On minimizing sequences for k-centres

Jüri Lember[*]

EURANDOM, PO Box 513-5600 MB Eindhoven, The Netherlands
E-mail: lember@eurandom.tue.nl

1. PRELIMINARIES

1.1.   Introduction

Several basic notions of probability theory can be defined and investigated in terms of approximation theory. As a classical example, the mean of a probability measure P on $\Re$ minimizes the mean squared loss $\int |x - a|^2 P(dx)$ over $\Re$. Hence, in a a sense, the mean is the best one-point approximation of P. The minimum value of mean squared loss is the variance of P - yet another important characteristic of P.

The idea of approximation of P by k points (k is a fixed integer) has led to the notion of "k-centre" (also called k-mean or principal points) of the distribution [3, 5, 8, 10, 11, 23, 24, 30]. Often the square function is replaced by another power function, or by some more complicated discrepancy function ϕ.

In this paper we consider a probability measure P on a separable metric space (E, d) equipped with the Borel σ-algebra. We fix an integer k, a discrepancy function ϕ, and we define the loss function

$$A \mapsto \int \min_{a \in A} \phi(d(x,a)) P(dx) =: W_k(A, P), \qquad (1)$$

where $A \in \mathcal{E}_k := \{A \subset E : |A| \leq k\}$. A k-centre of P is any element of $\mathcal{E}_k$ that minimizes (1). Thus, a k-centre is the best approximation of P by k points at most.

The main objects of the paper are the sequences $\{A_n\}$, A_n in $\mathcal{E}_k$ such that

$$W_k(A_n, P) \to \inf_{A \in \mathcal{E}_k} W_k(A, P) =: W_k(P). \qquad (2)$$

A sequence satisfying (2) is called minimizing (for $W_k(\cdot, P)$). The paper is devoted to the study of convergence of minimizing sequences. We replace

the usual assumption of compactness by the sequential compactness in another suitably chosen topology τ, and then we study conditions that allow strong convergence to be deduced from τ-convergence. For example, if E is a separable dual space the weak* topology serves for τ. The main result of that kind is Theorem 2.1.

The problem of the convergence of minimizing sequence arises from the investigation of the consistency of empirical k-centres. More precisely, let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed E-valued random elements defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and having common distribution P. Let $\{P_n\}$ be the sequence of empirical measure. Thus, P_n is a random measure on $(E, \mathcal{B})$ defined as follows

$$P_n^\omega(B) = \frac{1}{n} \sum_{i=1}^{n} I_B(X_i(\omega)), \quad \omega \in \Omega, \ B \in \mathcal{B}. \qquad (3)$$

Here I_B is the indicator function. Often the empirical measure P_n^ω (for some n and ω) is the only information about P. Hence, the k-centres of P_n^ω – the empirical k-centres – are the natural estimators of the k-centres of P, which in the present context are called theoretical k-centres. We call the empirical k-centres consistent if every sequence of empirical k-centres converges (in the Hausdorff sense) to the set of theoretical k-centres, $\mathbf{P}$-a.s. The consistency of k-centres is a relatively widely investigated topic; for an overview of that problem as well as the precise definitions of the consistency see Subsection 3.1 below. It turns out that $\mathbf{P}$-a.s., every sequence of the empirical k-centres is minimizing for $W_k(\cdot, P)$ (Lemma 3.1). By continuity of $W_k(\cdot, P)$ (Lemma 2.2) the empirical k-centres are consistent if every minimizing sequence is relatively compact. This observation justifies the study of convergence of a more general class of sequences than that of empirical k-centres – the class of minimizing sequences.

The described scheme for proving the consistency of an estimator obtained by minimizing the empirical loss-function is standard – using the classical tools of probability theory the sequence of empirical estimators is showed to be minimizing for (unknown) non-random loss-function. The question of the convergence of minimizing sequences now arises. At this stage the sample-caused randomness is irrelevant, important are the properties of the parameter space and loss function (in our case, $\mathcal{E}_k$ and $W_k(\cdot, P)$). In the present paper the k-centres are considered, but we believe that the introduced methods can be applied for a more general class of estimators.

The paper is organized as follows. In Section 1 we introduce the main notions: loss-function, k-variance, minimizing sequence. Section 2 is devoted to the properties of minimizing sequences. In Subsection 2.1 we show that every minimizing sequence is bounded (Lemma 2.1). In Subsection 2.2 we define the τ-convergence and we prove an important auxiliary lemma

(Lemma 2.2). In subsection 2.3 we investigate the conditions (in terms of P and E) under which τ-convergent minimizing sequences are convergent. Those conditions form the main convergence results for minimizing sequences (Theorem 2.1). Several counterexamples demonstrate the role of the conditions. In Section 3 we use the main result of previous section to prove a consistency theorem for empirical k-centres (Theorem 3.1). The latter generalizes previous results of this type.

We remark that the paper is based on the authors doctoral dissertation [17]. The main results can be found in the dissertation, but the proofs in the paper are better organized, shorter and more readable.

1.2. The Loss Function

Throughout the paper we assume that (E, d) is a separable metric space, $\mathcal{B}$ is its Borel σ-algebra and P is a probability measure on $(E, \mathcal{B})$. Let T denote the support of P.

We consider a discrepancy function $\phi : \Re^+ \mapsto \Re^+$ which is assumed to be continuous, nondecreasing, $\phi(0) = 0$ and such that for an $x_o \in E$, $\int \phi(d(x, x_o))P(dx) < \infty$ implies that for all $c \geq 0$

$$\int \phi(d(x, x_o) + c)P(dx). < \infty \qquad (4)$$

The purpose of (4) is to ensure the finiteness of the loss-function. It holds if ϕ has the Δ_2-property (there exists $u_o \geq 0$ and $\lambda < \infty$ such that $\phi(2u) \leq \lambda\phi(u)$ if $u > u_o$). However (4) can also be true without the Δ_2-property.

Denote by $\mathcal{E}$ the class of all finite subsets of E and define the *loss function*

$$W(\cdot, P) : \mathcal{E} \mapsto \Re^+, \quad W(A, P) = \int \min_{a \in A} \phi(d(x, a))P(dx).$$

Let $d(x, A) := \min_{a \in A} d(x, a)$ be the distance of x from the set A. Since ϕ is monotone, $W(A, P) = \int \phi(d(x, A))P(dx)$.

We now list some general properties of $W(\cdot, P)$. The proofs of them are straightforward and can be found in [17].

