Tilburg University

Projection Estimates of Constrained Functional Parameters

Fils-Villetard, A.; Guillou, A.; Segers, J.

Publication date:
2005

CentER

Discussion Paper

No. 2005–111

PROJECTION ESTIMATES OF CONSTRAINED FUNCTIONAL PARAMETERS

By Amélie Fils-Villetard, Armelle Guillou, Johan Segers

October 2005

ISSN 0924-7815

TILBURG UNIVERSITY

PROJECTION ESTIMATES OF CONSTRAINED FUNCTIONAL PARAMETERS

Amélie Fils-Villetard, Armelle Guillou and Johan Segers*

Université Pierre et Marie Curie, Université Pierre et Marie Curie and Tilburg University

Curve estimation problems can often be formulated in terms of a closed and convex parameter set embedded in a real Hilbert space. This is the case, for instance, if the curve of interest is a monotone or convex density or regression function, the support function of a convex set, or the Pickands dependence function of an extreme-value copula. The topic of this paper is the estimator that results when an arbitrary initial estimator possibly falling outside the parameter set is projected onto this parameter set. If direct computation of the projection is infeasible, the full parameter set can be replaced by an approximating sequence of finite-dimensional subsets. Asymptotic properties of the initial estimator sequence in the Hilbert space topology transfer easily to those of the projected sequence and its approximating sequence.

1 INTRODUCTION

Suppose we wish to estimate a function or a vector of functions subject to shape constraints. The functions could for instance be regression functions, probability density functions, hazard rates, and so on, and the shape constraint could for instance be that the functions are monotone, convex, non-negative, or a combination thereof. We have at our disposal an estimator, but unfortunately this estimator is not guaranteed to satisfy the constraints. Then how to modify this estimator so that the constraints are met?

If all relevant information in the sample is already contained in the initial estimator, then the modified estimator should depend on the data only through this initial estimator. Moreover, the modification should be as small as possible, to be measured along some metric on the appropriate function class.

Consider for instance the problem of estimating a regression function that is known to be non-decreasing. Mammen [18] proposes a two-step procedure whereby an initial Nadaraya-Watson kernel estimator is isotonized

Running title: *Shape-Constrained Functions*. This version: October 13, 2005.

*Corresponding author. Supported by a VENI grant of the Netherlands Organization for Scientific Research (NWO).

AMS 2000 subject classifications: Primary 62G05, 62G07, 62G08, 62G20; secondary 62G32

JEL: C13, C14

Keywords: convex function, curve estimation, extreme-value copula, Pickands dependence function, projection, shape constraint, support function, tangent cone.

using the pool adjacent violator algorithm. The resulting estimator turns out to be the non-decreasing function with smallest L^2-distance to the initial smooth. Applying the same kind of minimization procedure to the derivative of the piece-wise linear interpolation of the empirical distribution function of a sample leads to the Grenander estimator [12] for a monotone probability density.

Suppose then that the parameter set, that is, the set of functions satisfying the model constraints, can be identified with a closed, convex subset of a real Hilbert space and that the initial estimator takes its values in this Hilbert space as well. A natural idea is then to project the initial estimator onto the parameter set, the projection being defined as the minimizer of the Hilbert space distance between the initial estimator and members of the parameter set. Because the latter is assumed to be closed and convex, the projection operator is well-defined. As projection operators are non-expansive, the estimation error, measured in the Hilbert space distance, of the projection estimator is never larger than the one of the initial estimator.

The aim of this paper, then, is to develop an abstract framework for the projection estimator, focusing on the relationship between the asymptotic properties of the initial estimator sequence and the corresponding projection estimator sequence. The main tool here is the functional delta method in combination with a result on the Hadamard differentiability of projection operators.

Moreover, as the definition of the projection estimator involves a typically not explicitly solvable infinite-dimensional minimization problem, we propose a computational tool replacing the latter optimization problem by a sequence of quadratic programs. The dimension of the quadratic program serves as a tuning parameter, sufficiently large values of which guarantee the approximate projection estimator sequence to have the same asymptotics as the true projection estimator sequence.

Finally, we work out the details for three examples: convex functions on the positive half-line (section 4), bivariate extreme-value copulas (section 5), and support functions of convex, compact regions in the plane (section 6). The corresponding quadratic programs are written down explicitly, and the smallest dimensions needed for the approximate projection estimators to be asymptotically equivalent to the true projection estimators are derived in terms of the rate of convergence of the initial estimator sequence. Simulations serve to assess the finite-sample properties of the estimators.

A different though related general projection framework has been proposed by Mammen, Marron, Turlach, and Wand in [19]. The major difference is that in their approach, the Hilbert space changes with the sample. For instance, in a non-parametric regression setting with random design, the Hilbert space is an L^2-space with the underlying measure equal to the distribution corresponding to a non-parametric density estimate of the explanatory variable; see their proposition 1.

We should emphasize that the feasibility of the projection methodology crucially depends on the convexity of the parameter set. For instance, the problem of estimating a unimodal function with unknown mode, see for instance [24], falls outside the scope of the method, since convex combinations of unimodal functions with different modes are in general no longer unimodal.

All our asymptotic results are with respect to the Hilbert space topology. In the common case that the Hilbert space is an L^2-space, this means that the topology in force is the one of convergence in quadratic mean. In particular, we do not obtain results on point-wise or uniform convergence. Of course, in particular cases, it may be true that the projection estimator sequence does satisfy certain limit laws in these other topologies, but to establish such properties would require case-specific arguments not pursued here.

The projection estimator and the tool to compute it are presented in section 2. Section 3 states the asymptotic properties of the projection estimator and its finite-dimensional approximations. Sections 4 to 6 contain the three examples and can be read independently of one another. All proofs are deferred to the appendix.

2 THE ESTIMATORS

Throughout this article, the 'parameter' set $\mathcal{F}$ is a non-empty, closed and convex subset of a real Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. The aim is to estimate f_0, known to be in $\mathcal{F}$. There's available an estimate $\hat{f}$, but unfortunately $\hat{f}$ does not belong to $\mathcal{F}$. The subject of this article is the estimator that arises as the minimizer of the function $f \mapsto \|\hat{f} - f\|$ over $\mathcal{F}$, the projection of $\hat{f}$ on $\mathcal{F}$.

2.1 Projections on closed, convex sets

In view of their fundamental importance in this article, we quickly review a number of useful and well-known properties of projections on closed, convex sets in a Hilbert space; see for instance [32], chapter 2. A first matter is that of their existence and uniqueness.

LEMMA 2.1. *Let $\mathcal{F}$ be a non-empty, closed and convex subset of a real Hilbert space $\mathbb{H}$. For $h \in \mathbb{H}$ there exists a unique $f \in \mathcal{F}$ such that $\|h - f\| = \inf\{\|h - g\| : g \in \mathcal{F}\}$.*

The unique $f \in \mathcal{F}$ in Lemma 2.1 is called the (orthogonal) projection of h on $\mathcal{F}$; notation $f = \Pi(h \mid \mathcal{F})$. In geometrical terms, the projection $\Pi(h \mid \mathcal{F})$ of h on $\mathcal{F}$ is the unique point $f \in \mathcal{F}$ such that the hyperplane passing through f and orthogonal to $h - f$ separates h from $\mathcal{F}$. In the special case that $\mathcal{F}$ is affine, $\mathcal{F}$ is actually contained in this separating hyperplane.

LEMMA 2.2. *Let $\mathcal{F}$ be a non-empty, closed and convex subset of a real Hilbert space $\mathbb{H}$. For $h \in \mathbb{H}$ and $f \in \mathcal{F}$, the following are equivalent:*

(i) $f = \Pi(h \mid \mathcal{F})$;

(ii) $\langle h - f, f - g \rangle \geq 0$ *for all* $g \in \mathcal{F}$;

(iii) $\|h - f\|^2 + \|f - g\|^2 \leq \|h - g\|^2$ *for all* $g \in \mathcal{F}$.

If $\mathcal{F}$ is also a cone, then another equivalent statement is

(iv) $\langle h - f, f \rangle = 0$ *and* $\langle h - f, g \rangle \leq 0$ *for all* $g \in \mathcal{F}$.

If $\mathcal{F}$ is also affine, then the inequalities in (ii) and (iii) become equalities.

A familiar special case arises if $\mathcal{F}$ is a closed linear subspace: In that case, the projection operator is linear and bounded. In general, the projection operator on a closed, convex set is non-expansive: the distance between the projections of two points is at most as large as the original distance between those two points. In particular, projections are continuous.

LEMMA 2.3. *For $g, h \in \mathbb{H}$, we have* $\|\Pi(g \mid \mathcal{F}) - \Pi(h \mid \mathcal{F})\| \leq \|g - h\|$.

2.2 The projection estimator

The projection estimator based on the initial estimator $\hat{f}$ can now concisely be written as

$$\hat{f}^{\mathrm{P}} = \Pi(\hat{f} \mid \mathcal{F}) = \arg\min_{f \in \mathcal{F}} \|f - \hat{f}\|. \tag{2.1}$$

By definition, the projection estimator always belongs to the parameter set, $\hat{f}^{\mathrm{P}} \in \mathcal{F}$. If the Hilbert space is an L^2 space, then $\hat{f}^{\mathrm{P}} = \arg\min_{f \in \mathcal{F}} \int (f - \hat{f})^2$; therefore, we sometimes also use the phrase 'least-squares estimator'.

With respect to the Hilbert space norm, the projection estimator is always at least as good as the initial estimator: by Lemma 2.2(iii),

$$\|\hat{f}^{\mathrm{P}} - f_0\| \leq \|\hat{f} - f_0\|, \tag{2.2}$$

where $f_0 \in \mathcal{F}$ denotes the true but unknown parameter. In other words, there can be no harm in performing the projection.

Since projections are continuous (Lemma 2.3), the projection estimator $\hat{f}^{\mathrm{P}}$ is a measurable element of $\mathbb{H}$ as soon as the initial estimator $\hat{f}$ is one. If for instance $\mathbb{H} = L^2 = L^2([0,1], \mathrm{d}x)$ and if $\hat{f}$ is a random element in $C[0,1]$, then, since the embedding of $C[0,1]$ into L^2 is continuous, $\hat{f}$ is indeed also a measurable element of L^2. However, even if $\mathcal{F}$ can be identified with a subset of $C[0,1]$, then still nothing can in general be said about the measurability or continuity properties of the restriction of the projection operator to $C[0,1]$.

2.3 Examples

Further on in this paper, we will investigate in some detail the application of the projection methodology to the estimation of a convex function on the positive half-line (section 4), a bivariate extreme-value copula (section 5) and the support function of a compact, convex set in the plane (section 6). In those cases, the corresponding projection operator cannot be expressed explicitly, a possible solution being the successive approximation of $\mathcal{F}$ by means of a sieve of finite-dimensional subsets $\mathcal{F}_m$, to be explained in section 2.4. Here, we list a number of examples for which the projection does admit an explicit formula.

Bounded functions. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measurable space and let $L^2 = L^2(\mathcal{X}, \mathcal{A}, \mu)$ be the real Hilbert space of equivalence classes of real-valued, square-integrable functions on $\mathcal{X}$. Let $\mathcal{F} = \{f \in L^2 : f \geq 0\}$ be the cone of μ-almost everywhere non-negative, square-integrable functions on $\mathcal{X}$. By criterion (ii) in Lemma 2.2, the projection operator on $\mathcal{F}$ is given by $\Pi(h \mid \mathcal{F}) = \max(h, 0)$ for $h \in L^2$. For $g, g_1, g_2 \in L^2$, simple generalizations are $\mathcal{F} = \{f \in L^2 : f \geq g\}$ with $\Pi(h \mid \mathcal{F}) = \max(h, g)$ or $\mathcal{F} = \{f \in L^2 : g_1 \leq f \leq g_2\}$ with $\Pi(h \mid \mathcal{F}) = \min\{\max(h, g_1), g_2\}$.

Pairs of parallel functions. Let again $(\mathcal{X}, \mathcal{A}, \mu)$ be a measurable space and consider the real Hilbert space $L^2 \oplus L^2$, the direct sum of $L^2(\mathcal{X}, \mathcal{A}, \mu)$ with itself. Elements of $L^2 \oplus L^2$ can be thought of as column vectors $(h_1, h_2)'$ with $h_1, h_2 \in L^2$. For a fixed $g \in L^2$, let $\mathcal{F}$ be the closed linear subspace of $L^2 \oplus L^2$ consisting of those $(f_1, f_2)'$ such that $f_2 - f_1 = \alpha g$ for some arbitrary real α. The orthogonal projection from $L^2 \oplus L^2$ onto $\mathcal{F}$ is given by $\Pi\left((h_1, h_2)' \mid \mathcal{F}\right) = (f - (\alpha/2)g, f + (\alpha/2)g)'$ where $f = (h_1 + h_2)/2$ and $\alpha = \int (h_2 - h_1)g / \int g^2$.

If $\mu(\mathcal{X}) < \infty$, we can take for g the constant function equal to one. The space $\mathcal{F}$ then consists of pairs of functions at a fixed, arbitrary distance. The corresponding projection operator is useful, for instance, when estimating two regression curves which are believed to be parallel, see for instance the onions example in [23], [4], chapter 6.5, and [19], section 2.2.

Observed derivatives. Given estimates for a pair of functions, one of which is known to be the derivative of the other, how to modify these estimates as little as possible such that the new estimates have the same property?

