of [14].

Then the remainder of the proof follows as in the proof of Theorem 6.3 of [24] (see also [34]) and of Theorem 5.8 of [14]. By Lemma 4 below, $n^{1/5} \tau_{n,R} = O_p(1)$, and this allows us to also conclude that $\tilde{H}_{n,R}$ and its first, second, and third derivatives are all tight in appropriate metric spaces. Then, by Prohorov's theorem, for any subsequence we can find a subsubsequence of $\tilde{H}_{n,R}$ that converges to a limit process, H_R. The processes H_R and Y_R can be shown to satisfy $H_R(t) - Y_R(t) \leq 0$ for $t \geq 0$ and $\int_{[0,\infty)} (H_R - Y_R)\, d(H_R)^{(3)} = 0$ by (69). Arguing analogously for left-side processes, we can see that there are limit processes H_L and $Y_L \equiv Y_{L,a,\sigma}$ satisfying $H_L(t) -$

$Y_L(t) \le 0$, and $\int_{(-\infty,0]}(H_L - Y_L)d(H_L)^{(3)} = 0$ by (70), and by (68) that $H_L(0) - Y_L(0) = H_R(0) - Y_R(0)$. This shows conditions 1, 2, and 3 of Theorem 3 hold for the processes H_L and H_R. Therefore, the limit processes H_L and H_R are unique, so are identical along all subsequences. That is, we can conclude that $\tilde{H}_{n,L}$ and $\tilde{H}_{n,R}$ converge to the unique processes H_L and H_R given by Theorem 3. In particular, we have shown that $n^{2/5}\widehat{r}_n^0(tn^{-1/5}) = (\tilde{H}_{n,R})''(t) \equiv (\tilde{H}_{n,L})''(t)$ converges to $(H_L)'' \equiv (H_R)''$ and so (recalling that $x_0 = 0$ and $r(0) = r'(0) = 0$ by assumption) the proof is complete. □

Lemma 4 *Let the assumptions and terminology of* Theorem 4 *hold, and let* $\tau_{n,R}$ *be the smallest nonnegative bend point of* $\widehat{r}_n^0$. *Then* $n^{1/5}\tau_{n,R} = O_p(1)$.

Proof The proof is by a perturbation argument in the spirit of Theorem 4.3 (and Lemma 4.4) of [3] and Proposition 7.3 of [14] (which in turn are inspired by Lemma 8 of [34]). If 0 is itself a knot of $\widehat{r}_n^0$ then there is nothing to show (because $\tau_{n,R} = 0$). Thus we assume 0 is not a knot of $\widehat{r}_n^0$. We will construct a 'perturbation' $\Delta\colon \mathbb{R} \to \mathbb{R}$ such that $\langle \text{eval}_{n_0}\Delta, \nabla_{n_0}\phi_n(\underline{\widehat{r}}_n^0)\rangle = \langle \text{eval}_{n_0}\Delta, \underline{\widehat{r}}_n^0 - \underline{Y}_n\rangle \ge 0$ as in (17), where $\underline{Y}_n := (Y_{n,1}, \ldots, Y_{n,n})'$ (recalling $Y_{n,k^0} := 0$ if $n_0 = n+1$). This implies

$$\int \Delta(u)\left(\widehat{r}_n^0(u) - r_0(u)\right) dF_n(u) \ge \int \Delta(u)(Y(u) - r_0(u))\, dF_n(u) \tag{71}$$

using the notation developed in the proof of Theorem 4. The approach is to find a Δ such that the quantity on the left side of (71) is a positive constant times $-\tau_{n,R}^4 < 0$, and the quantity on the right side of (71) is $O_p(n^{-4/5})$. The conclusion then follows.

Let $\tau_{n,L} < 0 < \tau_{n,R}$ be the largest negative and smallest positive knots of $\widehat{r}_n^0$, respectively. Assume $\tau_{n,R} \le |\tau_{n,L}|$, without loss of generality.

Let

$$\Delta_1(t) := t\,\mathbb{1}_{[\tau_{n,L},0)} + \left(\frac{\tau_{n,L}}{\tau_{n,R}}\right)^3 t\,\mathbb{1}_{[0,\tau_{n,R}]}$$

which satisfies $\Delta_1(0) = 0$. A simple argument (see Lemma A.4 of [16]) shows that even though Δ_1 is discontinuous, the conclusion of (71) holds, meaning

$$\int \Delta_1(u)\left(\widehat{r}_n^0(u) - r_0(u)\right) dF_n(u) \ge \int \Delta_1(u)(Y(u) - r_0(u))\, dF_n(u). \tag{72}$$

