Journal of Convex Analysis
Volume 17 (2010), No. 1, 211–227

A Mathematical Programming Approach to Strong Separation in Normed Spaces

Marco A. López
Alicante University, Alicante, Spain

Soon-Yi Wu
National Cheng Kung University, Tainan, Taiwan

Chen Ling
Zhejiang University of Finance and Economics, Hangzhou, China

Liqun Qi
Polytechnic University of Hong Kong, Hong Kong, China

Received: December 28, 2008
Revised manuscript received: February 1, 2009

This paper deals with an infinite-dimensional optimization approach to the strong separation of two bounded sets in a normed space. We present an approximation procedure, called Algorithm (A), such that a semi-infinite optimization problem must be solved at each step. Its global convergence is established under certain natural assumptions, and a stopping criterion is also provided. The particular case of strong separation in the space $L_p(\mathbb{X}, \mathcal{A}, \mu)$ is approached in detail. We also propose Algorithm (B), which is an implementable modification of Algorithm (A) for separating two bounded sets in $L_p([a, b])$, with $[a, b]$ being an interval in $\mathbb{R}$. Some illustative computational experience is reported, and a particular stopping criterion is provided for the case of functions of bounded variation in $L_2([a, b])$.

Keywords: Strong separation, infinite dimensional optimization, semi-infinite programming

2000 Mathematics Subject Classification: 90C48, 46A22, 90C90

1. Introduction

Separation of sets, either in the Euclidean space or in an infinite dimensional space, is a key topic in many fields of applied mathematics like artificial intelligence, pattern recognition, weather forescasting, data mining, neural networks, etc., as in [1], [3], [6], [7], and [8] (as a reduced sample of representative papers) is pointed out.

Let us consider two nonempty subsets, F and G, of a real normed space X. If the topological dual space of X is represented by X^*, and the null vector in both X and X^* is denoted by θ, we can give the following definition:

Definition 1.1. An affine subspace H in X is called a *topological hyperplane* if there exist a linear functional $u \in X^* \setminus \{\theta\}$ and a positive scalar γ such that $H = \{x \in X : u(x) = \gamma\}$.

ISSN 0944-6532 / \$ 2.50

We say that H *separates strongly* F and G if there exists a scalar ε such that

$$u(z) \geq \gamma + \varepsilon > \gamma - \varepsilon \geq u(y) \quad \text{for all } y \in F \text{ and all } z \in G.$$

In [5] the following infinite-dimensional optimization problem is associated with the strong separation of F and G:

$$\begin{array}{lll} (P) & \text{Inf} & \beta - \alpha \\ & \text{s.t.} & u(y) + \alpha \leq 0, \ y \in F, \\ & & u(z) + \beta \geq 0, \ z \in G, \\ & & u \in \mathbb{B}_{X^*}, \ \alpha \in \mathbb{R}, \ \beta \in \mathbb{R}, \end{array}$$

where $\mathbb{B}_{X^*}$ is the closed unit ball in X^*.

The optimal value of (P), denoted by $v(P)$, satisfies the inequality $v(P) \leq 0$, since $(u, \alpha, \beta) = (\theta, 0, 0)$ is a feasible solution of (P). In [5, Proposition 2.2(ii)] the following result is proved:

Theorem 1.2. *Given two nonempty subsets F and G of a normed space X, and the associated problem (P), there exists a topological hyperplane in X separating strongly F and G if and only if $v(P) < 0$.*

In the case that $v(P) < 0$, in order to establish that F and G can be strongly separated we do not need to reach an optimal solution of (P). In fact, finding a feasible solution (u, α, β) of (P) such that $\beta - \alpha < 0$ is sufficient, since this inequality entails

$$u(z) \geq -\beta > -\alpha \geq u(y) \quad \text{for all } y \in F \text{ and all } z \in G,$$

and $u \neq \theta$; therefore

$$H = \{x \in X \mid u(x) = -(1/2)(\alpha + \beta)\}$$

is a topological hyperplane that strongly separates F and G.

In [5] a cutting plane algorithm for the strong separation of two compact sets in a *separable* normed space is proposed. At each iteration of this algorithm, a subproblem of (P), obtained by replacing the sets of indices F and G by finite subsets (grids), has to be solved. If the algorithm does not stop at the current iteration, new points of F and G are aggregated to the current grids according to certain standard rule (two cuts are performed on the current feasible set).

The main drawback of the algorithm given in [5] is that each added cut requires to solve a *global optimization* problem. This is why in this paper we propose a less costly alternative approach, not requiring any global optimization, and yielding the strong separation hyperplane whitout reaching optimality. More precisely, our procedure is based on the following approximation scheme. Given a sequence $(E_n)_{n=1}^{\infty}$ of finite dimensional subspaces of X^*, we approach $v(P)$ by means of the optimal values, $v(P_n)$, where (P_n) is the semi-infinite optimization problem

$$\begin{array}{lll} (P_n) & \text{Inf} & \beta - \alpha \\ & \text{s.t.} & u(y) + \alpha \leq 0, \ y \in F, \\ & & u(z) + \beta \geq 0, \ z \in G, \\ & & u \in \mathbb{B}_{E_n}, \ \alpha \in \mathbb{R}, \ \beta \in \mathbb{R}, \end{array}$$

and $\mathbb{B}_{E_n}$ is the closed unit ball in E_n, the norm in E_n being induced by the norm in X^*. In Section 2 we introduce *condition* (C) and we prove that $v(P_n) \to v(P)$ as $n \to \infty$, provided that either F and G are bounded and that (C) holds. A stopping criterion is also discussed. In Section 3 an optimally-convergent algorithm for solving the problem of separating strongly two bounded sets in the space $L_p(\mathbb{X}, \mathcal{A}, \mu)$ is presented, and a couple of conditions entailing (C) are provided. In this section we also provide an implementable modification of our method for the strong separation of two sets in $L_p([a, b])$, where $[a, b]$ is an interval in $\mathbb{R}$. An ending criterion for this method is given in the case that the involved functions are of bounded variation, and some illustrative computational experience is reported.

2. Strong separation

It is straightforward that problem (P) is equivalent to

$$(P') \quad \begin{array}{ll} \text{Inf} & \sigma_F(u) + \sigma_{-G}(u), \\ \text{s.t.} & u \in \mathbb{B}_{X^*}, \end{array}$$

where we use the *support function*

$$\sigma_S(u) := \sup_{x \in S} u(x).$$

By equivalent problems we mean here that $\overline{u}$ is optimal for (P') if and only if $(\overline{u}, -\sigma_F(\overline{u}), \sigma_{-G}(\overline{u}))$ is optimal for (P), and the optimal values coincide, i.e. $v(P) = v(P')$. Since the functions σ_F and σ_{-G} are w^*-lsc and $\mathbb{B}_{X^*}$ is w^*-compact (Alaoglu's theorem), the optimal value of (P'), denoted by $v(P')$, is finite and attainable.

