

Review

Isotonic and Convex Regression: A Review of Theory, Algorithms, and Applications

Eunji Lim 

Department of Decision Sciences and Marketing, Adelphi University, Garden City, NY 11530, USA;
elim@adelphi.edu

Abstract

Shape-restricted regression provides a flexible framework for estimating an unknown relationship between input variables and a response when little is known about the functional form, but qualitative structural information is available. In many practical settings, it is natural to assume that the response changes in a systematic way as inputs increase, such as increasing, decreasing, or exhibiting diminishing returns. Isotonic regression incorporates monotonicity constraints, requiring the estimated function to be nondecreasing with respect to its inputs, while convex regression imposes convexity constraints, capturing relationships with increasing or decreasing marginal effects. These shape constraints arise naturally in a wide range of applications, including economics, operations research, and modern data-driven decision systems, where they improve interpretability, stability, and robustness without relying on parametric model assumptions or tuning parameters. This review focuses on isotonic and convex regression as two fundamental examples of shape-restricted regression. We survey their theoretical properties, computational formulations based on optimization, efficient algorithms, and practical applications, and we discuss key challenges such as non-smoothness, boundary overfitting, and scalability. Finally, we outline open problems and directions for future research.

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

MSC: 62G08



Academic Editors: David F. Muñoz,
Federico Trigos and Jingzheng Ren

Received: 17 November 2025

Revised: 18 December 2025

Accepted: 28 December 2025

Published: 30 December 2025

Copyright: © 2025 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In many statistical and data-driven problems, the goal is to understand how a response variable changes as one or more input variables vary. In practice, the exact functional relationship is often unknown and difficult to specify parametrically. Nevertheless, domain knowledge frequently provides qualitative information about the structure of the relationship. For example, demand typically decreases as price increases, waiting times in queueing systems decrease as service capacity increases, and costs often exhibit diminishing returns as resources grow.

Shape-restricted regression is a class of nonparametric methods that incorporate qualitative prior knowledge by constraining the shape of the estimated function, rather than specifying a parametric form. Common shape constraints include monotonicity and convexity, which arise naturally in many applications.

Imposing such constraints fundamentally changes the fitted function. Isotonic regression enforces monotonicity and typically yields a piecewise constant estimator, while

convex regression produces a piecewise linear fit with monotone slopes. These estimators can be viewed as projections of the data onto a space of functions satisfying the prescribed shape constraints.

By preventing local fluctuations and enforcing global structural properties, shape-restricted regression often improves stability, interpretability, and robustness relative to unconstrained nonparametric methods. When qualitative relationships are known a priori, such as increasing risk with exposure or diminishing returns, these methods provide flexible, data-driven models that remain aligned with domain knowledge. Isotonic and convex regression are two fundamental and widely studied examples of this broader framework; Figures 1 and 2 later in the paper provide visual illustrations of how monotonicity and convexity constraints shape the fitted functions in practice.

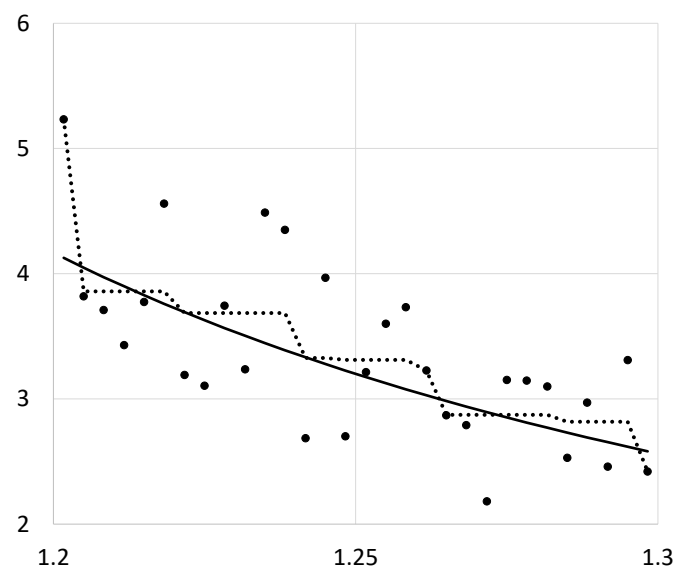


Figure 1. The solid line is f_0 , the circles are the (X_i, Y_i) values, and the dotted line is the isotonic regression estimator.

