Journal of Convex Analysis
Volume 10 (2003), No. 1, 169–184

Convexity and the Natural Best Approximation in Spaces of Integrable Young Measures*

Zvi Artstein†
Department of Mathematics,
The Weizmann Institute of Science, Rehovot 76100, Israel
zvi.artstein@weizmann.ac.il

Cristian Constantin Popa
Department of Mathematics,
The Weizmann Institute of Science, Rehovot 76100, Israel
popa@wisdom.weizmann.ac.il

Received June 8, 2001
Revised manuscript received January 18, 2002

The natural best approximation in function spaces singles, out of the family of best L_1-approximation of an integrable function in a convex set, the element which is the limit as p converges to 1+, of the unique best L_p-approximation of the function. The present paper extends the result to convex sets in spaces of integrable Young measures. Such spaces lack a standard affine structure. In this paper convexity is considered via a limiting procedure. Consequently, the proof of the existence of a natural best approximation does not rely on tools like weak convergence, available in an ordinary function space. Rather, the interplay of compactness and convexity in the relaxed setting plays a major role.

1. Introduction

Probability measure valued maps, also called Young measures, play a prime role in depicting solutions and parameters when instantaneous changes and oscillations are expected in variational problems. Such solutions were introduced by L.C. Young as generalized curves in the calculus of variations, and by J. Warga in the framework of optimal control theory. see Young [31] and Warga [30] for the general theory and historical accounts. More recent applications include solutions of partial differential equations and variational problems, and the description of limit behavior of systems when some parameters are rapidly oscillating. Consult Tartar [28], Demengel and Temam [12], Valadier [29], Artstein [2]. While in the mentioned applications there was no need for the Young measures to be equipped with a linear or with a convexity structure, more recent applications call for that. Hence a study of notions for linear variations and convexity structures of Young measures has begun. Useful developments in these directions are offered in Pedregal [22], Roubiček [24], and in the monograph Roubiček [26]. A study of norm-like structures was carried out by Balder [7], [8], Kružik and Roubiček [15] Piccinini and Valadier [23], Artstein [4], [5]. The last two papers develop notions of linearity and convexity of Young measures which are based on norm-like approximations of ordinary functions, and examine the problem of a best approximation of an ordinary map within a convex set of L_p-Young measures. The present paper continues this effort. A particular case in [4] is the case of relaxed σ-fields

*Research supported by grants from the Israel Science Foundation, by the MINERVA Foundation, Germany, and by INTAS, Belgium.
†Incumbent of the Hettie H. Heineman Professorial Chair in Mathematics.

ISSN 0944-6532 / $ 2.50

when $p = 2$, namely, appropriate measure valued limits of standard σ-fields. Then the best approximation is the relaxed conditional expectation of the map. It is shown in [5] that when $p > 1$ the property that a unique best approximation exists and is unique, is carried over from the vector function framework to the Young measure setting.

In this paper we pay special attention to the case $p = 1$. As in the case of ordinary functions, best approximations may not exist, and when they exist, they may not be unique. We follow in this paper the important development initiated in D. Landers and L. Rogge [17], [18], who introduced the notion of natural best approximation. The notion singles out a unique best approximation with the property that it approximates the unique L_p approximation for $p = 1+\varepsilon$ with $\varepsilon > 0$ small enough. We show that under appropriate general conditions, a natural best integrable approximation can be identified within the Young measures framework.

To this end we need to further develop the basic convexity arguments within the Young measures framework. Specifically, the definition of convexity is not purely algebraic, rather, it depends on the norm in question. When the natural best approximation concept is examined, the arguments involve more than one L_p-norm. Therefore, we have to examine the effect of different norms on the convexity notion. The proof pertaining to the natural best approximation follows the general lines set out in Landers and Rogge [17], but since some of the tools available in function spaces are not available in the relaxed spaces, we have to modify the arguments accordingly.

The paper is organized as follows. Up front in Section 2, even before recalling the relevant notions and terminology, we state the main result concerning the natural minimal norm problem in the space of integrable Young measures. The natural best approximation is an equivalent problem, which in the Young measures setting needs some explanation. This result is displayed in Section 5. After stating the main result in Section 2, we recall, for the benefit of the readers who are not familiar with Young measures, the necessary notions, and set the framework for the results. In Section 3 we display some tools for the analysis of convexity in the relaxed framework, examining in particular the impact of different norms on the convexity notion for Young measures. The proof of the main result is given in Section 4. In the closing section we display comments and examples concerning possible extensions and applications.

2. The main result and the setting

We start this section with the underlying definition and the statement of the main result concerning a minimal norm problem. Following the statement, for the benefit of the readers who may not be familiar with the terminology it relates to, we recall the basic notions, and display the setting within which the result is verified.

We denote by L_{1+} the collection of functions which are p-integrable for some $p > 1$ and, likewise, $\mathcal{Y}_{1+}$ denotes the collection of Young measures which are p-integrable for some $p > 1$. The following terminology is in line with the terminology introduced in Landers and Rogge [17] for function spaces.

Definition 2.1. Let $C \subseteq \mathcal{Y}_1$ be such that for every $p > 1$ close enough to 1, there exists a unique element in C, say $\boldsymbol{\mu}_p(C)$, with minimal $\| \, \|_p$-norm in C. An element $\boldsymbol{\mu}_1 \in C$ which minimizes the $\| \, \|_1$-norm in C, is called the natural minimizer of the $\| \, \|_1$-norm in

C if the following two properties hold.

(i) For every $\boldsymbol{\nu} \in C$, $\boldsymbol{\nu} \neq \boldsymbol{\mu}_1$, there exists a $p(\boldsymbol{\nu}) > 1$ such that $\|\boldsymbol{\mu}_1\|_p < \|\boldsymbol{\nu}\|_p$ for every $1 < p \leq p(\boldsymbol{\nu})$, and

(ii) $\boldsymbol{\mu}_p(C)$ converges in the sn_1-metric, as $p \to 1+$, to $\boldsymbol{\mu}_1$.

Note that if $\boldsymbol{\mu}_1 \in C$ satisfies properties (i) and (ii), it still may not be a minimizer of the $\| \, \|_1$-norm in C. Therefore the latter property is assumed explicitly. It is clear that if a natural minimizer of the $\| \, \|_1$-norm in C exists, then it is unique. The terminology "natural" follows Landers and Rogge [17]; properties (i) and (ii) reflect well-posedness and stability with respect to a perturbation in the minimization criterion; the stability identifies, indeed, a natural choice among the minimizers.

The statement of the main result follows the result in Landers and Rogge [17], adapted here to the relaxed setting.

Theorem 2.2. *Let $C \subseteq \mathcal{Y}_1$ be a K_1-convex set. Let $\beta_1(C)$ be the collection of elements of minimal $\| \, \|_1$-norm in C. Suppose that $\beta_1(C)$ is not empty and contains elements in $\mathcal{Y}_{1+}$. Then for every $p > 1$ close enough to 1, there exists a unique element in C, say $\boldsymbol{\mu}_p(C)$, with minimal $\| \, \|_p$-norm in C. Moreover, the natural minimizer of the $\mathcal{Y}_1$-norm in C exists and it is the minimizer of the functional*

$$\int_0^1 \int_{R^d} \sum_{i=1}^{d} |x_i| \ln |x_i| \nu(t)(dx)dt \tag{1}$$

among all the elements in $\beta_1(C)$.

