

DUALITY RESULTS AND DUAL BUNDLE METHODS BASED ON THE DUAL METHOD OF CENTERS FOR MINIMAX FRACTIONAL PROGRAMS*

KARIMA BOUFI[†] AND AHMED ROUBI[†]

Abstract. We propose new duality results for generalized fractional programs (GFP) for a wide class of problems, not limited only to the convex case. Our approach does not use Lagrangian duality, but only an equivalent form of the GFP. We present a general approximating scheme, based on the proximal point algorithm, for solving this dual program. We take advantage of the convexity property of the dual, independently of the primal properties, to build implementable bundle methods with the support of the general scheme. However, it is well known that the principal difficulty with the duality is the evaluation of the dual function. To mitigate this difficulty, we propose bundle methods that need only approximate values and approximate subgradients of the objective dual function. We prove the convergence and the rate of convergence of these algorithms. As is the case for dual algorithms, the proposed methods generate a sequence of values that converges from below to the minimal value of the GFP, and a sequence of approximate solutions that converges to a solution of the dual problem. For certain classes of problems, the convergence is at least linear.

Key words. minimax fractional programs, dual method of centers, proximal point algorithm, bundle methods, quadratic programming

AMS subject classifications. 90C32, 90C25, 49K35, 49M29, 49M37

DOI. 10.1137/18M1199708

1. Introduction. This paper presents duality results and algorithms for solving a dual of the problem

$$(P) \quad \bar{\lambda} = \inf_{x \in X} \left\{ \lambda(x) := \max_{i \in I} \frac{f_i(x)}{g_i(x)} \right\},$$

where the constraints set X is given by

$$X = \{x \in Y \mid h_j(x) \leq 0, j \in J = \{1, \dots, p\}\}$$

with Y a nonempty subset of \mathbb{R}^n , $I = \{1, \dots, m\}$, and f_i , g_i , and h_j are real-valued functions defined on X . Problems of this type arise in many fields of applications such as economics, management applications of goal programming, multi-objective programming involving ratios of functions, stochastic programming, databases, physics, telecommunications, and numerical analysis [21, 49, 44, 24, 55].

For solving a generalized fractional program (GFP), there have been several primal Dinkelbach-type algorithms [16, 17, 7, 13, 47, 14, 55] and dual algorithms and results [30, 15, 4, 5, 6, 11] in the literature. These algorithms are based on auxiliary parametric problems having simpler structures than the original problem. For the primal algorithms, the auxiliary problems furnish sequences of approximate optimal values converging decreasingly to the optimal value of (P) , whereas the sequences of values generated by the dual algorithms converge increasingly towards the optimal value of (P) .

*Received by the editors July 10, 2018; accepted for publication (in revised form) March 14, 2019; published electronically June 4, 2019.

<http://www.siam.org/journals/siopt/29-2/M119970.html>

[†]Laboratoire MISI, Faculté des Sciences et Techniques, University Hassan 1, Settat, Morocco
(karima1-2008@hotmail.com, roubia@hotmail.com).

For computational considerations, especially when a GFP does not have a unique solution or if the feasible set is unbounded, primal regularizations [26, 48, 59, 1, 10] and dual regularizations [20, 19, 3] were proposed using the proximal point algorithm; see, e.g., [43, 40, 46, 27].

Another strategy was proposed in [59], which consists in applying bundle methods for solving a GFP. These methods consist in approximately solving the primal auxiliary problem associated with the GFP by using primal bundle methods. Very recently, since the last algorithm is intended to solve a linearly constrained GFP, another primal bundle method, based this time on the extended method of centers [47], was proposed in [2] to deal with nonlinearly constrained GFPs. For a detailed bibliography, see [56, 57, 58].

If the duality results for a GFP, such as in [5], have been obtained from Lagrangian duality, or by using alternative theorems, as in [30, 61, 60], a new dual for GFP was proposed in [11] that exploits properties of the auxiliary program from the generalized method of centers [47]. In this work, we obtain duality results under mild assumptions. In particular, we use a minimax equality assumption instead of the usually used convexity assumptions, since this minimax equality holds for other classes of functions satisfying, e.g., quasi-convexity, invexity, or convexity-like properties; see, e.g., [51, 31, 52, 38, 23, 54].

On the other hand, it is well known that the Lagrangian dual of an ordinary program is convex, which is not the case for the dual of a GFP. However, the latter enjoys several other interesting properties. In particular, its associate parametric problems are convex with simple constraints. Taking the convexity property into account, our purpose in what follows is to solve the dual of the GFP proposed in [11] by means of its associated parametric programs using proximal bundle ideas. For bundle methods in the context of nonsmooth convex optimization, we refer the reader to [36, 12, 41, 42, 25, 37, 33, 35, 50, 29, 39]. Remember that bundle methods require values and subgradients of the objective and constraints functions from previous steps to construct models that approximate the original program. These models are usually polyhedral functions. Stabilized by a quadratic term, a kind of prox-regularization, quadratic programs must be solved several times to approximate the original problem.

In spite of the apparent nice properties and advantages of the dual problem, the disadvantage of the latter is that the evaluation of the objective function and subgradient may be difficult or expensive, since they require solving an optimization problem. To mitigate these difficulties, in our dual bundle method we firstly allow the user the choice of a simple constraint set in the definition of the dual function, and secondly content ourselves with approximate values and approximate subgradients of the dual objective function instead of exactly solving the optimization program.

The resolution of the dual via bundle methods has already been applied in [8] to the dual problem presented in [4], and in [9] to the dual given in [30, 60, 61], and gave rise to dual bundle proximal methods. The proposed methods give the minimal value of the GFP and a solution to its dual counterpart.

The paper is organized as follows. In section 2, we define a dual problem of the GFP and give duality results under minimal assumptions. In subsection 3.1, we present a general approximation proximal method based on the notion of (strong) c -approximation functions. Subsection 3.2 is devoted to effective construction of dual bundle methods.

2. Duality without Lagrangian. In this section, we propose a new dual for the GFP and some related duality results. These results were already proposed in [11]

as bases for the dual method of the primal one stated in [47], but we present them differently, with slight improvements and with relatively general assumptions, discarding the convexity assumptions made in [11]. For this, for $\lambda \in \mathbb{R}$, we put

$$\mathcal{F}(x, \lambda) := \max_{i \in I, j \in J} \{f_i(x) - \lambda g_i(x), h_j(x)\}.$$

In all of what follows, we will consider a nonempty subset S of \mathbb{R}^n such that

$$(2.1) \quad X \subset S \subset Y.$$

To avoid complex notation, we will not mention the dependence of several functions, problems, and assumptions on this set S . Implicitly, this notation will refer to a unique set satisfying (2.1). Its choice will be arbitrary provided that certain assumptions, which will become clear later, are verified.

Now, given a set S as in (2.1), we consider the following parametric program:

$$(P_\lambda) \quad \inf_{x \in S} \mathcal{F}(x, \lambda).$$

The primal method of centers proposed in [47] is based upon this parametric problem with $S = Y$. A dual version of the last method was presented in [11]. To begin the description of this dual approach, we pose for every $x \in \mathbb{R}^n$ the vector-valued functions

$$\begin{aligned} f(x) &:= (f_1(x), \dots, f_m(x))^\top, & g(x) &:= (g_1(x), \dots, g_m(x))^\top, \\ h(x) &:= (h_1(x), \dots, h_p(x))^\top. \end{aligned}$$

With the set S , we define for $(\alpha, \beta) \in \Sigma$, and $\alpha \neq 0$, the function d by

$$d(\alpha, \beta) = \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\},$$

where

$$\Sigma = \left\{ (\alpha, \beta) \in \mathbb{R}^m \times \mathbb{R}^p \mid \sum_{i=1}^m \alpha_i + \sum_{j=1}^p \beta_j = 1, \alpha_i \geq 0, \beta_j \geq 0, i \in I, j \in J \right\},$$

and we consider the dual problem

$$(D) \quad \sup_{(\alpha, \beta) \in \Sigma_*} d(\alpha, \beta),$$

where

$$\Sigma_* = \{(\alpha, \beta) \in \Sigma \mid \alpha \neq 0\}.$$

For $\lambda \in \mathbb{R}$, we associate with (D) the following parametric problem:

$$(D_\lambda) \quad \sup_{(\alpha, \beta) \in \Sigma} \mathcal{G}(\alpha, \beta, \lambda),$$

where

$$\mathcal{G}(\alpha, \beta, \lambda) := \inf_{x \in S} \{\alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x)\}.$$

Throughout the paper we will assume that

$$\delta := \inf_{x \in S} \min_{i \in I} g_i(x) > 0 \text{ and } \Delta := \sup_{x \in S} \max_{i \in I} g_i(x) < \infty.$$

In [11], the authors define (D) as a dual problem for (P) . Under certain convexity/concavity assumptions, they obtain duality results for (P) and (D) , and for their associate parametric problems, and propose an algorithm to solve (D) . The relationship between the optimization problem (P) and its new dual (D) is detailed in [11, Theorem 3.1].