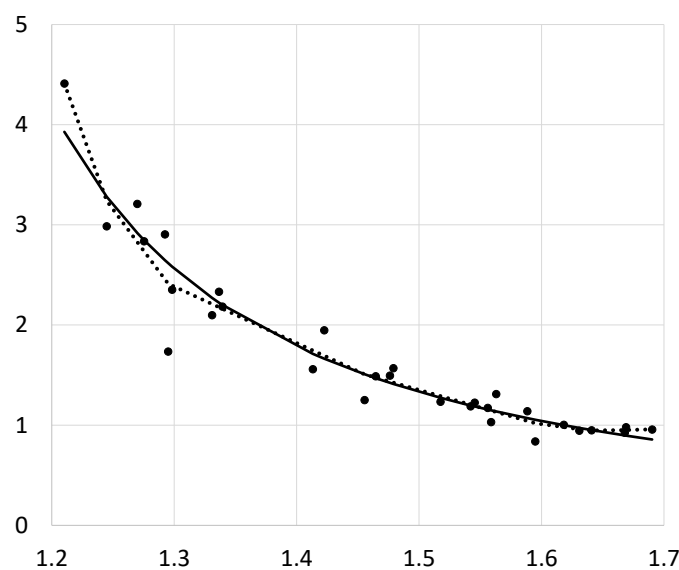


Figure 2. The solid line represents f_0 , the circles denote the observed data (X_i, Y_i) , and the dotted line corresponds to the convex regression estimator.

With this motivation in mind, we consider the problem of estimating an unknown regression function from noisy observations. Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ denote input variables and $Y_1, \dots, Y_n$ the corresponding responses, where each observation is subject to random noise.

Mathematically, we want to estimate an unknown function $f_0 : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ from 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, $\mathbf{x}$ may represent a parameter in a stochastic system, such as the service rate of a single-server queue, and $f_0(\mathbf{x})$ a performance metric, such as the steady-state mean waiting time in that single server queue when the service rate equals $\mathbf{x}$.

In many applications, there is no-closed form formula for $f_0(\mathbf{x})$. In such a case, a practical approach is to simulate the system at selected design points $\mathbf{x} = \mathbf{X}_1, \dots, \mathbf{X}_n$, 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 $d = 1$

We begin with the one-dimensional case to provide a simple and intuitive introduction to isotonic regression. In this setting, the monotonicity constraint reduces to a set of ordered inequalities on the fitted values, which clearly illustrates the basic structure of the estimator and its connection to quadratic programming. This case serves as a conceptual foundation for the more general multivariate formulation discussed later.

When f_0 is nondecreasing in $\mathbf{x}$, 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 $d = 1$

As with isotonic regression, we first consider convex regression in the one-dimensional setting to highlight its essential structure in the simplest possible form. When the covariate is scalar, convexity translates into monotonicity of successive slopes, leading to transparent linear inequality constraints. This formulation provides intuition for how convexity is enforced and motivates the multivariate extension presented in subsequent sections.

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. Scope, Motivation, and Organization of the Review

Over the past several decades, isotonic and convex regression have attracted sustained interest due to their ability to incorporate qualitative structural information into nonparametric estimation; see [3,4] for recent survey papers. Despite extensive theoretical and methodological development, the literature remains fragmented across statistical theory, optimization-based computation, and applied practice.

One central issue concerns computation. Although isotonic and convex regression estimators can be formulated as solutions to constrained least-squares problems, their practical implementation depends critically on efficient algorithms capable of handling large and possibly high-dimensional data sets. Understanding how these estimators can be computed numerically, and what algorithmic trade-offs arise in practice, is therefore a key focus of this review.

A second major theme involves statistical properties. A substantial body of work has established consistency, convergence rates, and limiting distributions under various assumptions, yet these results differ markedly between isotonic and convex regression and between univariate and multivariate settings. Clarifying these similarities and differences

is essential for understanding when and how shape-restricted regression methods can be effectively applied.

Beyond theory and computation, isotonic and convex regression play an important role in applications across economics, operations research, and modern data-driven decision systems. At the same time, practical use of these methods reveals recurring challenges, including non-smoothness of the estimators, boundary overfitting, and scalability limitations. Addressing these issues has motivated a range of extensions and regularization strategies, many of which introduce new theoretical and computational questions.

The objective of this paper is to provide an integrated review of isotonic and convex regression that connects these methodological, theoretical, and practical perspectives. We emphasize a unified optimization-based viewpoint, highlight common challenges, and identify open problems that remain insufficiently explored.

The remainder of the paper is organized as follows: Section 2 reviews quadratic programming formulations of isotonic and convex regression. Section 3 focuses on isotonic regression, covering statistical properties, algorithms, applications, and challenges. Section 4 examines convex regression in greater detail. Section 5 presents concluding remarks, and Section 6 discusses open challenges and directions for future research.