We present now, very briefly, the setting in which our problem is posed, and display the terminology which the main result refers to. For elaboration on Young measures in general see Valadier [29], Balder [9], Roubiček [26]. For details concerning the convexity-like arguments which we follow in the present paper, consult Artstein [4], [5]. References on weak convergence of measures are Billingsley [11] and Bertsekas and Shreve [10].

We work with measurable maps

$$f(\cdot) : [0,1] \to R^d, \tag{2}$$

with R^d the d-dimensional Euclidean space, and where $[0,1]$ is endowed with the Lebesgue measure (see Remark 6.3). We consider a function of the form (2) as a particular case of a mapping of the form

$$\mu(\cdot) : [0,1] \to \mathcal{P}, \tag{3}$$

where $\mathcal{P} = \mathcal{P}(R^d)$ is the space of probability measures on R^d endowed with the weak convergence of measures. This convergence is generated by a topology which is separable and metrizable by a metric which makes $\mathcal{P} = \mathcal{P}(R^d)$ a complete metric space. We demand that the measure-valued mappings be measurable with respect to the metric. Such maps are called Young measures. A function of the form (2) is associated with the Young measure which assigns to each t the Dirac measure $\delta_{\{f(t)\}}$. When $\mu(\cdot)$ is a Young measure then the value which the measure $\mu(t)$ assigns to the Borel set $B \subset R^d$ is denoted by $\mu(t)(B)$. Measurability of $\mu(\cdot)$ is equivalent to the measurability of $\mu(\cdot)(B)$ for every

Borel set B. When a function on R^d is integrated with respect to $\mu(t)$ we use the notation $\mu(t)(dx)$.

An alternative representation of a Young measure $\mu(\cdot)$ is its direct integral, namely, the measure $\boldsymbol{\mu}$ defined on $[0,1] \times R^d$ (the latter is considered with its Borel field), and determined by

$$\boldsymbol{\mu}(E \times B) = \int_E \mu(t)(B)dt. \tag{4}$$

Since $[0,1]$ has Lebesgue measure one, it follows that $\boldsymbol{\mu}$ is a probability measure on $[0,1] \times R^d$. We alternate freely between the notations $\boldsymbol{\mu}$ and $\mu(\cdot)$.

The space of Young measures is denoted by $\mathcal{Y}$. It is endowed with the weak convergence of measures, namely the weak convergence of the representations $\boldsymbol{\mu}$ on $[0,1] \times R^d$. This is the topology in which the maps (3) form a completion of the functions (2). The convergence on $\mathcal{Y}$ is also called the narrow convergence (to distinguish it from weak convergence notions for functions).

Next we define the spaces $\mathcal{Y}_p$. Recall that for each such $p \geq 1$ the space L_p is the space of functions of the form (2) which are p-integrable. We denote by $\|f\|_p$ the L_p-norm of $f(\cdot)$. The extension to Young measures is given by

$$\|\boldsymbol{\mu}\|_p = \left(\int_0^1 \int_{R^d} |x|_p^p \mu(t)(dx)dt \right)^{1/p}, \tag{5}$$

where $|x|_p$ is the l_p-norm of the vector $x = (x_1, \cdots, x_d)$ in R^d (however, when it is clear from the context which l_p space is under consideration, we suppress the subscript from the norm $|x|$ of the R^d-vectors). We refer to $\|\boldsymbol{\mu}\|_p$ as the $\mathcal{Y}_p$-norm or as the $\| \ \|_p$-norm of $\boldsymbol{\mu}$. On L_p-Young measures see Kružik and Roubiček [15] and Piccinini and Valadier [23]. It is clear that the L_p-norm and the $\mathcal{Y}_p$-norm coincide for functions.

Narrow convergence in $\mathcal{Y}_p$ plays, to some extent, the role of weak convergence in L_p, although a sequence in L_p may converge weakly in L_p and narrowly in $\mathcal{Y}_p$, with distinct limits. For a strong-type convergence we adopt the following.

Definition 2.3. The sequence $\boldsymbol{\mu}_k$ in $\mathcal{Y}_p$ converges in the p-strong-narrow sense to $\boldsymbol{\mu}_0 \in \mathcal{Y}_p$, if both, $\boldsymbol{\mu}_k$ converges narrowly to $\boldsymbol{\mu}_0$ and $\|\boldsymbol{\mu}_k\|_p \to \|\boldsymbol{\mu}_0\|_p$ as $k \to \infty$. This convergence is derived from a metric. We choose a specific metric to describe the p-strong-narrow convergence, as follows. The metric is denoted by $sn_p(\cdot,\cdot)$ and $sn_p(\boldsymbol{\mu},\boldsymbol{\nu})$ is defined to be the sum of $(\|\boldsymbol{\mu}\|_p - \|\boldsymbol{\nu}\|_p)$ and $n(\boldsymbol{\mu},\boldsymbol{\nu})$ where $n(\cdot,\cdot)$ is the metric which makes $\mathcal{Y}$ with the weak convergence a complete separable metric space.

The definition is inspired by a property which holds in uniformly convex spaces, e.g., in L_p spaces for $1 < p < \infty$; namely, that $f_k \to f_0$ weakly in L_p and $\|f_k\|_p \to \|f_0\|_p$ imply that $\|f_k - f_0\|_p \to 0$. See Dunford and Schwartz [13, II.4.28]. (In $\mathcal{Y}_p$, however, the difference $\boldsymbol{\mu}_k - \boldsymbol{\mu}_0$ is not defined, or when defined in an abstract manner, it may not belong to $\mathcal{Y}_p$, see Roubiček [26].) It is well known that this property is not valid in L_1. The definition, however, of p-strong-narrow convergence applies also to the case $p = 1$, and we use it in the $\mathcal{Y}_1$ case as well. In particular, when the p-strong-narrow convergence in $\mathcal{Y}_p$ is restricted to L_p, it coincides with the norm convergence in L_p, see Artstein [5, Proposition 2.6]. Thus, for each p (including $p = 1$), the normed space L_p is homeomorphically embedded in $\mathcal{Y}_p$ when the latter is endowed with the sn_p-metric. The spaces $\mathcal{Y}_p$ are not equipped with a natural affine structure which extends continuously the linear structure of the L_p spaces. In particular, the affine structure derived from the space of measures, namely, identifying a Young measure with a measure on the product space (see (4)) and considering the standard linear structure stemming from duality, does not work. Indeed, the average of two functions in L_p when considered as Young measures in the space of measures, does not coincide with the average of the same functions when considered as functions in L_p. We follow [5] and offer a convexity notion based on limit arguments as follows.

Definition 2.4. A set $C \subset \mathcal{Y}_p$ is K_p-convex if it is the K_p-limit of a sequence D_k of convex sets in L_p, where by K_p-limit we mean that

(i) every $\boldsymbol{\mu}_0 \in C$ is the sn_p-limit of a sequence $f_k \in D_k$, and
(ii) whenever $f_{k_i} \in D_{k_i}$ for a subsequence which converges in the sn_p-metric, say to $\boldsymbol{\mu}_0$, then $\boldsymbol{\mu}_0 \in C_0$.

The definition of K_p-limit amounts to the definition of the Kuratowski limit of sets in the space $\mathcal{Y}_p$. See Kuratowski [16], Klein and Thompson [14]. Another notion of convexity, namely the M-convexity, is mentioned in Remark 6.2.