Further,

$$\int \Delta_1(u)u\,du = 0, \tag{73}$$

which will later allow us to ignore a term in a Taylor expansion. (Note that in [3,14] the perturbation must satisfy $\int \Delta(x)dx = 0$; in the present case it turns out we do not need this to hold because of the constraint $\widehat{r}_n^0(0) = 0$. On the other hand, we must have $\Delta(0) = 0$.) Now, the empirical process argument used in the proof of Theorem 4 of [34] shows that the term on the right of (72) is $O_p(n^{-4/5})$. For the term on the left, we can show that $\int \Delta_1(u)(\widehat{r}_n^0(u) - r_0(u))\, dF_n(u) = (1 + o_p(1)) \int \Delta_1(u)(\widehat{r}_n^0(u) - r_0(u))\, du$ as in (67), by Assumption 2. Let $D := \widehat{r}_n^0 - r_0$. Since $\widehat{r}_n^0$ is linear on $(\tau_{n,L}, \tau_{n,R})$

and $\widehat{r}_n^0(0) = 0$, $\widehat{r}_n^0(u) = (\widehat{r}_n^0)'(0)u$ for $u \in [\tau_{n,L}, \tau_{n,R}]$ and so by (73) and a Taylor expansion of r_0 about 0 (recalling that λ is Lebesgue measure),

$$\int \Delta_1 D d\lambda = D(0) \int \Delta_1 d\lambda + D'(0) \int u \Delta_1(u) du$$
$$- \int \Delta_1(u) \frac{r_0''(0)}{2} u^2 (1 + o_p(1)) du \tag{74}$$
$$= - \int \Delta_1(u) \frac{r_0''(0)}{2} u^2 (1 + o_p(1)) du, \tag{75}$$

We compute that $\int u^2 \Delta_1(u) du = (-(\tau_{n,L})^4 + (\tau_{n,L})^3 \tau_{n,R})/4 \leq -(\tau_{n,L})^4/4 < 0$. Thus we can conclude that the quantity on the left of (72) equals $-C_{r_0}(\tau_{n,R} - \tau_{n,L})^4(1 + o_p(1))$ for a constant $C_{r_0} > 0$ (since $r_0''(0) < 0$ and $|\tau_{n,L}| \geq \tau_{n,R}$). Thus the proof is complete. □

4.1 The likelihood ratio statistic

Here we present a partial proof of Conjecture 1. We will break $2 \log \lambda_n$ into two terms, a "main" term and a "remainder" term. We focus on the main term, which drives the limit distribution (according to simulations), and do not analyze the remainder term (which Conjecture 1 and simulations would imply to be asymptotically negligible). To begin, we need to discuss certain rescalings of the processes studied in the previous sections. For $a, \sigma > 0$, let $X_{a,\sigma}(t) := \sigma W(t) - 4at^3$ as in Remark 1, and, correspondingly, let

$$Y_{a,\sigma}(t) := \sigma \int_0^t W(s) ds - at^4 \stackrel{d}{=} \sigma (\sigma/a)^{3/5} Y\left((a/\sigma)^{2/5} t\right), \tag{76}$$

where the equality in distribution can be checked using the fact that $W(\alpha \cdot)\alpha^{-1/2} \stackrel{d}{=} W(\cdot)$ for any $\alpha > 0$. Let $H_{a,\sigma}$ be the invelope process given by Theorem 6 based on $Y_{a,\sigma}$, and let $H_{a,\sigma}^0$ denote either of the (null hypothesis) invelope processes, H_R or H_L, given by Theorem 3 based on $Y_{a,\sigma}$. By (76),

$$H_{a,\sigma}(t) \stackrel{d}{=} \sigma (\sigma/a)^{3/5} H_{1,1}\left((a/\sigma)^{2/5} t\right), \quad \text{and} \quad H_{a,\sigma}^0(t) \stackrel{d}{=} \sigma (\sigma/a)^{3/5} H_{1,1}^0\left((a/\sigma)^{2/5} t\right).$$

Let $\widehat{r}_{a,\sigma}(t) := (H_{a,\sigma})''(t)$ and $\widehat{r}_{a,\sigma}^0(t) := (H_{a,\sigma}^0)''(t)$ (recall $H_L'' \equiv H_R''$). Then we have

$$\widehat{r}_{a,\sigma}(\cdot) \stackrel{d}{=} \sigma^{4/5} a^{1/5} \widehat{r}\left((a/\sigma)^{2/5} \cdot\right) =: \frac{1}{\gamma_1 \gamma_2^2} \widehat{r}(\cdot/\gamma_2), \tag{77}$$

$$\widehat{r}_{a,\sigma}^0(\cdot) \stackrel{d}{=} \sigma^{4/5} a^{1/5} \widehat{r}^0\left((a/\sigma)^{2/5} \cdot\right) =: \frac{1}{\gamma_1 \gamma_2^2} \widehat{r}^0(\cdot/\gamma_2), \tag{78}$$

where we let $\gamma_1 := (a/\sigma)^{3/5}/\sigma$ and $\gamma_2 := (\sigma/a)^{2/5}$. This allows us to relate the rescaled processes $\widehat{r}_{a,\sigma}$ and $\widehat{r}_{a,\sigma}^0$ (where a will later depend on r_0 and $\sigma^2 = \mathrm{Var}(\epsilon_{n,i})$) to the universal processes $\widehat{r}$ and $\widehat{r}^0$. For our future use, we note the relationship

$$\gamma_1 \gamma_2^{3/2} = \sigma^{-1}. \tag{79}$$

We have

$$0 \leq 2\log\lambda_n = 2\left(\phi_n\left(\widehat{r}_n^0\right) - \phi_n\left(\widehat{r}_n\right)\right) = -\sum_{i\in I}\widehat{r}_{n,i}^2 - \left(\widehat{r}_{n,i}^0\right)^2 - \left(2\widehat{r}_{n,i}Y_{n,i} - 2\widehat{r}_{n,i}^0 Y_{n,i}\right). \tag{80}$$