1.4. Notation and Definitions

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

For any convex set $S \subset \mathbb{R}^d$ and a convex function $g : S \rightarrow \mathbb{R}$, a vector $\boldsymbol{\beta} \in \mathbb{R}^d$ is said to be a subgradient of g at $\mathbf{x} \in S$ if $g(\mathbf{y}) \geq g(\mathbf{x}) + \boldsymbol{\beta}^T(\mathbf{y} - \mathbf{x})$ for all $\mathbf{y} \in S$. We denote a subgradient of f at $\mathbf{x} \in S$ by $\text{subgrad } g(\mathbf{x})$. The set of all subgradients of g at $\mathbf{x}$ is called the subdifferential of g at $\mathbf{x}$, denoted by $\partial g(\mathbf{x})$. When a function $g : S \rightarrow \mathbb{R}$ is differentiable at $\mathbf{x} \in [a, b]^d$, we denote its gradient at $\mathbf{x}$ by $\nabla g(\mathbf{x})$. 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:

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

for $\mathbf{v} = (v_1, \dots, v_d)$ and $\mathbf{w} = (w_1, \dots, w_d)$ in $\mathbb{R}^d$, where $\mathbf{v} \preceq \mathbf{w}$ 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).

To clarify the form of the resulting estimator, solving the quadratic program (4) yields fitted values $\hat{f}_1^m, \dots, \hat{f}_n^m$ at the observed design points $\mathbf{X}_1, \dots, \mathbf{X}_n$. The multivariate isotonic regression estimator is then defined as any nondecreasing function that interpolates these fitted values. In practice, the estimator is typically taken to be a piecewise constant function over regions of the covariate space determined by the order constraints.

As a result, the estimator can be readily evaluated at any point $\mathbf{x} \in \Omega$ by assigning it the fitted value corresponding to the maximal design point $\mathbf{X}_i \preceq \mathbf{x}$ under the partial order; i.e.,

$$\hat{f}_n^m(\mathbf{x}) = \begin{cases} \min_{1 \leq i \leq n} \{\hat{f}_i^m\} & \text{if } \mathbf{x} \preceq \mathbf{X}_i \text{ for all } 1 \leq i \leq n \\ \max_{1 \leq i \leq n} \{\hat{f}_i^m \mid \mathbf{X}_i \preceq \mathbf{x}\} & \text{otherwise} \end{cases}$$

This representation highlights both the interpretability and the practical usability of the multivariate isotonic regression estimator once the underlying optimization problem has been solved.

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 $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_n \in \mathbb{R}^d$ satisfying

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

see page 338 of Boyd and Vandenberghe [5]. Here, $\boldsymbol{\beta}_i$ 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 $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_n \in \mathbb{R}^d$:

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

Let $\hat{f}_1^c, \dots, \hat{f}_n^c, \hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_n$ 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:

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

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

In (2)–(5), we established the isotonic and convex regression estimators as solutions to QP problems. Since there are many efficient numerical algorithms to solve QP prob-

lems, 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 [6] established its strong consistency. More specifically, Brunk [6] 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 [7] established the asymptotic distribution theory for the isotonic regression estimator when $d = 1$. Brunk [7] proved that when $f'_0(x) \neq 0$,

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

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 [8] proved (6) under the assumption that the ε_i 's are normally distributed, and Zhang [9] 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 [10] established the strong consistency. Corollary 3.1 of Lim [10] proved that when $\Omega = (0, 1)^d$, for any $\epsilon > 0$

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

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

Lim [10] 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 $\hat{f}_n^m(\cdot)$, 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 $|f(\mathbf{x})| \leq C$ for some pre-set constant C for all $\mathbf{x} \in \Omega$.

Bagchi and Dhar [11] derived the pointwise limit distribution of the isotonic regression estimator in multiple dimensions. In particular, Bagchi and Dhar [11] proved

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

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

Han et al. [12] 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 $d \geq 1$.

These results reveal a pronounced dependence of the convergence behavior on the dimension d . In particular, the rapid deterioration of rates as d increases highlights the intrinsic difficulty of multivariate isotonic regression, even under strong shape constraints. This dimensional effect partially explains the limited practical adoption of multivariate isotonic regression and underscores the need for improved algorithms and regularization strategies.

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 (a package for disciplined convex programming), CPLEX (IBM ILOG CPLEX Optimization Studio), or MOSEK (a commercial optimization solver for large-scale convex problems).

However, there exist algorithms tailored to the specific setting of the isotonic regression estimator. In $d = 1$, Barlow et al. [13] 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. The 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 [14], Mammen [15], Friedman and Tibshirani [16], Ramsay [17], or Hall and Huang [18], 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 [19], Dykstra and Robertson [20], Lee [21], or Qian and Eddy [22], among others), but little work has been done for the case of random design.

Table 1 summarizes the main computational approaches discussed above.

Table 1. Comparison of computational algorithms for isotonic regression.