Let $L^2 = L^2([0, 1], \mathrm{d}t)$ be the real Hilbert space of equivalence classes of real-valued, Lebesgue square-integrable functions on the unit interval. Let $\mathcal{F}$ be the subset of $L^2 \oplus L^2$ consisting of those pairs $(f_1, f_2)'$ for which f_1 is almost everywhere equal to a primitive function of f_2. Let $P : L^2 \to L^2$ be the projection onto the orthogonal complement of the subspace of constant functions, so $Pf = f - \int_0^1 f$, and let $S : L^2 \to L^2$ be the operator mapping $f \in L^2$ to its primitive $t \mapsto Sf(t) = \int_0^t f$. Then a pair $(f_1, f_2)'$ belongs to

$\mathcal{F}$ if and only if the function $f_1 - Sf_2$ is constant, that is, if $A\left((f_1, f_2)'\right) := P(f_1 - Sf_2)$ is equal to the zero function. Hence $\mathcal{F}$ is equal to the kernel of the bounded linear operator A. In particular, $\mathcal{F}$ is a closed linear subspace.

The projection operator on $\mathcal{F}$ can be shown to be $\Pi\left((h_1, h_2)' \mid \mathcal{F}\right) = (h_1 - f, h_2 - Sf)$, where f is given by

$$f(t) = h(t) - \frac{1}{2}(e^t + e^{-t})\alpha + \frac{1}{2}\int_0^t (e^{t-s} - e^{s-t})h(s)\mathrm{d}s, \qquad \text{for } t \in [0, 1],$$

with $h = A\left((h_1, h_2)'\right)$ and $\alpha = (e - e^{-1})^{-1} \int_0^1 (e^{t-1} + e^{1-t})h(t)\mathrm{d}t$. The way to arrive at the stated expression for f is by writing the projection on $\mathcal{F} = N(A)$ in the form $\Pi = I - A^*(AA^*)^{-1}A$; by Banach's theorem ([16], Theorem 1.8.5), computing the inverse of the restriction of AA^* to the range of A boils down to solving an ordinary linear differential equation with constant coefficients [6]. However, given the stated expression for f, it is a routine matter to check that $(h_1 - f, h_2 - Sf)' \in \mathcal{F}$ and $\langle (f, Sf)', (g_1, g_2)' \rangle = 0$ for all $(g_1, g_2)' \in \mathcal{F}$, confirming the validity of the stated expression for Π.

2.4 Finite-dimensional approximations

The projection estimator (2.1) is defined as the solution of a minimization problem, so the question arises how to compute it. In many cases, direct computation is infeasible (sections 4, 5 and 6). Therefore, we present here a generally applicable sieve method to approximate the projection estimator up to any desired accuracy. In section 3 we state conditions under which the asymptotic behavior of the approximate projection estimator is the same as that of the true projection estimator.

Let $\mathcal{F}_m$ be a finite-dimensional subset of $\mathcal{F}$ of the form

$$\mathcal{F}_m = \left\{ \sum_{i=1}^{d_m} \lambda_i h_{im} : \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d_m})' \in \Lambda_m \right\} \tag{2.3}$$

where the h_{im} $(i = 1, \ldots, d_m)$ are fixed elements in $\mathbb{H}$ and where Λ_m is a closed and convex subset of d_m-dimensional Euclidean space. Typically, Λ_m is defined through linear equality and inequality constraints. The idea is to have a sequence of such finite-dimensional subsets which provide with increasing m increasingly more accurate approximations of the full set $\mathcal{F}$ in a sense to be made precise in the next section.

The approximate projection estimator based on initial estimator $\hat{f}$ is then defined as the projection of $\hat{f}$ on $\mathcal{F}_m$,

$$\hat{f}_m^{\mathrm{p}} = \Pi(\hat{f} \mid \mathcal{F}_m). \tag{2.4}$$

The approximate projection estimator is of the form $\hat{f}_m^{\mathrm{p}} = \sum_{i=1}^{d_m} \lambda_i h_{im}$, the vector $\boldsymbol{\lambda}$ being the solution to the quadratic program

$$\text{minimize } \boldsymbol{\lambda}'\boldsymbol{A}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\boldsymbol{b} \text{ subject to } \boldsymbol{\lambda} \in \Lambda_m \tag{2.5}$$

with the $d_m \times d_m$ matrix $\boldsymbol{A}$ given by $A_{ij} = \langle h_{im}, h_{jm} \rangle$ and with the vector $\boldsymbol{b}$ given by $b_i = \langle h_{im}, \hat{f} \rangle$. Note that $\boldsymbol{A}$ does not depend on $\hat{f}$, and that if $\mathbb{H}$ is an L^2-space, then the entries of $\boldsymbol{b}$ involve integrals for which numerical quadrature might be required.

3 ASYMPTOTICS

Let $\hat{f}_n$ be a sequence of estimators. Then the corresponding sequence of projection estimators is $\hat{f}_n^{\mathrm{p}} = \Pi(\hat{f}_n \mid \mathcal{F})$, while the approximate projection estimators are $\hat{f}_{n,m_n}^{\mathrm{p}} = \Pi(\hat{f}_n \mid \mathcal{F}_{m_n})$, with typically $m_n \to \infty$. In this section, we study the extent to which asymptotic properties of $\hat{f}_n$ are inherited by $\hat{f}_n^{\mathrm{p}}$ and $\hat{f}_{n,m_n}^{\mathrm{p}}$.

3.1 Rates of convergence

A first property is obvious: in view of (2.2), the rate of convergence of the projection estimator sequence $\hat{f}_n^{\mathrm{p}}$ in the Hilbert space norm is at least as fast as the rate of convergence of the initial estimator sequence $\hat{f}_n$.

For the approximate projections, $\hat{f}_{n,m_n}^{\mathrm{p}}$, however, the situation is different, as the true f_0 is possibly not contained in the finite-dimensional approximations $\mathcal{F}_{m_n}$ to $\mathcal{F}$. However, if $\mathcal{F}_{m_n}$ is in some sense close to $\mathcal{F}$, then $\hat{f}_{n,m_n}^{\mathrm{p}} = \Pi(\hat{f}_n \mid \mathcal{F}_{m_n})$ may be expected to be close to $\hat{f}_n^{\mathrm{p}} = \Pi(\hat{f}_n \mid \mathcal{F})$. The following lemma serves to get a handle on $\|\hat{f}_{n,m_n}^{\mathrm{p}} - \hat{f}_n^{\mathrm{p}}\|$. The proofs of this and the following results of this section are to be found in Appendix A.1.

Lemma 3.1. *If $\mathcal{F}$ and $\mathcal{G}$ are non-empty, closed and convex subsets of $\mathbb{H}$ and if $\mathcal{G} \subset \mathcal{F}$, then*

$$\|\Pi(h \mid \mathcal{G}) - \Pi(h \mid \mathcal{F})\| \le [\delta\{2\,\|h - \Pi(h \mid \mathcal{F})\| + \delta\}]^{1/2} \qquad \text{for } h \in \mathbb{H},$$

where $\delta = \|\Pi(h \mid \mathcal{F}) - \Pi(\Pi(h \mid \mathcal{F}) \mid \mathcal{G})\|$.

An immediate consequence of Lemma 3.1 is the following result on the rate of convergence of the approximate projection estimator $\hat{f}_{n,m_n}^{\mathrm{p}}$.

Theorem 3.2. *Assume that $\|\hat{f}_n - f_0\| = O_p(\varepsilon_n)$ or $o_p(\varepsilon_n)$ for some positive sequence ε_n tending to zero. If the positive integer sequence m_n is such that*

$$\|\hat{f}_n^{\mathrm{p}} - \Pi(\hat{f}_n^{\mathrm{p}} \mid \mathcal{F}_{m_n})\| = O_p(\varepsilon_n) \text{ or } o_p(\varepsilon_n), \text{ respectively,} \tag{3.1}$$

then also $\|\hat{f}_{n,m_n}^{\mathrm{p}} - f_0\| = O_p(\varepsilon_n)$ or $o_p(\varepsilon_n)$, respectively.

Theorem 3.2 suggests the following rule for the choice of m_n: choose m_n large enough such that (3.1) holds. Since, typically, the left-hand side will be decreasing with m_n, this indeed only imposes a lower bound on m_n. In practice, constraints in computing time and power will also impose an upper bound on m_n.

If the parameter space $\mathcal{F}$ is compact, then the left-hand side of (3.1) is bounded by the Hausdorff distance $\sup_{f \in \mathcal{F}} \|f - \Pi(f \mid \mathcal{F}_{m_n})\|$ between $\mathcal{F}$ and $\mathcal{F}_{m_n}$. A typical way of showing (3.1) is then by bounding this Hausdorff distance. In many cases, however, $\mathcal{F}$ is not compact and the Hausdorff distance between $\mathcal{F}$ and $\mathcal{F}_m$ is infinite for all m. In such cases, condition (3.1) needs to be checked by more subtle methods, for instance by using the fact that $\|\hat{f}_n^{\mathrm{p}} - f_0\| = o_p(1)$.

3.2 Limit distributions

Now assume that there exists a positive sequence ε_n tending to zero and a random element g in $\mathbb{H}$ such that

$$\frac{\hat{f}_n - f_0}{\varepsilon_n} \rightsquigarrow g, \qquad \text{in } \mathbb{H}. \tag{3.2}$$

Then what can we say about the limit distributions of the estimator sequences $\hat{f}_n^{\mathrm{p}}$ and $\hat{f}_{n,m_n}^{\mathrm{p}}$?

Notice that in the previous display we explicitly mentioned the space $\mathbb{H}$ in which the convergence in distribution takes place. However, if the Hilbert space is $L^2([0,1])$, then usually, the estimator $\hat{f}_n$ will live in the smaller space $D[0,1]$ or even in $C[0,1]$, and the convergence in distribution will hold in the finer topology of uniform convergence, thus implying convergence in distribution in the weaker Hilbert space topology. As said already in the beginning of this section, our method is general but is restricted to the Hilbert-space world. More refined limit results have to be pursued by more specialized methods adapted to individual cases.

Since $(\hat{f}_n^{\mathrm{p}} - f_0)/\varepsilon_n = \{\Pi(\hat{f}_n \mid \mathcal{F}) - \Pi(f_0 \mid \mathcal{F})\}/\varepsilon_n$, we can turn (3.2) into an asymptotic result for $\hat{f}_n^{\mathrm{p}}$ by the functional delta method. What is required, then, is the (one-sided) Hadamard derivative of the projection operator $\Pi(\,\cdot \mid \mathcal{F})$ in f_0. This derivative exists and is a projection as well, but now on the tangent cone of $\mathcal{F}$ at f_0, defined as

$$T_{\mathcal{F}}(f_0) = \overline{\{\lambda(f - f_0) : \lambda \geq 0,\, f \in \mathcal{F}\}}.$$

Lemma 3.3. *Let $f_0 \in \mathcal{F}$ and let $T_{\mathcal{F}}(f_0)$ be the tangent cone of $\mathcal{F}$ at f_0. If ε_n is a positive sequence tending to zero and if $g_n \to g$ in $\mathbb{H}$, then*

$$\lim_{n\to\infty} \frac{\Pi(f_0 + \varepsilon_n g_n \mid \mathcal{F}) - f_0}{\varepsilon_n} = \Pi(g \mid T_{\mathcal{F}}(f_0)).$$

We use the phrase 'one-sided Hadamard differentiability' because the sequence ε_n can only approach zero from the positive side. For instance, in

$\mathbb{R}$, the projection on $[0, \infty)$ is the function $x \mapsto \max(x, 0)$, a function whose left and right-hand derivatives at zero are different.

THEOREM 3.4. *If the initial estimator sequence $\hat{f}_n$ satisfies* (3.2), *with ε_n a positive sequence tending to zero and with g a random element in $\mathbb{H}$, then*

$$\frac{\hat{f}_n^{\mathrm{p}} - f_0}{\varepsilon_n} \rightsquigarrow \Pi\left(g \mid T_{\mathcal{F}}(f_0)\right).$$

Moreover, if the integer sequence m_n is such that (3.1) *holds, then also*

$$\frac{\hat{f}_{n,m_n}^{\mathrm{p}} - f_0}{\varepsilon_n} \rightsquigarrow \Pi\left(g \mid T_{\mathcal{F}}(f_0)\right).$$

Since $0 \in T_{\mathcal{F}}(f_0)$, Lemma 2.2(iii) implies

$$\|g\|^2 \geq \|g - \Pi(g \mid T_{\mathcal{F}}(f_0))\|^2 + \|\Pi(g \mid T_{\mathcal{F}}(f_0))\|^2 . \tag{3.3}$$

In this sense, the limiting random variable for the projection estimator sequence is 'smaller' than the one for the initial estimator sequence.

4 CONVEX FUNCTIONS

Let $\mathcal{C}$ be the set of real-valued, convex and Lebesgue square-integrable functions on the positive half-line, $(0, \infty)$. Such functions are necessarily non-negative and non-increasing, and they can diverge at zero. Since convex functions on $(0, \infty)$ that are equal almost everywhere must in fact be equal everywhere, we can view $\mathcal{C}$ as a subset of $L^2 = L^2((0, \infty), \mathrm{d}x)$.

First we will study the properties of the projection operator from L^2 onto $\mathcal{C}$. We will derive a number of characterizations which are reminiscent of Lemma 2.2 in Groeneboom, Jongbloed and Wellner [13]. Then we will show how to actually compute the projection through solving a quadratic program after approximating convex functions by piece-wise linear ones on appropriate non-equidistant grids. Finally, we will point out the relation between the projection approach and the least-squares estimator of a convex density in [13].

4.1 Characterizing the projection

The set $\mathcal{C}$ is an L^2-closed and convex cone; in particular, the projection operator on $\mathcal{C}$ is well-defined. Only the property that $\mathcal{C}$ is closed requires some thought. Let $(f_n)_n$ be a sequence in $\mathcal{C}$ for which there exists $g \in L^2$ such that $\int (f_n - g)^2 \to 0$. Then, along some subsequence $(n_k)_k$, we have $f_{n_k}(x) \to g(x)$ for almost every $x \in (0, \infty)$; see for instance [9], p. 68 and p. 90. By appropriately redefining the function g on the null set of those x for which the previous convergence does not hold, we can construct $f \in \mathcal{C}$

such that $f = g$ almost everywhere. Since necessarily $\int (f_n - f)^2 \to 0$, we find that $\mathcal{C}$ is indeed closed.