We mention first that the set Y is a part of the constraints—generally it is either the entire space \mathbb{R}^n or destined to represent simple, such as linear or bound, constraints—whereas the set S is an arbitrary set fulfilling (2.1). The introduction of such a set may be helpful from theoretical and computational points of view. For instance, if X is compact and Y is not bounded, one can choose S compact. If Y is nonconvex, the set S may be convex. This will also be useful for relaxing some assumptions we will need later. For numerical considerations, the set S may be chosen with simple structure to facilitate the evaluation of d and \mathcal{G} . It should be noted that the duality results we are going to develop remain valid regardless of S provided that certain hypotheses are fulfilled.

On the other hand, it is well known that the Lagrangian dual of an ordinary program is convex, which is not the case for the dual of a GFP. However, the latter enjoys several other interesting properties. In particular, its associated parametric problems are convex. Our purpose in what follows is to precisely solve (D) by means of its associated parametric programs using bundle proximal ideas. This approach has already been applied in [8] to the “dual” problem presented in [4], and in [9] to the dual given in [30, 60, 61], and gave rise to dual bundle proximal methods. Before developing our approach, we start by giving some results that are fundamental to our analysis.

In the following results we discuss some properties of the function \mathcal{G} .

PROPOSITION 2.1.

1. *The function $(\alpha, \beta) \mapsto \mathcal{G}(\alpha, \beta, \lambda)$ is concave for all $\lambda \in \mathbb{R}$.*
2. *For every $(\alpha, \beta) \in \Sigma$, where $\alpha \neq 0$, $\mathcal{G}(\alpha, \beta, \lambda) = 0$ if and only if $\lambda = d(\alpha, \beta)$.*
3. *For every $(\alpha, \beta) \in \Sigma$ we have $\mathcal{G}(\bar{\alpha}, \bar{\beta}, d(\alpha, \beta)) \geq 0$, where $(\bar{\alpha}, \bar{\beta})$ is any optimal solution of (D) .*

Proof. 1. The function $(\alpha, \beta) \mapsto \mathcal{G}(\alpha, \beta, \lambda)$ is concave for all $\lambda \in \mathbb{R}$ because it is the pointwise infimum of linear functions.

2. Let $(\alpha, \beta) \in \Sigma_*$ be such that $\mathcal{G}(\alpha, \beta, \lambda) = 0$. Then

$$(2.2) \quad \inf_{x \in S} \{ \alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x) \} = 0,$$

which implies that

$$\alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x) \geq 0 \quad \text{for all } x \in S.$$

Therefore, since by assumption $\delta > 0$, it follows that g is positive on S , and we have

$$\frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \geq \lambda \quad \text{for all } x \in S.$$

So, $d(\alpha, \beta) \geq \lambda$.

On the other hand, (2.2) entails that for all $\varepsilon > 0$, there exists $x_\varepsilon \in S$ such that

$$\alpha^\top [f(x_\varepsilon) - \lambda g(x_\varepsilon)] + \beta^\top h(x_\varepsilon) \leq \varepsilon.$$

Thus,

$$\frac{\alpha^\top f(x_\varepsilon) + \beta^\top h(x_\varepsilon)}{\alpha^\top g(x_\varepsilon)} \leq \lambda + \frac{\varepsilon}{\alpha^\top g(x_\varepsilon)}.$$

By using the inequality $\alpha^\top g(x_\varepsilon) \geq \delta \sum_{i \in I} \alpha_i$, we get

$$\inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \leq \lambda + \frac{\varepsilon}{\delta \sum_{i \in I} \alpha_i},$$

thereby implying that $d(\alpha, \beta) \leq \lambda$, and the equality follows.

3. If $(\bar{\alpha}, \bar{\beta}) \in \Sigma_*$ is an optimal solution of (D) , then for all $(\alpha, \beta) \in \Sigma_*$,

$$\begin{aligned} d(\alpha, \beta) &\leq d(\bar{\alpha}, \bar{\beta}) \\ &= \inf_{x \in S} \left\{ \frac{\bar{\alpha}^\top f(x) + \bar{\beta}^\top h(x)}{\bar{\alpha}^\top g(x)} \right\} \\ &\leq \frac{\bar{\alpha}^\top f(x) + \bar{\beta}^\top h(x)}{\bar{\alpha}^\top g(x)} \quad \text{for all } x \in S, \end{aligned}$$

which yields

$$0 \leq \bar{\alpha}^\top [f(x) - d(\alpha, \beta)g(x)] + \bar{\beta}^\top h(x) \quad \text{for all } x \in S.$$

This means that $\mathcal{G}(\bar{\alpha}, \bar{\beta}, d(\alpha, \beta)) \geq 0$. □

We will also need the following result from [10].

LEMMA 2.2. *Let $(\alpha, \beta) \in \Sigma$ and $\lambda, \mu \in \mathbb{R}$ be such that $\lambda \leq \mu$. Then we have*

$$\mathcal{G}(\alpha, \beta, \mu) + (\mu - \lambda) \delta \sum_{i \in I} \alpha_i \leq \mathcal{G}(\alpha, \beta, \lambda) \leq \mathcal{G}(\alpha, \beta, \mu) + (\mu - \lambda) \Delta \sum_{i \in I} \alpha_i.$$

The minimum requirements in term of hypotheses to establish duality results for the GFP are the following assumptions, which permit the abandonment of compactness of the constraints set and the convexity/concavity of the functions f_i , g_i , and h_j .

Assumption 2.3. Assume that $\bar{\lambda} > -\infty$ and $\mathcal{G}(\alpha, \beta, \bar{\lambda}) > -\infty$ for all $(\alpha, \beta) \in \Sigma$.

It is easy to see from Lemma 2.2 that, under the above conditions on δ and Δ , the assumption $\mathcal{G}(\alpha, \beta, \bar{\lambda}) > -\infty$ implies that $\mathcal{G}(\alpha, \beta, \lambda) > -\infty$ for all $\lambda \in \mathbb{R}$. This assumption will be made in all what follows.

Let us now go back to problem (P_λ) . By using the equality

$$(2.3) \quad \max_{i \in I, j \in J} \{f_i(x) - \lambda g_i(x), h_j(x)\} = \max_{(\alpha, \beta) \in \Sigma} \{\alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x)\},$$

we get

$$\inf_{x \in S} \mathcal{F}(x, \lambda) = \inf_{x \in S} \max_{(\alpha, \beta) \in \Sigma} \{\alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x)\}.$$

Let $\vartheta(\cdot)$ denote the optimal value of problem (\cdot) and let $\bar{\lambda}$ denote the optimal value of (P) .

PROPOSITION 2.4. *With the assumptions $\Delta < \infty$ and $\bar{\lambda} > -\infty$, we have*

$$\begin{aligned}\vartheta(P_{\bar{\lambda}}) &= \inf_{x \in S} \max_{(\alpha, \beta) \in \Sigma} \{ \alpha^\top [f(x) - \bar{\lambda}g(x)] + \beta^\top h(x) \} \\ &= \inf_{x \in X} \max_{(\alpha, \beta) \in \Sigma} \{ \alpha^\top [f(x) - \bar{\lambda}g(x)] + \beta^\top h(x) \} \\ &= 0\end{aligned}$$

for every subset S of \mathbb{R}^n such that $X \subset S$.

Proof. From the definition of $\bar{\lambda}$ we have $\bar{\lambda} \leq \lambda(x)$ for all $x \in X$. Let $S \subset \mathbb{R}^n$ be such that $X \subset S$, and let $x \in S$. If $x \in X$, let $i_0 \in I$ be such that $\lambda(x) = f_{i_0}(x)/g_{i_0}(x)$. Then, the previous inequality entails $0 \leq f_{i_0}(x) - \bar{\lambda}g_{i_0}(x)$. If $x \notin X$, there exists $j \in J$ such that $h_j(x) > 0$. Thus,

$$0 \leq \mathcal{F}(x, \bar{\lambda}) := \max_{i \in I, j \in J} \{ f_i(x) - \bar{\lambda}g_i(x), h_j(x) \} \quad \text{for all } x \in S.$$

This implies that $\vartheta(P_{\bar{\lambda}}) \geq 0$.