Now we consider the problems

$$(P'_n) \quad \begin{array}{ll} \text{Inf} & \sigma_F(u) + \sigma_{-G}(u), \\ \text{s.t.} & u \in \mathbb{B}_{E_n}, \end{array}$$

where $(E_n)_{n=1}^{\infty}$ is a given sequence of finite dimensional subspaces of X^*. Obviously $0 \geq v(P'_n) \geq v(P')$, because we take in E_n the norm induced by the norm in X^* and, so, $\mathbb{B}_{E_n} \subset \mathbb{B}_{X^*}$.

Let us introduce a key condition for the convergence of our method:

$$(C) \qquad \forall u \in X^*, \ \lim_{n \to \infty} d_*(u, E_n) = 0, \tag{1}$$

where $d_*(u, E_n) = \inf_{v \in E_n} \|u - v\|_*$ and $\|\cdot\|_*$ is the dual norm.

Condition (C) entails that $\cup_{n=1}^{\infty} E_n$ is dense in X^*, and thus X^* is separable. Moreover, when $(E_n)_{n=1}^{\infty}$ is an expansive sequence, i.e. $E_n \subset E_{n+1}$ for all n, then (C) holds if and only if $\cup_{n=1}^{\infty} E_n$ is dense in X^*. Observe also that if X^* has a *(Schauder) basis* $\{u_k, \ k = 1, 2, ...\}$ we can take $E_n = \text{span}\{u_1, u_2, ..., u_n\}, \ n = 1, 2, ...$

Proposition 2.1. *If F and G are bounded subsets of the normed space X and condition (C) holds, then*

$$\lim_{n \to \infty} v(P'_n) = v(P'). \tag{2}$$

Moreover, if there exists n_0 such that

$$E_{n_0}^{\perp} \cap \operatorname{span}(F \cup G) = \{\theta\}, \tag{3}$$

where

$$E_{n_0}^{\perp} := \{x \in X : \; u(x) = 0, \; \textit{for all } u \in E_{n_0}\},$$

then

$$v(P_1') = v(P_2') = ... = v(P_{n_0}') = 0 \Rightarrow v(P') = 0. \tag{4}$$

Proof. Since $0 \geq v(P_n') \geq v(P')$, (2) trivially holds when $v(P') = 0$. Assume then that $v(P') < 0$.

The boundedness of F and G entails the existence of a positive constant M such that $F \cup G \subset M \; \mathbb{B}_X$, with $\mathbb{B}_X$ denoting the closed unit ball in X.

Let $\overline{u} \in \mathbb{B}_{X^*}$ be an optimal solution of (P'), i.e. $\sigma_F(\overline{u}) + \sigma_{-G}(\overline{u}) = v(P')$. Since $v(P') < 0$ it is straightforward that $\|\overline{u}\|_* = 1$. Moreover, from (C), $\lim_{n \to \infty} d_*(\overline{u}, E_n) = 0$, and for every $\varepsilon > 0$ there will exist n_ε such that

$$d_*(\overline{u}, E_n) < \frac{\varepsilon}{4M}, \quad \text{for every } n \geq n_\varepsilon.$$

So, we can pick an element $u_n \in E_n$ such that

$$d_*(\overline{u}, u_n) < \frac{\varepsilon}{4M}, \quad \text{for every } n \geq n_\varepsilon.$$

Now we define, for $n \geq n_\varepsilon$,

$$v_n := \begin{cases} u_n, & \text{if } \|u_n\|_* \leq 1, \\ \dfrac{1}{\|u_n\|_*} u_n, & \text{if } \|u_n\|_* > 1. \end{cases}$$

Obviously $v_n \in \mathbb{B}_{E_n}$ and $\|u_n - v_n\|_* = \max\{0, \|u_n\|_* - 1\}$. Additionally

$$\|u_n\|_* \leq \|\overline{u}\|_* + \|u_n - \overline{u}\|_* = 1 + \|u_n - \overline{u}\|_*,$$

and, therefore,

$$\begin{aligned} \|\overline{u} - v_n\|_* &\leq \|\overline{u} - u_n\|_* + \|u_n - v_n\|_* \\ &\leq \|\overline{u} - u_n\|_* + \max\{0, \|u_n\|_* - 1\} \\ &\leq 2\,\|\overline{u} - u_n\|_* < \frac{\varepsilon}{2M}. \end{aligned}$$

Hence, for every $x \in M \; \mathbb{B}_X$, the following inequality holds

$$|(\overline{u} - v_n)(x)| = |\overline{u}(x) - v_n(x)| < \frac{\varepsilon}{2}.$$

Then, if $S \subset M \; \mathbb{B}_X$, it can easily be proved that

$$|\sigma_S(\overline{u}) - \sigma_S(v_n)| \leq \frac{\varepsilon}{2}.$$

The corresponding inequalities for $S = F$ and $S = -G$ yield

$$\begin{aligned}(\sigma_F(v_n) + \sigma_{-G}(v_n)) - v(P') &= (\sigma_F(v_n) + \sigma_{-G}(v_n)) - (\sigma_F(\overline{u}) + \sigma_{-G}(\overline{u})) \\ &\leq |(\sigma_F(\overline{u}) + \sigma_{-G}(\overline{u})) - (\sigma_F(v_n) + \sigma_{-G}(v_n)| \\ &\leq |\sigma_F(\overline{u}) - \sigma_F(v_n)| + |\sigma_{-G}(\overline{u}) - \sigma_{-G}(v_n)| \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,\end{aligned}$$

and

$$v(P'_n) \leq \sigma_F(v_n) + \sigma_{-G}(v_n) \leq v(P') + \varepsilon, \quad \text{for all } n \geq n_\varepsilon,$$

i.e. $v(P'_n) \to v(P')$ as $n \to \infty$.

Let us prove that (3) implies that for any element $u \in X^*$ it is possible to find an element $\overline{u} \in E_{n_0}$ such that

$$u(x) = \overline{u}(x), \quad \text{for all } x \in F \cup G. \tag{5}$$

Reasoning by contradiction, assume that there exists $u_0 \in X^* \diagdown \{\theta\}$ such that the system of equations

$$\{u(x) = u_0(x),\ x \in F \cup G\}$$

has no solution $u \in E_{n_0}$.

