Shape-Enforcing Operators for Generic Point and Interval Estimators of Functions

Xi Chen   XC13@STERN.NYU.EDU
Stern School of Business
New York University, New York, NY 10012, USA

Victor Chernozhukov   VCHERN@MIT.EDU
Department of Economics + Center for Statistics and Data Science
Massachusetts Institute of Technology, Cambridge, MA 02142, USA

Iván Fernández-Val   IVANF@BU.EDU
Department of Economics
Boston University, Boston, MA 02215-1403, USA

Scott Kostyshak   SKOSTYSHAK@UFL.EDU
Department of Economics
University of Florida, Gainesville, FL 32611, USA

Ye Luo   KURTLUO@HKU.HK
HKU Business School
The University of Hong Kong, Pok Fu Lam, Hong Kong

Editor: Garvesh Raskutti

Abstract

A common problem in econometrics, statistics, and machine learning is to estimate and make inference on functions that satisfy shape restrictions. For example, distribution functions are nondecreasing and range between zero and one, height growth charts are nondecreasing in age, and production functions are nondecreasing and quasi-concave in input quantities. We propose a method to enforce these restrictions *ex post* on generic unconstrained point and interval estimates of the target function by applying functional operators. The interval estimates could be either frequentist confidence bands or Bayesian credible regions. If an operator has reshaping, invariance, order-preserving, and distance-reducing properties, the shape-enforced point estimates are closer to the target function than the original point estimates and the shape-enforced interval estimates have greater coverage and shorter length than the original interval estimates. We show that these properties hold for six different operators that cover commonly used shape restrictions in practice: range, convexity, monotonicity, monotone convexity, quasi-convexity, and monotone quasi-convexity, with the latter two restrictions being of paramount importance. The main attractive property of the post-processing approach is that it works in conjunction with *any generic* initial point or interval estimate, obtained using any of parametric, semiparametric or nonparametric learning methods, including recent methods that are able to exploit either smoothness, sparsity, or other forms of structured parsimony of target functions. The post-processed point and interval estimates automatically inherit and provably improve these properties in finite samples, while also enforcing qualitative shape restrictions brought by scientific reasoning. We illustrate the results with two empirical applications

to the estimation of a height growth chart for infants in India and a production function for chemical firms in China.

Keywords: Shape Operator, Range, Monotonicity, Convexity, Quasi-Convexity, Rearrangement, Legendre-Fenchel, Confidence Bands, Credible Regions

1. Introduction

A common problem in econometrics, statistics, and machine learning is to estimate and make inference on functions that satisfy shape restrictions. These restrictions might arise either from the nature of the function and variables involved or from theoretical reasons. Examples of the first case include distribution functions, which are nondecreasing and range between zero and one, and height growth charts, which are nondecreasing in age. Examples of the second case include demand functions of utility-maximizing individuals, which are nonincreasing in price according to consumer demand theory; production functions of profit-maximizing firms, which are nondecreasing *and* quasi-concave in input quantities according to production theory *and* can also be concave in industries that exhibit diminishing returns to scale; bond yield curves, which are monotone and concave in time to maturity; and American and European call option prices, which are concave *and* monotone in the underlying stock price and increasing in volatility, according to the arbitrage pricing theory.[1]

We propose a method to enforce shape restrictions *ex post* on any initial generic point and interval estimates of functions by applying functional operators. If an operator has reshaping, invariance, order-preserving, and distance-reducing properties, enforcing the shape restriction improves the point estimates and improves the coverage property of the interval estimates. Thus, the shape-enforced point estimates are closer to the target function than the original point estimates under suitable distances, and the shape-enforced interval estimates have greater coverage and shorter length under suitable distances than the original interval estimates. We show that these properties hold for six different operators that enforce the following restrictions: range, convexity, monotonicity, joint convexity and monotonicity, quasi-convexity, and joint quasi-convexity and monotonicity, as well as for combinations of range with all of the above. We impose the range restriction with a natural operator that thresholds the estimates to the desired range. The double Legendre-Fenchel transform enforces convexity by transforming the estimates into their greatest convex minorants. We focus on the monotone rearrangement to enforce monotonicity (though projection on isotone class can also be used in all composition results, as well as convex combinations of isotone projection with rearrangement). We further develop a new operator to enforce quasi-convexity—a shape that has not been well explored in the literature, although it is common in applications. We also show that the compositions of the monotone rearrangement with the double Legendre-Fenchel and the new quasi-convexity operators yield monotone convex and monotone quasi-convex estimates, respectively. In other words, the application of the convex and quasi-convex operators does not affect the monotonicity of the function. We further demonstrate how to modify the operators to deal with concavity,

---

1. Different, but similar shape restrictions apply to put prices, with the American put price being log-concave in the stock price, for example.

quasi-concavity, their composition with the monotonicity and range operators, and shape restrictions on transformations of the function.

Our method is generic in that it can be applied to any point or interval estimator of the target function. For example, it works in combination with parametric, semi-parametric and nonparametric approaches to model and estimate the target function. It works with modern machine and deep learning methods that are able to exploit either smoothness or structured parsimony (e.g., approximate sparsity) of target functions. Hence our post-processed point and interval estimates automatically inherit the rates of convergence of these estimators and provably improve these properties in finite samples, while also enforcing qualitative shape restrictions brought by scientific reasoning. Moreover, our method applies without modification to any type of function including reduced form statistical objects such as conditional expectation, conditional density, conditional probability and conditional quantile functions, or causal and structural objects such as dose response, production, supply and demand functions identified and estimated using instrumental variable or other methods. The only requirement to obtain consistent point estimators or valid confidence bands is that the source point estimators be consistent or the source confidence bands be valid. There are many existing methods to construct such point estimators and confidence bands under general sampling conditions, including obtained through frequentist, Bayesian or approximate Bayesian methods.[2] Under misspecification these requirements may not be satisfied, but the shape-enforcing operators will bring improvements to the point estimators and confidence bands in a sense that we will make precise. To implement our method, we develop algorithms to compute the Legendre-Fenchel transform of multivariate functions and the new quasi-convexity enforcing operator.

We illustrate the theoretical results with two empirical applications to the estimation of a height growth chart for infants in India and a production function for chemical firms in China. In the case of the growth chart, we impose natural monotonicity in the effect of age, together with concavity that is plausible during early childhood. In the case of the production function, we enforce that a firm's output is nondecreasing and quasi-concave in labor and capital inputs according to standard production theory. We also consider imposing concavity in the effect of the inputs. In both applications we use series least squares methods to flexibly estimate the conditional expectation functions of interest, and construct confidence bands using bootstrap. We quantify the size of strict improvements that imposing shape restrictions bring to point and interval estimates in small samples through numerical simulations calibrated to the empirical applications.

Literature Review. Due to the wide range of applications of shape restrictions, shape-constrained estimation and inference have received a lot of attention in the statistics community. Classical examples include Hildreth (1954), Ayer et al. (1955), Brunk (1955), van Eeden (1956), Grenander (1956), Groeneboom et al. (2001), and Mammen (1991). We refer to Barlow et al. (1972) and Robertson et al. (1988) for classical references on isotonic

---

2. Bayesian methods are often used to quantify the uncertainty of complicated methods where the frequentist quantification is intractable, for example, in deep learning problems. Like in the classical approach, one may impose constraints directly during the estimation, though this is often quite cumbersome and is rarely done in practice. The post-processing can be applied to the unconstrained estimates and be justified on pragmatic grounds, ease of computation, or desire to analyze data without restrictions and accept a menu of restrictions ex-post only after validating them.

regression for monotonicity restrictions, and to Koenker and Mizera (2010) for the work on log-concave density estimation. In terms of risk bounds for estimation, please refer to Zhang (2002), Chatterjee et al. (2014), Han et al. (2019), and references therein for recent developments in isotonic regression; and Kuosmanen (2008), Seijo and Sen (2011), Guntuboyina and Sen (2015), Han and Wellner (2016), and Chen et al. (2020) for convex regression. Bellec (2018) established sharp oracle inequalities for least squares estimators, when only shape restrictions are known to hold. Moreover, Hengartner and Stark (1995), Dümbgen (2003), and Anevski and Hössjer (2006) considered the construction of confidence bands for univariate functions under monotonicity or convexity restrictions. Please refer to the book Groeneboom and Jongbloed (2014) and the survey paper Guntuboyina and Sen (2018) for more comprehensive reviews on estimation and inference under shape constraints.

Most existing works developed constrained methods via maximum likelihood methods for regression or density estimation that impose only shape restrictions and produce constrained estimates without further restrictions. We remark here that these direct approaches deliver advantages over our approach when such target functions are known to satisfy only the qualitative shape constraints. By contrast, our post-processing approach delivers advantages when *any generic* target function, in addition to satisfying qualitative constraints, satisfies smoothness or other structured parsimony restrictions (e.g., sparsity). Indeed, our method applies to generic problems, and is not tied to statistical parameters such as regression or density estimation. To explain where the advantages arise, we note that the direct isotone univariate regression converges to the true regression function at the $n^{-1/3}$ rate, which is minimax optimal for the parameter space of monotone functions. If the target function is known to lie in the space of smooth functions (Hölder or Sobolev with smoothness $s > 1$), the better and optimal rate $n^{-s/(2s+1)}$ can be achieved by an unconstrained estimator (e.g., Stone (1980)), making the pure isotonic regression suboptimal in this case. To fix the direct isotonic regression in this case, we would need to impose the smoothness constraints in the estimation directly, which ordinarily is not done in practice, let alone theoretically analyzed. (One exception here is Chernozhukov et al. (2015) that considered testing shape restrictions in Banach spaces, with the target function being (possibly partially) identified by general conditional moment condition problems, where shape restrictions induce a lattice structure on the space). Smooth cases and other problems, where unconstrained estimators achieve optimal rates, provide the chief motivation for our approach: in such cases, our method automatically inherits the optimal rate and improves the finite sample properties of the estimator through the distance-reducing properties. On the other hand, unlike constrained estimators, unconstrained estimators require delicate choices of tuning parameters to achieve the optimal rate, although adaptive estimation and inference of smooth functions is possible using the method of Lepskiĭ (1992) (e.g., Lepskiĭ and Spokoiny, 1997; Giné and Nickl, 2010b,a; Chernozhukov et al., 2014). It is worthwhile noting that Durot and Lopuhaa (2018) investigated asymptotic properties of smoothed isotonic estimators, and Jankowski and Wellner (2009) studied the method of rearrangements for obtaining discrete monotone distributions. However, these works only consider very specific classes of shape-constrained estimators, i.e., isotonic and/or discrete estimators.

Another recurrent problem with imposing shape restrictions in estimation is that the derivation of the statistical properties of the constrained estimators is involved and specific to the estimator and shape restriction. As a consequence, there exist very few distributional

results, mainly for univariate functions. The results available for the Grenander and isotonic regression estimators show that these estimators exhibit non-standard asymptotics (including relatively slow rates, since smoothness conditions are not exploited); see Guntuboyina and Sen (2018) for a recent review. Moreover, Horowitz and Lee (2017) and Freyberger and Reeves (2018) have recently pointed out the difficulties of developing inference methods from shape-constrained estimators with good uniformity properties with respect to the data generating process. They showed that for shape restrictions defined by inequalities, the distribution of the constrained estimator depends on where the inequalities are binding, which is unknown a priori. Inference based on this distribution therefore becomes sensitive to how close the inequalities are to binding relative to the sample size. We avoid all of these complications arising from the constrained estimators by enforcing the restrictions *ex post* and therefore relying on the distribution of the unconstrained estimators (whenever it is available) to construct the confidence bands. Our confidence interval method can also be applied on top of a different constrained estimator to provide potential improvements when the end-point functions of the generated confidence band do not themselves satisfy the restriction. It is worthwhile noting that the idea of *ex post* confidence bands was mentioned in Section 4.2 in Dümbgen (2003), which only discussed two cases on the univariate monotone function and univariate convex function. Moreover, the construction of confidence bands for the convex case in Dümbgen (2003) is quite different from ours (e.g., their lower bound is not necessarily a convex function).

Our paper generally follows the approach introduced in Chernozhukov et al. (2009), which focused on producing improved generic point and interval estimates of monotone functions using the monotone rearrangement. The class of shape enforcing operators covered by our paper is much bigger and much more useful, with analysis being much more challenging, and we view both aspects as a substantial contribution of our paper. Some of the operators that we consider have been analyzed previously in the literature. Dette and Volgushev (2008) apply a smoothed rearranged operator to kernel estimators for monotonization purposes and derive pointwise limit theory. Chernozhukov et al. (2010) applied the monotone rearrangement to deal with the quantile crossing problem and Belloni et al. (2019) to impose monotonicity in conditional quantile functions estimated using series quantile regression methods and construct monotonized uniform confidence bands. Beare and Fang (2017) used the double Legendre-Fenchel transform to construct point and interval estimates of univariate concave functions on the non-negative half-line. Other applications of the double Legendre-Fenchel transform include Delgado and Escanciano (2012), Beare and Moon (2015), Beare and Schmidt (2016). Chernozhukov et al. (2010) and Beare and Fang (2017) used an alternative approach to make inference on shape-constrained functions. Instead of applying the shape-enforcing operator to a confidence band constructed from an unconstrained estimator, they constructed confidence bands from the estimator after applying the shape-enforcing operator. To do so, they characterized the distribution of the constrained estimator from the distribution of the unconstrained via the delta method, after showing that the shape-enforcing operator is Hadamard or Hadamard directional differentiable. This approach usually yields narrower confidence bands than ours, but it is computationally more involved and requires additional assumptions and non-standard methods. For example, Beare and Fang (2017) showed that the bootstrap is inconsistent for the distribution of constrained estimators after applying the double Legendre-Fenchel

5

transform when the target function is not strictly concave. Finally, we refer to Matzkin (1994), and Chetverikov et al. (2018) for excellent, insightful up-to-date surveys on the use of shape restrictions in econometrics.

Relative to the literature, we summarize the major contributions of this paper as follows. (1) We introduce an operator to enforce quasi-convexity and deliver improved point and interval estimates of general multivariate quasi-convex functions. Quasi-convexity extends the notion of unimodality to multiple dimensions and generalizes convexity constraints. Despite its importance, the shape restriction of quasi-convexity has not been well studied in the literature and Guntuboyina and Sen (2018) listed quasi-convexity as an open area in shape-constrained estimation. (2) We extend the use of the Legendre-Fenchel transform to construct improved point and interval estimates of general multivariate convex functions. (3) We show that the composition of the monotone rearrangement with the Legendre-Fenchel transform can be used to construct improved point and interval estimates of monotone convex functions. (4) We show that the composition of the monotone rearrangement with our quasi-convex operator can be used to construct improved point and interval estimates of monotone quasi-convex functions. (The third and fourth contributions proved to be the most challenging and important steps, where the importance stems from shape restrictions often being a composition of monotonicity with convexity or quasi-concavity). (5) We provide a new algorithm to compute the Legendre-Fenchel transform of multivariate functions. (6) We develop an algorithm to compute our quasi-convex operator. The main advantage of our approach is that it works in conjunction with any generic point estimate (e.g., including recent machine and deep learning methods), or any generic interval estimate (that can be a frequentist confidence band or a Bayesian credible region). Because of genericity, it is able to exploit smoothness or other forms of structured parsimony through the use of the appropriate initial estimator. It inherits the rate properties of the initial estimator, while delivering better finite sample properties through distance-reducing inequalities.

Outline. The rest of the paper is organized as follows. Section 2 introduces the functional shape-enforcing operators and their properties, together with examples of operators that enforce the shape restrictions of interest. Section 3 discusses the use of shape-enforcing operators to obtain improved point and interval estimates of functions that satisfy shape restrictions. Section 4 provides algorithms to compute the shape-enforcing estimators. Section 5 reports the results of two empirical applications and numerical simulations calibrated to the applications. Section 6 concludes the paper. The proofs of the main results are gathered in the Appendix.

Notation. For any measurable function $f : \mathcal{X} \to \mathbb{R}$ and $p \geq 1$, let $\|f\|_p := \left\{ \int_{\mathcal{X}} |f(x)|^p dx \right\}^{1/p}$, the L^p-norm of f, with $\|f\|_{\infty} := \sup_{x \in \mathcal{X}} |f(x)|$, the L^{∞}-norm or sup-norm of f. We drop the subscript p for the Euclidean norm, i.e., $\|x\| := \|x\|_2$. For $p \geq 1$, let $\ell^p(\mathcal{X}) := \{f : \mathcal{X} \to \mathbb{R} : \|f\|_p < \infty\}$, the class of all measurable functions defined on $\mathcal{X}$ such that the L^p-norms of these functions is finite. For $x, x' \in \mathbb{R}^k$, we say $x \geq x'$ if every entry of x is no smaller than the corresponding entry in x'. For two functions f and g that map $\mathcal{X} \to \mathbb{R}$ we say that $f \leq g$ if $f(x) \leq g(x)$ for all $x \in \mathcal{X}$. We also use $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$ for any $a, b \in \mathbb{R}$. For two scalar sequences a_n and b_n, the notation $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$. For an operator $\mathbf{O} : \ell_0^{\infty}(\mathcal{X}) \to \ell_0^{\infty}(\mathcal{X})$, we use $\mathbf{O}^- f := -\mathbf{O}(-f)$. For

two operators $\mathbf{O_1} : \ell_0^\infty(\mathcal{X}) \to \ell_0^\infty(\mathcal{X})$ and $\mathbf{O_2} : \ell_0^\infty(\mathcal{X}) \to \ell_0^\infty(\mathcal{X})$, we define $\mathbf{O_1O_2}$ to be the composition $\mathbf{O_1} \circ \mathbf{O_2}$.

2. Functional Shape-Enforcing Operators

2.1 Properties of Shape-Enforcing Operators

Assume that the function of interest, f, is real-valued with domain $\mathcal{X} \subset \mathbb{R}^k$, for some positive integer k. Let $\ell^\infty(\mathcal{X})$ be the set of bounded *measurable* functions. Let $\ell_0^\infty(\mathcal{X})$ and $\ell_1^\infty(\mathcal{X})$ be two subspaces of $\ell^\infty(\mathcal{X})$, such that $\ell_1^\infty(\mathcal{X}) \subset \ell_0^\infty(\mathcal{X})$; and let $\mathbf{O} : \ell_0^\infty(\mathcal{X}) \to \ell_0^\infty(\mathcal{X})$ be a functional operator. In our case, $\ell_0^\infty(\mathcal{X})$ will be the class of unconstrained functions and $\ell_1^\infty(\mathcal{X})$ will be the subclass of functions that satisfy some shape restriction. We first introduce three properties that an operator must satisfy to be considered a shape-enforcing estimator.

Definition 1 (Shape-Enforcing Operator) *We say that an operator $\mathbf{O}$ is ℓ_1^∞-enforcing with respect to $\ell_0^\infty(\mathcal{X})$ if it satisfies the following properties:*

1. *Reshaping: the output of the operator is a function that satisfies the shape restriction:*

$$\mathbf{O}f \in \ell_1^\infty(\mathcal{X}), \text{ for any } f \in \ell_0^\infty(\mathcal{X}). \tag{2.1}$$

2. *Invariance: the operator should do nothing when the input function has already satisfied the shape restriction:*

$$\mathbf{O}f = f, \text{ for any } f \in \ell_1^\infty(\mathcal{X}). \tag{2.2}$$

3. *Order Preservation: the output functions preserve original order:*

$$\mathbf{O}f \leq \mathbf{O}g, \text{ for any } f, g \in \ell_0^\infty(\mathcal{X}) \text{ such that } f \leq g. \tag{2.3}$$

In addition to these properties, we consider the following "distance contraction" property.

Definition 2 (Distance-Reducing Operator) *Let ρ be a distance or semi-metric function on $\ell^\infty(\mathcal{X})$. We say that an operator $\mathbf{O}$ is a ρ-distance contraction if the output functions are weakly closer than input functions under the ρ:*

$$\rho(\mathbf{O}f, \mathbf{O}g) \leq \rho(f, g) \text{ for any } f, g \in \ell_0^\infty(\mathcal{X}). \tag{2.4}$$

Some particularly interesting cases of constrained classes $\ell_1^\infty(\mathcal{X})$ are subsets of functions that satisfy shape restrictions. In this paper, we focus on seven types of shape restrictions: (1) range, (2) convexity, (3) monotonicity, (4) monotone convexity, (5) quasi-convexity, (6) monotone quasi-convexity, and (7) compositions of range with all of the above. Our methods also apply to the restrictions of concavity and quasi-concavity by noting that if f is concave (quasi-concave), then $-f$ is convex (quasi-convex). In the case of monotonicity we focus on the case of monotonically nondecreasing functions. The methods also apply to monotonically nonincreasing functions noting that if f is nondecreasing then $-f$ is nonincreasing.