Method	Dimension	Design	Complexity
PAV algorithm [13]	$d = 1$	Ordered	$O(n)$
Graphical/cumulative sum	$d = 1$	Ordered	$O(n)$
General QP solvers	Any d	Any	Polynomial
Grid-based methods	$d > 1$	Grid	–

From a methodological standpoint, existing algorithms for isotonic regression reveal a clear tension between computational efficiency and general applicability. While the pool-adjacent-violators algorithm is exact and optimal in the univariate setting, its

reliance on an ordered design severely limits its extension to higher dimensions. Multivariate formulations, on the other hand, are conceptually straightforward but quickly become computationally prohibitive due to the explosion of monotonicity constraints. As a result, many practical implementations either restrict attention to structured designs or sacrifice exactness for scalability, highlighting a gap between theoretical formulations and real-world deployment.

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 [23], Matzkin [24], or Aït-Sahalia and Duarte [25], 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 [26]. 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. [27]) and the probability of contracting a cancer depends monotonically on certain factors such as smoking frequency, drinking frequency, and weight.

Illustration of Isotonic Regression

To illustrate how isotonic regression operates in practice, we consider a simple univariate example. We consider the case where f_0 represents the expected average waiting time of the first 5000 customers in an M/M/1 queue with a first in/first out discipline, unit arrival rate, and a service rate $x \in [1.2, 1.3]$, initialized as empty and idle. The service rate is set as $X_i = 1.2 + 0.1(i - 1)/n + 0.1/(2n)$ for $1 \leq i \leq n$. For each X_i , the corresponding Y_i is the average waiting time of the first 5000 customers in this M/M/1 queue with a unit arrival rate and a service rate of X_i . Once the pairs (X_i, Y_i) for $1 \leq i \leq n$ are generated, we computed the isotonic regression estimator by solving (2) using CVX.

Figure 1 displays the observed data along with the isotonic regression estimator. Its mean square error of the isotonic regression estimator, $\frac{1}{n} \sum_{i=1}^n (\hat{f}_n^m(X_i) - f_0(X_i))^2$, is 0.0596 while the mean square error of the Y_i 's, $\frac{1}{n} \sum_{i=1}^n (Y_i - f_0(X_i))^2$, is 0.2743. This shows improved performance of the isotonic regression estimator compared to the Y_i 's.

The estimator enforces monotonicity while fitting the data in a least-squares sense, resulting in a piecewise constant function. This example highlights two key features of isotonic regression: (i) the automatic enforcement of shape constraints without tuning parameters, and (ii) the tendency toward non-smoothness and boundary effects, which motivate the extensions discussed later in this paper.

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 [17] 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 [15] also considered the univariate case and proposed computing the isotonic regression estimator and then making it smooth by applying kernel regression. He and Shi [28] 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 [29] and Meyer [30] for other approaches in the univariate case.

In multiple dimensions, Dette and Scheder [31] 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.

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 [32] 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. [33] 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. [33] also proved similar results for the $d = 2$ case when the covariates are on a grid. Luss and Rosset [34] proposed an algorithm for solving the bounded isotonic regression for the multivariate case. On the other hand, Lim [32] 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 [32] 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 [35] 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 [36] and Groeneboom et al. [37]. In particular, Mammen [36] 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. [37] 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. [37] 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 [38] 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. [39]. Strong consistency over any compact subset of the interior of Ω was established by Seijo and Sen [40] and Lim and Glynn [41]. Seijo and Sen [40] 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 [42]. In particular, Lim [42] considered a variant of the convex regression estimator, say $\tilde{f}_n^c(\cdot)$, which minimizes the sum of squared errors $\sum_{i=1}^n (Y_i - f(X_i))^2$ over the set of convex functions f with subgradients bounded uniformly by a constant. Lim [42] 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 [43] 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 $\mathbf{X}_1$ and $\tilde{f}_n^c(\cdot)$ 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 [44] introduced a QP formulation of the convex regression estimator as in (3). Dykstra [45] developed an iterative algorithm to compute the convex regression estimator numerically. The procedure proposed by Dykstra [45] 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 [46] and Fraser and Massam [47].

For the multivariate setting, Allon et al. [39] 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. [48] 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 [49]; Aguilera et al. [50]; Hannah and Dunson [51]) 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 [52] 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 $\beta_1, \dots, \beta_n \in \mathbb{R}^d$:

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

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 $n = 10,000$ 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 [53] 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 $\beta_1, \dots, \beta_n \in \mathbb{R}^d$:

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

Luo and Lim [53] 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 .

Table 2 summarizes the main computational approaches discussed above.

Table 2. Comparison of computational algorithms for convex regression.