P1. *If $W(B, P) < \infty$ for a $B \in \mathcal{E}$, then $W(A, P) < \infty$ for every $A \in \mathcal{E}$.*

Because ϕ is monotone, **P1** follows from (4). It means that $W(\cdot, P)$ is either finite or always infinite on $\mathcal{E}$. Throughout the paper we assume that it is finite.

P2. *If $\sup_{x,y \in E} \phi(d(x, y)) =: \phi(\infty) > 0$, then $W(a, P) < \phi(\infty)$ for some $a \in E$.*

Clearly $\phi(\infty) < \infty$ if and only if $\phi \circ d$ is bounded.

P3. *There exists a sequence of finite sets $\{C_n\}$ such that $W(C_n, P) \to 0$.*

1.3. *k*-variance

From now on let $k \geq 1$ be a fixed integer, $\mathcal{E}_k := \{A \subset E : |A| \leq k\}$. The quantity

$$W_k(P) := \inf_{A \in \mathcal{E}_k} W(A, P)$$

is called the *k-variance of P*. It measures the mean discrepancy of approximating P by k points at most. The name "k-variance" reflects the fact that for $E = \Re$ and ϕ the square function, $W_1(P)$ is the ordinary variance. In most of what follows the function $W(\cdot, P)$ is considered on $\mathcal{E}_k$ only. To emphasize the role of k the restriction of $W(\cdot, P)$ to $\mathcal{E}_k$ is denoted $W_k(\cdot, P)$.

DEFINITION 1.1. Given an $\epsilon \geq 0$, any sequence of sets $\{A_n\}$ in $\mathcal{E}_k$ is said to be *ϵ-minimizing (for $W_k(\cdot, P)$)* if $\limsup_n W_k(A_n, P) \leq W_k(P) + \epsilon$. A 0-minimizing sequence is called *minimizing* and then $W_k(A_n, P) \to W_k(P)$.

Definition 1.1 and **P2** imply that $\phi(\infty) > W_1(P) \geq W_2(P) \geq \cdots$. In this paper we assume that these inequalities are strict, i.e.

$$W_k(P) < W_{k-1}(P) < \cdots < W_1(P) < \phi(\infty). \tag{5}$$

The following proposition gives a sufficient condition for (5). It generalizes the previous results of the same type [26, 27].

PROPOSITION 1.1. *If $W_{k-1}(P) > 0$, then (5) holds.*

Proof. It suffices to show that $W_k(P) = W_{k-1}(P)$ holds if and only if $W_{k-1}(P) = 0$. Let $W_k(P) = W_{k-1}(P) =: w$, and let $\{B_n\}$, in $\mathcal{E}_{k-1}$, be a minimizing sequence for $W_{k-1}(\cdot, P)$. Then for all $a \in E$ the sequence $\{\{B_n, a\}\}$ is minimizing for $W_k(\cdot, P)$ so that

$$W(B_n, P) - W(\{B_n, a\}, P) = \int_{S_n(a)} f_a^n(x) P(dx) \to 0, \tag{6}$$

where $f_a^n(x) := \phi(d(x, B_n)) - \phi(d(x, a))$ and $S_n(a) := \{x \in E | f_a^n(x) > 0\}$. Consider a finite set $C = \{c_1, \ldots, c_m\} \subset E$. Since (6) holds for every $a \in E$,

$$W(B_n, P) - W(\{B_n, C\}, P) \leq \sum_{i=1}^m \int_{S_n(c_i)} f_{c_i}^n(x) P(dx) \to 0.$$

Thus, $W(\{B_n, C\}, P) \to w$ for every $C \in \mathcal{E}$. Clearly, $W(\{B_n, C\}, P) \leq W(C, P)$, and **P3** finishes the proof. ▊

The inequalities (5) are important. If

$$\phi(u) = 0 \quad \text{if and only if} \quad u = 0, \tag{7}$$

then $W_{k-1}(P) > 0$ if and only if $T \notin \mathcal{E}_{k-1}$ [19, 23]. Hence, for strictly increasing ϕ, by assuming (5) we avoid the trivial case when P is concentrated on $k - 1$ or fewer points.

2. MINIMIZING SEQUENCES

Main task of the paper is to find the conditions for E and P that ensure the convergence of minimizing sequences. At first we present some general properties of minimizing sequences.

2.1. General Properties

Denote $W_0(P) := \phi(\infty)$ and $\delta_i := W_{i-1}(P) - W_i(P)$. By (5), $\delta_i > 0$ for all $i = 1, \ldots, k$.

Note first that if $\epsilon < \delta_k$ then every ϵ-minimizing sequence $\{A_n\}$ eventually consists of sets of exactly k points. Indeed, if $|A_{n_j}| < k$ along a subsequence $\{A_{n_j}\}$, we would reach a contradiction: $W_{k-1}(P) \leq \limsup_j W_{k-1}(A_{n_j}, P) \leq W_k(P) + \epsilon$.

The following important lemma states that, for sufficently small ϵ, every ϵ-minimizing sequence $\{A_n\}$ is bounded. The proof of it can be found in [17], we repeat the argument because of its importance. Fix an $x_o \in E$.

LEMMA 2.1. *Let $\{A_n\}$ be an ϵ-minimizing sequence for $W_k(\cdot, P)$. If $\epsilon < \delta_k$ then there exists a ball $B(x_o, r_k)$ such that $A_n \subset B(x_o, r_k)$ for all n.*

Proof. We may assume without loss of generality that $|A_n| = k$ for all $n = 1, 2, \ldots$. It will be shown that, for each $l = 0, 1, \ldots, k$, there exists r_l such that $|A_n \cap B(x_o, r_l)| \geq l$ for all n. For $l = 0$ there is nothing to prove. If the statement is not true then there exists a largest l in $\{0, 1, \ldots, k-1\}$ such that the condition is satisfied by some r_l. Then, for each $j = 1, 2, \ldots$, there exists n_j such that $n_j > n_{j-1}$ and

$$l \geq |A_{n_j} \cap B(x_o, r_l + j)| \geq |A_{n_j} \cap B(x_o, r_l)| \geq l.$$

Let $B_{n_j} = A_{n_j} \cap B(x_o, r_l)$. Then $A_{n_j} \backslash B_{n_j} \subset E \backslash B(x_o, r_l + j)$.
If $l = 0$ then $B_{n_j} = \emptyset$ for each j. Thus, for each $x \in E$, $\lim_j \phi(d(x, A_{n_j})) =$

$\phi(\infty)$. But $\{A_n\}$ is ϵ-minimizing and, therefore,

$$\phi(\infty) = \int \liminf_{j \to \infty} \phi(d(x, A_{n_j})) P(dx) \leq \limsup_{j \to \infty} W_k(A_{n_j}, P) < W_k + \epsilon \quad (8)$$

by Fatou's lemma. Hence $\phi(\infty) < \infty$ and, by the assumption on ϵ, the right side of (8) is $W_k + \epsilon < \phi(\infty)$ which is a contradiction.

