

Review

Not peer-reviewed version

Review on Isotonic and Convex Regression

[Eunji Lim](#)*

Posted Date: 20 November 2025

doi: 10.20944/preprints202511.1569.v1

Keywords: nonparametric regression; shape-restricted regression; convex regression; isotonic regression



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Review on Isotonic and Convex Regression

Eunji Lim 

Decision Sciences and Marketing, Adelphi University; elim@adelphi.edu

Abstract

Shape-restricted regression provides a framework for estimating an unknown regression function $f_0 : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ from noisy observations $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ when no explicit functional relationship between \mathbf{X} and Y is known, but f_0 is assumed to satisfy structural constraints such as monotonicity or convexity. In this work, we focus on these two shape constraints (monotonicity and convexity), and provide a review on the isotonic regression estimator, which is a least squares estimator under monotonicity, and the convex regression estimator, which is a least squares estimator under convexity. We review existing literature with an emphasis on the following key aspects: quadratic programming formulations of isotonic and convex regression, statistical properties of these estimators, efficient computational algorithms for computing them, their practical applications, and current challenges. Finally, we conclude with a discussion of open challenges and possible directions for future research.

Keywords: nonparametric regression; shape-restricted regression; convex regression; isotonic regression

1. Introduction

One of the central problems in statistics and simulation is to estimate an unknown function $f_0 : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ from noisy observations $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$, where we assume

$$Y_i = f_0(\mathbf{X}_i) + \varepsilon_i,$$

for $i = 1, \dots, n$, Ω is a closed convex subset of \mathbb{R}^d , and the ε_i 's are independent and identically distributed (iid) mean zero random variables with finite variance σ^2 . This type of problem arises in many settings. For instance, x may represent a parameter in a stochastic system, such as the service rate of a single-server queue, and $f_0(x)$ a performance metric, such as the steady-state mean waiting time in that single server queue when the service rate equals x .

In many applications, there is no closed form formula for $f_0(x)$. In such a case, a practical approach is to simulate the system at selected design points $x = \mathbf{X}_1, \dots, \mathbf{X}_n$, obtain simulated outputs Y_1, \dots, Y_n as estimates of $f_0(\mathbf{X}_1), \dots, f_0(\mathbf{X}_n)$, respectively, and use the data $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ to approximate f_0 : This is a typical regression problem. The problem becomes challenging when it is impossible to place the relationship into a simple parametric form. However, when f_0 is known to satisfy certain structural properties such as monotonicity or convexity, shape-restricted regression offers powerful estimators.

Given such prior shape information, a natural estimator of f_0 fits a function f with such a shape characterization to the data set and solves

$$\underset{f \in \mathcal{F}}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n (Y_i - f(\mathbf{X}_i))^2, \quad (1)$$

where \mathcal{F} is the class of functions satisfying the assumed shape constraint. For example, if f_0 is known to be nondecreasing, we can set $\mathcal{F} = \mathcal{F}_m$, where

$$\mathcal{F}_m = \{f : \Omega \rightarrow \mathbb{R} \mid f(\mathbf{v}) \leq f(\mathbf{w}) \text{ if } v_i \leq w_i \text{ for } i = 1, \dots, d, \\ \mathbf{v} = (v_1, \dots, v_d), \mathbf{w} = (w_1, \dots, w_d)\}.$$

On the other hand, if f_0 is known to be convex, then we can set $\mathcal{F} = \mathcal{F}_c$, where

$$\mathcal{F}_c = \{f : \Omega \rightarrow \mathbb{R} \mid f \text{ is convex, i.e.,} \\ f(\alpha \mathbf{v} + (1 - \alpha)\mathbf{w}) \leq \alpha f(\mathbf{v}) + (1 - \alpha)f(\mathbf{w}) \text{ for } \alpha \in [0, 1] \text{ and } \mathbf{v}, \mathbf{w} \in \Omega\}.$$

Throughout this paper, we will focus on these two types of problems, the one with $\mathcal{F} = \mathcal{F}_m$ which leads to isotonic regression and the other with $\mathcal{F} = \mathcal{F}_c$ which leads to convex regression.

One of the first questions that arises is how to compute the solution to (1) numerically when $\mathcal{F} = \mathcal{F}_m$ or $\mathcal{F} = \mathcal{F}_c$. Below, we describe how the solution to (1) can be obtained by solving quadratic programming (QP) problems.

1.1. QP Formulation for Isotonic Regression When $d = 1$

When f_0 is nondecreasing in x , as when price increases with demand, we estimate f_0 by fitting a nondecreasing function f that minimizes squared errors $\sum_{i=1}^n (Y_i - f(\mathbf{X}_i))^2$. Writing $f_i \triangleq f(\mathbf{X}_i)$ and assuming the design is ordered so $\mathbf{X}_1 < \dots < \mathbf{X}_n$, the existence of a nondecreasing fit through (\mathbf{X}_i, f_i) is equivalent to $f_1 \leq \dots \leq f_n$. Thus, the estimator solves the following quadratic programming (QP) problem:

$$\begin{aligned} & \underset{f_1, \dots, f_n \in \mathbb{R}}{\text{minimize}} && \frac{1}{n} \sum_{i=1}^n (Y_i - f_i)^2 \\ & \text{subject to} && f_1 \leq f_2 \leq \dots \leq f_n. \end{aligned} \quad (2)$$

Let $\hat{f}_1^m, \dots, \hat{f}_n^m$ be a solution to (2). Any nondecreasing function interpolating $((\mathbf{X}_i, \hat{f}_i^m) : 1 \leq i \leq n)$ defines the isotonic regression estimator and we will denote it by $\hat{f}_n^m(\cdot)$; see, for example, Brunk [1].

1.2. QP Formulation for Convex Regression When $d = 1$

Suppose instead f_0 is convex. Writing $f_i \triangleq f(\mathbf{X}_i)$ and assuming the design is ordered so $\mathbf{X}_1 < \dots < \mathbf{X}_n$, a convex fit through (\mathbf{X}_i, f_i) exists if and only if

$$\frac{f_2 - f_1}{\mathbf{X}_2 - \mathbf{X}_1} \leq \dots \leq \frac{f_n - f_{n-1}}{\mathbf{X}_n - \mathbf{X}_{n-1}}.$$

Hence, the least-squares convex fit solves

$$\begin{aligned} & \underset{f_1, \dots, f_n \in \mathbb{R}}{\text{minimize}} && \frac{1}{n} \sum_{i=1}^n (Y_i - f_i)^2 \\ & \text{subject to} && \frac{f_2 - f_1}{\mathbf{X}_2 - \mathbf{X}_1} \leq \dots \leq \frac{f_n - f_{n-1}}{\mathbf{X}_n - \mathbf{X}_{n-1}}. \end{aligned} \quad (3)$$

Let $\hat{f}_1^c, \dots, \hat{f}_n^c$ be a solution to (3). Any convex function interpolating $((\mathbf{X}_i, \hat{f}_i^c) : 1 \leq i \leq n)$ is a convex regression estimator and we will denote it by $\hat{f}_n^c(\cdot)$; see, for example, Hildreth [2].

