

# Existence and uniqueness of weighted generalized $\psi$ -estimators

MÁTYÁS BARCZY<sup>\*, $\diamond$</sup> , ZSOLT PÁLES<sup>\*\*</sup>

\* ELKH-SZTE Analysis and Applications Research Group, Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H-6720 Szeged, Hungary.

\*\* Institute of Mathematics, University of Debrecen, Pf. 400, H-4002 Debrecen, Hungary.

e-mail: barczy@math.u-szeged.hu (M. Barczy), pales@science.unideb.hu (Zs. Páles).

$\diamond$  Corresponding author.

## Abstract

We introduce the notions of generalized and weighted generalized  $\psi$ -estimators as unique points of sign change of some appropriate functions, and we give necessary as well as sufficient conditions for their existence. We also derive a set of sufficient conditions under which the so-called  $\psi$ -expectation function has a unique point of sign change. We present several examples from statistical estimation theory, where our results are well-applicable. For example, we consider the cases of empirical quantiles, empirical expectiles, some  $\psi$ -estimators that are important in robust statistics, and some examples from maximum likelihood theory as well. Further, we introduce Bajraktarević-type (in particular, quasi-arithmetic-type)  $\psi$ -estimators. Our results specialized to  $\psi$ -estimators with a function  $\psi$  being continuous in its second variable provide new results for (usual)  $\psi$ -estimators (also called Z-estimators).

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Notions and results on the existence and uniqueness of weighted generalized <math>\psi</math>-estimators</b>	<b>6</b>
<b>3</b>	<b>Existence and uniqueness of the point of sign change of <math>\psi</math>-expectation functions</b>	<b>14</b>

---

*2020 Mathematics Subject Classifications:* 62F10, 62F35, 26E60, 26A48

*Key words and phrases:* M-estimator,  $\psi$ -estimator, Z-estimator, likelihood equation, existence, uniqueness, Bajraktarević-type  $\psi$ -estimator, empirical quantile, empirical expectile, robust statistics.

Mátyás Barczy is supported by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, project no. TKP2021-NVA-09. Zsolt Páles is supported by the K-134191 NKFIH Grant.

4	Examples from statistical estimation theory	18
5	Proofs for Section 2	28
6	Proofs for Section 3	34
7	Proofs for Section 4	39
	References	41

# 1 Introduction

In statistics, M-estimators play a fundamental role, and a special subclass, the class of  $\psi$ -estimators (also called Z-estimators), is also in the heart of investigations. The M-estimators (where the letter M refers to "maximum likelihood-type") were introduced by Huber [8, 9]. Let  $(X, \mathcal{X})$  be a measurable space,  $\Theta$  be a Borel subset of  $\mathbb{R}$ , and  $\varrho : X \times \Theta \rightarrow \mathbb{R}$  be a function such that for each  $t \in \Theta$ , the function  $X \ni x \mapsto \varrho(x, t)$  is measurable with respect to the sigma-algebra  $\mathcal{X}$ . Let  $(\xi_n)_{n \geq 1}$  be a sequence of i.i.d. random variables with values in  $X$  such that the distribution of  $\xi_1$  depends on an unknown parameter  $\vartheta \in \Theta$ . For each  $n \geq 1$ , Huber [8, 9] introduced a natural estimator of  $\vartheta$  based on the observations  $\xi_1, \dots, \xi_n$  as a solution  $\widehat{\vartheta}_n := \widehat{\vartheta}_n(\xi_1, \dots, \xi_n)$  of the following minimization problem:

$$(1.1) \quad \inf_{t \in \Theta} \frac{1}{n} \sum_{i=1}^n \varrho(\xi_i, t),$$

provided that such a solution exists. One calls  $\widehat{\vartheta}_n$  an M-estimator of the unknown parameter  $\vartheta \in \Theta$  based on the i.i.d. observations  $\xi_1, \dots, \xi_n$ . For historical fidelity, we note that Huber [8] considered the special case when  $X := \mathbb{R}$ ,  $\Theta := \mathbb{R}$ , and the function  $\varrho$  depends only on  $x - t$ , i.e.,  $\varrho(x, t) := f(x - t)$ ,  $x \in \mathbb{R}$ ,  $t \in \Theta$ , with some given nonconstant function  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Turning back to the general case, under suitable regularity assumptions, the minimization problem (1.1) can be solved by setting the derivative of the objective function (with respect to the unknown parameter) equal to zero:

$$\frac{1}{n} \sum_{i=1}^n \partial_2 \varrho(\xi_i, t) = 0, \quad t \in \Theta.$$

In the statistical literature,  $\partial_2 \varrho$  is often denoted by  $\psi$ , and hence in this case the M-estimator is often called  $\psi$ -estimator, while other authors call it a Z-estimator (the letter Z refers to "zero"). For a detailed exposition of M-estimators and  $\psi$ -estimators (Z-estimators), see, e.g., Kosorok [12, Sections 2.2.5 and 13] or van der Vaart [20, Section 5].

Throughout this paper, let  $\mathbb{N} = \mathbb{Z}_{++}$ ,  $\mathbb{Z}_+$ ,  $\mathbb{Q}$ ,  $\mathbb{R}$ ,  $\mathbb{R}_+$  and  $\mathbb{R}_{++}$  denote the sets of positive integers, non-negative integers, rational numbers, real numbers, non-negative real numbers and

positive real numbers, respectively. An interval  $\Theta \subseteq \mathbb{R}$  will be called nondegenerate if it contains at least two distinct points. For a subset  $S \subseteq \mathbb{R}$ , the convex hull of  $S$  (which is the smallest interval containing  $S$ ) is denoted by  $\text{conv}(S)$ . For each  $n \in \mathbb{N}$ , let us also introduce the set  $\Lambda_n := \mathbb{R}_+^n \setminus \{(0, \dots, 0)\}$ . All the random variables are defined on an appropriate probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

To the best of our knowledge, the topic of existence and uniqueness of  $\psi$ -estimators is less addressed in the statistical literature. In the present paper, we are going to investigate two basic problems of this field as presented below. Roughly speaking, Problem 1 is about the existence and uniqueness of the newly introduced notions: generalized  $\psi$ -estimators and weighted generalized  $\psi$ -estimators. Problem 2 is devoted to the existence and uniqueness of a point of sign change for so-called  $\psi$ -expectation functions.

*Problem 1.* Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ . Let  $\Psi(X, \Theta)$  denote the class of real-valued functions  $\psi : X \times \Theta \rightarrow \mathbb{R}$  such that, for each  $x \in X$ , there exist  $t_+, t_- \in \Theta$  such that  $t_+ < t_-$  and  $\psi(x, t_+) > 0 > \psi(x, t_-)$ . Roughly speaking, a function  $\psi \in \Psi(X, \Theta)$  satisfies the following property: for each  $x \in X$ , the function  $t \ni \Theta \mapsto \psi(x, t)$  changes sign on the interval  $\Theta$  at least once. Given a function  $\psi \in \Psi(X, \Theta)$ ,  $n \in \mathbb{N}$  and  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ , let us consider the equation

$$(1.2) \quad \psi_{\mathbf{x}}(t) := \sum_{i=1}^n \psi(x_i, t) = 0, \quad t \in \Theta.$$

More generally, for  $n \in \mathbb{N}$ ,  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$  and  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$ , we also consider the weighted equation

$$(1.3) \quad \psi_{\mathbf{x}, \boldsymbol{\lambda}}(t) := \sum_{i=1}^n \lambda_i \psi(x_i, t) = 0, \quad t \in \Theta.$$

The basic question we are going to investigate now is to find necessary as well as sufficient conditions for the unique solvability of the equations (1.2) and (1.3), respectively. In a broader context, we are going to find necessary as well as sufficient conditions for the existence of a point of sign change (see Definition 2.1) for the functions  $\psi_{\mathbf{x}}$  and  $\psi_{\mathbf{x}, \boldsymbol{\lambda}}$  introduced by (1.2) and (1.3), respectively. It will turn out that the points of sign change in question are unique provided that they exist, and one can call them as a *generalized  $\psi$ -estimator* and *weighted generalized  $\psi$ -estimator*, respectively, for some unknown parameter in  $\Theta$  based on the realization  $(x_1, \dots, x_n) \in X^n$  and weights  $(\lambda_1, \dots, \lambda_n) \in \Lambda_n$ .

*Problem 2.* Let  $(X, \mathcal{X})$  be a measurable space,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi : X \times \Theta \rightarrow \mathbb{R}$  be a measurable function in its first variable, i.e., for each  $t \in \Theta$ , the mapping  $X \ni x \mapsto \psi(x, t)$  is measurable with respect to the sigma-algebra  $\mathcal{X}$ . Further, let  $\xi : \Omega \rightarrow X$  be a random variable defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  such that  $\mathbb{E}(|\psi(\xi, t)|) < \infty$  for each  $t \in \Theta$ . We investigate the question of existence and uniqueness of a point of sign change (see Definition 2.1) for the function

$$(1.4) \quad \Theta \ni t \mapsto \mathbb{E}(\psi(\xi, t)).$$

In the literature, we could not find a name for the function (1.4), however, we may call it as a  $\psi$ -expectation function. Under appropriate conditions, the  $\psi$ -estimator (Z-estimator) based on i.i.d. observations  $\xi_1, \dots, \xi_n$  is supposed to 'well-estimate' the zero of the function (1.4), provided that it exists uniquely, for more details, see, e.g., Kosorok [12, Sections 2.2.5 and 13].

In what follows, we discuss the connections between the Problems 1 and 2 introduced above. Under appropriate conditions, the generalized  $\psi$ -estimator based on i.i.d. observations  $\xi_1, \dots, \xi_n$  is supposed to 'well-estimate' the point of sign change of the function (1.4), provided that it exists. In this paper, we do not investigate this question, it could be a topic of future research. Further, note that if  $\xi$  is a simple random variable such that  $\mathbb{P}(\xi = x_i) = p_i$ ,  $i = 1, \dots, n$ , where  $n \in \mathbb{N}$ ,  $(x_1, \dots, x_n) \in X^n$  and  $p_1, \dots, p_n \geq 0$ ,  $p_1 + \dots + p_n = 1$ , then  $\mathbb{E}(\psi(\xi, t)) = \sum_{i=1}^n p_i \psi(x_i, t)$ ,  $t \in \Theta$ , and hence, in this special case, Problem 2 is a special case of Problem 1.

To mention some papers related to Problem 1, we can refer, for example, to Huber [8, Lemma 1], Tibshirani [19] and Ali and Tibshirani [1]. Tibshirani [19] considered the lasso problem, which is also known as the  $\ell_1$ -penalized linear regression. The lasso estimator is a popular tool in the theory of sparse linear regression, mathematically, it is a solution of a not necessarily strictly convex minimization problem, where a penalty term being the  $\ell_1$ -norm of the coefficient vector comes into play. Tibshirani [19] studied the question of uniqueness of the lasso estimator. Recently, Ali and Tibshirani [1] have studied the uniqueness of a generalized lasso estimator, where the penalty term in the corresponding minimization problem is the  $\ell_1$ -norm of a (penalty) matrix times the coefficient vector. To mention further papers related to Problem 2, we can refer to Huber [8, Lemma 2], Clarke [4] (for details, see Remark 3.6), Mathieu [14] (for details, see Examples 3.5 and 4.4), and to the very recent paper of Dimitriadis et al. [6, Propositions S1, S2 and S3], in which the authors, in particular, considered solvability of the equation  $\mathbb{E}(\psi(\zeta, \eta, t)) = 0$ ,  $t \in \Theta$ , where  $\psi : \mathbb{R} \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}$  is a measurable function,  $\Theta$  is a (non-empty) open parameter set of  $\mathbb{R}$ , and  $(\zeta, \eta)$  is a response-regression pair in some regression model.

Section 2 is about the existence and uniqueness of (weighted) generalized  $\psi$ -estimators (Problem 1). First, we introduce the required terminology: the notions of point of sign change and level of increase for a real-valued function defined on a non-degenerate open interval (see Definitions 2.1 and 2.6), and the notions of properties  $(T_n)$  and  $(T_n^\lambda)$  for a function in  $\Psi(X, \Theta)$  (see Definition 2.3, but we also present below). We say that a function  $\psi \in \Psi(X, \Theta)$  *possesses the property  $(T_n^\lambda)$  for some  $n \in \mathbb{N}$  and  $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$*  if there exists a mapping  $\vartheta_{n,\psi}^\lambda : X^n \rightarrow \Theta$  such that, for each  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$  and  $t \in \Theta$ ,

$$\psi_{\mathbf{x},\lambda}(t) = \sum_{i=1}^n \lambda_i \psi(x_i, t) \begin{cases} > 0 & \text{if } t < \vartheta_{n,\psi}^\lambda(\mathbf{x}), \\ < 0 & \text{if } t > \vartheta_{n,\psi}^\lambda(\mathbf{x}). \end{cases}$$

In case of  $\lambda_i = 1$ ,  $i = 1, \dots, n$  (or equivalently, in case of equal positive weights), the property  $(T_n^\lambda)$  is called property  $(T_n)$ . As the first main result of our paper, in Theorem 2.10, necessary as well as sufficient conditions are given for the properties  $(T_n)$  and  $(T_n^\lambda)$ . After Theorem 2.10, we present some properties of the property  $(T_n^\lambda)$ . For example, Proposition 2.13

is about a connection between the property  $(T_n)$  and the strictly  $\frac{1}{n}$ -increasingness of some appropriately defined functions, and in Proposition 2.14 we establish a 'grouping' property of the property  $(T_n^\lambda)$ . Examples 2.16, 2.17 and 2.20 highlight the role of the conditions in Theorem 2.10 and in Proposition 2.14. We also introduce a class of Bajraktarević-type (in particular, quasi-arithmetic-type)  $\psi$ -estimators (motivated by the representation of Bajraktarević means as special deviation means) for which our results are well-applicable, see Example 2.19.

Section 3 is devoted to study the existence and uniqueness of the point of sign change of the  $\psi$ -expectation function given in (1.4) (Problem 2). As the second main result of our paper, in Theorem 3.1, we give a set of sufficient conditions in order that the  $\psi$ -expectation function in question have a unique point of sign change. We apply our results for  $\psi$ -expectation functions, when  $\psi$  is a Bajraktarević-type function (see Example 3.4 and Proposition 3.4), and when  $\psi$  has the form used by Mathieu [14] (see Example 3.5 and Proposition 3.5). In Remark 3.6, restricted to a one-dimensional parameter set, we recall Theorem 3.2 in Clarke [4] on the local uniqueness of a root of the function (1.4). We will see that the assumptions of Theorem 3.2 in Clarke [4] are much more involved and quite different compared to those of our Theorem 3.1.

In Section 4 we present several examples from statistical estimation theory that demonstrate the applicability of our results in Sections 2 and 3. These examples may be divided into three main groups. The first group of examples includes several well-known descriptive statistics that can be considered as special  $\psi$ -estimators. Namely, the empirical median (Example 4.1), the empirical quantiles (Example 4.2) and the empirical expectiles (Example 4.3). In particular, in Proposition 4.2 we show that, given  $n \geq 2$ , the function  $\psi$  corresponding to the empirical  $\alpha$ -quantile has the property  $(T_n)$  if and only if  $\alpha \notin \{\frac{1}{n}, \dots, \frac{n-1}{n}\}$ . The second group of examples contains the class of  $\psi$ -estimators recently used by Mathieu [14] (Example 4.4), and some  $\psi$ -estimators that are important in robust statistics. In particular, in Proposition 4.4, we derive necessary and sufficient conditions under which the function  $\psi$  used in Mathieu [14] has the property  $(T_n^\lambda)$  for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ . We emphasize that in all the above examples, we investigate the existence and uniqueness of (weighted) generalized  $\psi$ -estimators, compared to the existing results that addressed  $\psi$ -estimators (Z-estimators). The third group of examples demonstrates the applicability of Theorem 2.10 together with Proposition 2.12 for proving existence and uniqueness of solutions of likelihood equations. In Example 4.5, we consider the maximum likelihood estimator (MLE) of one of the parameters of a normally distributed random variable supposing that its other parameter is known. In Example 4.7, we consider a mixture density function of the standard normally density function and the density function of a normally distributed random variable with mean  $m \in \mathbb{R}$  and variance  $\sigma^2 > 0$  with equal weights  $\frac{1}{2}$ , and we study the solutions of the likelihood equation for  $m$  provided that  $\sigma$  is known.

Sections 5, 6 and 7 are devoted to the proofs of the results in Sections 2, 3 and 4, respectively.

Finally, we summarize the novelties of the paper. We extensively discuss that, up to our knowledge, only few results are available for the existence and uniqueness of  $\psi$ -estimators and of the roots of  $\psi$ -expectation functions, and our paper can be considered as a theoretical contribution to this field. Another important feature of our paper is that we present a broad

variety of examples from statistical estimation theory, where our results can be well-applied.

## 2 Notions and results on the existence and uniqueness of weighted generalized $\psi$ -estimators

To investigate Problem 1 presented in the Introduction, we introduce the required terminology. First, we introduce the notion of a point of sign change for real-valued functions defined on an open interval.

**2.1 Definition.** Let  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ . For a function  $f : \Theta \rightarrow \mathbb{R}$ , consider the following three level sets

$$\Theta_{f>0} := \{t \in \Theta : f(t) > 0\}, \quad \Theta_{f=0} := \{t \in \Theta : f(t) = 0\}, \quad \Theta_{f<0} := \{t \in \Theta : f(t) < 0\}.$$

We say that  $\vartheta \in \Theta$  is a point of sign change (of decreasing type) for  $f$  if

$$f(t) > 0 \quad \text{for } t < \vartheta, \quad \text{and} \quad f(t) < 0 \quad \text{for } t > \vartheta.$$

**2.2 Remark.** Note that, if  $\vartheta \in \Theta$  is a point of sign change for  $f$ , then  $\Theta_{f>0}$  and  $\Theta_{f<0}$  are nonempty sets and  $\sup \Theta_{f>0} = \inf \Theta_{f<0} = \vartheta$ . Furthermore, there can exist at most one element  $\vartheta \in \Theta$  which is a point of sign change for  $f$ . If  $f$  is continuous at a point  $\vartheta$  of sign change, then  $f(\vartheta) = 0$ , moreover  $\Theta_{f=0} = \{\vartheta\}$ . Conversely, as an easy consequence of the Bolzano theorem, if  $f : \Theta \rightarrow \mathbb{R}$  is continuous, the sets  $\Theta_{f>0}$  and  $\Theta_{f<0}$  are nonempty and  $f$  has a unique zero  $\vartheta \in \Theta$ , i.e.,  $\Theta_{f=0} = \{\vartheta\}$  holds, then  $\vartheta$  is a point of sign change either for  $f$  or for  $(-f)$ . The continuity of  $f$ , however, is not necessary for the existence of a point of sign change for  $f$ . For example, if  $f$  is strictly decreasing and the sets  $\Theta_{f>0}$  and  $\Theta_{f<0}$  are nonempty, then it is easy to see that there exists a point of sign change for  $f$ .  $\square$

**2.3 Definition.** We say that a function  $\psi \in \Psi(X, \Theta)$

(i) is continuous in its second variable if, for each  $x \in X$ , the mapping  $\Theta \ni t \mapsto \psi(x, t)$  is continuous.