If $\{s_1, s_2, ..., s_p\}$ is a basis of E_{n_0}, the system above will not have solution $u \in E_{n_0}$ if and only if the system

$$\left\{\sum\nolimits_{i=1}^{p} \eta_i s_i(x) = u_0(x),\ x \in F \cup G\right\} \tag{6}$$

has no solution $(\eta_1, \eta_2, ..., \eta_p) \in \mathbb{R}^p$.

The semi-infinite system of linear equations in (6) has no solution if and only if

$$(0, ..., 0, 1) \in \operatorname{span}\left\{(s_1(x), ..., s_p(x), u_0(x)),\ x \in F \cup G\right\}. \tag{7}$$

(We have applied the Gale Alternative theorem (see, for instance, [4, Corollary 3.1.1])).

Finally, (7) entails the existence of a nonzero function $\varphi : F \cup G \to \mathbb{R}$ such that $\{x \in F \cup G :\ \varphi(x) \neq 0\}$ is finite, and

$$(0, ..., 0, 1) = \sum_{x \in F \cup G} \varphi(x)\left(s_1(x), ..., s_p(x), u_0(x)\right). \tag{8}$$

If we define

$$x_0 = \sum_{x \in F \cup G} \varphi(x) x \in \operatorname{span}(F \cup G),$$

from the first p equalities in (8) we get

$$s_1(x_0) = s_2(x_0) = ... = s_p(x_0) = 0 \Leftrightarrow x_0 \in E_{n_0}^{\perp},$$

whereas the last equality yields $x_0 \neq \theta$. In this way we have reached a contradiction with the assumption (3).

Now, we proceed with the proof of the stopping criterion (4). Reasoning again by contradiction, we assume that

$$v(P_1') = v(P_2') = ... = v(P_{n0}') = 0 \quad \text{and} \quad v(P') = v(P) < 0.$$

The last inequality implies the existence of $u \in X^* \diagdown \{\theta\}$ and $\alpha, \beta \in \mathbb{R}$ such that

$$u(z) \geq -\beta > -\alpha \geq u(y) \quad \text{for all } y \in F \text{ and all } z \in G.$$

According to the reasoning above, (3) leads us to the existence of $\overline{u} \in E_{n_0}$ satisfying (5), and therefore

$$\overline{u}(z) \geq -\beta > -\alpha \geq \overline{u}(y) \quad \text{for all } y \in F \text{ and all } z \in G.$$

It is obvious that $\overline{u} \neq \theta$, and $\overline{u}/\left\|\overline{u}\right\|$ is feasible for the problem (P_{n0}'). This yields the following contradiction:

$$v(P_{n_0}') \leq \frac{\beta - \alpha}{\|\overline{u}\|} < 0.$$

□

Remark 2.2. Let us discuss the computational implications of Proposition 2.1. In the following section we propose an algorithm for solving problem (P), approximating its optimal value $v(P)$ by means of the optimal values $v(P_n)$ (under condition (C)). Consequently, such algorithm will return the constant null sequence if and only if it is impossible to strongly separate F and G. Alternatively, if for a certain n_1 one has $v(P_{n_1}) < 0$, it turns out that it is possible to strongly separate both sets, and the algorithm stops because we have already computed a strong separation hyperplane (associated to a nonzero element $u \in E_{n_1}$). But if the algorithm keeps returning zeros as the optimal values of problems (P_n), we simply cannot decide whether it is because, despite the fact that $v(P) < 0$, the first nonnull $v(P_n)$ that the algoritm should return is yet to come.

In Section 3, we shall give an example for which our algorithm returns arbitrarily long sequences of zeros even for simple cases in which it is trivial to oberve that F and G can be strongly separated. According to this argument, it is a key matter to establish a stopping criterion like (4), and condition (3) gives rise to a such stopping criterion. In Section 4 we show how to check (3) in the particular setting $X = L_2([a, b]$.

Remark 2.3. Here we used a self-contained argument, but (2) can also be proved by using techniques of epiconvergence giving rise to the convergence of the optimal values of a sequence of unconstrained optimization problems associated with finitely-valued equi-Lipschitzian convex functions.

3. Strong separation in $L_p(\mathbb{X}, \mathcal{A}, \mu)$

Let $(\mathbb{X}, \mathcal{A}, \mu)$ be a measure space, i.e. $\mathcal{A}$ is a $\sigma-$algebra on $\mathbb{X}$ and μ is a measure on $(\mathbb{X}, \mathcal{A})$ $(\mu : \mathcal{A} \to [0, +\infty])$. For $1 < p < \infty$, consider the linear space of equivalence classes of real-valued $\mathcal{A}$-measurable functions such that $\int_{\mathbb{X}} |f|^p \, d\mu < \infty$, equipped with the norm

$$\|f\|_p = \left(\int_{\mathbb{X}} |f|^p \, d\mu \right)^{1/p}.$$

Here an equivalence class is formed by measurable functions that can differ from each other on a set of μ-measure zero. This is a Banach space, denoted by $L_p(\mathbb{X}, \mathcal{A}, \mu)$. Moreover, $L_p(\mathbb{X}, \mathcal{A}, \mu)$ is reflexive, and its dual is $L_q(\mathbb{X}, \mathcal{A}, \mu)$, with $q \in]1, \infty[$ satisfying $(1/p) + (1/q) = 1$. Although the elements of $L_p(\mathbb{X}, \mathcal{A}, \mu)$ are not functions, but equivalence classes of functions, this distinction is tacitly assumed and we shall speak of $L_p(\mathbb{X}, \mathcal{A}, \mu)$ as a space of functions. In particular, if $\mathbb{X}$ is a subset of the Euclidean space, $\mathbb{R}^k$, and μ is the Lebesgue measure, we write $L_p(\mathbb{X})$. For instance, we shall consider $L_p([a, b])$, where $[a, b]$ is an interval in $\mathbb{R}$, and some computational experience will be reported for functions in $L_2([a, b])$.

Given the measure space $(\mathbb{X}, \mathcal{A}, \mu)$, we say that $\mathcal{A}$ is *countably generated* if there is a countable subfamily $\mathcal{C}$ of $\mathcal{A}$ such that the σ-algebra generated by $\mathcal{C}$, represented by $\sigma(\mathcal{C})$ coincides with $\mathcal{A}$, i.e. $\sigma(\mathcal{C}) = \mathcal{A}$.