1.3. Key Questions

Isotonic and convex regression are nonparametric methods that avoid any bandwidth or tuning parameters in their basic form yet enjoy strong theoretical guarantees. In particular, as $n \rightarrow \infty$, these estimators converge to f_0 under standard conditions even when the noise variance $\text{Var}(Y_i)$ does not decrease. This is due to the fact that the estimators are required to satisfy the shape constraints,

and hence, just adding more and more data points enforces the estimators to converge to the true underlying function.

Over the past seven decades, extensive work was done to study their statistical properties, computational algorithms, and practical applications. This paper reviews the literature on these topics and highlights unresolved questions. We focus on answering the following questions:

1. How can the estimator be computed numerically from data $(X_1, Y_1), \dots, (X_n, Y_n)$?
2. What are its statistical properties?
3. Are there efficient algorithms that are tailored to the computation of the isotonic or convex regression estimators?
4. What applications benefit most from these approaches?
5. What are its limitations, and how can we overcome them?
6. What are the challenges that still lie in this field?

1.4. Organization of This Paper

The remainder of the paper is organized as follows. Section 2 reviews how the isotonic and convex regression estimators can be computed via QP problems. Section 3 focuses on isotonic regression while Section 4 examines convex regression in greater detail. Section 5 includes some conclusions while Section 6 discusses open challenges and future research directions.

1.5. Notation and Definitions

Throughout this paper, we view vectors as columns and write x^T to denote the transpose of a vector $x \in \mathbb{R}^d$. For $x \in \mathbb{R}^d$, we write its i th component as x_i , so $x = (x_1, \dots, x_d)^T$. We define $\|x\| = (x_1^2 + \dots + x_d^2)^{1/2}$ and $\|x\|_\infty = \max_{1 \leq i \leq d} |x_i|$.

For any convex set $S \subset \mathbb{R}^d$ and a convex function $g : S \rightarrow \mathbb{R}$, a vector $\beta \in \mathbb{R}^d$ is said to be a subgradient of g at $x \in S$ if $g(y) \geq g(x) + \beta^T(y - x)$ for all $y \in S$. We denote a subgradient of f at $x \in S$ by $\text{subgrad } g(x)$. The set of all subgradients of g at x is called the subdifferential of g at x , denoted by $\partial g(x)$. When a function $g : S \rightarrow \mathbb{R}$ is differentiable at $x \in [a, b]^d$, we denote its gradient at x by $\nabla g(x)$. When $g : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable, we denote its gradient at x by $g'(x)$.

For a sequence of random variables $(Z_n : n \geq 1)$ and a sequence of positive real numbers $(\alpha_n : n \geq 1)$, we say $Z_n = \mathcal{O}_p(\alpha_n)$ as $n \rightarrow \infty$ if, for any $\epsilon > 0$, there exist constants c and n_0 such that $\mathbb{P}(|Z_n/\alpha_n| > c) < \epsilon$ for all $n \geq n_0$.

For any sequences of real numbers $(a_n : n \geq 1)$ and $(b_n : n \geq 1)$, $a_n = O(b(n))$ if there exists a positive real number M and a real number n_0 such that $|a_n| \leq Mb(n)$ for all $n \geq n_0$.

For any sequence of random variables $(A_n : n \geq 1)$ and any random variable B , $A_n \Rightarrow B$ means A_n converges in distribution to B , or A_n converges weakly to B as $n \rightarrow \infty$.

2. The Basic QP Formulation

In this section, we describe how the isotonic and convex regression estimators can be found by solving QP problems. Since there are many efficient algorithms for solving QP problems, expressing the isotonic and convex regression estimators as solutions to QP problems offers a basic way to compute them numerically. In $d = 1$, (2) and (3) already provide QP formulations for the isotonic and convex regression estimators, respectively. We now present multivariate analogues.

2.1. Isotonic Regression When $d > 1$

The main challenge in finding the QP formulation for the multivariate isotonic regression estimator is how to express the monotonicity of a multivariate function in linear inequalities as in the constraints of (2). We will define a non-decreasing function g as follows:

$$g(v) \leq g(w) \quad \text{whenever } v \preceq w$$

for $v = (v_1, \dots, v_d)$ and $w = (w_1, \dots, w_d)$ in \mathbb{R}^d , where $v \preceq w$ indicates $v_i \leq w_i$ for $1 \leq i \leq d$.

Using this definition, the nondecreasing function with the minimum sum of squared errors can be found by solving the following QP problem in decision variables $f_1, \dots, f_n \in \mathbb{R}$:

$$\begin{aligned} & \text{minimize} && \frac{1}{n} \sum_{i=1}^n (Y_i - f_i)^2 \\ & \text{subject to} && f_i \leq f_j \quad \text{if } \mathbf{X}_i \preceq \mathbf{X}_j \text{ for } 1 \leq i, j \leq n. \end{aligned} \quad (4)$$

Let $\hat{f}_1^m, \dots, \hat{f}_n^m$ be a solution to (4). Any nondecreasing function interpolating $((\mathbf{X}_i, \hat{f}_i^m) : 1 \leq i \leq n)$ can serve as the isotonic regression estimator and we will denote it by $\hat{f}_n^m(\cdot)$

2.2. Convex Regression When $d > 1$

Again, the key question is how to express the existence of a convex function going through $(\mathbf{X}_1, f_1), \dots, (\mathbf{X}_n, f_n)$ as a set of linear constraints. One should notice that there exists a convex function passing through $(\mathbf{X}_1, f_1), \dots, (\mathbf{X}_n, f_n)$ if and only if there exist $\beta_1, \dots, \beta_n \in \mathbb{R}^d$ satisfying

$$f_i \geq f_j + \beta_j^\top (\mathbf{X}_i - \mathbf{X}_j) \quad \text{for } 1 \leq i, j \leq n;$$

see page 338 of Boyd and Vandenberghe [3]. Here, β_i serves as a subgradient of the convex function f at \mathbf{X}_i . Hence, a convex function with the least squares can be found by solving the following QP problem in decision variables $f_1, \dots, f_n \in \mathbb{R}$ and $\beta_1, \dots, \beta_n \in \mathbb{R}^d$:

$$\begin{aligned} & \text{minimize} && \frac{1}{n} \sum_{i=1}^n (Y_i - f_i)^2 \\ & \text{subject to} && f_i \geq f_j + \beta_j^\top (\mathbf{X}_i - \mathbf{X}_j) \quad \text{for } 1 \leq i, j \leq n. \end{aligned} \quad (5)$$