Now by (16), $\langle \widehat{r}_n, \nabla\phi_n(\widehat{r}_n)\rangle = 0$ and $\langle \widehat{r}_n^0, \nabla\phi_n(\widehat{r}_n^0)\rangle = 0$, so (80) equals

$$-\sum_{i\in I}\widehat{r}_{n,i}^2 - \left(\widehat{r}_{n,i}^0\right)^2 - 2\left(\widehat{r}_{n,i}^2 - \left(\widehat{r}_{n,i}^0\right)^2\right) = \sum_{i\in I}\widehat{r}_{n,i}^2 - \left(\widehat{r}_{n,i}^0\right)^2. \tag{81}$$

Now we expect that away from the constraint, $\widehat{r}_{n,i}$ and $\widehat{r}_{n,i}^0$ are asymptotically equivalent. In fact, we expect that (81) can be localized to a sum over indices corresponding to $O_p(n^{-1/5})$ neighborhoods of x_0. To discuss this, we note that (81) can be written as $n\int_{\mathbb{R}}\left(\widehat{r}_n(u)^2 - \widehat{r}_n^0(u)^2\right)dF_n(u)$ (recalling $F_n(x) := n^{-1}\sum_{i=1}^n \mathbb{1}_{[0,x]}(x_{n,i})$). Then we let $x_n(t) := x_0 + n^{-1/5}t$, and can then see that $2\log\lambda_n$ equals $\mathbb{D}_{n,b} + E_{n,b}$ where

$$\mathbb{D}_{n,b} := n\int_{x_n(-b)}^{x_n(b)}\left(\widehat{r}_n(u)^2 - \widehat{r}_n^0(u)^2\right)dF_n(u), \text{ and}$$
$$E_{n,b} := n\int_{\mathbb{R}\setminus[x_n(-b),x_n(b)]}\left(\widehat{r}_n(u)^2 - \widehat{r}_n^0(u)^2\right)dF_n(u).$$

We conjecture that $E_{n,b}$ is asymptotically negligible for large enough n and b. As was discussed in the introduction, proving that $E_{n,b}$ is asymptotically negligible may be quite challenging. A result of this sort was shown fully in [13,14] in the context of a likelihood ratio statistic for the mode of a log-concave density. In some contexts where the underlying shape constraint is one of monotonicity rather than convexity/concavity, the corresponding problem seems to often be simpler [4,6,22]. It is beyond the scope of the present paper to show $E_{n,b}$ is negligible; here, we focus on the non-negligible term $\mathbb{D}_{n,b}$.

Now, by Assumption 2, $\mathbb{D}_{n,b}$ is equal to [24, page 1695]

$$n\int_{x_n(-b)}^{x_n(b)}\widehat{r}_n(u)^2 - \widehat{r}_n^0(u)^2\,du + o(1) = n^{4/5}\int_{-b}^{b}\widehat{r}_n(x_n(v))^2 - \widehat{r}_n^0(x_n(v))^2 dv + o(1). \tag{82}$$

Let $a = |r_0''(x_0)|/24$. Let $\mathbb{S}_n(t) := n^{-1}\sum_{i=1}^n Y_{n,i}\mathbb{1}_{\{x_{n,i}\leq t\}}$, and let $Y_n(t) := \int_{x_0}^{x_n(t)}(\mathbb{S}_n(v) - \mathbb{S}_n(x_0))dv$. Then, for any $c > 0$, we can then check that Y_n converges weakly to $\sigma\int_0^t W(s)ds - at^4 = Y_{a,\sigma}(t)$ in the space of continuous functions on $[-c, c]$ with the uniform metric (see the proof of Theorem 4). Then, by (the proofs of) Theorem 7 and by Theorem 4 (recalling that $r_0(x_0) = 0$ and $r_0'(x_0) = 0$ by our data translation), $n^{2/5}\widehat{r}_n(x_n(\cdot))$ converges weakly to $\widehat{r}_{a,\sigma}$ and $n^{2/5}\widehat{r}_n^0(x_n(\cdot))$ converges

weakly to $\widehat{r}^0_{a,\sigma}$. Thus, the right side of (82) converges in distribution to