On other hand, since $\bar{\lambda} = \inf_{x \in X} \lambda(x)$, we have that for all $\varepsilon > 0$ there exists $x_\varepsilon \in X$ such that $\lambda(x_\varepsilon) \leq \bar{\lambda} + \varepsilon$. It follows that

$$\max_{i \in I} \{ f_i(x_\varepsilon) - \bar{\lambda}g_i(x_\varepsilon) \} \leq \varepsilon \max_{i \in I} g_i(x_\varepsilon) \leq \varepsilon \Delta.$$

We have, on the other hand, $h_j(x_\varepsilon) \leq 0$ for all $j \in J$, since $x_\varepsilon \in X$. Thus, $\mathcal{F}(x_\varepsilon, \bar{\lambda}) \leq \varepsilon \Delta$, which implies that

$$\inf_{x \in X} \mathcal{F}(x, \bar{\lambda}) \leq 0.$$

Since $X \subset S$, it follows that

$$\vartheta(P_{\bar{\lambda}}) = \inf_{x \in S} \mathcal{F}(x, \bar{\lambda}) \leq \inf_{x \in X} \mathcal{F}(x, \bar{\lambda}) \leq 0.$$

Therefore, since we showed that $\vartheta(P_{\bar{\lambda}}) \geq 0$, we get

$$\begin{aligned}\vartheta(P_{\bar{\lambda}}) &= \inf_{x \in S} \mathcal{F}(x, \bar{\lambda}) \\ &= \inf_{x \in X} \mathcal{F}(x, \bar{\lambda}) \\ &= 0,\end{aligned}$$

and we complete the proof by invoking (2.3). \square

Remark 2.5. If $\Delta = \infty$, the result of Proposition 2.4 may fail; see, e.g., Example 2.5 in [16], where $\Delta = \infty$ and $\vartheta(P_{\bar{\lambda}}) \neq 0$.

The usual assumption in dealing with convexity for a GFP is the convexity of the functions $x \mapsto \alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x)$ for all $(\alpha, \beta) \in \Sigma$. Conditions on f , g , and h to guarantee the last assumption are standard and may be found, e.g., in [30, 16, 17, 4, 5, 47, 59] and several other references.

To define duals and obtain duality results for the GFP, it seems that this kind of assumption only serves for using an equality minimax result similar to (2.4), which we state below in the next assumption. For this reason we use this assumption, knowing that it can be satisfied with other assumptions on f , g , and h such as

quasi-convexity [53], convexity-like assumptions [22, 53], and other general notions of convexity [51, 31, 52, 38, 23]. These references are only for illustration, and far from covering all the notions and results on the subject.

Assumption 2.6. Let $\lambda \in \mathbb{R}$ and $S \subset \mathbb{R}^n$. We will say that Assumption 2.6 is fulfilled for $\lambda \in \mathbb{R}$ and S if the minimax equality

$$(2.4) \quad \begin{aligned} \sup_{(\alpha,\beta) \in \Sigma} \inf_{x \in S} \{ \alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x) \} \\ = \inf_{x \in S} \sup_{(\alpha,\beta) \in \Sigma} \{ \alpha^\top [f(x) - \lambda g(x)] + \beta^\top h(x) \}, \end{aligned}$$

holds. This means that $\vartheta(P_\lambda) = \vartheta(D_\lambda)$.

To guarantee existence of solutions to the dual (D) , we need the following Slater condition.

Assumption 2.7. There exists some $\hat{x} \in Y$ such that $h(\hat{x}) < 0$.

PROPOSITION 2.8. *Assume that Assumption 2.6 is fulfilled with $\bar{\lambda}$ and some $S \supset X$. Then, with Assumptions 2.3 and 2.7, the parametric dual problem $(D_{\bar{\lambda}})$ has an optimal solution $(\alpha_{\bar{\lambda}}, \beta_{\bar{\lambda}}) \in \Sigma$ with $\alpha_{\bar{\lambda}} \neq 0$.*

Proof. Since

$$\vartheta(D_{\bar{\lambda}}) = \sup_{(\alpha,\beta) \in \Sigma} \inf_{x \in S} \{ \alpha^\top [f(x) - \bar{\lambda} g(x)] + \beta^\top h(x) \}$$

and the function $(\alpha, \beta) \mapsto \mathcal{G}(\alpha, \beta, \bar{\lambda})$ is continuous, as a finite concave function, there exists $(\alpha_{\bar{\lambda}}, \beta_{\bar{\lambda}}) \in \Sigma$ such that

$$\vartheta(D_{\bar{\lambda}}) = \inf_{x \in S} \{ \alpha_{\bar{\lambda}}^\top [f(x) - \bar{\lambda} g(x)] + \beta_{\bar{\lambda}}^\top h(x) \}.$$

But $\vartheta(D_{\bar{\lambda}}) = \vartheta(P_{\bar{\lambda}})$, by Assumption 2.6, and $\vartheta(P_{\bar{\lambda}}) = 0$ from Proposition 2.4. Then it follows that

$$0 \leq \alpha_{\bar{\lambda}}^\top [f(x) - \bar{\lambda} g(x)] + \beta_{\bar{\lambda}}^\top h(x) \quad \text{for all } x \in S.$$

From Assumption 2.7 we conclude that $\alpha_{\bar{\lambda}} \neq 0$, since otherwise we would have

$$0 \leq \beta_{\bar{\lambda}}^\top h(x) \quad \text{for all } x \in S.$$

And this inequality cannot hold for the $\hat{x} \in S$ specified in Assumption 2.7. \square

In the next theorem, we give a strong duality result between problem (P) and its associated dual (D) under Assumptions 2.3, 2.6, and 2.7, and we prove that the dual problem (D) always has optimal solutions, even if the primal problem (P) has an empty optimal solution set.

THEOREM 2.9. *With the assumptions that $\delta > 0$ and $\Delta < \infty$, and Assumption 2.3, we have $d(\alpha, \beta) > -\infty$ for all $(\alpha, \beta) \in \Sigma_*$. Moreover, if Assumption 2.7 is fulfilled and Assumption 2.6 is satisfied with $\lambda = \bar{\lambda}$ and $S \supset X$, then for every subset*

$S' \subset \mathbb{R}^n$ such that $X \subset S' \subset S$, it holds that

$$\begin{aligned}\vartheta(P) = \vartheta(D) &= \sup_{(\alpha, \beta) \in \Sigma} \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \\ &= \sup_{(\alpha, \beta) \in \Sigma} \inf_{x \in S'} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \\ &= \sup_{(\alpha, \beta) \in \Sigma} \inf_{x \in X} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\},\end{aligned}$$

and the optimal solution set of the dual problem (D) is nonempty.

Proof. Let $(\alpha, \beta) \in \Sigma_*$, and by virtue of Assumption 2.3, let $a \leq 0$ be such that

$$\inf_{x \in S} \{ \alpha^\top [f(x) - \bar{\lambda}g(x)] + \beta^\top h(x) \} \geq a.$$

Then $\alpha^\top [f(x) - \bar{\lambda}g(x)] + \beta^\top h(x) \geq a$ for all $x \in S$ and, consequently,

$$\begin{aligned}\frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} &\geq \frac{a}{\alpha^\top g(x)} + \bar{\lambda} \\ &\geq \frac{a}{\delta \sum_{i \in I} \alpha_i} + \bar{\lambda} \quad \text{for all } x \in S,\end{aligned}$$

where we use the inequality $\alpha^\top g(x) \geq \delta \sum_{i \in I} \alpha_i$ for all $x \in S$. This implies that $d(\alpha, \beta) > -\infty$.

Now, by using Assumption 2.6 with $\lambda = \bar{\lambda}$, and taking into account that $\vartheta(P_{\bar{\lambda}}) = 0$, by Proposition 2.4, we conclude that $\vartheta(D_{\bar{\lambda}}) = 0$. Hence, from Proposition 2.8, the parametric dual problem $(D_{\bar{\lambda}})$ has an optimal solution $(\alpha_{\bar{\lambda}}, \beta_{\bar{\lambda}}) \in \Sigma_*$. So, $\vartheta(D_{\bar{\lambda}}) = \mathcal{G}(\alpha_{\bar{\lambda}}, \beta_{\bar{\lambda}}, \bar{\lambda}) = 0$. From point 2 of Proposition 2.1, we conclude that

$$(2.5) \quad \begin{aligned}\bar{\lambda} &= d(\alpha_{\bar{\lambda}}, \beta_{\bar{\lambda}}) \\ &:= \inf_{x \in S} \left\{ \frac{\alpha_{\bar{\lambda}}^\top f(x) + \beta_{\bar{\lambda}}^\top h(x)}{\alpha_{\bar{\lambda}}^\top g(x)} \right\},\end{aligned}$$

which implies that

$$\bar{\lambda} \leq \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\}.$$

If a subset S' of \mathbb{R}^n satisfies $X \subset S' \subset S$, then

$$(2.6) \quad \bar{\lambda} \leq \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\}$$

$$(2.7) \quad \leq \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in S'} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\}$$

$$(2.8) \quad \leq \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in X} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\},$$

where the second and the third inequalities follow from the inclusions $X \subset S' \subset S$. On the other hand, from Proposition 2.4, we have

$$0 = \inf_{x \in X} \max_{(\alpha, \beta) \in \Sigma} \{ \alpha^\top [f(x) - \bar{\lambda}g(x)] + \beta^\top h(x) \},$$

which implies that for all $\varepsilon > 0$ there exists $x_\varepsilon \in X$ such that

$$\max_{(\alpha, \beta) \in \Sigma} \{ \alpha^\top [f(x_\varepsilon) - \bar{\lambda}g(x_\varepsilon)] + \beta^\top h(x_\varepsilon) \} \leq \varepsilon.$$