Let $\hat{f}_1^c, \dots, \hat{f}_n^c, \hat{\beta}_1, \dots, \hat{\beta}_n$ be a solution to (5). Any convex function going through $(\mathbf{X}_1, \hat{f}_1^c), \dots, (\mathbf{X}_n, \hat{f}_n^c)$ can serve as the convex regression estimator in the multidimensional case. However, to remove any ambiguity, we will define the convex regression estimator $\hat{f}_n^c(\cdot)$ as follows:

$$\hat{f}_n^c(\mathbf{x}) = \max_{1 \leq i \leq n} \{ \hat{f}_i^c + \hat{\beta}_i^\top (\mathbf{x} - \mathbf{X}_i) \}$$

for $\mathbf{x} \in \Omega$.

In (2), (3), (4), and (5), we established the isotonic and convex regression estimators as solutions to QP problems. Since there are many efficient numerical algorithms to solve QP problems, these QP formulations enable us to compute the isotonic and convex regression estimators numerically.

3. Isotonic Regression

In this section, we review statistical properties, interesting applications, and some challenges of isotonic regression estimator. We also review some algorithms for computing the isotonic regression estimator numerically.

3.1. Statistical Properties

3.1.1. Univariate Case

The isotonic regression estimator in one and multiple dimensions was first introduced by Brunk [1]. In one dimension, Brunk [4] established its strong consistency. More specifically, Brunk [4] proved that if $\Omega = (a, b) \subset \mathbb{R}$, for any $c > a$ and $d < b$,

$$\max_{c \leq x \leq d} |\hat{f}_n^m(x) - f_0(x)| \rightarrow 0$$

as $n \rightarrow \infty$ with probability one.

Brunk [5] established the asymptotic distribution theory for the isotonic regression estimator when $d = 1$. Brunk [5] proved that when $f'_0(x) \neq 0$,

$$n^{1/3}(\hat{f}_n^m(x) - f_0(x)) \Rightarrow \left[4\sigma^2 f'_0(x)\right]^{1/3} \operatorname{argmin}_{t \in \mathbb{R}} \{W(t) + t^2\}$$

as $n \rightarrow \infty$, where $W(\cdot)$ is the two-sided Brownian motion starting from 0.

The l_2 -risk of the isotonic regression estimator, defined by

$$R(f_0, \hat{f}_n^m) \triangleq \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n (\hat{f}_n^m(\mathbf{X}_i) - f_0(\mathbf{X}_i))^2 \right], \quad (6)$$

is proven to have the following bound

$$R(f_0, \hat{f}_n^m) \leq C\sigma^{4/3} V(f_0)^{2/3} n^{-2/3},$$

where $V(f_0)$ is the total variation of f_0 and $C > 0$ is a universal constant. Meyer and Woodroffe [6] proved (6) under the assumption that the ε_i 's are normally distributed, and Zhang [7] proved (6) under a more general assumption that the ε_i 's are iid with mean zero and variance σ^2 .

3.1.2. Multivariate Case

In multiple dimensions, Lim [8] established the strong consistency. Corollary 3.1 of Lim [8] proved that when $\Omega = (0, 1)^d$, for any $\epsilon > 0$

$$\sup_{x \in [\epsilon, 1-\epsilon]^d} |\hat{f}_n^m(x) - f_0(x)| \rightarrow 0$$

as $n \rightarrow \infty$ with probability one.

Lim [8] also established the rate of convergence and proved that

$$\mathbb{E} \left[(\tilde{f}_n^m(\mathbf{X}) - f_0(\mathbf{X}))^2 \right] = \begin{cases} \mathcal{O}_p(n^{-2/3}), & \text{if } d = 1 \\ \mathcal{O}_p(n^{-1/2}(\log n)^2), & \text{if } d = 2 \\ \mathcal{O}_p(n^{-\frac{1}{2(d-1)}}), & \text{if } d > 2, \end{cases}$$

where the expectation is taken with respect to \mathbf{X} , \mathbf{X} has the same distribution as the \mathbf{X}_i 's, and $\tilde{f}_n^m(\cdot)$, a variant of the isotonic regression estimator, minimizes the sum of squared errors $\sum_{i=1}^n (Y_i - f(\mathbf{X}_i))^2$ over all nondecreasing functions f satisfying $|f(x)| \leq C$ for some pre-set constant C for all $x \in \Omega$.

Bagchi and Dhar [9] derived the pointwise limit distribution of the isotonic regression estimator in multiple dimensions. In particular, Bagchi and Dhar [9] proved

$$n^{-1/(d+2)}(\hat{f}_n^m(x) - f_0(x)) \Rightarrow H$$

as $n \rightarrow \infty$ for some non-degenerate limit distribution H .

Han et al. [10] showed that the l_2 -risk of a multivariate isotonic regression estimator is bounded by a constant multiple of $n^{-1/d}(\log n)^4$ for all $d \geq 1$.

3.2. Computational Algorithms

Formulations (2) and (3) are QPs with linear constraints, and hence, can be solved by off-the-shelf solvers such as CVX, CPLEX, or MOSEK. However, there exist algorithms tailored to the specific setting of the isotonic regression estimator. In $d = 1$, Barlow et al. [11] discussed a graphical procedure to compute the isotonic regression estimator by identifying it as the greatest common minorant for a cumulative sum diagram, and proposed a tabular procedure called the pool-adjacent-violators

(PAV) algorithm. PAV algorithm computes the isotonic regression estimator in $O(n)$ steps. Various modifications of the PAV algorithm have been proposed in the literature; see Brunk [1], Mukerjee [12], Mammen [13], Friedman and Tibshirani [14], Ramsay [15], or Hall and Huang [16], among many others.

However, for regression functions of more than one variable, the problem of computing isotonic regression estimators is substantially more difficult, and good algorithms are not available except for special cases. Most of the literature deals with algorithms for multivariate isotonic regression on a grid (see Gebhardt [17], Dykstra and Robertson [18], Lee [19], or Qian and Eddy [20], among others), but little work has been done for the case of random design.