The measure μ in $(\mathbb{X}, \mathcal{A}, \mu)$ is said to be *σ-finite* if there exists a sequence $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ such that

$$\mathbb{X} = \bigcup_{n=1}^{\infty} A_n \quad \text{and} \quad \mu(A_n) < \infty, \ n = 1, 2, ... \tag{9}$$

Given $(\mathbb{X}, \mathcal{A}, \mu)$, if $\mathcal{A}$ is countably generated and μ is σ-finite, the space $L_p(\mathbb{X}, \mathcal{A}, \mu)$, $1 < p < \infty$, is separable (see, for instance, [2, Proposition 3.4.5]).

Obviously these two conditions are fulfilled whenever $\mathcal{A}$ is the Borel σ-algebra $\mathcal{B}(\mathbb{X})$ of a locally compact Hausdorff space $\mathbb{X}$, whose topology has a countable basis and $\mu(K) < +\infty$ for any compact subset $K \subset \mathbb{X}$. This is in particular the case of any nonempty subset $\mathbb{X} \subset \mathbb{R}^k$ and the Lebesgue measure induced on $\mathbb{X}$.

In our specific setting, Theorem 1.2 states that two nonempty subsets, F and G, of $L_p(\mathbb{X}, \mathcal{A}, \mu)$, $1 < p < \infty$, can be strongly separated by a topological hyperplane if and only if $v\,(P) < 0$, where (P) is the infinite-dimensional optimization problem

$$\begin{array}{lll} (P) & \text{Inf} & \beta - \alpha \\ & \text{s.t.} & \int_{\mathbb{X}} f\, h\, d\mu + \alpha \leq 0, \ f \in F, \\ & & \int_{\mathbb{X}} g\, h\, d\mu + \beta \geq 0, \ g \in G, \\ & & \|h\|_q \leq 1, \ \alpha \in \mathbb{R}, \ \beta \in \mathbb{R}, \end{array} \tag{10}$$

with $h \in L_q(\mathbb{X}, \mathcal{A}, \mu)$, $(1/p) + (1/q) = 1$. Remember that if $v\,(P) < 0$ and $\overline{h}$ is optimal for (P), then $\|\overline{h}\|_q = 1$.

In this section we provide an algorithm for solving (P) when F and G are bounded, and we establish a convergence result for it under the double assumption that $\mathcal{A}$ is countably generated and μ is σ-finite. To this aim, we make the following construction:

It is clear that, because $\mathcal{A}$ is countably generated, we can choose a countable subfamily $\mathcal{C}$ of $\mathcal{A}$ that generates $\mathcal{A}$ and contains the sets A_n, $n = 1, 2, ...$, in (9) (we just include them in $\mathcal{C}$).

Let $\widetilde{\mathcal{C}}$ consist of the sets in $\mathcal{C}$ together with their complements with respect to $\mathbb{X}$, and let $\widetilde{\mathcal{A}}$ be the algebra generated by $\widetilde{\mathcal{C}}$. Then, it is clear that $\widetilde{\mathcal{A}}$ is the set of finite unions of sets having the form $C_1 \cap C_2 \cap ... \cap C_k$, for some k and some choice of the sets C_1, C_2,..., C_k in $\widetilde{\mathcal{C}}$. We have $\sigma(\widetilde{\mathcal{A}}) = \mathcal{A}$, where $\sigma(\widetilde{\mathcal{A}})$ is the σ-algebra generated by $\widetilde{\mathcal{A}}$, and $A_n \in \widetilde{\mathcal{A}}$, $n = 1, 2, ...$ Moreover $\widetilde{\mathcal{A}}$ is countable, as well as the subfamily

$$\mathcal{D} := \{D \in \widetilde{\mathcal{A}} : \mu(D) < \infty\},$$

and we can write

$$\mathcal{D} = \{D_1, D_2, ..., D_k, ...\}. \tag{11}$$

Next we present the announced algorithm for separating two bounded sets, F and G, in $L_p(\mathbb{X}, \mathcal{A}, \mu)$, $1 < p < \infty$.

(A) Algorithm for solving (P) with F and G bounded.

Consider the sequence $\mathcal{D} = \{D_1, D_2, ..., D_k, ...\}$ in (11).

Step 1. *Set* $k = 1$.

Step 2. *Find an optimal solution* $(\lambda_{1k}, \lambda_{2k}, ..., \lambda_{kk}, \alpha_k, \beta_k)$ *of the semi-infinite programming problem*

$$\begin{array}{lll} (P_k) & \text{Inf} & \beta - \alpha \\ & \text{s.t.} & \int_{\mathbb{X}} f \left(\sum_{i=1}^k \lambda_i \mathbf{1}_{D_i}\right) d\mu + \alpha \leq 0, \ f \in F, \\ & & \int_{\mathbb{X}} g \left(\sum_{i=1}^k \lambda_i \mathbf{1}_{D_i}\right) d\mu + \beta \geq 0, \ g \in G, \\ & & \left\| \sum\limits_{i=1}^{k} \lambda_i \mathbf{1}_{D_i} \right\|_q \leq 1, \end{array} \tag{12}$$

with vector of variables $(\lambda_1, \lambda_2, ..., \lambda_k, \alpha, \beta) \in \mathbb{R}^{k+2}$, *and where* $\mathbf{1}_{D_i}$ *is the* characteristic function *of* D_i $(\mathbf{1}_{D_i}(x) = 1$ *if* $x \in D_i$ *and* $\mathbf{1}_{D_i}(x) = 0$ *if* $x \in \mathbb{X} \diagdown D_i)$, $i = 1, 2, ..., k, ...$

Consider $s_k := \sum_{i=1}^k \lambda_{ik} \mathbf{1}_{D_i} \in L_q(\mathbb{X}, \mathcal{A}, \mu)$.

Step 3. *If* $v(P_k) < 0$, *then* **stop**: $v(P) < 0$, $\|s_k\|_q = 1$, *and the topological hyperplane*

$$H_k := \left\{ h \in L_p(\mathbb{X}, \mathcal{A}, \mu) : \int_{\mathbb{X}} h\, s_k \, d\mu = -(1/2)(\alpha_k + \beta_k) \right\}$$

separates strongly F *and* G.