(ii) possesses the property  $(T_n)$  (briefly,  $\psi$  is a  $T_n$ -function) for some  $n \in \mathbb{N}$  if there exists a mapping  $\vartheta_{n,\psi} : X^n \rightarrow \Theta$  such that, for each  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$  and  $t \in \Theta$ ,

$$(2.1) \quad \psi_{\mathbf{x}}(t) = \sum_{i=1}^n \psi(x_i, t) \begin{cases} > 0 & \text{if } t < \vartheta_{n,\psi}(\mathbf{x}), \\ < 0 & \text{if } t > \vartheta_{n,\psi}(\mathbf{x}), \end{cases}$$

that is, for each  $\mathbf{x} \in X^n$ , the value  $\vartheta_{n,\psi}(\mathbf{x})$  is a point of sign change for the function  $\psi_{\mathbf{x}}$ . If there is no confusion, instead of  $\vartheta_{n,\psi}$  we simply write  $\vartheta_n$ . We may call  $\vartheta_{n,\psi}(\mathbf{x})$  as a generalized  $\psi$ -estimator for some unknown parameter in  $\Theta$  based on the realization  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ .

(iii) possesses the property  $(T_n^\lambda)$  for some  $n \in \mathbb{N}$  and  $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$  (briefly,  $\psi$  is a  $T_n^\lambda$ -function) if there exists a mapping  $\vartheta_{n,\psi}^\lambda : X^n \rightarrow \Theta$  such that, for each  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$  and  $t \in \Theta$ ,

$$(2.2) \quad \psi_{\mathbf{x},\lambda}(t) = \sum_{i=1}^n \lambda_i \psi(x_i, t) \begin{cases} > 0 & \text{if } t < \vartheta_{n,\psi}^\lambda(\mathbf{x}), \\ < 0 & \text{if } t > \vartheta_{n,\psi}^\lambda(\mathbf{x}), \end{cases}$$

that is, for each  $\mathbf{x} \in X^n$ , the value  $\vartheta_{n,\psi}^\lambda(\mathbf{x})$  is a point of sign change for the function  $\psi_{\mathbf{x},\lambda}$ . If there is no confusion, instead of  $\vartheta_{n,\psi}^\lambda$  we simply write  $\vartheta_n^\lambda$ . We may call  $\vartheta_{n,\psi}^\lambda(\mathbf{x})$  as a weighted generalized  $\psi$ -estimator for some unknown parameter in  $\Theta$  based on the realization  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$  and weights  $(\lambda_1, \dots, \lambda_n) \in \Lambda_n$ .

In the next remark, we present some basic facts about the properties  $(T_n)$  and  $(T_n^\lambda)$  given in Definition 2.3, which can be easily checked.

**2.4 Remark.** (i) If  $\psi$  is a  $T_n$ -function for some  $n \in \mathbb{N}$ , then for each  $x_1, \dots, x_n \in X$ , the equation (1.2) can have at most one solution.

(ii) If  $\psi$  is continuous in its second variable and is a  $T_n$ -function for some  $n \in \mathbb{N}$ , then  $t = \vartheta_n(x_1, \dots, x_n)$  is the unique solution to (1.2), and is called the  $\psi$ -estimator (Z-estimator) based on the observations  $x_1, \dots, x_n \in X$ . In particular, if  $\psi$  is continuous in its second variable and is a  $T_1$ -function, then, for each  $x \in X$ , the equation  $\psi(x, t) = 0$ ,  $t \in \Theta$ , has a unique solution  $\vartheta_1(x)$ .

(iii) If  $\lambda_1 = \dots = \lambda_n > 0$  with some  $n \in \mathbb{N}$ , then  $\vartheta_{n,\psi}^{(\lambda_1, \dots, \lambda_n)} = \vartheta_n$ . □

In the next remark, we point out an invariance property of the properties  $(T_n)$  and  $(T_n^\lambda)$  given in Definition 2.3.

**2.5 Remark.** We introduce a notion of equivalence in  $\Psi(X, \Theta)$  as follows. We say that the maps  $\psi, \varphi \in \Psi(X, \Theta)$  are *equivalent* (denoted as  $\psi \sim \varphi$ ) if there exists a positive function  $h : \Theta \rightarrow \mathbb{R}$  such that  $\psi(x, t) = h(t)\varphi(x, t)$  is valid for all  $(x, t) \in X \times \Theta$ . It is easy to see that  $\sim$  is an equivalence relation on  $\Psi(X, \Theta)$ , furthermore, the properties  $(T_n)$  and  $(T_n^\lambda)$  are invariant with respect to this equivalence, that is, if  $\psi \sim \varphi$  and  $\varphi$  possesses the property  $(T_n)$  (or the property  $(T_n^\lambda)$ ), then  $\psi$  also enjoys this property and  $\vartheta_{n,\psi} = \vartheta_{n,\varphi}$  (resp.  $\vartheta_{n,\psi}^\lambda = \vartheta_{n,\varphi}^\lambda$ ). □

**2.6 Definition.** Let  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$  and  $f : \Theta \rightarrow \mathbb{R}$  be a function. We say that  $y \in \mathbb{R}$  is a level of increase for  $f$  if the relations  $u \in \Theta$  and  $y \leq f(u)$  imply that  $y < f(v)$  for all  $v \in \Theta$  with  $u < v$ .

Observe that if  $y \in \mathbb{R}$  is a level of increase for  $f$ , then the inverse image  $f^{-1}(\{y\})$  is either empty or a singleton. In general, the converse of the previous statement is not true. To give a counterexample, let us consider the function  $f : \Theta \rightarrow \mathbb{R}$  given by  $f(\vartheta) := 1$  if  $\vartheta \in \Theta$  is rational, and  $f(\vartheta) := 0$  if  $\vartheta \in \Theta$  is irrational. Then  $f^{-1}(\{\frac{1}{2}\}) = \emptyset$ , but  $\frac{1}{2}$  is not a level of increase for  $f$ .

In the following lemma, we establish a connection between the notions of point of sign change and level of increase.

**2.7 Lemma.** *Let  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ ,  $f : \Theta \rightarrow \mathbb{R}$  be a function, and  $y \in \mathbb{R}$ . Then  $y$  is a level of increase for  $f$  if and only if one of the following assertions holds:*

- (i)  $y < f$  on  $\Theta$ .
- (ii)  $y > f$  on  $\Theta$ .
- (iii) There exists a point of sign change for the function  $y - f$ .

**2.8 Lemma.** *Let  $\Theta$  be a nondegenerate open interval and  $f : \Theta \rightarrow \mathbb{R}$  be a function. If the levels of increase for  $f$  form a dense subset in the convex hull of  $f(\Theta)$ , then  $f$  is increasing. The function  $f$  is strictly increasing if and only if every element of  $f(\Theta)$  is a level of increase for  $f$ . Furthermore, if  $g : H \rightarrow \mathbb{R}$  is a strictly increasing function, where  $H$  is a set containing  $f(\Theta)$ , and  $y \in H$  is level of increase for  $f$ , then  $g(y)$  is a level of increase for  $g \circ f$ .*

We recall also a definition due to Páles [16]: given a nondegenerate open interval  $\Theta$  of  $\mathbb{R}$  and  $\varepsilon > 0$ , a function  $f : \Theta \rightarrow \mathbb{R}$  satisfying the inequality  $f(u) \leq f(v) + \varepsilon$  for all  $u < v$ ,  $u, v \in \Theta$  is called  $\varepsilon$ -increasing. If the inequality is strict for all  $u < v$ ,  $u, v \in \Theta$ , then  $f$  is said to be *strictly  $\varepsilon$ -increasing*. We note that Páles [16, Theorem 3] offers the following simple characterization: a function  $f : \Theta \rightarrow \mathbb{R}$  is  $\varepsilon$ -increasing if and only if there exists an increasing function  $g : \Theta \rightarrow \mathbb{R}$  such that  $\|f - g\|_\infty := \sup_{u \in \Theta} |f(u) - g(u)| \leq \frac{\varepsilon}{2}$ .

The next lemma describes a connection between levels of increase for a function  $f : \Theta \rightarrow \mathbb{R}$  and its  $\varepsilon$ -increasingness property.

**2.9 Lemma.** *Let  $\Theta$  be a nondegenerate open interval, let  $n \in \mathbb{N}$  and let  $y_0 < \dots < y_n$  be real numbers. Assume that  $y_0, \dots, y_{n-1}$  are levels of increase for a function  $f : \Theta \rightarrow \mathbb{R}$  and  $f(\Theta) \subseteq [y_0, y_n]$ . Then  $f$  is strictly  $\varepsilon$ -increasing with  $\varepsilon := \max\{y_1 - y_0, \dots, y_n - y_{n-1}\}$ .*

Now, we state our first main result by presenting necessary as well as sufficient conditions for the properties  $(T_n)$  and  $(T_n^\lambda)$ .

**2.10 Theorem.** *Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi \in \Psi(X, \Theta)$  be a  $T_1$ -function.*

- (i) *If  $\psi$  is a  $T_2^{(\lambda_1, \lambda_2)}$ -function for some  $(\lambda_1, \lambda_2) \in (0, \infty)^2$ , then, for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the numbers  $\frac{\lambda_1}{\lambda_2}$  and  $\frac{\lambda_2}{\lambda_1}$  are levels of increase for the function*

$$(2.3) \quad (\vartheta_1(x), \vartheta_1(y)) \ni t \mapsto -\frac{\psi(x, t)}{\psi(y, t)}.$$

- (ii) *If  $\psi$  is a  $T_n^{(\lambda_1, \dots, \lambda_n)}$ -function for some  $n \in \mathbb{N} \setminus \{1\}$  and  $(\lambda_1, \dots, \lambda_n) \in (0, \infty)^n$ , then, for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the numbers  $\frac{\lambda_1 + \dots + \lambda_k}{\lambda_{k+1} + \dots + \lambda_n}$  and  $\frac{\lambda_{k+1} + \dots + \lambda_n}{\lambda_1 + \dots + \lambda_k}$ ,  $k \in \{1, \dots, n-1\}$ , are levels of increase for the function (2.3).*



- (iii) If  $\psi$  is a  $T_n$ -function for some  $n \in \mathbb{N} \setminus \{1\}$ , then, for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the elements of the set  $\{\frac{k}{n-k} \mid k \in \{1, \dots, n-1\}\}$  are levels of increase for the function (2.3).
- (iv) If  $\psi$  is a  $T_n$ -function for infinitely many  $n \in \mathbb{N}$ , then for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is increasing. In addition, if for each  $m \in \mathbb{N}$  there exists  $n \in \mathbb{N}$  such that  $m$  divides  $n$  and  $\psi$  is a  $T_n$ -function, then, for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , every positive rational number is a level of increase for the function (2.3).
- (v) If  $\psi$  is a  $T_2^\lambda$ -function for each  $\lambda \in \Lambda_2$ , then for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is strictly increasing.
- (vi) If, for each  $x \in X$ , the equality  $\psi(x, \vartheta_1(x)) = 0$  holds and, for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is strictly increasing, then  $\psi$  is a  $T_n^\lambda$ -function for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ .

Note that if  $\psi \in \Psi(X, \Theta)$  is continuous in its second variable, then the assertion (vi) of Theorem 2.10 provides a sufficient condition for the existence and uniqueness of a (usual)  $\psi$ -estimator.

The following statement establishes two equivalent conditions for  $\psi$  having the property  $(T_n^\lambda)$  for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ .

**2.11 Corollary.** *Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi \in \Psi(X, \Theta)$  be a  $T_1$ -function such that, for each  $x \in X$ , the equality  $\psi(x, \vartheta_1(x)) = 0$  holds. Then the following assertions are equivalent:*

- (i) For each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is strictly increasing.
- (ii) For each  $\lambda \in \Lambda_2$ , the function  $\psi$  has property  $(T_2^\lambda)$ .
- (iii) For each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ , the function  $\psi$  has property  $(T_n^\lambda)$ .

In the next proposition, we provide a sufficient condition (which does not involve the condition  $\psi(x, \vartheta_1(x)) = 0$  for all  $x \in X$ ) under which  $\psi$  has the property  $(T_n^\lambda)$  for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ .

**2.12 Proposition.** *Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi \in \Psi(X, \Theta)$  be a  $T_1$ -function.*

- (i) If for each  $x \in X$ , the function  $\Theta \ni t \mapsto \psi(x, t)$  is (strictly) decreasing, then for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is (strictly) increasing.
- (ii) If for each  $x \in X$ , the function  $\Theta \ni t \mapsto \psi(x, t)$  is strictly decreasing, then for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ , the function  $\psi$  has property  $(T_n^\lambda)$ .

The next proposition establishes a connection between the property  $(T_n)$  and the strictly  $\frac{1}{n}$ -increasingness of some appropriately defined functions.

**2.13 Proposition.** Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi \in \Psi(X, \Theta)$  be a  $T_1$ -function. If  $\psi$  is a  $T_n$ -function for some  $n \in \mathbb{N} \setminus \{1\}$ , then, for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function

$$(2.4) \quad (\vartheta_1(x), \vartheta_1(y)) \ni t \mapsto \frac{\psi(x, t)}{\psi(x, t) - \psi(y, t)}$$

is strictly  $\frac{1}{n}$ -increasing.

The following two results describe the hierarchy among the properties  $(T_n)_{n \in \mathbb{N}}$  and establish a kind of 'grouping' property of the property  $(T_n^\lambda)$ .

**2.14 Proposition.** Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi \in \Psi(X, \Theta)$ . If  $\psi$  is a  $T_n$ -function for some  $n \in \mathbb{N}$ , then  $\psi$  is a  $T_m$ -function for any  $m \in \{1, \dots, n\}$  that divides  $n$ .

**2.15 Proposition.** Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi \in \Psi(X, \Theta)$  be a  $T_n^\lambda$ -function, where  $n \in \mathbb{N}$  and  $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$ . Let  $m \in \{1, \dots, n\}$  and  $H_1, \dots, H_m$  be nonempty pairwise disjoint subsets of  $\{1, \dots, n\}$  such that  $H_1 \cup \dots \cup H_m = \{1, \dots, n\}$ . For each  $\alpha \in \{1, \dots, m\}$ , define  $\mu_\alpha := \sum_{i \in H_\alpha} \lambda_i$ . Then  $\mu := (\mu_1, \dots, \mu_m) \in \Lambda_m$  and  $\psi$  is a  $T_m^\mu$ -function.

Note that Proposition 2.15 implies Proposition 2.14. Indeed, let  $\psi \in \Psi(X, \Theta)$  be a  $T_n$ -function with some  $n \in \mathbb{N}$ . If  $m$  is a divisor of  $n$ , then  $n = km$  with some  $k \in \mathbb{N}$ , and hence  $H_j := \{(j-1)k + 1, \dots, jk\}$ ,  $j = 1, \dots, m$ , are nonempty pairwise disjoint subsets such that  $H_1 \cup \dots \cup H_m = \{1, \dots, n\}$ . With the choice  $\lambda := (\lambda_1, \dots, \lambda_n) := (1, \dots, 1) \in \Lambda_n$ , we have  $\psi$  is a  $T_n^\lambda$ -function and  $\mu_\alpha = k$ ,  $\alpha \in \{1, \dots, m\}$ . Hence Proposition 2.15 yields that  $\psi$  is a  $T_m^\mu$ -function. Since  $\mu_\alpha$ ,  $\alpha \in \{1, \dots, m\}$ , are all the same positive constant  $k$ , we have  $\psi$  is a  $T_m$ -function as well. Further, we note that Proposition 2.15 may be useful to see, for example, that the property  $(T_3^{(\lambda_1, \lambda_2, \lambda_3)})$  of  $\psi$  implies the property  $(T_2^{(\lambda_1 + \lambda_2, \lambda_3)})$  of  $\psi$ , where  $(\lambda_1, \lambda_2, \lambda_3) \in \Lambda_3$ .

The next example demonstrates that the property  $(T_2)$  can already fail to hold.

**2.16 Example.** Let  $m \in \mathbb{N}$ ,  $X := \{x_1, \dots, x_m\}$ ,  $\Theta := \mathbb{R}$  and let  $w_1, \dots, w_m > 0$ . Define  $\psi : X \times \Theta \rightarrow \mathbb{R}$  by

$$\psi(x_i, t) := \begin{cases} w_i & \text{if } t < i, \\ -w_i & \text{if } t \geq i, \end{cases} \quad i \in \{1, \dots, m\}.$$

Then  $\psi$  is a  $T_1$ -function and  $\vartheta_1(x_i) = i$  holds for all  $i \in \{1, \dots, m\}$ . One can easily see that the property  $(T_2)$  holds if and only if we have  $w_i \neq w_j$  for all distinct  $i, j \in \{1, \dots, m\}$ . Indeed, if  $1 \leq i < j \leq m$ , then

$$\psi(x_i, t) + \psi(x_j, t) = \begin{cases} w_i + w_j > 0 & \text{if } t < i, \\ -w_i + w_j & \text{if } i \leq t < j, \\ -w_i - w_j < 0 & \text{if } j \leq t. \end{cases}$$

This function has a point of sign change if and only if  $w_i \neq w_j$ , as desired. We also get that if  $\psi$  is a  $T_2$ -function, then

$$\vartheta_2(x_i, x_j) = \vartheta_2(x_j, x_i) = \begin{cases} i & \text{if } w_i > w_j, \\ j & \text{if } w_i < w_j. \end{cases}$$

Furthermore,  $\vartheta_2(x_i, x_i) = \vartheta_1(x_i) = i$  for all  $i \in \{1, \dots, m\}$ .  $\square$

In what follows, we give an example to point out that in part (vi) of Theorem 2.10 the assumption that  $\psi(x, \vartheta_1(x)) = 0$ ,  $x \in X$ , cannot be omitted.

**2.17 Example.** Let  $X := \{x_1, x_2\}$  (with  $x_1 \neq x_2$ ) and  $\Theta := \mathbb{R}$ . Let

$$\psi(x_1, t) := \begin{cases} 2 & \text{if } t < 1, \\ -t & \text{if } 1 \leq t \leq 2, \\ -2 & \text{if } t > 2, \end{cases}$$

and

$$\psi(x_2, t) := \begin{cases} 1 & \text{if } t < 2, \\ 2 & \text{if } t = 2, \\ -1 & \text{if } t > 2. \end{cases}$$

Then  $\psi$  is a  $T_1$ -function with  $\vartheta_1(x_1) = 1$  and  $\vartheta_1(x_2) = 2$ , and  $\psi(x_i, \vartheta_1(x_i)) \neq 0$  for  $i \in \{1, 2\}$ . We also note that the function

$$(\vartheta_1(x_1), \vartheta_1(x_2)) = (1, 2) \ni t \mapsto \frac{\psi(x_1, t)}{\psi(x_2, t)} = -t$$

is strictly decreasing. However,  $\psi$  is not a  $T_2$ -function, since

$$\psi(x_1, t) + \psi(x_2, t) = \begin{cases} 2 + 1 = 3 > 0 & \text{if } t < 1, \\ -1 + 1 = 0 & \text{if } t = 1, \\ -t + 1 < 0 & \text{if } 1 < t < 2, \\ -2 + 2 = 0 & \text{if } t = 2, \\ -2 - 1 = -3 < 0 & \text{if } t > 2, \end{cases}$$

which shows that  $\mathbb{R} \ni t \mapsto \psi(x_1, t) + \psi(x_2, t)$  does not have a point of sign change.  $\square$

In what follows, as an application of Proposition 2.12, we present an example of a large class of functions  $\psi : X \times \Theta \rightarrow \mathbb{R}$ , which possesses the property  $(T_n^\lambda)$  for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$  and for which the point of sign change  $\vartheta_{n,\psi}^\lambda(\mathbf{x})$  (where  $\mathbf{x} \in X^n$ ) has an explicit form. This class of functions may be called the class of Bajraktarević-type functions, motivated by the representation of Bajraktarević means as special deviation means. For the description of this class of functions, we need to recall the notion of generalized left inverse of a strictly monotone (but not necessarily continuous) function defined on a nondegenerate open interval of  $\mathbb{R}$  due to Grünwald and Páles [7, Lemma 1].