Method	Dimension	Remarks
QP formulation [2]	$d = 1$	Exact solution
Iterative projection [45]	$d = 1$	Guaranteed convergence
Standard multivariate QP [39]	$d > 1$	Severe computational burden
Augmented Lagrangian [48]	$d > 1$	Scales to $n \approx 10^3$
Cutting-plane methods [52]	$d > 1$	Handles $n \approx 10^4$
Hyperplane approximation	$d > 1$	Approximate estimator

Taken together, these algorithmic developments highlight a fundamental trade-off between statistical exactness and computational scalability. While the exact QP formulation provides a principled estimator with strong theoretical guarantees, its quadratic growth in constraints severely limits its applicability in large-scale settings. Approximate and penalized methods alleviate this burden, but at the cost of introducing additional tuning parameters or approximation error. Understanding this trade-off remains central to the practical deployment of convex regression.

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. [39]). Also, demand (Varian [54]) and utility (Varian [55]) 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 [26]. An additional example is provided in Demetriou and Tzitziris [56], 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 [4].

Illustration of Convex Regression

To illustrate how convex regression operates in practice, we consider a simple univariate example. We consider the case where $f_0(x)$ is the steady-state mean waiting time of a customer in a single server queue with the first-in/first-out discipline, where the service times are iid exponential random variables with a mean of $1/x$ for $x \in (1.2, 1.7)$, and the interarrival times are iid exponential random variables with a mean of 1.

The X_i 's are drawn uniformly from $(1.2, 1.7)$. For each X_i , Y_i is generated by averaging the waiting times of the first 5000 customers in the single server queue, initialized empty and idle, with the service rate of X_i . Once the (X_i, Y_i) 's are obtained, we computed the convex regression estimator by solving (3) using CVX.

Figure 2 displays the observed data along with the convex regression estimator. Its mean square error of the convex regression estimator, $\frac{1}{n} \sum_{i=1}^n (\hat{f}_n^c(X_i) - f_0(X_i))^2$, is 0.0132

while the mean square error of the Y_i 's, $\frac{1}{n} \sum_{i=1}^n (Y_i - f_0(X_i))^2$, is 0.0554. This shows improved performance of the convex regression estimator compared to the Y_i 's.

The estimator enforces convexity while fitting the data in a least-squares sense, resulting in a piecewise linear function. This example highlights two key features of convex regression: (i) the automatic enforcement of shape constraints without tuning parameters, and (ii) the tendency toward non-smoothness and boundary effects, which motivate the extensions discussed later in the next subsection.

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 in Figure 2, the convex regression estimator is piecewise linear and therefore non-smooth. It also exhibits boundary overfitting.

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 [29] 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 [25] suggested first estimating the convex regression estimator and then applying local polynomial smoothing to the estimator. Birke and Dette [57] 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. [58].

In the multivariate case, Mazumder et al. [48] 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 [59] 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 [42] 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. [48] 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. [60] and Bertsimas and Mundru [52] studied a penalized multivariate convex regression estimator defined as the solution to the following QP problem in decision variables $f_1, \dots, f_n \in \mathbb{R}, \beta_1, \dots, \beta_n \in \mathbb{R}^d$:

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

Lim [61] generalized this idea and proposed the following three alternative approaches:

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

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 [61] considered the case where

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

and $\partial f / \partial x_i(\mathbf{x})$ represents the partial derivative of f with respect to the i th component of $\mathbf{x}$. Lim [61] established the uniform strong consistency of $\hat{f}_n^A$ and $\hat{f}_n^C$ and their derivatives. Specifically, she showed that

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

and

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

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 [61] 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. Connections to Contemporary Machine Learning

Beyond their traditional applications in economics and operations research, isotonic and convex regression have found increasing use in contemporary machine learning, where structural constraints are often imposed to improve interpretability, stability, and compliance with domain knowledge.

A prominent example is probability calibration in classification and ranking systems. In large-scale recommender systems and online advertising platforms, raw model outputs (e.g., click-through rate (CTR) scores) are often poorly calibrated as probabilities. Isotonic regression is widely used as a post-hoc calibration method in CTR prediction and rec-

ommendation systems, enforcing monotonicity between model scores and observed click frequencies while preserving predictive ordering.

Shape constraints also arise naturally in learning-to-rank and monotone prediction problems, where certain features are known *a priori* to have a monotone effect on the response. Such constraints are increasingly emphasized in interpretable and trustworthy machine learning, particularly in high-stakes applications such as credit scoring, hiring, and healthcare, where monotonicity is often required for regulatory or ethical reasons.

From an optimization perspective, convex regression connects directly to learning convex functions and convex potentials, which appear in areas such as structured prediction, energy-based modeling, and optimal transport. In these settings, convex regression provides a principled way to estimate convex functions from data while preserving global shape properties. Related ideas also appear in modern machine learning through neural network architectures designed to learn convex functions. A notable example is Input Convex Neural Networks (ICNNs) [62], which impose architectural constraints to guarantee convexity with respect to the inputs. While ICNNs are not direct implementations of the nonparametric convex regression methods reviewed here, they are conceptually motivated by the same need to learn convex cost or loss functions from data. This parallel highlights the continued relevance of convexity constraints across both classical statistical estimation and contemporary machine learning.