By Lemma 2.1, for $h \in L^2$, there exists a unique

$$f = \arg\min_{g \in \mathcal{C}} \int (g - h)^2,$$

denoted by $f = \Pi(h \mid \mathcal{C})$. We are interested in the relation between f and h.

First, let us recall a few elementary facts about convex functions; see for instance [25], section 24. For every $f \in \mathcal{C}$ and every $0 < x < \infty$, the right-hand derivative

$$f'(x) := \lim_{\varepsilon \downarrow 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}, \qquad \text{for } 0 < x < \infty,$$

is well-defined. The function f' is non-positive, non-decreasing, vanishes at infinity, and f can be recovered from f' through $f(x) = -\int_x^\infty f'$ for $0 < x < \infty$. Moreover, f' is right-continuous with limits from the left; denote $f'(x-) = \lim_{y \uparrow x} f'(y)$ for $0 < x < \infty$. The differential $\mathrm{d}f'$ defines a non-negative measure on $(0, \infty)$ given by $\int_{(x,\infty)} \mathrm{d}f' = -f'(x)$ for $0 < x < \infty$. Moreover, by Fubini's theorem, the original function f can be expressed in terms of this measure through

$$f(x) = \int_{(0,\infty)} (z - x)_+ \, \mathrm{d}f'(z), \qquad \text{for } 0 < x < \infty. \tag{4.1}$$

Lemma 4.1. *Let $h \in L^2$ and $f \in \mathcal{C}$. For $0 < x < \infty$, put*

$$\begin{aligned} G(x) &= \int_0^x \int_0^y \{f(z) - h(z)\} \mathrm{d}z \, \mathrm{d}y \\ &= \int_0^\infty (x - z)_+ \{f(z) - h(z)\} \mathrm{d}z. \end{aligned}$$

The following are equivalent:

(i) *$f = \Pi(h \mid \mathcal{C})$;*

(ii) *$\int (f - h) f = 0$ and $G \geq 0$;*

(iii) *$\int_{\{G > 0\}} \mathrm{d}f' = 0$ and $G \geq 0$.*

Characterization (iii) in Lemma 4.1 is similar to the one of the least-squares estimator of a convex density in Groeneboom *et al.* [13].

We conclude this subsection with pointing out some relations between an arbitrary $h \in L^2$ and its projection $f = \Pi(h \mid C)$. First we mention some interesting things occurring at a point x in which f is not linear.

LEMMA 4.2. *Let $h \in L^2$ and $f = \Pi(h \mid \mathcal{C})$ with right-derivative f'. If $0 < x < \infty$ is such that $f'(s) < f'(t)$ for all $0 < s < x < t < \infty$, then*

$$\int_0^x f = \int_0^x h, \quad \int_0^x y f(y) \mathrm{d}y = \int_0^x y h(y) \mathrm{d}y, \quad \textit{and} \quad \int_0^x f^2 = \int_0^x hf.$$

If, additionally, h is continuous in x, then $f(x) \geq h(x)$.

Second, if h is a square-integrable probability density on $(0, \infty)$, then its projection $f = \Pi(h \mid \mathcal{C})$ is guaranteed to be a probability density as well. This property is a special case of (ii) in the following lemma.

LEMMA 4.3. *Let $h \in L^2$ and $f = \Pi(h \mid \mathcal{C})$.*

(i) If $\int |h| < \infty$, then $\int h \leq \int_{\{f>0\}} h = \int f$. In particular, $\int f \leq \int |h|$.

(ii) If $h \geq 0$, then $\int h = \int f$, finite or infinite, and $h\mathbf{1}_{\{f=0\}} = 0$ almost everywhere.

Finally, if $h \in L^2$ is uniformly bounded, then so is $f = \Pi(h \mid \mathcal{C})$.

LEMMA 4.4. *For $h \in L^2$ and $f = \Pi(h \mid \mathcal{C})$, we have $\|f\|_\infty \leq 2 \|h\|_\infty$.*

4.2 Piece-wise linear approximations

For positive integer m, let $\mathcal{C}_m$ be the set of those $f \in \mathcal{C}$ that are piece-wise linear with knots on the grid $\{(k/m)^2 : k = 0, 1, \ldots, m^2\}$. Necessarily $f \in \mathcal{C}_m$ vanishes on $[m^2, \infty)$. The following lemma asserts how well elements in $\mathcal{C}$ can be approximated by those in $\mathcal{C}_m$.

LEMMA 4.5. *Let $f \in \mathcal{C}$ be uniformly bounded and integrable. For every positive integer m, there exists $f_m \in \mathcal{C}_m$ such that*

$$\int (f_m - f)^2 \leq \frac{1}{m^2} \left(\|f\|_1^2 + \frac{8}{3} \|f\|_\infty^2 \right).$$

Since functions in $\mathcal{C}$ have the steepest slope near the origin, it makes sense to choose the grid in such a way that the density of points there is highest. If instead we would have taken the equidistant grid $\{k/m : k = 0, 1, \ldots, m^2\}$, then the approximation error in Lemma 4.5 would in general only be of the order $O(m^{-1})$. Lemma 4.5 leads to the following result on how to choose the integer m for the projections on $\mathcal{C}$ and $\mathcal{C}_m$ to be as close as desired.

LEMMA 4.6. *Let $(h_n)_n$ be a sequence in L^2 such that $\|h_n\|_\infty = O_p(1)$ and $\|h_n\|_1 = O_p(1)$. If*

$$\|h_n - \Pi(h_n \mid \mathcal{C})\|_2 = O_p(\varepsilon_n) \qquad \textit{and} \qquad \varepsilon_n m_n \to \infty$$

for some positive sequence $(\varepsilon_n)_n$ *and some positive integer sequence* $(m_n)_n$*, then*

$$\|\Pi(h_n \mid \mathcal{C}_{m_n}) - \Pi(h_n \mid \mathcal{C})\|_2 = o_p(\varepsilon_n).$$

Computing the projection of an arbitrary $h \in L^2$ on $\mathcal{C}_m$ amounts to solving a quadratic program of the form (2.5). For $k = 0, 1, 2, \ldots, m^2 - 1$, let $h_{k,m}$ be the real-valued function on $[0, \infty)$ which vanishes on $[m^2, \infty)$, is piece-wise linear on $[0, m^2]$ with knots on the grid $\{(l/m)^2 : l = 0, 1, \ldots, m^2\}$ and such that $h_{k,m}\big((k/m)^2\big) = 1$ and $h_{k,m}\big((l/m)^2\big) = 0$ for integer $l \neq k$. Any $f \in \mathcal{C}_m$ can be written as $f = \sum_{k=0}^{m^2-1} \lambda_k h_{k,m}$ with $\lambda_k = f\big((k/m)^2\big)$. In this way, we can identify $\mathcal{C}_m$ with the set Λ_m of m^2-dimensional vectors $\boldsymbol{\lambda}$ such that

$$\begin{aligned}
&(2k+1)\lambda_{k-1} - 4k\lambda_k + (2k-1)\lambda_{k+1} \geq 0, \qquad \text{for } k = 1, \ldots, m^2 - 2, \\
&(2m^2 - 1)\lambda_{m^2-2} - 4(m^2 - 1)\lambda_{m^2-1} \geq 0, \\
&\lambda_{m^2-1} \geq 0.
\end{aligned}$$

The m^2-by-m^2 matrix $\boldsymbol{A}$ in (2.5) with entries $A_{k,l} = \int h_{k,m} h_{l,m}$ is given by

$$\begin{aligned}
A_{0,0} &= 1/(3m^2), \\
A_{k,k} &= 4k/(3m^2), \qquad \text{for } k = 1, 2, \ldots, m^2 - 1, \\
A_{k,k+1} &= A_{k+1,k} = (2k+1)/(6m^2), \qquad \text{for } k = 0, 1, \ldots, m^2 - 2, \\
A_{k,l} &= 0, \qquad \text{if } |k - l| \geq 2.
\end{aligned}$$

Finally, the entries of the m^2-dimensional vector $\boldsymbol{b}$ in (2.5) are $b_k = \int h_{k,m} h$, where $h \in L^2$ is the function to be projected onto $\mathcal{C}_m$.

4.3 Estimating a convex probability density

Let $X_1, \ldots, X_n$ be an independent sample from a probability density function (pdf) f on the positive half-line, which is to be estimated under the assumption that it belongs to $\mathcal{C}$. This estimation problem was considered by Groeneboom, Jongbloed and Wellner (GJW) in [13], where the following estimator was proposed:

$$\hat{f}_n^{\mathrm{GJW}} = \arg\min_{g \in \mathcal{C}} Q_n(g) \tag{4.2}$$

where, for $g \in \mathcal{C}$,

$$Q_n(g) = \frac{1}{2} \int g^2 - \frac{1}{n} \sum_{i=1}^{n} g(X_i).$$

In [13], it is shown that $\hat{f}_n^{\mathrm{GJW}}$ is well-defined and is indeed a genuine pdf. The GJW estimator can be thought of as the projection of the empirical distribution of the sample onto $\mathcal{C}$. A related but different estimator is the non-parametric maximum likelihood estimator [1, 13].

Denote the order statistics of the sample by $0 < X_{1:n} < \ldots < X_{n:n}$. Let $\tilde{f}_n$ be the pdf of the distribution that puts mass $1/(n-1)$ uniformly on each of the intervals $(X_{i:n}, X_{i+1:n}]$, where $i = 1, \ldots, n-1$. This $\tilde{f}_n$ is a rather naive, even inconsistent, estimate of f. Still, let $\hat{f}_n^{\mathcal{C}} = \Pi(\tilde{f}_n \mid \mathcal{C})$ be the projection of $\tilde{f}_n$ onto $\mathcal{C}$. Since the distribution corresponding to $\tilde{f}_n$ is close to the empirical distribution of the sample, one may conjecture that $\hat{f}_n^{\mathcal{C}}$ and $\hat{f}_n^{\mathrm{GJW}}$ are close as well. This turns out to be the case; we do not even have to assume that the true density is convex.

LEMMA 4.7. *If $X_1, \ldots, X_n$ is a random sample from a uniformly bounded density f on $(0, \infty)$, then*

$$\int \left(\hat{f}_n^{\mathcal{C}} - \hat{f}_n^{\mathrm{GJW}}\right)^2 = O_p(n^{-1}).$$

The situation is different from but still similar to the one for the Grenander estimator for a monotone density [12], which is at the same time both the non-parametric maximum likelihood estimator as well as the projection of (a slightly different version of) the naive density estimator on the space of monotone functions, see for instance [34], chapter 24.

5 EXTREME-VALUE COPULAS

A copula is a multivariate distribution function with uniform$(0,1)$ margins [29, 21]. The statistical relevance of copulas is their role in the margin-free modelling of the dependence structure of a general multivariate distribution.

An interesting class of copulas is that of extreme-value copulas. The following representation for bivariate extreme-value copulas was discovered by Pickands in [22], building upon the seminal paper [14] by de Haan and Resnick; see also Deheuvels's instructive account [7]. A bivariate copula C is an extreme-value copula if and only if it admits the representation

$$C(u,v) = \exp\left\{\log(uv) A\left(\frac{\log(u)}{\log(uv)}\right)\right\}, \qquad \text{for } u, v \in (0,1).$$

The above expression defines a genuine copula if and only if the Pickands dependence function $A : [0,1] \to \mathbb{R}$ satisfies the following two properties:

(A1) A is convex;

(A2) $\max(t, 1-t) \le A(t) \le 1$ for all $t \in [0,1]$.

The class of Pickands dependence functions is denoted in the sequel by $\mathcal{A}$.

Estimation of the Pickands dependence function of an extreme-value copula is an important step in the analysis of multivariate extremes ([2], chapter 9). Non-parametric methods avoid the model risk associated with parametric methods like in [33], but unfortunately, the estimates they generate typically fail to satisfy (A1) or (A2) above. *Ad hoc* remedies are sometimes applied to enforce these properties [15, 30], but the effect of such modifications on the performance of the estimator remains most of the time unclear.

Our projection methodology yields a way to force such a non-parametric estimate $\hat{A}_n$ into $\mathcal{A}$ in a controllable and well-understood way. As $\mathcal{A}$ is a closed and convex subset of $L^2 = L^2([0,1], \mathrm{d}x)$ (with the usual identification of functions that differ on a null set only), the projection of $\hat{A}_n$ onto $\mathcal{A}$ is well-defined, yielding the projection estimator

$$\hat{A}_n^{\mathrm{p}} = \Pi(\hat{A}_n|\mathcal{A}) = \operatorname*{arg\,min}_{A \in \mathcal{A}} \int (\hat{A}_n - A)^2.$$

In the sequel, we focus on the computation and the asymptotic behavior of the projection estimator. Its performance will be illustrated by means of a small simulation study.

5.1 Piece-wise linear approximations

For positive integer m, let $\mathcal{A}_m$ be the class of Pickands dependence functions that are piece-wise linear with knots on the grid $\{k/m : k = 0, 1, \ldots, m\}$. The projection onto $\mathcal{A}_m$ can be computed by solving a quadratic program with linear constraints as in section 2.4. Each A in $\mathcal{A}_m$ admits the representation $A = \sum_{i=0}^m A(i/m) h_{im}$, where h_{im} is the unique piece-wise linear function on $[0,1]$ with knots on the given grid such that $h_{im}(i/m) = 1$ and $h_{im}(j/m) = 0$ for integer $j \neq i$.