So $l \geq 1$, $|B_{n_j}| = l$ for all j, and $\phi(d(x, A_{n_j}))$ and $\phi(d(x, B_{n_j}))$ are eventually equal for each $x \in E$. Furthermore, $\phi(d(x, B_{n_j})) \leq \phi(d(x, x_0) + r_l)$ which, by (4) and our finiteness assumption, is integrable. So, by Lebesgue's dominated convergence theorem,

$$W_l(B_{n_j}, P) - W_l(A_{n_j}, P) = \int \Big(\phi(d(x, B_{n_j})) - \phi(d(x, A_{n_j})) \Big) P(dx) \to 0$$

from which it follows that $\limsup_j W_l(B_{n_j}, P) = \limsup_j W_k(A_{n_j}, P) \leq W_k + \epsilon < W_l$, which is a contradiction. The proof is complete. $\blacksquare$

The following corollary follows almost directly from Lemma 2.1 [17].
For each finite $A = \{a_1, \ldots, a_l\} \subset E$ define the sets

$$S^o(a_i) := \{x \in E \,|\, d(x, a_i) < d(x, A \backslash \{a_j\})\},$$
$$S^o_\phi(a_i) := \{x \in E \,|\, \phi(d(x, a_i)) < \phi(d(x, A \backslash \{a_j\}))\}.$$

COROLLARY 2.1.   *Let $\{A_n\}$ be an ϵ-minimizing sequence for $W_k(\cdot, P)$. If $\epsilon < \delta_k$ then there exists an $\alpha > 0$ such that $P(S^o_\phi(a^n)) > \alpha$ for all $a_n \in A_n$.*

DEFINITION 2.1.   Let $\epsilon \geq 0$. Any $A \in \mathcal{E}_k$ satisfying $W_k(A, P) \leq W_k(P) + \epsilon$ is called an ϵ-optimal k-centre of P. The set of all ϵ-optimal k-centres is denoted $\mathcal{U}_k^\epsilon(P)$. A 0-optimal k-centre is called a k-centres of P. The set of all k-centres is $\mathcal{U}_k(P)$.

Generally the k-centres are not unique or they might not exist. The uniqueness is one of the crucial differences between 1-centres and k-centres in general: when E is a rotund normed space and ϕ is strictly convex then, unlike the cases $k > 1$, the 1-centre is always unique. In this paper we take account of the possible nonuniqueness, but we do not deal with existence problems. However, it is not hard to see that the assumption **B** used in Theorem 2.1 is sufficient for the existence of k-centres. We also remark that the existence of centres is a property that does not depend heavily on k. For a large class of Banach spaces the existence of 1-centres was proved in [12].

2.2. τ-convergence

Consider the metric space $(\mathcal{E}, h)$, where h stands for the Hausdorff metrics: $h(A, B) = \max\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\}$. We denote Hausdorff convergence of the sets by $A_n \to A$ and it will be referred to as strong convergence. For a class $\mathcal{U} \subset \mathcal{E}$ we write $A_n \to \mathcal{U}$ if $\inf_{B \in \mathcal{U}} h(A_n, B) \to 0$. Clearly $A_n \to \mathcal{U}$ if every subsequence of $\{A_n\}$ admits a subsequence which converges to an element of $\mathcal{U}$, but not conversely.

Suppose E is equipped with a Hausdorff topology τ (not necessarily comparable with the metric topology). The sets

$$\mathcal{V}(\mathcal{O}_1, \ldots, \mathcal{O}_l) := \{A \in \mathcal{E}_k | A \subset \mathcal{O}_1 \cup \ldots \cup \mathcal{O}_l, \quad A \cap \mathcal{O}_i \neq \emptyset, \ i = 1, \ldots, l\},$$

$l = 1, 2, \ldots,$ $\mathcal{O}_i \in \tau$ form a basis for a topology which is called Vietoris (or exponential) topology. If τ is the metric topology, then the Vietoris topology for $\mathcal{E}$ coincides with the topology generated by the Hausdorff metric. For finite sets, the convergence in the Vietoris topology generalizes Hausdorff convergence. Indeed, a sequence $\{A_n\}$ converges in the Vietoris topology to $A = \{a_1, \ldots, a_l\}$ if, for all sufficiently large n, there exists a partition $A_n = A_1^n \cup \cdots \cup A_l^n$ such that $|A_i^n| \geq 1$ and $\{A_i^n\}$ converges in τ to a_i for each $i = 1, \ldots, l$. This convergence is denoted $A_n \xrightarrow{\tau} A$.

From now on we assume that E is equipped with a Hausdorff topology τ, such that the mapping $d(\cdot, y) : (E, \tau) \mapsto \Re$ is sequentially lower semicontinuous or, equivalently, that every closed ball of E is sequentially τ-closed. Then, ϕ being continuous and monotonic, $\phi(d(\cdot, y))$ is also sequentially lower semicontinuous, i.e.

$$\liminf \phi(d(x_n, y)) \geq \phi(d(x, y)) \quad \text{for all } y \in E \tag{9}$$

if $x_n \xrightarrow{\tau} x$. In a normed space the weak topology satisfies (9). Similarly, when E is a dual, the weak* topology satisfies (9).

The next lemma is very important. A version of it can be found in [17].