$$\begin{aligned}\int_{-b}^{b} \widehat{r}^2_{a,\sigma} - \left(\widehat{r}^0_{a,\sigma}\right)^2 &= \int_{-b}^{b} \left(\frac{1}{\gamma_1\gamma_2^2}\right)^2 \left(\widehat{r}\left(\frac{s}{\gamma_2}\right)^2 - \widehat{r}^0\left(\frac{s}{\gamma_2}\right)^2\right) ds \\ &= \gamma_1^{-2}\gamma_2^{-3} \int_{-b/\gamma_2}^{b/\gamma_2} \left(\widehat{r}(u)^2 - \widehat{r}^0(u)^2\right) du \\ &= \sigma^2 \int_{-b/\gamma_2}^{b/\gamma_2} \left(\widehat{r}(u)^2 - \widehat{r}^0(u)^2\right) du, \end{aligned} \tag{83}$$

as $n \to \infty$, by (77) and (78), and recalling that $\gamma_1^2\gamma_2^3 = \sigma^{-2}$ by (79). Now if we let $b \to \infty$ then (83) converges to

$$\sigma^2 \int_{-\infty}^{\infty} \left(\widehat{r}(u)^2 - \widehat{r}^0(u)^2\right) du =: \mathbb{D}, \tag{84}$$

which does not depend on r_0, as desired. This shows that Conjecture 1 holds, assuming that $E_{n,b}$ is appropriately negligible. We thus now state Conjecture 1 as a theorem under the following assumption on the error term.

Assumption 3 For all small enough $\delta > 0$ there exists $b_\delta > 0$ such that $|E_{n,b_\delta}| \le \delta K$ where $K = O_p(1)$ does not depend on δ.

Theorem 5 *Assume the regression model* (1) *holds where* $Ee^{t\epsilon_{n,i}^2} < \infty$ *for some* $t > 0$. *Assume* r_0 *is concave,* $r_0(x_0) = y_0$, r_0 *is twice continuously differentiable in a neighborhood of* x_0, *and* $r_0''(x_0) < 0$. *Let* Assumption 1 *and* 2 *hold. Define* $2\log\lambda_n(y_0)$ *as in* (6). *If* Assumption 3 *holds, then* $2\log\lambda_n(y_0) \to_d \sigma^2\mathbb{D} := \sigma^2 \int_{-\infty}^{\infty} \widehat{r}(u)^2 - \widehat{r}^0(u)^2\, du$.

Proof For any $\delta > 0$, for a subsequence of $\{n\}_{n=1}^{\infty}$, there exists a subsubsequence such that along the subsubsequence $E_{n,b_\delta} \to_d \delta R$ where $|R| \le K$ almost surely, by Prohorov's theorem and Assumption 3. Thus since $\mathbb{D}_{n,b_\delta} \to_d \sigma^2 \int_{-b_\delta/\gamma_2}^{b_\delta/\gamma_2} \widehat{r}(u)^2 - \widehat{r}^0(u)^2\, du =: \sigma^2\mathbb{D}_{b_\delta}$ as $n \to \infty$ by (83), we see that $2\log\lambda_n \to_d \sigma^2\mathbb{D}_{b_\delta} + \delta R$ along the subsubsequence. Taking, say, $\delta = 1$, we see that $2\log\lambda_n$ has a (tight) limit, which we denote by $\sigma^2\mathbb{D}$, along the subsubsequence. Since K does not depend on δ, we can let $\delta \searrow 0$ so $\delta R \to_p 0$, and since then $b_\delta \nearrow \infty$ we thus see that $\sigma^2\mathbb{D} := \sigma^2\mathbb{D}_{b_\delta} + \delta R \to \sigma^2 \int_{-\infty}^{\infty} \widehat{r}(u)^2 - \widehat{r}^0(u)^2\, du$ so $\mathbb{D} = \int_{-\infty}^{\infty} \widehat{r}(u)^2 - \widehat{r}^0(u)^2\, du$. Thus, along the subsubsequence $2\log\lambda_n \to_d \sigma^2 \int_{-\infty}^{\infty} \widehat{r}(u)^2 - \widehat{r}^0(u)^2\, du$; since this holds for an arbitrary subsequence, the convergence holds along the original sequence. This completes the proof. □

5 Simulations

We now use simulation studies to assess our procedures. First, we give evidence in Fig. 1 that Conjecture 1 holds. We simulated from three different true concave

Table 1 Characteristics of the true concave regression function used in the Monte Carlo simulations

r_0	x_0	$r_0(x_0)$	$r_0'(x_0)$	$r_0''(x_0)$	$d(r_0)$
$-x^2$	0	0	0	-2	1.64
$\cos(x)$	$-\frac{1}{2}$	.878	.479	$-.878$	1.94
$-\exp(x)$	2	-7.39	-7.39	-7.39	1.27