**2.18 Lemma.** *Let  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , let  $f : \Theta \rightarrow \mathbb{R}$  be a strictly increasing function. Then there exists a uniquely determined monotone function  $g : \text{conv}(f(\Theta)) \rightarrow \Theta$  such that  $g$  is the left inverse of  $f$ , i.e.,*

$$(g \circ f)(x) = x, \quad x \in \Theta.$$

Furthermore,  $g$  is monotone in the same sense as  $f$ , is continuous, and the following relation holds:

$$(f \circ g)(y) = y, \quad y \in f(\Theta).$$

The function  $g : \text{conv}(f(\Theta)) \rightarrow \Theta$  described in Lemma 2.18 is called the *generalized left inverse of the strictly increasing function  $f : \Theta \rightarrow \mathbb{R}$*  and is denoted by  $f^{(-1)}$ . In fact, by the proof of Lemma 1 in Grünwald and Páles [7], it also turns out that

$$g(y) = \sup\{u \in \Theta : f(u) \leq y\}, \quad y \in \text{conv}(f(\Theta)).$$

It is clear that the restriction of  $f^{(-1)}$  to  $f(\Theta)$  is the inverse of  $f$  in the standard sense. Therefore,  $f^{(-1)}$  is the continuous and monotone extension of the inverse of  $f$  to the smallest interval containing the range of  $f$ , that is, to the convex hull of  $f(\Theta)$ .

**2.19 Example. (Bajraktarević-type functions)** Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ ,  $f : \Theta \rightarrow \mathbb{R}$  be a strictly increasing function, and  $p : X \rightarrow \mathbb{R}_{++}$  and  $\varphi : X \rightarrow \text{conv}(f(\Theta))$  be functions. In terms of these functions, define  $\psi : X \times \Theta \rightarrow \mathbb{R}$  by

$$(2.5) \quad \psi(x, t) := p(x)(\varphi(x) - f(t)), \quad x \in X, t \in \Theta.$$

**2.19 Proposition.** *Under the above assumptions,  $\psi$  is a  $T_n^\lambda$ -function for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ , and*

$$(2.6) \quad \vartheta_{n,\psi}^\lambda(\mathbf{x}) = f^{(-1)}\left(\frac{\lambda_1 p(x_1) \varphi(x_1) + \cdots + \lambda_n p(x_n) \varphi(x_n)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)}\right)$$

for each  $n \in \mathbb{N}$ ,  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$  and  $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$ . In particular,  $\vartheta_1 = f^{(-1)} \circ \varphi$  holds.

One may call the value  $\vartheta_{n,\psi}^\lambda(\mathbf{x})$  given by (2.6) as a Bajraktarević-type  $\psi$ -estimator of some unknown parameter in  $\Theta$  based on the realization  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$  and weights  $(\lambda_1, \dots, \lambda_n) \in \Lambda_n$  corresponding to the Bajraktarević-type function given by (2.5). In particular, if  $p = 1$  is a constant function in (2.5), then we speak about a quasi-arithmetic-type  $\psi$ -estimator.

Note that in case of  $X := \Theta$  and  $\varphi := f$ , Proposition 2.19 reduces to Theorem 3 in Grünwald and Páles [7] for Bajraktarević means. In addition, if  $p = 1$  is a constant function, then Proposition 2.19 is about generalized quasi-arithmetic means (here, we use the term 'generalized', since for usual quasi-arithmetic means, the function  $f$  is not only strictly increasing, but continuous as well).  $\square$

In the next example, we point out that, in general, one cannot omit the restriction that  $m$  divides  $n$  in Proposition 2.14. For another example, see the case of empirical median discussed in Example 4.1.

**2.20 Example.** Let  $X$  be an arbitrary set with at least two distinct elements and let  $X_1, X_2 \subseteq X$  be nonempty disjoint subsets such that  $X_1 \cup X_2 = X$  and let  $w_1, w_2 > 0$ . Define  $\psi : X \times \Theta \rightarrow \mathbb{R}$  by

$$\psi(x, t) := \begin{cases} w_i & t < i, \\ -w_i & t \geq i, \end{cases} \quad \text{if } x \in X_i.$$

Then  $\psi$  is a  $T_1$ -function and, for  $i \in \{1, 2\}$  and  $x \in X_i$ , we have  $\vartheta_1(x) = i$ .

Let  $n, k \in \mathbb{N}$  such that  $k$  is not a divisor of  $n$  (which implies that  $k \geq 2$ ). Then

$$(2.7) \quad \frac{1}{k} \notin \left\{ \frac{1}{n}, \dots, \frac{n-1}{n} \right\}.$$

Indeed, on the contrary, if the inclusion were valid, then  $1/k$  would be of the form  $m/n$  for some  $m \in \{1, \dots, n-1\}$ , yielding that  $n = km$ , which contradicts the assumption  $k \nmid n$ .

Assuming that  $w_1 = k - 1$  and  $w_2 = 1$ , we prove that  $\psi$  is a  $T_n$ -function, but it is not a  $T_k$ -function.

To show that  $\psi$  is a  $T_n$ -function, let  $\mathbf{y} := (y_1, \dots, y_n) \in X^n$ . For  $j \in \{1, 2\}$ , define the set  $S_j := \{i \in \{1, \dots, n\} : y_i \in X_j\}$ . Then  $\{S_1, S_2\}$  forms a partition of  $\{1, \dots, n\}$ . Let  $n_j$  denote the cardinality of  $S_j$ ,  $j \in \{1, 2\}$ . Then  $n = n_1 + n_2$  and

$$\psi_{\mathbf{y}}(t) := \sum_{i=1}^n \psi(y_i, t) = \begin{cases} n_1 w_1 + n_2 w_2 > 0 & \text{if } t < 1, \\ -n_1 w_1 + n_2 w_2 & \text{if } 1 \leq t < 2, \\ -n_1 w_1 - n_2 w_2 < 0 & \text{if } 2 \leq t. \end{cases}$$

Using condition (2.7) and  $k \geq 2$ , we have

$$-n_1 w_1 + n_2 w_2 = -n_1(k-1) + (n - n_1) = n - n_1 k \neq 0.$$

Therefore, the point of sign change for the function  $\psi_{\mathbf{y}}$  equals 1 if  $-n_1 w_1 + n_2 w_2 < 0$  and equals 2 if  $-n_1 w_1 + n_2 w_2 > 0$ . This proves that  $\psi$  possesses the property  $(T_n)$ .

To verify that  $\psi$  is not a  $T_k$ -function, let  $x_1 \in X_1$  and  $x_2 \in X_2$  be fixed and let  $\mathbf{z} := (x_1, x_2, \dots, x_2) \in X^k$ . Then

$$\psi_{\mathbf{z}}(t) := \psi(x_1, t) + (k-1)\psi(x_2, t) = \begin{cases} w_1 + (k-1)w_2 > 0 & \text{if } t < 1, \\ -w_1 + (k-1)w_2 = 0 & \text{if } 1 \leq t < 2, \\ -w_1 - (k-1)w_2 < 0 & \text{if } 2 \leq t. \end{cases}$$

Therefore, the function  $\psi_{\mathbf{z}}$  does not have a point of sign change, and, consequently,  $\psi$  is not a  $T_k$ -function, as desired.  $\square$

### 3 Existence and uniqueness of the point of sign change of $\psi$ -expectation functions

In this section, we investigate Problem 2 presented in the Introduction.

As it was mentioned in the Introduction, in case of simple random variables Problem 2 is a special case of Problem 1. More precisely, if  $\psi \in \Psi(X, \Theta)$  and  $\xi$  is a simple random variable such that  $\mathbb{P}(\xi = x_i) = p_i$ ,  $i = 1, \dots, n$ , where  $n \in \mathbb{N}$ ,  $(x_1, \dots, x_n) \in X^n$  and  $p_1, \dots, p_n \geq 0$ ,  $p_1 + \dots + p_n = 1$ , then  $\mathbb{E}(\psi(\xi, t)) = \sum_{i=1}^n p_i \psi(x_i, t)$ ,  $t \in \Theta$ . In addition, if  $\psi$  is a  $T_n^{(p_1, \dots, p_n)}$ -function, then, by definition, the function (1.4) has a unique point of sign change. Further, Theorem 2.10 provides some necessary as well as some sufficient conditions under which  $\psi$  possesses the property  $(T_n^{(p_1, \dots, p_n)})$ .

Next, we present our second main result in which we give a set of sufficient conditions in order that the function given by (1.4) have a unique point of sign change.

**3.1 Theorem.** *Let  $(X, \mathcal{X})$  be a measurable space,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ ,  $\psi : X \times \Theta \rightarrow \mathbb{R}$  be a function, and  $\xi : \Omega \rightarrow X$  be a random variable defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Let us suppose that*

- (i)  $\psi$  is a  $T_1$ -function and, for each  $x \in X$ , the equality  $\psi(x, \vartheta_1(x)) = 0$  holds,
- (ii) for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is strictly increasing,
- (iii)  $\psi$  is measurable in its first variable,
- (iv)  $\mathbb{E}(|\psi(\xi, t)|) < \infty$  for each  $t \in \Theta$ ,
- (v) there exist  $s_0, t_0 \in \Theta$  such that  $\mathbb{E}(\psi(\xi, s_0)) \geq 0$  and  $\mathbb{E}(\psi(\xi, t_0)) \leq 0$ .

Then the map  $\Theta \ni t \rightarrow \mathbb{E}(\psi(\xi, t))$  admits a unique point of sign change in  $\Theta$ .

Next, we provide a set of sufficient conditions (which does not involve the condition  $\psi(x, \vartheta_1(x)) = 0$  for each  $x \in X$ ) under which the map  $\Theta \ni t \rightarrow \mathbb{E}(\psi(\xi, t))$  also has a unique point of sign change.

**3.2 Proposition.** *Let  $(X, \mathcal{X})$  be a measurable space,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ ,  $\psi : X \times \Theta \rightarrow \mathbb{R}$  be a function, and  $\xi : \Omega \rightarrow X$  be a random variable defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Let us suppose that*

- (i) for each  $x \in X$ , the function  $\Theta \ni t \mapsto \psi(x, t)$  is strictly decreasing,
- (ii)  $\psi$  is measurable in its first variable,
- (iii)  $\mathbb{E}(|\psi(\xi, t)|) < \infty$  for each  $t \in \Theta$ ,

(iv) there exist  $s_0, t_0 \in \Theta$  such that  $\mathbb{E}(\psi(\xi, s_0)) \geq 0$  and  $\mathbb{E}(\psi(\xi, t_0)) \leq 0$ .

Then the function  $\Theta \ni t \rightarrow \mathbb{E}(\psi(\xi, t))$  admits a unique point of sign change in  $\Theta$ .

Next, we formulate a corollary of Theorem 3.1, which is in fact part (vi) of Theorem 2.10.

**3.3 Corollary.** *Let  $X$  be a nonempty set,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $\psi \in \Psi(X, \Theta)$  be a  $T_1$ -function. Let us suppose that, for each  $x \in X$ , the equality  $\psi(x, \vartheta_1(x)) = 0$  holds, and, for each  $x, y \in X$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is strictly increasing. Then  $\psi$  is a  $T_n^\lambda$ -function for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ .*

In what follows, we present an example to Proposition 3.2, which can be considered as a theoretical counterpart of Example 2.19 for Bajraktarević-type functions  $\psi$ .

**3.4 Example. (Bajraktarević-type functions)** Let  $(X, \mathcal{X})$  be a measurable space,  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ ,  $f : \Theta \rightarrow \mathbb{R}$  be a strictly increasing function, and  $p : X \rightarrow \mathbb{R}_{++}$  and  $\varphi : X \rightarrow \text{conv}(f(\Theta))$  be measurable functions. Let  $\psi : X \times \Theta \rightarrow \mathbb{R}$ ,

$$\psi(x, t) := p(x)(\varphi(x) - f(t)), \quad x \in X, t \in \Theta.$$

Further, let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a probability space,  $\xi : \Omega \rightarrow X$  be a random variable such that  $\mathbb{E}(p(\xi)|\varphi(\xi)) < \infty$  and  $\mathbb{E}(p(\xi)) < \infty$ .

**3.4 Proposition.** *Under the above assumptions, the function  $\Theta \ni t \rightarrow \mathbb{E}(\psi(\xi, t))$  admits a unique point of sign change in  $\Theta$  which is given by*

$$f^{(-1)}\left(\frac{\mathbb{E}(p(\xi)\varphi(\xi))}{\mathbb{E}(p(\xi))}\right).$$

The following auxiliary result is instrumental for the proof of Proposition 3.4.

**3.4 Lemma.** *Let  $(X, \mathcal{X})$  be a measurable space and  $p : X \rightarrow \mathbb{R}_{++}$  and  $\varphi : X \rightarrow \mathbb{R}$  be measurable functions. Further, let  $\xi$  be a random variable on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  such that  $\mathbb{P}(\xi \in X) = 1$ ,  $\mathbb{E}(p(\xi)) < \infty$  and  $\mathbb{E}(p(\xi)|\varphi(\xi)) < \infty$ . Then*

$$\frac{\mathbb{E}(p(\xi)\varphi(\xi))}{\mathbb{E}(p(\xi))} \in \text{conv}(\varphi(X)).$$

□

In the next example, we consider a particular form of  $\psi$  which has been recently investigated by Mathieu [14]: namely, let  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$(3.1) \quad \psi(x, t) := \text{sign}(x - t)f(|x - t|), \quad x, t \in \mathbb{R},$$

where  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ . Mathieu [14, Lemma 2] has derived some sufficient conditions on  $f$  and  $\xi$  under which the equation  $\mathbb{E}(\psi(\xi, t)) = 0$ ,  $t \in \mathbb{R}$ , has a unique solution, for more details and our new results in this special case, see the next Example 3.5.

**3.5 Example.** Let  $X := \mathbb{R}$ ,  $\Theta := \mathbb{R}$  and  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be given by (3.1). Given a random variable  $\xi$ , Mathieu [14] has recently considered the problem of finding a unique element  $t_0 \in \Theta$  such that  $\mathbb{E}(\psi(\xi, t_0)) = 0$  holds, where  $\psi$  has the form given in (3.1) such that  $f$  admits the following properties (called Assumption 2 in Mathieu [14]):

- (a)  $f$  is continuous and differentiable Lebesgue almost everywhere,
- (b)  $f(0) = 0$ ,
- (c)  $f$  is concave,
- (d) there exist  $\beta, \gamma > 0$  such that  $\gamma \mathbf{1}_{\{x \leq \beta\}} \leq f'(x) \leq 1$  Lebesgue a.e.  $x \geq 0$ .

For historical fidelity, we note that Mathieu [14] investigated a more general setup, he considered a random variable  $\xi$  having values in a Hilbert space  $\mathcal{H}$ , and a function  $\psi : \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}$ ,  $\psi(x, t) := \frac{x-t}{\|x-t\|} f(\|x-t\|)$  for  $x \neq t$ ,  $x, t \in \mathcal{H}$ , where  $\|\cdot\|$  is the norm of the Hilbert space  $\mathcal{H}$  (the value of  $\psi$  at  $(x, x)$ ,  $x \in \mathcal{H}$ , was not specified in Mathieu [14]).

Mathieu [14, Lemma 2] has shown that (formulating his result only in the case of  $\mathcal{H} = \mathbb{R}$ ) if  $f$  admits the properties (a)-(d),  $\mathbb{E}(|\xi|) < \infty$  and the inequality  $\mathbb{E}(\varrho(|\xi - \mathbb{E}(\xi)|)) < \varrho(\beta)$  holds, where  $\varrho(x) := \int_0^x f(u) du$ ,  $x \in \mathbb{R}_+$ , then there exists a unique element  $t_0 \in \Theta$  such that  $\mathbb{E}(\psi(\xi, t_0)) = 0$  holds. Mathieu [14] also noted that the assumptions under which existence and uniqueness of a solution in question was established are not the minimal ones, but he has not searched for possible minimal assumptions.

Note that one can rewrite  $\psi$  given by (3.1) as  $\psi(x, t) = \tilde{f}(x - t)$ ,  $x, t \in \mathbb{R}$ , where  $\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}$  denotes the odd extension of  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  to  $\mathbb{R}$ , which is given by

$$(3.2) \quad \tilde{f}(z) := \begin{cases} f(z) & \text{if } z > 0, \\ 0 & \text{if } z = 0, \\ -f(-z) & \text{if } z < 0. \end{cases}$$

As a new result, we have the following proposition.

**3.5 Proposition.** *If  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is continuous and strictly increasing with  $f(0) = 0$  and  $\lim_{z \rightarrow \infty} f(z) \in (0, \infty)$ , then, for any random variable  $\xi$ , we have that  $\mathbb{E}(|\psi(\xi, t)|) < \infty$ ,  $t \in \mathbb{R}$ , and the equation*

$$\mathbb{E}(\psi(\xi, t)) = \mathbb{E}(\tilde{f}(\xi - t)) = 0$$

*has a unique solution with respect to  $t \in \mathbb{R}$ .*

Finally, we compare the assumptions of Lemma 2 in Mathieu [14] and those of Proposition 3.5. Note that if a function  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  admits the properties (a)-(d), then it is not necessarily strictly increasing (for example, it may happen that  $f(x) = f(\beta)$  for  $x \geq \beta$ , see, e.g., the Huber function (4.9)), so we cannot say that Proposition 3.5 is a generalization



of Lemma 2 in Mathieu [14]. However, the conditions of Proposition 3.5 might be checked more easily than those of Lemma 2 in Mathieu [14] in order to prove that the equation  $\mathbb{E}(\psi(\xi, t)) = 0$  have a unique solution with respect to  $t \in \mathbb{R}$ . For example, if  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $f(z) := z/\sqrt{1+z^2/2}$ ,  $z \in \mathbb{R}_+$ , then  $f$  is a continuous and strictly increasing function starting from 0 and  $\lim_{z \rightarrow \infty} f(z) = \sqrt{2}$ . Indeed, we have  $f'(z) = (1+z^2/2)^{-3/2} > 0$  for each  $z \in \mathbb{R}_+$ . This special choice of  $f$  plays a role in robust statistics, for more details, see, e.g., Rey [18, Section 6.4] or Example 4.4.  $\square$

In the next remark, restricted to a one-dimensional parameter set, we recall Theorem 3.2 in Clarke [4] on the local uniqueness of a root of the function (1.4). We will see that the assumptions of Theorem 3.2 in Clarke [4] are much more involved and quite different compared to those of our Theorem 3.1.

**3.6 Remark.** Let  $X := \mathbb{R}$ ,  $\Theta$  be a non-degenerate open interval of  $\mathbb{R}$ , and  $\psi : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$  be a measurable function in its first variable. Given  $t_0 \in \Theta$  and a distribution function  $G : \mathbb{R} \rightarrow [0, 1]$ , let

$$I(\psi, G) := \left\{ t \in \Theta : \int_{\mathbb{R}} \psi(x, t) dG(x) = 0 \right\},$$

and let  $T(\psi, G) \in \Theta$  be a solution of the minimization problem

$$\min_{t \in I(\psi, G)} |t - t_0|,$$

provided that  $I(\psi, G)$  is nonempty.

In statistical estimation theory, a family of distribution functions  $\{F_t : \mathbb{R} \rightarrow [0, 1] : t \in \Theta\}$  is given, and one chooses  $G := F_{t_0}$  in the minimization problem above. We also note that the minimization problem above is a special case of a more general one given in (1.3) in Clarke [4], where a so-called selection function  $\varrho : \mathcal{E} \times \Theta \rightarrow \mathbb{R}$  comes into play, where  $\mathcal{E}$  denotes the set of distribution functions on  $\mathbb{R}$ . Namely, the minimization problem in (1.3) in Clarke [4] with the selection function  $\varrho(G, t) := |t - t_0|$ ,  $G \in \mathcal{E}$ ,  $t \in \Theta$ , gives the minimization problem above. In Clarke [5] one can find several interesting examples for other selection functionals, for example,  $\varrho_1(G, t) := \int_{\mathbb{R}} (G(x) - F_t(x))^2 dK(x)$ ,  $G \in \mathcal{E}$ ,  $t \in \Theta$ , where  $K : \mathbb{R} \rightarrow \mathbb{R}_+$  is a suitable weight function, or  $\varrho_2(G, t) := |\text{Med}(G) - t|$ ,  $G \in \mathcal{E}$ ,  $t \in \Theta$ , where  $\text{Med}(G)$  denotes the median of  $G$  (provided that it exists uniquely). Note that  $\varrho_2$  selects the root of  $\int_{\mathbb{R}} \psi(x, t) dG(x) = 0$ ,  $t \in \Theta$ , which is the closest to the median of  $G$ .