LEMMA 2.2. *Let $A_n \xrightarrow{\tau} A$ be a minimizing sequence for $W_k(\cdot, P)$. Then $A \in \mathcal{U}_k(P)$, and $\phi(d(x, A_n)) \to \phi(d(x, A))$, P-a.s..*

Proof. Let $A = \{a_1, \ldots, a_l\}$. Obviously, $l \leq k$. First we show that $A \in \mathcal{U}_k(P)$. The argument is adapted from Cuesta *et al.* [5]. Let $\{A_1^n, \ldots, A_l^n\}$ be the partition of A_n ensured by the definition of τ-convergence. Fix $x \in E$ and $i \in \{1, \ldots, l\}$. If $a_i^n \in A_i^n$ then $a_i^n \xrightarrow{\tau} a_i$ and so, by (9), $\liminf_n \phi(d(x, A_i^n)) \geq \phi(d(x, a_i))$. Hence, $\liminf_n \phi(d(x, A_n)) = \liminf_n \min_i \phi(d(x, A_i^n)) = \min_i \liminf_n \phi(d(x, A_i^n))$ and

$$\min_i \liminf_n \phi(d(x, A_i^n)) \geq \min_i \phi(d(x, a_i)) = \phi(d(x, A)). \tag{10}$$

By an appeal to Fatou's lemma,

$$W_k(A, P) = \int \phi(d(x, A)P(dx) \le \int \liminf_n \phi(d(x, A_n))P(dx)$$
$$\le \liminf_n \int \phi(d(x, A_n))P(dx) = \lim_n W_k(A_n, P) = W_k(P).$$

Consequently, $A \in \mathcal{U}_k(P)$ and the inequalities in the previous display are, in fact, equalities. This implies a.s. equality in (10). Thus $P(V) = 1$, where $V := \{x : \liminf_n \phi(d(x, A_n)) = \phi(d(x, A))\}$. We show that V is closed in the metric topology.

Let $x_m \in V$ and $x_m \to x$. Then, for each x_m,

$$|\phi(d(x, A)) - \liminf_n \phi(d(x, A_n))|$$
$$\le |\phi(d(x, A)) - \phi(d(x_m, A))| + |\liminf_n \phi(d(x_m, A_n)) - \liminf_n \phi(d(x, A_n))|$$
$$\le |\phi(d(x, A)) - \phi(d(x_m, A))| + \limsup_n |\phi(d(x_m, A_n)) - \phi(d(x, A_n))|$$

$$(11)$$

The sequence $\{A_n\}$ is bounded by Lemma 2.1, $|d(x_m, B) - d(x, B)| \le d(x, x_m)$ for any set B, and ϕ is uniformly continuous on bounded set. It follows that the right side of (11) tends to zero as $m \to \infty$ and so $x \in V$. Thus V is closed and $P(V) = 1$ so $T \subseteq V$.

Any subsequence $\{A_{n_j}\}$ is minimizing and τ-convergent to A. Hence, the corresponding set $V' := \{x : \liminf_{n_j} \phi(d(x, A_{n_j})) = \phi(d(x, A))\}$ contains T or, equivalently, $T \subseteq \cap_\alpha V_\alpha$, where the intersection is taken over the sets V corresponding to all subsequences of $\{A_n\}$. Thus, $T \subseteq \{x : \lim_n \phi(d(x, A_n) = \phi(d(x, A))\}$, or $\phi(d(x, A_n)) \to \phi(d(x, A))$, P-a.s. ∎

We assumed that (5) holds. Then any minimizing sequence eventually consists of sets of k elements. Also every k-centre is a set of k elements. Thus, by Lemma 2.2, any τ-limit of a minimizing sequence has exactly k-elements. Hence, if $A_n \xrightarrow{\tau} \{a_1, \ldots, a_k\}$ is a minimizing sequence, then for each n big enough there exists a labelling $A_n = \{a_1^n, \ldots, a_k^n\}$ such that $a_i^n \xrightarrow{\tau} a_i$ for all i. We now specify the convergence $\phi(d(x, A_n)) \to \phi(d(x, A))$ in terms of $\phi(d(x, a_i^n))$. The proof of the following corollary is straightforward and, therefore, omitted. It can be found in [17].

COROLLARY 2.2. *Let $A_n \xrightarrow{\tau} A$ be a minimizing sequence for $W_k(\cdot, P)$. Then $\phi(d(x, a_i^n)) \to \phi(d(x, a_i))$ for each $i = 1, \ldots, k$ and $x \in S_\phi^o(a_i) \cap T$. If ϕ is strictly increasing then $d(x, a_i^n) \to d(x, a_i)$ for each $x \in S^o(a_i) \cap T$.*

2.3.   Strong Convergence

In this section we study conditions under which τ-convergence of minimizing sequences implies strong convergence.

First Condition. Suppose ϕ satisfies (7) and let $A_n \xrightarrow{\tau} \{a_1, \ldots, a_k\}$ be a τ-convergent minimizing sequence. By Corollary 2.2, if $a_i \in T$ then $\phi(d(a_i^n, a_i)) \to 0$ and so $d(a_i^n, a_i) \to 0$. Hence, if $A \subset T$ then $a_i^n \to a_i$, for each i and $A_n \to A$.

It now follows, by Lemma 2.2, that in order to deduce strong convergence of a τ-convergent minimizing sequence one needs only that

$$A \subset T \quad \text{if} \quad A \in \mathcal{U}_k(P). \tag{12}$$

The pair of assumptions (12) and (7) is called the first condition. Obviously, (12) depends on P as well as on the setup of the problem. It might be meaningful to minimize the loss function only over the support of P. Then one can consider T as the whole space E and so (12) holds.

The condition (12) might seem technical and inessential. In the following example (12) fails and weak convergence does not imply strong convergence.

EXAMPLE 2.1.   Let $E = c_o$ and $P = \sum_{n=1}^{\infty} (\delta_{e_n} + \delta_{-e_n}) p_n$, where $e_n \in c_o$ consists of zeros except for 1 in the n-th place, and $p_n > 0$, $\sum_n p_n = \frac{1}{2}$. We take $k = 1$ and $\phi(x) = x^2$. Then $W(a, P) = \sum_{n=1}^{\infty} (\|e_n - a\|^2 + \|e_n + a\|^2) p_n$ for each $a \in E$. It is easy to see that $\{0\}$ is the unique 1-centre and that $W_1(P) = 1$. Obviously, the sequence $\{e_m\}$ is minimizing for $W_1(\cdot, P)$. Let τ be the weak topology. Then $e_n \xrightarrow{\tau} 0$ but $e_n \not\to 0$. However $0 \notin T$, (12) is not satisfied and the example does not contradict the first condition.

Condition (7) also cannot be omitted. We can modify Example 2.1 in such a way that $A_n \xrightarrow{\tau} A \subset T$ but $A_n \not\to A$.

EXAMPLE 2.2.   Let $E = c_o$, $k = 1$, and consider the following ϕ and P:

$$\phi(x) = \begin{cases} 0, & \text{if } x \in [0, 1] \\ (x-1)^2, & \text{if } x > 1 \end{cases}, \quad P = \sum_{n=1}^{\infty} (\delta_{2e_n} + \delta_{-2e_n}) p_n + \delta_0 p_0,$$

where $p_n > 0$, $n = 0, 1, 2, \ldots$, $p_0 + 2 \sum_{n=1}^{\infty} p_n = 1$.