3.3. Applications

In many applications of economics, operations research, biology, and medical sciences, monotonicity is often assumed to hold or is mathematically proven to hold. For example, in economics, economists often assume the downward slope of the demand as a function of the price. Also, monotonicity applies to indirect utility, expenditure production, profit, and cost functions; see Gallant and Golub [21], Matzkin [22], or Ait-Sahalia and Duarte [23], among many others. In operations research, the steady-state mean waiting time of a customer in a single-server queue is proven to be nonincreasing as a function of the service rate; see Weber [24]. In biology, the weight or height of growing objects over time are known to be nondecreasing over time. In medical sciences, the blood pressure is believed to be a monotone function of the use of tobacco and the body weight (see Moolchan et al. [25]) and the probability of contracting a cancer depends monotonically on certain factors such as smoking frequency, drinking frequency, and weight.

3.4. Challenges

Two common issues in the isotonic regression estimator are (i) non-smoothness (piecewise constant fits) and (ii) boundary overfitting (“spiking”). Figure 1 displays an instance of the true regression function f_0 , the (X_i, Y_i) 's, and the isotonic regression estimator. The isotonic regression estimator in Figure 1 is piecewise constant, and hence, non-smooth. Also, it overfits near the boundary of its domain.

3.4.1. Non-Smoothness of Isotonic Regression

To overcome the non-smoothness of the isotonic regression estimator, several researchers proposed variants of the isotonic regression estimator. Ramsay [15] considered the univariate case and proposed estimating f_0 as a nonnegative linear combination of the I-splines given the knot sequence X_1, \dots, X_n in \mathbb{R} . The I-splines are monotone by definition, so the corresponding estimator is guaranteed to be monotone. This method, however, can be sensitive to the number and placement of knots and the order of the I-splines. Mammen [13] also considered the univariate case and proposed computing the isotonic regression estimator and then making it smooth by applying kernel regression. He and Shi [26] considered the univariate case and proposed fitting normalized B-splines to the data set given the knot sequence X_1, \dots, X_n in \mathbb{R} . This method is also sensitive to the number and placement of the knots. See Mammen and Thomas-Agnan [27] and Meyer [28] for other approaches in the univariate case.

In multiple dimensions, Dette and Scheder [29] proposed a procedure that smooths the data points using a kernel-type method and then applies an isotonization step to each coordinate so that the resulting estimate maintains the monotonicity in each coordinate.

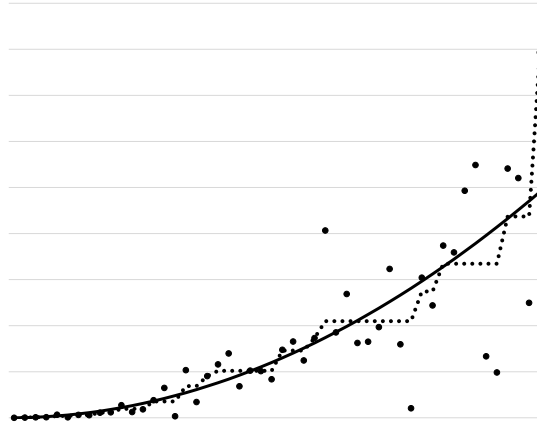


Figure 1. The solid line is f_0 , the circles are the (X_i, Y_i) values, and the dotted line is the isotonic regression estimator.

3.4.2. Overfitting of Isotonic Regression

The isotonic regression estimator is known to be inconsistent at boundaries; this is called the “spiking” problem. See Figure 1 for an illustration of this phenomenon. Proposition 1 of Lim [30] proved that when $\Omega = [a, b]$, $\hat{f}_n^m(a)$ does not converge to $f_0(a)$ in probability as $n \rightarrow \infty$. One popular remedy for this is using a penalized isotonic regression estimator, which minimizes the sum of squared errors plus a penalty term over all monotone functions. A penalized isotonic regression estimator can be expressed as the solution to the following problem:

$$\underset{f \in \mathcal{F}_m}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i))^2 + \lambda_n P_n(f), \quad (7)$$

where $\lambda_n \geq 0$ is a smoothing parameter and $P_n(\cdot)$ is a penalty term. Wu et al. [31] considered the case where $P_n(f)$ equals the range of f divided by n . They proved that when $d = 1$, the solution to (7) (which is often referred to as the “bounded isotonic regression”) evaluated at $x \in \Omega$ converges to $f_0(x)$ in probability as $n \rightarrow \infty$. This proves that the bounded isotonic regression estimator is consistent at the boundary unlike the traditional isotonic regression estimator. Wu et al. [31] also proved similar results for the $d = 2$ case when the covariates are on a grid. Luss and Rosset [32] proposed an algorithm for solving the bounded isotonic regression for the multivariate case. On the other hand, Lim [30] considered the case where $P_n(f) = \max_{x \in \Omega} |f'(x)|$ for the $d = 1$ case and established the strong consistency of the corresponding estimator at $x \in \Omega$, thereby proving that the estimator is consistent at the boundary of Ω . Lim [30] also computed the convergence rate of $n^{-1/3}$ for the proposed estimator.

Even though several approaches are proposed for the $d = 1$ case, limited work is available for the multivariate case. Isotonic regression estimator’s overfitting problem in the $d > 1$ case is a good research topic for future.

4. Convex Regression

In this section, we review statistical properties, interesting applications, and some challenges of convex regression estimator. We also review some algorithms for computing the convex regression estimator numerically.

4.1. Statistical Properties

4.1.1. Univariate Case

The convex regression estimator, along with its QP formulation in (3), was first introduced by Hildreth [2] for the $d = 1$ case. Hanson and Pledger [33] established strong consistency on compact subsets of Ω . They proved that, if $\Omega = (a, b)$, for any $c > a$ and $d < b$,

$$\max_{c \leq x \leq d} |\hat{f}_n^c(x) - f_0(x)| \rightarrow 0$$

as $n \rightarrow \infty$ with probability one.

Results on rates of convergence were obtained by Mammen [34] and Groeneboom et al. [35]. In particular, Mammen [34] showed that at any interior point $x \in \Omega$,

$$\hat{f}_n^c(x) - f_0(x) = \mathcal{O}_p(n^{-2/5})$$

as $n \rightarrow \infty$, provided that the second derivative of f_0 at x does not vanish.

Groeneboom et al. [35] identified the limiting distribution of $n^{2/5}(\hat{f}_n^c(x) - f_0(x))$ at a fixed point $x \in \Omega$. Groeneboom et al. [35] proved that if the errors are iid sub-Gaussian with mean zero and $f_0'(x) \neq 0$,

$$n^{2/5}(\hat{f}_n^c(x) - f_0(x)) \Rightarrow \mathbb{H}''(0),$$

where \mathbb{H} denotes the greatest convex minorant (i.e., envelope) of integrated Brownian motion with quartic drift h^4 , and $\mathbb{H}''(0)$ is its second derivative at 0. The envelope can be viewed as a cubic curve lying above and touching the integrated Brownian motion $+h^4$.