Remark 3 (Counterexample) *There are operators to impose shape restrictions that do not satisfy the conditions of Definitions 1 and 2. For example, $\mathbf{O}f = f/\|f\|_1$ is a natural operator to enforce that a non-negative function integrates to one in density estimation. This operator satisfies reshaping and invariance, but does not satisfy order preservation nor distance contraction for any L^p-norm.*

2.2 Range Restrictions

We first consider the subset of range-constrained functions $\ell_R^\infty(\mathcal{X}) := \{f \in \ell^\infty(\mathcal{X}) : \underline{f} \leq f(x) \leq \overline{f}$ for all $x \in \mathcal{X}\}$ for some constants $\underline{f} \leq \overline{f}$. A natural range-enforcing operator is as follows.

Definition 4 (R-Operator) *For any set $\mathcal{X} \subset \mathbb{R}^k$, the range operator $\mathbf{R} : \ell^\infty(\mathcal{X}) \to \ell^\infty(\mathcal{X})$ is defined by thresholding the values of the function f to $[\underline{f}, \overline{f}]$.*

$$\mathbf{R}f(x) := [\underline{f} \vee f(x)] \wedge \overline{f}, \text{ for any } x \in \mathcal{X}. \tag{2.5}$$

Let d_p be the distance measure induced by the L^p-norm, i.e., $d_p(f, g) = \|f - g\|_p$ for any $f, g \in \ell^p(\mathcal{X})$ and $p \geq 1$. The following theorem shows that $\mathbf{R}$ is indeed range-enforcing and distance-reducing with respect to d_p.

Theorem 5 (Range-Enforcing Operator) *The operator $\mathbf{R}$ is ℓ_R^∞-enforcing with respect to $\ell^\infty(\mathcal{X})$ and a d_p-distance contraction for any $p \geq 1$.*

2.3 Convexity

Let $\mathcal{X}$ be a convex subset of $\mathbb{R}^k$. Let $\ell_S^\infty(\mathcal{X}) := \{f \in \ell^\infty(\mathcal{X}) : \liminf_{x' \to x} f(x') \geq f(x)$ for all $x \in \mathcal{X}\}$, the set of bounded lower semi-continuous functions on $\mathcal{X}$, and $\ell_C^\infty(\mathcal{X}) := \{f \in \ell_S^\infty(\mathcal{X}) : f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha)f(x')$ for all $x, x' \in \mathcal{X}, \alpha \in [0,1]\}$, the subset of convex functions on $\mathcal{X}$. We consider the Double Legendre-Fenchel (DLF) transform as a convexity-enforcing operator. To define this operator, we first recall the definition of the Legendre-Fenchel transform (see, e.g., Hiriart-Urruty and Lemaréchal (2001)).

Definition 6 (Legendre-Fenchel transform) *For any convex set $\mathcal{X} \subset \mathbb{R}^k$ and $f \in \ell^\infty(\mathcal{X})$, let $f^*(\mathcal{X}) := \{\xi \in \mathbb{R}^k : \sup_{x \in \mathcal{X}}\{\xi'x - f(x)\} < \infty\}$. The Legendre-Fenchel transform $\mathbf{L}_\mathcal{X} : \ell^\infty(\mathcal{X}) \to \ell^\infty(f^*(\mathcal{X}))$ is defined by*

$$f^*(\xi) := \mathbf{L}_\mathcal{X} f(\xi) := \sup_{x \in \mathcal{X}}\{\xi'x - f(x)\}, \text{ for any } \xi \in f^*(\mathcal{X}).$$

The function $\xi \mapsto f^*(\xi)$ is a closed convex function (see Lemma 34 in the Appendix) which is also called the convex conjugate of f, and the Legendre-Fenchel transform $\mathbf{L}_\mathcal{X}$ is also called the conjugate operator. The Legendre-Fenchel transform is a functional operator that maps any function f to a function of its family of tangent planes, which is often referred to as the dual function of f.

Definition 7 (C-Operator) *For any convex set $\mathcal{X} \subset \mathbb{R}^k$, the double Legendre-Fenchel operator $\mathbf{C} : \ell_S^\infty(\mathcal{X}) \to \ell_S^\infty(\mathcal{X})$ is defined by the repeated application of the Legendre-Fenchel transform twice:*

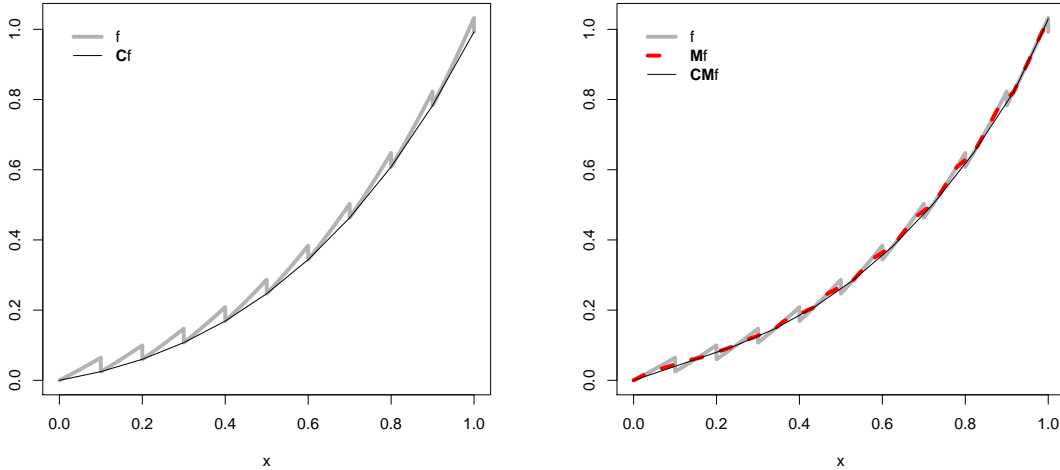$$\mathbf{C}f := \mathbf{L}_{f^*(\mathcal{X})} \circ \mathbf{L}_\mathcal{X} f.$$

Figure 1: Left: unconstrained and convexity-enforced functions. Right: unconstrained, monotonicity-enforced, and (convexity and monotonicity)-enforced functions. The original function is $f(x) = [10 \exp(3x/2) - \lfloor 10x \rfloor - 10]/25$.

Lemma 36 in the Appendix shows that the double Legendre-Fenchel operator maps any lower semi-continuous function f to its greatest convex minorant, i.e., the largest function $g \in \ell_C^\infty(\mathcal{X})$ such that $g \leq f$. The left panel of Figure 1 illustrates this property with a graphical example. We apply the operator $\mathbf{C}$ to the function $f(x) = [10 \exp(3x/2) - \lfloor 10x \rfloor - 10]/25$ on $\mathcal{X} = [0, 1]$, where $\lfloor \cdot \rfloor$ is the floor function. The convexity-enforced function $\mathbf{C}f$ is the greatest convex minorant of f, $\mathbf{C}f(x) = 10[f(\overline{x})(x - \underline{x}) + f(\underline{x})(\overline{x} - x)]$ if $\overline{x} \neq \underline{x}$ and $\mathbf{C}f(x) = f(x)$ otherwise, where $\overline{x} = \lceil 10x \rceil/10$, $\underline{x} = \lfloor 10x \rfloor/10$, and $\lceil \cdot \rceil$ is the ceiling function.

The following theorem is an immediate consequence of applying known results from convex analysis.

Theorem 8 (Convexity-Enforcing Operator) *For any convex set $\mathcal{X}$, the operator $\mathbf{C}$ is ℓ_C^∞-enforcing with respect to $\ell_S^\infty(\mathcal{X})$ and a d_∞-distance contraction.*

Remark 9 (Shifted Convexity-Enforcing Operator) *The convexity-enforced function is a minorant of the original function (see fig. 1). When the original function is estimated, the application of the $\mathbf{C}$-operator might introduce downward bias, especially in small samples. A way of reducing bias is by shifting the $\mathbf{C}$-operator, that is*

$$\mathbf{SC}f(x) := \mathbf{C}f(x) + \lambda(\mathcal{X})^{-1} \int_{\mathcal{X}} [f(x) - \mathbf{C}f(x)]dx,$$

where $\lambda(\mathcal{X})$ is the Lebesgue measure of $\mathcal{X}$. The $\mathbf{SC}$-operator is reshaping and invariant, but does not preserve order nor reduce distance. We compare the $\mathbf{C}$-operator with the $\mathbf{SC}$-operator in the numerical examples of Section 5.

9

2.4 Monotonicity

Let $\ell_M^\infty(\mathcal{X}) := \{f \in \ell^\infty(\mathcal{X}) : f(x') \le f(x) \text{ for all } x, x' \in \mathcal{X} \text{ such that } x' \le x\}$, the set of bounded nondecreasing, measurable functions on $\mathcal{X}$. We consider the multivariate monotone rearrangement of Chernozhukov et al. (2009) as a monotonicity-enforcing operator.

Definition 10 (M-Operator) *For any rectangular set $\mathcal{X}$ that is regular (i.e., has non-empty interior in $\mathbb{R}^k$), the multivariate increasing rearrangement operator $\mathbf{M} : \ell^\infty(\mathcal{X}) \to \ell_M^\infty(\mathcal{X})$ is defined by*

$$\mathbf{M}f := \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbf{M}_\pi f,$$

where $\pi = (\pi_1, \ldots, \pi_k)$ is a permutation of the integers $1, \ldots, k$, Π is a non-empty subset of all possible permutations π, and $\mathbf{M}_\pi f := \mathbf{M}_{\pi_1} \circ \cdots \circ \mathbf{M}_{\pi_k} f$, where $\mathbf{M}_j f$ is the one-dimensional increasing rearrangement applied to the function $x(j) \mapsto f(x(j), x(-j))$ defined by

$$\mathbf{M}_j f(x) := \inf \left\{ y \in \mathbb{R} : \int_{\mathcal{X}(j)} 1\{f(t, x(-j)) \le y\} dt \ge x(j) \right\},$$

the one-dimensional increasing rearrangement applied to the function $x(j) \mapsto f(x(j), x(-j))$. Here, we use $f(x(j), x(-j))$ to denote the dependence of f on $x(j)$, the jth-component of x, and all other arguments, $x(-j)$, and $\mathcal{X}(j)$ to denote the domain of $x(j) \mapsto f(x(j), x(-j))$.

Proposition 2 of Chernozhukov et al. (2009) showed that the multivariate increasing rearrangement is monotonicity-enforcing and distance-reducing with respect to d_p for any $p \ge 1$. We state this result as a theorem for the purpose of completeness.

Theorem 11 (Monotonicity Operator) *For any regular rectangular set $\mathcal{X}$, the operator $\mathbf{M}$ is ℓ_M^∞-enforcing with respect to $\ell^\infty(\mathcal{X})$ and is a d_p-distance contraction for any $p \ge 1$.*

Remark 12 (Isotonization Operators) *Isotonization operators, i.e., projections on the set of weakly increasing functions, can also be considered in place of rearrangement and the results below apply to them. Here we focus on the rearrangement for conciseness. Chernozhukov et al. (2009) showed, for the one-dimensional case, that isotonization and convex linear combinations of monotone rearrangement and isotonic regression are also ℓ_M^∞-enforcing operators with respect to $\ell^\infty(\mathcal{X})$ and d_p-distance contractions for any $p \ge 1$. Extension to the multivariate case follows analogously to Chernozhukov et al. (2009) by an induction argument.*

Remark 13 (Multivariate Distributions) *Multivariate distribution functions satisfy stronger shape restrictions than monotonicity. For example, in the bivariate case they are 2-increasing (supermodular) and grounded Nelsen (2007). The grounded restriction can be enforced using a simple variation of the range operator. We are not aware of any operator that enforces supermodularity.*

2.5 Convexity and Monotonicity

Let $\ell_{CM}^\infty(\mathcal{X}) := \ell_C^\infty(\mathcal{X}) \cap \ell_M^\infty(\mathcal{X})$ be the set of bounded convex and nondecreasing functions on $\mathcal{X}$. We consider the composition of the $\mathbf{C}$ and $\mathbf{M}$ operators to enforce both convexity and monotonicity.

Definition 14 (CM-Operator) *For any regular rectangular set $\mathcal{X}$, the convex rearrangement operator $\mathbf{CM} : \ell_S^\infty(\mathcal{X}) \to \ell_S^\infty(\mathcal{X})$ is defined by*

$$\mathbf{CM}f := \mathbf{C} \circ \mathbf{M}f.$$

Remark 15 (Rectangular Domain) *The $\mathbf{C}$-operator does not preserve monotonicity in general.[3] When $\mathcal{X}$ is a regular rectangle, the $\mathbf{C}$-operator can be obtained by separate application to each face of the rectangle and does not affect the monotonicity of the function; see Lemma 38 in the Appendix. From a practical point of view, we do not find this assumption very restrictive because the domains usually have the product form $\mathcal{X} = [a_1, b_1] \times \cdots \times [a_k, b_k]$ in applications. If the domain of the target function is not rectangular, we can either restrict the analysis to a rectangular subset of the domain or extend the function to a rectangular set that contains the domain.*

In the right panel of Figure 1, we apply the operators $\mathbf{M}$ and $\mathbf{CM}$ to the function $f(x) = [10\exp(3x/2) - \lfloor 10x \rfloor - 10]/25$ on $\mathcal{X} = [0, 1]$. The monotonicity-enforced function $\mathbf{M}f$ in dashed line is not convex, whereas the (convexity and monotonicity)-enforced function $\mathbf{CM}f$ is both monotone and convex. Indeed, $\mathbf{CM}f$ is the greatest convex minorant of $\mathbf{M}f$.

Theorem 16 (Convexity and Monotonicity Operator) *For any regular rectangular set $\mathcal{X}$, the operator $\mathbf{CM}$ is ℓ_{CM}^∞-enforcing with respect to $\ell_S^\infty(\mathcal{X})$ and a d_∞-distance contraction.*

Remark 17 (Proof of Theorem 16 and ordering of composition) *The proof of Theorem 16 does not follow from combining Theorems 8 and 11. As indicated in Remark 15, the argument is more subtle as we need to verify that the application of the $\mathbf{C}$-operator preserves monotonicity. Moreover, the order of the composition of the operators matters. Thus, $\mathbf{MC} := \mathbf{M} \circ \mathbf{C}$ is not ℓ_{CM}^∞-enforcing because the operator $\mathbf{M}$ does not preserve convexity in general.*

Remark 18 (Concavity and Monotonicity) *Using the notation for inverse operators given in the introduction, we can construct composite operators for all the combinations of concavity/convexity and increasing/decreasing monotonicity restrictions. Thus, the operator $\mathbf{CM}^- f = \mathbf{C}[-\mathbf{M}(-f)]$ enforces convexity and decreasing monotonicity, $\mathbf{C}^-\mathbf{M}f = -\mathbf{C}[-\mathbf{M}(f)]$ enforces concavity and increasing monotonicity, and $\mathbf{C}^-\mathbf{M}^- f = -\mathbf{C}[\mathbf{M}(-f)]$ enforces concavity and decreasing monotonicity. It can be shown that these operators satisfy analogous properties to $\mathbf{CM}$ by a straightforward modification of the proof of Theorem 16.*

---

3. Let $\mathcal{X} \subset \mathbb{R}^2$ be a triangular set with vertices at $(-1, 3)$, $(0, 0)$ and $(3, -1)$. Then, the function $f(x) = 1 + 3(x_1 + x_2)/2 - |x_1 - x_2|$ is increasing on $\mathcal{X}$, but its greatest convex minorant $\mathbf{C}f(x) = 1 - (x_1 + x_2)/2$ is decreasing on $\mathcal{X}$.

2.6 Quasi-convexity

Quasi-convexity is a global property of a function, which is weaker than convexity. A convex function must be quasi-convex, but a quasi-convex function is not necessarily convex. Intuitively, a function f defined on a convex domain is quasi-convex if and only if all its level sets are convex. Please also see its definition in Eq. (2.6) below. Quasi-convexity is commonly used in economics because it is an ordinal property, preserved by monotone transformations, which represents well economic relationships such as utility and production functions Guerraggio and Molho (2004); Koenker and Mizera (2010); Crouzeix (2017). Moreover, quasi-convex functions have good optimization properties.

Consider the set of bounded lower semi-continuous quasi-convex functions on $\mathcal{X}$:

$$\ell_Q^\infty(\mathcal{X}) := \{f \in \ell_S^\infty(\mathcal{X}) : f(\alpha x + (1-\alpha)x') \leq \max\{f(x), f(x')\} \text{ for all } x, x' \in \mathcal{X}, \alpha \in [0,1]\}.$$
(2.6)

We note that $\ell_C^\infty(\mathcal{X}) \subset \ell_Q^\infty(\mathcal{X})$, and that for any $f \in \ell_Q^\infty(\mathcal{X})$, the lower contour sets, $\mathcal{I}_f(y) := \{x \in \mathcal{X} : f(x) \leq y\}$, are convex for all $y \in \mathbb{R}$. For any set $\mathcal{Z} \subset \mathbb{R}^k$, let $\text{conv}(\mathcal{Z})$ denote the convex hull of $\mathcal{Z}$. We consider the following new operator to impose quasi-convexity:

Definition 19 (Q-Operator) *For any convex and compact set $\mathcal{X} \subset \mathbb{R}^k$, the quasi-convexity operator $\mathbf{Q} : \ell_S^\infty(\mathcal{X}) \to \ell_S^\infty(\mathcal{X})$ is defined by*

$$\mathbf{Q}f(x) := \min\{y \in \mathbb{R} : x \in \text{conv}[\mathcal{I}_f(y)]\}. \tag{2.7}$$

Remark 20 (Existence of Q-Operator) *The restriction of the operator $\mathbf{Q}$ to $\ell_S^\infty(\mathcal{X})$, where $\mathcal{X}$ is convex and compact, guarantees that the minimum in (2.7) exists (see Lemma 40 in the Appendix). When the set $\mathcal{X}$ is non-compact or the function $f \notin \ell_S^\infty(\mathcal{X})$, there exist counter examples such that the minimum in (2.7) does not exists.[4] In such cases, one might still define $\mathbf{Q}f(x) := \inf\{y \in \mathbb{R} : x \in \text{conv}[\mathcal{I}_f(y)]\}$, but this operator appears to lose the contraction property stated below.*

The operator $\mathbf{Q}$ transforms any bounded lower semi-continuous function into a quasi-convex function. To see this, recall that a function is quasi-convex if its domain and all its lower contour sets are convex. By construction, $x \in \text{conv}[\mathcal{I}_f(y)]$ if and only if $\mathbf{Q}f(x) \leq y$. Therefore, the lower contour set of $\mathbf{Q}f$ at any level $y \in \mathbb{R}$ is $\mathcal{I}_{\mathbf{Q}f}(y) = \{x \in \mathcal{X} : \mathbf{Q}f(x) \leq y\} = \text{conv}[\mathcal{I}_f(y)]$, which is a convex set.

The left panel of Figure 2 shows a graphical example. We apply the operator $\mathbf{Q}$ to the function $f(x) = [10\exp(3x/2) - \lfloor 10x \rfloor - 10]/25$ on $\mathcal{X} = [0,1]$. Here we can see that the function $\mathbf{Q}f$ is the greatest quasi-convex minorant of f, $\mathbf{Q}f(x) = \min\{f(x), f(\lceil 10x \rceil/10)\}$.