Again, $\mathcal{U}_1(P) = \{0\}$, $W_1(P) = 1 - p_0$, and the sequence $\{e_m\}$ is minimizing. The assumptions on P and $\{e_m\}$ are satisfied. But (7) is not satisfied and the strong convergence fails.

Second Condition. Suppose now ϕ is strictly increasing, but (12) is not satisfied. Then some additional assumptions should be made.

Let $(E, \|\cdot\|)$ be a normed space, and τ be a topological vector space topology on E that satisfies (9). This happens, for example, if E is a separable dual and τ is the weak* topology. Consider a τ-convergent minimizing sequence

$A_n \xrightarrow{\tau} \{a_1, \ldots, a_k\}$, and let $x_i \in S^o(a_i) \cap T$. By Corollary 2.1 $S^o(a_i) \cap T$ is not empty. The convergence $a_i^n \xrightarrow{\tau} a_i$ implies $x_i - a_i^n \xrightarrow{\tau} x_i - a_i$, and, by Corollary 2.2 $\|x_i - a_i^n\| \to \|x_i - a_i\|$, $i = 1, \ldots, k$. Consequently, for each $i = 1, \ldots, k$, there exists an x_i such that $x_i - a_i^n \xrightarrow{\tau} x_i - a_i$ and $\|x_i - a_i^n\| \to \|x_i - a_i\|$.

Sometimes (E, τ) possesses the property:

$$y_n \xrightarrow{\tau} y \quad \text{and} \quad \|y_n\| \to \|y\| \quad \text{imply} \quad y_n \to y. \tag{13}$$

If so, then $x_i - a_i^n \to x_i - a_i$ and $a_i^n \to a_i$. Hence, the following proposition holds (recall that τ satisfies (9)).

PROPOSITION 2.1.   *Let $(E, \|\cdot\|)$ be a normed space. Assume (E, τ) is a topological vector space, satisfying (13). Let $A_n \xrightarrow{\tau} A$ be a minimizing sequence for $W_k(\cdot, P)$. If ϕ is strictly increasing then $A_n \to A$.*

If (E, τ) satisfies conditions of Proposition 2.1 then (E, τ) is said to have τ-*Kadec-Klee* property [1, 15, 20]. Usually Kadec-Klee property is defined with respect to the weak topology and then it is sometimes called the *Radon-Riesz property* or *property H* [16, 21]. In the best approximation context the Kadec-Klee property is also called *property A* [22]. The τ-Kadec-Klee property is a straightforward generalization. Often $(E, \|\cdot\|)$ is assumed to be a Banach space and τ is weaker than norm-topology. In this case conditions (13) and (9) are equivalent to the following [14, 15]

$$\{x_n\} \subset B(0) \quad \& \quad x_n \xrightarrow{\tau} x \quad \& \quad \text{sep}\{x_n\} > 0 \quad \text{imply} \quad \|x\| < 1.$$

Here $B(0)$ is the unit ball in E and $\text{sep}\{x_n\} := \inf\{\|x_m - x_n\| : m \neq n\}$. For examples of Banach spaces with τ-Kadec-Klee property see [15].

EXAMPLE 2.3.   Proposition 2.1 needs not hold if ϕ is not strictly increasing. Let $E = l_2, k = 1$ and $\phi \equiv 0$ in $[0, \sqrt{2}]$ and $\phi(x) = (x - \sqrt{2})^2$ otherwise. Consider the P as in Example 2.1. Clearly $W_1(P) = 0$, and the sequence $\{e_m\}$ is minimizing. The space l_2 has Kadec-Klee property, $e_m \xrightarrow{\tau} 0$, but $e_m \not\to 0$.

The Main Theorem.   So far we have investigated the possibility of deducing strong convergence of a minimizing sequence from τ-convergence when τ satisfies (9). The first and second conditions provide sufficient conditions, in terms of P and E, respectively. How to ensure the τ-convergence of a minimizing sequence is a general and possible unsolvable problem. Be-

cause of the boundedness of the minimizing sequences it is enough that:

B: every closed ball of E is sequentially τ-compact.

Obviously, **B** is stronger than (9). The weak* topology in a separable dual satisfies **B**. In particular, the weak topology in a reflexive space satisfies **B**. The main theorem now directly follows from Lemma 2.1, first condition and Proposition 2.1.

THEOREM 2.1. *If E admits a topology τ which satisfies **B** and either*
1) *(7) and (12) are satisfied, or*
2) *ϕ is strictly increasing and E is a normed linear space such that (E, τ) has τ-Kadec-Klee property,*
then every minimizing sequence for $W_k(\cdot, P)$ has a subsequence that converges strongly to an element of $\mathcal{U}_k(P)$.

If E is a reflexive Banach space with the Kadec-Klee property then the assumptions of Theorem 2.1 are satisfied. Such a space is sometimes called a *Efimov-Stechkin* space [21, 22]. A Efimov-Stechkin space posesses several properties which are useful from the point of view of best approximation. For example, a Banach space is a Efimov-Stechkin space if and only if every weakly closed set is approximatively compact [13]. The Efimov-Stechkin spaces include all uniformly rotund spaces, the spaces $L_p(\Omega; \Re)$ and the Sobolev spaces $W_p^n(a, b)$ $(1 < p < \infty)$. A good overview of Efimov-Stechkin spaces and related geometrical properties can be found in [21].
Examples 2.1 and 2.2 show the importance of conditions **1)** of Theorem 2.1. The space c_o with the weak topology does not satisfy (13) so the same examples show that (13) cannot be omitted from **2)** in Theorem 2.1. If ϕ does not satisfy (7) then Theorem 2.1 needs not hold even in a Efimov-Stechkin space. To see this reconsider Example 2.3 and verify that $e_m \not\to \mathcal{U}_1(P)$. Indeed, $a = (a_1, a_2, \ldots) \in \mathcal{U}_1(P)$ if and only if $\|e_n - a\|^2 \leq 2$ and $\|e_n + a\|^2 \leq 2$ for each n. Then $|a_n| \leq \sqrt{2} - 1$ for $n = 1, 2, \ldots$, and $\|e_m - a\| \geq 2 - \sqrt{2}$.