regression functions, $-x^2$, $\cos(x)$, and $-\exp(x)$. We used a fixed design setting, with $n = 1000$ points uniformly spaced along an interval. For $-x^2$ and $\cos(x)$ the intervals were $[-1, 1]$. For $-\exp(x)$ the interval was $[1, 3]$. We used standard normal error terms. Figure 1 gives empirical cdfs based on $M = 5000$ Monte Carlo replications of the distribution of $2\log\lambda_{1000}$ for the three regression functions. The curves are visually indistinguishable, giving evidence in support of Conjecture 1. The curve labeled "limit" is based on simulating directly from the distribution of $\mathbb{D}$. To do this, we simulated the process $X(t) = W(t) - 4t^3$ and computed the limit process $\widehat{r}$ from Theorem 6 and $\widehat{r}^0$ from Theorem 2 based on the 'data' X. We then computed $\mathbb{D} = \int_{\mathbb{R}}(\widehat{r}^2(t) - (\widehat{r}^0)^2(t))dt$. The actual form of the limit is not fundamental to Conjecture 1. However the simulation results reported in Fig. 1 appear to indeed show that $\mathbb{D}$ has this form, since the "limit" curve is visually indistinguishable from the other three curves described above. The final curve is the cdf of a chi-squared distribution with 1 degree of freedom. This would be the limit of the likelihood ratio statistic if this were a regular parametric problem, but is distinct from the limit of our likelihood ratio statistic, in this nonparametric problem.

Thus, with Conjecture 1 in mind, we implemented our likelihood ratio test for the hypothesis test (4), rejecting when $2\log\lambda_n(y_0) > d_\alpha$, $\alpha \in (0, 1)$, where d_α is based on the simulated limit distribution in Fig. 1. Specifically, we used the curve based on $-x^2$ with a Gaussian error distribution as the limit distribution for $2\log\lambda$. We tested the level under the null hypothesis via Monte Carlo. Our simulations were based on sample sizes of either $n = 30$, $n = 100$, or $n = 1000$, and $M = 20{,}000$ Monte Carlo replications. We used the three r_0's of $-x^2$, $\cos(x)$, and $-\exp(x)$ again on the same intervals listed above. Two designs were used for each of the r_0's; a fixed design, uniformly spaced, and a random uniform design (although the random design is not covered by our theory). The reported results are for a standard normal error distribution. Table 1 gives the x_0 used for each function's hypothesis test, and $r_0(x_0)$, the true value (which was used for the null hypothesis). We also report smoothness characteristics of r_0 at x_0, which could in general affect inference procedures, including the constant $d(r_0) := (24/\sigma^4|r_0''(x_0)|)^{1/5}$ with $\sigma = 1$, from (3). Table 2 gives the simulated levels from the Monte Carlo experiments. The third and fourth columns give the Monte Carlo level of the test procedure for the two nominal levels of $\alpha = .05$ and $\alpha = .1$, respectively, in the fixed design setting. The fifth and sixth columns give the results in the random design settings. The results for $\cos(x)$ were generally the worst, which is perhaps attributable to having x_0 closer to the edge of the covariate design interval than in the scenarios for the other two regression functions. Shape constrained estimators suffer near the covariate domain boundary. We do not present simulation results for coverage of our confidence intervals, since by definition the

Table 2 Monte Carlo level of the likelihood ratio test procedure for nominal levels $\alpha = .05, .1$

r_0	$\alpha = .05$; f	$\alpha = .1$; f	$\alpha = .05$; r	$\alpha = .1$; r
$n = 1000$				
$-x^2$	.0466	.0964	.0579	.112
$\cos(x)$	.0501	.106	.0680	.131
$-\exp(x)$	.0464	.0964	.0469	.0957
$n = 100$				
$-x^2$	.0527	.108	.0514	.107
$\cos(x)$	.0670	.125	.0650	.126
$-\exp(x)$	.0451	.0990	.0447	.0979
$n = 30$				
$-x^2$	.0591	.116	.0578	.110
$\cos(x)$	.0683	.127	.0694	.127
$-\exp(x)$	.0495	.105	.0455	.0996

The column heading "f" denotes fixed design, and "r" denotes random design. Results are based on sample sizes of $n = 30$, 100, and 1000, and $M = 20{,}000$ Monte Carlo replications

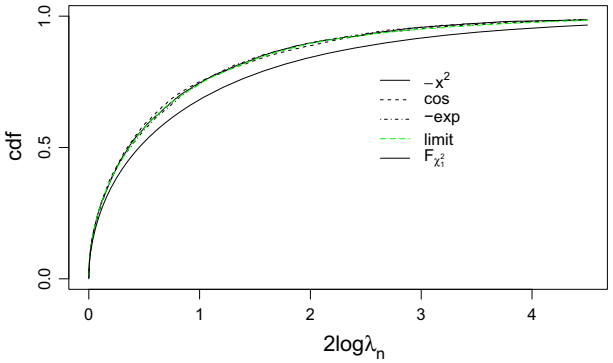


Fig. 1 Empirical distributions of $2 \log \lambda_n$ for three different true concave regression functions: $-x^2$, $\cos(x)$, and $-e^x$, all with $n = 1000$, $M = 5000$ replications

probability our confidence intervals fail to cover the truth is exactly equal to the level of the corresponding hypothesis test. We present in Fig. 2 a plot of our confidence interval procedure on a single instance of simulated data.

6 Conclusions and related problems

There are several problems related to the concave regression problem discussed in this paper. We mention two here: the problem of forming tests/CI's for the value of of a univariate log-concave density, and the problem of forming tests/CI's for the value of a concave/convex regression function with multivariate predictors.