By (A1)–(A2), a piece-wise linear function A on $[0,1]$ with knots on the given grid is a Pickands dependence function if and only if

$$\begin{cases} A(0) = A(1) = 1, \\ A(0) - A(\frac{1}{m}) \le \frac{1}{m}, \\ A(\frac{i-1}{m}) - 2A(\frac{i}{m}) + A(\frac{i+1}{m}) \ge 0, \qquad \text{for } i = 1, \ldots, m-1, \\ A(1) - A(\frac{m-1}{m}) \le \frac{1}{m}. \end{cases}$$

These constraints define the subset Λ_m of $\mathbb{R}^{m+1}$ to which the vector $\boldsymbol{\lambda}$ in (2.5) with entries $\lambda_i = A(i/m)$ should belong. The $(m+1)$-by-$(m+1)$ matrix $\boldsymbol{A}$ with elements $A_{ij} = \int h_{im} h_{jm}$ is easily computed to be

$$\boldsymbol{A} = \frac{1}{6m} \begin{pmatrix} 2 & 1 & & & & 0 \\ 1 & 4 & 1 & & & \\ & 1 & 4 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 4 & 1 \\ 0 & & & & 1 & 2 \end{pmatrix}. \tag{5.1}$$

Given an initial estimate $\hat{A}$, computing the vector $\boldsymbol{b}$ involves integrals of the form $b_i = \int h_{im}\hat{A}$. A simple procedure consists of interpolating the integrand on each interval of the form $[(i-1)/m, i/m]$ by a quadratic polynomial with interpolation points $(i-1)/m$, $(i-1/2)/m$ and i/m, resulting in the approximation

$$b_i \approx \begin{cases} \dfrac{1}{3m}\left\{\dfrac{1}{2}\hat{A}(0) + \hat{A}\left(\dfrac{1}{2m}\right)\right\} & \text{if } i = 0, \\ \dfrac{1}{3m}\left\{\hat{A}\left(\dfrac{2i-1}{2m}\right) + \hat{A}\left(\dfrac{i}{m}\right) + \hat{A}\left(\dfrac{2i+1}{2m}\right)\right\} & \text{if } i = 1, \ldots, m-1, \\ \dfrac{1}{3m}\left\{\hat{A}\left(\dfrac{2m-1}{2m}\right) + \dfrac{1}{2}\hat{A}(1)\right\} & \text{if } i = m. \end{cases}$$

Of course, in order this approximation to $\boldsymbol{b}$ is sufficiently accurate, we would need detailed sample-path properties of $\hat{A}$. We do not pursue this issue here and simple take the above formula as a convenient computational tool. In the further theoretical analysis, we assume that the vector $\boldsymbol{b}$ is computed exactly.

For the projection onto the subclass $\mathcal{A}_m$ to be close to the projection onto the full class $\mathcal{A}$, we need to have an upper bound on the Hausdorff distance between $\mathcal{A}_m$ and $\mathcal{A}$ (Lemma 3.1).

LEMMA 5.1. *For $A \in \mathcal{A}$, let $A_m \in \mathcal{A}_m$ be the Pickands dependence function obtained by linearly interpolating A at $\{0, 1/m, 2/m, \ldots, 1\}$. Then*

$$\int (A_m - A)^2 \leq 2m^{-5/2}.$$

Piece-wise linear functions are special cases of splines. Smoother estimates may be obtained with higher-order splines. However, in view of the results in section 3, there will be no difference asymptotically. Indeed, in the experiments we have run, the difference in output between fitting piece-wise linear functions or fitting higher-order splines was most of the time negligible.

5.2 Asymptotics

Let $\hat{A}_n$ be a sequence of estimators for an unknown Pickands dependence function A. Assume that the asymptotics of $\hat{A}_n$ are known and of the form

$$\frac{\hat{A}_n - A}{\varepsilon_n} \rightsquigarrow G \qquad \text{in } L^2, \tag{5.2}$$

with $0 < \varepsilon_n \to 0$. Typically, the convergence above holds in the stronger topology of uniform convergence. The limit process G is a centered Gaussian process, and ε_n is equal to the reciprocal of the square root of the effective sample size, that is, the number of block maxima [8, 5, 27] or the number of high-threshold exceedances [10].

By Theorem 3.4, equation (5.2) implies that the asymptotic distribution of the projection estimator $\hat{A}_n^{\mathrm{p}} = \Pi(\hat{A}_n \mid \mathcal{A})$ is given by

$$\frac{\hat{A}_n^{\mathrm{p}} - A}{\varepsilon_n} \rightsquigarrow \Pi\left(G \mid T_{\mathcal{A}}(A)\right) \qquad \text{in } L^2. \tag{5.3}$$

Here, $T_{\mathcal{A}}(A)$ is the tangent cone of $\mathcal{A}$ at A, defined as the set of limits (in L^2) of all sequences of the form $\lambda_n(A_n - A)$, with $\lambda_n \in [0, \infty)$ and $A_n \in A$. Furthermore, if m_n is a positive integer sequence such that $\varepsilon_n^{4/5} m_n \to \infty$, then, by Theorem 3.4, the approximate projection estimator $\hat{A}_{n,m_n}^{\mathrm{p}} = \Pi(\hat{A}_n \mid \mathcal{A}_{m_n})$ has the same asymptotic distribution as the projection estimator,

$$\frac{\hat{A}_{n,m_n}^{\mathrm{p}} - A}{\varepsilon_n} \rightsquigarrow \Pi\left(G \mid T_{\mathcal{A}}(A)\right) \qquad \text{in } L^2. \tag{5.4}$$

According to equation (3.3),

$$\int G^2 \geq \int \left\{\Pi\left(G \mid T_{\mathcal{A}}(A)\right)\right\}^2,$$

suggesting the improved performance of the projection estimator in comparison to the initial estimator. The amount of improvement depends on the tangent cone $T_{\mathcal{A}}(A)$: the smaller this tangent cone, the larger the improvement. In a number of special cases, the tangent cone admits an explicit description.

Lemma 5.2. *(i) If A is twice differentiable and $\inf_t A''(t) > 0$, then $T_{\mathcal{A}}(A) = L^2$.*
(ii) If $A = 1$, then $T_{\mathcal{A}}(A)$ is the set of non-positive, convex functions.

In particular, in case (i), the projection estimator has the same asymptotic distribution as the initial estimator on which it is based. In case (ii) of independent margins, the integrated squared error of the least-squares estimator will typically be smaller than the one of the corresponding initial estimator.

5.3 Simulation study

In order to illustrate the finite-sample properties of the projection estimator, we generated data from bivariate extreme-value distributions with the following Pickands dependence functions:

- the independent copula, $A(t) = 1$ (Figure 1, first row);
- the Gumbel or logistic model, $A(t) = \{t^{1/\alpha} + (1-t)^{1/\alpha}\}^{\alpha}$, with parameter value $\alpha = 0.9$ (Figure 1, second row);

- the asymmetric logistic model, $A(t) = \{(\psi_1(1-t))^{1/\alpha} + (\psi_2 t)^{1/\alpha}\}^\alpha + (\psi_1 - \psi_2)t + 1 - \psi_1$, with parameter vector $(\alpha, \psi_1, \psi_2) = (0.7, 0.5, 0.1)$ (Figure 1, third row).

In each case, 500 samples were generated of size $n = 100$. The performance of the estimators is visualised by their normalized point-wise root mean squared errors (RMSE), defined as

$$\left(\frac{n}{500} \sum_{i=1}^{500} \left\{ A_n^{(i)}(t) \right\}^2 \right)^{1/2}, \qquad \text{for } t \in [0,1],$$

this quantity being an estimate of the standard deviation of the limit distribution of $n^{1/2}\{A_n(t) - A(t)\}$, with A_n representing any of the estimators considered.

For the initial estimators we took the Capéraà-Fougères-Genest (CFG) estimator [5] with tuning parameter $p(t) = t$ and the Hall-Tajvidi (HT) estimator [15]. The marginal distributions were estimated by the empirical distribution functions.

For the actual computations, we implemented the approximation method of section 5.1. As the rate of convergence of the CFG and HT estimators is $O_p(n^{-1/2})$, their projections on the full class $\mathcal{A}$ and on the subclasses $\mathcal{A}_{m_n}$ will be asymptotically undistinguishable as soon as $n^{-2/5} m_n \to \infty$; see equation (5.4). In our simulations, we experimented with several values of m. It turned out that as soon as $m \geq 20$, there was no visible difference anymore between the approximate projection estimators. In Figure 1, only the results for $m = 20$ are shown.

For each model considered and every $t \in [0,1]$, the least-squares estimator had a smaller root mean squared error than the initial estimator on which it was based. In accordance to Lemma 5.2(ii), the improvement was largest in case of independence.

6 SUPPORT FUNCTIONS OF CONVEX SETS

Let C be a compact, convex subset of $\mathbb{R}^2$. The support function $f : (0, 2\pi] \to \mathbb{R}$ of C is defined by

$$f(\theta) = \sup_{x \in C} x' e(\theta), \qquad \text{for } \theta \in (0, 2\pi], \tag{6.1}$$

where $e(\theta) = (\cos\theta, \sin\theta)'$. The set C can be recovered from its support function through the relation

$$C = \bigcap_{\theta \in (0, 2\pi]} \{x \in \mathbb{R}^2 : x' e(\theta) \leq f(\theta)\}. \tag{6.2}$$

Imagine a situation where noisy measurements of the support function f are available; the aim is to estimate C. This estimation problem arises

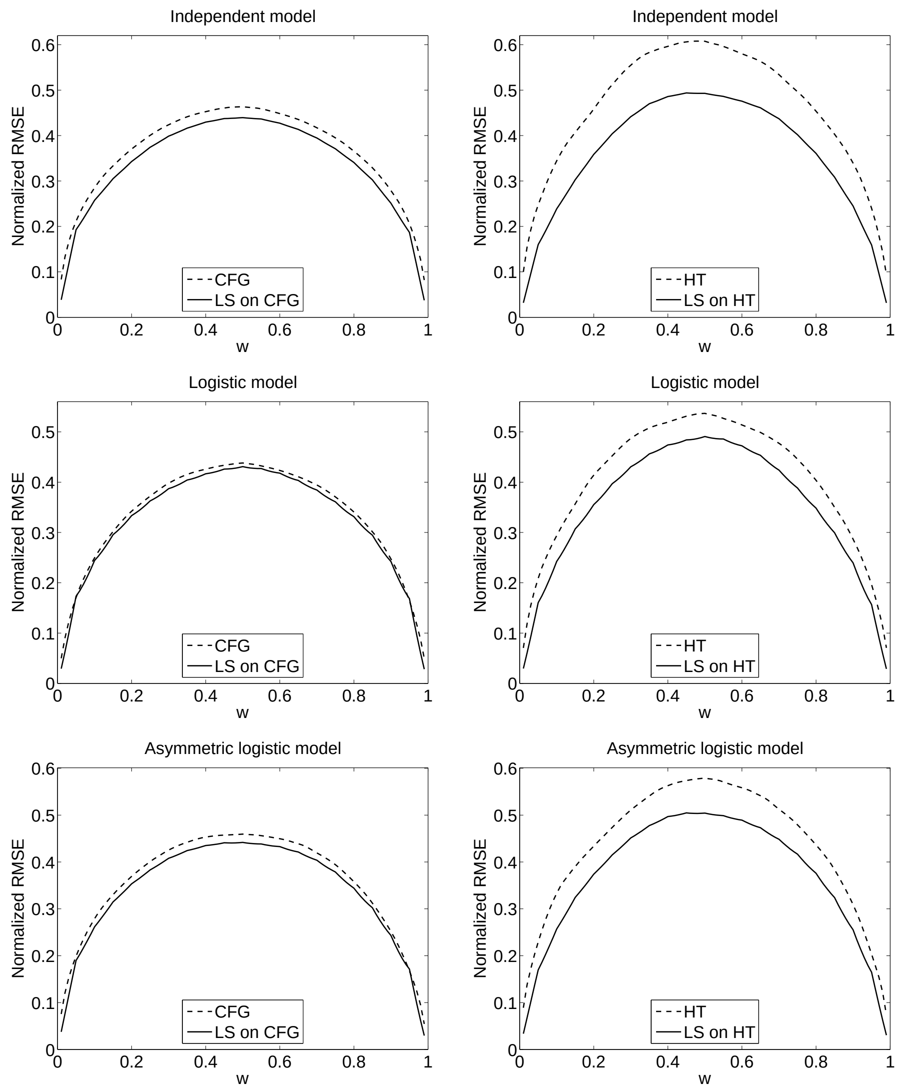

Figure 1: Normalized point-wise RMSE for the original Pickands dependence function estimators (dashed line: CFG on the left, HT on the right) and their projection versions (full line). First row: independent model; second row: logistic model; third row: asymmetric logistic model.

for instance in medical imaging [31] and robotic vision [17]. From an estimate $\hat{f}$ of f, an estimate $\hat{C}$ of C can be obtained by replacing f by $\hat{f}$ in equation (6.2). Even if $\hat{f}$ is itself not a support function, still, the estimate $\hat{C}$, being the intersection of half-planes, will be convex. However, ensuring that $\hat{f}$ is itself a genuine support function is likely to improve its accuracy and thus also the one of $\hat{C}$.

In this section, we show how to cast the support-function estimation problem in our set-up. The actual computation is done by projecting on the subclass of support functions of convex m-gons with faces in regulary spaced, fixed directions. Moreover, we derive how large the number of faces, m, should be at least for the approximation to be sufficiently accurate.

6.1 Basics

Let $L^2 = L^2((0, 2\pi], \mathrm{d}\theta)$ be the real Hilbert space of real-valued, Lebesgue square-integrable functions on the interval $(0, 2\pi]$. Let $\mathcal{S}$ be the set of support functions, that is, all real-valued functions f on $(0, 2\pi]$ for which there exists a compact, convex subset C of $\mathbb{R}^2$ such that (6.1) holds. Since support functions are uniformly bounded, we can view $\mathcal{S}$ as a subset of L^2. As usual, we tacitly identify functions that are equal almost everywhere.