Theorem 21 (Quasi-Convexity Operator) *For any convex and compact set $\mathcal{X}$, the operator $\mathbf{Q}$ is ℓ_Q^∞-enforcing with respect to $\ell_S^\infty(\mathcal{X})$ and a d_∞-distance contraction.*

---

4. For example, the function $f(x) = x[1(x \leq 0)/(2+x) - 1(x > 0)/(2-x)]$ if $|x| \neq 1$ and $f(1) = f(-1) = -1$ is not lower-semicontinuous on $\mathcal{X} = [-1,1]$. In this case, for any $x \in (-1,1)$, $\{y : x \in \text{conv}[\mathcal{I}_f(y)]\} = (-1, +\infty)$, so that $\mathbf{Q}(x)$ is not well-defined. The same problem arises if $\mathcal{X} := (-1,1)$, i.e., the domain $\mathcal{X}$ is not compact.
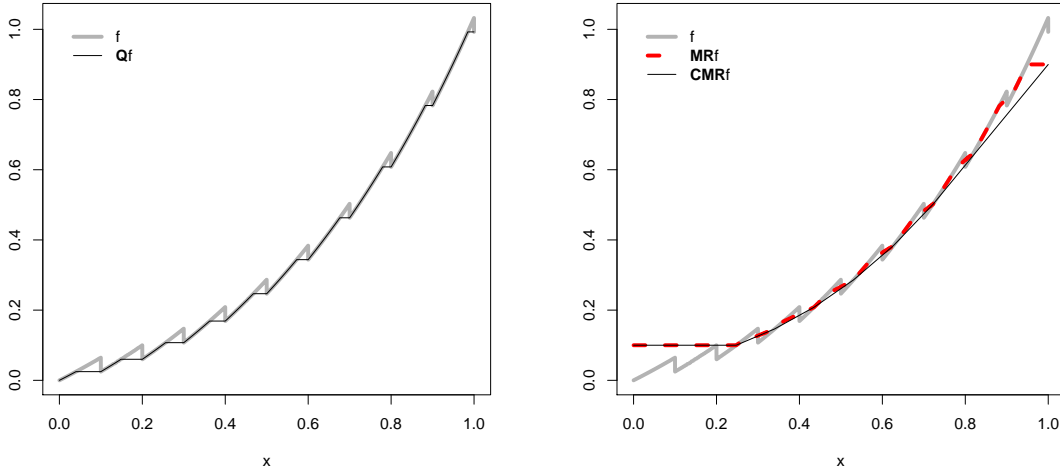
Figure 2: Left: unconstrained and quasiconvexity-enforced functions. Right: unconstrained and (convexity, monotonicity and range)-enforced functions. The original function is $f(x) = [10\exp(3x/2) - \lfloor 10x \rfloor - 10]/25$.

Remark 22 (Shifted Quasi-Convexity-Enforcing Operator) *By similar reasons to Remark 9, we introduce the shifted quasi-convexity-enforcing operator:*

$$\mathbf{SQ}f(x) := \mathbf{Q}f(x) + \lambda(\mathcal{X})^{-1} \int_{\mathcal{X}} [f(x) - \mathbf{Q}f(x)]dx,$$

where $\lambda(\mathcal{X})$ is the Lebesgue measure of $\mathcal{X}$.

2.7 Quasi-Convexity and Monotonicity

Let $\ell_{QM}^\infty(\mathcal{X}) := \ell_Q^\infty(\mathcal{X}) \cap \ell_M^\infty(\mathcal{X})$ be the set of bounded quasi-convex and partially nondecreasing functions on $\mathcal{X}$. This case is only relevant when $k > 1$ because univariate monotone functions are quasi-convex. We consider the composition of the $\mathbf{Q}$ and $\mathbf{M}$ operators to impose both quasi-convexity and monotonicity.

Definition 23 (QM-Operator) *For any regular rectangular set $\mathcal{X}$, the quasi-convex rearrangement operator $\mathbf{QM} : \ell_S^\infty(\mathcal{X}) \mapsto \ell_S^\infty(\mathcal{X})$ is defined by*

$$\mathbf{QM}f := \mathbf{Q} \circ \mathbf{M}f.$$

Theorem 24 (Quasi-Convexity and Monotonicity Operator) *For any regular rectangular set $\mathcal{X}$, the operator $\mathbf{QM}$ is ℓ_{QM}^∞-enforcing with respect to $\ell_S^\infty(\mathcal{X})$ and a d_∞-distance contraction.*

The comments and example in Remark 15 also apply to the $\mathbf{QM}$-operator. Thus, the assumption that $\mathcal{X}$ is a rectangular set is sufficient to guarantee that the $\mathbf{Q}$-operator preserves monotonicity.

Remark 25 (Quasi-Concavity and Monotonicity) *Similar to Remark 18, we can construct composite operators for all the combinations of quasi-concavity/quasi-convexity and increasing/decreasing monotonicity restrictions. Thus, the operator $\mathbf{QM}^- f = \mathbf{Q}[-\mathbf{M}(-f)]$ enforces quasi-convexity and decreasing monotonicity, $\mathbf{Q}^-\mathbf{M}f = -\mathbf{Q}[-\mathbf{M}(f)]$ enforces quasi-concavity and increasing monotonicity, and $\mathbf{Q}^-\mathbf{M}^- f = -\mathbf{Q}[\mathbf{M}(-f)]$ enforces quasi-concavity and decreasing monotonicity. It can be shown that these operators satisfy analogous properties to $\mathbf{QM}$ by a straightforward modification of the proof of Theorem 24.*

2.8 Range and Other Shape Restrictions

The following theorem shows that the operator $\mathbf{R}$ can be composed with $\mathbf{C}$, $\mathbf{M}$ and $\mathbf{Q}$ to produce range-constrained convex, monotone or quasi-convex functions. Let $\ell^\infty_{CR}(\mathcal{X}) := \ell^\infty_C(\mathcal{X}) \cap \ell^\infty_R(\mathcal{X})$, $\ell^\infty_{MR}(\mathcal{X}) := \ell^\infty_M(\mathcal{X}) \cap \ell^\infty_R(\mathcal{X})$, $\ell^\infty_{QR}(\mathcal{X}) := \ell^\infty_Q(\mathcal{X}) \cap \ell^\infty_R(\mathcal{X})$, $\mathbf{CR} := \mathbf{C} \circ \mathbf{R}$, $\mathbf{MR} := \mathbf{M} \circ \mathbf{R}$, and $\mathbf{QR} := \mathbf{Q} \circ \mathbf{R}$.

Theorem 26 (Composition with Range Operator) *(i) For any convex set $\mathcal{X}$, the operator $\mathbf{CR}$ is ℓ^∞_{CR}-enforcing with respect to $\ell^\infty_S(\mathcal{X})$ and a d_∞-distance contraction; (ii) for any regular rectangular set $\mathcal{X}$, the operator $\mathbf{MR}$ is ℓ^∞_{MR}-enforcing with respect to $\ell^\infty(\mathcal{X})$ and a d_p-distance contraction for any $p \geq 1$; and (iii) for any convex and compact set $\mathcal{X}$, the operator $\mathbf{QR}$ is ℓ^∞_{QR}-enforcing with respect to $\ell^\infty_S(\mathcal{X})$ and a d_∞-distance contraction.*

The operator $\mathbf{MR}$ can be composed with $\mathbf{C}$ and $\mathbf{Q}$ to produce range-constrained monotone convex or quasi-convex functions. The properties of the resulting operators $\mathbf{CMR} := \mathbf{C} \circ \mathbf{MR}$ and $\mathbf{QMR} := \mathbf{Q} \circ \mathbf{MR}$ follow from combining Theorem 26 with Theorems 16 and 24, respectively. Let $\ell^\infty_{CMR}(\mathcal{X}) := \ell^\infty_{CM}(\mathcal{X}) \cap \ell^\infty_R(\mathcal{X})$ and $\ell^\infty_{QMR}(\mathcal{X}) := \ell^\infty_{QM}(\mathcal{X}) \cap \ell^\infty_R(\mathcal{X})$.

Corollary 27 (Composition with Range and Monotonicity Operators) *(i) For any regular rectangular set $\mathcal{X}$, the operator $\mathbf{CMR}$ is ℓ^∞_{CMR}-enforcing with respect to $\ell^\infty_S(\mathcal{X})$ and a d_∞-distance contraction; and (ii) for any regular rectangular set $\mathcal{X}$, the operator $\mathbf{QMR}$ is ℓ^∞_{QMR}-enforcing with respect to $\ell^\infty_S(\mathcal{X})$ and a d_∞-distance contraction.*

In the right panel of Figure 2, we apply the operators $\mathbf{MR}$ and $\mathbf{CMR}$ to the function $f(x) = [10\exp(3x/2) - \lfloor 10x \rfloor - 10]/25$ on $\mathcal{X} = [0,1]$. We enforce that the range be in the interval $[0.1, 0.9]$. The (monotonicity and range)-enforced function $\mathbf{MR}f$ in dashed line satisfies the monotonicity and range restrictions but is not convex. The (convexity, monotonicity and range)-enforced function $\mathbf{CMR}f$ satisfies the three shape restrictions.

2.9 Shape Restrictions on Transformations

The shape operators can be combined with other functions to enforce shape restrictions on transformations of the function f. An example is log-concavity where we assume that $\log f$ is concave.[5] Let h be a real-valued bijection with inverse function h^{-1}. We consider the operator $\mathbf{O}_h$ that applies the operator $\mathbf{O}$ to the transformation $h(f)$ and then recovers the shape-constrained version of f by inversion, that is

$$\mathbf{O}_h f = h^{-1} \circ \mathbf{O}(h \circ f).$$

---

5. Bagnoli and Bergstrom (2005) discussed applications of log-concavity to economics and statistics, and analyzed the log-concavity properties of common distributions.

For example, if f is log-concave, then $h(x) = \log x$ and

$$\mathbf{O}_h f = \mathbf{C}_{\log}^- f = \exp[\mathbf{C}^-(\log f)].$$

The following theorem gives conditions under which the transformations h and h^{-1} preserve the properties of the operator $\mathbf{O}$. Define $\ell_{h,j}^\infty(\mathcal{X}) = \{f \in \ell_0^\infty(\mathcal{X}) : h \circ f \in \ell_j^\infty(\mathcal{X})\}$ for $j \in \{0, 1\}$.

Theorem 28 (Properties of $\mathbf{O}_h$-Operator) *Let $\mathbf{O}$ be a $\ell_1^\infty(\mathcal{X})$-enforcing operator with respect to $\ell_0^\infty(\mathcal{X})$, and $y \mapsto h(y)$ be a real valued strictly monotonic bijection on the domain $\mathcal{Y} \subset \mathbb{R}$. Then, $\mathbf{O}_h$ is a $\ell_{h,1}$-enforcing operator with respect to $\ell_{h,0}^\infty(\mathcal{X})$. Moreover, if $\mathbf{O}$ is a ρ-distance contraction, then $\mathbf{O}_h$ is a ρ_h-distance contraction for $\rho_h(f, g) := \rho(h \circ f, h \circ g)$.*

2.10 Other Ways of Generating Shape Enforcing Operators

When $\ell_0^\infty(\mathcal{X})$ is a Hilbert space and $\ell_1^\infty(\mathcal{X})$ is a closed set in the L^2 metric, it is possible to construct generic shape-enforcing operators via L^2-projection in $\ell_1^\infty(\mathcal{X})$:

Definition 29 (P-Operator) *The L^2-projection operator on the Hilbert space $\ell_0^\infty(\mathcal{X})$, $\mathbf{P} : \ell_0^\infty(\mathcal{X}) \to \ell_0^\infty(\mathcal{X})$, is defined by*

$$\mathbf{P}f(x) := \arg \min_{g \in \ell_1^\infty(\mathcal{X})} \|f - g\|_2. \tag{2.8}$$

The **P**-operator involves an infinite dimensional optimization program that can be computationally challenging except for special cases. For example, when $\ell_1^\infty(\mathcal{X}) = \ell_M^\infty(\mathcal{X})$, the **P**-operator corresponds to the isotonization operator that can be computed using the pool adjacent violators algorithm described in Barlow et al. (1972). The following result, discussed on p.45 of Chetverikov et al. (2018) and stated here as a theorem for the purpose of completeness, shows that **P** is range-enforcing and distance-reducing with respect to d_2 under some conditions on $\ell_1^\infty(\mathcal{X})$.

Theorem 30 (L^2-Projection Operator) *If $\ell_0^\infty(\mathcal{X})$ is a Hilbert space, $\ell_1^\infty(\mathcal{X})$ is a closed and convex set, and for any $f_1, f_2 \in \ell_1^\infty(\mathcal{X})$ the pointwise maximum and minimum of f_1 and f_2 belongs to $\ell_1^\infty(\mathcal{X})$, then the operator $\mathbf{P}$ is ℓ_1^∞-enforcing with respect to $\ell_0^\infty(\mathcal{X})$ and a d_2-distance contraction.*

The condition that $\ell_0^\infty(\mathcal{X})$ is a Hilbert space is satisfied when $\mathcal{X}$ is bounded. The sets $\overline{\ell}_C^\infty(\mathcal{X})$, $\overline{\ell}_M^\infty(\mathcal{X})$, $\overline{\ell}_Q^\infty(\mathcal{X})$ and their intersections are convex and closed.[6] The condition on the maximum and minimum is a more restrictive Hilbert lattice property. It is satisfied by $\overline{\ell}_M^\infty(\mathcal{X})$, but not by $\overline{\ell}_C^\infty(\mathcal{X})$ and $\overline{\ell}_Q^\infty(\mathcal{X})$. Theorem 30 therefore covers the isotonization operator, but not the convex and quasi-convex projections.

Remark 31 (Proof of Theorem 30) *Chetverikov et al. (2018) referred to Lemma 2.4 in Nishimura and Ok (2012) for the order-preserving and to Lemma 46.5.4 in Zeidler (1984) for the distance contraction. Reshaping and invariance hold trivially by the definition of the operator.*

---

6. The set $\overline{\ell}_O^\infty(\mathcal{X})$ denotes the intersection of $\ell_O^\infty(\mathcal{X})$ with the set of uniformly bounded functions, for $O \in \{C, M, Q\}$. The intersection ensures that $\overline{\ell}_O^\infty(\mathcal{X})$ is closed in the L^2 metric. Alternatively, we can ensure that $\ell_O^\infty(\mathcal{X})$ is closed in the L^2 metric by restricting $\mathcal{X}$ to be a countable set.

3. Improved Point and Interval Estimation

We show how to use shape-enforcing operators to improve point and interval estimators of a shape-constrained function. Let $f_0 : \mathcal{X} \to \mathbb{R}$ be the target function, which is known to satisfy a shape restriction, i.e., $f_0 \in \ell_1^\infty(\mathcal{X})$. Assume we have a point estimator f of f_0, and an interval estimator or uniform confidence band $[f_l, f_u]$ for f_0. These estimators are unconstrained and therefore do not necessarily satisfy the shape restrictions, i.e., $f, f_l, f_u \in \ell_0^\infty(\mathcal{X})$ but $f, f_l, f_u \notin \ell_1^\infty(\mathcal{X})$ in general.

There are many different ways to obtain these initial estimators, ranging from parametric to modern adaptive nonparametric methods (e.g., Fan and Gijbels, 1996; Li and Racine, 2007; Hastie et al., 2009). These methods can be tailored to properties of the target function such as smoothness or sparsity. A common frequentist confidence band for the function f_0 is constructed as

$$f_l(x) = f(x) - c_p s(x), \quad f_u(x) = f(x) + c_p s(x),$$

where $s(x)$ is the standard error of $f(x)$ and c_p is a critical value chosen such that

$$\mathrm{P}(f_0 \in [f_l, f_u]) \geq p,$$

for some confidence level p, where event $f_0 \in [f_l, f_u]$ means $\{f_0(x) \in [f_l(x), f_u(x)]$ for all $x \in \mathcal{X}\}$. Wasserman (2006) provides an excellent overview of methods for constructing the critical value; see also Giné and Nickl (2010a) and Chernozhukov et al. (2014) for constructions of adaptive confidence bands in low-dimensional smooth nonparametric models and Bach et al. (2020) for a recent proposal in high-dimensional generalized additive models. With a slight abuse of notation, an initial Bayesian credible region $[f_l, f_u]$ can be constructed similarly with the constant c_p determined such that

$$\Pi\{f_0 \in [f_l, f_u] \mid S\} \geq p,$$

where S denotes data (can be a set of statistics derived from data in robust Bayes procedures, for example, means or empirical moment functions), $[f_l, f_u]$ is a measurable function of S, and $\Pi(\cdot \mid S)$ denotes posterior distribution of parameter f_0 (viewed as a random element in the Bayesian approach), induced by S and a prior distribution over potential values f_0 can take. We give empirical and numerical examples in Section 5.

To enforce the shape restriction, we apply a suitable shape-enforcing operator to the original point estimator and end-point functions of the confidence band. The resulting estimator, $\mathbf{O}f$, and confidence band, $[\mathbf{O}f_l, \mathbf{O}f_u]$, improve over f and $[f_l, f_u]$ in the sense that f lies weakly closer to f_0 and the width of the band $[\mathbf{O}f_l, \mathbf{O}f_u]$ is weakly smaller than that of $[f_l, f_u]$, while the coverage is weakly greater. These properties of $\mathbf{O}f$ and $[\mathbf{O}f_l, \mathbf{O}f_u]$ are a corollary of Definition 1:

Corollary 32 (Improved Point and Interval Estimators) *Suppose we have a target function $f_0 \in \ell_1^\infty(\mathcal{X})$, an estimator $f \in \ell_0^\infty(\mathcal{X})$ a.s., and a confidence band $[f_l, f_u]$ such that $f_l, f_u \in \ell_0^\infty(\mathcal{X})$ a.s. If the operator $\mathbf{O}$ is ℓ_1^∞-enforcing with respect to $\ell_0^\infty(\mathcal{X})$, then a.s.*

(1) the ℓ_1^∞-enforced confidence band $[\mathbf{O}f_l, \mathbf{O}f_u]$ has weakly greater coverage than $[f_l, f_u]$:

$$1\{f_0 \in [\mathbf{O}f_l, \mathbf{O}f_u]\} \geq 1\{(f_0 \in [f_l, f_u]\}.$$

If in addition $\mathbf{O}$ is a ρ-distance contraction, then a.s.

(2) the ℓ_1^∞-enforced estimator $\mathbf{O}f$ is weakly closer to f_0 than f with respect to the distance ρ,

$$\rho(\mathbf{O}f, f_0) \leq \rho(f, f_0);$$

(3) and the ℓ_1^∞-enforced confidence band $[\mathbf{O}f_l, \mathbf{O}f_u]$ is weakly shorter than $[f_l, f_u]$ with respect to the distance ρ,

$$\rho(\mathbf{O}f_l, \mathbf{O}f_u) \leq \rho(f_l, f_u).$$

Part (1) shows that $[\mathbf{O}f_l, \mathbf{O}f_u]$ provides a coverage improvement over $[f_l, f_u]$ in that $[\mathbf{O}f_l, \mathbf{O}f_u]$ contains f_0 whenever $[f_l, f_u]$ does. Part (2) shows that the shape-enforced point estimator improves over the original estimator in terms of estimation error measured by the ρ-distance between the estimator and the target function. Parts (1) and (3) show that the shape-enforced confidence band not only has greater coverage but also is shorter with respect to the ρ-distance than the original band. These improvements apply to any sample size. In particular, they imply that enforcing the shape restriction preserves the statistical properties of the point and interval estimators. Thus, the shape-enforced estimator inherits the rate of consistency of the original estimator, and the shape-enforced confidence band has coverage at least p in large samples if the original band has coverage p in large samples. Corollary 32 can therefore be coupled with Theorems 5–30 to yield improved inference on a function that satisfies any of the shape restrictions considered in the previous section. It is also worthwhile noting that further quantifying the exact size of improvement depends on f_0 and properties of the obtained estimators f and $[f_l, f_u]$.