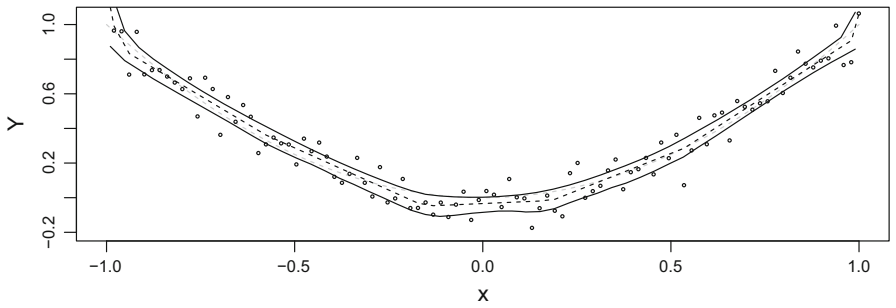


Fig. 2 Pointwise confidence interval (solid lines) at each $x \in (-.99, .99)$ via our log likelihood ratio procedure. Gray dashed line is the true concave regression function $r_0(x) = x^2$; circular points are the simulated data, where $Y_i = r_0(x_i) + \epsilon_i, \epsilon_i \overset{\text{iid}}{\sim} N(0, .1^2), i = 1, \ldots, 100$; black dashed line is the ALSE

A likelihood ratio for the value of a log-concave density on $\mathbb{R}$ In the problem of univariate log-concave density estimation, it is known that the limit distribution of the (univariate) LSE for concave regression [24] and the (univariate) maximum likelihood estimator for log-concave density estimation [3] have the same universal component (they differ in terms of problem-dependent constants). For studying a height-constrained estimator in the log-concave density problem, the class of interest does not immediately form a convex cone, but by translation of the log-densities one can arrive at a convex cone. Consider now $\tilde{X}_1, \ldots, \tilde{X}_n \overset{\text{iid}}{\sim} \tilde{f}_0 = e^{\tilde{\varphi}_0}$ on $\mathbb{R}$ where $\tilde{\varphi}_0 \in \mathcal{C}$. Assume $\tilde{f}_0(x_0) = e^{y_0}$, or $\tilde{\varphi}_0(x_0) = y_0$. The nonparametric log likelihood is $f \mapsto \sum_{i=1}^n \log f(X_i)$. Following [42], we modify this by a Lagrange term (which allows us to optimize over all concave φ without regard to the constraint that $\int e^{\varphi(z)} dz = 1$). Optimizing over $\varphi = \log f$, the unconstrained log-concave MLE [38] is $\widehat{\varphi}_n := \text{argmax}_{\varphi \in \mathcal{C}} \frac{1}{n} \sum_{i=1}^n \varphi(\tilde{X}_i) - \int e^{\varphi(z)} dz$. As in the concave regression problem, we let $\mathcal{C}^0 := \{\varphi \in \mathcal{C} \mid \varphi(x_0) = 0\}$. We can then consider defining $\widehat{\varphi}_n^0 := \text{argmax}_{\varphi \in \mathcal{C}^0} \frac{1}{n} \sum_{i=1}^n \varphi(\tilde{X}_i) + y_0 - \int e^{y_0 + \varphi(z)} dz$. We can combine $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$ to form a likelihood ratio statistic for testing $H_0 : \tilde{\varphi}(x_0) = y_0$ against $H_1 : \tilde{\varphi}(x_0) \neq y_0$. We expect that $\widehat{\varphi}_n^0$ will share features with $\widehat{r}_n^0$ and that the likelihood ratio statistic formed from $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$ will share features with the likelihood ratio statistic (6) discussed in this paper. We would expect that it will in fact have the same universal limit distribution $\mathbb{D}$, independent of nuisance parameters.

A likelihood ratio for the value of a multivariate concave regression function Consider the regression model

$$\tilde{Y}_i = \tilde{r}_0(\tilde{x}_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{85}$$

where ϵ_i are mean 0 and now $\tilde{x}_i \in \mathbb{R}^d$ with $d > 1$. We are again interested in assuming $\tilde{r}_0$ is concave and consider estimating it by least-squares, as in [31,33,41]. We could also consider a constrained estimator as in (12), and form a likelihood ratio statistic for inference about $\tilde{r}_0(x_0)$ at a fixed point $x_0 \in \mathbb{R}^d$. Unfortunately, in the multidimensional case there is no easy analog for Proposition 2 describing the generators of the set of concave functions [8,30]. Thus, it is unlikely that there are

easy analogs of Proposition 4 and, in the constrained case, Theorem 1. In fact, while [33,41] give proofs of consistency of the estimators, pointwise limit distribution results are still unknown. Again, this is in part because of the lack of simple generators for the class of multivariate concave functions, so that there is no analog of Theorem 2 in the constrained estimator case (or analog of the simpler process studied in [23,24] in the unconstrained case). Thus, when $d > 1$, making progress in pointwise asymptotics for the estimators and in studying a likelihood ratio statistic may require new tools or a different approach.

A Appendix: Technical formulas and other results

Here is a statement of an integration by parts formulas for functions of bounded variation. See, e.g., page 102 of [19] for the definition of bounded variation.