In addition to the lack of an inherent convexity notion, the spaces $\mathcal{Y}$ and $\mathcal{Y}_p$ lack a natural linear structure compatible with the completion. We refer to this problem in Section 5 and display in Theorem 5.3 the best approximation facet of the main result.

The minimal norm notion mentioned in the statement of the main result refers to the norm given in (5). In particular, an element $\boldsymbol{\mu} \in C$ is of minimal $\| \, \|_p$-norm in C if

$$\|\boldsymbol{\mu}\|_p = \min\{\|\boldsymbol{\nu}\|_p : \boldsymbol{\nu} \in C\}. \tag{6}$$

The set of minimal $\| \, \|_p$-norms is denoted by $\beta_p(C)$.

It is clear from (6) that when $C \subset L_p$, the minimal norm reduces to the classical notion of minimal norm in L_p spaces. On norm minimizers and best approximations in function spaces see, e.g., Singer [27]. In $\mathcal{Y}_p$, however, the problem is not of metric projection, this since (6) is not induced by a metric compatible with respect to the p-strong-narrow convergence.

It is proved in [5] that if C is K_p-convex and $p > 1$, the set $\beta_p(C)$ contains exactly one point. The set $\beta_1(C)$ may be empty, and may contain more than one point. Following Landers and Rogge, Theorem 2.2 singles out, under the displayed conditions, the aforementioned stable point in $\beta_1(C)$.

3. Compactness and convexity in $\mathcal{Y}_p$

In this section we display some observations which are needed for the proof of the main result. They are concerned with compactness and K_p-convexity arising when applying different norms in the relaxed framework.

We start with some compactness issues. Recall the notion of uniform p-integrability of a family of Young measures, namely, a family F of Young measures in $\mathcal{Y}_p$ (for $1 \le p < \infty$)

is called uniformly p-integrable if for every $\varepsilon > 0$ there exists an $N > 0$ such that

$$\int_0^1 \int_{|x|\geq N} |x|^p \mu(t)(dx)dt \leq \varepsilon \tag{7}$$

for every $\boldsymbol{\mu} \in F$. It is shown in [4, Proposition 6.9] that for $1 \leq p < \infty$, a family $F \subset \mathcal{Y}_p$ is relatively compact (namely, its closure is compact) in the sn_p-topology if and only if it is uniformly p-integrable. The following result is a de la Vallée-Poussin type estimate (see e.g., Natanson [20, Chapter IV]) for Young measures.

Lemma 3.1. *Let $\gamma(t,x)$ be a continuous function from $[0,1]\times R^d$ to R, which grows faster than the p-power in $|x|$ (namely, $\gamma(t,x)|x|^{-p}$ tends to $+\infty$ as $|x| \to \infty$, uniformly in t). Let $F \subset \mathcal{Y}_p$ be such that*

$$\sup_{\boldsymbol{\nu}\in F} \int_{[0,1]\times R^d} \gamma(t,x)d\boldsymbol{\nu} < \infty. \tag{8}$$

Then F is relatively compact in the sn_p-topology.

Proof. Suppose that the criterion (7) does not hold, i.e., suppose that for some $\varepsilon_0 > 0$

$$\int_0^1 \int_{|x|\geq N} |x|^p \mu(t)(dx)dt \geq \varepsilon_0 \tag{9}$$

for arbitrarily large N and some $\boldsymbol{\mu} \in F$. Then

$$\int_0^1 \int_{|x|\geq N} \gamma(t,x)\mu(t)(dx)dt \geq \alpha(N)\varepsilon_0 \tag{10}$$

with $\alpha(N)$ a lower bound on $\gamma(t,x)|x|^{-p}$ for $|x| \geq N$. Since $\alpha(N) \to \infty$ as $N \to \infty$, we get a contradiction to (8). This verifies the claim of the lemma.

Corollary 3.2. *Let F be a bounded set in $\mathcal{Y}_q$. Then it is relatively compact in the sn_p-topology of $\mathcal{Y}_p$ for $1 \leq p < q$.*

Proof. Use the previous lemma with $\gamma(t,x) = |x|^q$.

The following result is trivial in L_p spaces, but needs a proof in the relaxed framework.

Corollary 3.3. *Suppose that $\boldsymbol{\mu}_j \to \boldsymbol{\mu}_0$ in the sn_q-metric, then $\boldsymbol{\mu}_j \to \boldsymbol{\mu}_0$ in the sn_p-metric for all $1 \leq p \leq q$.*

Proof. By Corollary 3.2 the sequence is relatively compact in $\mathcal{Y}_p$, hence convergent subsequences exist. The narrow convergence to $\boldsymbol{\mu}_0$ implies that it is the unique limit.

We need the following truncation notion.

Notation 3.4. For $N > 0$ and a vector $x = (x_1, \ldots, x_d)$ define the vector $N(x)$ by $N(x)_i = x_i$ if $|x_i| \leq N$, $N(x)_i = N$ if $x_i > N$ and $N(x)_i = -N$ if $x_i < -N$.

Lemma 3.5. *Let $\gamma(t,x) : [0,1] \times R^d \to [0,\infty)$ be continuous, and let $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ be two Young measures such that $\int \gamma(t,x)d\boldsymbol{\nu}_k$ is finite for $k = 1,2$. Let $f_{1,j}$ and $f_{2,j}$ be sequences converging narrowly, as $j \to \infty$, to $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ respectively, and assume that*

$f_{3,j} = \frac{1}{2}(f_{1,j} + f_{2,j})$ *converges narrowly, say to* $\boldsymbol{\nu}_3$*. Then there exist sequences* $h_{1,j}$ *and* $h_{2,j}$*, which converge narrowly to* $\boldsymbol{\nu}_1$ *and* $\boldsymbol{\nu}_2$ *respectively, and such that* $h_{3,j} = \frac{1}{2}(h_{1,j} + h_{2,j})$ *converges narrowly to* $\boldsymbol{\nu}_3$*, and furthermore,*

$$\int_0^1 \gamma(t, h_{k,j}(t))dt \to \int_{[0,1]\times R^d} \gamma(t,x)d\boldsymbol{\nu}_k \tag{11}$$

for $k = 1, 2, 3$.

Proof. Using Notation 3.4, the sequences $N(f_{1,j})$ and $N(f_{2,j})$ converge narrowly to $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ when both, j and N, tend to ∞. Since, when considered as Young measures, in particular as measures on the product space (see (4)), the sequences are tight (for tightness of collections of measures see, e.g., Billingsley [11]), it follows that $N(f_{3,j})$ converges narrowly to $\boldsymbol{\nu}_3$ as j and N tend to ∞. It is also easy to see that

$$\lim_{N\to\infty} \lim_{j\to\infty} \int_0^1 \gamma(t, N(f_{k,j}(t)))dt = \int_{[0,1]\times R^d} \gamma(t,x)d\boldsymbol{\nu}_k \tag{12}$$

for $k = 1, 2, 3$. From (12) it follows that a relation $N(j)$ can be determined such that (11) holds for the diagonal sequences $h_{k,j} = N(f_{k,N(j)})$ for $k = 1, 2, 3$.