Remark 33 (Model Misspecification) *Let f_∞ denote the probability limit of the estimator f, provided that the limit exists. Model misspecification occurs when f_∞ is different from the target function f_0. In this case the results of Corollary 32 still apply. Moreover, if f_∞ does not satisfy the shape restriction, $f_\infty \notin \ell_1^\infty(\mathcal{X})$, then enforcing this restriction also improves estimation and inference on $\mathbf{O}f_\infty$. Thus, the probability limit of the shape-enforced estimator, $\mathbf{O}f_\infty$, is closer to f_0 in ρ-distance than f_∞, and the shape-enforced confidence band, $[\mathbf{O}f_l, \mathbf{O}f_u]$, covers $\mathbf{O}f_\infty$ with at least the same probability as $[f_l, f_u]$ covers f_∞ and $[\mathbf{O}f_l, \mathbf{O}f_u]$ is shorter than $[f_l, f_u]$ in ρ-distance.*

4. Implementation Algorithms

We provide implementation algorithms for the different shape-enforcing operators based on a sample or grid of n points $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ with corresponding values of f given by the array $\mathcal{Y}_n = \{y_1, \ldots, y_n\}$ with $y_i = f(x_i)$. Computation of the $\mathbf{R}$-operator is trivial, as it amounts to thresholding the elements of $\mathcal{Y}_n$ to be between $\underline{f}$ and $\overline{f}$, i.e.,

$$\mathbf{R}f(x_i) = (\underline{f} \vee y_i) \wedge \overline{f}.$$

When $k = 1$, Chernozhukov et al. (2009) showed that the $\mathbf{M}$-operator sorts the elements of $\mathcal{Y}_n$. Thus, assume that $x_1 \leq x_2 \leq \ldots \leq x_n$ and let $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ denote the sorted array of $\mathcal{Y}_n$. Then,

$$\mathbf{M}f(x_i) = y_{(i)}.$$

When $k > 1$, each $\mathbf{M}_j$-operator in Definition 10 can be computed by applying the same sorting procedure to the dimension j sequentially for each possible value of the other dimensions. We refer to Chernozhukov et al. (2009) for more details on computation. We next develop new algorithms for the $\mathbf{C}$ and $\mathbf{Q}$ operators.

4.1 Computation of C-Operator

When $k = 1$, we can obtain the greatest convex minorant using the standard method based on the pool adjacent violators algorithm described in Barlow et al. (1972). We provide an algorithm for the case where $k > 1$. By Definitions 6 and 7, the DFL transform of f is the solution to

$$\mathbf{C}f(x) = \sup_{\xi \in f^*(\mathcal{X})} \inf_{\tilde{x} \in \mathcal{X}} \{\xi'(x - \tilde{x}) + f(\tilde{x})\}.$$

This is a saddle point problem that might be difficult to tackle directly. However, when $\mathcal{X}$ is replaced by the finite grid $\mathcal{X}_n$, the problem has a convenient linear programming representation:

$$\mathbf{C}f(x) = \max_{v \in \mathbb{R}, \xi \in \mathbb{R}^k} \quad v \tag{4.9}$$
$$\text{s.t.} \quad v + \xi'(x_i - x) \leq f(x_i), \quad i = 1, 2, \ldots, n.$$

This program can be solved using standard linear programming methods. In particular, the computational complexity of the standard interior point method for solving (4.9) is $O((k+1)(n + (k+1))^{1.5})$, where $k + 1$ is the number of decision variables and n is the number of constraints.

The following algorithm summarizes the computation of the $\mathbf{C}$-Operator.

Algorithm 1 (C-Operator) *(1) Pick a dense enough grid of size n in $\mathcal{X}$, denoted as $\mathcal{X}_n$. One natural choice is the set of values of x observed in the data. (2) For each $x \in \mathcal{X}_n$, solve the linear programming problem stated in (4.9) to obtain $\mathbf{C}f(x)$.*

An alternative to step (2) would be to solve for all points on the grid in a single stacked linear program at the cost of increased memory due to a large constraint matrix.

4.2 Computation of Q-Operator

We propose a method to compute the $\mathbf{Q}$ operator based on solving problem (2.7) on a finite grid, namely

$$\mathbf{Q}f(x) = \min \{y \in \mathcal{Y}_n : x \in \text{conv}[\mathcal{I}_{f,n}(y)]\},$$

where $\mathcal{I}_{f,n}(y) = \{x_i : y_i = f(x_i) \leq y, i = 1, 2, \ldots, n\}$. We find the solution to the program using the following bisection search algorithm:

Algorithm 2 (Q-Operator) *For a given $x \in \mathcal{X}_n$: (1) Initialize $y_L = y_{(1)}$ and $y_U = y_{(n)}$. (2) Find the median of $\{y \in \mathcal{Y}_n : y_L \leq y \leq y_U\}$ and assign it to y^*. (3) Compute the lower contour set $\mathcal{I}_{f,n}(y^*)$. (4) If $x \in \text{conv}[\mathcal{I}_{f,n}(y^*)]$ (which indicates $y^* \geq \mathbf{Q}f(x)$), set $y_U = y^*$; otherwise, set $y_L = y^*$. (5) Repeat (2)–(4) until $y_U = y_L$ and report $\mathbf{Q}f(x) = y_U$.*

The binary search algorithm for the **Q**-Operator runs in $O(\log(n))$ iterations. The major computational cost within each iteration is the check of whether x is in the convex hull in step (4). This check does not require construction of the actual convex hull, which is computationally expensive especially in high dimensions. Instead, it is sufficient to check the existence of a feasible solution of a linear program.

We further note that each of the above two algorithms can be run in parallel across the grid points, because the output of the algorithm for one grid point does not depend on the output for any other grid point. This parallelizability allows for efficient computation on nontrivial grids.

5. Numerical Examples

5.1 Univariate Case

We consider an empirical application to growth charts and a calibrated simulation where the target function f_0 is univariate.

5.1.1 HEIGHT GROWTH CHARTS FOR INDIAN CHILDREN

Since their introduction by Quetelet in the 19th century, reference growth charts have become common tools to assess an individual's health status. These charts describe the evolution of individual anthropometric measures, such as height, weight, and body mass index, across different ages. See Cole (1988) for a classical work on the subject, and Wei et al. (2006) for a detailed analysis and additional references. Here we consider the estimation of height growth charts imposing monotonicity and concavity restrictions. These restrictions are plausible, since an individual's height is nondecreasing in age at a nonincreasing growth rate during early childhood; see, e.g., Tanner et al. (1966) and the growth child standards of the World Health Organization at `https://www.who.int/childgrowth/en/`.

We use the data from Fenske et al. (2011) and Koenker (2011) on childhood malnutrition in India. These data include a measure of height in centimeters, Y, age in months, X, and 22 covariates, Z, for 37,623 Indian children. All of the children have ages between 0 and 5 years, i.e., $X \in \mathcal{X} = \{0, 1, \ldots, 59\}$. The covariates Z include the mother's body mass index, the number of months the child was breastfed, and the mother's age (as well as the square of the previous three covariates); the mother's years of education and the father's years of education; indicator variables for the child's sex, whether the child was a single birth or part of a multiple birth, whether the mother was unemployed, whether the mother's residence is urban or rural, and whether the mother has each of: electricity, a radio, a television, a refrigerator, a bicycle, a motorcycle, and a car; and factor variables for birth order of the child, the mother's religion and quintiles of wealth.

We assume a partially linear model for the conditional expectation of Y given X and Z, namely

$$\mathrm{E}[Y \mid X = x, Z = z] = f_0(x) + z'\gamma.$$

The target function is the conditional average growth chart $x \mapsto f_0(x)$, which we assume to be nondecreasing and concave. Since X is discrete, we can express $f_0(x) = P(x)'\beta$, where $P(x)$ is a vector of indicators for each value in $\mathcal{X}$, i.e., $P(X) = [1(X = 0), 1(X =$
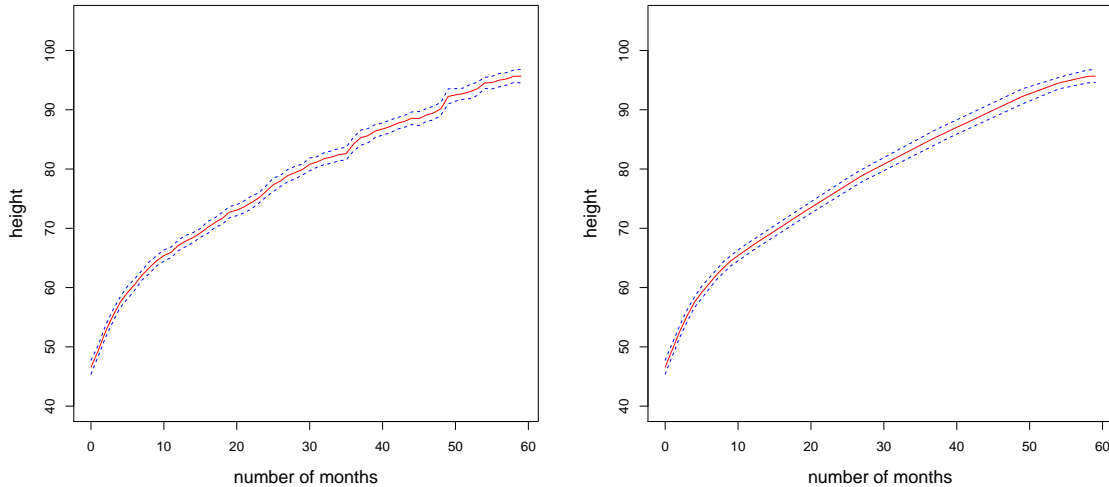
Figure 3: Entire Sample: Estimates and 95% confidence bands. Left: f and $[f_l, f_u]$. Right: $\mathbf{C}^-\mathbf{M}f$ and $[\mathbf{C}^-\mathbf{M}f_l, \mathbf{C}^-\mathbf{M}f_u]$

$1), \ldots, 1(X = 59)]'$. We estimate β and γ by least squares of Y on $P(X)$ and Z, and construct a confidence band for f_0 on $\mathcal{X}$ using weighted bootstrap with standard exponential weights and 200 repetitions Præstgaard and Wellner (1993); Hahn (1995). Standard errors are estimated using bootstrap rescaled interquartile ranges Chernozhukov et al. (2013), and the critical value is the bootstrap 0.95-quantile of the maximal t-statistic. Weighted bootstrap is computationally convenient in this application because it is less sensitive than empirical bootstrap to singular designs, which are likely to arise in the bootstrap resampling because Z and $P(X)$ contain many indicators.

Figures 3 and 4 report the point estimates and 95% confidence bands of f_0 for the entire sample and a random extract with 1,000 observations, respectively. We use the subsample to illustrate the deviations from the shape restrictions that are more apparent when the sample size is small. The original estimates are displayed in the left panels, and the estimates imposing monotonicity and concavity in the right panels. The original estimates in the entire sample are nondecreasing in age except at 45 months, and deviate from concavity in some areas. The $\mathbf{M}$ and $\mathbf{C}^-$ operators correct these deviations. The estimates in the random extract of the data clearly show deviations from both monotonicity and concavity. The $\mathbf{M}$ and $\mathbf{C}^-$ operators fix these deviations and produce point estimates that are closer to the estimates in the entire sample.

5.1.2 CALIBRATED MONTE CARLO SIMULATION

We quantify the finite-sample improvement in the point and interval estimates of enforcing shape restrictions using simulations calibrated to the growth chart application. The child's height, Y, is generated by

$$Y_i = \mathbf{C}^-\mathbf{M}[P(X_i)'\hat{\beta}] + Z_i'\hat{\gamma} + \hat{\sigma}\epsilon_i, \quad i = 1, \ldots, n,$$
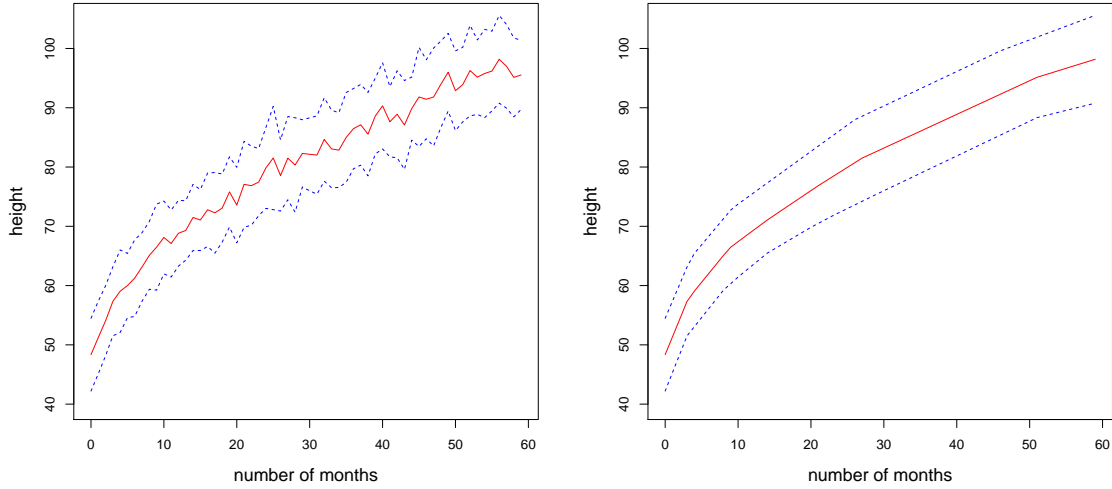
20

Figure 4: Subsample with 1,000 observations: Estimates and 95% confidence bands. Left: f and $[f_l, f_u]$. Right: $\mathbf{C}^-\mathbf{M}f$ and $[\mathbf{C}^-\mathbf{M}f_l, \mathbf{C}^-\mathbf{M}f_u]$

where $P(X_i)$ is the vector of indicators for all the values of $\mathcal{X}$; $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\sigma}$ are the least squares estimates of β, γ and the residual standard deviation in the growth chart data; and ϵ_i are independent draws from the standard normal distribution. The application of the $\mathbf{C}^-\mathbf{M}$-operator guarantees that the target function $f_0(x) = \mathbf{C}^-\mathbf{M}[P(x)'\hat{\beta}]$ is monotone and concave. We consider six sample sizes, $n \in \{500, 1{,}000, 2{,}000, 4{,}000, 8{,}000, 37{,}623\}$, where $n = 37{,}623$ is the same sample size as in the empirical application. The values of X_i and Z_i are randomly drawn from the data without replacement. The results are based on 500 simulations. In each simulation we construct point and band estimates of f_0 using the same methods as in the empirical application.

Table 1 reports simulation averages of the d_∞-distance between the estimates and target function, coverage of the target function by the confidence band and d_∞-length of the confidence band for the original and shape-enforced estimators. We consider enforcing concavity with the $\mathbf{C}^-$-operator, monotonicity with the $\mathbf{M}$-operator, and both concavity and monotonicity with the $\mathbf{C}^-\mathbf{M}$-operator. The improvements from imposing the shape restrictions are decreasing in the sample size, but there are substantial benefits in estimation error even with the largest sample size. Enforcing monotonicity has generally stronger effects than enforcing concavity, but both help improve the estimates. Thus, the $\mathbf{C}^-\mathbf{M}$-operator produces the best point and interval estimators for every sample size. For the smallest sample size, the reduction in estimation error is almost 37% and the improvement in length of the confidence band is more than 20%. The gains in coverage probability are also substantial, especially for the smaller sample sizes. Overall, the simulation results clearly showcase the benefits of enforcing shape restrictions, even with large sample sizes.

We compare the d_∞-error of several estimators in the simplified design

$$Y_i = \mathbf{C}^-\mathbf{M}[P(X_i)'\hat{\beta}] + \hat{\sigma}\epsilon_i, \quad i = 1, \ldots, n,$$

21

Table 1: Finite-Sample Properties

	$\|f - f_0\|_\infty$	$\|f_u - f_l\|_\infty$	$P(f_0 \in [f_l, f_u])$	$\|f - f_0\|_\infty$	$\|f_u - f_l\|_\infty$	$P(f_0 \in [f_l, f_u])$
		$n = 500$			$n = 1,000$	
Original	7.45	26.31	0.69	4.83	18.45	0.84
$\mathbf{C}^-$	6.28	21.77	0.79	4.16	16.95	0.89
$\mathbf{M}$	5.23	21.24	0.92	3.72	16.58	0.95
$\mathbf{C}^-\mathbf{M}$	4.73	20.15	0.93	3.30	15.97	0.95
		$n = 2,000$			$n = 4,000$	
Original	3.35	13.55	0.90	2.29	9.61	0.93
$\mathbf{C}^-$	2.93	13.21	0.94	1.98	9.53	0.96
$\mathbf{M}$	2.79	12.89	0.96	2.00	9.43	0.97
$\mathbf{C}^-\mathbf{M}$	2.48	12.64	0.97	1.76	9.34	0.98
		$n = 8,000$			$n = 37,623$	
Original	1.61	6.92	0.93	0.72	3.22	0.95
$\mathbf{C}^-$	1.40	6.90	0.97	0.64	3.21	0.97
$\mathbf{M}$	1.47	6.89	0.96	0.71	3.22	0.96
$\mathbf{C}^-\mathbf{M}$	1.30	6.87	0.98	0.63	3.21	0.98

Notes: Based on 1,000 simulations. Nominal level of the confidence bands is 95%.

Confidence bands constructed by weighted bootstrap with standard exponential weights and 200 repetitions.

where $P(X_i)$, $\hat{\beta}$, $\hat{\sigma}$, ϵ_i and n are the same as for Table 1. We consider unconstrained, shape-constrained, shape-enforced, and combinations of shape-enforced and shape-constrained estimators. The unconstrained estimators include the same estimator as in Table 1 (Piecewise Constant) and a locally linear estimator with data-driven choice of bandwidth (Locally Linear).[7] We consider two classical shape-constrained estimators: the isotonic regression estimator (Isoreg) that imposes monotonicity and the concave regression estimator (Conreg) that imposes concavity.[8] We illustrate how to combine shape-enforced operators with shape-constrained estimators by applying the $\mathbf{C}^-$-operator to the isotonic regression estimator to enforce monotonicity and concavity. Finally, we compare the $\mathbf{C}^-$-operator with the $\mathbf{SC}^-$-operator defined in Remark 9.

Table 2 shows the results based on 5,000 simulations. The comparison between shape-constrained and shape-enforced estimators produces mixed results, which vary with the

---

7. The piecewise constant estimator can be viewed as a locally constant estimator with bandwidth equal to zero. The locally linear estimator is computed using the package `KernSmooth` Wand (2019) with the bandwidth chosen by the plug-in method of Ruppert et al. (1995).

8. We compute the isotonic regression using the `R` command `isoreg` R Core Team (2019), and the concave regression using the package `cobs` Ng and Maechler (2020).

Table 2: Comparison of Estimators

	500	1,000	2,000	4,000	8,000	37,623
Piecewise Constant	5.85	3.93	2.66	1.82	1.27	0.58
$\mathbf{C}^-$ PC	5.32	3.65	2.49	1.71	1.19	0.54
$\mathbf{SC}^-$ PC	4.89	3.41	2.34	1.59	1.10	0.52
$\mathbf{M}$ PC	3.73	2.83	2.13	1.55	1.15	0.57
$\mathbf{C}^-\mathbf{M}$ PC	3.54	2.67	2.01	1.46	1.08	0.53
$\mathbf{SC}^-\mathbf{M}$ PC	3.29	2.45	1.84	1.33	0.97	0.50
Conreg	3.04	2.27	1.64	1.16	0.83	0.42
Isoreg	3.82	2.95	2.21	1.60	1.17	0.57
$\mathbf{C}^-$ Isoreg	3.52	2.75	2.08	1.50	1.10	0.53
$\mathbf{SC}^-$ Isoreg	3.28	2.53	1.90	1.37	1.00	0.50
Locally Linear	2.66	2.03	1.53	1.15	0.87	0.52
$\mathbf{C}^-$ LL	2.64	2.02	1.52	1.14	0.87	0.52
$\mathbf{SC}^-$ LL	2.61	1.99	1.51	1.13	0.87	0.53
$\mathbf{M}$ LL	2.60	1.99	1.50	1.13	0.87	0.52
$\mathbf{C}^-\mathbf{M}$-LL	2.58	1.98	1.50	1.13	0.87	0.52
$\mathbf{SC}^-\mathbf{M}$ LL	2.54	1.94	1.48	1.12	0.86	0.52

Notes: Based on 5,000 simulations. Entries are $\|f - f_0\|_\infty$.

estimator, shape restriction and sample size. Thus, the $\mathbf{M}$-operator outperforms Isoreg for both unconstrained estimators, whereas Conreg outperforms the $\mathbf{C}$-operator applied to Piecewise Constant. The unconstrained locally linear estimator outperforms Isoreg, Conreg and the shape-enforced estimators applied to Piecewise Constant for most sample sizes, despite the target function not being smooth. This finding highlights the benefit of using estimators that exploit smoothness when the sample size is not large. On the other hand, the shape-enforcing operators are more effective when applied to estimators such as Piece Constant and Isoreg that do not rely on smoothness. Shifting the $\mathbf{C}^-$-operator to deal with potential bias generally reduces estimation error for all the estimators considered.