Otherwise, set $k \leftarrow k + 1$, *and go to* **Step 2.**

The following result yields the conclusion that, if the previous algorithm does not terminate, F and G cannot be strongly separated. Here we are handling the expansive sequence of finite dimensional subspaces of $L_q(\mathbb{X}, \mathcal{A}, \mu)$

$$E_n := \left\{ \sum_{i=1}^{n} \lambda_i \mathbf{1}_{D_i} \mid \lambda_i \in \mathbb{R}, \ i = 1, 2, ..., n \right\}, \ n = 1, 2, ... \tag{13}$$

Proposition 3.1. *If* $v(P)$ *and* $v(P_k)$ *represent the optimal values of the problems* (P) *and* (P_k) *considered in* (10) *and* (12), *then we have*

$$\lim_{k \to \infty} v(P_k) = v(P). \tag{14}$$

Consequently, if the sets F and G can be strongly separated ***Algorithm*** **(A)** *terminates in a finite number of iterations.*

Moreover, if there exists k_0 such that for each $h \in \operatorname{span}(F \cup G)$, $h \neq \theta$, there is an associated $\ell_h \in E_{k_0}$ such that

$$\int_{\mathbb{X}} h\, \ell_h \, d\mu \neq 0, \tag{15}$$

then

$$v(P_1) = v(P_2) = ... = v(P_{k_0}) = 0 \Rightarrow v(P) = 0.$$

Proof. For the expansive sequence of finite dimensional subspaces defined in (13), it is proved in [2, Proposition 3.4.5, p. 112] that $\cup_{n=1}^{\infty} E_n$ is dense in $L_q(\mathbb{X}, \mathcal{A}, \mu)$ and, so, condition (C) holds. Applying Proposition 2.1, we conclude

$$\lim_{k\to\infty} v(P_k) = \lim_{k\to\infty} v(P_k') = v(P') = v(P).$$

Obviously, condition (15) is equivalent, in this particular setting, to (3), and the second part of Proposition 2.1 applies. □

Let us consider the particular case of the strong separation of two bounded sets F and G in $L_p([a,b])$, where $[a,b]$ is an interval in $\mathbb{R}$. Now we shall deal with the problem

$$\begin{array}{lll} (P) & \text{Inf} & \beta - \alpha \\ & \text{s.t.} & \int_a^b f\, h\, dt + \alpha \leq 0,\ f \in F, \\ & & \int_a^b g\, h\, dt + \beta \geq 0,\ g \in G, \\ & & \int_a^b |h|^q \, dt \leq 1, \end{array} \tag{16}$$

with variables $h \in L_q([a,b])$, $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, and the integrals are in the sense of Lebesgue.

For $k = 1, 2, ...,$ we consider the k open subintervals of $[a,b]$

$$D_i^k = \left] a + \frac{b-a}{k}(i-1), a + \frac{b-a}{k} i \right[, \quad i = 1, 2, \cdots, k,$$

and introduce the following specific algorithm:

(B) Algorithm for solving (P) with F and G bounded in $L_p([a,b])$.

Consider the sequence $\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_k, ...,$ with $\mathcal{D}_k := \{D_i^k,\ i = 1, 2, \cdots, k\}$.

Step 1. *Set $k = 1$.*

Step 2. *Given the problem with real variables λ_i, $i = 1, 2, \cdots, k$, α, and β,*

$$\begin{array}{lll} \left(\widetilde{P}_k\right) & \text{Inf} & \beta - \alpha \\ & \text{s.t.} & \int_a^b f \left(\sum_{i=1}^k \lambda_i \mathbf{1}_{D_i^k}\right) dt + \alpha \leq 0,\ f \in F, \\ & & \int_a^b g \left(\sum_{i=1}^k \lambda_i \mathbf{1}_{D_i^k}\right) dt + \beta \geq 0,\ g \in G, \\ & & \frac{b-a}{k} \sum_{i=1}^{k} |\lambda_i|^q \leq 1, \end{array}$$

find optimal values of $\left(\widetilde{P}_k\right)$: λ_{ik}, $i = 1, 2, \cdots, k$, α_k, and β_k.

Define $s_k := \sum_{i=1}^{k} \lambda_{ik} \mathbf{1}_{D_i^k}$.

Step 3. *If* $v(\widetilde{P}_k) < 0$, *then* **stop**: $v(P) < 0$ *and the topological hyperplane*

$$H_k := \left\{ h \in L_p([a,b]) : \int_a^b h\, s_k\, dt = -(1/2)(\alpha_k + \beta_k) \right\},$$

separates strongly F *and* G.

Otherwise, set $k \leftarrow k+1$, *and go to* **Step 2.**

The problem $(\widetilde{P}_k)$ can be equivalently written

$$\begin{array}{lll} \left(\widetilde{P}_k\right) & \text{Inf} & \beta - \alpha \\ & \text{s.t.} & \sum_{i=1}^{k} \lambda_i \int_{D_i^k} f(t)\, dt + \alpha \leq 0, \; f \in F, \\ & & \sum_{i=1}^{k} \lambda_i \int_{D_i^k} g(t)\, dt + \beta \geq 0, \; g \in G, \\ & & \sum\limits_{i=1}^{k} |\lambda_i|^q \leq \frac{k}{b-a}. \end{array} \tag{17}$$

Next we present the convergence theorem relative to **Algorithm (B)**. Now we are considering the sequence of finite dimensional subspaces of $L_q([a,b])$

$$\widetilde{E}_n := \left\{ \sum_{i=1}^{n} \lambda_i \mathbf{1}_{D_i^n} \mid \lambda_i \in \mathbb{R}, \; i = 1,2,...,n \right\}, \; n = 1,2,.... \tag{18}$$

(Remember that $\mathcal{D}_n := \{D_i^n, \; i = 1,2,\cdots,n\}$.)

Proposition 3.2. *If* $v(\widetilde{P}_k)$ *represent the optimal values of the problems* $(\widetilde{P}_k)$ *considered in* (17), *then we have*

$$\lim_{k\to\infty} v(\widetilde{P}_k) = v(P). \tag{19}$$

Moreover, if $v(P) < 0$ ***Algorithm* (B)** *terminates in a finite number of iterations.*

Additionally, if there exists k_0 *such that for each* $h \in \operatorname{span}(F \cup G)$, $h \neq \theta$, *there is a function* $\ell_h \in \widetilde{E}_{k_0}$ *such that*

$$\int_a^b h(t)\, \ell_h(t)\, dt \neq 0, \tag{20}$$

then

$$v(\widetilde{P}_1) = v(\widetilde{P}_2) = ... = v(\widetilde{P}_{k_0}) = 0 \Longrightarrow v(P) = 0.$$

Proof. If $s : [a,b] \to \mathbb{R}$ is a nonzero step function, there will be numbers $\alpha_0, \alpha_1, ..., \alpha_J$ such that $a = \alpha_0 < \alpha_1 < ... < \alpha_J = b$ and s will take the value a_j at every point of the interval $]\alpha_{j-1}, \alpha_j[$, $j = 1,2,...,J$. Given the properties of the Lebesgue integral, we can neglect the values of s at the points α_j and consider

$$s = \sum_{j=1}^{J} a_j \mathbf{1}_{]\alpha_{j-1},\alpha_j[}.$$

Since s is nonzero, at least one of the scalars a_j, $j = 1,2,...,J$, is different from zero.