This means that

$$\alpha^\top [f(x_\varepsilon) - \bar{\lambda}g(x_\varepsilon)] + \beta^\top h(x_\varepsilon) \leq \varepsilon \quad \text{for all } (\alpha, \beta) \in \Sigma.$$

So, by using the inequality $\alpha^\top g(x_\varepsilon) \geq \delta \sum_{i \in I} \alpha_i$, we obtain

$$\frac{\alpha^\top f(x_\varepsilon) + \beta^\top h(x_\varepsilon)}{\alpha^\top g(x_\varepsilon)} \leq \bar{\lambda} + \frac{\varepsilon}{\delta \sum_{i \in I} \alpha_i} \quad \text{for all } (\alpha, \beta) \in \Sigma_*.$$

Therefore, for all $\varepsilon > 0$, one has

$$\inf_{x \in X} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \leq \bar{\lambda} + \frac{\varepsilon}{\delta \sum_{i \in I} \alpha_i} \quad \text{for all } (\alpha, \beta) \in \Sigma_*.$$

This gives

$$\inf_{x \in X} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \leq \bar{\lambda} \quad \text{for all } (\alpha, \beta) \in \Sigma_*,$$

implying that

$$(2.9) \quad \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in X} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \leq \bar{\lambda}.$$

Finally, from (2.9) and inequalities (2.6)–(2.8), we conclude that

$$\begin{aligned} \bar{\lambda} &= \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \\ &= \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in S'} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \\ &= \sup_{(\alpha, \beta) \in \Sigma_*} \inf_{x \in X} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\}. \end{aligned}$$

It is clear that if Assumption 2.7 is satisfied, $d(\alpha, \beta) \rightarrow -\infty$, when $\alpha \rightarrow 0$, in such a way that we will have

$$\begin{aligned} \bar{\lambda} &= \sup_{(\alpha, \beta) \in \Sigma} \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \\ &= \sup_{(\alpha, \beta) \in \Sigma} \inf_{x \in S'} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \\ &= \sup_{(\alpha, \beta) \in \Sigma} \inf_{x \in X} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\}. \end{aligned}$$

We return now to (2.5) to conclude that $(\alpha_{\bar{\lambda}}, \beta_{\bar{\lambda}})$ is an optimal solution for the dual (D) , which achieves the proof. \square

3. Dual proximal bundle method. In this section we present a dual bundle-type algorithm based on the bundle methods, the proximal point algorithm, and the dual method of centers to solve the dual problem (D) of the GFP.

Primal bundle methods for GFPs were first proposed in [59]. These methods solve approximately auxiliary problems by solving simpler ones, assuming certain convexity/concavity properties. The convexity is a key property, since bundle methods require knowledge of at least a subgradient and the value of the objective function at each point. For bundle methods for convex problems, see, e.g., [36, 32, 12, 29]. Also, there are several recent works dealing with duality for GFPs (see, e.g., [4, 5, 20, 19, 3, 11, 8, 9]); in particular, dual bundle methods were proposed in [8] to solve the dual problem presented in [4], and other dual bundle methods were presented in [9] to solve the Lagrangian dual problem given in [5].

An advantage of the dual problems is that their parametric subproblems are convex, since the objective function is concave and the constraint set is convex. This is always true without convexity assumptions on the objective functions and on the constraints of the primal problem. This permits us to use subgradients, but an inconvenience is that their calculation may be expensive since it requires solving an optimization problem. For this reason we will content ourselves with approximate subgradients, since their calculation requires only approximately solving the optimization problem.

More precisely, let $(\hat{\alpha}, \hat{\beta}) \in \Sigma$. Then, if there exists $\hat{x} \in S$ such that

$$\begin{aligned}\mathcal{G}(\hat{\alpha}, \hat{\beta}, \lambda) &:= \inf_{x \in S} \left\{ \hat{\alpha}^\top [f(x) - \lambda g(x)] + \hat{\beta}^\top h(x) \right\} \\ &:= \hat{\alpha}^\top [f(\hat{x}) - \lambda g(\hat{x})] + \hat{\beta}^\top h(\hat{x}),\end{aligned}$$

then

$$-\begin{pmatrix} f(\hat{x}) - \lambda g(\hat{x}) \\ h(\hat{x}) \end{pmatrix} \in \partial[-\mathcal{G}](\hat{\alpha}, \hat{\beta}, \lambda),$$

where $\partial[-\mathcal{G}](\hat{\alpha}, \hat{\beta}, \lambda)$ denotes the subdifferential of $-\mathcal{G}(\cdot, \cdot, \lambda)$ at $(\hat{\alpha}, \hat{\beta})$; see, e.g., [28, Definition 1.2.1, Page 241].

But, as mentioned above, even if the dual problem (D_λ) is convex, the exact evaluation of $\mathcal{G}(\hat{\alpha}, \hat{\beta}, \lambda)$ can be a source of difficulty. For this reason, we will content ourselves with an approximate evaluation of the function $\mathcal{G}(\cdot, \cdot, \lambda)$ at the desired points to obtain approximate subgradients.

Indeed, let $(\hat{\alpha}, \hat{\beta}) \in \mathbb{R}^m \times \mathbb{R}^p$, $\varepsilon \geq 0$, and let $\hat{x} \in S$ be a point satisfying

$$\hat{\alpha}^\top [f(\hat{x}) - \lambda g(\hat{x})] + \hat{\beta}^\top h(\hat{x}) \leq \mathcal{G}(\hat{\alpha}, \hat{\beta}, \lambda) + \varepsilon.$$

If we define

$$s_\lambda(\hat{\alpha}, \hat{\beta}) := -\begin{pmatrix} f(\hat{x}) - \lambda g(\hat{x}) \\ h(\hat{x}) \end{pmatrix},$$

then $s_\lambda(\hat{\alpha}, \hat{\beta}) \in \partial_\varepsilon[-\mathcal{G}](\hat{\alpha}, \hat{\beta}, \lambda)$, where $\partial_\varepsilon[-\mathcal{G}](\hat{\alpha}, \hat{\beta}, \lambda)$ is the ε -subdifferential of the convex function $-\mathcal{G}(\cdot, \cdot, \lambda)$ at $(\hat{\alpha}, \hat{\beta})$. That is,

$$-\mathcal{G}(\alpha, \beta, \lambda) \geq -\mathcal{G}(\hat{\alpha}, \hat{\beta}, \lambda) + \langle s_\lambda(\hat{\alpha}, \hat{\beta}), (\alpha, \beta) - (\hat{\alpha}, \hat{\beta}) \rangle - \varepsilon \quad \text{for all } (\alpha, \beta) \in \mathbb{R}^m \times \mathbb{R}^p.$$

At this stage, one can compute an ε -subgradient and an approximate value of the function at desired points of Σ . This is the weak assumption required for constructing bundle methods. So, having such information at several points one can construct a bundle approximation of $\mathcal{G}(\cdot, \cdot, \lambda)$. Before doing so, we will begin by introducing a general approximation scheme based only on the notion of c -approximation functions. Existence and construction of such functions will be discussed later.

3.1. General dual approximating proximal algorithm. Before formally stating our general approximating proximal algorithm for solving the dual problem (D) , we first give some helpful results which we will use later to interpret the algorithm. Let us start with this interesting result, which may be shown by simple calculations.

PROPOSITION 3.1. *Let $\lambda \in \mathbb{R}$ and $(\bar{\alpha}, \bar{\beta}) \in \Sigma$. For all $\varepsilon \geq 0$, let $x_\varepsilon \in S$ be such that*

$$\begin{aligned}\mathcal{G}(\bar{\alpha}, \bar{\beta}, \lambda) &:= \inf_{x \in S} \{ \bar{\alpha}^\top [f(x) - \lambda g(x)] + \bar{\beta}^\top h(x) \} \\ &\geq \bar{\alpha}^\top [f(x_\varepsilon) - \lambda g(x_\varepsilon)] + \bar{\beta}^\top h(x_\varepsilon) - \varepsilon.\end{aligned}$$

Then

$$-\begin{pmatrix} f(x_\varepsilon) - \lambda g(x_\varepsilon) \\ h(x_\varepsilon) \end{pmatrix} \in \partial_\varepsilon[-\mathcal{G}](\bar{\alpha}, \bar{\beta}, \lambda),$$

where $\partial_\varepsilon[-\mathcal{G}](\bar{\alpha}, \bar{\beta}, \lambda)$ is the ε -subdifferential of $-\mathcal{G}(\cdot, \cdot, \lambda)$ at $(\bar{\alpha}, \bar{\beta})$.