Given  $t_0 \in \Theta$  and a distribution function  $G : \mathbb{R} \rightarrow [0, 1]$ , suppose the following assumptions:

- (A0)  $t_0 \in I(\psi, G)$ . In this case, we have  $T(\psi, G) = t_0$ .
- (A1)  $\psi$  has a continuous partial derivative with respect to its second variable on  $\mathbb{R} \times D$ , where  $D \subseteq \Theta$  is a non-degenerate compact interval containing  $t_0$  in its interior.
- (A2) there exist a function  $g : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $\varepsilon > 0$  such that

- $|\psi(x, t)| \leq g(x)$  for each  $x \in \mathbb{R}$  and  $t \in D$ ,

- $|\partial_2\psi(x, t)| \leq g(x)$  for each  $x \in \mathbb{R}$  and  $t \in D$ ,
- $\int_{\mathbb{R}} g(x) dF(x) < \infty$  for each  $F \in K(G, \varepsilon)$ , where  $F \in K(G, \varepsilon)$  denotes the open neighbourhood of  $G$  with radius  $\varepsilon$  with respect to the Kolmogorov's distance  $d_K$  of distribution functions given by  $d_K(F, \tilde{F}) := \sup_{x \in \mathbb{R}} |F(x) - \tilde{F}(x)|$  for distribution functions  $F$  and  $\tilde{F}$ .

$$(A3) \int_{\mathbb{R}} \partial_2\psi(x, t_0) dG(x) \neq 0.$$

(A4) for each  $\delta > 0$  there exists  $\varepsilon > 0$  such that for each  $F \in K(G, \varepsilon)$ , we have that

$$\sup_{t \in D} \left| \int_{\mathbb{R}} \psi(x, t) dF(x) - \int_{\mathbb{R}} \psi(x, t) dG(x) \right| < \delta$$

and

$$\sup_{t \in D} \left| \int_{\mathbb{R}} \partial_2\psi(x, t) dF(x) - \int_{\mathbb{R}} \partial_2\psi(x, t) dG(x) \right| < \delta.$$

In particular, Condition (A4) implies that, for each  $t \in D$ , the functionals  $\mathcal{E} \ni F \mapsto \int_{\mathbb{R}} \psi(x, t) dF(x)$  and  $\mathcal{E} \ni F \mapsto \int_{\mathbb{R}} \partial_2\psi(x, t) dF(x)$  are continuous at  $G$  with respect to the Kolmogorov's distance  $d_K$ .

Given  $t_0 \in \Theta$  and a distribution function  $G : \mathbb{R} \rightarrow [0, 1]$ , under the assumptions (A0) – (A4), Clarke [4, Theorem 3.2] proved that for any  $\kappa > 0$  there exists an  $\varepsilon > 0$  such that  $T(\psi, F)$  exists for each  $F \in K(G, \varepsilon)$  and  $T(\psi, F) \in (t_0 - \kappa, t_0 + \kappa)$ . Further, for this  $\varepsilon$  there exists a  $\kappa^* > 0$  such that

$$I(\psi, F) \cap (t_0 - \kappa^*, t_0 + \kappa^*) = T(\psi, F),$$

that is, the equation  $\int_{\mathbb{R}} \psi(x, t) F(dx) = 0$ ,  $t \in \Theta$ , has a unique solution in the interval  $(t_0 - \kappa^*, t_0 + \kappa^*)$ . In particular, by choosing  $F := G$ , one can see that the equation  $\int_{\mathbb{R}} \psi(x, t) dG(x) = 0$ ,  $t \in \Theta$ , has a unique solution *locally*, provided that it has a solution.

We note that, if the Kolmogorov's distance in the assumptions (A2) and (A4) is replaced by the Lévy's distance for distribution functions, then an analogous statement holds (see Clarke [4, page 1197]).  $\square$

## 4 Examples from statistical estimation theory

In this section, we present several examples from statistical estimation theory, where our results in Sections 2 and 3 can be well-applied. For example, we consider the cases of the empirical median, the empirical quantiles, the empirical expectiles,  $\psi$ -estimators recently used by Mathieu [14], some  $\psi$ -estimators that are important in robust statistics, and we also study some examples from maximum likelihood theory.

**4.1 Example. (Empirical median)** Let  $X := \mathbb{R}$ ,  $\Theta := \mathbb{R}$  and  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $\psi(x, t) := \text{sign}(x - t)$ ,  $x, t \in \mathbb{R}$ . For each  $x \in \mathbb{R}$ , the function  $\mathbb{R} \ni t \mapsto \psi(x, t)$  is decreasing, but not strictly decreasing. Then for each  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , the equation (1.2) takes the form

$$(4.1) \quad \sum_{i=1}^n \text{sign}(x_i - t) = 0, \quad t \in \mathbb{R}.$$

In this special case, the function  $\psi$  is not continuous in its second variable, and the corresponding equation (4.1) has an important role in statistics. Namely, one can check that  $\text{Med}_n : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$(4.2) \quad \text{Med}_n(x_1, \dots, x_n) := \frac{1}{2} \left( x_{\lceil \frac{n}{2} \rceil}^* + x_{\lfloor \frac{n}{2} + 1 \rfloor}^* \right) = \begin{cases} x_{k+1}^* & \text{if } n = 2k + 1, \\ \frac{1}{2}(x_k^* + x_{k+1}^*) & \text{if } n = 2k, \end{cases} \quad k \in \mathbb{Z}_+,$$

is a solution of the equation (4.1), where  $x_1^* \leq x_2^* \leq \dots \leq x_n^*$  denotes the ordered sample of  $x_1, \dots, x_n \in \mathbb{R}$ , and  $\lceil z \rceil$  and  $\lfloor z \rfloor$  denote the upper and lower integer parts of a real number  $z \in \mathbb{R}$ , respectively. Of course, if  $n = 2k$ , then there are other solutions of the equation (4.1). For example, if  $x_1 < x_2 < \dots < x_{2k}$ , then we have

$$(4.3) \quad \left\{ t \in \mathbb{R} : \sum_{i=1}^{2k} \text{sign}(x_i - t) = 0 \right\} = [x_k, x_{k+1}],$$

where  $[x_k, x_{k+1}]$  is not a singleton. Note that  $\text{Med}_n(x_1, \dots, x_n)$  is nothing else but the well-known empirical median of  $x_1, \dots, x_n$ .

Further, we have that  $\psi$  is a  $T_n$ -function for each  $n = 2k + 1$ ,  $k \in \mathbb{Z}_+$ , with  $\vartheta_n(\mathbf{x}) = x_{k+1}^*$ ,  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ . Note also that  $\psi$  is not a  $T_n$ -function for any  $n = 2k$ ,  $k \in \mathbb{N}$ . Indeed, if  $x_1 < x_2 < \dots < x_{2k}$ , then (4.3) implies that the function  $\psi_{(x_1, \dots, x_{2k})}$  does not have a point of sign change. Furthermore, for each  $x, y \in \mathbb{R}$  with  $\vartheta_1(x) = x < y = \vartheta_1(y)$ , the function (2.4) takes the form

$$(x, y) \ni t \mapsto \frac{\psi(x, t)}{\psi(x, t) - \psi(y, t)} = \frac{\text{sign}(x - t)}{\text{sign}(x - t) - \text{sign}(y - t)} = \frac{-1}{-1 - 1} = \frac{1}{2}.$$

This function is a rational constant, in particular strictly  $\frac{1}{\ell}$ -increasing for each  $\ell \in \mathbb{N}$ . It underlines the fact that the strictly  $\frac{1}{n}$ -increasing property of the function (2.4) in Proposition 2.13 is only a necessary, but not a sufficient condition for  $\psi$  being a  $T_n$ -function. Indeed, in the present example, the function (2.4) is strictly  $\frac{1}{\ell}$ -increasing for each  $\ell \in \mathbb{N}$ , but it is not a  $T_n$ -function for  $n = 2k$ ,  $k \in \mathbb{N}$ . Moreover, for each  $x, y \in \mathbb{R}$  with  $\vartheta_1(x) = x < y = \vartheta_1(y)$ , the function (2.3) takes the form

$$(x, y) \ni t \mapsto -\frac{\psi(x, t)}{\psi(y, t)} = -\frac{\text{sign}(x - t)}{\text{sign}(y - t)} = 1.$$

This function is a rational constant, in particular, increasing, but not strictly increasing. It is in accordance with part (iv) of Theorem 2.10 and part (i) of Proposition 2.12 as well, since  $\psi$  is a  $T_n$ -function for infinitely many  $n \in \mathbb{N}$ , namely, for each  $n = 2k + 1$ ,  $k \in \mathbb{N}$ ; and for each  $x \in \mathbb{R}$ ,

the function  $\mathbb{R} \ni t \mapsto \psi(x, t)$  is decreasing. Note also that it does not hold that for each  $m \in \mathbb{N}$  there exists an  $n \in \mathbb{N}$  such that  $m$  divides  $n$  and  $\psi$  is a  $T_n$ -function (indeed, in case of an even  $m \in \mathbb{N}$  one cannot choose such an  $n$ ). Moreover, one cannot apply part (vi) of Theorem 2.10. This underlines that the increasing property of the function (2.3) is only a necessary, but not a sufficient condition in order that  $\psi$  be a  $T_n$ -function for each  $n \in \mathbb{N}$ . Finally, we mention that the present example also shows that in Proposition 2.14 the restriction that  $m \in \{1, \dots, n\}$  divides  $n$  cannot be removed in general.  $\square$

**4.2 Example. (Empirical quantiles)** Given  $\alpha \in (0, 1)$ ,  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , an empirical  $\alpha$ -quantile based on  $x_1, \dots, x_n$  is defined as any solution of the minimization problem:

$$(4.4) \quad \begin{aligned} \min_{t \in \mathbb{R}} \sum_{i=1}^n \varphi_\alpha(x_i - t) &= \min_{t \in \mathbb{R}} \sum_{i=1}^n \left( \alpha \mathbf{1}_{\{x_i \geq t\}} + (\alpha - 1) \mathbf{1}_{\{x_i < t\}} \right) (x_i - t) \\ &= \min_{t \in \mathbb{R}} \sum_{i=1}^n \frac{1}{2} \left( |x_i - t| + (2\alpha - 1)(x_i - t) \right), \end{aligned}$$

where  $\varphi_\alpha : \mathbb{R} \rightarrow \mathbb{R}$  is the so-called  $\alpha$ -quantile check function given by

$$\varphi_\alpha(x) := |\alpha - \mathbf{1}_{\{x < 0\}}| |x| = (\alpha \mathbf{1}_{\{x \geq 0\}} + (\alpha - 1) \mathbf{1}_{\{x < 0\}}) x = \frac{1}{2} \left( |x| + (2\alpha - 1)x \right), \quad x \in \mathbb{R},$$

see, e.g., Koenker and Bassett [11, Section 3]. Some authors call a solution of (4.4) as a geometric  $\alpha$ -quantile, see, e.g., Passeggeri and Reid [17, Section 2].

It is known that  $q \in \mathbb{R}$  is an empirical  $\alpha$ -quantile based on  $x_1, \dots, x_n$  if and only if the following two inequalities hold

$$(4.5) \quad \frac{1}{n} \sum_{i: x_i < q} 1 \leq \alpha \quad \text{and} \quad \alpha \leq \frac{1}{n} \sum_{i: x_i \leq q} 1,$$

see, e.g., Lange [13, Problems 12.12/13.].

It is also known that, given  $\alpha \in (0, 1)$ ,  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , an empirical  $\alpha$ -quantile given as a solution of the minimization problem (4.4) is uniquely defined if and only if  $\alpha \notin \{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}\}$ , and in case of uniqueness, we have that it is given by

$$(4.6) \quad q_n^{(\alpha)}(x_1, \dots, x_n) := \frac{1}{2} (x_{\lceil n\alpha \rceil}^* + x_{\lfloor n\alpha + 1 \rfloor}^*),$$

see Passeggeri and Reid [17, Lemma 4.1]. We also mention an interesting result of Passeggeri and Reid [17, Lemma 4.2], which states that, given  $\alpha \in (0, 1)$  and  $n \in \mathbb{N}$ , the function  $q_n^{(\alpha)} : \mathbb{R}^n \rightarrow \mathbb{R}$  given by (4.6) is Lipschitz continuous:

$$|q_n^{(\alpha)}(x_1, \dots, x_n) - q_n^{(\alpha)}(y_1, \dots, y_n)| \leq \max_{j \in \{1, \dots, n\}} |x_j - y_j|$$

for each  $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$ .

Note that if  $\alpha = \frac{1}{2}$ , then the empirical median  $\text{Med}_n(x_1, \dots, x_n)$  of  $x_1, \dots, x_n$  given in (4.2) is a solution of the minimization problem

$$\min_{t \in \mathbb{R}} \sum_{i=1}^n |x_i - t|.$$

Indeed, if  $\alpha = \frac{1}{2}$ , then  $\varphi_{1/2}(x) = \frac{1}{2}|x|$ ,  $x \in \mathbb{R}$ , and hence the minimization problem (4.4) with  $\alpha = \frac{1}{2}$  is equivalent to  $\min_{t \in \mathbb{R}} \sum_{i=1}^n |x_i - t|$ , and, as we have recalled,  $q \in \mathbb{R}$  is a solution of this minimization problem if and only if the inequalities (4.5) hold for  $q$  with  $\alpha = \frac{1}{2}$ . Further, one can easily check that the empirical median  $\text{Med}_n(x_1, \dots, x_n)$  based on  $x_1, \dots, x_n$  satisfies the inequalities (4.5) with  $\alpha = \frac{1}{2}$ .

Motivated by the minimization problem (4.4), we investigate the function  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$(4.7) \quad \psi(x, t) := \begin{cases} \alpha & \text{if } x > t, \\ 0 & \text{if } x = t, \\ \alpha - 1 & \text{if } x < t. \end{cases}$$

For each  $x \in \mathbb{R}$ , the function  $\mathbb{R} \ni t \mapsto \psi(x, t)$  is decreasing, but not strictly decreasing. By choosing  $X := \Theta := \mathbb{R}$ , we have that  $\psi$  is a  $T_1$ -function with  $\vartheta_1(x) := x$ ,  $x \in \mathbb{R}$ . Analogously to the result of Passeggeri and Reid [17, Lemma 4.1], we are going to show the following result.

**4.2 Proposition.** *Given  $\alpha \in (0, 1)$ , for each  $n \geq 2$ , the function  $\psi$  defined by (4.7) has the property  $(T_n)$  if and only if  $\alpha \notin \{\frac{1}{n}, \dots, \frac{n-1}{n}\}$ .*

□

**4.3 Example. (Empirical expectiles)** Let  $\alpha \in (0, 1)$ ,  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ . The empirical  $\alpha$ -expectile based on  $x_1, \dots, x_n$  is defined as any solution of the minimization problem:

$$\min_{t \in \mathbb{R}} \sum_{i=1}^n \tilde{\varphi}_\alpha(x_i - t) = \min_{t \in \mathbb{R}} \sum_{i=1}^n \left( \alpha \mathbf{1}_{\{x_i \geq t\}} + (1 - \alpha) \mathbf{1}_{\{x_i < t\}} \right) (x_i - t)^2,$$

where  $\tilde{\varphi}_\alpha : \mathbb{R} \rightarrow \mathbb{R}$  is given by

$$\tilde{\varphi}_\alpha(x) := |\alpha - \mathbf{1}_{\{x < 0\}}| x^2 = (\alpha \mathbf{1}_{\{x \geq 0\}} + (1 - \alpha) \mathbf{1}_{\{x < 0\}}) x^2, \quad x \in \mathbb{R},$$

see, e.g., Newey and Powell [15]. Expectiles are also called smoothed versions of quantiles or least asymptotically weighted squares estimators.

Motivated by this minimization problem, we may investigate the applicability of Theorem 2.10 for the function  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\psi(x, t) := \begin{cases} \alpha(x - t) & \text{if } x > t, \\ 0 & \text{if } x = t, \\ (1 - \alpha)(x - t) & \text{if } x < t. \end{cases}$$

For each  $x \in \mathbb{R}$ , the function  $\mathbb{R} \ni t \mapsto \psi(x, t)$  is strictly decreasing. By choosing  $X := \mathbb{R}$  and  $\Theta := \mathbb{R}$ , we have  $\psi$  is a  $T_1$ -function with  $\vartheta_1(x) := x$ ,  $x \in \mathbb{R}$ , and for each  $x, y \in \mathbb{R}$  with  $\vartheta_1(x) = x < y = \vartheta_1(y)$ , the function (2.3) takes the form

$$(x, y) \ni t \mapsto -\frac{\psi(x, t)}{\psi(y, t)} = -\frac{(1 - \alpha)(x - t)}{\alpha(y - t)} = -\frac{1 - \alpha}{\alpha} \left(1 - \frac{y - x}{y - t}\right),$$

which is a strictly increasing function. It is in accordance with part (i) of Proposition 2.12. Since  $\psi(x, \vartheta_1(x)) = 0$ ,  $x \in \mathbb{R}$ , by part (vi) of Theorem 2.10, we have  $\psi$  is a  $T_n^\lambda$ -function for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ . In particular, using that  $\psi$  is continuous in its second variable, we also have, for each  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , that the equation  $\sum_{i=1}^n \psi(x_i, t) = 0$ ,  $t \in \mathbb{R}$ , has a unique solution, see part (ii) of Remark 2.4.  $\square$

Next, we apply Theorem 2.10 for a function  $\psi$  that has been recently used for constructing M-estimators by Mathieu [14] (see also Example 3.5).

**4.4 Example.** Let  $X := \mathbb{R}$ ,  $\Theta := \mathbb{R}$  and  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$(4.8) \quad \psi(x, t) := \text{sign}(x - t)f(|x - t|), \quad x, t \in \mathbb{R},$$

where  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  is an arbitrary function with  $f(0) = 0$ . Note that if  $f$  is continuous as well, then  $\psi$  is continuous in its second variable. Next, we discuss the applicability of Theorem 2.10 for the function  $\psi$  given in (4.8). Namely, we prove the following statement.

**4.4 Proposition.** *Let  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  be a function with  $f(0) = 0$  and let  $\psi$  be given by (4.8). Then we have that*

- (a)  $\psi$  is a  $T_1$ -function if and only if  $f(z) > 0$  for all  $z > 0$ , and, in this case,  $\vartheta_1(x) = x$  and  $\psi(x, \vartheta_1(x)) = 0$  hold for all  $x \in \mathbb{R}$ .
- (b) if  $\psi$  is a  $T_1$ -function and  $T_n$ -function for infinitely many  $n \in \mathbb{N}$ , then  $f$  is increasing.
- (c)  $\psi$  is a  $T_2^\lambda$ -function for each  $\lambda \in \Lambda_2$  if and only if  $f$  is strictly increasing.
- (d)  $\psi$  is a  $T_n^\lambda$ -function for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$  if and only if  $f$  is strictly increasing.

As a consequence of Proposition 4.4, if  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a strictly increasing and continuous function such that  $f(0) = 0$ , then  $\psi$  given in (4.8) is continuous in its second variable, and hence for each  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , the equation  $\sum_{i=1}^n \psi(x_i, t) = 0$ ,  $t \in \mathbb{R}$  (i.e., the equation (1.2)) has a unique solution.