Finally, the computational challenges discussed in this paper—scalability, overfitting, and the need for efficient algorithms—are closely aligned with current concerns in large-scale machine learning. Recent developments such as augmented Lagrangian methods, cutting-plane algorithms, and penalized formulations mirror optimization techniques widely used in modern ML, highlighting shape-restricted regression as a natural and increasingly relevant component of the contemporary machine learning toolkit.

6. Future Directions

This section outlines several open problems that, in our view, remain both important and insufficiently explored. The goal of this section is not only to list unresolved questions, but also to provide readers with a research roadmap highlighting where significant theoretical, computational, and applied advances can be made. Each problem below is motivated by practical limitations of existing methods and points to concrete opportunities for future research.

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

Modern applications increasingly involve data sets with millions or even billions of observations. For example, in large-scale hiring platforms or online advertising systems, isotonic regression is routinely used for score calibration, yet existing algorithms are unable to scale to such data sizes.

Despite decades of progress in computational optimization, no algorithm is currently capable of computing isotonic or convex regression estimators efficiently at this scale under general designs. Developing scalable algorithms—possibly through distributed optimization, streaming methods, or approximation schemes—remains a critical challenge. This problem is particularly relevant to researchers working at the intersection of optimization, machine learning, and large-scale data analysis.

6.2. Problem 2—Algorithms for Multivariate Isotonic Regression

While univariate isotonic regression enjoys highly efficient algorithms such as the pool-adjacent-violators method, the multivariate case remains far less understood. In

particular, for isotonic regression under random designs in higher dimensions, efficient and practically implementable algorithms are largely unavailable.

Most existing work focuses on grid-based designs or highly structured settings, leaving a substantial gap between theory and practice. Closing this gap would significantly expand the applicability of isotonic regression in modern data-driven problems involving multiple covariates.

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

Although many global theoretical properties of convex regression estimators have been established, a central question remains unresolved: the pointwise limiting distribution of the multivariate convex regression estimator.

Specifically, it is unknown whether

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

converges weakly to a non-degenerate limit for an appropriate normalization r_n , and if so, what the form of this limit might be. Resolving this problem would place multivariate convex regression on a theoretical footing comparable to its univariate counterpart and would substantially deepen our understanding of the local behavior of shape-restricted estimators.

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

A well-known limitation of isotonic and convex regression is their inherent lack of smoothness, resulting in piecewise constant or piecewise linear estimators. While several smoothing and regularization techniques have been proposed, most are either heuristic or limited to low-dimensional settings.

Developing principled methods that simultaneously enforce shape constraints and smoothness in multiple dimensions remains an open challenge. Progress in this direction would be particularly valuable for applications requiring stable derivatives, such as sensitivity analysis, optimization, and decision-making systems.

6.5. Problem 5—Overfitting in Isotonic and Convex Regression

Both isotonic and convex regression are known to suffer from overfitting near the boundaries of the design space. Existing remedies typically rely on penalization or hard constraints on the estimator or its subgradients, which introduce tuning parameters that are often selected in an ad-hoc or problem-specific manner.

This raises several fundamental questions: How should such tuning parameters be chosen in a principled way? Are there alternative approaches to controlling overfitting that avoid explicit penalization altogether? Addressing these questions would improve both the theoretical robustness and practical usability of shape-restricted regression methods.

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

While penalization has proven effective in mitigating non-smoothness and boundary overfitting, the theoretical understanding of penalized shape-restricted estimators remains incomplete. For isotonic regression, most existing results are confined to the univariate setting, and extensions to higher dimensions are largely unexplored.

For convex regression, many foundational questions remain open, including optimal convergence rates, adaptive behavior, and the interaction between penalization and dimensionality. Advancing the theory of penalized isotonic and convex regression would provide essential guidance for both methodological development and practical implementation.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