In order to easily solve the prox-regularized problem of (D_{d_k}) introduced in [10] and given by

$$(D_{\eta_k}(\alpha_k, \beta_k, d_k)) \quad \max_{(\alpha, \beta) \in \Sigma} \left\{ \mathcal{G}(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\},$$

we propose in this section, based on the bundle methods concept, approximating the function $\mathcal{G}(\cdot, \cdot, d_k)$ from above by a concave function $\psi(\cdot, \cdot, d_k)$. So, instead of solving $(D_{\eta_k}(\alpha_k, \beta_k, d_k))$, a sequence of approximating problems

$$(\mathcal{AD}_{\eta_k}(\alpha_k, \beta_k, d_k)) \quad \max_{(\alpha, \beta) \in \Sigma} \left\{ \psi(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\},$$

where (α_k, β_k) is the point computed at the precedent iteration, will be solved until an approximate solution, in a sense that will be defined later, of the prox-regularized problem $(D_{\eta_k}(\alpha_k, \beta_k, d_k))$ is reached.

DEFINITION 3.2. *Letting $c \in]0, 1[$ be a given parameter, a concave function $\psi(\cdot, \cdot, d_k)$ is a c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at $(\alpha_k, \beta_k) \in \Sigma$ if*

$$\psi(\alpha, \beta, d_k) \geq \mathcal{G}(\alpha, \beta, d_k) \quad \text{for all } (\alpha, \beta) \in \Sigma,$$

and if

$$\psi(\alpha_{k+1}, \beta_{k+1}, d_k) \leq \frac{1}{c} \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k),$$

where

$$(\alpha_{k+1}, \beta_{k+1}) := \operatorname{argmax}_{(\alpha, \beta) \in \Sigma} \left\{ \psi(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}.$$

PROPOSITION 3.3. *Let $\psi(\cdot, \cdot, d_k)$ be a c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at $(\alpha_k, \beta_k) \in \Sigma$, and $(\alpha_{k+1}, \beta_{k+1})$ be the unique solution of the approximate problem*

$$\max_{(\alpha, \beta) \in \Sigma} \left\{ \psi(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}.$$

Suppose that Assumption 2.7 is fulfilled. Then $\alpha_{k+1} \neq 0$.

Proof. From the definition of ψ given in Definition 3.2, we have

$$\mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \leq \psi(\alpha_{k+1}, \beta_{k+1}, d_k) \leq \frac{1}{c} \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k).$$

Since $c \in]0, 1[$, it follows that $\mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \geq 0$. This implies that

$$\alpha_{k+1}^\top [f(x) - d_k g(x)] + \beta_{k+1}^\top h(x) \geq 0 \quad \text{for all } x \in S.$$

If $\alpha_{k+1} = 0$, the last inequality becomes

$$\beta_{k+1}^\top h(x) \geq 0 \quad \text{for all } x \in S.$$

But this inequality cannot hold with $\hat{x} \in X$ of Assumption 2.7. So, $\alpha_{k+1} \neq 0$. \square

We are now ready to state our general approximating algorithm to solve the dual problem (D) .

Algorithm 3.1 General approximating algorithm.

0. Let $c \in]0, 1[$ be a given parameter. Choose a starting point $(\alpha_0, \beta_0) \in \Sigma$ such that $\alpha_0 \neq 0$ and $\eta_0 > 0$. Compute $d_0 := d(\alpha_0, \beta_0)$ and set $k = 0$.
 1. At step k , we have (α_k, β_k) and d_k . Then, construct a c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at (α_k, β_k) and find the unique maximum $(\alpha_{k+1}, \beta_{k+1})$ of the problem $(\mathcal{AD}_{\eta_k}(\alpha_k, \beta_k, d_k))$.
 2. Set $d_{k+1} = d(\alpha_{k+1}, \beta_{k+1})$, choose $\eta_{k+1} > 0$, increase k by 1, and go back to step 1.
-

3.1.1. Convergence of Algorithm 3.1. In this section we analyze the convergence properties of Algorithm 3.1. So, we will show that the sequence $\{(\alpha_k, \beta_k)\}$ converges to an optimal solution of problem (D) , and that the sequence $\{d_k\}$ increasingly converges to the minimal value of (P) . For this purpose, we shall provide the following lemmas, where we establish some preliminary results for analyzing convergence of the algorithm.

LEMMA 3.4. *Assume that Assumption 2.7 is fulfilled. Then,*

1. *the sequence $\{d_k\}$ is increasing and bounded from above by $\bar{\lambda}$, the optimal value of (P) ;*
2. *we have*

$$\begin{aligned} \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) &\rightarrow 0, \quad \psi(\alpha_k, \beta_k, d_k) \rightarrow 0, \\ \text{and} \quad \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 &\rightarrow 0 \end{aligned}$$

as $k \rightarrow 0$.

Proof. 1. Recall that we showed in the proof of Proposition 3.3 that

$$(3.1) \quad \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \geq 0.$$

This implies that

$$\alpha_{k+1}^\top [f(x) - d_k g(x)] + \beta_{k+1}^\top h(x) \geq 0 \quad \text{for all } x \in S.$$

Or equivalently, since $\alpha_{k+1} \neq 0$ from Proposition 3.3,

$$\frac{\alpha_{k+1}^\top f(x) + \beta_{k+1}^\top h(x)}{\alpha_{k+1}^\top g(x)} \geq d_k \quad \text{for all } x \in S.$$

Therefore, $d_{k+1} \geq d_k$ for all $k \in \mathbb{N}$.

Let us now prove that $d_k \leq \bar{\lambda}$. It is easy to show (see, e.g., [11, Lemma 3.1]) that

$$\frac{\alpha_k^\top f(x)}{\alpha_k^\top g(x)} \leq \max_{i \in I} \frac{f_i(x)}{g_i(x)} \quad \text{for all } x \in X.$$

Since $h(x) \leq 0$ for $x \in X$, the previous inequality gives

$$\frac{\alpha_k^\top f(x) + \beta_k^\top h(x)}{\alpha_k^\top g(x)} \leq \frac{\alpha_k^\top f(x)}{\alpha_k^\top g(x)} \leq \max_{i \in I} \frac{f_i(x)}{g_i(x)} \quad \text{for all } x \in X.$$

This implies that

$$\inf_{x \in X} \left\{ \frac{\alpha_k^\top f(x) + \beta_k^\top h(x)}{\alpha_k^\top g(x)} \right\} \leq \inf_{x \in X} \left\{ \max_{i \in I} \frac{f_i(x)}{g_i(x)} \right\} =: \bar{\lambda}.$$

Using the fact that $X \subset S$, we deduce from the previous inequality that $d_k \leq \bar{\lambda}$.

2. From the second inequality in Lemma 2.2, with $\mu = d_{k+1}$, $\lambda = d_k$, and $(\alpha, \beta) = (\alpha_{k+1}, \beta_{k+1})$, we get

$$\mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \leq \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_{k+1}) + (d_{k+1} - d_k) \Delta \sum_{i \in I} \alpha_{k+1}^i.$$

Since $\mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_{k+1}) = 0$ from the second point of Proposition 2.1, with (3.1) the last inequality implies that

$$(3.2) \quad 0 \leq \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \leq \Delta(d_{k+1} - d_k),$$

where the last inequality follows since $\sum_{i \in I} \alpha_{k+1}^i \leq 1$. This implies that

$$(3.3) \quad \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \rightarrow 0 \text{ when } k \rightarrow \infty.$$

On the other hand, from the definition of $(\alpha_{k+1}, \beta_{k+1})$ we have, for all $(\alpha, \beta) \in \Sigma$,

$$\begin{aligned} \psi(\alpha_{k+1}, \beta_{k+1}, d_k) - \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \\ \geq \psi(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2. \end{aligned}$$

For $(\alpha, \beta) = (\alpha_k, \beta_k)$ we get

$$\psi(\alpha_{k+1}, \beta_{k+1}, d_k) - \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \geq \psi(\alpha_k, \beta_k, d_k).$$

The second condition in the definition of a c -approximation gives

$$\begin{aligned} \frac{1}{c} \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) &\geq \psi(\alpha_{k+1}, \beta_{k+1}, d_k) \\ &\geq \psi(\alpha_{k+1}, \beta_{k+1}, d_k) - \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \\ &\geq \psi(\alpha_k, \beta_k, d_k) \\ &\geq 0, \end{aligned}$$

where the last inequality follows from the first condition in the definition of a c -approximation, and the fact that $\mathcal{G}(\alpha_k, \beta_k, d_k) = 0$. Therefore, by using (3.3), we obtain

$$\psi(\alpha_{k+1}, \beta_{k+1}, d_k) \rightarrow 0 \text{ and } \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \rightarrow 0 \quad \text{when } k \rightarrow \infty.$$

This achieves the proof. \square

We shall also need the following results.