The set $\mathcal{S}$ is a convex cone. For, if f and g are the support functions of the convex, compact sets C and D, respectively, and if $\lambda \in [0, \infty)$, then λf is the support function of the set $\lambda C = \{\lambda x : x \in C\}$ and $f + g$ is the support function of the Minkowski sum of C and D, defined as $C + D = \{x + y : x \in C,\, y \in D\}$, see for instance [3].

The set $\mathcal{S}$ is also closed in L^2. The argument relies on the following characterization of support functions: a function $f : (0, 2\pi] \to \mathbb{R}$ belongs to $\mathcal{S}$ if and only if the function $h_f : \mathbb{R}^2 \to \mathbb{R}$ defined by $h_f\left((0,0)'\right) = 0$ and $h_f(\rho e(\theta)) = \rho f(\theta)$ (for $0 < \rho < \infty$ and $\theta \in (0, 2\pi]$) is subadditive, that is, $h_f(x) + h_f(y) \le h_f(x + y)$ for all $x, y \in \mathbb{R}^2$; see [3], chapter 4. (As the function h_f is by definition positively homogeneous, it is then necessarily also convex.) The argument for closedness of $\mathcal{S}$ now runs along the same lines as the corresponding argument for the class $\mathcal{C}$ in the beginning of section 4.1. In the special case that f is twice continuously differentiable, a necessary and sufficient condition for f to be a support function is that $f + f'' \ge 0$; see e.g. [26], p. 2.

Given a square-integrable initial estimate $\hat{f}$ of an unknown support function f, the corresponding projection or least-squares estimator can now be defined by

$$\hat{f}^{\mathrm{p}} = \Pi(\hat{f} \mid \mathcal{S}) = \operatorname*{arg\,min}_{s \in \mathcal{S}} \int \left(s - \hat{f}\right)^2. \tag{6.3}$$

Computing the projection estimator requires solving an infinite-dimensional optimization problem. Alternatively, the method in section 2.4 prescribes to replace the full class $\mathcal{S}$ by a smaller, finite-dimensional one, reducing the optimization problem to a quadratic program.

6.2 Approximations by polygons

What we need is a sequence of subsets $\mathcal{S}_m$ of $\mathcal{S}$ that provide with increasing positive integer m increasingly accurate approximations of $\mathcal{S}$. Piece-wise linear functions, used for instance in sections 4 and 5, are useless here, since such functions are not support functions.

Instead, a natural class to consider is the class $\mathcal{S}_m$ of support functions g of polygons P with m faces perpendicular to the equidistant grid of directions $\theta_{i,m} = (i/m)2\pi$ for $i = 1, \ldots, m$, a class already considered in [17]. A support function g in $\mathcal{S}_m$ is completely determined by its values on $\{\theta_{i,m} : i = 1, \ldots, m\}$, as the polygon P it corresponds to is given by

$$P = \bigcap_{i=1}^{m} \{x \in \mathbb{R}^2 : x' e(\theta_{i,m}) \leq g(\theta_{i,m})\}. \tag{6.4}$$

The vertices $v_1, \ldots, v_m$ of P are given by

$$v_i = \frac{1}{\sin\left(\frac{2\pi}{m}\right)} \begin{pmatrix} \sin(\theta_{i+1,m}) g(\theta_{i,m}) - \sin(\theta_{i,m}) g(\theta_{i+1,m}) \\ -\cos(\theta_{i+1,m}) g(\theta_{i,m}) + \cos(\theta_{i,m}) g(\theta_{i+1,m}) \end{pmatrix}. \tag{6.5}$$

Since P is equal to the convex hull of its vertices, the value of its support function g in $\theta \in [\theta_{i,m}, \theta_{i+1,m}]$ is $g(\theta) = \max_{j=1,\ldots,m} v_j' e(\theta) = v_i' e(\theta)$ or

$$g(\theta) = \frac{1}{\sin\left(\frac{2\pi}{m}\right)} \left\{ \sin(\theta_{i+1,m} - \theta) g(\theta_{i,m}) + \sin(\theta - \theta_{i,m}) g(\theta_{i+1,m}) \right\}. \tag{6.6}$$

Note that g is a piece-wise trigonometric function with knots on the grid $\{\theta_{i,m} : i = 1, \ldots, m\}$. Since every vertex v_i must be contained in the intersection of the two half-spaces $\{x \in \mathbb{R}^2 : x' e(\theta) \leq g(\theta)\}$ for $\theta \in \{\theta_{i-1,m}, \theta_{i+1,m}\}$, a necessary and sufficient condition on a vector $\boldsymbol{g} = (g_1, \ldots, g_m)'$ to be equal to $(g(\theta_{1,m}), \ldots, g(\theta_{m,m}))'$ for some $g \in \mathcal{S}_m$ is that for $i = 1, \ldots, m$,

$$g_{i-1} \sin(2\pi/m) - g_i \sin(4\pi/m) + g_{i+1} \sin(2\pi/m) \geq 0, \tag{6.7}$$

where $g_0 = g_m$ and $g_{m+1} = g_1$; see [17], equation (3). In particular, $\mathcal{S}_m$ is a closed and convex cone in L^2.

Now, given an initial estimate $\hat{f}$ of an unknown support function f, with f not necessarily in $\mathcal{S}_m$, the approximate projection or least-squares estimator is defined by

$$\hat{f}_m^{\mathrm{p}} = \Pi(\hat{f} \mid S_m) = \operatorname*{arg\,min}_{g \in \mathcal{S}_m} \int \left(g - \hat{f}\right)^2. \tag{6.8}$$

The two questions to be answered then are, first, how to compute $\hat{f}_m^{\mathrm{p}}$, and second, how large to take m. We treat these two questions in turn.

Since, by definition, $\hat{f}_m^{\mathrm{p}}$ belongs to $\mathcal{S}_m$, it is of the form stated in equation (6.6) for some vector $\boldsymbol{g} \in \mathbb{R}^m$ with $g_i = g(\theta_{i,m}) = \hat{f}_m^{\mathrm{p}}(\theta_{i,m})$. Substituting the right-hand side of equation (6.6) into the integral on the right-hand side of equation (6.8) and expanding the quadratic function, we obtain after some calculations that the vector $\boldsymbol{g}$ is the solution to the quadratic program

$$\text{minimize } \boldsymbol{g}'\boldsymbol{A}\boldsymbol{g} - 2\boldsymbol{g}'\boldsymbol{b} \text{ subject to (6.7).}$$

The m-by-m matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \begin{pmatrix} 2a & b & & & b \\ b & 2a & b & & \\ & \ddots & \ddots & \ddots & \\ & & b & 2a & b \\ b & & & b & 2a \end{pmatrix}$$

(entries not mentioned explicitly being zero), where

$$\begin{aligned} a &= \frac{1}{\sin^2\left(\frac{2\pi}{m}\right)} \left\{ \frac{\pi}{m} - \frac{1}{4} \sin\left(\frac{4\pi}{m}\right) \right\}, \\ b &= -a \cos\left(\frac{2\pi}{m}\right) + \frac{1}{2} \sin\left(\frac{2\pi}{m}\right), \end{aligned}$$

and where the vector $\boldsymbol{b} = (b_1, \ldots, b_m)'$ is given by

$$b_i = \int_{\theta_{i-1,m}}^{\theta_{i,m}} \frac{\sin(\theta - \theta_{i-1,m})}{\sin\left(\frac{2\pi}{m}\right)} \hat{f}(\theta) \mathrm{d}\theta + \int_{\theta_{i,m}}^{\theta_{i+1,m}} \frac{\sin(\theta_{i+1,m} - \theta)}{\sin\left(\frac{2\pi}{m}\right)} \hat{f}(\theta) \mathrm{d}\theta,$$

the interval $[2\pi, \theta_{m+1,m}]$ to be identified with the interval $[0, \theta_{1,m}]$. The entries of the vector $\boldsymbol{b}$ have to be calculated by numerical quadrature. A simple approximation consists of replacing $\sin(u)$ by u for $0 \le u \le 2\pi/m$ and to interpolate the resulting integrands by quadratic polynomials in the mid-points and end-points of each interval $[\theta_{i,m}, \theta_{i+1,m}]$, resulting in

$$b_i \approx \frac{2\pi}{3m} \left\{ \hat{f}\left((i - 1/2)\frac{2\pi}{m}\right) + \hat{f}\left(i\frac{2\pi}{m}\right) + \hat{f}\left((i + 1/2)\frac{2\pi}{m}\right) \right\}.$$

Next, we treat the question how large one should take the integer m. The following lemma quantifies how closely $\mathcal{S}$ can be approximated by $\mathcal{S}_m$ (proof of this and the following lemma in Appendix A.4).

LEMMA 6.1. *For all $c \in (0, \infty)$ and all integer $m \ge 3$,*

$$\sup_{f \in \mathcal{S}: \|f\|_\infty \le c} \inf_{g \in \mathcal{S}_m} \|f - g\|_\infty \le \frac{6\pi c}{m}.$$

Although the constant 6π in the upper bound is not optimal, the order $O(m^{-1})$ cannot be improved upon, not even if we replace the L^∞-norm by the L^2-norm.

From Lemmas 3.1 and 6.1, we can deduce a minimal sufficient speed at which a positive integer sequence m_n should tend to infinity so that the distance between $\Pi(h_n \mid \mathcal{S})$ and $\Pi(h_n \mid \mathcal{S}_{m_n})$ for a given (random) sequence h_n converges to zero at a specified rate.

LEMMA 6.2. *Let h_n be a sequence of random elements in L^2. If $\|h_n\|_\infty = O_p(1)$ and*

$$\|h_n - \Pi(h_n \mid \mathcal{S})\|_2 = O_p(\varepsilon_n) \qquad \textit{and} \qquad \varepsilon_n m_n \to \infty$$

for some positive sequence ε_n and some positive integer sequence m_n, then

$$\|\Pi(h_n \mid \mathcal{S}_{m_n}) - \Pi(h_n \mid \mathcal{S})\|_2 = o_p(\varepsilon_n).$$

The condition $\|h_n - \Pi(h_n \mid \mathcal{S})\|_2 = O_p(\varepsilon_n)$ in Lemma 6.2 is implied by $\|h_n - f\|_2 = O_p(\varepsilon_n)$ for some $f \in \mathcal{S}$. Hence, what is needed to apply Lemma 6.2 on a sequence of initial estimators $\hat{f}_n$ is a result on its rate of convergence in integrated squared error sense.

6.3 Simulations

By way of illustration, we now present a small simulation example. The compact, convex set to be estimated is the ellipse $E = \{(x,y)' : x^2/a^2 + y^2/b^2 \leq 1\}$ with support function $f(\theta) = (a^2 \cos^2\theta + b^2 \sin^2\theta)^{1/2}$ for $\theta \in (0, 2\pi]$. We consider two cases: $(a,b) = (1,1)$, the unit disk, and $(a,b) = (4,1)$, a strongly eccentric ellipse with the longer major axis in horizontal position. Data are observed as pairs (θ_i, Y_i), $1, \ldots, n$. The θ_i form an independent sample uniformly distributed on $(0, 2\pi]$, while $Y_i = f(\theta_i) + \varepsilon_i$, the errors ε_i being sampled independently from the normal distribution with mean zero and standard deviation 0.25.

The initial estimator $\hat{f}_n$ is taken to be the one in equation (3) in Fisher *et al.* [11], that is, a circular analog of a local linear smooth. Its kernel is chosen to be the von Mises density on the circle, $K_\kappa(\theta) = C(\kappa)\exp(\kappa\cos\theta)$. The smoothing parameter $\kappa > 0$ plays the same role as the inverse of the square of the bandwidth for a kernel density estimator on the line. The normalizing constant $C(\kappa)$ does not appear in the final expression of the estimator.

We project this initial estimator on the space $\mathcal{S}_m$ of support functions of convex m-gons with faces perpendicular to the directions $(i/m)2\pi$, $i = 1, \ldots, m$. If $\kappa = \kappa_n$ is of the order $n^{2/5}$, then the L^2-convergence rate of the initial estimator sequence $\hat{f}_n$ is the optimal one, $n^{-2/5}$; see [11], Theorem 4.1. For such κ_n, if the positive integer sequence m_n satisfies $n^{-2/5}m_n \to \infty$, then, by Lemma 6.2, the integrated squared difference between the projections of $\hat{f}_n$ on $\mathcal{S}_{m_n}$ and $\mathcal{S}$ is of smaller order than $n^{-2/5}$.

So for sample size $n = 500$, the number of faces, m, should be much larger than $n^{2/5} \approx 12$. Figure 2 shows the results for $n = 500$ and $m = 25$. We experimented with larger m, only to get visually undistinguishable results. The top line shows the true support functions and the observations. Secondly, the middle line shows the true support functions, the Fisher *et al.* [11] smooths and their least-squares versions. Finally, the bottom line shows the true shapes together with the polygonal shapes reconstructed from the projection estimates of the support functions. The smoothing parameter for the Fisher *et al.* [11] estimator was taken as $\kappa = n^{2/5} \approx 12$ for the circle (left column) and $\kappa = 200$ for the ellipse (right column), corresponding to bandwidths $h = \kappa^{-1/2}$ equal to 0.29 and 0.07, respectively. The reason for the relatively small bandwidth of 0.07 in the second case is to keep the bias under control; see also [19], section 2.3. For larger bandwidths, the initial Fisher *et al.* smooth is itself already a genuine support function, but the bias of the estimate is of larger order than its standard deviation, ruining the estimate, see Figures 1 and 2 and Theorem 4.1 in [11].