To complete this example, we recall some special choices for the function  $f$  appearing in (4.8) such that the corresponding  $\psi$ -estimator has an important role in (robust) statistics:

- (i) the Huber function  $f_H : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,

$$(4.9) \quad f_H(z) := z\mathbf{1}_{\{z \leq \beta\}} + \beta\mathbf{1}_{\{z > \beta\}}, \quad z \in \mathbb{R}_+,$$

where  $\beta > 0$  (see Huber [8]), which is a continuous and increasing (but not strictly increasing) function starting from 0. Then the function  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  given in (4.8) takes the form

$$\psi(x, t) = \begin{cases} x - t & \text{if } |x - t| \leq \beta, \\ \beta \operatorname{sign}(x - t) & \text{if } |x - t| > \beta, \end{cases} \quad x, t \in \mathbb{R}.$$

In general,  $\psi$  is not a  $T_2$ -function. Indeed, for example, if  $\beta := 1$ ,  $x_1 := 0$  and  $x_2 := 3$ , then

$$psi_{(x_1, x_2)}(t) = \sum_{i=1}^2 \psi(x_i, t) = \begin{cases} 1 + 1 = 2 > 0 & \text{if } t \leq -1, \\ -t + 1 > 0 & \text{if } -1 < t \leq 1, \\ -1 + 1 = 0 & \text{if } 1 < t \leq 2, \\ -1 + 3 - t = 2 - t < 0 & \text{if } 2 < t < 4, \\ -1 - 1 = -2 < 0 & \text{if } t \geq 4. \end{cases}$$

Consequently, the function  $\psi_{(x_1, x_2)}$  does not have a point of sign change, and, thus  $\psi$  is not a  $T_2$ -function, and, by Proposition 2.14, it yields that  $\psi$  is not a  $T_{2k}$ -function for any  $k \in \mathbb{N}$ . On the other hand, for each  $k \in \mathbb{N}$ ,  $\psi$  is a  $T_{2k+1}$ -function. Indeed, let  $k \in \mathbb{N}$  and  $x_1, \dots, x_{2k+1} \in \mathbb{R}$  be arbitrary. Then the function  $\mathbb{R} \ni t \mapsto \sum_{i=1}^{2k+1} \psi(x_i, t)$  is decreasing (because each of its terms is decreasing), and, on the contrary, let us assume that it does not have a point of sign change. Then it is constant on a proper subinterval  $I$  of  $\mathbb{R}$ . In this case, each term in question must be also constant on this subinterval, and, by taking into account the form of  $\psi$ , it must be equal to  $\beta$  or to  $-\beta$  on  $I$ . However, a  $(2k+1)$ -term sum whose terms are equal to  $\beta$  or to  $-\beta$  cannot be equal to zero, which leads us to a contradiction. All in all,  $\psi$  is a  $T_n$ -function if  $n \in \mathbb{N}$  is odd, and  $\psi$  is not a  $T_n$ -function if  $n \in \mathbb{N}$  is even.

(ii) the Catoni function  $f_C : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,

$$f_C(z) := \ln \left( 1 + \frac{z}{b} + \frac{1}{2} \left( \frac{z}{b} \right)^2 \right), \quad z \in \mathbb{R}_+,$$

where  $b > 0$  (see Catoni [2]), which is continuous and strictly increasing starting from 0.

(iii) a polynomial function  $f_P : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,

$$f_P(z) := \frac{z}{1 + \left( \frac{z}{\beta} \right)^{1 - \frac{1}{p}}}, \quad z \in \mathbb{R}_+,$$

where  $p \in \mathbb{N}$  and  $\beta > 0$ , which is continuous and strictly increasing starting from 0.

(iv) another Catoni-type function  $\tilde{f}_C : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,

$$\tilde{f}_C(z) := \ln \left( 1 + z + \frac{z^\alpha}{\alpha} \right), \quad z \in \mathbb{R}_+,$$

where  $\alpha \in (1, 2)$  (see Chen et al. [3]), which is continuous and strictly increasing starting from 0.

(v)  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $f(z) := z/\sqrt{1+z^2/2}$ ,  $z \in \mathbb{R}_+$ , which is a continuous and strictly increasing function starting from 0. Indeed, we have  $f'(z) = (1+z^2/2)^{-3/2} > 0$  for each  $z \in \mathbb{R}_+$ . Then the function  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  given in (4.8) takes the form

$$\psi(x, t) = \frac{x - t}{\sqrt{1 + \frac{(x-t)^2}{2}}}, \quad x, t \in \mathbb{R}.$$

In robust statistics, one calls  $\psi$  as the  $L_1$ - $L_2$  function, see, e.g., Rey [18, Section 6.4].

(vi)  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $f(z) := z/(1+z)$ ,  $z \in \mathbb{R}_+$ , which is a continuous and strictly increasing function starting from 0. Then the function  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  given in (4.8) takes the form

$$\psi(x, t) = \frac{x - t}{1 + |x - t|}, \quad x, t \in \mathbb{R},$$

which is called a "fair"-type function in robust statistics, see, e.g., Rey [18, Section 6.4].

In the above recalled special cases (ii)–(vi), by part (d) of Proposition 4.4, the corresponding function  $\psi$  given in (4.8) is a  $T_n^\lambda$ -function for all  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ . In particular, since in the cases (ii)–(vi),  $\psi$  is continuous in its second variable, for each  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , the equation (1.2) with the given function  $\psi$  has a unique solution, see part (ii) of Remark 2.4.  $\square$

In the remaining part of this section, we investigate the applicability of Theorem 2.10 for finding solutions of likelihood equations in the theory of MLEs. Let  $\Theta$  be a nondegenerate open interval of  $\mathbb{R}$ , and  $f : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$  be a function such that for each  $t \in \Theta$ , the function  $\mathbb{R} \ni x \mapsto f(x, t)$  is a density function. Let us introduce the set

$$\mathcal{X}_f := \{x \in \mathbb{R} : f(x, t) > 0, \forall t \in \Theta\},$$

and suppose that  $\mathcal{X}_f$  is nonempty. Note that, in general, it can happen that  $\mathcal{X}_f = \emptyset$ . For example, if  $\Theta = (0, \infty)$  and  $f : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$ ,

$$f(x, t) := \begin{cases} \frac{1}{t} & \text{if } x \in (0, t), \\ 0 & \text{if } x \notin (0, t), \end{cases} \quad t \in (0, \infty),$$

i.e., for each  $t \in (0, \infty)$ , the function  $\mathbb{R} \ni x \mapsto f(x, t)$  is the density function of a uniformly distributed random variable on  $(0, t)$ , then  $\mathcal{X}_f = \emptyset$ . Turning back to the case when  $\mathcal{X}_f \neq \emptyset$ , and supposing that  $\partial_2 f$  exists, the equation (1.2) with  $X := \mathcal{X}_f$  and the function  $\psi : \mathcal{X}_f \times \Theta \rightarrow \mathbb{R}$  defined by

$$(4.10) \quad \psi(x, t) := \partial_2(\ln(f(x, t))) = \frac{\partial_2 f(x, t)}{f(x, t)}, \quad (x, t) \in \mathcal{X}_f \times \Theta,$$

is nothing else but the likelihood equation based on the observations  $x_1, \dots, x_n$  in the theory of MLEs.

In the next examples, we demonstrate the applicability of Theorem 2.10 together with Proposition 2.12 for proving existence and uniqueness of a solution of the likelihood equation (1.2) corresponding to the function  $\psi$  given in (4.10).



**4.5 Example.** Let  $\xi$  be a normally distributed random variable with mean  $m \in \mathbb{R}$  and with variance  $\sigma^2$ , where  $\sigma > 0$ . Let  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$  be a realization of a sample of size  $n$  for  $\xi$ . Here by a sample of size  $n$ , we mean independent and identically distributed random variables  $\xi_1, \dots, \xi_n$  with common distribution as that of  $\xi$ . It is known that, supposing that  $\sigma$  is known, there exists a unique MLE of  $m$  based on  $x_1, \dots, x_n \in \mathbb{R}$ , and it takes the form  $\widehat{m}_n := \frac{x_1 + \dots + x_n}{n}$ . We will establish the existence and uniqueness of a solution of the corresponding likelihood equation using Theorem 2.10 together with Proposition 2.12. In this case, we have  $\Theta = \mathbb{R}$  and  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$f(x, m) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x, m \in \mathbb{R},$$

and consequently  $\mathcal{X}_f = \mathbb{R}$ . Then  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\psi(x, m) = \frac{\frac{1}{\sigma^2}(x-m)f(x, m)}{f(x, m)} = \frac{1}{\sigma^2}(x-m), \quad x, m \in \mathbb{R}.$$

Hence  $\psi$  is a  $T_1$ -function with  $\vartheta_1(x) := x$ ,  $x \in \mathbb{R}$ ,  $\psi(x, \vartheta_1(x)) = 0$ ,  $x \in \mathbb{R}$ , and  $\psi$  is continuous in its second variable. Further, using Proposition 2.12 and Theorem 2.10 (with  $X := \mathcal{X}_f = \mathbb{R}$ ), we can conclude that for each  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , the (likelihood) equation (1.2) with the given function  $\psi$  has a unique solution, which is equal to  $\vartheta_n(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n} = \widehat{m}_n$ , as desired.

It is also known that, supposing that  $m$  is known, there exists a unique MLE of  $\sigma^2$  based on  $x_1, \dots, x_n \in \mathbb{R}$ , and it takes the form  $\widehat{\sigma}_n^2 := \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$ . We will establish the existence and uniqueness of a solution of the corresponding likelihood equation using Theorem 2.10. In this case, we have  $\Theta = (0, \infty)$ , and  $f : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$ ,

$$f(x, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, \sigma^2 > 0,$$

and consequently  $\mathcal{X}_f = \mathbb{R}$ . Then  $\psi : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \psi(x, \sigma^2) &= \frac{1}{f(x, \sigma^2)} \left( \frac{1}{\sqrt{2\pi}} \left( -\frac{1}{2} \right) (\sigma^2)^{-\frac{3}{2}} e^{-\frac{(x-m)^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \frac{(x-m)^2}{2} (\sigma^2)^{-2} \right) \\ &= \frac{1}{2(\sigma^2)^2} ((x-m)^2 - \sigma^2), \quad x \in \mathbb{R}, \sigma^2 > 0. \end{aligned}$$

Hence  $\psi$  is a  $T_1$ -function with  $\vartheta_1(x) := (x-m)^2$ ,  $x \in \mathbb{R}$ ,  $\psi(x, \vartheta_1(x)) = 0$ ,  $x \in \mathbb{R}$ , and  $\psi$  is continuous in its second variable. Further, for each  $x, y \in \mathbb{R}$  with  $\vartheta_1(x) < \vartheta_1(y)$ , i.e.,  $(x-m)^2 < (y-m)^2$ , we get that the function

$$((x-m)^2, (y-m)^2) \ni \sigma^2 \mapsto -\frac{\psi(x, \sigma^2)}{\psi(y, \sigma^2)} = -\frac{(x-m)^2 - \sigma^2}{(y-m)^2 - \sigma^2} = -1 + \frac{(y-m)^2 - (x-m)^2}{(y-m)^2 - \sigma^2}$$

is strictly increasing. Consequently, by part (vi) of Theorem 2.10 (with  $X := \mathcal{X}_f = \mathbb{R}$ ), we conclude that for each  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$ , the (likelihood) equation (1.2) with the given

$\psi$  has a unique solution, which is equal to  $\vartheta_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 = \widehat{\sigma}_n^2$ , as desired. Finally, we present an alternative argument. Note that the equation (1.2) with the given function  $\psi$  has a solution if and only if the equation (1.2) with the function  $\tilde{\psi} : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$ ,

$$\tilde{\psi}(x, \sigma^2) := 2(\sigma^2)^2 \psi(x, \sigma^2) = (x - m)^2 - \sigma^2, \quad x \in \mathbb{R}, \sigma^2 > 0,$$

has a solution, and the two sets of solutions coincide. Further,  $\tilde{\psi}$  is a  $T_1$ -function and the corresponding  $\vartheta_1(x)$  is equal to  $(x - m)^2$  for each  $x \in \mathbb{R}$ . The function  $\tilde{\psi}$  is strictly decreasing in its second variable, and hence part (i) of Proposition 2.12 can be applied to  $\tilde{\psi}$ . This, together with part (vi) of Theorem 2.10, yield that  $\tilde{\psi}$  is a  $T_n$ -function for each  $n \in \mathbb{N}$ , and hence, trivially,  $\psi$  is a  $T_n$ -function for each  $n \in \mathbb{N}$  as well, as expected.  $\square$

**4.6 Example.** Let  $\alpha > 0$  and let  $\xi$  be an absolutely continuous random variable with a density function

$$f_\xi(x) := \begin{cases} 2\alpha x(1 - x^2)^{\alpha-1} & \text{if } x \in (0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

Then one can check that given  $n \in \mathbb{N}$  and a realization  $x_1, \dots, x_n \in (0, 1)$  of a sample of size  $n$  for  $\xi$ , there exists a unique MLE of  $\alpha$  and it takes the form

$$\widehat{\alpha}_n := -\frac{n}{\sum_{i=1}^n \ln(1 - x_i^2)}.$$

We will establish the existence and uniqueness of a solution of the corresponding likelihood equation using Theorem 2.10 together with Proposition 2.12. In this case, we have  $\Theta = (0, \infty)$  and  $f : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$ ,

$$f(x, \alpha) := \begin{cases} 2\alpha x(1 - x^2)^{\alpha-1} & \text{if } x \in (0, 1), \alpha > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and consequently  $\mathcal{X}_f = (0, 1)$ . Then  $\psi : (0, 1) \times (0, \infty) \rightarrow \mathbb{R}$ ,

$$\psi(x, \alpha) = \frac{2x \left( (1 - x^2)^{\alpha-1} + \alpha(1 - x^2)^{\alpha-1} \ln(1 - x^2) \right)}{2\alpha x(1 - x^2)^{\alpha-1}} = \frac{1}{\alpha} + \ln(1 - x^2), \quad x \in (0, 1), \alpha > 0.$$

Hence  $\psi$  is a  $T_1$ -function with  $\vartheta_1(x) := -\frac{1}{\ln(1-x^2)}$ ,  $x \in (0, 1)$ ,  $\psi(x, \vartheta_1(x)) = 0$ ,  $x \in (0, 1)$ , and  $\psi$  is continuous in its second variable. Further, using Theorem 2.10 and Proposition 2.12 (with  $X := \mathcal{X}_f = (0, 1)$ ), we can conclude that for each  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in (0, 1)$ , the (likelihood) equation (1.2) with the given  $\psi$  has a unique solution, which is equal to  $\vartheta_n(x_1, \dots, x_n) = -\frac{n}{\sum_{i=1}^n \ln(1-x_i^2)} = \widehat{\alpha}_n$ , as desired.  $\square$

**4.7 Example.** Let  $\xi$  be an absolutely continuous random variable with a density function

$$f_\xi(x) := \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

where  $m \in \mathbb{R}$  and  $\sigma > 0$ . Note that  $f_\xi$  is a mixture density function of the standard normally density function and the density function of a normally distributed random variable with mean  $m$  and variance  $\sigma^2$  with equal  $\frac{1}{2}$  weights. Let  $n \in \mathbb{N}$  and  $x_1, \dots, x_n \in \mathbb{R}$  be a realization of a sample of size  $n$  for  $\xi$ . In what follows, we assume that  $\sigma$  is known, and, using Theorem 2.10, we show that, in general, the corresponding likelihood equation for  $m$  may have more solutions. In this case, we have  $\Theta = \mathbb{R}$  and  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$f(x, m) := \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x, m \in \mathbb{R},$$

and consequently  $\mathcal{X}_f = \mathbb{R}$ . Then  $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\psi(x, m) = \frac{\frac{1}{2\sigma^3\sqrt{2\pi}}(x-m)e^{-\frac{(x-m)^2}{2\sigma^2}}}{f(x, m)}, \quad x, m \in \mathbb{R}.$$

Hence  $\psi$  is a  $T_1$ -function with  $\vartheta_1(x) := x$ ,  $x \in \mathbb{R}$ ,  $\psi(x, \vartheta_1(x)) = 0$ ,  $x \in \mathbb{R}$ , and  $\psi$  is continuous in its second variable. Consequently, for each  $n \in \mathbb{N}$  and  $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ , the likelihood equation (1.2) has at least one solution in  $\mathbb{R}$ . Further, for each  $x \in \mathbb{R}$ , we have  $\lim_{m \rightarrow \pm\infty} \psi(x, m) = 0$ . In what follows, we check that it is not true that for each  $x, y \in \mathbb{R}$  with  $\vartheta_1(x) < \vartheta_1(y)$ , i.e.,  $x < y$ , the function (2.3) is increasing. For each  $x, y \in \mathbb{R}$  with  $x < y$ , the function (2.3) takes the form

$$(x, y) \ni m \mapsto -\frac{\psi(x, m)}{\psi(y, m)} = \frac{m-x}{y-m} \cdot \frac{f(y, m)}{f(x, m)} \cdot \frac{e^{-\frac{(y-m)^2}{2\sigma^2}}}{e^{-\frac{(x-m)^2}{2\sigma^2}}} = \frac{(m-x) \left( \sigma e^{-\frac{y^2}{2} + \frac{(y-m)^2}{2\sigma^2}} + 1 \right)}{(y-m) \left( \sigma e^{-\frac{x^2}{2} + \frac{(x-m)^2}{2\sigma^2}} + 1 \right)},$$

and, for each  $m \in (x, y)$ , one can check that

$$(4.11) \quad \begin{aligned} \frac{d}{dm} \left( -\frac{\psi(x, m)}{\psi(y, m)} \right) &= \frac{1}{(y-m)^2 \left( \sigma e^{-\frac{x^2}{2} + \frac{(x-m)^2}{2\sigma^2}} + 1 \right)^2} \\ &\times \left[ (y-x) \left( \sigma e^{-\frac{x^2}{2} + \frac{(x-m)^2}{2\sigma^2}} + 1 \right) \left( \sigma e^{-\frac{y^2}{2} + \frac{(y-m)^2}{2\sigma^2}} + 1 \right) \right. \\ &\quad \left. - \frac{1}{\sigma} (m-x)(y-m)^2 e^{-\frac{y^2}{2} + \frac{(y-m)^2}{2\sigma^2}} \left( \sigma e^{-\frac{x^2}{2} + \frac{(x-m)^2}{2\sigma^2}} + 1 \right) \right. \\ &\quad \left. - \frac{1}{\sigma} (m-x)^2 (y-m) e^{-\frac{x^2}{2} + \frac{(x-m)^2}{2\sigma^2}} \left( \sigma e^{-\frac{y^2}{2} + \frac{(y-m)^2}{2\sigma^2}} + 1 \right) \right]. \end{aligned}$$

If for each  $x, y \in \mathbb{R}$  with  $x < y$ , the function (2.3) were increasing, then we would have that

$$\frac{d}{dm} \left( -\frac{\psi(x, m)}{\psi(y, m)} \right) \geq 0, \quad m \in (x, y), \quad x < y, \quad x, y \in \mathbb{R},$$

which, by (4.11), is equivalent to

$$\sigma(y-x) \geq \frac{(m-x)(y-m)^2}{\sigma + e^{\frac{y^2 - (y-m)^2}{2\sigma^2}}} + \frac{(m-x)^2(y-m)}{\sigma + e^{\frac{x^2 - (x-m)^2}{2\sigma^2}}}, \quad m \in (x, y), \quad x < y, \quad x, y \in \mathbb{R}.$$