Corollary 3.6. *Let* C *be a* K_p*-convex set in* $\mathcal{Y}_p$*. Let* $\boldsymbol{\nu}_1$ *and* $\boldsymbol{\nu}_1$ *be in* C *such that both are in* $\mathcal{Y}_q$ *for some* $q > p$*. let* $f_{1,j}$ *and* $f_{2,j}$ *be sequences converging narrowly, as* $j \to \infty$*, to* $\boldsymbol{\nu}_1$ *and* $\boldsymbol{\nu}_2$ *respectively, and assume that* $f_{3,j} = \frac{1}{2}(f_{1,j} + f_{2,j})$ *converges narrowly, say to* $\boldsymbol{\nu}_3$*. Then there exist sequences* $h_{1,j}$ *and* $h_{2,j}$ *in* $\mathcal{Y}_q$*, which converge in the* sn_q*-metric to* $\boldsymbol{\nu}_1$ *and* $\boldsymbol{\nu}_2$ *respectively, and such that* $h_{3,j} = \frac{1}{2}(h_{1,j} + h_{2,j})$ *converges in the* sn_q*-metric to* $\boldsymbol{\nu}_3$*.*

Proof. Apply Lemma 3.5 with $\gamma(t,x) = |x|^q$.

As mentioned already, the definition of convexity in the relaxed setting depends on the choice of the norm. The following result shows that under a compactness condition, the convexity notions for different norms coincide (in view of Corollary 3.2, the compactness in the following result may be replaced in some cases by boundedness).

Proposition 3.7. *Let* C *be a* K_p*-convex set in* $\mathcal{Y}_p$*. Suppose that for some* $1 \le q < \infty$*, the family* C *is uniformly* q*-integrable. Then* C *is* K_q*-convex in* $\mathcal{Y}_q$*.*

Proof. Corollary 3.3 covers the case where $p \ge q$. Assume that $p < q$. Let D_k be the sequence of convex subsets of L_p, establishing, according to Definition 2.4, the K_p-convexity of C. Let $\varepsilon_j \to 0$. Since C is relatively compact in $\mathcal{Y}_q$, it follows that for every $\varepsilon_j > 0$ there exists a finite set, say $\{\boldsymbol{\mu}_{j,1}, \dots, \boldsymbol{\mu}_{j,r_j}\}$ which is an ε_j net in C in the sn_q-metric (namely, it ε_j-approximates every element in C). For a fixed j, let $f_{k,j,1}, \dots, f_{k,j,r_j}$ be r_j sequences such that $f_{k,j,l} \in D_k$ and $f_{k,j,l}$ converge as $k \to \infty$, in the sn_p-metric, to $\boldsymbol{\mu}_{j,l}$.

Let N_j be the estimate guaranteed when the test (7) of uniform p-integrability is applied when $\varepsilon = \frac{\varepsilon_j}{r_j}$, and when $F = \{f_{k,j,l} : l = 1, \dots, r_j,\ k = 1, 2, \dots\}$, namely a finite number r_j of convergent sequences. Using Notation 3.4 we define

$$g_{k,j,l} = N_j(f_{k,j,l}). \tag{13}$$

Let $E_{k,j}$ be the convex hull of $g_{k,j,l}$ for $l = 1, \ldots, r_j$. We construct a diagonal sequence $E_{k(j),j}$ as follows. Let $k(j)$ be such that $sn_q(\boldsymbol{\mu}_{j,l}, g_{k(j),j,l}) < 2\varepsilon_j$ and $\|f_{k(j),j,l} - g_{k(j),j,l}\|_p < 2\varepsilon_j$ for all $l = 1, \ldots, r_j$. An index $k(j)$ validating the latter estimates exists due to the relative compactness of C in $\mathcal{Y}_q$ and the assumed convergence in $\mathcal{Y}_p$. We now establish the K_q-convexity of C by the K_q-convergence of $E_{k(j),j}$ to C. Indeed, let $\boldsymbol{\mu} \in C$. Then $\boldsymbol{\mu}$ is a limit of elements $\boldsymbol{\mu}_{j,l(j)}$ for an appropriate choice of $l(j)$. Then $g_{k(j),j,l(j)}$ converges in the sn_q-metric to $\boldsymbol{\mu}$, as condition (i) in Definition 2.4 requires. To check condition (ii) of the definition, we should consider a convergent subsequence, say also indexed by j. Let $h_j \in E_{k(j),j}$ converge in the sn_q-metric, say to $\boldsymbol{\mu}_0$. Then each h_j is a convex combination of the elements $g_{k(j),j,1}, \ldots, g_{k(j),j,r_j}$, and in particular of at most r_j elements. The choice of N_j as guaranteeing an $\frac{\varepsilon_j}{r_j}$ estimated in (7), implies that the corresponding convex combination with the same weights of $f_{k(j),j,1}, \ldots, f_{k(j),j,r_j}$, will have the same sn_p-limit. By the K_p convexity, this limit point is in C. This completes the proof.

4. Proof of Theorem 2.2

The proof follows on the lines of Landers and Rogge, [17, Theorem 2], but with the required modifications. We shall not make specific comparisons, but note here that a prime tool used in [17, Theorem 2] is not available in our setting, namely, that convex sets closed in norm are also closed with respect to weak convergence. We overcome this difficulty by establishing enough compactness with respect to the strong-narrow topology. Another difficulty is that averages of Young measures are not defined. The way to overcome this obstacle is to employ averages in the limit process that determines the convexity.

For each $\boldsymbol{\nu} \in \beta_1(C)$ denote

$$\phi_{\boldsymbol{\nu}}(p) = \|\boldsymbol{\nu}\|_p^p, \tag{14}$$

namely, the p-norm of the Young measure $\nu(\cdot)$ raised to the p-power, see (5). It follows from the assumptions that there exists an element $\boldsymbol{\nu} \in \beta_1(C)$ such that $\phi_{\boldsymbol{\nu}}(p) < \infty$ for p in a non-degenerate interval $[1, p(\boldsymbol{\nu})]$.

Claim 4.1. *For any two elements $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in $\beta_1(C)$*

$$\phi_{\boldsymbol{\nu}}(1) = \phi_{\boldsymbol{\mu}}(1). \tag{15}$$

Proof. Trivial, since all the elements in $\beta_1(C)$ share the same $\| \|_1$-norm.

Claim 4.2. *For $\boldsymbol{\nu}$ in $\beta_1(C)$ either $\phi_{\boldsymbol{\nu}}(p) = \infty$ for all $p > 1$, or the function $\phi_{\boldsymbol{\nu}}(\cdot)$ is differentiable as a function of p at $p = 1$, and the derivative is*

$$\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1) = \int_0^1 \int_{R^d} \left(\sum_{i=1}^{d} |x_i| \ln |x_i| \right) \nu(t)(dx)dt. \tag{16}$$

Proof. We spell out the expression for $\phi_{\boldsymbol{\nu}}$ as an integral over $[0, 1] \times R^d$, as follows (compare with (5)).

$$\phi_{\boldsymbol{\nu}}(p) = \int_0^1 \int_{R^d} |x|_p^p \, d\boldsymbol{\nu}. \tag{17}$$

The integrand $|x|_p^p = |x_1|^p + \cdots |x_d|^p$ is clearly differentiable. Its derivative with respect to p is $\sum_{i=1}^d |x_i|^p \ln |x_i|^p$. This derivative is uniformly bounded from below (by the constant

$-\frac{d}{e}$). If the integral of this derivative diverges, then $\phi_{\boldsymbol{\nu}}(p) = \infty$ for $p > 1$. Otherwise the expression is $\boldsymbol{\nu}$-integrable for all p in a half open interval $[1, p(\boldsymbol{\nu}))$. In the latter case we can change the order of integration and differentiation and get the desired expression (16). This completes the proof.