5.2 Multivariate Case

We consider an empirical application to production functions and a calibrated simulation where the target function f_0 is bivariate.

5.2.1 PRODUCTION FUNCTIONS OF CHINESE FIRMS

The production function is a fundamental relationship in economics that maps the quantity of inputs, such as capital, labor and intermediate goods, to the quantity of output of a firm. When there are only two inputs, the law of diminishing marginal rate of technical substitution dictates that the production function of a firm is nondecreasing and quasi-concave in the inputs Hicks and Allen (1934). If in addition the industry exhibits diminishing returns to scale, then the production function is concave in the inputs. We use the data from Jacho-Chávez et al. (2010), and Horowitz and Lee (2017) to estimate the production

function of Chinese firms in the chemical industry. These data contain information on real value added (output), real fixed assets (capital) and number of employees (labor) for 1,638 firms in 2001.[9] We estimate a production function using these data and enforce the monotonicity and quasi-concavity restrictions. We provide results from enforcing concavity only for illustrative purposes because the chemical industry might exhibit increasing returns to scale at some levels of the inputs.

Figure 5 shows 3-dimensional estimates and 95% confidence bands for the average production function, together with upper contour sets for the point estimates. The estimates and bands are displayed in a region defined by the tensor product of two grids for labor and capital. Each grid includes 20 equidistant points from the 10% to the 90% sample percentiles of the corresponding variable. We obtain the unconstrained estimates from least squares with the tensor product of third-degree global polynomials as the two marginal bases for capital and labor. The confidence bands are constructed using weighted bootstrap with standard exponential weights and 500 repetitions. Standard errors are estimated using bootstrap rescaled interquartile ranges and the critical value is the bootstrap 0.95-quantile of the maximal t-statistic. Many of the upper contour sets of the unconstrained point estimates are far from being convex, and thus imply a violation of quasi-concavity. In fact, violations of monotonicity occur over a considerable area—most notably, see the positive slopes of the contour curves at low levels of labor and high levels of capital (the upper-left region of the contour plot).

The second row of Figure 5 shows the results after the $\mathbf{Q^-M}$-operator is applied to the point estimates and to each end-point function of the confidence band to ensure monotonicity and quasi-concavity. The contour curves are convex by construction, and thus satisfy the quasi-concavity restrictions. Finally, the third row of Figure 5 shows the results after the $\mathbf{C^-M}$-operator is applied to enforce monotonicity and concavity. Although quasi-concavification of a production function estimate is always reasonable, whether restriction to concavity is appropriate depends on prior knowledge of the industry.

5.2.2 CALIBRATED MONTE CARLO SIMULATION

Similar to the univariate case, we now explore the finite-sample improvements from enforcing shape restrictions via simulations calibrated to the production function application. The output, Y, of each firm is generated by

$$Y_i = \hat{\gamma} + \hat{\beta}_1 L_i + \hat{\beta}_2 K_i + \hat{\sigma}\epsilon_i, \quad i = 1, \ldots, n,$$

where $\hat{\gamma}$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}$ are calibrated to the least squares estimates and the residual standard deviation of this linear regression model in the production function data; ϵ_i are independent draws from the standard normal distribution; and n is the sample size of the simulated data. The vector (L, K) of labor and capital is drawn without replacement from the original data. The target function is

$$f_0(l, k) := \mathrm{E}[Y \mid L = l, K = k] = \hat{\gamma} + \hat{\beta}_1 l + \hat{\beta}_2 k,$$

which is increasing and concave in the capital and labor inputs because $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 > 0$. We consider five sample sizes, $n \in \{100, 200, 500, 1{,}000, 1{,}638\}$, where $n = 1{,}638$ is the same

---

9. Following Jacho-Chávez et al. (2010) and Horowitz and Lee (2017), we drop observations with a capital-to-labor ratio below the 0.025 sample quantile or above the 0.975 sample quantile.
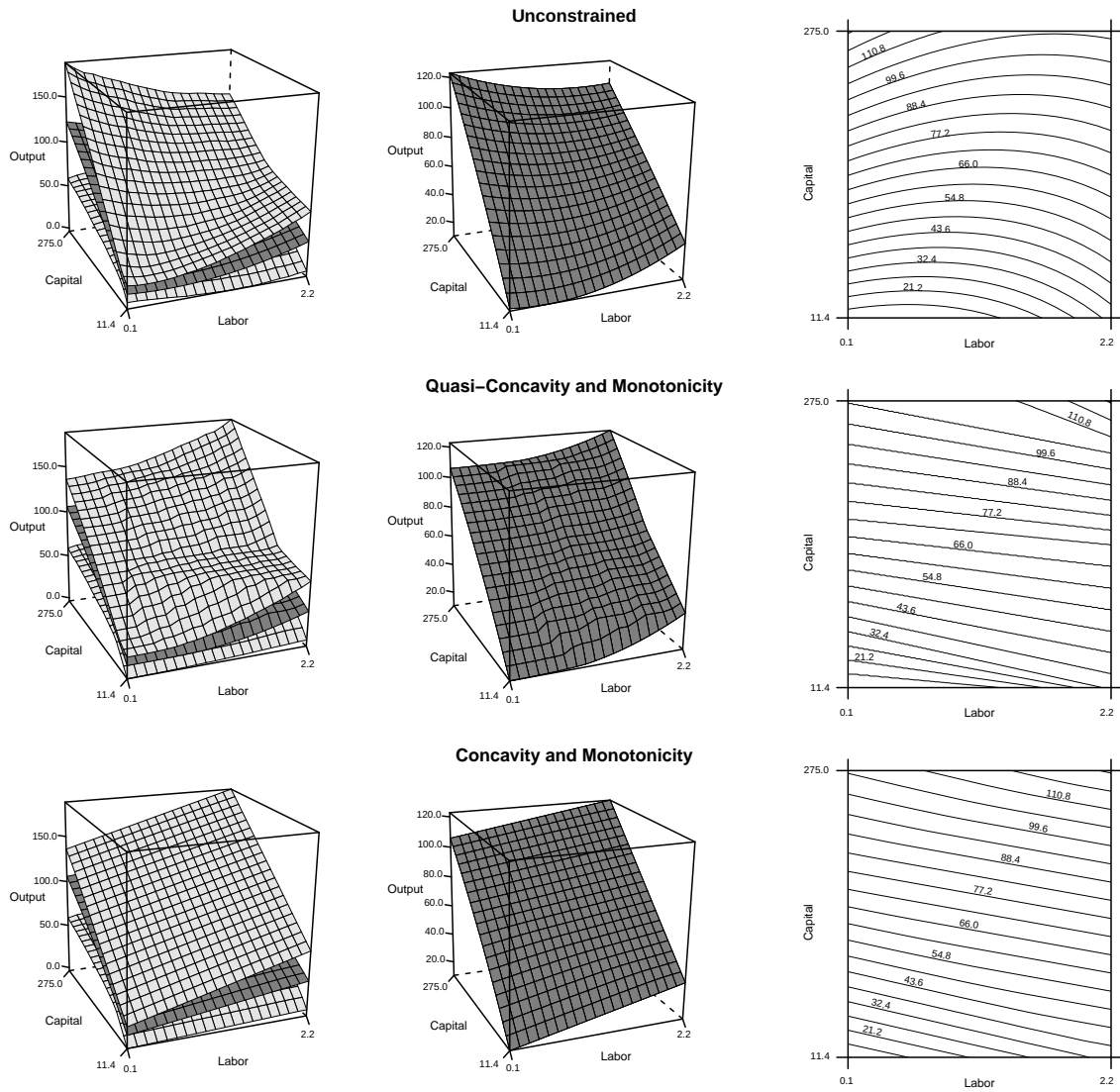
Figure 5: Confidence Bands and Countour Maps of Shape-Constrained and Unconstrained Estimates of the Production Function. Output and capital are measured in millions of 2000 yuan, and labor is measured in thousands of workers.

Table 3: Finite-Sample Properties

		n				
		100	200	500	1,000	1,638
$\mathrm{P}(f_0 \in [f_l, f_u])$	Original	0.83	0.89	0.93	0.94	0.95
	$\mathbf{Q}^-\mathbf{M}$	0.93	0.97	0.98	0.97	0.96
	$\mathbf{C}^-\mathbf{M}$	0.93	0.97	0.99	0.98	0.97
$\|f - f_0\|_\infty$	Original	297	88.9	36.5	20.4	13.2
	$\mathbf{Q}^-\mathbf{M}$	271	67.1	26.6	17.6	12.7
	$\mathbf{SQ}^-\mathbf{M}$	270	66.7	26.2	17.4	12.6
	$\mathbf{C}^-\mathbf{M}$	270	66.3	25.7	17.3	12.6
	$\mathbf{SC}^-\mathbf{M}$	250	63.0	25.2	17.3	12.6
$\|f_u - f_l\|_\infty$	Original	2297	484	190	109	72.0
	$\mathbf{Q}^-\mathbf{M}$	1441	332	135	82.8	62.0
	$\mathbf{C}^-\mathbf{M}$	1440	331	135	82.8	62.0

Notes: Based on 5,000 simulations. Nominal level of the confidence bands is 95%. Confidence bands constructed by weighted bootstrap with standard exponential weights and 500 repetitions.

sample size as in the empirical application. The results are based on 5,000 simulations. In each simulation we construct point and band estimates of f_0 using the same methods as in the empirical application.

Table 3 reports the same diagnostics as Table 1. The operators $\mathbf{Q}^-\mathbf{M}$ and $\mathbf{C}^-\mathbf{M}$ perform similarly in this case. Both bring substantial gains in estimation and inference, and the shifted variants bring additional gains. Shifting the $\mathbf{C}^-\mathbf{M}$ operator in particular has a notable effect on estimation error for small sample sizes. The operators reduce estimation error between 5% and 31% and the width of the confidence band between 14% and 37% in the sup-norm, depending on the sample size. The operators also improve the coverage of the confidence bands, especially for the smaller sample sizes. Indeed, enforcing the constraints compensates for the undercoverage of the unconstrained estimates for most of the sample sizes considered.

6. Conclusion

In this paper, we investigate a pool of shape-enforcing operators, including range, rearrangement, double Legendre-Fenchel, quasi-convexification, composition of rearrangement and double Legendre-Fenchel, and composition of rearrangement and quasi-convexification operators. We show that enforcing the shape restrictions through these operators improves point and interval estimators, and provide computational algorithms to implement these shape-enforcing operators. It would be useful to develop operators to enforce other shape restrictions, such as supermodularity or the Slutsky conditions for demand functions. We leave this extension to future research.

Acknowledgments

Appendix A. Proofs

Proof of Theorem 5

We first show that $\mathbf{R}$ satisfies the three properties of Definition 1.

(1) Reshaping: it holds because for any $f \in \ell^\infty(\mathcal{X})$, $\mathbf{R}f \in \ell_R(\mathcal{X})$ by construction.

(2) Invariance: it holds trivially because $\mathbf{R}f = f$ for any $f \in \ell_R(\mathcal{X})$ by definition of $\mathbf{R}$.

(3) Order preservation: assume that $f, g \in \ell^\infty(\mathcal{X})$ are such that $f \geq g$. For any $x \in \mathcal{X}$ there are three possible cases. (a) If $f(x) \geq \overline{f}$, then $\mathbf{R}f(x) = \overline{f} \geq \mathbf{R}g(x)$. (b) If $f(x) \leq \underline{f}$, then $g(x) \leq \underline{f}$, and $\mathbf{R}f(x) = \underline{f} = \mathbf{R}g(x)$. (c) if $\underline{f} < f(x) < \overline{f}$, then $\mathbf{R}f(x) = f(x) \geq \max(g(x), \underline{f}) = \mathbf{R}g(x)$ because $g(x) < \overline{f}$. Thus, $\mathbf{R}f(x) \geq \mathbf{R}g(x)$ for any $x \in \mathcal{X}$.

We next show Definition 2 for $\rho = d_p$ for any $p \geq 1$. For any $f, g \in \ell^\infty(\mathcal{X})$ assume without loss of generality that $f(x) \geq g(x)$ for some $x \in \mathcal{X}$. We need to show that $|\mathbf{R}f(x) - \mathbf{R}g(x)| \leq f(x) - g(x)$. There are five possible cases. (a) If $f(x) \geq g(x) \geq \overline{f}$, then $|\mathbf{R}f(x) - \mathbf{R}g(x)| = |\overline{f} - \overline{f}| = 0 \leq f(x) - g(x)$. (b) If $f(x) \geq \overline{f} \geq g(x)$, then $\mathbf{R}f(x) = \overline{f} \leq f(x)$, and $\mathbf{R}g(x) \geq g(x)$. By the order preservation property proved in (3), $0 \leq |\mathbf{R}f(x) - \mathbf{R}g(x)| \leq f(x) - g(x)$. (c) If $\overline{f} > f(x) \geq g(x) \geq \underline{f}$, then $\mathbf{R}f(x) = f(x)$ and $\mathbf{R}g(x) = g(x)$, and $|\mathbf{R}f(x) - \mathbf{R}g(x)| = f(x) - g(x)$. (d) If $\overline{f} > f(x) > \underline{f} \geq g(x)$, then $0 \leq |\mathbf{R}f(x) - \mathbf{R}g(x)| = f(x) - \underline{f} \leq f(x) - g(x)$. (e) If $\underline{f} \geq f(x) \geq g(x)$, then $|\mathbf{R}f(x) - \mathbf{R}g(x)| = |\underline{f} - \underline{f}| = 0 \leq f(x) - g(x)$. ∎

Proof of Theorem 8

Before we prove Theorem 8, we recall some useful geometric properties of the Legendre-Fenchel transform.

Lemma 34 (Properties of Legendre-Fenchel transformation) *Given a convex set $\mathcal{X} \subset \mathbb{R}^k$, suppose that $f, g \in \ell_S^\infty(\mathcal{X})$. Then:*

(1) Lower semi-continuity: $\mathbf{L}_\mathcal{X} f \in \ell_S^\infty(\mathbb{R}^k)$.

(2) Convexity: $\mathbf{L}_\mathcal{X} f$ is closed convex on $\mathbb{R}^k$.

(3) Order reversing: If $f \geq g$, then $\mathbf{L}_\mathcal{X} f \leq \mathbf{L}_\mathcal{X} g$.

(4) d_∞-Distance reducing: $\|\mathbf{L}_\mathcal{X} f - \mathbf{L}_\mathcal{X} g\|_\infty \leq \|f - g\|_\infty$.

Proof [Proof of Lemma 34] (1) For any $\xi \in f^*(\mathcal{X})$ and $\epsilon > 0$, there must exist $x_0 \in \mathcal{X}$ such that $\xi' x_0 - f(x_0) \geq \mathbf{L}_\mathcal{X} f(\xi) - \epsilon/2$. Then, for any ξ_1 such that $\|\xi - \xi_1\|_2 \leq \min[1, \epsilon/(2\|x_0\|)]$, we have:

$$\mathbf{L}_\mathcal{X} f(\xi_1) \geq \xi_1' x_0 - f(x_0) = (\xi_1 - \xi)' x_0 + \xi' x_0 - f(x_0) \geq -\|\xi - \xi_1\|_2 \|x_0\| + \mathbf{L}_\mathcal{X} f(\xi) - \frac{\epsilon}{2}$$

$$\geq -\frac{\epsilon}{2} + \mathbf{L}_\mathcal{X} f(\xi) - \frac{\epsilon}{2} \geq \mathbf{L}_\mathcal{X} f(\xi) - \epsilon.$$

Hence, $\mathbf{L}_\mathcal{X} f(\xi)$ is lower semi-continuous at ξ. Since ξ can be arbitrary, we conclude that $\mathbf{L}_\mathcal{X} f$ is a lower semi-continuous function.

Properties (2) and (3) are shown in Theorem 1.1.2 and Proposition 1.3.1 in Chapter E of Hiriart-Urruty and Lemaréchal (2001). For (4), it is easy to check that $\|\mathbf{L}_\mathcal{X} f - \mathbf{L}_\mathcal{X} g\|_\infty = \sup_{\xi \in \mathbb{R}^k} |\mathbf{L}_\mathcal{X} f(\xi) - \mathbf{L}_\mathcal{X} g(\xi)| \leq \sup_{\xi \in \mathbb{R}^k} \sup_{x \in \mathcal{X}} |\{\xi' x - f(x)\} - \{\xi' x - g(x)\}| \leq \|f - g\|_\infty$.

■

Remark 35 *The Legendre-Fenchel transform $\xi \mapsto \mathbf{L}_{\mathcal{X}} f(\xi)$ is locally Lipshitz, because any convex function is locally Lipshitz. Statement (4) of Lemma 34 is known as Marshall's Lemma Marshall (1970).*

Next, we derive some properties for the Double Legendre-Fenchel transformation.

Lemma 36 (Properties of C-Operator) *Given a convex set $\mathcal{X} \subset \mathbb{R}^k$, suppose that $f \in \ell_S^\infty(\mathcal{X})$. Then:*

(1) $\mathbf{C}f$ is the greatest convex minorant of f, i.e., the largest function $g \in \ell_C^\infty(\mathcal{X})$ such that $g \leq f$.

(2) If $\mathcal{X}$ is compact, for any $x \in \mathcal{X}$ there exist $d \leq k+2$ points $x_1, x_2, \ldots, x_d$ and scalars $(\alpha_1, \alpha_2, \ldots, \alpha_d) \in \Delta_{d-1}$, where Δ_{d-1} is the $(d-1)$-simplex

$$\Delta_{d-1} := \left\{ \alpha \in \mathbb{R}^d : \alpha_j \geq 0, j = 1, 2, \ldots, d, \sum_{j=1}^d \alpha_j = 1 \right\},$$

such that

$$\mathbf{C}f(x) = \sum_{i=1}^d \alpha_i f(x_i), \tag{A.10}$$

where $x = \sum_{i=1}^d \alpha_i x_i$ and $f(x_i) = \mathbf{C}f(x_i)$, $1 \leq i \leq d$.

(3) We say that f is convex at $x \in \mathcal{X}$ if there exists a supporting hyperplane with direction ξ such that $f(\tilde{x}) \geq f(x) + \xi'(\tilde{x} - x)$ for all $\tilde{x} \in \mathcal{X}$. Then, $f(x) = \mathbf{C}f(x)$ if and only if f is convex at x. Furthermore, if f is convex at every $x \in \mathcal{X}$, then f is a convex function.

Proof [Proof of Lemma 36] Statement (1): Recall that $\mathbf{C}f := \mathbf{L}_{f^*(\mathcal{X})} \circ \mathbf{L}_{\mathcal{X}} f$. We first show that $\mathbf{C}f \in \ell_C^\infty(\mathcal{X})$ and $\mathbf{C}f \leq f$. For any $g \in f^*(\mathcal{X})$, $\mathbf{L}_{f^*(\mathcal{X})} g$ is a closed convex function by Lemma 34(2), so that $\mathbf{C}f \in \ell_C^\infty(\mathcal{X})$. Let $f^* := \mathbf{L}_{\mathcal{X}} f$. For any $x \in \mathcal{X}$, $\mathbf{C}f(x) = \mathbf{L}_{f^*(\mathcal{X})} f^*(x) = \sup_{\xi \in f^*(\mathcal{X})} \{\xi' x - f^*(\xi)\} \leq \sup_{\xi \in f^*(\mathcal{X})} (\xi' x - (\xi' x - f(x))) = f(x)$ because $f^*(\xi) \geq \xi' x - f(x)$ for any $\xi \in f^*(\mathcal{X})$ by definition of f^*.