A PROOFS

A.1 Proofs for section 3

PROOF OF LEMMA 3.1. By Lemma 2.2(iii), since $\Pi(h \mid \mathcal{G}) \in \mathcal{G} \subset \mathcal{F}$,

$$\|\Pi(h \mid \mathcal{F}) - \Pi(h \mid \mathcal{G})\|^2 \le \|h - \Pi(h \mid \mathcal{G})\|^2 - \|h - \Pi(h \mid \mathcal{F})\|^2 .$$

Since $\Pi(\Pi(h \mid \mathcal{F}) \mid \mathcal{G}) \in \mathcal{G}$ and by an application of the triangle inequality and the definition of δ,

$$\begin{aligned}\|h - \Pi(h \mid \mathcal{G})\| &\le \|h - \Pi(\Pi(h \mid \mathcal{F}) \mid \mathcal{G})\| \\ &\le \|h - \Pi(h \mid \mathcal{F})\| + \delta.\end{aligned}$$

Combine the previous two displays to get

$$\begin{aligned}\|\Pi(h \mid \mathcal{F}) - \Pi(h \mid \mathcal{G})\|^2 &\le \{\|h - \Pi(h \mid \mathcal{F})\| + \delta\}^2 - \|h - \Pi(h \mid \mathcal{F})\|^2 \\ &= 2\delta \, \|h - \Pi(h \mid \mathcal{F})\| + \delta^2.\end{aligned}$$

Take square roots on both sides of the display to finish the proof. $\square$

PROOF OF LEMMA 3.3. By Lemma 2.3, we have

$$\left\| \frac{\Pi(x + \varepsilon_n y_n \mid \mathcal{F}) - x}{\varepsilon_n} - \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} \right\| \le \|y_n - y\| .$$

Hence, without loss of generality we may take assume $y_n = y$ for all n.

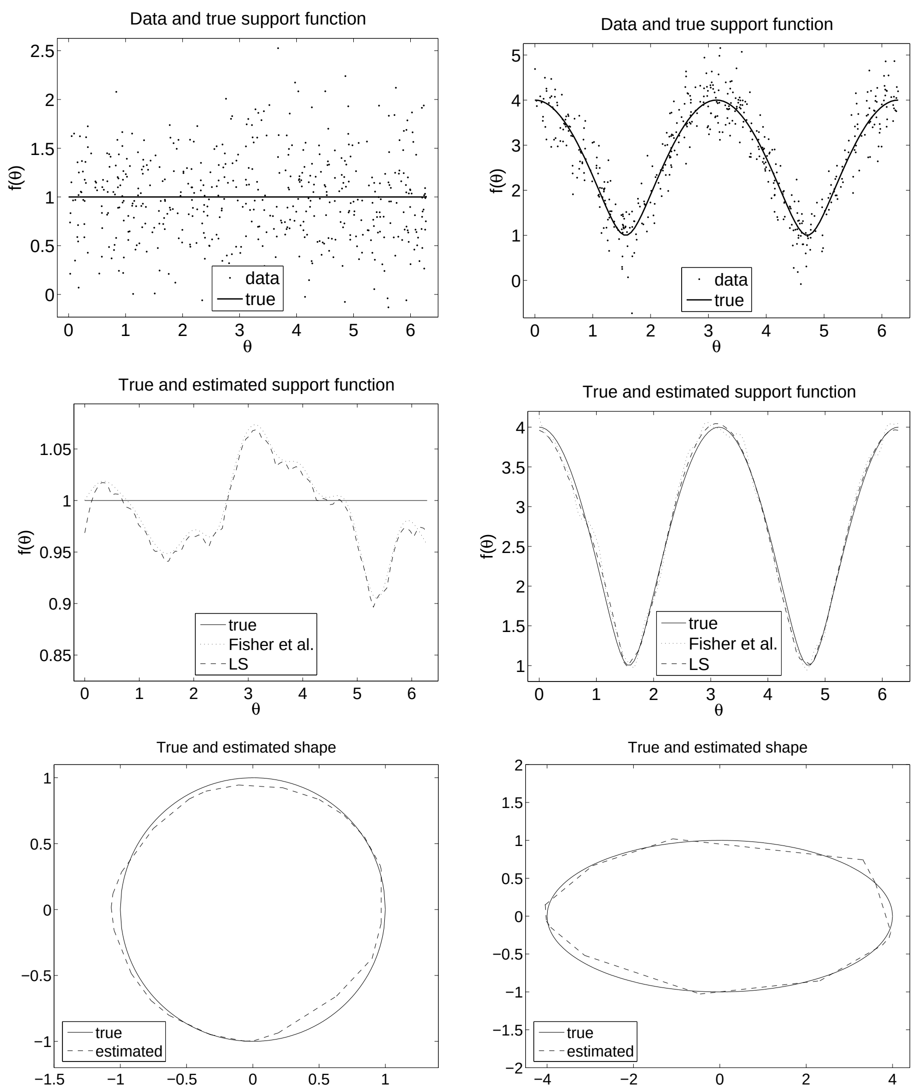

Figure 2: *Top:* Observations and true support functions. *Middle:* Support functions (full: true; dotted: initial estimate; dashed: projection estimate). *Bottom:* True shape (full) and shape reconstructed from projection estimate of support function (dashed). Ellipses with parameter values $(a, b) = (1, 1)$ (left) and $(4, 1)$ (right).

First, assume that $y \in T_{\mathcal{F}}(x)$. If $y = 0$ there is nothing to prove, so suppose $y \neq 0$. There exists a sequence $z_n = \lambda_n(x_n - x)$ with $\lambda_n > 0$ and $x_n \in \mathcal{F}$ such that $z_n \to y$ as $n \to \infty$. Without loss of generality we may assume that $\lambda_n \varepsilon_n \leq 1$. We have

$$\begin{aligned}
&\left\| \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} - y \right\| \\
&\quad \leq \quad \left\| \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} - \frac{\Pi(x + \varepsilon_n z_n \mid \mathcal{F}) - x}{\varepsilon_n} \right\| \\
&\qquad\quad + \left\| \frac{\Pi(x + \varepsilon_n z_n \mid \mathcal{F}) - x}{\varepsilon_n} - z_n \right\| + \|z_n - y\| .
\end{aligned}$$

By Lemma 2.3, the first term on the right is at most $\|y - z_n\|$. Moreover, $x + \varepsilon_n z_n = x + \varepsilon_n \lambda_n (x_n - x) \in \mathcal{F}$, so the middle term on the right-hand side of the previous display vanishes. Hence

$$\left\| \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} - y \right\| \leq 2 \|z_n - y\| \to 0,$$

as required.

Next, take a general $y \in \mathbb{H}$ and put $v = \Pi(y \mid T_{\mathcal{F}}(x))$ and $v_n = \varepsilon_n^{-1}\{\Pi(x + \varepsilon_n v \mid \mathcal{F}) - x\}$. We have

$$\begin{aligned}
\left\| y - \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} \right\| &= \varepsilon_n^{-1} \|(x + \varepsilon_n y) - \Pi(x + \varepsilon_n y \mid \mathcal{F})\| \\
&\leq \varepsilon_n^{-1} \|(x + \varepsilon_n y) - \Pi(x + \varepsilon_n v \mid \mathcal{F})\| \\
&= \|y - v_n\| .
\end{aligned}$$

By the previous paragraph, $v_n \to v$, whence

$$\limsup_{n \to \infty} \left\| y - \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} \right\| \leq \|y - v\| .$$

On the other hand, since $\varepsilon_n^{-1}\{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x\} \in T_{\mathcal{F}}(x)$ and $v = \Pi(y \mid T_{\mathcal{F}}(x))$, Lemma 2.2(iii) implies

$$\left\| y - \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} \right\|^2 \geq \|y - v\|^2 + \left\| v - \frac{\Pi(x + \varepsilon_n y \mid \mathcal{F}) - x}{\varepsilon_n} \right\|^2 .$$

Combine the previous two displays to obtain that the second term on the right-hand side of the last display converges to zero, as required. □

A.2 Proofs for section 4

Proof of Lemma 4.1. *(i) and (ii) are equivalent.* By Lemma 2.2(iv), (i) is equivalent to $\int (f - h) f = 0$ and $\int (f - h) g \geq 0$ for all $g \in \mathcal{C}$. For fixed

$0 < x < \infty$, the function $z \mapsto (x-z)_+$ belongs to $\mathcal{C}$, whence (i) implies (ii). Conversely, let $g \in \mathcal{C}$ with right-derivative g'. By (4.1) and Fubini's theorem,

$$\begin{aligned}\int (f-h)g &= \int_0^\infty \{f(x)-h(x)\} \int_{(0,\infty)} (y-x)_+ \mathrm{d}g'(y)\,\mathrm{d}x \\ &= \int_{(0,\infty)} G(y)\,\mathrm{d}g'(y).\end{aligned}$$

Hence $G \geq 0$ implies $\int (f-h)g \geq 0$.

(ii) and (iii) are equivalent. Choose g equal to f in the previous display to get $\int (f-h)f = \int_{(0,\infty)} G\,\mathrm{d}f'$. Hence, if $G \geq 0$ and $\int (f-h)f = 0$, then necessarily $\int_{\{G>0\}} \mathrm{d}f' = 0$. Conversely, if $G \geq 0$ and $\int_{\{G>0\}} \mathrm{d}f' = 0$, then $\int (f-h)f = 0$. □

Proof of Lemma 4.2. Define G as in Lemma 4.1. Since $G \geq 0$, G is continuous, and $\int G\mathrm{d}f' = 0$, necessarily $G(x) = 0$. Since moreover G is absolutely continuous with continuous derivative $G'(t) = \int_0^t (f-h)$, also $G'(x) = 0$. This gives the first two equalities. Further, for all $0 < t < \infty$,

$$\begin{aligned}\int_0^t (f-h)f &= \int_0^t \{f(u)-h(u)\} \int_{(0,\infty)} (v-u)_+\,\mathrm{d}f'(v)\,\mathrm{d}u \\ &= \int_{(0,\infty)} \int_0^t \{f(u)-h(u)\}(v-u)_+\,\mathrm{d}u\,\mathrm{d}f'(v) \\ &= \int_{(0,t]} G(v)\,\mathrm{d}f'(v) \\ &\quad + \int_{(t,\infty)} \int_0^t \{f(u)-h(u)\}\{(v-t)+(t-u)\}\,\mathrm{d}u\,\mathrm{d}f'(v) \\ &= 0 + G'(t)f(t) - G(t)f'(t).\end{aligned}$$

Hence $\int_0^x (f-h)f = 0$. Finally, if h is continuous in x, then by Taylor's theorem, $0 \leq G(x+y) = \frac{1}{2}y^2\{f(x)-h(x)\} + o(y^2)$ as $y \to 0$, whence $f(x) \geq h(x)$. □

Proof of Lemma 4.3. Let G be as in Lemma 4.1. Since $G \geq 0$, also $G(x)/x \geq 0$ for every $0 < x < \infty$, whence

$$\int_0^\infty (1-y/x)_+ f(y)\mathrm{d}y \geq \int_0^\infty (1-y/x)_+ h(y)\mathrm{d}y, \qquad \text{for } 0 < x < \infty.$$

Let $x \to \infty$; for the left-hand side, apply the monotone convergence theorem, while on the right-hand side, apply the dominated convergence theorem in case (i) and the monotone convergence theorem in case (ii) to get $\int f \geq \int h$.

If $f = 0$, then $\int h \leq 0 = \int_{\{f>0\}} h = \int f$. In case (ii), this also implies $h = 0$ almost everywhere.

So assume that f is not identically zero. Let $x_f = \sup\{x : f'(x) < 0\}$.

If $x_f < \infty$, then $f'(s) < 0 = f'(t)$ for all $0 < s < x_f < t < \infty$, whence, by Lemma 4.2, $\int f = \int_0^{x_f} f = \int_0^{x_f} h = \int_{\{f>0\}} h$; if $h \geq 0$, then, since also $\int h \leq \int f$, necessarily $\int_{\{f=0\}} h = 0$, whence $h\mathbf{1}_{\{f=0\}} = 0$ almost everywhere.

If $x_f = \infty$, then $f > 0$ and $f' < 0$ on $(0, \infty)$. Since $f'(x) \to 0$ as $x \to \infty$, there exists by Lemma 4.2 a sequence $0 < x_n \to \infty$ such that $\int_0^{x_n} f = \int_0^{x_n} h$ for all n. Let $n \to \infty$ to get $\int f = \int h$ in both cases (i) and (ii). $\square$

Proof of Lemma 4.4. Without loss of generality, assume $\|h\|_\infty < \infty$. Fix $0 < x < \infty$. We will show that $f(x) \leq 2\|h\|_\infty$. If $f(x) = 0$, there is nothing to prove, so assume $f(x) > 0$. Define $f_x \in \mathcal{C}$ by $f_x(y) = f(x+y)$ for $0 < y < \infty$. By Lemma 2.2(ii) we have $\int (h - f)(f - f_x) \geq 0$, or

$$\int h(f - f_x) \geq \int f(f - f_x).$$

Since $f - f_x \geq 0$,

$$\int h(f - f_x) \leq \|h\|_\infty \int (f - f_x) = \|h\|_\infty \int_0^x f.$$

On the other hand,

$$\begin{aligned}
\textstyle\int f(f - f_x) &\geq \textstyle\int f^2 - \left(\int f^2\right)^{1/2} \left(\int f_x^2\right)^{1/2} \\
&= \textstyle\left(\int f^2\right)^{1/2} \left(\left(\int f^2\right)^{1/2} - \left(\int_x^\infty f^2\right)^{1/2}\right) \\
&= \textstyle\left(\int f^2\right)^{1/2} \dfrac{\int_0^x f^2}{\left(\int f^2\right)^{1/2} + \left(\int_x^\infty f^2\right)^{1/2}} \\
&\geq \frac{1}{2} \int_0^x f^2 \geq \frac{1}{2} f(x) \int_0^x f.
\end{aligned}$$

Combine the last three displays to conclude the proof. $\square$

Proof of Lemma 4.5. Let $h = \big(f - f(m^2)\big)_+$. Define the function f_m on $[0, \infty)$ as the linear interpolation of h on $\{(k/m)^2 : k = 0, 1, \ldots, m^2\}$ and vanishing on $[m^2, \infty)$. Clearly, f_m belongs to $\mathcal{C}_m$. We have

$$\int (f_m - f)^2 \leq 2 \int (f_m - h)^2 + 2 \int (h - f)^2.$$

Now

$$\int (h - f)^2 = m^2 f^2(m^2) + \int_{m^2}^\infty f^2.$$

Since f is convex, it must stay above its tangent lines, whence $2tf(t) \leq \int f$ or $f(t) \leq (2t)^{-1} \int f$ for all $t > 0$. Substitute this inequality into the previous display to get

$$\int (h - f)^2 \leq \frac{1}{2m^2} \left(\int f\right)^2.$$

In the remainder of the proof, we derive an upper bound for $\int (f_m - h)^2$. The following notations will be useful: Put $p = m^2$ and, for $k = 0, 1, \ldots, p+1$, put $t_k = (k/m)^2$ and $h_k = h(t_k)$. Further, for $k = 0, 1, \ldots, p$, put $\Delta t_k = t_{k+1} - t_k$ and $\Delta h_k = h_{k+1} - h_k$.