LEMMA 3.5. *For all $k \in \mathbb{N}$, define*

$$\tau_k = \psi(\alpha_{k+1}, \beta_{k+1}, d_k) - 2\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2.$$

Then, we have

1. $\tau_k \geq 0$;
2. *for all $(\alpha, \beta) \in \Sigma$,*

$$\begin{aligned} & \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha, \beta)\|^2 \\ & \leq \frac{1}{\eta_k} \left[\tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 - \mathcal{G}(\alpha, \beta, d_k) \right] \\ & \quad + \|(\alpha_k, \beta_k) - (\alpha, \beta)\|^2; \end{aligned}$$

3. *the series $\sum_{k \geq 0} (\tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2)$ is convergent.*

Proof. 1. By definition we have

$$(\alpha_{k+1}, \beta_{k+1}) := \operatorname{argmax}_{(\alpha, \beta) \in \Sigma} \left\{ \psi(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}.$$

This implies (see, e.g., [18, Proposition 2.2, Page 37]) that, for all $(\alpha, \beta) \in \Sigma$,

$$(3.4) \quad \begin{aligned} -\psi(\alpha, \beta, d_k) & \geq -\psi(\alpha_{k+1}, \beta_{k+1}, d_k) \\ & \quad - 2\eta_k \langle (\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k), (\alpha, \beta) - (\alpha_{k+1}, \beta_{k+1}) \rangle. \end{aligned}$$

For $(\alpha, \beta) = (\alpha_k, \beta_k)$ we get

$$-\psi(\alpha_k, \beta_k, d_k) \geq -\psi(\alpha_{k+1}, \beta_{k+1}, d_k) + 2\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2.$$

Thus, since $\psi(\alpha_k, \beta_k, d_k) \geq 0$, we obtain $\tau_k \geq 0$.

2. From the definition of the c -approximation $\psi(\cdot, \cdot, d_k)$ we have

$$\psi(\alpha, \beta, d_k) \geq \mathcal{G}(\alpha, \beta, d_k) \quad \text{for all } (\alpha, \beta) \in \Sigma.$$

Including this inequality in (3.4) we obtain

$$\begin{aligned} \mathcal{G}(\alpha, \beta, d_k) & \leq \psi(\alpha_{k+1}, \beta_{k+1}, d_k) + 2\eta_k \langle (\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k), (\alpha, \beta) - (\alpha_{k+1}, \beta_{k+1}) \rangle \\ & = \psi(\alpha_{k+1}, \beta_{k+1}, d_k) - 2\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \\ & \quad + 2\eta_k \langle (\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k), (\alpha, \beta) - (\alpha_k, \beta_k) \rangle \\ & = \tau_k + 2\eta_k \langle (\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k), (\alpha, \beta) - (\alpha_k, \beta_k) \rangle. \end{aligned}$$

By using the equality

$$\begin{aligned} 2 \langle (\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k), (\alpha, \beta) - (\alpha_k, \beta_k) \rangle \\ = - \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha, \beta)\|^2 + \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 + \|(\alpha_k, \beta_k) - (\alpha, \beta)\|^2, \end{aligned}$$

we get

$$\begin{aligned}\mathcal{G}(\alpha, \beta, d_k) &\leq \tau_k - \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha, \beta)\|^2 + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \\ &\quad + \eta_k \|(\alpha_k, \beta_k) - (\alpha, \beta)\|^2.\end{aligned}$$

Rearranging terms, we obtain the desired inequality.

3. From the definition of τ_k we have

$$\tau_k + 2\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 = \psi(\alpha_{k+1}, \beta_{k+1}, d_k).$$

So, this equality, the definition of $\psi(\cdot, \cdot, d_k)$, and (3.2) give

$$\begin{aligned}\tau_k + 2\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 &\leq \frac{1}{c} \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \\ &\leq \frac{\Delta}{c} (d_{k+1} - d_k).\end{aligned}$$

Now, summing over $k = 0, \dots, q$, we get

$$\sum_{k=0}^q [\tau_k + 2\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2] \leq \frac{\Delta}{c} (d_{q+1} - d_0).$$

Since

$$0 \leq \tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \leq \tau_k + 2\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2,$$

we obtain the result. \square

In the next lemma we recall an important result which will be used later in the proof of our main results.

LEMMA 3.6. *Let $\{\mu_k\}$ and $\{\beta_k\}$ be two sequences of nonnegative reals such that*

$$\sum_{i=0}^{\infty} \mu_i < \infty, \quad \sum_{i=0}^{\infty} \beta_i < \infty,$$

and let $\{u_k\}$ be a sequence of reals such that

$$u_{k+1} \leq (1 + \mu_k)u_k + \beta_k.$$

Then, the sequence $\{u_k\}$ converges to some $u \in \mathbb{R} \cup \{-\infty\}$.

Proof. See, e.g., [45, Lemma 2] and [48, Lemma 2.1]. \square

With these key results, we may now state our principal convergence results.

THEOREM 3.7. *Suppose that $\sum_k 1/\eta_k = \infty$. If the assumptions of Theorem 2.9 are fulfilled, then the sequence $\{d_k\}$ converges to $\bar{\lambda}$, the optimal value of (P).*

Proof. First, we recall that we use the notation $d_k = d(\alpha_k, \beta_k)$. From Lemma 3.4 we know that the sequence $\{d_k\}$ is increasing and bounded from above by $\bar{\lambda}$, and hence it converges to some limit, say $\hat{d} \leq \bar{\lambda}$. Let us prove that $\hat{d} \geq \bar{\lambda}$.

From Lemma 2.2 with $\mu = \hat{d}$ and $\lambda = d_k$, we get, for all $(\alpha, \beta) \in \Sigma$ and $k \in \mathbb{N}$,

$$(\hat{d} - d_k) \delta \sum_{i \in I} \alpha_i \leq \mathcal{G}(\alpha, \beta, d_k) - \mathcal{G}(\alpha, \beta, \hat{d}) \leq (\hat{d} - d_k) \Delta \sum_{i \in I} \alpha_i.$$

This shows that $\mathcal{G}(\alpha, \beta, d_k) \rightarrow \mathcal{G}(\alpha, \beta, \hat{d})$ as $k \rightarrow \infty$.

On the other hand, point 3 of Lemma 3.5 implies that

$$\tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

In order to obtain a contradiction, we assume that for some $(\hat{\alpha}, \hat{\beta}) \in \Sigma$ we have $\mathcal{G}(\hat{\alpha}, \hat{\beta}, \hat{d}) > 0$. It follows that there exist $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ such that

$$(3.5) \quad \tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 - \mathcal{G}(\hat{\alpha}, \hat{\beta}, d_k) < -\varepsilon \quad \text{for all } k \geq k_0.$$

Writing the second part of Lemma 3.5 with $(\alpha, \beta) = (\hat{\alpha}, \hat{\beta})$, we get

$$\begin{aligned} \|(\alpha_{k+1}, \beta_{k+1}) - (\hat{\alpha}, \hat{\beta})\|^2 &\leq \|(\alpha_k, \beta_k) - (\hat{\alpha}, \hat{\beta})\|^2 \\ &\quad + \frac{1}{\eta_k} [\tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 - \mathcal{G}(\hat{\alpha}, \hat{\beta}, d_k)]. \end{aligned}$$

Using this inequality and (3.5), we obtain

$$\|(\alpha_{k+1}, \beta_{k+1}) - (\hat{\alpha}, \hat{\beta})\|^2 \leq \|(\alpha_k, \beta_k) - (\hat{\alpha}, \hat{\beta})\|^2 - \frac{1}{\eta_k} \varepsilon.$$

Summing the previous inequality from $k = k_0$ to $k = q$ yields

$$\sum_{k=k_0}^q \|(\alpha_{k+1}, \beta_{k+1}) - (\hat{\alpha}, \hat{\beta})\|^2 \leq \sum_{k=k_0}^q \|(\alpha_k, \beta_k) - (\hat{\alpha}, \hat{\beta})\|^2 - \varepsilon \sum_{k=k_0}^q \frac{1}{\eta_k}.$$

This implies that

$$\|(\alpha_{q+1}, \beta_{q+1}) - (\hat{\alpha}, \hat{\beta})\|^2 \leq \|(\alpha_{k_0}, \beta_{k_0}) - (\hat{\alpha}, \hat{\beta})\|^2 - \varepsilon \sum_{k=k_0}^q \frac{1}{\eta_k}.$$

Taking the limit when $q \rightarrow \infty$, we get a contradiction from the assumption that $\sum_{k \geq 0} 1/\eta_k = \infty$. Our assumption is therefore impossible, which implies that necessarily $\mathcal{G}(\alpha, \beta, \hat{d}) \leq 0$ for all $(\alpha, \beta) \in \Sigma$.

On the other hand, from the definition of $\mathcal{G}(\alpha, \beta, \hat{d})$, we have that for every $\varepsilon > 0$ there exists $x_\varepsilon \in S$ such that

$$(3.6) \quad \begin{aligned} \alpha^\top [f(x_\varepsilon) - \hat{d}g(x_\varepsilon)] + \beta^\top h(x_\varepsilon) &\leq \mathcal{G}(\alpha, \beta, \hat{d}) + \varepsilon \\ &\leq \varepsilon, \end{aligned}$$

where the last inequality follows from the fact that $\mathcal{G}(\alpha, \beta, \hat{d}) \leq 0$.