Next, we show that $\mathbf{C}f$ is the convex minorant of f, i.e. $\mathbf{C}f \geq h$ for any $h \in \ell_C^\infty(\mathcal{X})$ such that $h \leq f$. If $h \in \ell_C^\infty(\mathcal{X})$, for any $x \in \mathcal{X}$, there exists $\xi \in f^*(\mathcal{X})$ such that $h(\tilde{x}) \geq h(x) + \xi'(\tilde{x} - x)$ for all $\tilde{x} \in \mathcal{X}$. Since $h(\tilde{x}) \leq f(\tilde{x})$,

$$\xi' x - h(x) \geq \xi' \tilde{x} - h(\tilde{x}) \geq \xi' \tilde{x} - f(\tilde{x}). \tag{A.11}$$

By definition, $f^*(\xi) = \sup_{\tilde{x} \in \mathcal{X}} \{\xi' \tilde{x} - f(\tilde{x})\}$, so that for any $\epsilon > 0$, there must exist $\tilde{x} \in \mathcal{X}$ such that $\xi' \tilde{x} - f(\tilde{x}) \geq f^*(\xi) - \epsilon$, which combined with (A.11) gives $\xi' x - h(x) \geq f^*(\xi) - \epsilon$, or rearranging terms, $\xi' x - f^*(\xi) \geq h(x) - \epsilon$. Then, $\mathbf{C}f(x) = \sup_{\tilde{\xi} \in f^*(\mathcal{X})} \{\tilde{\xi}' x - f^*(\tilde{\xi})\} \geq \xi' x - f^*(\xi) \geq h(x) - \epsilon$. Since ϵ can be arbitrarily small, we conclude that $\mathbf{C}f(x) \geq h(x)$.

29

Statement (2): by Proposition 2.5.1. in Chapter B of Hiriart-Urruty and Lemaréchal (2001),

$$\mathbf{C}f(x) = \inf\left\{\sum_{j=1}^{k+2}\alpha_j f(x_j) : \sum_{j=1}^{k+2}\alpha_j x_j = x, \alpha = (\alpha_1,\ldots,\alpha_{k+2}) \in \Delta_{k+1}\right\}, \qquad \text{(A.12)}$$

where Δ_{k+1} is the $(k+1)$-simplex.

By (A.12), there exists a sequence $(x^t, \alpha^t) \in \mathcal{X} \times \Delta_k$ such that $\sum_{j=1}^{k+2}\alpha_j^t x_j^t = x$ and $\mathbf{C}f(x) \leq \sum_{j=1}^{k+2}\alpha_j^t f(x_j^t) + \frac{1}{t}$. Since $\mathcal{X} \times \Delta_{k+1}$ is compact, there must exist a limit point $(x^0, \alpha^0) \in \mathcal{X} \times \Delta_{k+1}$ of the sequence (x^t, α^t) such that $\sum_{j=1}^{k+2}\alpha_j^0 x_j^0 = \lim_{t\to\infty}\sum_{j=1}^{k+2}\alpha_j^t x_j^t = x$, and by lower semi-continuity of f, $\sum_{j=1}^{k+2}\alpha_j^0 f(x_j^0) \leq \lim_{t\to\infty}\sum_{j=1}^{k+2}\alpha_j^t f(x_j^t) = \mathbf{C}f(x)$. Then, it follows from (A.12) that $\sum_{j=1}^{k+2}\alpha_j^0 f(x_j^0) = \mathbf{C}f(x)$. Equivalently, $\sum_{j=1}^{k+2}\alpha_j^0 x_j^0 = x$, and

$$\sum_{j=1}^{k+2}\alpha_j^0 f(x_j^0) = \mathbf{C}f(x). \qquad \text{(A.13)}$$

Let $(\alpha_1,\ldots,\alpha_d)$ and $(x_1,\ldots,x_d)$ denote the subsets of $(\alpha_1^0,\ldots,\alpha_{k+2}^0)$ and $(x_1^0,\ldots,x_{k+2}^0)$ corresponding to the components with $\alpha_j^0 > 0$, where $j = 1, 2, \ldots, k+2$ and $d \leq k+2$. Next, we show that $f(x_j) = \mathbf{C}f(x_j)$, for $1 \leq j \leq d$. By statement (1), since $\mathbf{C}f$ is the convex minorant of f, it follows that $x \mapsto \mathbf{C}f(x)$ is convex and $\mathbf{C}f(x) \leq f(x)$ for all $x \in \mathcal{X}$. In particular,

$$\mathbf{C}f(x) \leq \sum_{j=1}^{d}\alpha_j \mathbf{C}f(x_j) \leq \sum_{j=1}^{d}\alpha_j f(x_j).$$

By (A.13), the two inequalities imply that

$$\sum_{j=1}^{d}\alpha_j \mathbf{C}f(x_j) = \sum_{j=1}^{d}\alpha_j f(x_j).$$

Since $\mathbf{C}f(x_j) \leq f(x_j)$ for all $j = 1, 2, \ldots, d$, it follows that $f(x_j) = \mathbf{C}f(x_j)$ for all $j = 1, 2, \ldots, d$.

This completes the proof of statement (2).

Statement (3): $\mathbf{C}f(x) \leq f(x)$ for all $x \in \mathcal{X}$ by (1). If there exists a $\xi \in \mathbb{R}^k$ such that $f(\tilde{x}) \geq f(x) + \xi'(\tilde{x} - x)$ for any $\tilde{x} \in \mathcal{X}$, then $g(\tilde{x}) = f(x) + \xi'(\tilde{x} - x)$ is a convex function that lies below f. By (1), $\mathbf{C}f(x) \geq g(x) = f(x)$. Therefore, $\mathbf{C}f(x) = f(x)$. On the other hand, suppose that $\mathbf{C}f(x) = f(x)$. Since $x \mapsto \mathbf{C}f(x)$ is convex on $\mathcal{X}$, there must exist $\xi \in \mathbb{R}^k$ such that $\mathbf{C}f(\tilde{x}) \geq \mathbf{C}f(x) + \xi'(\tilde{x} - x)$ for any $\tilde{x} \in \mathcal{X}$. By definition of greatest convex minorant, $f(\tilde{x}) \geq \mathbf{C}f(\tilde{x}) \geq \mathbf{C}f(x) + \xi'(\tilde{x} - x) = f(x) + \xi'(\tilde{x} - x)$ for any $\tilde{x} \in \mathcal{X}$. So $\mathbf{C}f(x) = f(x)$ implies that f is convex at x.

If f is convex at every $x \in \mathcal{X}$, then by the results above, $f(x) = \mathbf{C}f(x)$ for every $x \in \mathcal{X}$. That is, $f = \mathbf{C}f$, which implies that f is convex on $\mathcal{X}$ because $\mathbf{C}f$ is convex on $\mathcal{X}$. ∎

Theorem 8 follows from the properties in Lemmas 34 and 36. The properties (1) and (3) in Definition 1 are implied by properties (2) and (3) of Lemma 34 applied to $\mathbf{L}_\mathcal{X}f$ and

using that $\mathbf{L}_{\mathcal{X}} f \in \ell_S^\infty(\mathbb{R}^k)$ by property (1) of Lemma 34. The property (2) in Definition 1 is implied by property (3) of Lemma 36. Moreover, the d_∞-contraction property is given by property (4) in Lemma 34 again applied to $\mathbf{L}_{\mathcal{X}} f$ and using that $\mathbf{L}_{\mathcal{X}} f \in \ell_S^\infty(\mathbb{R}^k)$ by property (1) of Lemma 34. ■

Proof of Theorem 16

We start by demonstrating that the **C**-operator on a rectangle can be computed separately at each face of the rectangle.

Definition 37 (C-Operator Restricted to a Face of a Rectangle) *For any regular rectangular set $\mathcal{X} := [a_1, b_1] \times \ldots \times [a_k, b_k]$, a set $\mathcal{F}_m$ is an m-dimensional face of $\mathcal{X}$ if there exists a set of indexes $i(\mathcal{F}) \subset \{1, 2, \ldots, k\}$ with m elements such that $\mathcal{F}_m = \{x \in \mathcal{X} : x_j \in [a_j, b_j], \text{ for any } j \in i(\mathcal{F}), x_j \in \{a_j, b_j\}, \text{ for any } j \notin i(\mathcal{F})\}$. For every $x \in \mathcal{F}_m$, we can define the **C**-operator restricted to the face $\mathcal{F}_m$ by applying the Legendre-Fenchel transform only to each of the coordinates of x that are in $i(\mathcal{F}_m)$. Thus, let*

$$\mathbf{L}_{\mathcal{X}|\mathcal{F}_m} f(\xi) := \sup_{x \in \mathcal{F}_m} \{\xi' x_{i(\mathcal{F}_m)} - f(x_{i(\mathcal{F}_m)}, x_{i^c(\mathcal{F}_m)})\},$$

*where we partition x into the coordinates with indexes in $i(\mathcal{F}_m)$, $x_{i(\mathcal{F}_m)}$, and the rest of the coordinates, $i^c(\mathcal{F}_m)$. Then, the **C**-operator restricted to the face $\mathcal{F}_m$ of $f \in \ell_S^\infty(\mathcal{X})$ is*

$$\mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x) := \mathbf{L}_{f^*(\mathcal{X}|\mathcal{F}_m)} \circ \mathbf{L}_{\mathcal{X}|\mathcal{F}} f(x),$$

where $f^*(\mathcal{X} \mid \mathcal{F}_m) := \{\xi \in \mathbb{R}^m : \mathbf{L}_{\mathcal{X}|\mathcal{F}_m} f(\xi) < \infty\}$. Moreover, by Proposition 2.5.1 of Hiriart-Urruty and Lemaréchal (2001), $\mathbf{C}_{\mathcal{X}|\mathcal{F}} f(x)$ is a linear combination of the f-images of $m + 2$ elements of $\mathcal{F}_m$, that is

$$\mathbf{C}_{\mathcal{X}|\mathcal{F}} f(x) = \inf \left\{ \sum_{j=1}^{m+2} \alpha_j f(x_j) : x_j \in \mathcal{F}_m, (\alpha_1, \ldots, \alpha_{m+2}) \in \Delta_{m+1} \right\}, \quad \text{for any } x \in \mathcal{F}_m,$$

where Δ_{m+1} is the $(m + 1)$-simplex.

Lemma 38 (C-Operator on a Regular Rectangular Set) *For any regular rectangular set $\mathcal{X}$ and $f \in \ell_S^\infty(\mathcal{X})$, if $x \in \mathcal{F}_m$ with $m > 0$, then*

$$\mathbf{C} f(x) = \mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x).$$

Proof [Proof of Lemma 38] Suppose that $\mathcal{X}$ is a regular rectangle in $\mathbb{R}^k$. Let $\mathcal{F}_m$ be a face of $\mathcal{X}$ with dimension m such that $x \in \mathcal{F}_m$. The result follows from the following facts:

First, $x \mapsto \mathbf{C} f(x)$ is a convex function and lies below $x \mapsto f(x)$ on $\mathcal{X}$, so that $x \mapsto \mathbf{C} f(x)$ is a convex function and lies below $x \mapsto f(x)$ on $\mathcal{F}_m \subset \mathcal{X}$. By definition, $\mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f$ is the convex minorant of f restricted on $\mathcal{F}_m$, i.e., the largest possible convex function lying below f restricted on $\mathcal{F}_m$. Therefore, it must be that $\mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f \geq \mathbf{C} f(x)$ for all $x \in \mathcal{F}_m$.

Second, by statement (2) of Lemma 36, for any $x \in \mathcal{F}_m$, there exist $d \leq k + 2$ points $x_1, \ldots, x_d$ and $\alpha_i > 0$, $1 \leq i \leq d$, $\sum_{i=1}^d \alpha_i = 1$, such that $\mathbf{C} f(x_i) = f(x_i), \sum_{i=1}^d \alpha_i x_i = x$, and $\sum_{i=1}^d \alpha_i f(x_i) = f(x)$. It must be that $x_i \in \mathcal{F}_m$, $1 \leq i \leq d$, since $x \in \mathcal{F}_m$.

Third, by definition of greatest convex minorant, on the face $\mathcal{F}_m$, $f(x) \geq \mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x)$ for any $x \in \mathcal{F}_m$. Since $\mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x)$ is the convex minorant of $f(x)$ restricted on $\mathcal{F}_m$, and $\mathbf{C}f$ is a convex function on $\mathcal{F}_m$, it follows that $\mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x) \geq \mathbf{C}f(x)$ for any $x \in \mathcal{F}_m$. Therefore,

$$f(x) \geq \mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x) \geq \mathbf{C}f(x) \tag{A.14}$$

for all $x \in \mathcal{F}_m$.

Fourth, for each x_i, $i = 1, 2, \ldots, d$, we know that $f(x_i) = \mathbf{C}f(x_i)$. Applying equation (A.14), it must be that $f(x_i) = \mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x_i) = \mathbf{C}f(x_i)$. Therefore,

$$\mathbf{C}f(x) = \sum_{i=1}^{d} \alpha_i f(x_i) = \sum_{i=1}^{d} \alpha_i \mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x_i) \geq \mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x), \tag{A.15}$$

where the inequality follows from convexity of $x \mapsto \mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x)$.

Combining inequalities (A.14) and (A.15), we conclude that $\mathbf{C}_{\mathcal{X}|\mathcal{F}_m} f(x) = \mathbf{C}f(x)$. ∎

Before stating the main proof of Theorem 16, we require a lemma to show that $\mathbf{M}$ maps a function in $\ell_S^\infty(\mathcal{X})$ to $\ell_S^\infty(\mathcal{X})$.

Lemma 39 *Suppose $\mathcal{X} = [0,1]^k$. The rearrangement operator maps any function $f \in \ell_S^\infty(\mathcal{X})$ to $\ell_S^\infty(\mathcal{X})$.*

Proof First, it is easy to see that for any $f_1, f_2 \in \ell_S^\infty(\mathcal{X})$ and $a, b \geq 0$, $af_1 + bf_2 \in \ell_S^\infty(\mathcal{X})$. Therefore, to show that $\mathbf{M}$ maps a function $f \in \ell_S^\infty(\mathcal{X})$ to $\ell_S^\infty(\mathcal{X})$, it suffices to show that $\mathbf{M}_\pi$ maps a function $f \in \ell_S^\infty(\mathcal{X})$ to $\ell_S^\infty(\mathcal{X})$, since $\mathbf{M}f = \sum_{\pi \in \Pi} \mathbf{M}_\pi f / |\Pi|$. Denote $\pi = (\pi_1, \ldots, \pi_k)$, so $\mathbf{M}_\pi = \mathbf{M}_{\pi_1} \circ \cdots \circ \mathbf{M}_{\pi_k}$. For any function $f \in \ell_S^\infty(\mathcal{X})$ and $j = 1, 2, \ldots, k$, we would like to prove that $\mathbf{M}_j f \in \ell_S^\infty(\mathcal{X})$. If the statement above is true, then it follows that $\mathbf{M}_\pi f = \mathbf{M}_{\pi_1} \circ \cdots \circ \mathbf{M}_{\pi_k} f \in \ell_S^\infty(\mathcal{X})$. Consequently, the conclusion of the lemma is true.

Second, we prove that for any function $f \in \ell_S^\infty(\mathcal{X})$ and $j = 1, 2, \ldots, k$, $\mathbf{M}_j f \in \ell^\infty(\mathcal{X})$. Without loss of generality, we can assume $j = 1$. By definition,

$$\mathbf{M}_1 f(x(1), x(-1)) = \inf \left\{ y \in \mathbb{R} : \int_{\mathcal{X}(1)} 1\{f(t, x(-1)) \leq y\} dt \geq x(1) \right\}.$$

For any $x(1) \in [0,1]$ and $x(-1) \in [0,1]^{k-1}$, $\int_{\mathcal{X}(1)} 1\{f(t, x(-1)) \leq y_{max}\} dt = 1 \geq x(1)$, and $\int_{\mathcal{X}(1)} 1\{f(t, x(-1)) \leq (y_{min} - \epsilon)\} dt = 0 < x(1)$, where $y_{max} = \sup_{x \in \mathcal{X}} f(x)$, $y_{min} = \inf_{x \in \mathcal{X}} f(x)$ and $\epsilon > 0$ can be any arbitrarily small constant. Since $f \in \ell_S^\infty(\mathcal{X})$, y_{min} and y_{max} exist. Therefore, $\inf \left\{ y \in \mathbb{R} : \int_{\mathcal{X}(1)} 1\{f(t, x(-1)) \leq y\} dt \geq x(1) \right\}$ must be well defined and bounded by y_{max} from above and by $y_{min} - \epsilon$ from below. We conclude that $\mathbf{M}_1 f \in \ell^\infty(\mathcal{X})$.

Third, we show that $\mathbf{M}_1 f \in \ell_S^\infty(\mathcal{X})$ if $f \in \ell_S^\infty(\mathcal{X})$. We prove this by contradiction: suppose that $\mathbf{M}_1 f$ is not lower semi-continuous at a point $x^0 \in \mathcal{X}$. There must exist a sequence $x^1, \ldots, x^n, \ldots$ in $\mathcal{X}$ and a constant $\epsilon > 0$ such that $\|x^n - x^0\| \to 0$ as $n \to \infty$ and $\mathbf{M}_1 f(x^n) \leq \mathbf{M}_1 f(x^0) - \epsilon$ for all $n \geq 1$. Let $y^0 := \mathbf{M}_1 f(x^0)$. By definition of $\mathbf{M}_1 f$, it must be that $\int_{\mathcal{X}(1)} 1\{f(t, x^0(-1)) \leq y^0 - \frac{\epsilon}{2}\} dt < x^0(1)$, and $\int_{\mathcal{X}(1)} 1\{f(t, x^n(-1)) \leq y^0 - \epsilon\} dt \geq$

$x^n(1)$ for all $n \geq 1$. For any $t \in [0,1]$, since $f \in \ell_S^\infty(\mathcal{X})$ and $(t, x^n(-1)) \to (t, x^0(-1))$, $\lim_{n\to\infty} f(t, x^n(-1)) \geq f(t, x^0(-1))$. Therefore, there exists N large enough such that $f(t, x^n(-1)) \geq f(t, x^0(-1)) - \frac{\epsilon}{2}$ for all $n \geq N$. Consequently, $1\{f(t, x^n(-1)) \leq y^0 - \epsilon\} \leq 1\{f(t, x^0(-1)) \leq y^0 - \frac{\epsilon}{2}\}$ for all $n \geq N$. Then,

$$\limsup_{n\to\infty} 1\{f(t, x^n(-1)) \leq y^0 - \epsilon\} \leq 1\{f(t, x^0(-1)) \leq y^0 - \frac{\epsilon}{2}\}$$

holds for all t. By reverse Fatou's Lemma,

$$\limsup_{n\to\infty} \int_{\mathcal{X}(1)} 1\{f(t, x^n(-1)) \leq y^0 - \epsilon\} dt \leq \int_{\mathcal{X}(1)} \limsup_{n\to\infty} 1\{f(t, x^n(-1)) \leq y^0 - \epsilon\} dt$$

$$\leq \int_{\mathcal{X}(1)} 1\{f(t, x^0(-1)) \leq y^0 - \frac{\epsilon}{2}\} dt. \quad \text{(A.16)}$$

However, $\int_{\mathcal{X}(1)} 1\{f(t, x^0(-1)) \leq y^0 - \frac{\epsilon}{2}\} dt < x(1)^0$, while $\limsup_{n\to\infty} \int_{\mathcal{X}(1)} 1\{f(t, x^n(-1)) \leq y^0 - \epsilon\} dt \geq \limsup_{n\to\infty} x^n(1) = x_1^0$. Hence,

$$\limsup_{n\to\infty} \int_{\mathcal{X}(1)} 1\{f(t, x^n(-1)) \leq y^0 - \epsilon\} dt \geq x_1^0 > \int_{\mathcal{X}(1)} 1\{f(t, x^0(-1)) \leq y^0 - \frac{\epsilon}{2}\} dt,$$

which contradicts (A.16). Therefore, we conclude that $\mathbf{M}_1 f \in \ell_S^\infty(\mathcal{X})$ if $f \in \ell_S^\infty(\mathcal{X})$. ■

We now start the proof of Theorem 16.

(1) We first show that $\mathbf{CM}$ satisfies the reshaping property (1) of Definition 1.

We know that $\mathbf{CM}f = \mathbf{C}(\mathbf{M}f)$. By Lemma 39, for any $f \in \ell_S^\infty(\mathcal{X})$, $\mathbf{M}f \in \ell_S^\infty(\mathcal{X})$. Consequently, $\mathbf{M}f \in \ell_S^\infty(\mathcal{X}) \cap \ell_M^\infty(\mathcal{X})$.