Since h is convex, we have for $k = 0, 1, \ldots, p-1$ and $t_k < t \leq t_{k+1}$,

$$\begin{aligned} h(t) &\leq h(t_{k+1}) + \frac{\Delta h_k}{\Delta t_k}(t - t_{k+1}) = f_m(t), \\ h(t) &\geq h(t_{k+1}) + \frac{\Delta h_{k+1}}{\Delta t_{k+1}}(t - t_{k+1}), \end{aligned}$$

whence

$$0 \leq f_m(t) - h(t) \leq \left(\frac{\Delta h_{k+1}}{\Delta t_{k+1}} - \frac{\Delta h_k}{\Delta t_k} \right) (t_{k+1} - t).$$

Integrating over t yields

$$\begin{aligned} \int (f_m - h)^2 &\leq \frac{1}{3} \sum_{k=0}^{p-1} \left(\frac{\Delta h_{k+1}}{\Delta t_{k+1}} - \frac{\Delta h_k}{\Delta t_k} \right)^2 (\Delta t_k)^3 \\ &= \frac{1}{3} \sum_{k=0}^{p-1} (\Delta t_k \Delta h_{k+1} - \Delta t_{k+1} \Delta h_k)^2 \frac{\Delta t_k}{(\Delta t_{k+1})^2}. \end{aligned}$$

Since, by convexity, every $\Delta t_k \Delta h_{k+1} - \Delta t_{k+1} \Delta h_k$ is non-negative, the inequality $\sum_k a_k b_k \leq (\sum_k a_k)(\sum_k b_k)$ for non-negative a_k and b_k yields

$$\begin{aligned} \int (f_m - h)^2 &\leq \frac{1}{3} \sum_{k=0}^{p-1} (\Delta t_k \Delta h_{k+1} - \Delta t_{k+1} \Delta h_k) \\ &\quad \times \sum_{k=0}^{p-1} \left(\left(\frac{\Delta t_k}{\Delta t_{k+1}} \right)^2 \Delta h_{k+1} - \frac{\Delta t_k}{\Delta t_{k+1}} \Delta h_k \right). \end{aligned} \tag{A.1}$$

Since $h_p = h_{p+1} = 0$, the first summation on the right-hand side of the previous display is equal to

$$\begin{aligned} &\Delta t_1 h_0 + \{\Delta t_2 - (\Delta t_0 + \Delta t_1)\} h_1 \\ &\quad + \sum_{k=2}^{p-1} \{(\Delta t_{k+1} - \Delta t_k) - (\Delta t_{k-1} - \Delta t_{k-2})\} h_k. \end{aligned}$$

As the function $k \mapsto t_k$ is a second-order polynomial in k, all the terms in the summation on the second line of the previous display vanish. Since also $h_0 \geq h_1$ and $\Delta t_2 > \Delta t_0 + \Delta t_1$,

$$\sum_{k=0}^{p-1} (\Delta t_k \Delta h_{k+1} - \Delta t_{k+1} \Delta h_k) \leq (\Delta t_2 - \Delta t_0) h_0 = \frac{4}{m^2} h_0.$$

Furthermore, since h is non-increasing and Δt_k is increasing in k, the second factor on the right-hand side of (A.1) is bounded by

$$\begin{aligned}\sum_{k=0}^{p-1}\left(\left(\frac{\Delta t_k}{\Delta t_{k+1}}\right)^2 \Delta h_{k+1} - \frac{\Delta t_k}{\Delta t_{k+1}}\Delta h_k\right) &\leq \sum_{k=0}^{p-1}\frac{\Delta t_k}{\Delta t_{k+1}}(h_k - h_{k+1}) \\ &\leq \sum_{k=0}^{p-1}(h_k - h_{k+1}) = h_0.\end{aligned}$$

Summing up, we obtain

$$\int (f_m - h)^2 \leq \frac{4}{3m^2}h_0^2 \leq \frac{4}{3m^2}f^2(0).$$

Combine the bounds for $\int(f_m - h)^2$ and $\int(h - f)^2$ to finish the proof. $\square$

Proof of Lemma 4.6. By Lemma 3.1, for all positive integer m,

$$\|\Pi(h_n \mid \mathcal{C}_m) - \Pi(h_n \mid \mathcal{C})\|_2 \leq \left(2\delta_m \|h_n - \Pi(h_n \mid \mathcal{C})\|_2 + \delta_m^2\right)^{1/2}$$

where

$$\delta_m = \|\Pi(h_n \mid \mathcal{C}) - \Pi(\Pi(h_n \mid \mathcal{C}) \mid \mathcal{C}_m)\|_2 .$$

By Lemma 4.3(i), $\|\Pi(h_n \mid \mathcal{C})\|_1 \leq \|h_n\|_1 = O_p(1)$. Moreover, by Lemma 4.4, $\|\Pi(h_n \mid \mathcal{C})\|_\infty \leq 2\|h_n\|_\infty = O_p(1)$. Now, by Lemma 4.5, for all positive integer m,

$$\delta_m^2 \leq \frac{1}{m^2}\left(\|h_n\|_1^2 + \frac{8}{3}\|h_n\|_\infty^2\right).$$

Since $1/m_n = o_p(\varepsilon_n)$, we find $\delta_{m_n} = o_p(\varepsilon_n)$. Combine this with the first display in this proof to arrive at the stated result. $\square$

Proof of Lemma 4.7. Note that

$$\hat{f}_n^{\mathcal{C}} = \arg\min_{g\in\mathcal{C}} \int (g - \tilde{f}_n)^2 = \arg\min_{g\in\mathcal{C}} \tilde{Q}_n(g)$$

with

$$\tilde{Q}_n(g) = \frac{1}{2}\int g^2 - \int g\tilde{f}_n, \qquad \text{for } g \in \mathcal{C}.$$

For $g \in \mathcal{C}$, we have on the one hand

$$\begin{aligned}\int g\tilde{f}_n &= \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{1}{X_{i+1:n} - X_{i:n}}\int_{X_{i:n}}^{X_{i+1:n}} g \\ &\leq \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{g(X_{i:n}) + g(X_{i+1:n})}{2} \\ &= \frac{1}{2(n-1)}g(X_{1:n}) + \frac{1}{n-1}\sum_{i=2}^{n-1}g(X_{i:n}) + \frac{1}{2(n-1)}g(X_{n:n}),\end{aligned}$$

while on the other hand

$$\int g\tilde{f}_n \geq \frac{1}{n-1}\sum_{i=2}^{n} g(X_{i:n}).$$

Hence, for $g \in \mathcal{C}$,

$$\begin{aligned}
&Q_n(g) - \tilde{Q}_n(g) \\
&\quad= \int g\tilde{f}_n - \frac{1}{n}\sum_{i=1}^{n} g(X_{i:n}) \\
&\quad\leq -\frac{n-2}{2(n-1)n} g(X_{1:n}) + \frac{1}{(n-1)n}\sum_{i=2}^{n-1} g(X_{i:n}) - \frac{n-2}{2(n-1)n} g(X_{n:n}) \\
&\quad\leq \frac{1}{(n-1)n}\sum_{i=\lfloor n/2\rfloor+1}^{n-1} g(X_{i:n}) \leq \frac{1}{2n} g(X_{\lfloor n/2\rfloor:n})
\end{aligned}$$

as well as

$$Q_n(g) - \tilde{Q}_n(g) \geq \frac{1}{n-1}\sum_{i=2}^{n} g(X_{i:n}) - \frac{1}{n}\sum_{i=1}^{n} g(X_{i:n}) \geq -\frac{1}{n} g(X_{1:n}).$$

By Lemma 2.2(iii),

$$\int \left(\hat{f}_n^{\mathrm{GJW}} - \tilde{f}_n\right)^2 \geq \int \left(\hat{f}_n^{\mathrm{GJW}} - \hat{f}_n^{\mathcal{C}}\right)^2 + \int \left(\hat{f}_n^{\mathcal{C}} - \tilde{f}_n\right)^2,$$

whence

$$\frac{1}{2}\int \left(\hat{f}_n^{\mathrm{GJW}} - \hat{f}_n^{\mathcal{C}}\right)^2 \leq \tilde{Q}_n(\hat{f}_n^{\mathrm{GJW}}) - \tilde{Q}_n(\hat{f}_n^{\mathcal{C}}).$$

Combine this with the upper and lower bounds on $Q_n - \tilde{Q}_n$ above to get

$$\begin{aligned}
&\frac{1}{2}\int \left(\hat{f}_n^{\mathrm{GJW}} - \hat{f}_n^{\mathcal{C}}\right)^2 \\
&\quad\leq Q_n(\hat{f}_n^{\mathrm{GJW}}) + \frac{1}{n}\hat{f}_n^{\mathrm{GJW}}(X_{1:n}) - Q_n(\hat{f}_n^{\mathcal{C}}) + \frac{1}{2n}\hat{f}_n^{\mathcal{C}}(X_{\lfloor n/2\rfloor:n}) \\
&\quad\leq \frac{1}{n}\hat{f}_n^{\mathrm{GJW}}(X_{1:n}) + \frac{1}{2n}\hat{f}_n^{\mathcal{C}}(X_{\lfloor n/2\rfloor:n}).
\end{aligned}$$

From [13], equation (2.6), case $k = 1$, we infer that

$$\hat{f}_n^{\mathrm{GJW}}(0) \leq 2 \max_{i=1,\ldots,n} \frac{i/n}{X_{i:n}}.$$

Moreover, since $\hat{f}_n^{\mathcal{C}}$ is a convex pdf on $(0,\infty)$, necessarily $\hat{f}_n^{\mathcal{C}}(x) \leq 1/(2x)$ for $0 < x < \infty$; see for instance [13], equation (3.1). All in all,

$$\frac{1}{2}\int \left(\hat{f}_n^{\mathrm{GJW}} - \hat{f}_n^{\mathcal{C}}\right)^2 \leq \frac{4}{n} \max_{i=1,\ldots,n} \frac{i/n}{X_{i:n}}.$$

It remains to show that

$$\max_{i=1,\ldots,n} \frac{i/n}{X_{i:n}} = O_p(1).$$

Since $F^{\leftarrow}(u) \geq u/c$ for $0 < u < 1$, where $F^{\leftarrow}(u) = \inf\{x : F(x) \geq u\}$ is the generalized inverse of F and $c = \|f\|_\infty$, the displayed statement holds true for F as soon as it holds in the special case of the uniform distribution. For this case, the property follows from the well-known representation of uniform order statistics as ratios of partial sums of unit exponential random variables, see e.g. [28], proposition 8.2.1, p. 335. □

A.3 Proofs for section 5

Proof of Lemma 5.1. Put $d_k = m\{A(k/m) - A((k-1)/m)\}$ for $k = 1, \ldots, m$ and put $d_0 = -1$. Furthermore, put $e_k = d_k - d_{k-1}$ for $k = 1, \ldots, m$.