Thus, for every $(\alpha, \beta) \in \Sigma_*$, it follows from (3.6) and the inequality $\alpha^\top g(x) \geq \delta \sum_{i \in I} \alpha_i$ that

$$\begin{aligned} \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} &\leq \frac{\alpha^\top f(x_\varepsilon) + \beta^\top h(x_\varepsilon)}{\alpha^\top g(x_\varepsilon)} \\ &\leq \hat{d} + \frac{\varepsilon}{\alpha^\top g(x_\varepsilon)} \\ &\leq \hat{d} + \frac{\varepsilon}{\delta \sum_{i \in I} \alpha_i}. \end{aligned}$$

The inequality holds for all $\varepsilon > 0$, and so

$$\inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \leq \hat{d} \quad \text{for all } (\alpha, \beta) \in \Sigma.$$

Thus, this inequality with Theorem 2.9 gives

$$\bar{\lambda} = \sup_{(\alpha, \beta) \in \Sigma} \inf_{x \in S} \left\{ \frac{\alpha^\top f(x) + \beta^\top h(x)}{\alpha^\top g(x)} \right\} \leq \hat{d}.$$

Finally, $\hat{d} = \bar{\lambda}$. \square

THEOREM 3.8. Suppose that $\sum_k 1/\eta_k = \infty$. If the assumptions of Theorem 2.9 are fulfilled, then we have

1. every cluster point of the sequence $\{(\alpha_k, \beta_k)\}$ solves the dual problem (D) ;
2. if $\eta_k \geq \underline{\eta} > 0$, then the sequence $\{(\alpha_k, \beta_k)\}$ converges to an optimal solution of (D) .

Proof. 1. Recall first that the sequence $\{(\alpha_k, \beta_k)\}$ satisfies $\{(\alpha_k, \beta_k)\} \subset \Sigma$, and so it is bounded. Let $(\hat{\alpha}, \hat{\beta}) \in \Sigma$ be one of its cluster points.

Also, remember that by point 2 of Proposition 2.1, $\mathcal{G}(\alpha_k, \beta_k, d_k) = 0$, and that, from Assumption 2.3 and point 1 of Proposition 2.1, the function $(\alpha, \beta) \mapsto \mathcal{G}(\alpha, \beta, \bar{\lambda})$ is continuous as a finite concave function. So, by using Lemma 2.2 with $\mu = \bar{\lambda}$, $\lambda = d_k$, and $(\alpha, \beta) = (\alpha_k, \beta_k)$, and by passing to the limit on a subsequence, we show that $\mathcal{G}(\hat{\alpha}, \hat{\beta}, \bar{\lambda}) = 0$. Thus,

$$\hat{\alpha}^\top [f(x) - \bar{\lambda}g(x)] + \hat{\beta}^\top h(x) \geq 0 \quad \text{for all } x \in S.$$

This proves that necessarily $\hat{\alpha} \neq 0$ when Assumption 2.7 holds. Now, from point 2 of Proposition 2.4 and the fact that $\mathcal{G}(\hat{\alpha}, \hat{\beta}, \bar{\lambda}) = 0$, we conclude that $\bar{\lambda} = d(\hat{\alpha}, \hat{\beta})$, implying, by referring to Theorem 2.9, that $(\hat{\alpha}, \hat{\beta})$ is a solution of the dual (D) .

2. Let $(\hat{\alpha}, \hat{\beta})$ be a cluster point of the sequence $\{(\alpha_k, \beta_k)\}$. It follows from the previous point that $(\hat{\alpha}, \hat{\beta})$ is an optimal solution of (D) . From point 3 of Proposition 2.1, we have $\mathcal{G}(\hat{\alpha}, \hat{\beta}, d_k) \geq 0$, since $d_k = d(\alpha_k, \beta_k)$. On the other hand, from the second point of Lemma 3.5, with $(\alpha, \beta) = (\hat{\alpha}, \hat{\beta})$, we have

$$\begin{aligned} \|(\alpha_{k+1}, \beta_{k+1}) - (\hat{\alpha}, \hat{\beta})\|^2 &\leq \|(\alpha_k, \beta_k) - (\hat{\alpha}, \hat{\beta})\|^2 \\ &\quad + \frac{1}{\eta_k} [\tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 - \mathcal{G}(\hat{\alpha}, \hat{\beta}, d_k)] \\ &\leq \|(\alpha_k, \beta_k) - (\hat{\alpha}, \hat{\beta})\|^2 \\ &\quad + \frac{1}{\underline{\eta}} [\tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2]. \end{aligned}$$

Since the series $\sum_{k \geq 0} (\tau_k + \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2)$ is convergent, it follows from Lemma 3.6 that the sequence $\{\|(\alpha_k, \beta_k) - (\hat{\alpha}, \hat{\beta})\|^2\}$ is convergent, and since it has zero as a cluster point, it follows that the sequence $\{(\alpha_k, \beta_k)\}$ converges to $(\hat{\alpha}, \hat{\beta})$, which is an optimal solution of (D) . \square

3.1.2. Rate of convergence of Algorithm 3.1. Now, we will analyze the convergence rate of Algorithm 3.1 under the following assumption.

Assumption 3.9. There exist $r > 0$, $\kappa > 0$ such that

$$-\mathcal{G}(\alpha, \beta, \bar{\lambda}) \geq \kappa \text{dist}((\alpha, \beta), \bar{\Sigma})^2 \quad \text{for all } (\alpha, \beta) \in B(\bar{\Sigma}, r) \cap \Sigma,$$

where

$$\bar{\Sigma} = \operatorname{argmax}_{(\alpha, \beta) \in \Sigma} \mathcal{G}(\alpha, \beta, \bar{\lambda}), \quad \text{dist}((\alpha, \beta), \bar{\Sigma}) = \inf_{(\bar{\alpha}, \bar{\beta}) \in \bar{\Sigma}} \|(\bar{\alpha}, \bar{\beta}) - (\alpha, \beta)\|,$$

$$B((\bar{\alpha}, \bar{\beta}), r) = \{(\alpha, \beta) \in \mathbb{R}^m \times \mathbb{R}^p \mid \|(\alpha, \beta) - (\bar{\alpha}, \bar{\beta})\| \leq r\},$$

and

$$B(\bar{\Sigma}, r) = \bigcup_{(\bar{\alpha}, \bar{\beta}) \in \bar{\Sigma}} B((\bar{\alpha}, \bar{\beta}), r).$$

THEOREM 3.10. Suppose that $\sum_{k \geq 0} 1/\eta_k = \infty$, that $\eta_k \geq \underline{\eta} > 0$ for all $k \in \mathbb{N}$, and that the assumptions of Theorem 3.8 hold. Assume also that the function $\mathcal{G}(\cdot, \cdot, \bar{\lambda})$ satisfies Assumption 3.9. Then, for η_k sufficiently small, the sequence $\{d_k\}$ converges linearly to $\bar{\lambda}$.

Proof. From the definition of $(\alpha_{k+1}, \beta_{k+1})$ we have, for all $(\alpha, \beta) \in \Sigma$,

$$\begin{aligned} \psi(\alpha_{k+1}, \beta_{k+1}, d_k) - \eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2 \\ \geq \psi(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2. \end{aligned}$$

Using the definition of ψ we obtain, for all $(\alpha, \beta) \in \Sigma$,

$$(3.7) \quad \frac{1}{c} \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) \geq \mathcal{G}(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2.$$

From Theorem 3.8, the sequence $\{(\alpha_k, \beta_k)\}$ converges to some $(\bar{\alpha}, \bar{\beta}) \in \bar{\Sigma}$. It follows that $(\alpha_k, \beta_k) \in B(\bar{\Sigma}, r)$ for k large enough.