Denote the integrand in (16) by $\Phi(x)$, namely

$$\Phi(x) = \sum_{i=1}^{d} |x_i| \ln |x_i|. \tag{18}$$

Claim 4.3. *There exists a Young measure in $\beta_1(C)$ which minimizes the expression $\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1)$.*

Proof. Since we assumed that $\beta_1(C)$ contains elements in $\mathcal{Y}_{1+}$ it follows from the preceding claim that $\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1) < \infty$ for some $\boldsymbol{\nu} \in \beta_1(C)$. Let $\boldsymbol{\nu}_j$ be a minimizing sequence of $\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1)$ in $\beta_1(C)$. In particular, the condition (8) holds with $F = \{\boldsymbol{\nu}_j\}$ and $\gamma(t, x) = \Phi(x)$, with respect to $p = 1$. Hence there exists a subsequence which converges in the sn_1-topology, with a limit, say $\boldsymbol{\nu}_0$. The closedness of $\beta_1(C)$ implies that $\boldsymbol{\nu}_0 \in \beta_1(C)$. Since $\Phi(x)$ is a convex function for $|x|$ large enough, it follows that $\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1)$, which is given by (16), is lower semi-continuous in the sn_1-topology. The compactness together with the lower semi-continuity imply that $\boldsymbol{\nu}_0$ is a minimizer of the given expression.

Claim 4.4. *The minimizer in $\beta_1(C)$ of the expression $\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1)$ is unique.*

Proof. Suppose that there are two distinct minimizers, say $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$. Since C is K_1-convex, it follows that there exist two sequences $f_{1,j}$ and $f_{2,j}$ which sn_1-converge to $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$, respectively, and such that any sn_1-limit point of the averages

$$f_{3,j} = \frac{1}{2}(f_{1,j} + f_{2,j}) \tag{19}$$

is in $\beta_1(C)$. The compactness criterion given in (7) implies that the sequence $f_{3,j}$ is relatively compact in the sn_1-topology, hence a converging subsequence exists. We assume that it is the sequence itself, and denote the limit by $\boldsymbol{\nu}_3$. In view of Lemma 3.5 we may assume that

$$\lim_{j\to\infty} \int_0^1 \Phi(f_{k,j}(t))dt = \frac{d}{dp}\phi_{\boldsymbol{\nu}_k}(1) \tag{20}$$

for $k = 1, 2, 3$. We plan to show that the strict convexity of Φ is uniform enough to get, in the limit, a contradiction to the assumption that $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ are distinct minimizers. But we should proceed with care since Φ is convex only on a pointed orthant of R^d.

Since the $\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1) < \infty$ both $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$, it follows that both belong to $\mathcal{Y}_q$ for some $q > 1$. Therefore, by Corollary 3.6, we may assume that the strong-narrow convergence as $j \to \infty$ of the sequences $f_{k,j}(\cdot)$ for $k = 1, 2, 3$ is actually in sn_q, and in particular, (by [4, Proposition 6.9]) the three sequences are uniformly q-integrable.

Denote

$$Q = \{(x, y) : x_i y_i < 0 \ \text{ for some coordinate } \ i\} \tag{21}$$

and let

$$T_j = \{t : (f_{1,j}(t), f_{2,j}(t)) \in Q\}. \tag{22}$$

For $t \in T_j$ let $I_j(t)$ be the set of indices i in $\{1, \ldots, d\}$ such that the i-th coordinates of $f_{1,j}(t)$ and $f_{2,j}(t)$ have opposite signs, and then let

$$g_j(t) = \max_{i \in I_j(t)} \min\{|f_{1,j}(t)_i|, |f_{2,j}(t)_i|\}. \tag{23}$$

Define $g_j(t) = 0$ if $t \notin T_j$. A simple geometrical observation implies that

$$|f_{3,j}(t)| \le \frac{1}{2}(|f_{1,j}(t)| + |f_{2,j}(t)|) - g_j(t) \tag{24}$$

(equality holds when $t \notin T_j$). Since the L_1-norms of the three sequences $f_{k,j},\ k = 1, 2, 3$, converge to the same limit, it follows from (24) that

$$\int_0^1 g_j(t)dt \to 0 \quad \text{as} \quad j \to \infty. \tag{25}$$

The uniform q-integrability of $f_{k,j}$ for some $q > 1$ implies that g_j are uniformly q-integrable for the same q.

The next step is to show that for some $\delta > 0$

$$\int_0^1 \Phi(f_{3,j}(t))dt \le \frac{1}{2}\left(\int_0^1 \Phi(f_{1,j}(t))dt + \int_0^1 \Phi(f_{2,j}(t))dt\right) - \delta. \tag{26}$$

Once (26) is verified, it implies that the limit (20) contradicts the assumption that $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ are distinct minimizers, and the proof of the claim would be complete.

To verify (26) we employ an estimate for the convexity gap, as follows. A simple comparison with quadratic functions shows that for any convex real valued function, say $\gamma(\cdot)$, one can estimate $\frac{1}{2}(\gamma(s) + \gamma(r)) - \gamma(\frac{1}{2}(s + r))$ by $\frac{1}{8}|s - r|^2\eta$ when η is the infimum of the second derivative of γ on the interval $[r, s]$ (say $r < s$). Applying this estimate to the function $\gamma(s) = |s| \ln |s|$ (notice that this function is convex on the half line only) one gets that for real numbers r and s which have the same sign, namely $rs \ge 0$, and say $|r| < |s|$, then

$$\frac{|r + s|}{2} \ln(\frac{|r + s|}{2}) \le \frac{1}{2}(|r| \ln |r| + |s| \ln |s|) - \frac{1}{8}(s - r)^2|s|^{-1}. \tag{27}$$

In the case where $rs < 0$, and, say, $|r| \le |s|$, the function is not convex anymore on $[r, s]$ and a correction argument is needed. In this case one can use the convexity on a half line and the fact that $0 \ln 0$ is 0 to get (say $r < 0 < s$) the inequality $\frac{1}{2}(r + s) \ln \frac{1}{2}(r + s) \le |r| \ln |r| + \frac{1}{2}(|r| + s) \ln \frac{1}{2}(|r| + s)$. Applying now the preceding estimate for the convexity gap on $[|r|, s]$, employing the fact that the function is symmetric around 0 and manipulating a bit the terms, one gets the expression (valid also when $s < 0 < r$)

$$\frac{|r + s|}{2} \ln(\frac{|r + s|}{2}) \le \frac{1}{2}(|r| \ln |r| + |s| \ln |s|) - \frac{1}{8}(s - r)^2|s|^{-1} + \frac{1}{2}|r| - |r| \ln |r|. \tag{28}$$

Consequently, if we denote by $h_j(t)$ the largest absolute value of a coordinate of either $f_{1,j}(t)$ or $f_{2,j}(t)$, we get the estimate

$$\Phi(f_{3,j}(t)) \le \frac{1}{2}(\Phi(f_{1,j}(t)) + \Phi(f_{2,j}(t))) - \frac{1}{8}|f_{1,j}(t) - f_{2,j}(t)|_2^2 h_j(t)^{-1} + c_j(t) \tag{29}$$

(if $h_j(t) = 0$ then both sides equal 0) where the correction term $c_j(t)$ is given by

$$c_j(t) = \frac{1}{2}|g_j(t)| - dg_j(t) \ln g_j(t) \tag{30}$$

(in the above displayed expression $|\cdot|_2$ is the l_2-norm, d is the dimension of the Euclidean space and $g_j(t)$ is as defined in (23); in particular, when $t \notin T_j$ we have $g_j(t) = 0$, hence $c_j(t) = 0$). Now we integrate the expressions (29) over $[0, 1]$ and check the limit as $j \to \infty$.