COROLLARY 2.3. *If the conclusion of Theorem 2.1 holds and there is a unique k-centre then any minimizing sequence for $W_k(\cdot, P)$ converges to A.*

Remark 2. 1. If there is a unique k-centre the word "sequentially" can be skipped in **B**. Let $\{A_n\}$ be a minimizing sequence. Let B be a closed ball containing $\{A_n\}$. Take $E := B$. Now (E, τ) is τ-compact, which implies

the compactness of $\mathcal{E}_k$ for Vietoris topology [9, 3.12.26]. The loss function $W_k(\cdot, P)$ is lower semicontinuous on $\mathcal{E}_k$ by Fatou's lemma, and attains its minimum on a (countably) compact subset of $\mathcal{E}_k$. Now it is easy to see that every minimizing sequence $\{A_n\}$ satisfies $A_n \xrightarrow{\tau} \mathcal{U}_k(P) = A$, and both convergence conditions apply.

3. EMPIRICAL K-CENTRES

3.1. The Consistency Problem

Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed E-valued random elements defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and having common distribution P. Without loss of generality we may assume that $(\Omega, \mathcal{F}, \mathbf{P})$ is complete, i.e. every subset of a 0-measure set is measurable. Let $\{P_n\}$ be the sequence of empirical measures as defined in (3). The k-centres of P_n^ω and P are called *empirical k-centres* and *theoretical k-centres*, respectively. Hence, (for fixed n and ω) the set of all empirical k-centres is $\mathcal{U}_k(P_n^\omega)$ and the set of all theoretical k-centres is $\mathcal{U}_k(P)$. Calculated from the sample of n observations, $\mathcal{U}_k(P_n^\omega)$ can be regarded as an estimator of (usually unknown) $\mathcal{U}_k(P)$. Is such an estimator consistent, i.e. does $\mathcal{U}_k(P_n)$ (in some sense) converge to $\mathcal{U}_k(P)$, $\mathbf{P}$-a.s.? In this paper we study the following sense of convergence: let $\{A_n\}$ be an arbitrary sequence of empirical k-centres, i.e. $A_n \in \mathcal{U}_k(P_n^\omega)$ for each n. We call empirical k-centres *consistent* if $h(A_n, \mathcal{U}_k(P)) \to 0$ for $\mathbf{P}$-a.e. ω. Thus, empirical k-centres are consistent if

$$\mathbf{P}\left\{\omega \,\Big|\, \sup_{A_n \in \mathcal{U}_k(P_n^\omega)} h(A_n, \mathcal{U}_k(P)) \to 0\right\} = 1. \tag{14}$$

Clearly (14) holds if every subsequence of empirical k-centres has a further subsequence converging to a theoretical k-centre, $\mathbf{P}$-a.s.. Then

$$\mathbf{P}\left\{\omega \,\Bigg|\, \begin{array}{l} \text{any subsequence } \{A_{n_j}\}, \text{ where } A_{n_j} \in \mathcal{U}_k(P_{n_j}^\omega) \text{ has} \\ \text{a subsequence} \{A_{n_{j_l}}\} \text{ such that } A_{n_{j_l}} \to A \in \mathcal{U}_k(P) \end{array}\right\} = 1. \tag{15}$$

Also, the k-variance $W_k(P_n)$ of an empirical measure is an estimator of $W_k(P)$, which justifies the investigation of the convergence $W_k(P_n) \to W_k(P)$, $\mathbf{P}$-a.s.. If the latter holds, then empirical k-variances are called consistent. Since $|\phi(d(x, A_n)) - \phi(d(x, B_n))| \to 0$ if $h(A_n, B_n) \to 0$, it is not hard to see that by the strong law of large numbers (SLLN) the consistency of empirical k-centres implies that of empirical k-variances. In

Lemma 3.1 we show the latter without (14).

The problem of consistency of k-centres has attracted much attention. In [23], the consistency of k-centres and k-variances have been proved for finite dimensional spaces. For trimmed k-centres in Euclidean spaces, similar results were obtained in [8] (by calculation of a trimmed k-centre of P in level $\delta \in (1, 0)$, a restricted part of P is used. This restriction is chosen such that it has total mass at least $1 - \delta$ and minimizes the loss function over all such restrictions.) The strongly related case of compact metric spaces were considered in [29]. The leading consistency results for k-centres in uniformly convex spaces were obtained by Cuesta *et al.* in [5, 6]. They used Skorohod Representation together with a version of Proposition 2.1. The first condition was introduced in [19]. There we also developed the connection between the consistency of k-centres and best approximation theory. This allowed us to consider the consistency of k-centres as the semicontinuity of a metric projection in the Lebsgue-Bochner space. This approach has its origins in [4, 5, 7] and was studied also in [18]. The disadvantage of it is the restriction on ϕ - the latter is usually a power function. In the present paper we try to keep ϕ as general as possible.

If the theoretical k-centre is unique, i.e. $\mathcal{U}_k(P) = \{A\}$, then (14) and (15) reduce to the following

$$\mathbf{P}\{\omega \,|\, \text{for any sequence } \{A_n\}, \ A_n \in \mathcal{U}_k(P_n^\omega) \text{ it holds } A_n \to A\} = 1. \ (16)$$

In previous papers discussing consistency of k-centres [5, 6, 8, 23, 24] the uniquess of theoretical k-centres was assumed, i.e. only (16) was considered. We deal with the generalized versions of consistency.

3.2.   Main Consistency Results

In order to apply the results of Section 2, we show that every sequence of empirical k-centres is minimizing for $W_k(\cdot, P)$, $\mathbf{P}$-a.s..

PROPOSITION 3.1.   *Let $l, N > 0$ be integers and denote $\mathcal{E}_l^N := \{A \in \mathcal{E}_l : A \subset B(x_0, N)\}$. If $\{P_n\}$ is a sequence of probability measures converging weakly to P then for a sequence $\{B_n\}$ such that $B_n \in \mathcal{E}_l^N$, it holds*

$$\lim_n |W_l(B_n, P_n) - W_l(B_n, P)| = 0. \tag{17}$$

Proof.   Define $\mathcal{A} := \{f_A : A \in \mathcal{E}_l^N\}$, where $f_A(x) = \phi(d(x, A))$. The family $\mathcal{A}$ consists of continuous functions, bounded by $g(x) := \phi(d(x, x_o) + N)$;

$\mathcal{A}$ is equicontinuous [25]. If P_n converges to P weakly and $\int g dP_n \to \int g dP < \infty$, then $\lim_{n\to\infty} \sup_{f\in\mathcal{A}} |\int f dP_n - \int f dP| = 0$, by the Ranga Rao theorem [28]. That proves the proposition. ∎

LEMMA 3.1. *For* **P**-*a.e.* ω,
i) $W_k(P_n^\omega) \to W_k(P)$,
ii) every sequence $\{A_n\}$ *where* $A_n \in \mathcal{U}_k(P_n^\omega)$ *is minimizing for* $W_k(\cdot, P)$.