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. [\[CrossRef\]](#)
2. Hildreth, C. Point estimates of ordinates of concave functions. *J. Am. Stat. Assoc.* **1954**, *49*, 598–619. [\[CrossRef\]](#)
3. Guntuboyina, A.; Sen, B. Nonparametric Shape-Restricted Regression. *Stat. Sci.* **2018**, *33*, 568–594. [\[CrossRef\]](#)
4. Johnson, A.L.; Jiang, D.R. Shape Constraints in Economics and Operations Research. *Stat. Sci.* **2018**, *33*, 527–546. [\[CrossRef\]](#)
5. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
6. Brunk, H.D. On the estimation of parameters restricted by inequalities. *Ann. Math. Stat.* **1958**, *29*, 437–454. [\[CrossRef\]](#)
7. Brunk, H.D. Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference*; Cambridge University Press: New York, NY, USA, 1970; pp. 177–195.
8. Meyer, M.; Woodroffe, M. On the degrees of freedom in shape-restricted regression. *Ann. Stat.* **2000**, *28*, 1083–1104. [\[CrossRef\]](#)
9. Zhang, C.H. Risk bounds in isotonic regression. *Ann. Stat.* **2002**, *30*, 528–555. [\[CrossRef\]](#)
10. Lim, E. The Limiting Behavior of Isotonic and Convex Regression Estimators When The Model Is Misspecified. *Electron. J. Stat.* **2020**, *14*, 2053–2097. [\[CrossRef\]](#)
11. Bagchi, P.; Dhar, S.S. A study on the least squares estimator of multivariate isotonic regression function. *Scand. J. Stat.* **2020**, *47*, 1192–1221. [\[CrossRef\]](#)
12. Han, Q.; Wang, T.; Chatterjee, S.; Samworth, R.J. Isotonic Regression in General Dimensions. *Ann. Stat.* **2019**, *47*, 2440–2471. [\[CrossRef\]](#)
13. Barlow, R.E.; Bartholomew, R.J.; Bremner, J.M.; Brunk, H.D. *Statistical Inference Under Order Restrictions*; Wiley: New York, NY, USA, 1972.
14. Mukerjee, H. Monotone nonparametric regression. *Ann. Stat.* **1988**, *16*, 741–750. [\[CrossRef\]](#)
15. Mammen, E. Estimating a smooth monotone regression function. *Ann. Stat.* **1991**, *19*, 724–740. [\[CrossRef\]](#)
16. Friedman, J.; Tibshirani, R. The monotone smoothing of scatterplots. *Technometrics* **1984**, *26*, 243–250. [\[CrossRef\]](#)
17. Ramsay, J.O. Monotone regression splines in action. *Stat. Sci.* **1988**, *3*, 425–461. [\[CrossRef\]](#)
18. Hall, P.; Huang, L.S. Nonparametric kernel regression subject to monotonicity constraints. *Ann. Stat.* **2001**, *29*, 624–647. [\[CrossRef\]](#)
19. Gebhardt, F. An algorithm for monotone regression with one or more independent variables. *Biometrika* **1970**, *57*, 263–271. [\[CrossRef\]](#)
20. Dykstra, R.L.; Robertson, T. An algorithm for isotonic regression for two or more independent variables. *Ann. Stat.* **1982**, *10*, 708–716. [\[CrossRef\]](#)
21. Lee, C.I.C. The min-max algorithm and isotonic regression. *Ann. Stat.* **1983**, *11*, 467–477. [\[CrossRef\]](#)
22. Qian, S.; Eddy, W.F. An algorithm for isotonic regression on ordered rectangular grids. *J. Comput. Graph. Stat.* **1996**, *5*, 225–235. [\[CrossRef\]](#)
23. Gallant, A.; Golub, G.H. Imposing curvature restrictions on flexible functional forms. *J. Econ.* **1984**, *26*, 295–321. [\[CrossRef\]](#)
24. Matzkin, R.L. Restrictions of economic theory in nonparametric methods. In *Handbook of Econometrics*; North-Holland: Amsterdam, The Netherlands, 1994; Volume IV, pp. 2523–2558.
25. Ait-Sahalia, Y.; Duarte, J. Nonparametric option pricing under shape restrictions. *J. Econom.* **2003**, *116*, 9–47. [\[CrossRef\]](#)
26. Weber, R.R. A note on waiting times in single server queues. *Oper. Res.* **1983**, *31*, 950–951. [\[CrossRef\]](#)
27. 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.
28. He, X.; Shi, P. Monotone B-spline smoothing. *J. Am. Stat. Assoc.* **1998**, *93*, 643–650. [\[CrossRef\]](#)
29. Mammen, E.; Thomas-Agnan, C. Smoothing splines and shape restrictions. *Scand. J. Stat.* **1999**, *26*, 239–252. [\[CrossRef\]](#)
30. Meyer, M.C. Inference using shape-restricted regression splines. *Ann. Stat.* **2008**, *2*, 1013–1033. [\[CrossRef\]](#)
31. Dette, H.; Scheder, R. Strictly Monotone and Smooth Nonparametric Regression for Two or More Variables. *Can. J. Stat.* **2006**, *34*, 535–561. [\[CrossRef\]](#)
32. Lim, E. An estimator for isotonic regression with boundary consistency. *Stat. Probab. Lett.* **2025**, *226*, 110513. [\[CrossRef\]](#)
33. Wu, J.; Meyer, M.C.; Opsomer, J.D. Penalized isotonic regression. *J. Stat. Plan. Inference* **2015**, *161*, 12–24. [\[CrossRef\]](#)
34. Luss, R.; Rosset, S. Bounded isotonic regression. *Electron. J. Stat.* **2017**, *11*, 4488–4514. [\[CrossRef\]](#)
35. Hanson, D.L.; Pledger, G. Consistency in concave regression. *Ann. Stat.* **1976**, *4*, 1038–1050. [\[CrossRef\]](#)
36. Mammen, E. Nonparametric regression under qualitative smoothness assumptions. *Ann. Stat.* **1991**, *19*, 741–759. [\[CrossRef\]](#)
37. Groeneboom, P.; Jongbloed, G.; Wellner, J.A. Estimation of a convex function: Characterizations and Asymptotic Theory. *Ann. Stat.* **2001**, *29*, 1653–1698. [\[CrossRef\]](#)