The uniform q-integrability of g_j for some $q > 1$ implies that that in addition to (25) the convergence $\int_0^1 dg_j(t) \ln g_j(t)dt \to 0$ as $j \to \infty$ holds as well. Hence $\int_0^1 c_j(t)dt \to 0$ as $j \to \infty$. Since $f_{1,j}$ and $f_{2,j}$ converge to distinct Young measures, it follows that the integrals $\int_0^1 |f_{1,j}(t) - f_{2,j}(t)|_2^2 dt$ are bounded away from 0 as $j \to \infty$. The uniform 1-integrability of the sequences, which implies that h_j are uniformly integrable, implies that $\int_0^1 \frac{1}{4}|f_{1,j}(t) - f_{2,j}(t)|_2^2 h_j(t)^{-1}dt$ is bounded away from 0 as $j \to \infty$. This verifies (26), and as already noted, this completes the proof of Claim 4.4.

Notice that we have proved in the previous claims that there exists a unique element in $\beta_1(C)$, which we denote by $\boldsymbol{\mu}_1$, such that

$$\frac{d}{dp}\phi_{\boldsymbol{\mu}_1}(1) < \frac{d}{dp}\phi_{\boldsymbol{\nu}}(1) \tag{31}$$

for all $\boldsymbol{\nu} \in \beta_1(C)$ (the right hand side of (31) could be equal to $+\infty$). This inequality together with Claim 4.1 verify (i) of Definition 2.1 and the minimization of (1) in Theorem 2.2. It remains to verify property (ii) in Definition 2.1.

Claim 4.5. *For $p > 1$ close to 1, there exists a unique minimizer of the $\mathcal{Y}_p$-norm in C.*

Proof. The assumption $\beta_1(C) \bigcap \mathcal{Y}_{1+} \neq \emptyset$ implies that for $p > 1$ close to 1 the intersection of $\mathcal{Y}_p$ and C is not empty. Let $\boldsymbol{\nu}_j$ be a sequence in C minimizing the $\mathcal{Y}_p$-norm. Applying Lemma 3.1 with $\gamma(t, x) = |x|_p^p$ implies that the sequence $\boldsymbol{\nu}_j$ is relatively compact in the sn_1-topology. Hence, the sequence has a subsequence, say it is the sequence itself, which converges, say to $\boldsymbol{\mu}_p$, in C. The semi-continuity of the $\mathcal{Y}_p$-norm with respect to sn_1-convergence in C implies that $\boldsymbol{\mu}_p$ is a minimizer of the $\mathcal{Y}_p$-norm in C. Applying Lemma 3.5 with $\gamma(t, x) = |x|_p^p$, and using the strict convexity of the latter function, implies that $\boldsymbol{\mu}_p$ is a unique minimizer.

Claim 4.6.

$$\limsup_{p \to 1+} \int_{[0,1]\times R^d} \Phi(x) d\boldsymbol{\mu}_p \leq \int_{[0,1]\times R^d} \Phi(x) d\boldsymbol{\mu}_1. \tag{32}$$

Proof. Recall the definition of $\Phi(x)$ from (18). By the mean value theorem,

$$\Phi(x) \leq \sum_{i=1}^{d} \frac{|x_i|^p - |x_i|}{p - 1}. \tag{33}$$

Integrating with respect to $\boldsymbol{\mu}_p$ yields

$$\int_{[0,1]\times R^d} \Phi(x) d\boldsymbol{\mu}_p \leq \frac{1}{p-1} \left(\int_{[0,1]\times R^d} |x|_p^p d\boldsymbol{\mu}_p - \int_{[0,1]\times R^d} |x|_1 d\boldsymbol{\mu}_p \right). \tag{34}$$

Using the fact that $\boldsymbol{\mu}_p$ is the minimizer of the $\| \ \|_p$-norm, and $\boldsymbol{\mu}_1$ is the minimizer of the $\| \ \|_1$-norm, we can replace the measures with respect to which we integrate in the right hand side of (34) and obtain

$$\int_{[0,1]\times R^d} \Phi(x) d\boldsymbol{\mu}_p \leq \frac{1}{p-1} \left(\int_{[0,1]\times R^d} |x|_p^p d\boldsymbol{\mu}_1 - \int_{[0,1]\times R^d} |x|_1 d\boldsymbol{\mu}_1 \right). \tag{35}$$

With $p \to 1+$, the relation (35) verifies the claim.

Claim 4.7. *Any sequence $\boldsymbol{\mu}_{p_n}$ for $p_n \to 1+$ is relatively compact in the sn_1-topology. Any sn_1-limit point of such a sequence is in $\beta_1(C)$.*

Proof. The first part follows from Claim 4.6 together with Lemma 3.1 when $\gamma(t,x) = \Phi(x)$. The second part follows from the continuity of the $\mathcal{Y}_1$-norm.

Claim 4.8. *Any sn_1-limit of a sequence $\boldsymbol{\mu}_{p_n}$ for $p_n \to 1+$, coincides with $\boldsymbol{\mu}_1$.*

Proof. By Claim 4.7 any sn_1-limit, say $\boldsymbol{\nu}$, of a sequence as described, is in $\beta_1(C)$. By Claim 4.6 the derivative $\frac{d}{dp}\phi_{\boldsymbol{\nu}}(1)$ is less than or equal to the derivative $\frac{d}{dp}\phi_{\boldsymbol{\mu}_1}(1)$. The uniqueness established in Claim 4.4 implies that $\boldsymbol{\nu} = \boldsymbol{\mu}_1$.

With the last claim the proof of the Theorem 2.2 is complete. Indeed, the sn_1-compactness of $\boldsymbol{\mu}_p$ as $p \to 1+$, together with the fact that any subsequence converges in the sn_1-topology to $\boldsymbol{\mu}_1$, imply that (ii) of Definition 2.1 is satisfied; since (i) of Definition 2.1 and the desired minimization of (1) were established with Claim 4.4, the proof is complete.

5. Natural best approximations

In this section we show how the result on the natural minimizer within the Young measures framework can be used to define a natural best approximation, following the lines of Landers and Rogge [17]. Some preparations need, however, to be carried out.

In addition to the lack of an inherent convexity notion, the spaces $\mathcal{Y}$ and $\mathcal{Y}_p$ lack a natural linear structure compatible with the completion, as was explained before introducing Definition 2.4. A partial linear structure compatible with the L_p-operations can be identified, namely when an ordinary function interacts with a Young measure, and when a Young measure is multiplied by a scalar, as follows.