More recently, Chatterjee [36] derived an upper bound on the l_2 -risk of the convex regression estimator. Under the assumption of iid Gaussian errors with mean zero, they showed that

$$R(f_0, \hat{f}_n^c) \triangleq \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n (\hat{f}_n^c(\mathbf{X}_i) - f_0(\mathbf{X}_i))^2 \right] \leq Cn^{-4/5}$$

for all $n \geq 1$, where C is a constant.

4.1.2. Multivariate Case

The multivariate convex regression estimator and its QP formulation in (5) were first introduced by Allon et al. [37]. Strong consistency over any compact subset of the interior of Ω was established by Seijo and Sen [38] and Lim and Glynn [39]. Seijo and Sen [38] proved that for any compact subset A of the interior of Ω ,

$$\max_{x \in A} |\hat{f}_n^c(x) - f_0(x)| \rightarrow 0$$

as $n \rightarrow \infty$ with probability one under some modest assumptions.

Rates of convergence were obtained by Lim [40]. In particular, Lim [40] considered a variant of the convex regression estimator, say $\tilde{f}_n^c(\cdot)$, which minimizes the sum of squared errors $\sum_{i=1}^n (Y_i - f(\mathbf{X}_i))^2$ over the set of convex functions f with subgradients bounded uniformly by a constant. Lim [40] showed that

$$\frac{1}{n} \sum_{i=1}^n (\tilde{f}_n^c(\mathbf{X}_i) - f_0(\mathbf{X}_i))^2 = \begin{cases} \mathcal{O}_p(n^{-4/(4+d)}), & \text{if } d < 4 \\ \mathcal{O}_p((\log n)n^{-1/2}), & \text{if } d = 4 \\ \mathcal{O}_p(n^{-2/d}), & \text{if } d > 4 \end{cases}$$

as $n \rightarrow \infty$.

A bound on the global risk has been established by Han and Wellner [41] as follows:

$$\mathbb{E} \left[\int_{\Omega} (\tilde{f}_n^c(x) - f_0(x))^2 \lambda(x) dx \right] = \begin{cases} \mathcal{O}(n^{-4/(4+d)}), & \text{if } d < 4 \\ \mathcal{O}((\log n)n^{-1/2}), & \text{if } d = 4 \\ \mathcal{O}(n^{-2/d}), & \text{if } d > 4 \end{cases}$$

as $n \rightarrow \infty$, where $\lambda(\cdot)$ is the density function of X_1 and $\hat{f}_n^c(\cdot)$ minimizes the sum of squared errors $\sum_{i=1}^n (Y_i - f(X_i))^2$ over the set of convex functions f bounded uniformly by a constant.

The limiting distribution of $r_n^{-1}(\hat{f}_n^c(x) - f_0(x))$ for an appropriate rate of convergence r_n has not been established so far and this is still an open problem.

4.2. Computational Algorithms

In the univariate case, Dent [42] introduced a QP formulation of the convex regression estimator as in (3). Dykstra [43] developed an iterative algorithm to compute the convex regression estimator numerically. The procedure proposed by Dykstra [43] is guaranteed to converge to the convex regression estimator as the number of iterations approaches infinity. For other algorithms for solving the QP in (3), see Wu [44] and Fraser and Massam [45].

For the multivariate setting, Allon et al. [37] first presented the QP formulation given in (5). However, this formulation requires $O(n^2)$ linear constraints, which limits its applicability to data sets of only a few hundred observations under current computational capabilities. To address this, Mazumder et al. [46] proposed an algorithm based on the augmented Lagrangian method. Their approach was able to solve (5) for $n = 1000$ and $d = 10$ within a few seconds.

To mitigate the computational burden, several authors (Magnani and Boyd [47]; Aguilera et al. [48]; Hannah and Dunson [49]) proposed approximating the convex regression estimator by the maximum of a small set of hyperplanes. Specifically, they solved the optimization problem

$$(\alpha^*, \beta^*, K^*) = \operatorname{argmin}_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}^d, K \in \mathbb{N}} \sum_{i=1}^n \left[Y_i - \max_{k=1, \dots, K} (\alpha_k + \beta_k^\top x_i) \right]^2,$$

where $(\alpha, \beta) \in \mathbb{R} \times \mathbb{R}^d$ defines a hyperplane and \mathbb{N} is the set of positive integers. Their estimator $\hat{f}_n(\cdot)$ is then given by the maximum over these hyperplanes:

$$\hat{f}_n(x) = \max_{k=1, \dots, K^*} (\alpha_k^* + \beta_k^{*\top} x).$$

More recently, to handle larger data sets, Bertsimas and Mundru [50] introduced a penalized convex regression estimator defined as the solution to the following QP problem in decision variables $f_1, \dots, f_n \in \mathbb{R}$ and $\beta_1, \dots, \beta_n \in \mathbb{R}^d$:

$$\begin{aligned} & \text{minimize} && \frac{1}{n} \sum_{i=1}^n (Y_i - f_i)^2 + \frac{\lambda_n}{n} \sum_{i=1}^n \|\beta_i\|^2 \\ & \text{subject to} && f_i \geq f_j + \beta_j^\top (X_i - X_j), \quad 1 \leq i, j \leq n, \end{aligned} \quad (8)$$

where $\lambda_n > 0$ is a smoothing parameter. They proposed a cutting-plane algorithm for solving (8) and reported that it can handle data sets with $n = 10000$ and $d = 10$ within minutes.

Another strategy to alleviate the computational burden of the convex regression estimator is to reformulate the problem as a linear programming (LP) problem rather than a quadratic program. Luo and Lim [51] proposed estimating the convex function that minimizes the sum of absolute deviations (instead of the sum of squared errors). The resulting estimator can be obtained by solving the following LP problem in decision variables $f_1, \dots, f_n \in \mathbb{R}$ and $\beta_1, \dots, \beta_n \in \mathbb{R}^d$:

$$\begin{aligned} & \text{minimize} && \frac{1}{n} \sum_{i=1}^n |Y_i - f_i| \\ & \text{subject to} && f_i \geq f_j + \beta_j^\top (X_i - X_j) \quad \text{for } 1 \leq i, j \leq n. \end{aligned} \quad (9)$$

Luo and Lim [51] further showed that the corresponding estimator converges a.s. to the true function as $n \rightarrow \infty$ over any compact subset of the interior of Ω .