38. Chatterjee, S. An improved global risk bound in concave regression. *Electron. J. Stat.* **2016**, *10*, 1608–1629. [\[CrossRef\]](#)
39. Allon, G.; Beenstock, M.; Hackman, S.; Passy, U.; Shapiro, A. Nonparametric estimation of concave production technologies by entropy. *J. Appl. Econom.* **2007**, *22*, 795–816. [\[CrossRef\]](#)
40. Seijo, E.; Sen, B. Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Stat.* **2011**, *39*, 1633–1657. [\[CrossRef\]](#)
41. Lim, E.; Glynn, P.W. Consistency of Multi-dimensional Convex Regression. *Oper. Res.* **2012**, *60*, 196–208. [\[CrossRef\]](#)
42. Lim, E. On convergence rates of convex regression in multiple dimensions. *INFORMS J. Comput.* **2014**, *26*, 616–628. [\[CrossRef\]](#)
43. Han, Q.; Wellner, J.A. Multivariate convex regression: Global risk bounds and adaptation. *arXiv* **2016**. [\[CrossRef\]](#)
44. 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. [\[CrossRef\]](#)
45. Dykstra, R.L. An algorithm for restricted least squares regression. *J. Am. Stat. Assoc.* **1983**, *78*, 837–842. [\[CrossRef\]](#)
46. Wu, C.F. Some algorithms for concave and isotonic regression. *TIMS Stud. Manag. Sci.* **1982**, *19*, 105–116.
47. 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.
48. 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. [\[CrossRef\]](#)
49. Magnani, A.; Boyd, S. Convex piecewise-linear fitting. *Optim. Eng.* **2009**, *10*, 1–17. [\[CrossRef\]](#)
50. Aguilera, N.; Forzani, L.; Morin, P. On uniform consistent estimators for convex regression. *J. Nonparametr. Stat.* **2011**, *23*, 897–908. [\[CrossRef\]](#)
51. Hannah, L.A.; Dunson, D.B. Multivariate Convex Regression with Adaptive Partitioning. *J. Mach. Learn. Res.* **2013**, *14*, 3261–3294.
52. Bertsimas, D.; Mundru, N. Sparse Convex Regression. *INFORMS J. Comput.* **2021**, *33*, 262–279. [\[CrossRef\]](#)
53. Luo, Y.; Lim, E. On consistency of absolute deviations estimators of convex functions. *Int. J. Stat. Probab.* **2016**, *5*, 1–18. [\[CrossRef\]](#)
54. Varian, H.R. The nonparametric approach to demand analysis. *Econometrica* **1982**, *50*, 945–973. [\[CrossRef\]](#)
55. Varian, H.R. The nonparametric approach to production analysis. *Econometrica* **1984**, *52*, 579–597. [\[CrossRef\]](#)
56. Demetriou, I.; Tzitziris, P. Infant mortality and economic growth: Modeling by increasing returns and least squares. In Proceedings of the World Congress on Engineering, London, UK, 5–7 July 2017; Volume 2.
57. Birke, M.; Dette, H. Estimating a convex function in nonparametric regression. *Scand. J. Stat.* **2007**, *34*, 384–404. [\[CrossRef\]](#)
58. Du, P.; Parmeter, C.F.; Racine, J.S. Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Stat. Sin.* **2013**, *23*, 1347–1371.
59. Ghosal, P.; Sen, B. On univariate convex regression. *Sankhya A* **2017**, *79*, 215–253. [\[CrossRef\]](#)
60. Chen, X.; Lin, Q.; Sen, B. On degrees of freedom of projection estimators with applications to multivariate nonparametric regression. *J. Am. Stat. Assoc.* **2020**, *115*, 173–186. [\[CrossRef\]](#)
61. Lim, E. Convex Regression with a Penalty. *arXiv* **2025**, arXiv:2509.19788. [\[CrossRef\]](#)
62. Amos, B.; Kolter, J.Z. Input Convex Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR Inc.: Cambridge, MA, USA, 2017.

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.