Let $\boldsymbol{\mu} \in \mathcal{Y}$, let α be a scalar and let $f : [0,1] \to R^d$ be measurable. The Young measure $\alpha\boldsymbol{\mu}$ is defined by

$$(\alpha\mu)(t)(B) = \mu(t)(\{x : \alpha x \in B\}), \tag{36}$$

and the Young measure $\boldsymbol{\mu} + f$ is defined by

$$(\mu + f)(t)(B) = \mu(t)(\{x : (x + f(t)) \in B\}) \tag{37}$$

(notice that the latter is not an operation on the entire product space $\mathcal{Y} \times \mathcal{Y}$). Similar operations were used in the literature; see Roubiček [25].

We call $\boldsymbol{\mu} + f$ the translation of $\boldsymbol{\mu}$ by f; obviously, $\boldsymbol{\mu} - f = \boldsymbol{\mu} + (-f)$. The expressions given in (36) and (37) are continuous with respect to convergence of scalars, the strong convergence in L_p and the sn_p-convergence in $\mathcal{Y}_p$. With the aid of (37) we can formulate the best approximation problem as follows.

Definition 5.1. Let $C \subset \mathcal{Y}_p$ and $g \in L_p$. An element $\boldsymbol{\mu} \in C$ is a $\| \, \|_p$-best approximation of g in C, if

$$\|\boldsymbol{\mu} - g\|_p = \min\{\|\boldsymbol{\nu} - g\|_p : \boldsymbol{\nu} \in C\}. \tag{38}$$

As it is evident from (38), when $C \subset L_p$ the best approximation reduces to the classical notion of the best approximation in L_p spaces. Note, however, that we do not define the best approximation problem for a Young measure and a set of Young measures. Indeed, the difference $\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2$ is not defined in our context (in other available definitions, see Roubiček [26], the difference $\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2$ may not belong to $\mathcal{Y}$).

The translation defined in (37) enables reduction of the best approximation problem to a minimal norm problem. For completeness we state the relevant definition and the result extending the main result in Landers and Rogge [17] to Young measures, as follows.

Definition 5.2. Let $f \in L_{1+}$ and $C \subseteq \mathcal{Y}_1$ be such that for every $p > 1$ close enough to 1, there exists a unique $\| \, \|_p$-best approximation of f in C, say $\boldsymbol{\mu}_p(f, C)$. An element $\boldsymbol{\mu}_1 \in C$ which is a $\| \, \|_1$-best approximation of f in C, is called the natural best approximation of f in C if the following two properties hold.

(i) For every $\boldsymbol{\nu} \in C$, $\boldsymbol{\nu} \neq \boldsymbol{\mu}_1$, there exists a $p(\boldsymbol{\nu}) > 1$ such that $\|\boldsymbol{\mu}_1 - f\|_p < \|\boldsymbol{\nu} - f\|_p$ for every $1 < p \leq p(\boldsymbol{\nu})$, and

(ii) $\boldsymbol{\mu}_p(f, C)$ converges in the sn_1-metric, as $p \to 1+$, to $\boldsymbol{\mu}_1$.

Theorem 5.3. *Assume that $f \in L_{1+}$ and let $C \subseteq \mathcal{Y}_1$ be a K_1-convex set. Let $\beta_1(f, C)$ be the collection of measures which minimize the expression $\|\boldsymbol{\mu} - f\|_1$ in C. Suppose that $\beta_1(f, C) \bigcap \mathcal{Y}_{1+}$ is not empty. Then for every $p > 1$ close enough to 1, there exists a unique element in C, say $\boldsymbol{\mu}_p(f, C)$, such that $\|\boldsymbol{\mu}_p(f, C) - f\|_p$ is minimal among the elements in C. Furthermore, the natural best approximation of f in C exists, and it is the minimizer of the functional*

$$\int_0^1 \int_{R^d} \sum_{i=1}^{d} |f(t)_i - x_i| \ln |f(t)_i - x_i| \nu(t)(dx)dt \tag{39}$$

among the Young measures in $\beta_1(f, C)$.

Proof. The continuity of the translation of $\boldsymbol{\mu}$ by f reduces the natural best approximation problem to the problem of natural minimizer addressed in Theorem 2.2.

6. Comments and Examples

Remark 6.1. A particular case of the natural best approximation is the case where C is the set of measurable functions with respect to a given σ-field, say $\mathcal{B}$. The L_1-approximation is then the conditional median of f given $\mathcal{B}$. See Landers and Rogge [18]. A concept of relaxed σ-fields, as an appropriate completion of ordinary σ-fields, was developed in Artstein [4], extending to the relaxed framework the continuity of the conditional expectation as established in, e.g., Alonso and Bramila-Paz [1]. A relaxed conditional median of a function, given a relaxed σ-field, is then a $\mathcal{Y}_1$-best approximation. The application of the present paper to a relaxed σ-field identifies the natural relaxed conditional median of an ordinary function.

Remark 6.2. In addition to the K-convexity given in Definition 2.4, another convexity notion is useful, as follows.

A set $C \subset \mathcal{Y}_p$ is M_p-convex if it is the M_p-limit of a sequence D_k of convex sets in L_p, where for the M_p-limit we require the conditions of K_p-limit, and in addition

(iii) whenever $f_{k_i} \in D_{k_i}$, for a subsequence k_i, is bounded in the L_p-norm and converges in the narrow topology to $\boldsymbol{\mu}_0$, then $\boldsymbol{\mu}_0 \in C_0$.

While the K_p-limit amounts to the definition of the Kuratowski limit of sets, the definition of the M_p-limit is the generalization of the Mosco-convergence in L_p-spaces; for the latter concept see Attouch [6], Mosco [19]. For an analysis of this relaxed convexity notion see [5]. It is clear that a set which is M_p-convex is also K_p-convex; indeed, (iii) in this remark is strictly stronger than (ii) in Definition 2.4. Hence Theorems 2.2 and 5.3 hold for M-convex sets. It is also not difficult to see that the result established in Proposition 3.7, guaranteeing that, in the case of compactness, convexity does not depend on the choice of the norm, holds for M-convexity as well.

Remark 6.3. The choice of $[0, 1]$ as the underlying space in our main result is for convenience of notations only. The results can be extended with no effort to any compact metric space endowed with an atomless measure, and with some effort, along the lines of Balder [9], to abstract measure spaces satisfying some regularity conditions. We leave out the details.

Remark 6.4. Of interest in applications of the best approximation problem, say of f in C, is the dependence of the best approximation $\boldsymbol{\mu}_p(f, C)$ on variations in f and in C (where on the family of sets we consider, say, the Hausdorff distance with respect to the sn_1-metric; for the Hausdorff distance see, e.g. [14]). The problem was addressed in [5] within the Young measures framework for $p > 1$. The natural best approximation for $p = 1$ does not depend continuously on variations in C, even in a finite dimensional space. Indeed, the unique element of minimal l_1-norm in the interval joining the points $(0, 1)$ and $(1 + \eta, 0)$ in R^2 for $\eta > 0$ is the point $(0, 1)$, while the natural minimizer of the l_1-norm of the limit set as $\eta \to 0$ is the point $(\frac{1}{2}, \frac{1}{2})$. The natural best approximation $\boldsymbol{\mu}_1(f, C)$ is a continuous function of f when the latter is endowed with the L_p-norm for $p > 1$, and the range is endowed with the sn_1-topology. This is a direct consequence of the characterization of the natural best approximation as the unique minimizer of the functional (39).