4.3. Applications

Convexity and concavity play an important role across various fields, including economics and operations research. In economics, for instance, it is common to assume diminishing marginal productivity with respect to input resources, which implies that the production function is concave in its inputs (Allon et al. [37]). Also, demand (Varian [52]) and utility (Varian [53]) functions are often assumed to be convex/concave. In operations research, the steady-state waiting time in a single-server queue has been shown to be convex in the service rate; see, for example, Weber [24]. An additional example is provided in Demetriou and Tzitziris [54], where infant mortality rates are modeled as exhibiting increasing returns with respect to the gross domestic product (GDP) per capita, leading to the assumption that the infant mortality function is convex in GDP per capita. For further applications of convex regression, see Johnson and Jiang [55].

4.4. Challenges

The convex regression estimator faces two primary limitations: (i) non-smoothness (piecewise linear fits) and (ii) overfitting near the boundary of its domain (“spiking”). Figure 2 illustrates these issues by displaying the true regression function f_0 , the observed data (X_i, Y_i) , and the convex regression estimator. As shown, the convex regression estimator is piecewise linear and therefore non-smooth. It also exhibits boundary overfitting.

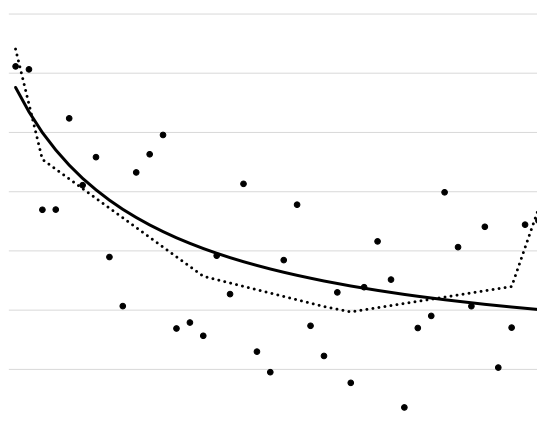


Figure 2. The solid line represents f_0 , the circles denote the observed data (X_i, Y_i) , and the dotted line corresponds to the convex regression estimator.

4.4.1. Non-Smoothness of Convex Regression

In the univariate setting, several approaches have been proposed to address the lack of smoothness. Mammen and Thomas-Agnan [27] developed a two-step procedure in which the data are first smoothed using nonparametric smoothing methods, followed by convex regression applied to the smoothed values. Alternatively, Ait-Sahalia and Duarte [23] suggested first estimating the convex regression estimator and then applying local polynomial smoothing to the estimator. Birke and Dette [56] proposed a different strategy: smoothing the data with a nonparametric method such as the kernel, local polynomial, series, or spline methods, computing its derivative, applying isotonic regression to the derivative, and integrating to obtain a convex estimator. Other methods for the univariate case are discussed in Du et al. [57].

In the multivariate case, Mazumder et al. [46] constructed a smooth approximation to the convex regression estimator that remains uniformly close to the original estimator.

4.4.2. Overfitting of Convex Regression

As with isotonic regression, convex regression tends to overfit near domain boundaries, often producing extremely large subgradients and inconsistent estimates of f_0 near the boundary. For $d = 1$, Ghosal and Sen [58] showed that the convex regression estimator fails to converge in probability to the

true value at boundary points and that its subgradients at the boundary are unbounded in probability as $n \rightarrow \infty$.

A common strategy to mitigate overfitting is to constrain the subgradients. Lim [40] proposed bounding the $\|\cdot\|_\infty$ norm of the subgradients, imposing $\|\beta_i\|_\infty \leq C$ for all i for some fixed $C > 0$ to the constraints of (5). Instead, Mazumder et al. [46] added $\|\beta_i\| \leq C$ to the constraints of (5).

Another widely used approach is penalization, which adds a penalty term to the objective function of (1). Chen et al. [59] and Bertsimas and Mundru [50] studied a penalized multivariate convex regression estimator defined as the solution to the following QP problem in decision variables $f_1, \dots, f_n \in \mathbb{R}, \beta_1, \dots, \beta_n \in \mathbb{R}^d$:

$$\begin{aligned} & \text{minimize} && \frac{1}{n} \sum_{i=1}^n (Y_i - f_i)^2 + \frac{1}{n} \sum_{i=1}^n \|\beta_i\|^2 \\ & \text{subject to} && f_i \geq f_j + \beta_j^\top (\mathbf{X}_i - \mathbf{X}_j) \quad \text{for } 1 \leq i, j \leq n. \end{aligned} \quad (10)$$

Lim [60] generalized this idea and proposed the following three alternative approaches:

$$\begin{aligned} \text{(A)} \quad \hat{f}_n^A &= \operatorname{argmin}_{f \in \mathcal{F}_c} \frac{1}{n} \sum_{i=1}^n (Y_i - f(\mathbf{X}_i))^2 + \lambda_n J(f) \\ \text{(B)} \quad \hat{f}_n^B &= \operatorname{argmin}_{f \in \mathcal{F}_c} \frac{1}{n} \sum_{i=1}^n (Y_i - f(\mathbf{X}_i))^2 && \text{subject to } J(f) \leq u_n \\ \text{(C)} \quad \hat{f}_n^C &= \operatorname{argmin}_{f \in \mathcal{F}_c} J(f) && \text{subject to } \frac{1}{n} \sum_{i=1}^n (Y_i - f(\mathbf{X}_i))^2 \leq s_n \end{aligned}$$

for some smoothing parameter $\lambda_n \geq 0$ and constants $u_n, s_n \geq 0$, where $J(f)$ is a penalty term measuring the overall magnitude of the subgradient of f .

In particular, Lim [60] considered the case where

$$J(f) = \max_{1 \leq i \leq d} \sup_{x \in \Omega} \left| \frac{\partial f}{\partial x_i}(x) \right|$$

and $\partial f / \partial x_i(x)$ represents the partial derivative of f with respect to the i th component of x . Lim [60] established the uniform strong consistency of \hat{f}_n^A and \hat{f}_n^C and their derivatives. Specifically, she showed that

$$\sup_{x \in \Omega} |\hat{f}_n^A(x) - f_0(x)| \rightarrow 0 \quad (11)$$

and

$$\sup_{x \in \Omega} \sup_{\beta \in \partial \hat{f}_n^A(x)} \|\beta - \nabla f_0(x)\|_\infty \rightarrow 0 \quad (12)$$

as $n \rightarrow \infty$ with probability one. Analogous results were shown for \hat{f}_n^C . These results confirm that penalization successfully eliminates boundary overfitting.