Lemma 5 ([19]) *Assume that F and G are of bounded variation on a set $[a, b]$ where $-\infty < a < b < \infty$. If at least one of F and G is continuous, then*

$$\int_{(a,b]} F dG + \int_{(a,b]} G dF = F(b)G(b) - F(a)G(a).$$

Theorem 6 ([23, Theorem 2.1]) *Let $\sigma, a > 0$. Let $X(t) = \sigma W(t) - 4at^3$ where $W(t)$ is standard two-sided Brownian motion starting from 0, and let Y be the integral of X satisfying $Y(0) = 0$. Thus $Y_{a,\sigma}(t) = \sigma \int_0^t W(s)ds - at^4$ for $t \in \mathbb{R}$. Then, with probability 1, there exists a uniquely defined random continuous function $H_{a,\sigma}$ satisfying the following:*

1. *The function $H_{a,\sigma}$ satisfies $H_{a,\sigma}(t) \le Y(t)$ for all $t \in \mathbb{R}$.*
2. *The function $H_{a,\sigma}$ has a concave second derivative, $\widehat{r}_{a,\sigma} := H''_{a,\sigma}$.*
3. *The function $H_{a,\sigma}$ satisfies $\int_{\mathbb{R}} (H_{a,\sigma}(t) - Y_{a,\sigma}(t)) d H^{(3)}_{a,\sigma}(t) = 0$.*

Theorem 7 ([24, Theorem 6.3]) *Suppose that the regression model (9) holds, that $\epsilon_{n,1}, \ldots, \epsilon_{n,n}$ are i.i.d. with $E^{\epsilon^2_{n,1} t} < \infty$ for some $t > 0$, that $r_0 \in \mathcal{C}$, that $r''_0(x_0) < 0$, and that r''_0 is continuous in a neighborhood of x_0. Let* Assumptions 1 *and* 2 *hold. Let $a := |r''_0(x_0)|/24$ and $\sigma^2 := \mathrm{Var}(\epsilon_{n,i})$. Then*

$$n^{2/5}\left(\widehat{r}_n\left(x_0 + tn^{-1/5}\right) - r_0(x_0) - r'_0(x_0)tn^{-1/5}\right) \to_d \widehat{r}_{a,\sigma}(t)$$

in $L^p[-K, K]$ for all $K > 0$.

References

1. Aït-Sahalia, Y., Duarte, J.: Nonparametric option pricing under shape restrictions. J. Econom. **116**(1–2), 9–47 (2003)
2. Allon, G., Beenstock, M., Hackman, S., Passy, U., Shapiro, A.: Nonparametric estimation of concave production technologies by entropic methods. J. Appl. Econom. **22**(4), 795–816 (2007)