Let us introduce the subintervals

$$C_j^k = \bigcup \left\{ D_i^k \in \mathcal{D}_k : D_i^k \subset]\alpha_{j-1}, \alpha_j[\right\}, \quad j = 1, 2, ..., J.$$

Obviously

$$\mu \left( ]\alpha_{j-1}, \alpha_j[\diagdown C_j^k \right) \leq \frac{b-a}{k}, \quad j = 1, 2, ..., J, \ k = 1, 2, ...,$$

where μ is here the Lebesgue measure. Then, given $\varepsilon > 0$, there will exist k_ε such that, for every $k \geq k_\varepsilon$ one has

$$\frac{b-a}{k} \leq \left(\frac{\varepsilon}{|a_j| \, J} \right)^q, \quad \text{for all } a_j \neq 0, \ j = 1, 2, ..., J. \tag{21}$$

Now if we consider the function $s_k : [a, b] \to \mathbb{R}$

$$s_k = \sum_{j=1}^{J} a_j \mathbf{1}_{C_j^k},$$

then

$$\begin{aligned} \|s - s_k\|_q &= \left\| \sum_{j=1}^{J} a_j \left(\mathbf{1}_{]\alpha_{j-1}, \alpha_j[} - \mathbf{1}_{C_j^k} \right) \right\|_q \\ &\leq \sum_{j=1}^{J} |a_j| \left\| \mathbf{1}_{]\alpha_{j-1}, \alpha_j[} - \mathbf{1}_{C_j^k} \right\|_q \\ &= \sum_{j=1}^{J} |a_j| \, \mu \left( ]\alpha_{j-1}, \alpha_j[\diagdown C_j^k \right)^{1/q} \\ &\leq \sum_{j=1, a_j \neq 0}^{J} |a_j| \frac{\varepsilon}{|a_j| \, J} \\ &\leq \varepsilon. \end{aligned} \tag{22}$$

Given the sequence of finite dimensional subspaces of $L_q([a, b])$ introduced in (18), $(\widetilde{E}_n)_{n=1}^{\infty}$, our previous reasoning proves that, for every step function s, one has

$$\lim_{n \to \infty} d_{L_q}(s, \widetilde{E}_n) = 0.$$

(If $s = 0$, then $s \in \widetilde{E}_n$ for all n.)

Since the subspace of step functions is dense in $L_q([a, b])$, a straightforward application of the triangular inequality leads us to conclude that condition (C) holds. Applying Proposition 2.1, we get $\lim_{k \to \infty} v(\widetilde{P}_k) = v(P)$.

For proving the second statement in this proposition, just observe that (20) is nothing else but the particular version of (15) in this specific context. □

Computational experience. We coded Algorithm (B) in Matlab 7.0.1 for strongly separate two finite sets of functions in $L_2([a,b])$. To solve the subproblems $(\widetilde{P}_k)$ we used the function **fmincon**, a nonlinear programming solver in Matlab toolbox.

Problem 3.3. *Let* $[a,b]=[1,100]$, *and consider the sets*

$$F=\left\{t,t^2,t^3,t^4,t^5,t^6,t^7,t^8,t^9,t^{10}\right\}$$

and

$$G=\{1,\sin t,\cos t,\sin 2t,\cos 2t,\sin 3t,\cos 3t,\sin 4t,\cos 4t\}.$$

These two sets are easy to be separated since Algorithm (B) *yields the solution in* $k=1$. *The optimal solution of* $(\widetilde{P}_1)$ *is*

$$\lambda_{11}=-0.1005,\ \alpha_1=502.4687,\ \beta_1=9.9499,$$

and the topological hyperplane strongly separating F *and* G *is*

$$H_1=\left\{h\in L_2([1,100])\ :\ \int_1^{100} h(t)\ dt=2549.3\right\}.$$

Problem 3.4. *Let* $[a,b]=[0,3]$, $F=\{1,t,t^2,t^4\}$, *and* $G=\{t^3,t^5,t^7,t^8\}$.
Applying our algorithm, we obtain for $k=3$, $\alpha_3=0.4839$, $\beta_3=0.4645$, *and*

$$\lambda_3=\ (-0.9970,-0.0776,0.0047)^T.$$

Hence, a topological hyperplane strongly separating F *and* G *is*

$$H_3=\left\{h\in L_2([0,3])\ :\ \int_0^{3} h(t)s_3(t)\ dt=-0.4742\right\},$$

where $s_3(t)=\sum_{i=1}^3\lambda_{i3}\mathbf{1}_{D_i^3}(t)$ *and* $D_i^3=]i-1,i[,\ i=1,2,3$.

Problem 3.5. *Let* $[a,b]=[1,9]$, $F=\{1,\cos t,\cos 2t,\cdots,\cos 15t\}$, *and* $G=\{\sin t,\sin 2t,\cdots,\sin 12t\}$.

The number of subintervals needed is $k=3$, *and the solution of* $(\widetilde{P}_3)$ *is* $(\alpha_3,\beta_3)=(0.0008,-0.0022)$ *and*

$$\lambda_3=\ (-0.0450,-0.4089,0.4536)^T.$$

The topological hyperplane strongly separating F *and* G *is*

$$H_3=\left\{h\in L_2([1,9])\ :\ \int_1^{9} h(t)s_3(t)\ dt=0.0007\right\},$$

where $s_3(t)=\sum_{i=1}^3\lambda_{i3}\mathbf{1}_{D_i^3}(t)$ *and* $D_i^3=]1+8(i-1)/3,1+8i/3[,\ i=1,2,3$.

Problem 3.6. *Let* $[a,b]=[1,10]$, *and the sets of functions*

$$F=\left\{\begin{array}{l} t\tan\frac{1}{t},\ \frac{t}{1+t^3},\ \cos 3t,\ \sin t^2,\ t^4,\ t-\ln(1+t), \\ \frac{1}{t}+e^t,\ 4t-\sin t,\ \tan\left(\frac{1}{t}+1\right),\ e^{-t^2}-t^2 \end{array}\right\},$$

and

$$G=\left\{5,\ 5t,\ t^2,\ \ln(1+t^2),\ \sin 5t,\ \cos 7t,\ t^5+e^{-t^4},\ \frac{t}{\sqrt{1+t^2}}\right\}.$$

The number of subintervals involved is $k=9$, *and the solution of* $(\widetilde{P}_9)$ *is* $(\alpha_9,\beta_9)=(0.0032,-0.0024)$ *and*

$$\lambda_9=(-0.1334,0.5220,-0.6116,0.0014,0.4927,-0.2988,-0.0146,0.0574,-0.0122)^T.$$

The topological hyperplane strongly separating F *and* G *is*

$$H_9=\left\{h\in L_2([1,10])\ :\ \int_1^{10} h(t)s_9(t)\ dt=-0.0004\right\},$$

where $s_9(t)=\sum_{i=1}^{9}\lambda_{i9}\mathbf{1}_{D_i^9}(t)$ *and* $D_i=]i,1+i[$, $i=1,2,\cdots,9$.