Proof.   At first we show the existence of a set $\Omega_o \in \mathcal{F}$ such that $\mathbf{P}(\Omega_o) = 1$ and for every $\omega \in \Omega_o$,
1) $\limsup_n W_k(P_n^\omega) \le W_k(P)$,
2) $\int \phi(d(x, x_0) + N) dP_n^\omega \to \int \phi(d(x, x_0) + N) dP$, for each $N = 1, 2, \ldots$,
3) $P_n^\omega \Rightarrow P$, where $\Rightarrow$ stands for the weak convergence of probability laws.
Let $A \in \mathcal{U}_k(P)$. By SLLN, $W(A, P_n) \to W(A, P) = W_k(P)$, **P**-a.s.. Since $W_k(P_n) \le W(A, P_n)$, we now obtain that $\limsup_n W_k(P_n) \le W_k(P)$, **P**-a.s.. Let Ω_1 be the set of **P**-measure 1 where the latter holds.
By (4), $\int \phi(d(x, x_0) + N) P(dx) < \infty$ for every $N = 1, 2, \ldots$. By SLLN, for each N there exists a set $\Omega_N \in \mathcal{F}$ such that $\mathbf{P}(\Omega_N) = 1$ and $\int \phi(d(x, x_0) + N) dP_n^\omega \to \int \phi(d(x, x_0) + N) dP$ if $\omega \in \Omega_N$. Take $\Omega_2 := \cap_N \Omega_N$. Now $\mathbf{P}(\Omega_2) = 1$ and 2) holds for each $\omega \in \Omega_2$.
Finally, due to the well-known result of Varadarajan (see, e.g. [2]) $\mathbf{P}(\Omega_3) = 1$, where $\Omega_3 := \{\omega : P_n^\omega \Rightarrow P\}$. Now take $\Omega_o := \Omega_1 \cap \Omega_2 \cap \Omega_3$.
We shall show that Ω_o is the required set. Let $\omega \in \Omega_o, P_n := P_n^\omega$, and consider a sequence $\{A_n\}$ such that $A_n \in \mathcal{U}_k(P_n)$, $\forall n$. The first step is to prove the boundedness of $\{A_n\}$. For we that proceed as in the proof of Lemma 2.1, i.e. we show that, for each $l = 0, 1, \ldots, k$ there exists r_l such that $|A_n \cap B(x_o, r_l)| \ge l$ for all n. Suppose r_1 does not exists. Then there exists a subsequence $\{A_{n_j}\}$ such that $\lim_{j\to\infty} h_j(x) = \phi(\infty)$, where $h_j : E \mapsto [-\infty, \infty]$, $h_j(x) = \phi(d(x, A_{n_j}))$. Note that the convergence $x_j \to x$ yields $h_j(x_j) \to \phi(\infty)$. Hence, $P_{n_j} h_j^{-1} \Rightarrow \delta_{\phi(\infty)}$ [2, Thm. 5.5], implying that $\phi(\infty) \le \liminf_j \int h_j dP_{n_j} = \liminf_j W_k(A_{n_j}, P_{n_j}) \le W_k(P) < \phi(\infty)$ [2, Thm. 5.3]. This is a contradiction, so r_1 exists.
Suppose now that r_l exists, and let $B_n = A_n \cap B(x_o, r_l)$. Suppose r_{l+1} does not exist. Then, for each $j = 1, 2, \ldots$, there exists n_j such that $n_j > n_{j-1}$ and $f_j(x) := \phi(d(x, B_{n_j})) - \phi(d(x, A_{n_j}))$ is eventually zero for each $x \in E$. Furthermore, $\phi(d(x, B_{n_j})) \le \phi(d(x, x_o) + N) =: g(x)$, where $N \ge r_l$. It is not hard to see that $f_j(x_j) \to 0$, if $x_j \to x$, implying that $P_{n_j} f_j^{-1} \Rightarrow \delta_0$.

Since for every j, $f_j \leq g$ and, by 2), $\int g dP_{n_j} \to \int g dP$, it holds

$$\int f_j dP_{n_j} = W_l(B_{n_j}, P_{n_j}) - W_k(P_{n_j}) \to 0. \tag{18}$$

The convergence (18) follows almost immediately from the general theory of weak convergence, see e.g. [17, Cor. 3.2.2] or [2, Thm. 5.4].
From (18) and 1) follows that

$$\limsup_j W_l(B_{n_j}, P_{n_j}) = \limsup_j W_l(A_{n_j}, P_{n_j}) \leq W_k(P).$$

On the other hand, by (17), $\limsup_j W_l(B_{n_j}, P_{n_j}) = \limsup_j W_l(B_{n_j}, P) \geq W_l(P)$. Hence, $W_l(P) \leq W_k(P)$ which is a contradiction. Therefore, $\{A_n\}$ is bounded and (3.1) yields

$$|W_k(P_n) - W_k(A_n, P)| \to 0. \tag{19}$$

From (19) and 1) we get i). From i) and (19) we now get ii). ∎

Remark 3. 1. Just like in Lemma 2.1, the argument of Lemma 3.1 holds also for the ϵ-optimal empirical k-centres, provided $\epsilon < \delta_k$. So the existence of empirical k-centres is not needed for proving the consistency of k-variance.

For finite-dimensional spaces or compact metric spaces, the statements of Lemma 3.1 were proved in [23, 29], respectively. In their proofs the Δ_2-property of ϕ was assumed. In [25] these ideas were used to prove *i)* for general separable metric space E. However, our proof is much shorter and more general. It generalizes the approach introduced in [5]. For trimmed k-centres in the space $\Re^m$, a counterpart of Lemma 3.1 can be found in [8]. We also remark that if E is a normed space, and $\phi(x) = x^p$, $p \geq 1$, then Lemma 3.1 follows from best approximation theory [18, 19].
From Theorem 2.1 and Lemma 3.1 we have the following theorem.

THEOREM 3.1. *Suppose E admits a topology τ, satisfying **B** and either*
1) *ϕ satisfies (7), and (12) are satisfied, or*
2) *ϕ is strictly increasing and E is a normed linear space such that (E, τ) has τ-Kadec-Klee property,*
then (15) holds.

The assumption **2)** of Theorem 3.1 ensures (15) in a Efimov-Stechkin space. This generalizes the consistency results for k-centres in [23, 29, 26, 5], where finite-dimensional space, compact metric space, Hilbert space or

uniformly convex spaces were considered, respectively. The assumption **1)** ensures (15) in separable duals, provided (12) and (7) hold. We are not aware of previous consistency results, where no additional assumptions on E has been made.