3. Balabdaoui, F., Rufibach, K., Wellner, J.A.: Limit distribution theory for maximum likelihood estimation of a log-concave density. Ann. Stat. **37**(3), 1299–1331 (2009)
4. Banerjee, M.: Likelihood ratio tests for monotone functions. Ph.D. thesis, University of Washington (2000)
5. Banerjee, M.: Likelihood based inference for monotone response models. Ann. Stat. **35**(3), 931–956 (2007). https://doi.org/10.1214/009053606000001578
6. Banerjee, M., Wellner, J.A.: Likelihood ratio tests for monotone functions. Ann. Stat. **29**(6), 1699–1731 (2001)
7. Birke, M., Dette, H.: Estimating a convex function in nonparametric regression. Scand. J. Stat. **34**(2), 384–404 (2007)
8. Bronšteĭn, E.M.: Extremal convex functions. Sibirsk. Mat. Ž. **19**(1), 10–18, 236 (1978)
9. Brunk, H.D.: Estimation of isotonic regression. In: Nonparametric techniques in statistical inference (Proceedings of Symposium, Indiana University, Bloomington, IN, 1969), pp. 177–197. Cambridge University Press, London (1970)
10. Cai, T.T., Low, M.G., Xia, Y.: Adaptive confidence intervals for regression functions under shape constraints. Ann. Stat. **41**(2), 722–750 (2013)
11. Dent, W.: A note on least squares fitting of functions constrained to be either nonnegative, nondecreasing or convex. Manag. Sci. **20**, 130–132 (1973/74)
12. Doss, C.R., Wellner, J.A.: Global rates of convergence of the MLEs of log-concave and s-concave densities. Ann. Stat. **44**(3), 954–981 (2016)
13. Doss, C.R., Wellner, J.A.: Inference for the mode of a log-concave density. Submitted to the Annals of Statistics (2018). arXiv:1611.10348
14. Doss, C.R., Wellner, J.A.: Log-concave density estimation with symmetry or modal constraints. Submitted to Annals of Statistics (2018). arXiv:1611.10335v2
15. Dümbgen, L.: Optimal confidence bands for shape-restricted curves. Bernoulli **9**(3), 423–449 (2003)
16. Dümbgen, L., Rufibach, K.: Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. Bernoulli **15**(1), 40–68 (2009)
17. Dykstra, R.L.: An algorithm for restricted least squares regression. J. Am. Stat. Assoc. **78**(384), 837–842 (1983)
18. Efron, B.: Bootstrap methods: another look at the jackknife. Ann. Stat. **7**(1), 1–26 (1979)
19. Folland, G.B.: Real Analysis, Pure and Applied Mathematics, 2nd edn. Wiley, New York (1999)
20. Fraser, D.A.S., Massam, H.: A mixed primal-dual bases algorithm for regression under inequality constraints. application to concave regression. Scand. J. Stat. **16**(1), 65–74 (1989)
21. Groeneboom, P.: Lectures on inverse problems. In: Bernard, P. (ed.) Lectures on Probability Theory, Ecole d'Eté de Probabilités de Saint-Flour XXIV-1994, pp. 67–164. Springer, Berlin (1996)
22. Groeneboom, P., Jongbloed, G.: Nonparametric confidence intervals for monotone functions. Ann. Stat. **43**(5), 2019–2054 (2015)
23. Groeneboom, P., Jongbloed, G., Wellner, J.A.: A canonical process for estimation of convex functions: the "invelope" of integrated brownian motion $+t^4$. Ann. Stat. **29**(6), 1620–1652 (2001)
24. Groeneboom, P., Jongbloed, G., Wellner, J.A.: Estimation of a convex function: characterizations and asymptotic theory. Ann. Stat. **29**(6), 1653–1698 (2001)
25. Hall, P.: Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. Ann. Stat. **20**(2), 675–694 (1992)
26. Hannah, L.A., Dunson, D.B.: Multivariate convex regression with adaptive partitioning. J. Mach. Learn. Res. **14**, 3261–3294 (2013)
27. Hanson, D.L., Pledger, G.: Consistency in concave regression. Ann. Stat. **4**(6), 1038–1050 (1976)
28. Hildreth, C.: Point estimates of ordinates of concave functions. J. Am. Stat. Assoc. **49**(267), 598–619 (1954)
29. Hudson, D.J.: Least-squares fitting of a polynomial constrained to be either non-negative nondecreasing or convex. J. R. Stat. Soc. B **31**(1), 113–118 (1969)
30. Johansen, S.: The extremal convex functions. Math. Scand. **34**, 61–68 (1974)
31. Kuosmanen, T.: Representation theorem for convex nonparametric least squares. Econom. J. **11**(2), 308–325 (2008)
32. Lim, E.: Response surface computation via simulation in the presence of convexity. In: Johansson, B., Jain, S., Montoya-Torres, J., Hugan, J., Yücesan, E. (eds.) 2010 Winter Simulation Conference, pp. 1246–1254 (2010)

33. Lim, E., Glynn, P.W.: Consistency of multidimensional convex regression. Oper. Res. **60**(1), 196–208 (2012)
34. Mammen, E.: Nonparametric regression under qualitative smoothness assumptions. Ann. Stat. **19**(2), 741–759 (1991)
35. Meyer, M.C.: Inference using shape-restricted regression splines. Ann. Appl. Stat. **2**(3), 1013–1033 (2008)
36. Meyer, M.C.: Constrained penalized splines. Can. J. Stat. **40**(1), 190–206 (2012)
37. Monti, M.M., Grant, S., Osherson, D.N.: A note on concave utility functions. Mind Soc. **4**(1), 85–96 (2005)
38. Pal, J.K., Woodroofe, M., Meyer, M.: Estimating a Polya frequency function$_2$. In: Complex Datasets and Inverse Problems, IMS Lecture Notes Monograph Series, vol. 54, pp. 239–249. Institute of Mathematics Statistics, Beachwood (2007)
39. Pflug, G., Wets, R.J.B.: Shape-restricted nonparametric regression with overall noisy measurements. J. Nonparametr. Stat. **25**(2), 323–338 (2013)
40. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
41. Seijo, E., Sen, B.: Nonparametric least squares estimation of a multivariate convex regression function. Ann. Stat. **39**(3), 1633–1657 (2011)
42. Silverman, B.: On the estimation of a probability density function by the maximum penalized likelihood method. Ann. Stat. **10**(3), 795–810 (1982)
43. Sinai, Y.G.: Statistics of shocks in solutions of inviscid burgers equation. Commun. Math. Phys. **148**(3), 601–621 (1992)
44. Topaloglu, H., Powell, W.B.: An algorithm for approximating piecewise linear concave functions from sample gradients. Oper. Res. Lett. **31**(1), 66–76 (2003)
45. Toriello, A., Nemhauser, G., Savelsbergh, M.: Decomposing inventory routing problems with approximate value functions. Naval Res. Logist. **57**(8), 718–727 (2010)
46. Wang, J.C., Meyer, M.C.: Testing the monotonicity or convexity of a function using regression splines. Can. J. Stat. **39**(1), 89–107 (2011)
47. Wasserman, L.: All of Nonparametric Statistics. Springer Texts in Statistics. Springer, New York (2006)
48. Wu, C.F.: Some algorithms for concave and isotonic regression. Stud. Manag. Sci. **19**, 105–116