However, this inequality does not hold in general, since its left hand side tends to 0 as  $\sigma \downarrow 0$ , but its right hand side tends to  $\infty$  as  $\sigma \downarrow 0$ . To give an example, for example, on Figure 1, we plotted the function  $(x, y) \ni m \mapsto -\frac{\psi(x, m)}{\psi(y, m)}$  with  $x = 1$ ,  $y = 5$  and  $\sigma = 1$ , which is not increasing. All in all, it is not true that for each  $x, y \in \mathbb{R}$  with  $x < y$ , the function (2.3) is

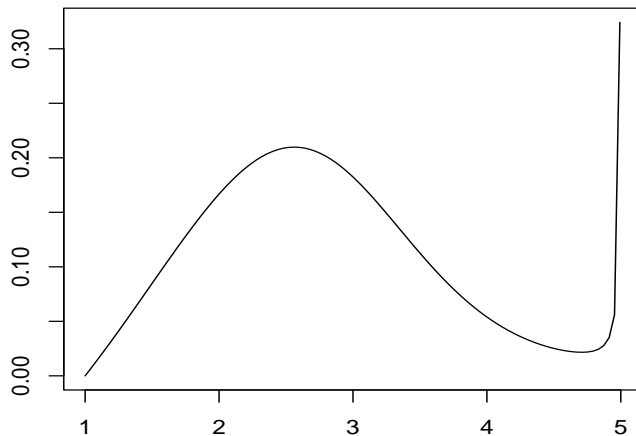


Figure 1: The function  $(1, 5) \ni m \mapsto -\frac{\psi(x, m)}{\psi(y, m)}$  with  $\sigma = 1$ .

increasing. Hence, by part (iv) of Theorem 2.10 (with  $X := \mathcal{X}_f = \mathbb{R}$ ), we get that there exists  $n_0 \in \mathbb{N}$  such that  $\psi$  is not a  $T_n$ -function for any  $n \geq n_0$ ,  $n \in \mathbb{N}$ . In particular, it yields that there exists  $n_0 \in \mathbb{N}$  such that for each  $n \geq n_0$ , there exist real numbers  $x_1, \dots, x_n \in \mathbb{R}$  such that the solution of the corresponding likelihood equation (1.2) (based on  $\mathbf{x} = (x_1, \dots, x_n)$ ) is not unique. By part (v) of Theorem 2.10, we also get that there exists a  $\boldsymbol{\lambda} \in \Lambda_2$  such that  $\psi$  is not a  $T_2^\lambda$ -function.  $\square$

## 5 Proofs for Section 2

**Proof of Lemma 2.7.** Assume that  $y$  is a level of increase for  $f$  and (i) and (ii) are not valid. Define

$$A := \{v \in \Theta \mid y \geq f(v)\} \quad \text{and} \quad B := \{u \in \Theta \mid y \leq f(u)\}.$$

Then  $A \cup B = \Theta$  and  $A, B$  are nonempty, since (i) and (ii) do not hold. If  $v \in A$  and  $u \in B$ , then the inequality  $u < v$  cannot be true, since  $y$  is a level of increase for  $f$ , thus  $v \leq u$  must be valid. Consequently,  $\sup A \leq \inf B$ , and, using that  $A \cup B = \Theta$ , we have that  $\sup A = \inf B =: \vartheta$ . If

$t < \vartheta$ , then  $t \notin B$ , which implies that  $y > f(t)$ . Similarly, if  $\vartheta < t$ , then  $t \notin A$ , which yields  $y < f(t)$ . Hence, we have proved that  $\vartheta$  is a point of sign change for the function  $y - f$ , i.e., (iii) must hold.

Conversely, if (i) or (ii) hold, then almost obviously, using the definition of a level of increase, we have that  $y$  is a level of increase for  $f$ . Finally, assume that (iii) is valid, i.e., there exists  $\vartheta \in \Theta$  which is a point of sign change for the function  $y - f$ , and, on the contrary, suppose that  $y$  is not a level of increase for  $f$ . Then there exist  $u, v \in \Theta$  such that  $u < v$ ,  $y \leq f(u)$  and  $y \geq f(v)$ . Therefore,  $y - f(u) \leq 0$  and  $y - f(v) \geq 0$ , which imply that  $v \leq \vartheta \leq u$  contradicting  $u < v$ .  $\square$

**Proof of Lemma 2.8.** Assume that the levels of increase for  $f$  form a dense subset in the convex hull of  $f(\Theta)$ , but  $f$  is not increasing. Then there exist  $u, v \in \Theta$  with  $u < v$  such that  $f(u) > f(v)$ . The convex hull of  $f(\Theta)$  contains the open interval  $(f(v), f(u))$ . Hence, by the assumption, one can find an element  $y \in (f(v), f(u))$  which is a level of increase for  $f$ . Then we have that  $u, v \in \Theta$  with  $u < v$  and  $y \leq f(u)$ , which imply  $y < f(v)$ , and this contradicts  $y \in (f(v), f(u))$ . The second statement of the lemma readily follows from the definitions of strictly increasing property and level of increase. For the last statement of the lemma, assume that  $u \in \Theta$  satisfies  $g(y) \leq (g \circ f)(u) = g(f(u))$ . Since  $g$  is strictly increasing and  $y, f(u) \in H$ , we have that  $y \leq f(u)$ . Using that  $y$  is a level of increase for  $f$ , for all  $v \in \Theta$  with  $u < v$ , we get that  $y < f(v)$ . Hence the strictly increasing property of  $g$  together with  $y, f(v) \in H$  yield that  $g(y) < g(f(v))$  for all  $v \in \Theta$  with  $u < v$ , as desired.  $\square$

**Proof of Lemma 2.9.** To verify the  $\varepsilon$ -increasingness property of  $f$ , let  $u, v \in \Theta$  be arbitrary such that  $u < v$ . The intervals  $J_1 := [y_0, y_1], \dots, J_n := [y_{n-1}, y_n]$  cover the image  $f(\Theta)$ . Therefore, for some  $i \in \{1, \dots, n\}$ , we have that  $f(u) \in J_i$ , which implies  $y_{i-1} \leq f(u) \leq y_i$ . Since  $y_{i-1}$  is a level of increase for  $f$ , we have that  $y_{i-1} < f(v)$ . On the other hand, due to the definition of  $\varepsilon$ , it follows that  $y_i \leq y_{i-1} + \varepsilon$ . Thus, we get

$$f(u) \leq y_i \leq y_{i-1} + \varepsilon < f(v) + \varepsilon,$$

which was to be proved.  $\square$

**Proof of Theorem 2.10.** (i): Let  $x, y \in X$  be such that  $\vartheta_1(x) < \vartheta_1(y)$ . Since  $\psi$  is a  $T_1$ -function, for all  $t \in (\vartheta_1(x), \vartheta_1(y))$ , we have  $\psi(x, t) < 0$  and  $\psi(y, t) > 0$ .

To the contrary, assume that  $\frac{\lambda_2}{\lambda_1}$  is not a level of increase for the function (2.3). Then there exist  $u, v \in \mathbb{R}$  such that  $\vartheta_1(x) < u < v < \vartheta_1(y)$  and

$$-\frac{\psi(x, v)}{\psi(y, v)} \leq \frac{\lambda_2}{\lambda_1} \leq -\frac{\psi(x, u)}{\psi(y, u)}.$$

Rearranging these inequalities, we get that

$$\lambda_1 \psi(x, u) + \lambda_2 \psi(y, u) \leq 0 \leq \lambda_1 \psi(x, v) + \lambda_2 \psi(y, v).$$

In view of the property  $(T_2^{(\lambda_1, \lambda_2)})$  of  $\psi$ , this implies that

$$v \leq \vartheta_{2, \psi}^{(\lambda_1, \lambda_2)}(x, y) \leq u,$$

which contradicts the inequality  $u < v$ .

To the contrary, assume that  $\frac{\lambda_1}{\lambda_2}$  is not a level of increase for the function (2.3). Similarly as before, we have that there exist  $u, v \in \mathbb{R}$  such that  $\vartheta_1(x) < u < v < \vartheta_1(y)$  and

$$\lambda_1\psi(y, u) + \lambda_2\psi(x, u) \leq 0 \leq \lambda_1\psi(y, v) + \lambda_2\psi(x, v).$$

In view of the property  $(T_2^{(\lambda_1, \lambda_2)})$  of  $\psi$ , this implies that  $v \leq \vartheta_{2, \psi}^{(\lambda_1, \lambda_2)}(y, x) \leq u$ , which again contradicts the inequality  $u < v$ .

(ii): Let  $\psi$  be a  $T_n^{(\lambda_1, \dots, \lambda_n)}$ -function for some  $n \in \mathbb{N} \setminus \{1\}$  and  $(\lambda_1, \dots, \lambda_n) \in (0, \infty)^n$ . Then for each  $k \in \{1, \dots, n-1\}$ , we have  $\psi$  is a  $T_2^{(\sum_{i=1}^k \lambda_i, \sum_{i=k+1}^n \lambda_i)}$ -function, since for each  $x, y \in X$  and  $k \in \{1, \dots, n-1\}$ , it holds that

$$\left( \sum_{i=1}^k \lambda_i \right) \psi(x, t) + \left( \sum_{i=k+1}^n \lambda_i \right) \psi(y, t) = \sum_{i=1}^n \lambda_i \psi(x_i, t), \quad t \in \Theta,$$

where  $x_i := x$ ,  $i \in \{1, \dots, k\}$  and  $x_i := y$ ,  $i \in \{k+1, \dots, n\}$ . Consequently, the assertion readily follows from part (i) of the present theorem.

(iii): If  $\psi$  is a  $T_n$ -function for some  $n \in \mathbb{N} \setminus \{1\}$ , then it is a  $T_n^{(\lambda_1, \dots, \lambda_n)}$ -function with  $\lambda_1 = \dots = \lambda_n := 1$ , thus the assertion follows from part (ii) of the present theorem.

(iv): Let  $(n_i)_{i \in \mathbb{N}} \subseteq \mathbb{N}$  be a strictly increasing sequence such that  $\psi$  is a  $T_{n_i}$ -function for all  $i \in \mathbb{N}$ . Let  $x, y \in X$  such that  $\vartheta_1(x) < \vartheta_1(y)$ . Then, by assertion (iii) of this theorem we have that the numbers

$$\left\{ \frac{k}{n_i - k} \mid k \in \{1, \dots, n_i - 1\}, i \in \mathbb{N} \right\} = \left\{ \frac{1}{n_i - 1}, \frac{2}{n_i - 2}, \dots, \frac{n_i - 2}{2}, n_i - 1 \mid i \in \mathbb{N} \right\}$$

are levels of increase for the function (2.3). We are going to apply Lemma 2.8. The convex hull of the range of the function (2.3) is contained in  $(0, \infty)$ , so if we check that the set  $\left\{ \frac{k}{n_i - k} \mid k \in \{1, \dots, n_i - 1\}, i \in \mathbb{N} \right\}$  is dense in  $(0, \infty)$ , then Lemma 2.8 will imply that the function (2.3) is increasing. Since the function  $g : (0, \infty) \rightarrow (0, 1)$ ,  $g(u) := \frac{u}{u+1}$ ,  $u > 0$ , is bijective, it is enough to check that the set

$$\left\{ g \left( \frac{k}{n_i - k} \right) \mid k \in \{1, \dots, n_i - 1\}, i \in \mathbb{N} \right\} = \left\{ \frac{k}{n_i} \mid k \in \{1, \dots, n_i - 1\}, i \in \mathbb{N} \right\}$$

is dense in  $g((0, \infty)) = (0, 1)$ . This readily follows, since if  $(a, b) \subseteq (0, 1)$  is an open interval, then there exists  $i_0 \in \mathbb{N}$  such that  $\frac{1}{n_{i_0}} < b - a$  (due to  $n_i \rightarrow \infty$  as  $i \rightarrow \infty$ ), and hence there exists  $k \in \{1, \dots, n_{i_0} - 1\}$  such that  $\frac{k}{n_{i_0}} \in (a, b)$ .

Now we turn to prove the second statement of the assertion (iv). Let  $\ell, m \in \mathbb{N}$ . We show that  $\frac{\ell}{m}$  is a level of increase for the function (2.3). By assumption, there exists  $i_0 \in \mathbb{N}$  such that  $m + \ell$  divides  $n_{i_0}$  and  $\psi$  is a  $T_{n_{i_0}}$ -function. By part (iii) of this theorem, we have that the elements of the set

$$\left\{ \frac{k}{n_{i_0} - k} \mid k \in \{1, \dots, n_{i_0} - 1\} \right\}$$

are levels of increase for the function (2.3). By choosing  $k := \frac{\ell n_{i_0}}{m+\ell} \in \{1, \dots, n_{i_0} - 1\}$ , we have

$$\frac{k}{n_{i_0} - k} = \frac{\ell n_{i_0}}{(m + \ell)n_{i_0} - \ell n_{i_0}} = \frac{\ell}{m},$$

yielding that  $\frac{\ell}{m}$  is a level of increase for the function (2.3), as desired.

(v): Assume that  $\psi$  is a  $T_2^\lambda$ -function for each  $\lambda \in \Lambda_2$ . Let  $x, y \in X$  be such that  $\vartheta_1(x) < \vartheta_1(y)$ . By assertion (i), it follows that  $\lambda_1/\lambda_2$  is a level of increase for the function (2.3) for each  $\lambda_1, \lambda_2 > 0$ . Since  $\lambda_1$  and  $\lambda_2$  are arbitrary positive numbers, we get that each positive number is a level of increase for the positive function (2.3). In view of the second statement of Lemma 2.8, this implies that the function (2.3) is strictly increasing.

(vi): This assertion is an immediate consequence of Theorem 3.1 as stated in Corollary 3.3, therefore its proof is omitted here.  $\square$

**Proof of Corollary 2.11.** If (i) holds, then part (vi) of Theorem 2.10 implies that (iii) is valid as well. If (iii) holds, then (ii) is readily satisfied. Finally, if (ii) holds, then part (v) of Theorem 2.10 implies the validity of (i).  $\square$

**Proof of Proposition 2.12.** (i): First, let us suppose that for each  $x \in X$ , the function  $\Theta \ni t \mapsto \psi(x, t)$  is decreasing. Let  $\vartheta_1(x) < s < t < \vartheta_1(y)$ . Then, since  $\psi$  is a  $T_1$ -function, we have

$$0 > \psi(x, s) \geq \psi(x, t) \quad \text{and} \quad \psi(y, s) \geq \psi(y, t) > 0.$$

Consequently, we get

$$0 < -\psi(x, s)\psi(y, t) \leq -\psi(x, t)\psi(y, s),$$

which is equivalent to

$$-\frac{\psi(x, s)}{\psi(y, s)} \leq -\frac{\psi(x, t)}{\psi(y, t)},$$

yielding that the function (2.3) is increasing. The case when the function  $\Theta \ni t \mapsto \psi(x, t)$  is strictly decreasing for each  $x \in X$  can be handled similarly.

(ii): Let us suppose that for each  $x \in X$ , the function  $\Theta \ni t \mapsto \psi(x, t)$  is strictly decreasing. Let  $n \in \mathbb{N}$ ,  $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$  and  $(x_1, \dots, x_n) \in X^n$ . Since  $\lambda_i \geq 0$ ,  $i \in \{1, \dots, n\}$  and  $\lambda_1 + \dots + \lambda_n > 0$ , we get that the function  $\Theta \ni t \mapsto \sum_{i=1}^n \lambda_i \psi(x_i, t)$  is strictly decreasing. Using that  $\psi$  is a  $T_1$ -function, we have that

$$\sum_{i=1}^n \lambda_i \psi(x_i, t) \begin{cases} > 0 & \text{if } t < \min(\vartheta_1(x_1), \dots, \vartheta_1(x_n)), \\ < 0 & \text{if } t > \max(\vartheta_1(x_1), \dots, \vartheta_1(x_n)). \end{cases}$$

Consequently, we have

$$t^* := \sup \left\{ t \in \Theta : \sum_{i=1}^n \lambda_i \psi(x_i, t) > 0 \right\} \leq \max(\vartheta_1(x_1), \dots, \vartheta_1(x_n)),$$

$$t_* := \inf \left\{ t \in \Theta : \sum_{i=1}^n \lambda_i \psi(x_i, t) < 0 \right\} \geq \min(\vartheta_1(x_1), \dots, \vartheta_1(x_n)),$$

yielding that  $t^* \leq t_*$ . Using the definition of infimum, supremum and that the map  $\Theta \ni t \mapsto \sum_{i=1}^n \lambda_i \psi(x_i, t)$  is strictly decreasing, we get  $t^* = t_*$ . Indeed, if  $t^* < t_*$  were true, then, by the definition of infimum and supremum,  $\sum_{i=1}^n \lambda_i \psi(x_i, t) = 0$ ,  $t \in (t^*, t_*)$ , would hold, contradicting the strictly decreasing property of the function  $\Theta \ni t \mapsto \sum_{i=1}^n \lambda_i \psi(x_i, t)$ . All in all, we get  $\vartheta_{n,\psi}^\lambda(x_1, \dots, x_n) = t^* = t_*$ , and then  $\psi$  possesses the property  $(T_n^\lambda)$ .  $\square$

**Proof of Proposition 2.13.** Let  $x, y \in X$  be such that  $\vartheta_1(x) < \vartheta_1(y)$ . Since  $\psi$  is a  $T_1$ -function, for all  $t \in (\vartheta_1(x), \vartheta_1(y))$ , we have  $\psi(x, t) < 0$  and  $\psi(y, t) > 0$ , and hence the function (2.4) takes values in  $(0, 1)$ . By part (iii) of Theorem 2.10, we have that the elements of the set  $\{\frac{k}{n-k} \mid k \in \{1, \dots, n-1\}\}$  are levels of increase for the function (2.3). Note that

$$\frac{\psi(x, t)}{\psi(x, t) - \psi(y, t)} = g\left(-\frac{\psi(x, t)}{\psi(y, t)}\right), \quad t \in (\vartheta_1(x), \vartheta_1(y)),$$

where  $g : (0, \infty) \rightarrow \mathbb{R}$ ,  $g(u) := \frac{u}{u+1}$ ,  $u > 0$ . Since  $g$  is strictly increasing and  $(0, \infty)$  contains the range of the function (2.3), by Lemma 2.8, we get that the elements of the set

$$\left\{g\left(\frac{k}{n-k}\right) \mid k \in \{1, \dots, n-1\}\right\} = \left\{\frac{k}{n} \mid k \in \{1, \dots, n-1\}\right\}$$

are levels of increase for the function (2.4). Since the function (2.3) is positive, we readily have that 0 is also a level of increase for the function (2.3). Consequently, using that the function (2.4) takes values in  $(0, 1)$ , the conditions of Lemma 2.9 are satisfied with the choices  $y_k := \frac{k}{n}$ ,  $k \in \{0, 1, \dots, n\}$ , and then we get that the function (2.4) is strictly  $\varepsilon$ -increasing with  $\varepsilon := \max\left\{\frac{k}{n} - \frac{k-1}{n} : k \in \{1, \dots, n\}\right\} = \frac{1}{n}$ , as desired.  $\square$

**Proof of Proposition 2.14.** Now let us suppose that  $m \in \{1, \dots, n\}$  and  $m$  is a divisor of  $n$ . Then there exists  $k \in \mathbb{N}$  such that  $n = km$ . Consequently, for each  $(y_1, \dots, y_m) \in X^m$ , with the notation

$$(x_1, \dots, x_n) := (\underbrace{y_1, \dots, y_1}_k, \underbrace{y_2, \dots, y_2}_k, \dots, \underbrace{y_m, \dots, y_m}_k) \in X^n,$$

using that  $\psi$  is a  $T_n$ -function, we have that

$$k \sum_{i=1}^m \psi(y_i, t) = \sum_{i=1}^n \psi(x_i, t) \begin{cases} > 0 & \text{if } t < \vartheta_n(x_1, \dots, x_n), \\ < 0 & \text{if } t > \vartheta_n(x_1, \dots, x_n). \end{cases}$$