Problem 3.7. *Let* $[a,b]=[1,6]$,

$$F=\left\{1,\ t,\ t^4,\ te^t,\ \sin t,\ \ln t,\ \cos 4t,\ \cot\frac{1}{3t}\right\},$$

and

$$G=\left\{t^2,\ t^3,\ \frac{1}{1+t^2},\ te^{-t},\ \sin(t-1),\ \tan\frac{1}{t}\right\}.$$

The number of subintervals is $k=7$, *and the solution of* $(\widetilde{P}_7)$ *is* $(\alpha_7,\beta_7)=(0.0067, 0.0037)$ *and*

$$\lambda_7=\ (0.1020,-0.4309,0.7258,-0.7001,0.4102,-0.1354,0.0189)^T.$$

Then the topological hyperplane strongly separating F *and* G *is*

$$H_7=\left\{h\in L_2([1,6])\ :\ \int_1^{6} h(t)s_7(t)\ dt=-0.0052\right\},$$

where $s_7(t)=\sum_{i=1}^{7}\lambda_{i7}\mathbf{1}_{D_i^7}(t)$ *and* $D_i^7=]1+5(i-1)/7,1+5i/7[$, $i=1,2,\cdots,7$.

The test results for Problem 3.3–3.7 are summarized in Table 3.1, where k is the number of subintervals in which we divided $[a,b]$. The results reported in Table 3.1 show a fast convergence of Algorithm (B) in all these test problems.

Problem 3.8. *Let* $[a,b]=[1,3]$, $F=\{1,\ t,\ t^2,\ t^4\}$, *and* $G=\{t^3,\ t^5,\ t^7,\ t^8,\ 0.1t^2+0.9t^4\}$.

It is clear that F *and* G *cannot be strongly separated by any topological hyperplane since the function* $0.1t^2+0.9t^4$ *in* G *is a convex combination of the functions* t^2 *and* t^4 *in* F. *Consequently, Algorithm* (B) *stopped when* $k=150$ *(using different starting points) with* $v(\widetilde{P}_k)=0$, $k=1,2,...,150$.

Problem	$(\lambda_0, \alpha_0, \beta_0)$	k	$v(\widetilde{P}_k)$
3.3	$(0, \cdots, 0, -1, 1)$	1	-492.5188
3.4	$(0, \cdots, 0, -1, 1)$	1	0
		2	0
		3	-0.0194
3.5	$(0, \cdots, 0, -1, 1)$	1	0
		2	0
		3	-0.0030
3.6	$(0, \cdots, 0, -1, 1)$	7	0
		8	0
		9	-0.0056
3.7	$(0, \cdots, 0, -1, 1)$	5	0
		6	0
		7	-0.0030

Table 3.1: The last three iterates generated by Algorithm (B)

The following example shows that Algorithm (B) can return arbitrarily large sequences of zeros even for very simple problems involving sets F and G which can easily be strongly separated.

Example 3.9. *Let $F = \{0\}$ and $G = \{g_n\}$ two singletons from $L_2([0, 2\pi])$, with*

$$g_n(t) := \frac{1}{\sqrt{\pi}} \sin(n!t).$$

The L_2-norm of g_n is 1, and it is obvious that F and G can be strongly separated by using the topological hyperplane

$$H = \left\{ \ell \in L_2([0, 2\pi]) : \int_0^{2\pi} \ell(t) \sin(n!t)\ dt = \frac{\sqrt{\pi}}{2} \right\}.$$

However, when we use Algorithm (B) to separate F from G, we obtain $v(\widetilde{P}_1) = v(\widetilde{P}_2) = \ldots = v(\widetilde{P}_n) = 0$ since, for all $k \leq n$ and $i = 1, 2, \ldots, k$,

$$\begin{aligned}
\int_{D_i^k} g_n(t)\ dt &= \frac{1}{\sqrt{\pi}} \int_{\frac{i-1}{k}2\pi}^{\frac{i}{k}2\pi} \sin(n!t)\ dt \\
&= \frac{1}{n!\sqrt{\pi}} \int_{\frac{i-1}{k}2\pi n!}^{\frac{i}{k}2\pi n!} \sin(s)\ ds \\
&= 0.
\end{aligned}$$

So, it is straightforward that $v(\widetilde{P}_k) = 0$. In fact, we have proved that condition (3) fails for every $k \leq n$, since

$$g_n \in E_k^{\perp} = \{\mathbf{1}_{D_i^k},\ i = 1, 2, \ldots, k\}^{\perp}.$$

Next we provide a particular stopping criterion for the special case in which F and G are sets of bounded variation functions.

Let us recall that the *total variation* of a measurable function h in the interval $[a,b]$ is

$$\mathbb{V}(h) := \sup_{p\in\mathbb{N}} \sup_{a\leq\alpha_0<\ldots<\alpha_p\leq b} \sum_{i=0}^{p-1} (\sup_{[\alpha_i,\alpha_{i+1}]} h - \inf_{[\alpha_i,\alpha_{i+1}]} h),$$

where $\sup_{[\alpha_i,\alpha_{i+1}]} h$ and $\inf_{[\alpha_i,\alpha_{i+1}]} h$ are the *essential supremum* and the *essential infimum* of h in the corresponding subinterval. Since h is actually a class of functions, any two of them identical except on a set of null measure, the total variation defined above is nothing else but the infimum of the classically-defined total variation among all the real functions whithin the class h. Next we give a lemma which is needed later on.