Remark 6.5. The continuous dependence situation is slightly different when along with variations in (f, C) we consider also the convergence, as $p \to 1+$, of $\boldsymbol{\mu}_p(f, C)$ to $\boldsymbol{\mu}_1(f, C)$ (when the latter is the natural best approximation). Indeed, it is not difficult to show that if $p_j \to 1+$ as numbers, $f_j \to f$ in the L_1-norm and $C_j \to C$ in the Hausdorff metric, then $\boldsymbol{\mu}_{p_j}(f_j, C_j)$ converge to $\boldsymbol{\mu}_1(f, C)$, provided the convergence of (f_j, C_j) to (f, C) is faster (in the appropriate sense) than the convergence of p to 1. The arguments are straightforward.

Example 6.6. We conclude with an illustration of the limit processes leading to sets of Young measures, and the notion of the natural best approximation. Such processes appear in variational problems depending on a parameter, see [3],[4]. In this framework, the observation in Remark 6.5 has a natural interpretation. Here and in the next example we restrict the discussion to minimum norm problems.

For a function $h(\cdot) : [0,1] \to R$ denote by $U(h)$ the family of measurable functions, say g, satisfying

$$|g(t)| \geq |h(t)| \text{ and } h(t)g(t) \geq 0. \tag{40}$$

Let $f_1(t) = t$ and $f_2(t) = 1 - t$. For $k = 1, 2$, let $h_{k,j}(t) =$ sign $\cos(jt)f_k(t)$. Let D_j be the convex hull of the union of $U(h_{1,j})$ and $U(h_{2,j})$.

It is easy to verify the K_1-convergence of the family of integrable functions in D_j to the set C in $\mathcal{Y}_1$ given as follows. Each $\boldsymbol{\nu} \in C$ is related to a number α in $[0,1]$ such that $\nu(t)$ is a probability measure supported on the union of the half lines $[\alpha t + (1-\alpha)(1-t), \infty)$ and $(-\infty, -\alpha t - (1-\alpha)(1-t)]$, and the values of $\nu(t)$ on the positive half line is the reflection of the values on the negative half line. In particular, C is K_1-convex. It is easy to compute the natural $\mathcal{Y}_1$-norm minimizer. Indeed, it is the constant probability measure, equally distributed on $\{\frac{1}{2}, -\frac{1}{2}\}$. In addition, this Young measure is the sn_1-limit of the natural L_1-norm minimizers of D_j.

Example 6.7. A variant of the preceding example would alter slightly the results. Replace in Example 6.6 the functions f_1 by $\rho_j f_1$ where $\rho_j \to 1+$. The limit set is exactly as in Example 6.6, and likewise is the natural $\mathcal{Y}_1$-norm minimizer. There is, however, a unique L_1-norm minimizer in D_j, namely $h_{2,j}$. The unique minimizer does not converge to the natural norm minimizer of the relaxed limit. If, however, $p_j \to 1+$ is given and $\rho_j \to 1+$ fast enough, then (see Remark 6.5) the unique L_{p_j}-norm minimizer in D_j would converge in the sn_1-metric to the $\mathcal{Y}_1$-norm minimizer.

Acknowledgements. We thank an anonymous referee for useful references and for enlightening remarks which have led to an improved presentation.

References

[1] A. Alonso, F. Bramila-Paz: L^p-continuity of conditional expectations, J. Math. Anal. and Appl. 221 (1998) 161–176.

[2] Z. Artstein: Chattering variational limits of control systems, Forum Mathematicum 5 (1993) 369–403.

[3] Z. Artstein: Variational limits constrained by measure-valued multifunctions, in: Calculus of Variations and Optimal Control, A. Ioffe, S. Reich, I. Shafrir (eds.), Chapman & Hall, London (1999) 1–23.

[4] Z. Artstein: Compact convergence of σ-fields and relaxed conditional expectation, Prob. Theory and Related Fields 120 (2001) 369–394.

[5] Z. Artstein: Projections on convex sets in the relaxed limit, Set-Valued Analysis 9 (2001) 13–34.

[6] H. Attouch: Variational Convergence of Functions and Operators, Pitman, Boston (1984).

[7] E. J. Balder: A general denseness result in relaxed control theory, Bull. Austral. Math. Soc. 30 (1984) 463–465.

[8] E. J. Balder: Consequences on denseness of Dirac Young measures, J. Math. Anal. Appl. 207 (1997) 536–540.

[9] E. J. Balder: Lectures on Young measure theory and its applications to economics, Rend. Istit. Mat. Univ. Trieste 31, supp. 1 (2000) 1–69.

[10] D. P. Bertsekas, S. E. Shreve: Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York (1978).

[11] P. Billingsley: Convergence of Probability Measures, Wiley, New York (1968).

[12] F. Demengel, R. Temam: Convex functions of a measure and applications, Indiana Univ. Math. J. 33 (1984) 673–709.

[13] N. Dunford, J. T. Schwartz: Linear Operators, Wiley, New York (1964).

[14] E. Klein, A. C. Thompson: Theory of Correspondences, Wiley, New York (1984).

[15] M. Kružik, T. Roubiček: Explicit characterization of L_p-Young measures, J. Math. Anal. and Appl. 198 (1996) 830–843.

[16] K. Kuratowski: Topology, Academic Press, New York (1966).

[17] D. Landers, L. Rogge: Natural choices for L_1 approximations, J. Approx. Theory 33 (1981) 268–280.

[18] D. Landers, L. Rogge: The natural median, Ann. Prob. 9 (1981) 1041–1042, Ann. Prob. 10 (1982) 1092.

[19] U. Mosco: Convergence of convex sets and of solutions of variational inequalities, Adv. in Math. 3 (1969) 510–585.

[20] I. P. Natanson: Theory of Functions of a Real Variable, Vol. I, Frederick Ungar Pub., New York (1955).

[21] P. Pedregal: Parameterized Measures and Variational Principles, Birkhäuser Verlag, Basel (1997).

[22] P. Pedregal: Equilibrium conditions for Young measures, SIAM J. Control Opt. 36 (1998) 797–813.

[23] L. Piccinini, M. Valadier: Uniform integrability and Young measures, J. Math. Anal. Appl. 195 (1995) 428–439.

[24] T. Roubiček: Convex compactification and special extensions of optimization problems, Nonlinear Anal. 16 (1991) 1117–1126.

[25] T. Roubiček: Numerical approximation of relaxed variational problems, J. Convex Anal. 3 (1996) 329–347.

[26] T. Roubiček: Relaxation in Optimization Theory and Variational Calculus, Walter de Gruyter, Berlin (1997).

[27] I. Singer: Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces, Springer-Verlag, Berlin (1970).

[28] L. Tartar: Compensated compactness and applications in partial differential equations, in: Nonlinear Analysis and Mechanics, R. J. Knops (ed.), Pitman, London (1979) 136–212.

[29] M. Valadier: A course on Young measures, Rend. Istit. Mat. Univ. Trieste 26, supp. (1994) 349–394.

[30] J. Warga: Optimal Control of Differential and Functional Equations, Academic Press, New York (1972).

[31] L. C. Young: Lectures on the Calculus of Variations and Optimal Control Theory, Saunders, New York (1969).