Hence  $\psi$  is a  $T_m$ -function with  $\vartheta_m(y_1, \dots, y_m) := \vartheta_n(x_1, \dots, x_n)$ .  $\square$

**Proof of Proposition 2.15.** For each  $\mathbf{y} := (y_1, \dots, y_m) \in X^m$  and  $t \in \Theta$ , we have

$$(5.1) \quad \psi_{\mathbf{y}, \boldsymbol{\mu}}(t) := \sum_{\alpha=1}^m \mu_\alpha \psi(y_\alpha, t) = \sum_{\alpha=1}^m \left( \sum_{i \in H_\alpha} \lambda_i \right) \psi(y_\alpha, t) = \sum_{j=1}^n \lambda_j \psi(x_j, t) = \psi_{\mathbf{x}, \boldsymbol{\lambda}}(t),$$

where  $\mathbf{x} := (x_1, \dots, x_n) \in X^n$  is such that  $x_j := y_\alpha$  if  $j \in H_\alpha$ . By the assumption, the value  $\vartheta_n^\lambda(x_1, \dots, x_n)$  is a point of sign change for the function  $\psi_{\mathbf{x}, \boldsymbol{\lambda}}$ . Therefore, by (5.1), we can see that the function  $\psi_{\mathbf{y}, \boldsymbol{\mu}}$  has the same point of sign change, and hence we have

$$\vartheta_m^\mu(y_1, \dots, y_m) = \vartheta_n^\lambda(x_1, \dots, x_n),$$



yielding that  $\psi$  is a  $T_m^\mu$ -function.  $\square$

**Proof of Proposition 2.19.** First, we check that  $\psi$  is a  $T_1$ -function with  $\vartheta_1 = f^{(-1)} \circ \varphi$ . Let  $x \in X$  be fixed. If  $t < (f^{(-1)} \circ \varphi)(x)$ ,  $t \in \Theta$ , then  $\psi(x, t) > 0$ , since otherwise  $\psi(x, t) \leq 0$  would yield that  $\varphi(x) \leq f(t)$ , and hence, by Lemma 2.18, we would have that  $(f^{(-1)} \circ \varphi)(x) \leq (f^{(-1)} \circ f)(t) = t$ , leading us to a contradiction. Similarly, if  $t > (f^{(-1)} \circ \varphi)(x)$ ,  $t \in \Theta$ , then  $\psi(x, t) < 0$ , since otherwise  $\psi(x, t) \geq 0$  would yield that  $\varphi(x) \geq f(t)$ , and hence, by Lemma 2.18, we would have that  $(f^{(-1)} \circ \varphi)(x) \geq (f^{(-1)} \circ f)(t) = t$ , leading us to a contradiction. All in all, for each  $x \in X$ , we have that

$$\psi(x, t) \begin{cases} > 0 & \text{if } t < (f^{(-1)} \circ \varphi)(x), t \in \Theta, \\ < 0 & \text{if } t > (f^{(-1)} \circ \varphi)(x), t \in \Theta, \end{cases}$$

as desired.

Consequently, using also that for each  $x \in X$ , the function  $\Theta \ni t \mapsto \psi(x, t) = p(x)(\varphi(x) - f(t))$  is strictly decreasing, part (ii) of Proposition 2.12 implies that the function  $\psi$  has the property  $(T_n^\lambda)$  for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$ .

It remains to check that (2.6) holds. First, note that the right hand side of (2.6) is well-defined, since

$$\begin{aligned} & \frac{\lambda_1 p(x_1) \varphi(x_1) + \cdots + \lambda_n p(x_n) \varphi(x_n)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)} \\ &= \frac{\lambda_1 p(x_1)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)} \varphi(x_1) + \cdots + \frac{\lambda_n p(x_n)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)} \varphi(x_n) \\ & \in \text{conv}(\varphi(X)) \subseteq \text{conv}(f(\Theta)), \end{aligned}$$

and  $f^{(-1)}$  is defined on  $\text{conv}(f(\Theta))$  (see Lemma 2.18). Let  $n \in \mathbb{N}$ ,  $(x_1, \dots, x_n) \in X^n$  and  $\lambda \in \Lambda_n$  be fixed. If

$$t < f^{(-1)} \left( \frac{\lambda_1 p(x_1) \varphi(x_1) + \cdots + \lambda_n p(x_n) \varphi(x_n)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)} \right), \quad t \in \Theta,$$

then  $\sum_{i=1}^n \lambda_i \psi(x_i, t) > 0$ , since otherwise  $\sum_{i=1}^n \lambda_i \psi(x_i, t) \leq 0$  would yield that

$$\frac{\lambda_1 p(x_1) \varphi(x_1) + \cdots + \lambda_n p(x_n) \varphi(x_n)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)} \leq f(t).$$

Hence, by Lemma 2.18, we would have that

$$f^{(-1)} \left( \frac{\lambda_1 p(x_1) \varphi(x_1) + \cdots + \lambda_n p(x_n) \varphi(x_n)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)} \right) \leq (f^{(-1)} \circ f)(t) = t,$$

leading us to a contradiction. Similarly, we can easily see that the inequality

$$t > f^{(-1)} \left( \frac{\lambda_1 p(x_1) \varphi(x_1) + \cdots + \lambda_n p(x_n) \varphi(x_n)}{\lambda_1 p(x_1) + \cdots + \lambda_n p(x_n)} \right), \quad t \in \Theta,$$

implies  $\sum_{i=1}^n \lambda_i \psi(x_i, t) < 0$ . These two properties together with the fact that  $\psi$  has the property  $(T_n^\lambda)$  yield the equality (2.6).  $\square$

## 6 Proofs for Section 3

**Proof of Theorem 3.1.** Define the sets  $U, V \subseteq \Theta$  by

$$U := \{s \in \Theta : \mathbb{E}(\psi(\xi, s)) \geq 0\} \quad \text{and} \quad V := \{t \in \Theta : \mathbb{E}(\psi(\xi, t)) \leq 0\}.$$

Then, in view of assumption (v), we have that  $s_0 \in U$  and  $t_0 \in V$ . In what follows, we show that  $s \leq t$  holds for all  $s \in U$  and  $t \in V$ . To the contrary, assume that  $t < s$ , and, for any Borel subset  $H \subseteq \Theta$ , let us define

$$\Omega_H := \{\omega \in \Omega : \vartheta_1(\xi(\omega)) \in H\}.$$

Then  $\Omega_H \in \mathcal{A}$  due to the measurability of  $\vartheta_1 : X \rightarrow \Theta$  and  $\xi : \Omega \rightarrow X$ . Indeed, for each  $r \in \Theta$  we have that

$$\vartheta_1^{-1}((-\infty, r)) = \{x \in X : \vartheta_1(x) < r\} = \{x \in X : \psi(x, r) < 0\} \in \mathcal{X},$$

where we used assumptions (i), (iii) and that the sigma-algebra generated by the family  $\{(-\infty, r) \cap \Theta, r \in \Theta\}$  coincides with the Borel sigma-algebra on  $\Theta$ .

Consider the following partition of  $\Theta$ , which is induced by  $t$  and  $s$ :

$$I := \Theta \cap (-\infty, t), \quad J := [t, s], \quad K := \Theta \cap (s, \infty).$$

Then, using assumption (i), we have

$$\begin{aligned} \Omega_I &= \{\omega \in \Omega : \vartheta_1(\xi(\omega)) < t\} = \{\omega \in \Omega : \psi(\xi(\omega), t) < 0\}, \\ \Omega_K &= \{\omega \in \Omega : \vartheta_1(\xi(\omega)) > s\} = \{\omega \in \Omega : \psi(\xi(\omega), s) > 0\}. \end{aligned}$$

We show that  $\mathbb{P}(\Omega_I) > 0$  and  $\mathbb{P}(\Omega_K) > 0$ . Indeed, on the contrary, if  $\mathbb{P}(\Omega_I) = 0$ , then  $\mathbb{P}(\psi(\xi, t) \geq 0) = 1$ , which implies that  $\mathbb{E}(\psi(\xi, t)) \geq 0$ . By the inclusion  $t \in V$ , we also have that  $\mathbb{E}(\psi(\xi, t)) \leq 0$  and hence  $\mathbb{E}(\psi(\xi, t)) = 0$ . Therefore,  $\mathbb{P}(\psi(\xi, t) = 0) = 1$ , i.e.,  $\mathbb{P}(\vartheta_1(\xi) = t) = 1$ . It follows from the inequality  $t < s$  that  $\mathbb{P}(\vartheta_1(\xi) < s) = 1$ , and hence  $\mathbb{P}(\psi(\xi, s) < 0) = 1$ . This implies that  $\mathbb{E}(\psi(\xi, s)) < 0$ , which contradicts that  $s$  belongs to  $U$ . The equality  $\mathbb{P}(\Omega_K) = 0$  leads to a contradiction similarly.

The inequalities  $\mathbb{P}(\Omega_I) > 0$  and  $\mathbb{P}(\Omega_K) > 0$  imply that  $\Omega_I \neq \emptyset$  and  $\Omega_K \neq \emptyset$ . Then, for all  $\omega' \in \Omega_I$  and  $\omega'' \in \Omega_K$ , we have that

$$\vartheta_1(\xi(\omega')) < t < s < \vartheta_1(\xi(\omega'')).$$

Therefore, using assumption (ii) with  $x := \xi(\omega')$  and  $y := \xi(\omega'')$ , the function (2.3) is strictly increasing, hence we get

$$(6.1) \quad \frac{\psi(\xi(\omega'), s)}{\psi(\xi(\omega''), s)} < \frac{\psi(\xi(\omega'), t)}{\psi(\xi(\omega''), t)}.$$

Using that  $\psi(\xi(\omega''), s) > 0$  and  $\psi(\xi(\omega''), t) > 0$ , we can obtain that

$$(6.2) \quad \psi(\xi(\omega'), s)\psi(\xi(\omega''), t) < \psi(\xi(\omega'), t)\psi(\xi(\omega''), s), \quad (\omega', \omega'') \in \Omega_I \times \Omega_K.$$

Integrating on  $\Omega_I$  and then on  $\Omega_K$  with respect to  $\mathbb{P}$ , it follows that

$$\int_{\Omega_I} \psi(\xi(\omega'), s) d\mathbb{P}(\omega') \cdot \int_{\Omega_K} \psi(\xi(\omega''), t) d\mathbb{P}(\omega'') < \int_{\Omega_I} \psi(\xi(\omega'), t) d\mathbb{P}(\omega') \cdot \int_{\Omega_K} \psi(\xi(\omega''), s) d\mathbb{P}(\omega''),$$

that is,

$$(6.3) \quad \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_I}) \cdot \mathbb{E}(\psi(\xi, t)\mathbf{1}_{\Omega_K}) < \mathbb{E}(\psi(\xi, t)\mathbf{1}_{\Omega_I}) \cdot \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_K}).$$

The inequality in (6.3) is indeed strict because the left hand side of (6.2) is strictly smaller than its right hand side over the set  $\Omega_I \times \Omega_K$  which has positive measure with respect to the product probability  $\mathbb{P} \otimes \mathbb{P}$ .

Furthermore, using also that  $t < s$ , for each  $\omega' \in \Omega_J$  and  $\omega'' \in \Omega_K$ , we have that  $\psi(\xi(\omega'), s) \leq 0$ ,  $\psi(\xi(\omega''), t) > 0$ ,  $\psi(\xi(\omega'), t) \geq 0$ , and  $\psi(\xi(\omega''), s) > 0$ . Therefore,

$$\psi(\xi(\omega'), s)\psi(\xi(\omega''), t) \leq 0 \leq \psi(\xi(\omega'), t)\psi(\xi(\omega''), s), \quad (\omega', \omega'') \in \Omega_J \times \Omega_K.$$

Integrating on  $\Omega_J$  and then on  $\Omega_K$  with respect to  $\mathbb{P}$ , it follows that

$$(6.4) \quad \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_J}) \cdot \mathbb{E}(\psi(\xi, t)\mathbf{1}_{\Omega_K}) \leq \mathbb{E}(\psi(\xi, t)\mathbf{1}_{\Omega_J}) \cdot \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_K}).$$

Adding up the inequalities (6.3) and (6.4), and using that  $\Omega_I$  and  $\Omega_J$  are disjoint, we get

$$A(s)B(t) := \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_I \cup \Omega_J}) \cdot \mathbb{E}(\psi(\xi, t)\mathbf{1}_{\Omega_K}) < \mathbb{E}(\psi(\xi, t)\mathbf{1}_{\Omega_I \cup \Omega_J}) \cdot \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_K}) =: A(t)B(s).$$

Further, we have  $A(t) + B(t) = \mathbb{E}(\psi(\xi, t)) \leq 0$  and  $A(s) + B(s) = \mathbb{E}(\psi(\xi, s)) \geq 0$ , since  $t \in V$ ,  $s \in U$ , and  $\Omega_I \cup \Omega_J$  and  $\Omega_K$  are disjoint.

To summarize,  $t, s \in \Theta$  are such that  $t < s$  and the following inequalities hold

$$(6.5) \quad A(s)B(t) < B(s)A(t), \quad A(t) + B(t) \leq 0, \quad A(s) + B(s) \geq 0.$$

Here  $B(t) > 0$  because it equals the integral of a positive function over the set  $\Omega_K$  which has positive measure with respect to the probability  $\mathbb{P}$ . On the other hand,  $A(s) < 0$ , because

$$A(s) = \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_I \cup \Omega_J}) = \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_I}) + \mathbb{E}(\psi(\xi, s)\mathbf{1}_{\Omega_J})$$

and the first term is negative being equal to the integral of a negative function over the set  $\Omega_I$  (which has positive measure with respect to  $\mathbb{P}$ ) and the second term is nonpositive being equal to the integral of a nonpositive function over the set  $\Omega_J$ .

Consequently, by the last two inequalities of (6.5), we get

$$0 < B(t) \leq -A(t) \quad \text{and} \quad 0 < -A(s) \leq B(s),$$

yielding that

$$0 < -A(s)B(t) \leq -A(t)B(s),$$

i.e.,  $A(s)B(t) \geq B(s)A(t)$ . This contradicts to the first inequality in (6.5).

Consequently, we have that  $s_0 \leq u_0 := \sup U \leq \inf V =: v_0 \leq t_0$ . It remains to show that  $u_0 = v_0$ . If, to the contrary, we assume that  $u_0 < v_0$ , then for each  $r \in (u_0, v_0)$ , we get  $r \notin U$  and  $r \notin V$ , yielding that  $\mathbb{E}(\psi(\xi, r)) < 0$  and  $\mathbb{E}(\psi(\xi, r)) > 0$ , respectively, which is a contradiction.

All in all,  $u_0 = v_0$  is a unique point of sign change for the function  $\Theta \ni t \mapsto \mathbb{E}(\psi(\xi, t))$ , as desired.  $\square$

**Proof of Proposition 3.2.** By the assumption (i), for each  $\omega \in \Omega$ , we have that the function  $\Theta \ni t \mapsto \psi(\xi(\omega), t)$  is strictly decreasing. By the monotonicity of the expectation, it implies that the function  $\Theta \ni t \mapsto \mathbb{E}(\psi(\xi, t))$  is decreasing, and in fact, it is strictly decreasing. Indeed, if  $t_1 < t_2$ ,  $t_1, t_2 \in \Theta$ , are such that  $\mathbb{E}(\psi(\xi, t_1)) = \mathbb{E}(\psi(\xi, t_2))$ , then  $\mathbb{E}(\psi(\xi, t_1) - \psi(\xi, t_2)) = 0$ , where  $\mathbb{P}(\psi(\xi, t_1) - \psi(\xi, t_2) \geq 0) = 1$ . Consequently,  $\mathbb{P}(\psi(\xi, t_1) - \psi(\xi, t_2) = 0) = 1$ , leading us to a contradiction, since  $\psi(\xi(\omega), t_1) > \psi(\xi(\omega), t_2)$ ,  $\omega \in \Omega$ .

Define the sets  $U, V \subseteq \Theta$  by

$$U := \{s \in \Theta : \mathbb{E}(\psi(\xi, s)) \geq 0\} \quad \text{and} \quad V := \{t \in \Theta : \mathbb{E}(\psi(\xi, t)) \leq 0\}.$$

By the assumption (iv), we have that  $s_0 \in U$  and  $t_0 \in V$ . Since the function  $\Theta \ni t \mapsto \mathbb{E}(\psi(\xi, t))$  is strictly decreasing, we can easily see that  $s \leq t$  for all  $s \in U, t \in V$ . Indeed, if for some  $s \in U$  and  $t \in V$ , the inequality  $s > t$  were true, then we would have that  $0 \leq \mathbb{E}(\psi(\xi, s)) < \mathbb{E}(\psi(\xi, t)) \leq 0$ , leading us to a contradiction. Hence we have  $s_0 \leq u_0 := \sup U \leq \inf V =: v_0 \leq t_0$ . It remains to show that  $u_0 = v_0$ . If, to the contrary, we assume that  $u_0 < v_0$ , then for each  $r \in (u_0, v_0)$ , we get  $r \notin U$  and  $r \notin V$ , yielding that  $\mathbb{E}(\psi(\xi, r)) < 0$  and  $\mathbb{E}(\psi(\xi, r)) > 0$ , respectively, which is a contradiction. Consequently,  $u_0 = v_0$  is a unique point of sign change for the function  $\Theta \ni t \mapsto \mathbb{E}(\psi(\xi, t))$ , as desired.  $\square$

**Proof of Corollary 3.3.** To verify the statement, we have to show that, for each  $n \in \mathbb{N}$ ,  $x_1, \dots, x_n \in X$  and  $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$ , the function  $\Theta \ni t \mapsto \sum_{i=1}^n \lambda_i \psi(x_i, t)$  has a unique point of sign change in  $\Theta$ . Without loss of generality, we may assume that  $x_1, \dots, x_n$  are pairwise distinct elements of  $X$  and  $\lambda_1, \dots, \lambda_n > 0$  with  $\lambda_1 + \dots + \lambda_n = 1$ .

Define the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  by

$$\Omega := \{x_1, \dots, x_n\}, \quad \mathcal{A} := 2^\Omega, \quad \mathbb{P}(\{x_i\}) := \lambda_i \quad i \in \{1, \dots, n\},$$

and the random variable  $\xi : \Omega \rightarrow \Omega$  by  $\xi(\omega) := \omega$ ,  $\omega \in \Omega$ .