Lemma 3.10. *Let $n \in \mathbb{N}$ and $h \in L_2([a,b])$ such that*

$$\int_a^b h(t)\ell(t)\ dt = 0, \quad \textit{for all } \ell \in E_n, \tag{23}$$

where $E_n = \operatorname{span}\{\mathbf{1}_{D_i^n},\ i = 1,2,...,n\}$. *Then*

$$\mathbb{V}(h) \geq \|h\|_{L_2} \sqrt{\frac{n}{b-a}}. \tag{24}$$

Proof. (24) is trivially satisfied when $h = 0$; therefore we can assume that $\|h\|_{L_2} > 0$. Set

$$\lambda_i := \frac{\int_{D_i^n} h^2(t)\ dt}{\|h\|_{L_2}}.$$

It is obvious that $\lambda_i \geq 0,\ i = 1,2,...,n$, and that $\sum_{i=1}^n \lambda_i = 1$. At the same time, a standard argument in measure theory allows us to ensure the existence of a set of positive measure $K \subset D_i^n$ such that

$$|h| \geq \sqrt{\frac{n\lambda_i}{b-a}}\, \|h\|_{L_2} \quad \text{on } K. \tag{25}$$

Moreover condition (23) implies that $\int_a^b h(t)\ dt = 0$, and there must exist two subsets of D_i^n of positive measure, say U and V, such that h is nonpositive on U and nonnegative on V. Hence, $\sup_{D_i^n} h \geq 0$ and $\inf_{D_i^n} h \leq 0$ and, from (25) either $\sup_{D_i^n} h \geq \sqrt{\frac{n\lambda_i}{b-a}}\, \|h\|_{L_2}$ or $\inf_{D_i^n} h \leq -\sqrt{\frac{n\lambda_i}{b-a}}\, \|h\|_{L_2}$. Thus

$$\sup_{D_i^n} h - \inf_{D_i^n} h \geq \sqrt{\frac{n\lambda_i}{b-a}}\, \|h\|_{L_2},$$

and, consequently,

$$\begin{aligned} \mathbb{V}(h) &\geq \|h\|_{L_2} \sqrt{\frac{n}{b-a}} \sum_{i=1}^n \sqrt{\lambda_i} \\ &\geq \|h\|_{L_2} \sqrt{\frac{n}{b-a}} \sum_{i=1}^n \lambda_i \\ &= \|h\|_{L_2} \sqrt{\frac{n}{b-a}}. \end{aligned}$$

□

Proposition 3.11 (Stopping criterion). *Assume that there exists $M > 0$ such that*

$$\mathbb{V}(h) \leq M \quad \text{for all } h \in F \cup G,$$

and an algebraic basis of $\operatorname{span}(F \cup G)$, $\{h_i,\ i \in I\}$, *for which the scalar*

$$\alpha := \inf \left\{ \left\| \sum_{i \in I} \lambda_i h_i \right\|_{L_2} : \begin{array}{c} \sum_{i \in I} |\lambda_i| = 1 \text{ and only} \\ \text{finitely many } \lambda_i \text{ are nonzero} \end{array} \right\}$$

is positive. Finally, consider any

$$k_0 > \frac{(b-a)M^2}{\alpha^2}.$$

Then,

$$v(\widetilde{P}_1) = v(\widetilde{P}_2) = \ldots = v(\widetilde{P}_{k_0}) = 0 \Longrightarrow v(P) = 0.$$

Remark 3.12 (Before the proof). *If* $\operatorname{span}(F \cup G)$ *is finite-dimensional, i.e. if the cardinal $|I|$ is finite, the continuity of the norm and the compactness of the set*

$$\left\{ (\lambda_1, \lambda_2, \ldots, \lambda_{|I|}) : \sum_{i \in I} |\lambda_i| = 1 \right\} \subset \mathbb{R}^{|I|},$$

yields the positiveness of α.

Proof. Since $\mathbb{V} : \operatorname{span}(F \cup G) \to \mathbb{R}_+$ is a norm, we have for every set of coefficients λ_i, $i \in I$, such that $\sum_{i \in I} |\lambda_i| = 1$ and only finitely many of them are nonzero, the following inequality

$$\mathbb{V}\left(\sum_{i \in I} \lambda_i h_i \right) \leq M.$$

Therefore, for every $h = \sum_{i \in I} \eta_i h_i \in \operatorname{span}(F \cup G) \diagdown \{\theta\}$,

$$\mathbb{V}(h) = \mathbb{V}\left(\sum_{j \in I} |\eta_j| \sum_{i \in I} \frac{\eta_i}{\sum_{j \in I} |\eta_j|} h_i \right) \leq M \sum_{j \in I} |\eta_j| .$$

According to this and the definition of α one has

$$\|h\|_{L_2} \geq \alpha \sum_{j \in I} |\eta_j| \geq \frac{\alpha \mathbb{V}(h)}{M}$$

and, so,

$$\frac{\mathbb{V}(h)}{\|h\|_{L_2}} \leq \frac{M}{\alpha}. \tag{26}$$

Suppose, now, that for all $h \in \operatorname{span}(F \cup G)$ we have

$$\sqrt{\frac{k_0}{b-a}} > \frac{\mathbb{V}(h)}{\|h\|_{L_2}}. \tag{27}$$

Then, by the last lemma, for each $h \in \operatorname{span}(F \cup G)$ there exists a certain $\ell_h \in E_{k_0}$ such that

$$\int_a^b h(t)\ell_h(t)\, dt \neq 0.$$

Consequently, $v(\widetilde{P}_1) = v(\widetilde{P}_2) = ... = v(\widetilde{P}_{k_0}) = 0$ will imply $v(P) = 0$, but from (26), we observe that (27) holds when

$$\sqrt{\frac{k_0}{b-a}} > \frac{M}{\alpha},$$

or equivalently, if

$$k_0 > \frac{(b-a)M^2}{\alpha^2}.$$

□

Acknowledgements. The authors are thankful to Profs. L. Thibault and C. Zălinescu, as well as to an anonymous referee, for their unvaluable comments and suggestions for improvements, mainly for those ideas about the stopping criterion.

References

[1] K. P. Bennet: Decision tree construction via linear programming, in: Proc. 4th Midwest Artificial Intelligence Cognitive Sci. Soc. Conf. (1992) 97–101.

[2] D. L. Cohn: Measure Theory, Birhhäuser, Boston (1980).

[3] B. D. Coleman: Finitely convergent learning programs for the separation of shaded figures, Biol. Cybern. 25 (1976) 57–60.

[4] M. Á. Goberna, M. A. López: Linear Semi-Infinite Optimization, Wiley, Chichester (1998).

[5] M. Á. Goberna, M. A. López, S.-Y. Wu: Separation by hyperplanes: a linear semi-infinite programming approach, in: Semi-Infinite Programming. Recent Advances, M. A. Goberna, M. A. López (eds.), Nonconvex Optim. Appl. 57, Kluwer, Dordrecht (2001) 255–269.

[6] O. L. Mangasarian: Linear and nonlinear separation of patterns by linear programming, Oper. Res. 13 (1965) 444–452.

[7] O. L. Mangasarian: Mathematical programming in neural networks, ORSA J. Comput. 5 (1993) 349–360.

[8] J. B. Rosen: Pattern separation by convex programming, J. Math. Anal. Appl. 10 (1965) 123–134.