We use induction to prove that $\mathbf{C}f \in \ell_{CM}^\infty(\mathcal{X})$ for any $f \in \ell_S^\infty(\mathcal{X}) \cap \ell_M^\infty(\mathcal{X})$, where $\mathcal{X} \subset \mathbb{R}^k$ is a regular rectangular set. Without loss of generality, assume that $\mathcal{X} = [0,1]^k$. Since $\mathbf{C}f \in \ell_C^\infty(\mathcal{X})$ by Theorem 8, we only need to show that $\mathbf{C}f \in \ell_M^\infty(\mathcal{X})$.

For dimension $k = 1$, $\mathcal{X}$ is a closed interval. We prove that $\mathbf{C}f$ is nondecreasing. Assume, by contradiction, that there exists a pair of points $x, x' \in \mathcal{X}$ such that $x < x'$ and $\mathbf{C}f(x) > \mathbf{C}f(x')$. Let $\underline{x}$ be the left end point of the interval $\mathcal{X}$. By convexity, $\mathbf{C}f(\underline{x}) \geq \mathbf{C}f(x) > \mathbf{C}f(x')$. By Lemma 38, $\mathbf{C}f(\underline{x}) = \mathbf{C}_{\mathcal{X}|\mathcal{F}_0} f(\underline{x}) = f(\underline{x})$. By statement (2) of Lemma 36, there exist $x_1, \ldots, x_d \in \mathcal{X}$ and $\alpha_1, \ldots, \alpha_d > 0, \sum_{j=1}^d \alpha_j = 1$ such that $\mathbf{C}f(x') = \mathbf{C}f(x') = \sum_{j=1}^d \alpha_j f(x_j)$. Since f is nondecreasing, we have $\sum_{j=1}^d \alpha_j f(x_j) \geq \sum_{j=1}^d \alpha_j f(\underline{x}) = f(\underline{x})$, which contradicts that $\mathbf{C}f(\underline{x}) > \mathbf{C}f(x')$. Hence, for any $x < x'$, it must be that $\mathbf{C}f(x) \leq \mathbf{C}f(x')$. We conclude that $x \mapsto \mathbf{C}f(x)$ is nondecreasing.

Suppose that $x \mapsto \mathbf{C}f(x)$ is nondecreasing for $(k-1)$-dimensional regular rectangles, $k \geq 2$. Let $\mathcal{X}$ be a k-dimensional rectangle. Assume, by contradiction, that there exists $x \leq x'$ $(x \neq x')$ such that $\mathbf{C}f(x) > \mathbf{C}f(x')$. Consider the radial originated from x' that passes through x, denoted as L. L can be written as $\{z \in \mathbb{R}^k : z = \gamma x' + (1-\gamma)x, \gamma \leq 1\}$. Therefore, there exists a $\gamma_0 \leq 0$ such that $\gamma x' + (1-\gamma)x \in \mathcal{X} \cap L$ if and only if $1 \geq \gamma \geq \gamma_0$. Denote $\underline{l} = \gamma_0 x' + (1 - \gamma_0)x$. By convexity of $x \mapsto \mathbf{C}f(x)$, it must be that

$$\mathbf{C}f(\underline{l}) \geq \mathbf{C}f(x) > \mathbf{C}f(x'). \quad \text{(A.17)}$$

33

By statement (2) of Lemma 36, there are d points $x_1, \ldots, x_d \in \mathcal{X}$ and $\alpha_1, \ldots, \alpha_d > 0$, $\sum_{i=1}^{d} \alpha_i = 1$, such that $\sum_{j=1}^{d} \alpha_j x_j = x'$ and $\sum_{j=1}^{d} \alpha_j f(x_j) = \mathbf{C}f(x')$. The point $\underline{l}$ must be on a $k-1$ dimensional face of $\mathcal{X}$, denoted by $\mathcal{F}_{k-1}$. Since $\mathcal{X} = [0,1]^k$, $\mathcal{F}_{k-1}$ can be expressed as $A_1 \times A_2 \times \ldots \times A_k$, where $A_i = [0,1]$ for $i \in i(\mathcal{F}_{k-1})$, $i(\mathcal{F}_{k-1}) \subset \{1, 2, \ldots, k\}$, $|i(\mathcal{F}_{k-1})| = k-1$, and $A_i = \{0\}$ or $\{1\}$ if $i \notin i(\mathcal{F}_{k-1})$. Without loss of generality, we can assume that $i(\mathcal{F}_{k-1}) = \{1, 2, \ldots, k-1\}$. Denote $A_k = \{w\}$, so $w = 0$ or 1.

Let s be the projection mapping from $\mathcal{X} = [0,1]^k$ to $\mathcal{F}_{k-1}$, so $s : (x(1), \ldots, x(k)) \mapsto (x(1), \ldots, x(k-1), w)$ for any $x(1), \ldots, x(k) \in [0,1]$. Since $\underline{l} \in \mathcal{F}_{k-1}$, it must be that $s(\underline{l}) = \underline{l}$. If $w = 0$, $s(x) \leq x$ for any $x \in \mathcal{X}$. Therefore, $s(x_1) \leq x_1, \ldots, s(x_d) \leq x_d$. Then, since $x \mapsto f(x)$ is nondecreasing, $f(x_i) \geq f(s(x_i))$ for all $i = 1, 2, \ldots, d$. If $w = 1$, then any point $x \in \mathcal{F}_{k-1}$ satisfies $x(k) = 1$, including $\underline{l}$. Since $x' \geq \underline{l}$, the kth entry of x' must equal to 1. By $\sum_{j=1}^{d} \alpha_j x_j = x'$, $x_j \in [0,1]^k$, it must be that the kth entry of x_j equals to 1 for all $i = 1, 2, \ldots, d$. Therefore, $s(x_j) = x_j$, and $s(x') = x'$. Therefore, regardless of the value of w, $x_j \geq s(x_j)$, $x' \geq s(x')$. Since $x' \geq \underline{l}$, it must be that $\underline{l} \leq s(x') = \sum_{j=1}^{d} \alpha_j s(x_j)$. By Lemma 9, $\mathbf{C}_{\mathcal{X}|\mathcal{F}} f(\underline{l}) = \mathbf{C}f(\underline{l})$ and by (A.17),

$$\mathbf{C}_{\mathcal{X}|\mathcal{F}} f(\underline{l}) = \mathbf{C}f(\underline{l}) > \mathbf{C}f(x') = \sum_{i=1}^{d} \alpha_i f(x_i) \geq \sum_{i=1}^{d} \alpha_i f(s(x_i))$$

$$\geq \sum_{i=1}^{d} \alpha_i \mathbf{C}_{\mathcal{X}|\mathcal{F}} f(s(x_i)) \geq \mathbf{C}_{\mathcal{X}|\mathcal{F}} f(s(x')),$$

where the second inequality holds by monotonicity of $x \mapsto f(x)$, the third inequality by $\mathbf{C}_{\mathcal{X}|\mathcal{F}}$ being the convex minorant of f, and the fourth by convexity of $x \mapsto \mathbf{C}_{\mathcal{X}|\mathcal{F}} f(x)$. Therefore,

$$\mathbf{C}_{\mathcal{X}|\mathcal{F}} f(\underline{l}) > \mathbf{C}_{\mathcal{X}|\mathcal{F}} f(s(x')). \tag{A.18}$$

By induction, $\mathbf{C}_{\mathcal{X}|\mathcal{F}} f$ restricted on the $k-1$ dimensional regular rectangle $\mathcal{F}$ is nondecreasing. Since $s(x') \geq \underline{l}$, it must be that $\mathbf{C}_{\mathcal{X}|\mathcal{F}} f(s(x')) \geq \mathbf{C}_{\mathcal{X}|\mathcal{F}} f(\underline{l})$, which contradicts (A.18). Hence, the induction is complete. $x \mapsto \mathbf{C}f(x)$ is nondecreasing if $x \mapsto f(x)$ is nondecreasing. Therefore, for any $f \in \ell_S^\infty(\mathcal{X})$, $\mathbf{CM}f$ is monotonically increasing.

We next show that $\mathbf{CM}$ satisfies the rest of the properties of Definition 1 and distance reduction.

(2) To show invariance, note that if $f \in \ell_{CM}^\infty(\mathcal{X})$, then $\mathbf{M}f = f$ by Theorem 11, and therefore $\mathbf{CM}f = \mathbf{C}(\mathbf{M}f) = \mathbf{C}f = f$ by definition of $\mathbf{CM}$ and Theorem 8.

(3) Similarly, $\mathbf{CM}$ is order preserving because if $f \geq g$ then $\mathbf{M}f \geq \mathbf{M}g$ by Theorem 11, and therefore $\mathbf{CM}f = \mathbf{C}(\mathbf{M}f) \geq \mathbf{C}(\mathbf{M}g) = \mathbf{CM}g$ by definition of $\mathbf{CM}$ and Theorem 8.

(4) Finally, $\mathbf{CM}$ is a $d_\infty-$distance contraction because

$$d_\infty(\mathbf{CM}f, \mathbf{CM}g) = d_\infty(\mathbf{C}[\mathbf{M}f], \mathbf{C}[\mathbf{M}g]) \leq d_\infty(\mathbf{M}f, \mathbf{M}g) \leq d_\infty(f, g),$$

where the first equality follows from definition of $\mathbf{CM}$, the first inequality by Theorem 8, and the second inequality by Theorem 11. ∎

Proof of Theorem 21

We start with a lemma establishing that the operator $\mathbf{Q}$ is well-defined.

Lemma 40 (Properties of Operator Q) *For any convex and compact set $\mathcal{X} \subset \mathbb{R}^k$, the operator $\mathbf{Q}$ defined in (2.7) is well-defined in that the minimum of the set $\{y \in \mathbb{R} : x \in \mathrm{conv}[\mathcal{I}_f(y)]\}$ exists for all $x \in \mathcal{X}$ and $\mathbf{Q}f \in \ell_S^\infty(\mathcal{X})$ for any $f \in \ell_S^\infty(\mathcal{X})$.*

Proof Define $\mathcal{Q}_f(x) := \{y \in \mathbb{R} : x \in \mathrm{conv}[\mathcal{I}_f(y)]\}$. We first show that $\min \mathcal{Q}_f(x)$ exists. Obviously, $\inf \mathcal{Q}_f(x) \geq f(x) > -\infty$, because $\mathcal{I}_f(y) \subset \mathrm{conv}[\mathcal{I}_f(y)]$ and $x \in \mathcal{I}_f(f(x))$. Hence, $\mathcal{Q}_f(x)$ is bounded from below. Let $y^0 := \inf \mathcal{Q}_f(x)$. We need to show that $\min \mathcal{Q}_f(x) = y^0$, i.e, there exists a sequence of $y^1, \ldots, y^i, \ldots$ such that $y^i \in \mathcal{Q}_f(x)$, and $y^i \to y^0$ as $i \to \infty$. For each y^i, $x \in \mathrm{conv}[\mathcal{I}_f(y^i)]$. Hence, by Carathéodory's theorem, there exist $d_x \leq k+1$ points $X^i := (x_1^i, \ldots, x_{d_x}^i)$ where $x_j^i \in \mathcal{I}_f(y^i), j = 1, 2, \ldots, d_x$, and $\alpha^i := (\alpha_1^i, \ldots, \alpha_{d_x}^i) \in \Delta_{d_x-1}$ where Δ_{d_x-1} is the (d_x-1)-simplex, such that $x = \sum_{j=1}^{d_x} \alpha_j^i x_j^i$. Since $\mathcal{X}$ and Δ_{d_x-1} are both compact, there must exist a subsequence of (X^i, α^i) that converges to a limit point (X^0, α^0) where $X^0 = (x_1^0, \ldots, x_{d_x}^0)$, $x_j^0 \in \mathcal{X}$, $j = 1 \ldots, d_x$, and $\alpha^0 \in \Delta_{d_x-1}$. For simplicity, let us just assume that (X^i, α^i) converges to (X^0, α^0). Consequently, $x = \sum_{j=1}^{d_x} \alpha_j^0 x_j^0$. By $f \in \ell_S^\infty(\mathcal{X})$ and $\lim_{i\to\infty} x_j^i = x_j^0$, $f(x_j^0) \leq \liminf_{i\to\infty} f(x_j^i) \leq \liminf_{i\to\infty} y_i = y^0$, where the second inequality follows from $f(x_j^i) \leq y^i$ for each x_j^i by definition of $\mathcal{I}_f(y^i)$. Hence, $x_j^0 \in \mathcal{I}_f(y^0)$, for all j. Since $x = \sum_{j=1}^{d_x} \alpha_j^0 x_j^0$, and $x_j^0 \in \mathcal{I}_f(y^0)$, it must be that $x \in \mathrm{conv}[\mathcal{I}_f(y^0)]$. Therefore, $y^0 \in \mathcal{Q}_f(x)$. We conclude that $\min(\mathcal{Q}_f(x)) = y^0$ because $y^0 = \inf \mathcal{Q}_f(x)$ by definition.

We next show that $\mathbf{Q}f \in \ell_S^\infty(\mathcal{X})$ for any $f \in \ell_S^\infty(\mathcal{X})$. It is easy to see that $\mathbf{Q}f \in \ell^\infty(\mathcal{X})$ because $\inf_{x\in\mathcal{X}} f(x) \leq \mathbf{Q}f(x) \leq \sup_{x\in\mathcal{X}} f(x)$. We prove the result by contradiction. Suppose that $\mathbf{Q}f \notin \ell_S^\infty(\mathcal{X})$, i.e., there exists $x^0 \in \mathcal{X}$, and a sequence $x^1, \ldots, x^n, \ldots \in \mathcal{X}$ such that $x^i \to x^0$ and $\liminf_{i\to\infty} \mathbf{Q}f(x^i) < y^0 - \epsilon$ for some constant $\epsilon > 0$ and $y^0 := \mathbf{Q}f(x_0)$. Since $\mathcal{X}$ is compact, there must exist a subsequence of $\{x^i\}_{i=1}^\infty$, denoted as $\{x^{n_i}\}_{i=1}^\infty$, such that for all i,

$$y^i := \mathbf{Q}f(x^{n_i}) \to c < y^0 - \epsilon. \tag{A.19}$$

For simplicity, we can assume that $x^{n_i} = x^i$ for all $i = 1, 2, \ldots$. Similar to the proof above, each x^i can be written as $\sum_{j=1}^{d_x} \alpha_j^i x_j^i$, where $\alpha^i = (\alpha_j^i, \ldots, \alpha_{d_x}^i) \in \Delta_{d_x-1}$ and $x_j^i \in \mathcal{I}_f(y_i)$. By compactness of $\mathcal{X} \times \Delta_{d_x-1}$, there exist subsequences of $X^i := (x_1^i, \ldots, x_{d_x}^i)$ and α^i, $i = 1, 2, \ldots$, such that they converge to $X^0 := (x_1^0, \ldots, x_{d_x}^0)$ and $\alpha^0 \in \Delta_{d_x-1}$. Again, for simplicity, we can assume that the subsequences are the sequences X^i and α^i. Since $x^i \to x^0$, it is easy to see that $x^0 = \sum_{j=1}^{d_x} \alpha_j^0 x_j^0$. By $f \in \ell_S^\infty(\mathcal{X})$, for $j = 1, \ldots, d_x$,

$$f(x_j^0) \leq \liminf_{i\to\infty} f(x_j^i). \tag{A.20}$$

Moreover, $f(x_j^i) \leq y^i$ for all i because $x_j^i \in \mathcal{I}_f(y_i)$. Combining this result with (A.19) and (A.20) yields that $f(x_i^0) \leq \liminf_{i\to\infty} f(x_j^i) \leq \liminf_{i\to\infty} y^i < y^0 - \epsilon$ for all $i = 1, 2, \ldots, d_x$. Hence, $x^0 \in \mathcal{I}_f(y^0 - \epsilon)$ as $x^0 = \sum_{j=1}^{d_x} \alpha_j^0 x_j^0$. By definition of $\mathbf{Q}f$, it must be that $\mathbf{Q}f(x^0) \leq y^0 - \epsilon < y^0$, which leads to a contradiction with $y^0 = \mathbf{Q}f(x^0) = \min \mathcal{Q}_f(x^0)$. $\blacksquare$

We now proceed to prove Theorem 21. We first show that $\mathbf{Q}$ satisfies the three properties of Definition 1.

(1) For any $f \in \ell_S^\infty(\mathcal{X})$, the lower contour set of $\mathbf{Q}f$ at level y is defined as $\mathcal{I}_{\mathbf{Q}f}(y) = \{x \in \mathcal{X} : \mathbf{Q}f(x) \leq y\} := \mathrm{conv}[\mathcal{I}_f(y)]$, where $\mathcal{I}_f(y)$ is the lower contour set of f at level y. Since $\mathcal{I}_{\mathbf{Q}f}(y)$ is convex for any y, $\mathbf{Q}f \in \ell_Q^\infty(\mathcal{X})$.

(2) If $f \in \ell_Q^\infty(X)$, then $\mathrm{conv}[\mathcal{I}_f(y)] = \mathcal{I}_f(y)$ for any $y \in \mathbb{R}$. Thus, the lower contour set of f agrees with the lower contour set of $\mathbf{Q}f$ at any level y, which implies that $f = \mathbf{Q}f$.

(3) If $f \geq g$, then $\mathcal{I}_f(y) \supset \mathcal{I}_g(y)$ at any level $y \in \mathbb{R}$. If follows that $\mathrm{conv}[\mathcal{I}_f(y)] \supseteq \mathrm{conv}[\mathcal{I}_g(y)]$, which means that the level set of $\mathbf{Q}f$ contains the level set of $\mathbf{Q}g$ at any level y, i.e., $\mathbf{Q}f \geq \mathbf{Q}g$.

We next show that $\mathbf{Q}$ is d_∞-distance contraction. For any $f, g \in \ell_S^\infty(\mathcal{X})$, let $\epsilon := \|f - g\|_\infty$. Then, $g(x) - \epsilon \leq f(x) \leq g(x) + \epsilon$. It is easy to see that $\mathbf{Q}(g + c) = \mathbf{Q}g + c$ for any constant c. By order preserving property of $\mathbf{Q}$, $\mathbf{Q}g(x) - \epsilon \leq \mathbf{Q}f(x) \leq \mathbf{Q}g(x) + \epsilon$. It follows that

$$\|\mathbf{Q}g - \mathbf{Q}f\|_\infty \leq \epsilon = \|f - g\|_\infty.$$

∎

Proof of Theorem 24

Without loss of generality, we can assume that the domain $\mathcal{X} = [0, 1]^k$. For a vector $w \in \mathbb{R}^k$, denote $w(i)$ as the ith entry of w.

(1) We first prove that $\mathbf{QM}$ is reshaping.

For any $f \in \ell_S^\infty(\mathcal{X})$, $\mathbf{QM}f \in \ell_Q^\infty(\mathcal{X})$ by Theorem 21. Therefore, we only need to show that $\mathbf{QM}f \in \ell_M^\infty(\mathcal{X})$. Let $g := \mathbf{M}f$. By Theorem 11, $g \in \ell_M^\infty(\mathcal{X})$, so that for any $y \in \mathbb{R}$, the lower contour set $\mathcal{I}_g(y)$ satisfies:

$$\text{For any } x \in \mathcal{I}_g(y) \text{ and } x' \in \mathcal{X} \text{ such that } x' \leq x, x' \in \mathcal{I}_g(y). \tag{A.21}$$

Therefore, we need to prove that for any x, y such that $x \in \mathrm{conv}[\mathcal{I}_g(y)]$, $x' \in \mathrm{conv}[\mathcal{I}_g(y)]$ for any $x' \in \mathcal{X}$ such that $x' \leq x$.