EXAMPLE 3.1.   To emphasize the role of the definition (15) of consistency we consider the example given in [5]. Let $k = 2$, and P be a probability measure on $(\Re, \mathcal{B})$ such that $P(A) = \frac{1}{6} Leb(A \cap ([0, 2] \cup [3, 4])) + \frac{1}{2} I_A(2)$. Let $\phi \equiv 0$, in $[0, 1]$, and $\phi(x) = (x - 1)^2$ if $x > 1$. Then $A \in \mathcal{U}_2(P)$ if and only if $W_2(A, P) = 0$. The 2-centres depend on the support P. Such type of k-centres are sometimes called *best k-nets*. Now, $\mathcal{U}_2(P) = \{\{1, b\}|b \in [3, 4]\}$ and the empirical best 2-net is unique, **P**-a.s.. Then, for **P**-a.e. ω, it holds $\mathcal{U}_2(P_n^\omega) = A_n :=: \{a_1^n, a_2^n\}$, where the points in A_n are ordered such that $a_1^n < a_2^n$. In this case $a_1^n \to 1$ and the sequence $\{a_2^n\}$ has two cluster points: 3 and 4.5. Therefore, every subsequence of best empirical 2-nets has a subsubsequence converging either to $\{1, 3\} \in \mathcal{U}_k(P)$ or to $\{1, 4.5\} \in \mathcal{U}_k(P)$, **P**-a.s.. Originally this example was to demonstrate that the a.s. convergence of empirical k-centres needs not hold, if $\mathcal{U}_k(P)$ has more than one set. However, we still have the a.s. convergence to the set $\mathcal{U}_k(P)$. So, in space $\Re$ the a.s. convergence of empirical k-centres in the sense (16) might not take place; because of Theorem 3.1, the consistency in a wider sense (15) always holds.

REFERENCES

1. M. Besbes, S.J. Dilworth, P.N. Dowling, C.J. Lennard, New convexity and fixed point properties in Hardy and Lebesgue-Bochner spaces, *J.Funct. Anal.* **119** (1994), 340–357.

2. P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1999.

3. J.A. Cuesta-Albertos Medidas de centralización multidimensionales (Ley Fuerte de los Grandes Números, *Trab. Estad. e Invest. Oper.* **35**, (1984), 3–16.

4. J.A. Cuesta, C. Matran, Strong laws of large numbers in abstract spaces via Skorohod's Representation Theorem, *Sankhya Ser. A* **84**, (1986), 98–103.

5. J.A. Cuesta, C. Matran, The strong law of large numbers for k-means and best possible nets of Banach valued random variables, *Prob. Th. Rel. Fields* **78**, (1988a), 523–534.

6. J.A. Cuesta, C. Matran, Uniform consistency of r-means, *Stat. Probab. Lett.* **6**, (1988b), 65–71.

7. J.A. Cuesta-Albertos, A. Gordaliza, C. Matran, Consistency of L_p-best monotone approximations, *J. Stat. Plann. Inference* **47**, (1995), 297–318.

8. J.A. Cuesta-Albertos, A. Gordaliza, C. Matran, Trimmed k-means: An attempt to robustify quantizers, *Ann. Math. Statist.* **25**, (1997), 553–576.

9. R. Engelking, "General Topology" (Russian), Mir, Moskva, 1986.

10. S. Graf, H. Luschgy, "Foundations of Quantization for Probability Distributions", Springer, 2000

11. R. Gray, Quantization, *IEEE Trans. Inform, Theory* **44**, 2325–2383.

12. N. Herrndorf, Approximation of vector-valued random variables by constants, *J. Approx. Theory* **37**, (1983), 175–181.

13. R.B. Holmes, "A Course on Optimization and Best Approximation," Singer, New York-Heidelberg-Berlin, 1972.

14. R. Huff, Banach spaces which are nearly uniformly convex, *Rocky Mountain J. Math* **10** (1980), 743–749.

15. M.A. Khamsi, On uniform Opal condition and uniform Kadec-Klee property in Banach and metric spaces, *Nonliner Anal.* **26**, 1733–1748.

16. B.-L. Lin, P.-K. Lin, Property (H) in Lebesgue-Bochner function space, *Proc. Amer. Math. Soc.* **95**, (1985), 581–584.

17. J. Lember, Consistency of Empirical k-centres (doctoral dissertation), *Dissertationes Mathematicae Universitatis Tartuensis* **20**, Tartu, 1999.

18. J. Lember, Consistency of k-centres via metric projection, *in*: "Proceedings of the Fourth Hungarian Colloquium on Limit Theorems of Probability and Statistics" (2000) (accepted).

19. J. Lember, K. Pärna, Strong consistency of k-centres in reflexive spaces, *in*: "Probability Theory and Mathematical Statistics", pp. 441–452, VSP/TEV Vilnius-Utrecht-Tokyo, 1999.

20. C. Lennard, A new convexity property that implies a fixed point property for L_1, *Studia Math.* **100**, (1991), 95–108.

21. R.E. Megginson, "An Introduction to Banach Space Theory", Springer, 1998.

22. E.V. Oshman, Continuity of metric projection, (Russian) *Mat. Zametki* **37**, (1985), 200–211.

23. D. Pollard, Strong consistency of k-means clustering, *Ann. Math. Statist.* **9**, (1981), 135–140.

24. D. Pollard, A central limit theorem for k-means clustering, *Ann. Probab.* **10**, (1982), 919–926.

25. K. Pärna, On the existence and weak convergence of k-centres in Banach spaces, *Acta et Comment Univ. Tartuensis* **798**, (1986), 19-36.

26. K. Pärna, Strong consistency of k-means clustering criterion, *Acta et Comment Univ. Tartuensis* **733**, (1990), 86-96.

27. K. Pärna, J. Lember, A. Viiart, Approximation of distributions by sets, *in:* "Classification in the information Age", pp. 215-224, Springer-Verlag, Heidelberg, 1999.

28. R. Ranga Rao, Relations between weak and uniform convergence of measures with applications, *Ann. Math. Statist.* **33**, (1962), 659–680.

29. H. Sverdrup-Thygeson, Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces, *Ann. Math. Statist.* **9**, (1981), 141–145.

30. T, Tarpey, L. Li, B. Flury, Principal points and self-consistent points of elliptical distributions, *Ann. Math. Statist.* **23**, (1995), 103–112.