Finally, Lim [60] derived convergence rates for \hat{f}_n^A and \hat{f}_n^C . For \hat{f}_n^A , she proved

$$\frac{1}{n} \sum_{i=1}^n (\hat{f}_n^A(\mathbf{X}_i) - f_0(\mathbf{X}_i))^2 = \begin{cases} \mathcal{O}_p(n^{-\frac{4}{4+d}}), & \text{if } d < 4 \\ \mathcal{O}_p(n^{-1/2} \log n), & \text{if } d = 4 \\ \mathcal{O}_p(n^{-2/d}), & \text{if } d > 4 \end{cases}$$

as $n \rightarrow \infty$, and for \hat{f}_n^C , she established

$$\frac{1}{n} \sum_{i=1}^n (\hat{f}_n^C(\mathbf{X}_i) - f_0(\mathbf{X}_i))^2 = \begin{cases} \mathcal{O}_p(n^{-1/2} \sqrt{\log n}), & \text{if } d < 4 \\ \mathcal{O}_p(n^{-1/2} \log n), & \text{if } d = 4 \\ \mathcal{O}_p(n^{-2/d}), & \text{if } d > 4 \end{cases}$$

as $n \rightarrow \infty$.

5. Conclusions

Shape-restricted regression offers a way to incorporate qualitative information, such as monotonicity and convexity, into nonparametric function estimation without relying on bandwidths or other tuning parameters. In this review, we have focused on the isotonic and convex regression, and have highlighted how they arise naturally in areas such as statistics, economics, and operations research. By expressing these estimators as solutions to QP problems, we established a unified optimization framework that facilitates their numerical computation.

On the theoretical side, a substantial body of work has established strong consistency, convergence rates, and risk bounds for isotonic and convex regression. At the same time, the literature has identified important limitations of these estimators, including their inherent non-smoothness and boundary overfitting (“spiking”), which can severely distort behavior near the edges of the design space.

These shortcomings have motivated a variety of methodological extensions. For isotonic regression, smoothing via splines, kernel methods, and other post- or pre-processing steps, as well as penalization schemes such as bounded isotonic regression, have been proposed to improve smoothness and boundary behavior. For convex regression, bounding or penalizing subgradients has emerged as an effective strategy to control overfitting while preserving convexity.

6. Future Directions

This section outlines several open problems that, in our view, remain both important and underexplored. Although some of these topics have appeared in prior work, they call for substantially deeper investigation.

6.1. Problem 1. Scalable Algorithms for Large-Scale Data

Modern applications frequently involve data sets with millions of observations. For example, when a hiring manager receives an application through LinkedIn, the platform is expected to infer the applicant’s seniority level using massive amounts of historical data. Problems of this nature essentially reduce to isotonic regression with millions of data points. At present, no computational algorithm exists that can efficiently compute the isotonic regression estimator at such a scale. In the era of “Big Data,” there is an urgent need for scalable algorithms capable of handling large-scale isotonic and convex regression problems.

6.2. Problem 2. Algorithms for Multivariate Isotonic Regression

In the multivariate setting, efficient algorithms for isotonic regression under random designs remain largely unavailable.

6.3. Problem 3. Pointwise Limit Distribution of Multivariate Convex Regression

Although most major theoretical properties of shape-restricted estimators have been established over the past seven decades, one central question remains open: the pointwise limiting distribution of the multivariate convex regression estimator. Specifically, it is unknown whether

$$r_n^{-1}(\hat{f}_n^c(\mathbf{x}) - f_0(\mathbf{x}))$$

converges weakly to a non-degenerate limit for some appropriate normalization r_n , and if so, what the form of that limit should be. Resolving this would close one of the most significant remaining theoretical gaps in convex regression.

6.4. Problem 4. Incorporating Smoothness into Shape-restricted Regression in Multiple Dimensions

Although several studies have explored ways to address the inherent non-smoothness of shape-restricted regression, existing methods remain preliminary. More effective techniques are needed for incorporating smoothness into multivariate isotonic and convex regression estimators.

6.5. Problem 5. Overfitting in Isotonic and Convex Regression

Only a handful of methods exist to mitigate overfitting in isotonic and convex regression. The most common strategies involve (i) adding a penalty term to the least squares objective function in (1), or (ii) imposing a hard bound on the subgradients (in the convex regression case). Both approaches introduce tuning parameters such as the smoothing constants or gradient bounds that must be chosen in practice, often arbitrarily or in a data-dependent ad-hoc manner. This raises two fundamental questions: How can these parameters be optimally determined? Are there alternative strategies for preventing overfitting without relying on externally imposed parameters?

6.6. Problem 6. Theory for Penalized Isotonic and Convex Regression

There remains substantial scope for advancing the theoretical foundations of penalized isotonic and convex regression. For penalized isotonic regression, most existing results address only the univariate case; extending these insights to higher dimensions remains open. For penalized convex regression, many fundamental questions about its theoretical behavior are still unresolved.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Brunk, H.D. Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **1955**, *26*, 607—616.
2. Hildreth, C. Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **1954**, *49*, 598—619.
3. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
4. Brunk, H.D. On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **1958**, *29*, 437—454.
5. Brunk, H.D. Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference*; Cambridge Univ. Press: New York, 1970; pp. 177–195.
6. Meyer, M.; Woodroffe, M. On the degrees of freedom in shape-restricted regression. *Ann. Statist.* **2000**, *28*, 1083–1104.
7. Zhang, C.H. Risk bounds in isotonic regression. *Ann. Statist.* **2002**, *30*, 528–555.
8. Lim, E. The Limiting Behavior of Isotonic and Convex Regression Estimators When The Model Is Misspecified. *Electron. J. Stat.* **2020**, *14*, 2053–2097.
9. Bagchi, P.; Dhar, S.S. A study on the least squares estimator of multivariate isotonic regression function. *Scand. J. Stat.* **2020**, *47*, 1192–1221.
10. Han, Q.; Wang, T.; Chatterjee, S.; Samworth, R.J. Isotonic Regression in General Dimensions. *Ann. Statist.* **2019**, *47*, 2440–2471.
11. Barlow, R.E.; Bartholomew, R.J.; Bremner, J.M.; Brunk, H.D. *Statistical Inference under Order Restrictions*; Wiley: New York, 1972.
12. Mukerjee, H. Monotone nonparametric regression. *Ann. Statist.* **1988**, *16*, 741–750.
13. Mammen, E. Estimating a smooth monotone regression function. *Ann. Statist.* **1991**, *19*, 724–740.
14. Friedman, J.; Tibshirani, R. The monotone smoothing of scatterplots. *Technometrics* **1984**, *26*, 243–250.
15. Ramsay, J.O. Monotone regression splines in action. *Statist. Sci.* **1988**, *3*, 425–461.
16. Hall, P.; Huang, L.S. Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* **2001**, *29*, 624–647.
17. Gebhardt, F. An algorithm for monotone regression with one or more independent variables. *Biometrika* **1970**, *57*, 263–271.
18. Dykstra, R.L.; Robertson, T. An algorithm for isotonic regression for two or more independent variables. *Ann. Statist.* **1982**, *10*, 708–716.
19. Lee, C.I.C. The min-max algorithm and isotonic regression. *Ann. Statist.* **1983**, *11*, 467–477.