Then the conditions (i) and (ii) of Theorem 3.1 follow from our assumptions. The measurability condition (iii) of Theorem 3.1 is trivial due to the fact that  $\mathcal{A} = 2^\Omega$ . Since

$$\mathbb{E}(|\psi(\xi, t)|) = \sum_{i=1}^n \lambda_i |\psi(x_i, t)|, \quad t \in \Theta,$$

the condition (iv) of Theorem 3.1 is obviously valid. Finally, the condition (v) of Theorem 3.1 is satisfied by

$$s_0 := \min\{\vartheta_1(x_1), \dots, \vartheta_1(x_n)\} \quad \text{and} \quad t_0 := \max\{\vartheta_1(x_1), \dots, \vartheta_1(x_n)\}.$$

Indeed, for each  $i \in \{1, \dots, n\}$ , we have that  $\psi(x_i, s_0) \geq 0 \geq \psi(x_i, t_0)$ , since  $\psi$  is a  $T_1$ -function and  $\psi(x, \vartheta_1(x)) = 0$ ,  $x \in X$ . This implies that

$$\mathbb{E}(\psi(\xi, s_0)) = \sum_{i=1}^n \lambda_i \psi(x_i, s_0) \geq 0 \geq \sum_{i=1}^n \lambda_i \psi(x_i, t_0) = \mathbb{E}(\psi(\xi, t_0)).$$

Therefore, according to the conclusion of Theorem 3.1, the mapping  $\Theta \ni t \mapsto \mathbb{E}(\psi(\xi, t)) = \sum_{i=1}^n \lambda_i \psi(x_i, t)$  has a unique point of sign change in  $\Theta$ , as desired.  $\square$

**Proof of Lemma 3.4.** Let us define the probability measure  $\mathbb{Q}$  on the measurable space  $(\Omega, \mathcal{A})$  by

$$\mathbb{Q}(A) := \int_A \frac{p(\xi)}{\mathbb{E}(p(\xi))} \, d\mathbb{P} = \frac{\mathbb{E}(p(\xi)\mathbf{1}_A)}{\mathbb{E}(p(\xi))}, \quad A \in \mathcal{A}.$$

By denoting the expectation with respect to  $\mathbb{Q}$  by  $\mathbb{E}_{\mathbb{Q}}$ , we have

$$\mathbb{E}_{\mathbb{Q}}(|\varphi(\xi)|) = \frac{\mathbb{E}(p(\xi)|\varphi(\xi)|)}{\mathbb{E}(p(\xi))},$$

and hence, by the assumptions,  $\varphi(\xi)$  is integrable with respect to  $\mathbb{Q}$ , and we also get

$$\mathbb{E}_{\mathbb{Q}}(\varphi(\xi)) = \frac{\mathbb{E}(p(\xi)\varphi(\xi))}{\mathbb{E}(p(\xi))}.$$

Applying Lemma 1 in Janković and Merkle [10] (for integrable one-dimensional random variables), we have that  $\mathbb{E}_{\mathbb{Q}}(\varphi(\xi)) \in \text{conv}(\varphi(\xi(\Omega))) \subseteq \text{conv}(\varphi(X))$ , yielding the statement.  $\square$

**Proof of Proposition 3.4.** We apply Proposition 3.2. The assumptions (i), (ii) and (iii) of Proposition 3.2 readily hold.

To verify the assumption (iv) of Proposition 3.2, we first show that, for any  $y \in J := \text{conv}(f(\Theta))$ , there exist  $s_0, t_0 \in \Theta$  such that  $f(s_0) \leq y \leq f(t_0)$ . By the Carathéodory's Theorem on convex hulls, there exist at most two elements  $y_1, y_2 \in f(\Theta) \subseteq J$  with  $y_1 \leq y_2$  such that  $y$  can be represented as a convex combination of  $y_1$  and  $y_2$ . This also yields that  $y_1 \leq y \leq y_2$ , and therefore, there exist  $s_0, t_0 \in \Theta$  such that  $f(s_0) \leq y \leq f(t_0)$ . Now observe that

$$f^{(-1)}\left(\frac{\mathbb{E}(p(\xi)\varphi(\xi))}{\mathbb{E}(p(\xi))}\right)$$

is well-defined, since, by Lemma 3.4, we get that

$$\frac{\mathbb{E}(p(\xi)\varphi(\xi))}{\mathbb{E}(p(\xi))} \in \text{conv}(\varphi(X)) \subseteq \text{conv}(f(\Theta)),$$

and, by Lemma 2.18,  $f^{(-1)}$  is defined on  $\text{conv}(f(\Theta))$ . Next, for  $y := \mathbb{E}(p(\xi)\varphi(\xi))/\mathbb{E}(p(\xi))$ , let us choose  $s_0, t_0 \in \Theta$  as it was described above. Then

$$\mathbb{E}(\psi(\xi, s_0)) = \mathbb{E}(p(\xi)\varphi(\xi)) - f(s_0)\mathbb{E}(p(\xi)) = \mathbb{E}(p(\xi))(y - f(s_0)) \geq 0,$$

and

$$\mathbb{E}(\psi(\xi, t_0)) = \mathbb{E}(p(\xi)\varphi(\xi)) - f(t_0)\mathbb{E}(p(\xi)) = \mathbb{E}(p(\xi))(y - f(t_0)) \leq 0.$$

Therefore, the assumption (iv) of Proposition 3.2 holds as well, and, according to the conclusion of Proposition 3.2, we get that the function  $\Theta \ni t \rightarrow \mathbb{E}(\psi(\xi, t))$  admits a unique point of sign change in  $\Theta$ . It remains to check that this unique point of sign change takes the form given in the proposition.

If, for some  $t \in \Theta$ , we have  $t < f^{(-1)}(y)$ , then  $\mathbb{E}(\psi(\xi, t)) > 0$ , since otherwise  $\mathbb{E}(\psi(\xi, t)) \leq 0$  would yield that  $\mathbb{E}(p(\xi)\varphi(\xi)) \leq f(t)\mathbb{E}(p(\xi))$ , i.e.,  $y \leq f(t)$ . Then, by Lemma 2.18, we would get  $f^{(-1)}(y) \leq f^{(-1)}(f(t)) = t$  leading us to a contradiction.

If for some  $t \in \Theta$ , we have  $t > f^{(-1)}(y)$ , then one can similarly argue to obtain that  $\mathbb{E}(\psi(\xi, t)) < 0$ .

Consequently, the unique point of sign change in question is  $f^{(-1)}(y)$ , as desired.  $\square$

**Proof of Proposition 3.5.** First, we give a direct proof. Denote the limit  $\lim_{z \rightarrow \infty} f(z)$  by  $f_\infty \in (0, \infty)$ . In view of the increasingness of  $f$ , it follows that  $0 \leq f(z) \leq f_\infty$  for all  $z \in \mathbb{R}_+$ . Therefore,  $|\tilde{f}(z)| \leq f_\infty$  for all  $z \in \mathbb{R}$ , which implies that  $|\psi(x, t)| \leq f_\infty$  for all  $x, t \in \mathbb{R}$ . Hence, for any random variable  $\xi$  and for any  $t \in \mathbb{R}$ , we have that  $\mathbb{E}(|\psi(\xi, t)|) < \infty$ .

Since for each  $x \in \mathbb{R}$ , the function  $\mathbb{R} \ni t \mapsto \psi(x, t)$  is strictly decreasing, we have the function  $\mathbb{R} \ni t \mapsto \mathbb{E}(\psi(\xi, t))$  is strictly decreasing. Indeed, if  $s < t$ ,  $s, t \in \mathbb{R}$ , then we have  $\psi(\xi(\omega), s) > \psi(\xi(\omega), t)$ ,  $\omega \in \Omega$ , yielding that  $\mathbb{E}(\psi(\xi, s)) \geq \mathbb{E}(\psi(\xi, t))$ . Here the equality cannot hold, since otherwise  $\mathbb{E}(\psi(\xi, s) - \psi(\xi, t)) = 0$  would be valid yielding that  $\mathbb{P}(\psi(\xi, s) - \psi(\xi, t) = 0) = 1$ . This leads us to a contradiction, since  $\psi(\xi(\omega), s) - \psi(\xi(\omega), t) > 0$ ,  $\omega \in \Omega$ . Using that  $\lim_{t \rightarrow \pm\infty} \psi(\xi(\omega), t) = \mp f_\infty$ ,  $\omega \in \Omega$ , and  $|\psi(\xi(\omega), t)| \leq f_\infty$ ,  $\omega \in \Omega$ ,  $t \in \mathbb{R}$ , the dominated convergence theorem implies that

$$(6.6) \quad \lim_{t \rightarrow \infty} \mathbb{E}(\psi(\xi, t)) = \mathbb{E}\left(\lim_{t \rightarrow \infty} \psi(\xi, t)\right) = \mathbb{E}(-f_\infty) = -f_\infty < 0,$$

and

$$(6.7) \quad \lim_{t \rightarrow -\infty} \mathbb{E}(\psi(\xi, t)) = \mathbb{E}\left(\lim_{t \rightarrow -\infty} \psi(\xi, t)\right) = \mathbb{E}(f_\infty) = f_\infty > 0.$$

Since  $f$  is continuous and  $f(0) = 0$ , we have that  $\psi$  is continuous in its second variable. Thus, by the dominated convergence theorem, it follows that the function  $\mathbb{R} \ni t \mapsto \mathbb{E}(\psi(\xi, t))$  is also continuous. All in all, the function  $\mathbb{R} \ni t \mapsto \mathbb{E}(\psi(\xi, t))$  is strictly decreasing, continuous, and changes sign, and hence there exists a unique  $t_0 \in \mathbb{R}$  such that  $\mathbb{E}(\psi(\xi, t_0)) = 0$ , as desired.

Finally, we present an alternative proof of Proposition 3.5 using Theorem 3.1. We check that the assumptions of Theorem 3.1 hold. Since  $f(0) = 0$  and  $f$  is strictly increasing, the

assumption (i) of Theorem 3.1 holds with  $\vartheta_1(x) = x$ ,  $x \in \mathbb{R}$ . Using that  $f$  is strictly increasing, by part (d) of Proposition 4.4, we have that  $\psi$  is a  $T_2^\lambda$ -function for all  $\lambda \in \Lambda_2$ . Consequently, part (v) of Theorem 2.10 yields that, for each  $x, y \in \mathbb{R}$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is strictly increasing, i.e., the assumption (ii) of Theorem 3.1 holds. The assumption (iii) of Theorem 3.1 readily holds. The first part of the direct proof of the present proposition implies that the assumption (iv) of Theorem 3.1 holds. Using (6.6) and (6.7) we have that the assumption (v) of Theorem 3.1 holds as well. All in all, we can apply Theorem 3.1, and it yields that the function  $\mathbb{R} \ni t \mapsto \mathbb{E}(\psi(\xi, t)) = \mathbb{E}(\text{sign}(\xi - t)f(|\xi - t|))$  has a (unique) point of sign change. Since the function  $\mathbb{R} \ni t \mapsto \mathbb{E}(\psi(\xi, t))$  is strictly decreasing and continuous (see the direct proof), we have that the equation  $\mathbb{E}(\psi(\xi, t)) = 0$  has a unique solution with respect to  $t \in \mathbb{R}$ , as desired.  $\square$

## 7 Proofs for Section 4

**Proof of Proposition 4.2.** Assume that  $n \geq 2$  and  $\psi$  has the property  $(T_n)$ . Then, according to part (iii) of Theorem 2.10, for each  $x, y \in \mathbb{R}$  with  $x < y$ , the number  $\frac{n-k}{k}$  must be a level of increase for the function (2.3) if  $k \in \{1, \dots, n-1\}$ . Note that for each  $x, y \in \mathbb{R}$  with  $x < y$ , the function (2.3) takes the form

$$(x, y) \ni t \mapsto -\frac{\psi(x, t)}{\psi(y, t)} = \frac{1 - \alpha}{\alpha} > 0.$$

Therefore, we have that

$$\frac{1 - \alpha}{\alpha} \neq \frac{n - k}{k} \quad \text{for each } k \in \{1, \dots, n - 1\},$$

which implies that  $\alpha \neq \frac{k}{n}$  for each  $k \in \{1, \dots, n - 1\}$ .

Conversely, if  $\alpha \notin \{\frac{1}{n}, \dots, \frac{n-1}{n}\}$ , where  $n \geq 2$ , then we have that

$$\frac{k}{n} < \alpha < \frac{k + 1}{n}$$

for some  $k \in \{0, \dots, n - 1\}$ , and hence  $k < n\alpha < k + 1$ . Let  $x_1, \dots, x_n \in \mathbb{R}$  be arbitrary. If  $t \in \mathbb{R}$  with  $t < x_{k+1}^*$ , then we have that  $\psi(x_i^*, t) \geq \alpha - 1$ ,  $i = 1, \dots, k$ , and  $\psi(x_i^*, t) = \alpha$ ,  $i = k + 1, \dots, n$ , yielding that

$$\sum_{i=1}^n \psi(x_i, t) = \sum_{i=1}^k \psi(x_i^*, t) + \sum_{i=k+1}^n \psi(x_i^*, t) \geq k(\alpha - 1) + (n - k)\alpha = n\alpha - k > 0.$$

If  $t \in \mathbb{R}$  with  $t > x_{k+1}^*$ , then we have that  $\psi(x_i^*, t) = \alpha - 1$ ,  $i = 1, \dots, k + 1$ , and  $\psi(x_i^*, t) \leq \alpha$ ,  $i = k + 2, \dots, n$ , yielding that

$$\sum_{i=1}^n \psi(x_i, t) = \sum_{i=1}^{k+1} \psi(x_i^*, t) + \sum_{i=k+2}^n \psi(x_i^*, t) \leq (k + 1)(\alpha - 1) + (n - k - 1)\alpha = n\alpha - k - 1 < 0.$$

Therefore  $\vartheta_n(x_1, \dots, x_n)$  exists and equals  $x_{k+1}^*$ . This proves that  $\psi$  is indeed a  $T_n$ -function.  $\square$

**Proof of Proposition 4.4.** Part (a): It follows from the facts that for each  $x \in \mathbb{R}$ , we have  $\psi(x, x) = 0$ ;  $\psi(x, t) > 0$  holds for each  $t < x$  if and only if  $f(z) > 0$  for each  $z > 0$ ; and  $\psi(x, t) < 0$  holds for each  $t > x$  if and only if  $f(z) > 0$  for each  $z > 0$ .

Part (b): Let us suppose that  $\psi$  is a  $T_1$ -function and  $T_n$ -function for infinitely many  $n \in \mathbb{N}$ . By part (iv) of Theorem 2.10, for each  $x, y \in \mathbb{R}$  with  $\vartheta_1(x) < \vartheta_1(y)$ , the function (2.3) is increasing. Since  $\psi$  is a  $T_1$ -function, by part (a) of the present proposition, we have  $f(z) > 0$  for each  $z > 0$  and  $\vartheta_1(x) = x$ . Consequently, for each  $x < y$ ,  $x, y \in \mathbb{R}$ , the function (given by (2.3))

$$(7.1) \quad (x, y) \ni t \mapsto -\frac{\psi(x, t)}{\psi(y, t)}$$

is increasing. Hence the statement of part (b) follows by the following observation (that we check below): provided that  $f(z) > 0$  for each  $z > 0$ , the function (7.1) is (strictly) increasing for each  $x < y$ ,  $x, y \in \mathbb{R}$  if and only if  $f$  is (strictly) increasing. Since the function  $\psi$  given in (4.8) depends only on  $x - t$ , it is enough to check that

$$(7.2) \quad \text{the function } (0, z) \ni t \mapsto \frac{\psi(0, t)}{\psi(z, t)} = -\frac{f(t)}{f(z-t)} \text{ is (strictly) decreasing for each } z > 0$$

holds if and only if  $f$  is (strictly) increasing. Indeed, for each  $x < y$ ,  $x, y \in \mathbb{R}$ , and  $t \in (x, y)$ , we have

$$\frac{\psi(x, t)}{\psi(y, t)} = -\frac{f(t-x)}{f(y-t)} = -\frac{f(t-x)}{f(y-x-(t-x))} = \frac{\psi(0, t-x)}{\psi(y-x, t-x)}, \quad t-x \in (0, y-x).$$

Thus, the property (7.2) holds if and only if

$$\frac{f(s)}{f(z-s)} (<) \leq \frac{f(t)}{f(z-t)} \quad \text{for each } s, t, z \in \mathbb{R} \text{ with } 0 < s < t < z,$$

which is equivalent to

$$(7.3) \quad f(s)f(z-t) (<) \leq f(t)f(z-s) \quad \text{for each } s, t, z \in \mathbb{R} \text{ with } 0 < s < t < z.$$

Using the nonnegativity of  $f$ , it yields that (7.2) holds if and only if

$$(7.4) \quad f(s) (<) \leq f(t) \quad \text{for each } s, t \in \mathbb{R} \text{ with } 0 < s < t,$$

i.e.,  $f$  is (strictly) increasing on  $\mathbb{R}_+$ . Indeed, if (7.2) holds, then (7.3) holds as well, and, by choosing  $z = s + t$ , we get that  $f(s)^2 (<) \leq f(t)^2$ , which implies (7.4), since  $f$  is nonnegative. If (7.4) holds, then for each  $s, t, z \in \mathbb{R}$  with  $0 < s < t < z$ , we have  $0 \leq f(s) (<) \leq f(t)$  and  $0 \leq f(z-t) (<) \leq f(z-s)$ , implying (7.3), and hence (7.2) as well.

Parts (c) and (d): Let us suppose that  $\psi$  is a  $T_2^\lambda$ -function for all  $\lambda \in \Lambda_2$ . In particular,  $\psi$  is a  $T_1$ -function, and hence, by part (a) of the present proposition, we have  $f(z) > 0$  for each  $z > 0$



and  $\vartheta_1(x) = x$ ,  $x \in \mathbb{R}$ . Consequently, by part (v) of Theorem 2.10, for each  $x < y$ ,  $x, y \in \mathbb{R}$ , the function (7.1) is strictly increasing. By the proof of part (b) of the present proposition, it implies that  $f$  is strictly increasing, as desired. Conversely, let us suppose that  $f$  is strictly increasing. Then for each  $x \in \mathbb{R}$ , the function  $\mathbb{R} \ni t \mapsto \psi(x, t)$  is strictly decreasing. Hence Proposition 2.12 implies that  $\psi$  is a  $(T_n^\lambda)$ -function for each  $n \in \mathbb{N}$  and  $\lambda \in \Lambda_n$  (in particular, a  $T_2^\lambda$ -function for each  $\lambda \in \Lambda_n$ ), as desired.  $\square$

## References

- [1] A. Ali and R. J. Tibshirani. The generalized lasso problem and uniqueness. *Electron. J. Stat.*, 13(2):2307–2347, 2019.
- [2] O. Catoni. Challenging the empirical mean and empirical variance: a deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(4):1148–1185, 2012.
- [3] P. Chen, X. Jin, X. Li, and L. Xu. A generalized Catoni’s M-estimator under finite  $\alpha$ -th moment assumption with  $\alpha \in (1, 2)$ . *Electron. J. Stat.*, 15(2):5523–5544, 2021.
- [4] B. R. Clarke. Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.*, 11(4):1196–1205, 1983.
- [5] B. R. Clarke. The selection functional. *Probab. Math. Statist.*, 11(2):149–156 (1991), 1990.
- [6] T. Dimitriadis, T. Fissler, and J. Ziegel. Characterizing M-estimators. arXiv 2208.08108, 2022.
- [7] R. Grünwald and Zs. Páles. On the equality problem of generalized Bajraktarević means. *Aequationes Math.*, 94(4):651–677, 2020.
- [8] P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35:73–101, 1964.
- [9] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)*, Vol. I: Statistics, pages 221–233. Univ. California Press, Berkeley, Calif., 1967.
- [10] S. Janković and M. Merkle. A mean value theorem for systems of integrals. *J. Math. Anal. Appl.*, 342(1):334–339, 2008.
- [11] R. Koenker and Jr. G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.
- [12] M. R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York, 2008.
- [13] K. Lange. *Numerical Analysis for Statisticians*. Statistics and Computing. Springer, New York, second edition, 2010.

- [14] T. Mathieu. Concentration study of M-estimators using the influence function. *Electron. J. Stat.*, 16(1):3695–3750, 2022.
- [15] W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847, 1987.
- [16] Zs. Páles. On approximately convex functions. *Proc. Amer. Math. Soc.*, 131(1):243–252, 2003.
- [17] R. Passeggeri and N. Reid. On quantiles, continuity and robustness. 2022. Available on *Arxiv*: 2206.06998.
- [18] William J. J. Rey. *Introduction to Robust and Quasirobust Statistical Methods*. Universitext. Springer-Verlag, Berlin, 1983.
- [19] R. J. Tibshirani. The lasso problem and uniqueness. *Electron. J. Stat.*, 7:1456–1490, 2013.
- [20] A. W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.