First, we show the following:

$$\text{If } x' = x - te_i \text{ for some } t \geq 0, \text{ then } x' \in \mathrm{conv}[\mathcal{I}_g(y)], \tag{A.22}$$

where e_i is defined as the ith standard unit vector, $i = 1, 2, \ldots, k$. Without loss of generality, we can simply assume that $i = 1$, so x' and x are the same for all entries except for the first one. By assumption that $\mathcal{X} = [0, 1]^k$, we know that the first entry of x', denoted as $x'(1)$, must be non-negative. Since $x \in \mathrm{conv}[\mathcal{I}_g(y)]$, by Carathéodory's theorem, there exists a finite set of points $x_1, \ldots, x_{d_x}$ such that $d_x \leq k + 1$, $x_j \in \mathcal{I}_g(y)$, $j = 1, 2, \ldots, d_x$, and $(\alpha_1, \ldots, \alpha_{d_x}) \in \Delta_{d_x - 1}$, such that $\sum_{j=1}^{d_x} \alpha_j x_j = x$. Define $\widetilde{x}_j = (0, x_j(2), \ldots, x_j(k))$ as a vector which is constructed by replacing the first entry of x_j with 0. Therefore, $\widetilde{x} := \sum_{j=1}^{d_x} \alpha_j \widetilde{x}_j = (0, x(2), \ldots, x(k))$ is a vector such that $\widetilde{x} \leq x' \leq x$. Therefore, there must exist $x_1^*, \ldots, x_{d_x}^*$ such that $x_j^* = (x_j^*(1), x_j(2), \ldots, x_j(k))$ with $x_j^*(1) \in [0, x_j(1)]$ such that $\sum_{j=1}^{d_x} \alpha_j x_j^*(1) = x'(1) \in [0, x(1)]$. By construction, $x_j^* \in \mathcal{X}$. Since $x_j^* \leq x_j \in \mathcal{I}_g(y)$, (A.21) implies that $x_j^* \in \mathcal{I}_g(y)$. It follows that $\sum_{j=1}^{d_x} \alpha_j x_j^* = x'$, and therefore $x' \in \mathrm{conv}[\mathcal{I}_g(y)]$.

Now, for any $x' \in \mathcal{X}$ such that $x' \leq x$, denote $v := x - x' \geq 0$. Since $x \in \mathrm{conv}[\mathcal{I}_g(y)]$, it follows that $x - v(1)e_1 \in \mathrm{conv}[\mathcal{I}_g(y)]$, and then that $(x - v(1)e_1) - v(2)e_2 \in \mathrm{conv}[\mathcal{I}_g(y)]$, ... Therefore, after applying (A.22) for k times, $x' = x - v(1)e_1 - v(2)e_2 - \cdots - v(k)e_k \in$

conv$[\mathcal{I}_g(y)]$. By (2.7), $\mathbf{Q}g(x) := \min\{y \in \mathbb{R} : x \in \text{conv}[\mathcal{I}_g(y)]\}$. Let $y' := \mathbf{Q}g(x)$ so that $x \in \text{conv}[\mathcal{I}_g(y')]\}$. Then, for any $x' \in \mathcal{X}$ such that $x' \leq x$, $x' \in \text{conv}[\mathcal{I}_g(y'))\}$. That implies $\mathbf{Q}g(x') \leq y' = \mathbf{Q}g(x)$. Therefore, we conclude that $\mathbf{Q}g = \mathbf{QM}f$ is nondecreasing.

(2) Next, we can show that $\mathbf{QM}$ satisfies the rest of the properties of Definition 1 using the same argument as in the proof of Theorem 16, replacing $\mathbf{C}$ with $\mathbf{Q}$. We omit it for brevity.

(3) Finally, since $\mathbf{M}$ and $\mathbf{Q}$ are both d_∞-distance contractions by Theorems 11 and 21, the composite map $\mathbf{QM}$ is also a d_∞-distance contraction. $\blacksquare$

Proof of Theorem 26

We first show that if $f \in \ell_S^\infty(\mathcal{X})$, then $\mathbf{R}f \in \ell_S^\infty(\mathcal{X})$. This result is used in parts (i) and (iii) to ensure that we apply the $\mathbf{C}$ and $\mathbf{Q}$ operators to lower semi-continuous functions. By $f \in \ell_S^\infty(\mathcal{X})$, for any sequence $x^1, ..., x^n, ... \in \mathcal{X}$ such that $\lim_{i \to \infty} x^i = x^0 \in \mathcal{X}$, $f(x^0) \leq \liminf_{i \to \infty} f(x^i)$. Therefore, for any $\epsilon > 0$, there exists N large enough such that for any $i > N$, $f(x^i) > f(x^0) - \epsilon$. It follows that for any $i > N$, $\mathbf{R}f(x^i) > \mathbf{R}f(x^0) - \epsilon$. Hence, $\liminf_{i \to \infty} \mathbf{R}f(x^i) \geq \mathbf{R}f(x^0)$, i.e., $\mathbf{R}f \in \ell_S^\infty(\mathcal{X})$.

We now proceed to prove each of the parts of the Lemma.

Part (i): $\mathbf{CR}f \in \ell_C^\infty(\mathcal{X})$ by the definition of $\mathbf{C}$ applied to $\mathbf{R}f$ and $\mathbf{R}f \in \ell_S^\infty(\mathcal{X})$. Moreover, by statement (2) of Lemma 36, there exist $d \leq k + 2$ points $x_1, \ldots, x_d \in \mathcal{X}$ and $\alpha_1 > 0, \ldots, \alpha_d > 0$, $\sum_{i=1}^d \alpha_i = 1$ such that $\mathbf{CR}f(x) = \sum_{i=1}^d \alpha_i \mathbf{R}f(x_i)$, where $x = \sum_{i=1}^d \alpha_i x_i$. Therefore, $\mathbf{CR}f \in \ell_R^\infty(\mathcal{X})$ because $\mathbf{R}f(x_i) \in [\underline{f}, \bar{f}]$ for all $x_i \in \mathcal{X}$. It is easy to see that $\mathbf{CR}$ also satisfies invariance and order preservation because it is a composition of two operators that satisfy these properties. Indeed, if $f \in \ell_{CR}^\infty(\mathcal{X})$ then $\mathbf{CR}f = \mathbf{C}(\mathbf{R}f) = \mathbf{C}f = f$, and if $g \geq f$, $g, f \in \ell_S^\infty(\mathcal{X})$, then $\mathbf{R}g \geq \mathbf{R}f$, $\mathbf{R}g, \mathbf{R}f \in \ell_S^\infty(\mathcal{X})$, and $\mathbf{C}(\mathbf{R}g) \geq \mathbf{C}(\mathbf{R}f)$. Hence, $\mathbf{CR}$ is ℓ_{CR}^∞-enforcing with respect to $\ell_S^\infty(\mathcal{X})$. By Theorems 5 and 8, both $\mathbf{C}$ and $\mathbf{R}$ are d_∞-distance contractions. Therefore, the composite map $\mathbf{CR}$ must be a d_∞-distance contraction.

Part (ii): $\mathbf{R}f \in \ell_R^\infty(\mathcal{X})$ by the definition of $\mathbf{R}$. The $\mathbf{M}$ operator is the average of sorting operators, where each sorting operator does not change the range of the function. Therefore, $\mathbf{MR}f \in \ell_{MR}^\infty(\mathcal{X})$. As in part (i), it is easy to see that $\mathbf{MR}$ also satisfies invariance and order preservation because it is a composition of two operators that satisfy these properties. Hence, $\mathbf{MR}$ is ℓ_{MR}^∞-enforcing with respect to $\ell^\infty(\mathcal{X})$. Since $\mathbf{M}$ and $\mathbf{R}$ are both d_p-distance contractions for any $p \geq 1$, it must be that $\mathbf{MR}$ is d_p-distance contraction for any $p \geq 1$.

Part (iii): let $f \in \ell_S^\infty(\mathcal{X}) \cap \ell_R^\infty(\mathcal{X})$. By definition $\mathbf{Q}f(x) = \min\{y \in \mathbb{R} : x \in \text{conv}[\mathcal{I}_f(y)]\}$, so that $\bar{f} \in \mathcal{Q}_f(x) := \{y \in \mathbb{R} : x \in \text{conv}[\mathcal{I}_f(y)]\}$ and $\mathbf{Q}f(x) \leq \bar{f}$. Moreover, for any $y < \underline{f}$, $\mathcal{I}_f(y) = \emptyset$, so that $y \notin \mathcal{Q}_f(x)$ and $\mathbf{Q}f(x) \geq \underline{f}$. Consequently, $\mathbf{Q}f \in \ell_R^\infty(\mathcal{X})$. On the other hand, if $f \in \ell_S^\infty(\mathcal{X})$ then $\mathbf{Q}f \in \ell_S^\infty(\mathcal{X})$ by Theorem 21. Combining the previous results, $\mathbf{Q}f \in \ell_S^\infty(\mathcal{X}) \cap \ell_R^\infty(\mathcal{X})$. Since $\mathbf{R}f \in \ell_S^\infty(\mathcal{X}) \cap \ell_R^\infty(\mathcal{X})$ for any $f \in \ell_S^\infty(\mathcal{X})$, then $\mathbf{QR}f \in \ell_S^\infty(\mathcal{X}) \cap \ell_R^\infty(\mathcal{X})$. As in part (i), it is easy to see that $\mathbf{QR}$ also satisfies invariance and order preservation because it is a composition of two operators that satisfy these properties. Hence, $\mathbf{QR}$ is ℓ_{QR}^∞-enforcing with respect to $\ell_S^\infty(\mathcal{X})$. By Theorems 5 and 21, both $\mathbf{R}$ and $\mathbf{Q}$ are d_∞-distance contractions. Therefore, the composite map $\mathbf{QR}$ must be a d_∞-distance contraction. $\blacksquare$

Proof of Theorem 28

We need to show the 3 properties of Definition 1. (1) By definition, $\widetilde{f} := \mathbf{O}(h \circ f) \in \ell_1^\infty(\mathcal{X})$. Therefore, $\mathbf{O}_h(f) = h^{-1} \circ f \in \ell_{h,1}^\infty(\mathcal{X})$ and reshaping holds. (2) If $f \in \ell_{h,1}^\infty(\mathcal{X})$, then $h \circ f \in \ell_1^\infty(\mathcal{X})$ and $\mathbf{O}(h \circ f) = h \circ f$ by the invariance property of $\mathbf{O}$. Hence, $\mathbf{O}_h = h^{-1} \circ h \circ f = f$. (3) Since h is a real-valued bijection, it must be that h and h^{-1} are both strictly increasing or strictly decreasing. We can assume that they are strictly increasing without loss of generality. For any $f, g \in \ell_{h,0}^\infty(\mathcal{X})$ such that $f \geq g$, $h \circ f \geq h \circ g$, $\mathbf{O}(h \circ f) \geq \mathbf{O}(h \circ g)$, and $h^{-1} \circ \mathbf{O}(h \circ f) \geq h^{-1}\mathbf{O}(h \circ g)$. Therefore, $\mathbf{O}_h$ is order-preserving.

To show contractivity, let $\widetilde{f} := \mathbf{O}_h(f)$ and $\widetilde{g} := \mathbf{O}_h(g)$. Then, $\rho_h(\widetilde{f}, \widetilde{g}) = \rho(h \circ \widetilde{f}, h \circ \widetilde{g}) = \rho(\mathbf{O}(h \circ f), \mathbf{O}(h \circ g))$. Since $\mathbf{O}$ is a ρ-distance contraction, $\rho(\mathbf{O}(h \circ f), \mathbf{O}(h \circ g)) \leq \rho(h \circ f, h \circ g) = \rho_h(f, g)$. Hence, $\mathbf{O}_h$ is a ρ_h-distance contraction. ∎

Proof of Corollary 32

(1) We show that the event $\{f_l \leq f_0 \leq f_u\}$ implies the event $\{\mathbf{O}f_l \leq f_0 \leq \mathbf{O}f_u\}$ by the properties of the $\mathbf{O}$-operator. Indeed, by order preservation, $\{f_l \leq f_0 \leq f_u\}$ implies that $\{\mathbf{O}f_l \leq \mathbf{O}f_0 \leq \mathbf{O}f_u\}$, which is equivalent to $\{\mathbf{O}f_l \leq f_0 \leq \mathbf{O}f_u\}$ because $\mathbf{O}f_0 = f_0$ by invariance.

(2) The result follows from $\rho(\mathbf{O}f_0, \mathbf{O}f) \leq \rho(f_0, f)$ by ρ-distance contraction and $\mathbf{O}f_0 = f_0$ by invariance.

(3) The result follows directly by ρ-distance contraction. ∎

References

D. Anevski and O. Hössjer. A general asymptotic scheme for inference under order restrictions. *Ann. Statist.*, 34(4):1874–1930, 08 2006.

Miriam Ayer, H Daniel Brunk, George M Ewing, William T Reid, and Edward Silverman. An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.*, 26(4):641–647, 1955.

Philipp Bach, Sven Klaassen, Jannis Kueck, and Martin Spindler. Uniform inference in high-dimensional generalized additive models, 2020.

Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic theory*, 26(2):445–469, 2005.

RE Barlow, DJ Bartholomew, JM Bremner, and HD Brunk. *Statistical inference under order restrictions*. Wiley, New York, 1972.

Brendan K Beare and Zheng Fang. Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electron. J. Stat.*, 11(2):3841–3870, 2017.

Brendan K Beare and Jong-Myun Moon. Nonparametric tests of density ratio ordering. *Econometric Theory*, 31(3):471–492, 2015.

Brendan K Beare and Lawrence DW Schmidt. An empirical test of pricing kernel monotonicity. *J. Appl. Econometrics*, 31(2):338–356, 2016.

Pierre C Bellec. Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.*, 46(2):745–780, 2018.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Iván Fernández-Val. Conditional quantile processes based on series or many regressors. *J. Econometrics*, 213(1): 4–29, 2019.

Hugh D Brunk. Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.*, 26(4):607–616, 1955.

Sabyasachi Chatterjee, Adityanand Guntuboyina, and Bodhisattva Sen. On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.*, 2014. to appear.

Xi Chen, Qihang Lin, and Bodhisattva Sen. On degrees of freedom of projection estimators with applications to multivariate shape restricted regression. *J. Amer. Statist. Assoc.*, 115(529):173–186, 2020.

Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2009.

Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125, 2010.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.*, 42(5):1787–1818, 10 2014.

Victor Chernozhukov, Whitney K. Newey, and Andres Santos. Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311*, 2015.

Denis Chetverikov, Andres Santos, and Azeem M Shaikh. The econometrics of shape restrictions. *Annual Review of Economics*, 10:31–63, 2018.

T. J. Cole. Fitting smoothed centile curves to reference data. *J. Roy. Statist. Soc. Ser. A*, 151:385–418, 1988.

Jean Pierre Crouzeix. *Quasi-Concavity*, pages 1–3. Palgrave Macmillan UK, London, 2017.

Miguel A Delgado and Juan Carlos Escanciano. Distribution-free tests of stochastic monotonicity. *J. Econometrics*, 170(1):68–75, 2012.

Holger Dette and Stanislav Volgushev. Non-crossing non-parametric estimates of quantile curves. *J. Roy. Statist. Soc. Ser. B*, 70(3):609–627, 2008.

Lutz Dümbgen. Optimal confidence bands for shape-restricted curves. *Bernoulli*, 9(3): 423–449, 2003.

Cecile Durot and Hendrik P. Lopuhaa. Limit theory in monotone function estimation. *Statist. Sci.*, 33(4):547–567, 2018.

Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.

Nora Fenske, Thomas Kneib, and Torsten Hothorn. Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *J. Amer. Statist. Assoc.*, 106(494):494–510, 2011.

Joachim Freyberger and Brandon Reeves. Inference under shape restrictions. Technical report, University of Wisconsin-Madison, 2018.

Evarist Giné and Richard Nickl. Confidence bands in density estimation. *Ann. Statist.*, 38 (2):1122–1170, 04 2010a.

Evarist Giné and Richard Nickl. Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. *Bernoulli*, 16(4):1137–1163, 11 2010b.

Ulf Grenander. On the theory of mortality measurement. *Scand. Actuar. J.*, 1956(1):70–96, 1956.

Piet Groeneboom and Geurt Jongbloed. *Nonparametric estimation under shape constraints*. Cambridge University Press, 2014.

Piet Groeneboom, Geurt Jongbloed, and Jon A Wellner. Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.*, 26(4):1653–1698, 2001.

Angelo Guerraggio and Elena Molho. The origins of quasi-concavity: a development between mathematics and economics. *Historia Mathematica*, 31(1):62 – 75, 2004.

Adityanand Guntuboyina and Bodhisattva Sen. Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields*, 163:379–411, 2015.

Adityanand Guntuboyina and Bodhisattva Sen. Nonparametric shape-restricted regression. *Statist. Sci.*, 33(4):568–594, 2018.

Jinyong Hahn. Bootstrapping quantile regression estimators. *Econometric Theory*, 11(1): 105–121, 1995.

Qiyang Han and Jon A. Wellner. Multivariate convex regression: global risk bounds and adaptation. *arXiv preprint arXiv:1601.06844*, 2016.

Qiyang Han, Tengyao Wang, Sabyasachi Chatterjee, and Richard J Samworth. Isotonic regression in general dimensions. *Ann. Statist.*, 47(5):2440–2471, 2019.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media, 2009.

Nicolas W Hengartner and Philip B Stark. Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.*, pages 525–550, 1995.

John R Hicks and Roy GD Allen. A reconsideration of the theory of value. Part I. *Economica*, 1(1):52–76, 1934.

Clifford Hildreth. Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.*, 49(267):598–619, 1954.

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of Convex Analysis.* Springer, 2001.

Joel L Horowitz and Sokbae Lee. Nonparametric estimation and inference under shape restrictions. *J. Econometrics*, 201(1):108–126, 2017.

David Jacho-Chávez, Arthur Lewbel, and Oliver Linton. Identification and nonparametric estimation of a transformed additively separable model. *J. Econometrics*, 156(2):392–407, 2010.

Hanna K. Jankowski and Jon A. Wellner. Estimation of a discrete monotone distribution. *Electron. J. Stat.*, 3:1567–1605, 2009.

Roger Koenker. Additive models for quantile regression: Model selection and confidence bandaids. *Braz. J. Probab. Stat.*, 25(3):239–262, 11 2011.

Roger Koenker and Ivan Mizera. Quasi-concave density estimation. *Ann. Statist.*, 38(5):2998–3027, 2010.

Timo Kuosmanen. Representation theorem for convex nonparametric least squares. *Econom. J.*, 11(2):308–325, 2008.

Oleg V. Lepskiĭ. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992.

Oleg V. Lepskiĭ and Vladimir Grigor'evich Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, 25(6):2512–2546, 12 1997.

Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics: theory and practice.* Princeton University Press, 2007.

Enno Mammen. Estimating a smooth monotone regression function. *Ann. Statist.*, 19(2):724–740, 1991.

AW Marshall. Discussion on Barlow and van Zwet's paper. *Nonparametric Techniques in Statistical Inference*, 1969:174–176, 1970.

Rosa L Matzkin. Restrictions of economic theory in nonparametric methods. *Handbook of econometrics*, 4:2523–2558, 1994.

Roger B Nelsen. *An introduction to copulas.* Springer Science & Business Media, 2007.

Pin T. Ng and Martin Maechler. *COBS – Constrained B-splines (Sparse matrix based)*, 2020. R package version 1.3-4.

Hiroki Nishimura and Efe A. Ok. Solvability of variational inequalities on hilbert lattices. *Mathematics of Operations Research*, 37(4):608–625, 2012.

Jens Præstgaard and Jon A Wellner. Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, 21(4):2053–2086, 1993.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2019.

Tim Robertson, FT Wright, and RL Dykstra. *Order restricted statistical inference.* Wiley, New York, 1988.

D. Ruppert, S. J. Sheather, and M. P. Wand. An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, 90(432):1257–1270, 1995.

Emilio Seijo and Bodhisattva Sen. Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.*, 39(3):1633–1657, 2011.

Charles J. Stone. Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, 8(6):1348–1360, 1980.

J Tanner, R Whitehouse, and M Takaishi. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children 1965 ii. *Archives of disease in childhood*, 41:454–71, 11 1966.

C. van Eeden. Maximum likelihood estimation of ordered probabilities. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen A*, 59:444–455, 1956.

Matt Wand. *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*, 2019.

Larry Wasserman. *All of Nonparametric Statistics.* Springer, 2006.

Ying Wei, Anneli Pere, Roger Koenker, and Xuming He. Quantile regression methods for reference growth charts. *Stat. Med.*, 25(8):1369–1382, 2006.

Eberhard Zeidler. *Nonlinear functional analysis and its applications: III: variational methods and optimization.* Springer Science & Business Media, 1984.

Cun-Hui Zhang. Risk bounds in isotonic regression. *Ann. Statist.*, 30(2):528–555, 2002.