Let $t \in (0, 1]$ and let $k = 1, \ldots, m$ be such that $(k-1)/m < t \leq k/m$. By convexity and the fact that $A(t) \geq 1 - t$,

$$A(t) \begin{cases} \leq & A\left(\frac{k-1}{m}\right) + \left(t - \frac{k-1}{m}\right) d_k = A_m(t), \\ \geq & A\left(\frac{k-1}{m}\right) + \left(t - \frac{k-1}{m}\right) d_{k-1}, \end{cases}$$

whence

$$0 \leq A_m(t) - A(t) \leq \left(t - \frac{k-1}{m}\right) e_k.$$

Integrate over t to obtain

$$\int (A_m - A)^2 \leq \frac{1}{3m^3} \sum_{k=1}^m e_k^2.$$

Since $-1 = d_0 \leq d_1 \leq \ldots \leq d_m \leq 1$, we have $0 \leq e_k \leq 2$ for all $k = 1, \ldots, m$. Moreover, as $\sum_{k=1}^m d_k = m\{A(1) - A(0)\} = 0$, we have $\sum_{k=1}^m (m-k+1) e_k = \sum_{k=1}^m \sum_{l=1}^k e_l = \sum_{k=1}^m d_k - m d_0 = m$. Hence

$$\sum_{k=1}^m e_k^2 \leq \sup\left\{\sum_{k=1}^m x_k^2 : (x_1, \ldots, x_m) \in [0,2]^m, \sum_{k=1}^m k x_k = m\right\}.$$

This quadratic program can be solved using the knapsack heuristic, see e.g. [20]. Its solution consists in fixing as many of the first variables as possible at their maximum values, $x_1 = \ldots = x_{\ell_m} = 2$, the next one, x_{ℓ_m+1}, at a value between 0 and 2 such as to ensure $\sum_{k=1}^{\ell_m+1} k x_k = m$, and the remaining variables, the x_i for $i \geq \ell_m + 2$, at zero. Formally, the supremum on the right-hand side of the previous display is therefore attained at the vector $(y_1, \ldots, y_m)$ defined as

$$y_k = \begin{cases} 2 & \text{if } k = 1, \ldots, \ell_m, \\ \dfrac{m}{\ell_m + 1} - \ell_m & \text{if } k = \ell_m + 1, \\ 0 & \text{if } k = \ell_m + 2, \ldots, m, \end{cases}$$

where $\ell_m = \max\{\ell = 1, \ldots, m : \ell(\ell+1) \le m\}$. Since $\ell_m \le m^{1/2}$,

$$\sum_{k=1}^m e_k^2 \le \sum_{k=1}^m y_k^2 \le 4(\ell_m + 1) \le 4(m^{1/2} + 1).$$

In case $m \ge 4$, the right-hand side is bounded by $6m^{1/2}$, yielding $\int (A_m - A)^2 \le 2m^{-5/2}$. In case $m \in \{1, 2, 3\}$, the inequality stated in the lemma is trivially fulfilled as $\int (A_m - A)^2 \le \int_0^1 \{1 - \max(t, 1-t)\}^2 \mathrm{d}t = 1/12$. □

Proof of Lemma 5.2. (i) Fix $f \in L^2$. By the dominated convergence theorem, $\|f - f\mathbf{1}_{[1/n,1-1/n]}\|_2 \to 0$ as $n \to \infty$. Define $g_n \in L^2$ by $g_n(t) = f(t)t^{-2}(1-t)^{-2}\mathbf{1}_{[1/n,1-1/n]}(t)$. Since the set of polynomials is dense in L^2, there exist polynomials p_n such that $\|g_n - p_n\|_2 \to 0$ as $n \to \infty$. Define $q_n \in L^2$ by $q_n(t) = t^2(1-t)^2 p_n(t)$. Then

$$\begin{aligned} \|f - q_n\|_2 &\le \|f - f\mathbf{1}_{[1/n,1-1/n]}\|_2 + \|f\mathbf{1}_{[1/n,1-1/n]} - q_n\|_2 \\ &\le \|f - f\mathbf{1}_{[1/n,1-1/n]}\|_2 + \|g_n - p_n\|_2 \\ &\to 0, \qquad n \to \infty. \end{aligned}$$

Now, let $\lambda_n > 0$ be large enough so that

$$\lambda_n^{-1} \sup\left\{|q_n''(t)| : t \in [0,1]\right\} \le \inf\left\{A''(t) : t \in [0,1]\right\}.$$

Define $A_n = A + \lambda_n^{-1} q_n$. Clearly, $\|f - \lambda_n(A_n - A)\|_2 \to 0$ as $n \to \infty$. Moreover, $A_n \in \mathcal{A}$, as follows from $A_n(0) = A(0) = 1$, $A_n(1) = A(1) = 1$, $A_n'(0) = A'(0) \in [-1, 0]$, $A_n'(1) = A'(1) \in [0, 1]$, and $A_n'' = A'' + \lambda_n^{-1} q_n'' \ge 0$. Together, $f \in T_{\mathcal{A}}(A)$.

(ii) Let f be a convex, non-positive function on $[0, 1]$. For positive integer n, define

$$f_n(x) = \begin{cases} nxf(1/n) & \text{if } 0 \le x \le 1/n, \\ f(x) & \text{if } 1/n \le x \le 1 - 1/n, \\ n(1-x)f(1-1/n) & \text{if } 1 - 1/n \le x \le 1. \end{cases}$$

The function f_n is convex, and since $f \le f_n \le 0$ and $f_n(t) \to f(t)$ for all $0 < t < 1$, also $\|f_n - f\|_2 \to 0$ as $n \to \infty$ by dominated convergence. Let $\lambda_n > 0$ be such that $\lambda_n \ge n \max\{|f(1/n)|, |f(1-1/n)|\}$ and define $A_n = 1 + \lambda_n^{-1} f_n$. Then $\|f - \lambda_n(A_n - 1)\|_2 \to 0$ as $n \to \infty$. Moreover,

$A_n \in \mathcal{A}$ as $A_n(0) = 1$, $A_n(1) = 1$, $A_n'(0) = \lambda_n^{-1} n f(1/n) \in [-1, 0]$, and $A_n'(1) = \lambda_n^{-1} n |f(1 - 1/n)| \in [0, 1]$. Hence $f \in T_{\mathcal{A}}(1)$.

Conversely, let $f \in T_{\mathcal{A}}(1)$. Then $\|f - \lambda_n(A_n - 1)\|_2 \to 0$ for some sequence λ_n of positive numbers and some sequence A_n in $\mathcal{A}$. Each $f_n = \lambda_n(A_n - 1)$ is non-positive and convex. Then, along some subsequence $(n_k)_k$, we have $f_{n_k}(t) \to f(t)$ as $k \to \infty$ for almost every $t \in [0, 1]$ (see [9], p. 68 and p. 90). Hence f is almost everywhere equal to some non-positive, convex function. □

A.4 Proofs for section 6

Proof of Lemma 6.1. By the scaling property of support functions, we can without loss of generality assume that $c = 1$. Let C be a compact, convex set in $\mathbb{R}^2$ with support function f. The assumption that $\|f\|_\infty \le 1$ means that C is contained in the unit disk.

Fix an integer $m \ge 3$. Let g be the support function in $\mathcal{S}_m$ given by the right-hand side of equation (6.6) with $g(\theta_{i,m}) = f(\theta_{i,m})$. Note that g is the support function of the polygon P in equation (6.4). Since C is contained in P, necessarily $f \le g$.

Because C is compact and the function $x \mapsto x'e(\theta)$ is continuous, there exist $x_i \in C$ such that $x_i' e(\theta_{i,m}) = f(\theta_{i,m})$ for $i = 1, \ldots, m$. Further, let $v_1, \ldots, v_m$ be the vertices of P. The support function g is determined by these vertices through equation (6.6).

Let $\theta \in (0, 2\pi]$. There exists $i = 1, \ldots, m$ such that $\theta \in (\theta_{i-1,m}, \theta_{i,m}]$. Since $v_i' e(\theta_{i,m}) = g(\theta_{i,m}) = f(\theta_{i,m}) = x_i' e(\theta_{i,m})$,

$$\begin{aligned} 0 \le g(\theta) - f(\theta) &\le (v_i - x_i)' e(\theta) \\ &= (v_i - x_i)' \left(e(\theta) - e(\theta_{i,m})\right) \\ &\le \|v_i - x_i\| \, \|e(\theta) - e(\theta_{i,m})\| , \end{aligned}$$

where $\|y\| = \sqrt{y'y}$ is the Euclidean norm of $y \in \mathbb{R}^2$. Since $\theta - \theta_{i,m}$ is the length of the arc of the unit circle connecting $e(\theta_{i,m})$ and $e(\theta)$,

$$\|e(\theta) - e(\theta_{i,m})\| \le \theta - \theta_{i,m} \le \frac{2\pi}{m}.$$

Moreover, writing $v_i = \|v_i\| \, e(\alpha_i)$, we have by the fact that $v_i \in P$,

$$\begin{aligned} 1 &\ge f(\theta_{j,m}) = g(\theta_{j,m}) \\ &\ge v_i' e(\theta_{j,m}) = \|v_i\| \cos(\alpha_i - \theta_{j,m}), \qquad \text{for } j = 1, \ldots, m. \end{aligned}$$

There exists j such that $|\alpha_i - \theta_{j,m}| \le \pi/m$. For this j, the inequality in the above display implies $\|v_i\| \le 1/\cos(\pi/m) \le 1/\cos(\pi/3) = 2$. By the triangle inequality and the fact that x_i lies in the unit disk, we arrive at

$$0 \le g(\theta) - f(\theta) \le \frac{6\pi}{m}.$$

Since θ was arbitrary, the proof of the lemma is finished. □

LEMMA A.1. *There exists a positive constant c such that $\|\Pi(h \mid \mathcal{S})\|_\infty \leq c\,\|h\|_\infty$ for every uniformly bounded, measurable $h : (0, 2\pi] \to \mathbb{R}$.*

PROOF. Put $a = \|h\|_\infty$ and $f = \Pi(h \mid \mathcal{S})$. If $\|f\|_\infty \leq a$, then there is nothing to prove, so assume that $\|f\|_\infty > a$. Let C be the convex, compact set with support function f. There exist $1 < b < \infty$ and $0 < \alpha \leq 2\pi$ such that $x = ab\, e(\alpha)$ belongs to C.

On the one hand, as the zero function is the support function of the origin in $\mathbb{R}^2$, necessarily

$$\int_0^{2\pi} (f - h)^2 \leq \int_0^{2\pi} h^2 \leq 2\pi a^2.$$

On the other hand, $f(\theta) \geq x' e(\theta) = ab \cos(\theta - \alpha)$ for all $\theta \in (0, 2\pi]$, whence

$$\int_0^{2\pi} (f - h)^2 \geq \int_0^{2\pi} \{ab \cos(\theta - \alpha) - a\}_+^2 \mathrm{d}\theta = a^2 \int_0^{2\pi} (b \cos\theta - 1)_+^2 \mathrm{d}\theta.$$

Combining the two displays above yields $\int_0^{2\pi} (b \cos\theta - 1)_+^2 \mathrm{d}\theta \leq 2\pi$. Hence, a possible choice for c is the supremum of all $b \in (1, \infty)$ for which the latter inequality holds. □

PROOF OF LEMMA 6.2. The proof is completely analogous to the proof of Lemma 4.6, this time using Lemma A.1 above. □

References

[1] ANEVSKI, D. (2003). Estimating the derivative of a convex density. *Statist. Neerlandica* **57**, 245–257.

[2] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications.* Wiley, New York.

[3] BONNESEN, T., and FENCHEL, W. (1987). *Theory of convex bodies.* BCS, Moscow, Idaho.

[4] BOWMAN, A.W. and AZZALINI, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford University Press, Oxford.

[5] CAPÉRAÀ, P., FOUGÈRES, A.L., and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**, 567–577.

[6] CODDINGTON, E.A. and CARLSON, R. (1997). *Linear Ordinary Differential Equations.* SIAM, Philadelphia.

[7] Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In: *Statistical Extremes and Applications* (ed. J. Tiago de Oliveira), p. 117–130. Reidel.

[8] Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statist. Probab. Lett.* **12**, 429–439.

[9] Doob, J.L. (1994). *Measure theory.* Springer-Verlag, New York.

[10] Einmahl, J.H.J., de Haan, L., and Piterbarg, V.I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29**, 1401–1423.

[11] Fisher, N.I., Hall, P., Turlach, B.A., and Watson, G.S (1997). On the estimation of a convex set from noisy data on its support function. *J. Amer. Statist. Assoc.* **92**, 84-91.

[12] Grenander, U. (1956). On the theory of mortality measurement, part II. *Skandinavisk Aktuarietidskrift* **39**, 129–153.

[13] Groeneboom, P., Jongbloed, G., and Wellner, J. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29**, 1653–1698.

[14] de Haan, L. and Resnick, S.I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40**, 317–337.

[15] Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* **6**, 835–844.

[16] Kadison, R.V. and Ringrose, J.R. (1983). *Fundamentals of the Theory of Operator Algebras, Volume I.* Academic Press, New York.

[17] Lele, A.S., Kulkarni, S.R., and Willsky, A.S. (1992). Convex-polygon estimation from support-line measurements and applications to target reconstruction from laser-radar data. *Journal of the Optical Society of America A* **9**, 1693–1714.

[18] Mammen, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19**, 724–740.

[19] Mammen, E., Marron, J.S., Turlach, B.A., and Wand, M.P. (2001). A general projection framework for constrained smoothing. *Statist. Sci.* **16**, 232–248.

[20] Martello, S. and Toth, P. (1990). *Knapsack problems.* Wiley, New York.

[21] Nelsen, R.B. (1999). *An introduction to copulas.* Lecture Notes in Statistics **139**, Springer Verlag, New York.

[22] Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin of the International Statistical Institute*, 859–878.

[23] Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling.* Dekker, New York.

[24] Reboul, L. (2005). Estimation of a function under shape restrictions. Applications to reliability. *Ann. Statist.* **33**, 1330–1356.

[25] Rockafeller, R.T. (1970). *Convex Analysis.* Princeton University Press, Princeton.

[26] Santaló, L.A. (1976). *Integral Geometry and Geometric Probability.* Addison-Wesley, Reading, MA.

[27] Segers, J. (2004). Non-parametric inference for bivariate extreme-value copulas. *CentER Discussion Paper* 2004-91, Tilburg University, www.center.nl.

[28] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[29] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* **8**, 229–231.

[30] Smith, R.L., Tawn, J.A. and Yuen, H.K. (1990). Statistics of multivariate extremes. *International Statistical Review* **58**, 47–58.

[31] Stark, H. and Peng, H. (1988). Shape estimation in computer tomography from minimal data. *Journal of the Optical Society of America A* **5**, 331–343.

[32] Stark, H. and Yang, Y. (1998). *Vector Space Projections.* Wiley, New York.

[33] Tawn, J.A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397–415.

[34] van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Amélie Fils-Villetard, Armelle Guillou
Université Pierre et Marie Curie
Laboratoire de Statistique Théorique et Appliquée
175, rue du Chevaleret, 75013 Paris
France
E-mail: *fils@ccr.jussieu.fr, guillou@ccr.jussieu.fr*

Johan Segers
Tilburg University
Department of Econometrics and Operations Research
PO Box 90153, NL-5000 LE Tilburg
the Netherlands
E-mail: *jsegers@uvt.nl*