20. Qian, S.; Eddy, W.F. An algorithm for isotonic regression on ordered rectangular grids. *J. Comput. Graph. Statist.* **1996**, *5*, 225–235.
21. Gallant, A.; Golub, G.H. Imposing curvature restrictions on flexible functional forms. *Journal of Economics* **1984**, *26*, 295–321.
22. Matzkin, R.L. Restrictions of economic theory in nonparametric methods. In *Handbook of Econometrics Vol IV*; North-Holland: Amsterdam, 1994; pp. 2523–2558.
23. Ait-Sahalia, Y.; Duarte, J. Nonparametric option pricing under shape restrictions. *J. Econometrics* **2003**, *116*, 9–47.
24. Weber, R.R. A note on waiting times in single server queues. *Oper. Res.* **1983**, *31*, 950–951.
25. Moolchan, E.T.; Hudson, D.L.; Schroeder, J.R.; Sehnert, S.S. Heart rate and blood pressure responses to tobacco smoking among African-American adolescents. *J. Natl. Med. Assoc.* **2004**, *96*, 767.
26. He, X.; Shi, P. Monotone B-spline smoothing. *J. Amer. Statist. Assoc.* **1998**, *93*, 643–650.
27. Mammen, E.; Thomas-Agnan, C. Smoothing splines and shape restrictions. *Scand. J. Stat.* **1999**, *26*, 239–252.
28. Meyer, M.C. Inference using shape-restricted regression splines. *Ann. Statist.* **2008**, *2*, 1013–1033.
29. Dette, H.; Scheder, R. Strictly Monotone and Smooth Nonparametric Regression for Two Or More Variables. *Canad. J. Statist.* **2006**, *34*, 535–561.
30. Lim, E. An estimator for isotonic regression with boundary consistency. *Statist. Probab. Lett.* **2025**, *226*, 110513.
31. Wu, J.; Meyer, M.C.; Opsomer, J.D. Penalized isotonic regression. *J. Statist. Plann. Inference* **2015**, *161*, 12–24.
32. Luss, R.; Rosset, S. Bounded isotonic regression. *Electron. J. Statist.* **2017**, *11*, 4488–4514.
33. Hanson, D.L.; Pledger, G. Consistency in concave regression. *Ann. Statist.* **1976**, *4*, 1038–1050.
34. Mammen, E. Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **1991**, *19*, 741–759.
35. Groeneboom, P.; Jongbloed, G.; Wellner, J.A. Estimation of a convex function: Characterizations and Asymptotic Theory. *Ann. Statist.* **2001**, *29*, 1653–1698.
36. Chatterjee, S. An improved global risk bound in concave regression. *Electron. J. Stat.* **2016**, *10*, 1608–1629.
37. Allon, G.; Beenstock, M.; Hackman, S.; Passy, U.; Shapiro, A. Nonparametric estimation of concave production technologies by entropy. *J. Appl. Econometrics* **2007**, *22*, 795–816.
38. Seijo, E.; Sen, B. Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.* **2011**, *39*, 1633–1657.
39. Lim, E.; Glynn, P.W. Consistency of Multi-dimensional Convex Regression. *Oper. Res.* **2012**, *60*, 196–208.
40. Lim, E. On convergence rates of convex regression in multiple dimensions. *INFORMS J. Comput.* **2014**, *26*, 616–628.
41. Han, Q.; Wellner, J.A. Multivariate convex regression: Global risk bounds and adaptation. <https://arxiv.org/abs/1601.06844>, 2016. arXiv preprint.
42. Dent, W. A note on least squares fitting of function constrained to be either non-negative, nondecreasing, or convex. *Manag. Sci.* **1973**, *20*, 130–132.
43. Dykstra, R.L. An algorithm for restricted least squares regression. *J. Am. Stat. Assoc.* **1983**, *78*, 837–842.
44. Wu, C.F. Some algorithms for concave and isotonic regression. *TIMS Studies in Management Science* **1982**, *19*, 105–116.
45. Fraser, D.A.S.; Massam, H. A mixed primal-dual bases algorithm for regression under inequality constraints. Application to concave regression. *Scand. J. Stat.* **1989**, *16*, 65–74.
46. Mazumder, R.; Iyengar, A.C.G.; Sen, B. A Computational Framework for Multivariate Convex Regression and Its Variants. *J. Am. Stat. Assoc.* **2019**, *114*, 318–331.
47. Magnani, A.; Boyd, S. Convex piecewise-linear fitting. *Optim. Eng.* **2009**, *10*, 1–17.
48. Aguilera, N.; Forzani, L.; Morin, P. On uniform consistent estimators for convex regression. *J. Nonparametr. Stat.* **2011**, *23*, 897–908.
49. Hannah, L.A.; Dunson, D.B. Multivariate Convex Regression with Adaptive Partitioning. *J. Mach. Learn. Res.* **2013**, *14*, 3261–3294.
50. Bertsimas, D.; Mundru, N. Sparse Convex Regression. *INFORMS J. Comput.* **2021**, *33*, 262–279.
51. Luo, Y.; Lim, E. On consistency of absolute deviations estimators of convex functions. *International Journal of Statistics and Probability* **2016**, *5*, 1–18.
52. Varian, H.R. The nonparametric approach to demand analysis. *Econometrica* **1982**, *50*, 945–973.
53. Varian, H.R. The nonparametric approach to production analysis. *Econometrica* **1984**, *52*, 579–597.

54. Demetriou, I.; Tzitziris, P. Infant mortality and economic growth: Modeling by increasing returns and least squares. In Proceedings of the Proceedings of the World Congress on Engineering, 2017, Vol. 2.
55. Johnson, A.L.; Jiang, D.R. Shape Constraints in Economics and Operations Research. *Statist. Sci.* **2018**, *33*, 527–546.
56. Birke, M.; Dette, H. Estimating a convex function in nonparametric regression. *Scand. J. Statist.* **2007**, *34*, 384–404.
57. Du, P.; Parmeter, C.F.; Racine, J.S. Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statist. Sinica* **2013**, *23*, 1347–1371.
58. Ghosal, P.; Sen, B. On univariate convex regression. *Sankhya A* **2017**, *79*, 215–253.
59. Chen, X.; Lin, Q.; Sen, B. On degrees of freedom of projection estimators with applications to multivariate nonparametric regression. *J. Amer. Statist. Assoc.* **2020**, *115*, 173–186.
60. Lim, E. Convex Regression with a Penalty. <http://arxiv.org/abs/2509.19788>, 2025. arXiv preprint.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.