Now, let $(\hat{\alpha}_k, \hat{\beta}_k) \in \bar{\Sigma}$ be such that

$$\|(\hat{\alpha}_k, \hat{\beta}_k) - (\alpha_k, \beta_k)\| = \text{dist}((\alpha_k, \beta_k), \bar{\Sigma}).$$

By Assumption 3.9 we get

$$(3.8) \quad -\kappa \|(\hat{\alpha}_k, \hat{\beta}_k) - (\alpha_k, \beta_k)\|^2 \geq \mathcal{G}(\alpha_k, \beta_k, \bar{\lambda}).$$

Taking into account that the third point of Proposition 2.1 gives $\mathcal{G}(\hat{\alpha}_k, \hat{\beta}_k, \bar{\lambda}) = 0$, Lemma 2.2 with $\mu = \bar{\lambda}$, $\lambda = d_k$, and $(\alpha, \beta) = (\hat{\alpha}_k, \hat{\beta}_k)$ entails that

$$(3.9) \quad \mathcal{G}(\hat{\alpha}_k, \hat{\beta}_k, d_k) \geq (\bar{\lambda} - d_k) \delta \sum_{i \in I} \hat{\alpha}_k^i.$$

Again using Lemma 2.2, but with $\mu = \bar{\lambda}$, $\lambda = d_k$, and $(\alpha, \beta) = (\alpha_k, \beta_k)$, and recalling that $\mathcal{G}(\alpha_k, \beta_k, d_k) = 0$, we get

$$(3.10) \quad \mathcal{G}(\alpha_k, \beta_k, \bar{\lambda}) \geq (d_k - \bar{\lambda}) \Delta \sum_{i \in I} \alpha_k^i.$$

On the other hand, since $\mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_{k+1}) = 0$, Lemma 2.2 with $\mu = d_{k+1}$, $\lambda = d_k$, and $(\alpha, \beta) = (\alpha_{k+1}, \beta_{k+1})$ gives

$$(3.11) \quad (d_{k+1} - d_k) \Delta \sum_{i \in I} \alpha_{k+1}^i \geq \mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k).$$

For $(\alpha, \beta) = (\hat{\alpha}_k, \hat{\beta}_k)$ in (3.7), by using (3.8)–(3.11) we obtain

$$\frac{1}{c} (d_{k+1} - d_k) \Delta \sum_{i \in I} \alpha_{k+1}^i \geq (\bar{\lambda} - d_k) \delta \sum_{i \in I} \hat{\alpha}_k^i + (d_k - \bar{\lambda}) \frac{\eta_k \Delta}{\kappa} \sum_{i \in I} \alpha_k^i.$$

Or equivalently,

$$d_{k+1} - d_k \geq \frac{c\delta}{\Delta} \frac{\sum_{i \in I} \hat{\alpha}_k^i}{\sum_{i \in I} \alpha_{k+1}^i} (\bar{\lambda} - d_k) + \frac{c\eta_k}{\kappa} \frac{\sum_{i \in I} \alpha_k^i}{\sum_{i \in I} \alpha_{k+1}^i} (d_k - \bar{\lambda}).$$

This implies that

$$d_{k+1} - \bar{\lambda} \geq (d_k - \bar{\lambda}) \left(1 - \frac{c\delta}{\Delta} \frac{\sum_{i \in I} \hat{\alpha}_k^i}{\sum_{i \in I} \alpha_{k+1}^i} + \frac{c\eta_k}{\kappa} \frac{\sum_{i \in I} \alpha_k^i}{\sum_{i \in I} \alpha_{k+1}^i} \right).$$

Since $d_k < \bar{\lambda}$, for all $k \in \mathbb{N}$, the last inequality can be rewritten as

$$(3.12) \quad \frac{d_{k+1} - \bar{\lambda}}{d_k - \bar{\lambda}} \leq 1 - \frac{c\delta}{\Delta} \frac{\sum_{i \in I} \hat{\alpha}_k^i}{\sum_{i \in I} \alpha_{k+1}^i} + \frac{c\eta_k}{\kappa} \frac{\sum_{i \in I} \alpha_k^i}{\sum_{i \in I} \alpha_{k+1}^i}.$$

On the other hand, we have

$$\begin{aligned} \|(\hat{\alpha}_k, \hat{\beta}_k) - (\alpha_k, \beta_k)\| &= \text{dist}((\alpha_k, \beta_k), \bar{\Sigma}) \\ &:= \inf_{(\alpha, \beta) \in \bar{\Sigma}} \|(\alpha, \beta) - (\alpha_k, \beta_k)\| \\ &\leq \|(\alpha, \beta) - (\alpha_k, \beta_k)\| \quad \text{for all } (\alpha, \beta) \in \bar{\Sigma}. \end{aligned}$$

For $(\alpha, \beta) = (\bar{\alpha}, \bar{\beta})$, the limit of the sequence $\{(\alpha_k, \beta_k)\}$, we obtain

$$\|(\hat{\alpha}_k, \hat{\beta}_k) - (\alpha_k, \beta_k)\| \leq \|(\bar{\alpha}, \bar{\beta}) - (\alpha_k, \beta_k)\|.$$

This also shows that the sequence $\{(\hat{\alpha}_k, \hat{\beta}_k)\}$ converges to $(\bar{\alpha}, \bar{\beta})$. Therefore, referring to (3.12), we get

$$\limsup_{k \rightarrow \infty} \frac{d_{k+1} - \bar{\lambda}}{d_k - \bar{\lambda}} \leq 1 - \frac{c\delta}{\Delta} + \frac{c}{\kappa} \limsup_{k \rightarrow \infty} \eta_k.$$

Hence, if

$$\limsup_{k \rightarrow \infty} \eta_k < \frac{\delta\kappa}{\Delta},$$

the sequence $\{d_k\}$ converges at least linearly to $\bar{\lambda}$. \square

3.2. Construction of the c -approximations. Before describing the method of constructing c -approximation functions, we begin by defining, as in [59], strong c -approximations.

DEFINITION 3.11. Letting $c \in]0, 1[$ be a given parameter, a concave function $\psi^k(\cdot, \cdot, d_k)$ is a strong c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at $(\alpha_k, \beta_k) \in \Sigma$ if

$$\psi^k(\alpha, \beta, d_k) \geq \mathcal{G}(\alpha, \beta, d_k) \quad \text{for all } (\alpha, \beta) \in \Sigma$$

and if

$$\mathcal{G}(\alpha_{k+1}, \beta_{k+1}, d_k) - \psi^k(\alpha_{k+1}, \beta_{k+1}, d_k) \geq 2(c-1)\eta_k \|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|^2,$$

where

$$(\alpha_{k+1}, \beta_{k+1}) := \operatorname{argmax}_{(\alpha, \beta) \in \Sigma} \left\{ \psi^k(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}.$$

Remark 3.12. It is easy to see that if $\psi^k(\cdot, \cdot, d_k)$ is a strong c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at $(\alpha_k, \beta_k) \in \Sigma$, then it is a c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at $(\alpha_k, \beta_k) \in \Sigma$.

Instead of directly solving the problem $(\mathcal{D}_{\eta_k}(\alpha_k, \beta_k, d_k))$, a set of approximating and easier problems, indexed by $\ell = 1, \dots, \ell(k)$,

$$(3.13) \quad \max_{(\alpha, \beta) \in \Sigma} \left\{ \psi^\ell(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}$$

will be solved until an approximate solution $(\alpha_k^{\ell(k)}, \beta_k^{\ell(k)}) \in \Sigma$ of the subproblem $(\mathcal{D}_{\eta_k}(\alpha_k, \beta_k, d_k))$ is reached. Then iteration $k+1$ will be performed by approximately solving problem $(\mathcal{D}_{\eta_{k+1}}(\alpha_{k+1}, \beta_{k+1}, d_{k+1}))$ with $(\alpha_{k+1}, \beta_{k+1}) = (\alpha_k^{\ell(k)}, \beta_k^{\ell(k)})$ and $d_{k+1} = d(\alpha_{k+1}, \beta_{k+1})$ by the same procedure.

To obtain the approximate solution $(\alpha_k^{\ell(k)}, \beta_k^{\ell(k)})$ of $(\mathcal{D}_{\eta_{k+1}}(\alpha_{k+1}, \beta_{k+1}, d_{k+1}))$ one may construct, at each iteration ℓ , an approximation $\psi^\ell(\cdot, \cdot, d_k)$ of $\mathcal{G}(\cdot, \cdot, d_k)$ and solve the approximating problem (3.13) to obtain the solution $(\alpha_k^\ell, \beta_k^\ell)$. The procedure then stops with ℓ such that $\psi^\ell(\cdot, \cdot, d_k)$ is a (strong) c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at (α_k, β_k) , and we set $\ell(k) = \ell$.

For $\ell = 1, 2, \dots$, we define the affine function $\mathcal{L}_{k,\ell}(\cdot, \cdot)$, on Σ , by

$$\mathcal{L}_{k,\ell}(\alpha, \beta) = \psi^\ell(\alpha_k^\ell, \beta_k^\ell, d_k) - 2\eta_k \langle (\alpha_k^\ell, \beta_k^\ell), (\alpha, \beta) - (\alpha_k^\ell, \beta_k^\ell) \rangle.$$

Next, we discuss the construction of c -approximation functions of $\mathcal{G}(\cdot, \cdot, d_k)$. To guarantee their existence, these functions must satisfy the following classical properties:

- (C1) $\psi^\ell(\alpha, \beta, d_k) \geq \mathcal{G}(\alpha, \beta, d_k)$ for all $\ell = 1, \dots, \ell(k)$ and all $(\alpha, \beta) \in \Sigma$;
- (C2) $\psi^{\ell+1}(\alpha, \beta, d_k) \leq \mathcal{L}_{k,\ell}(\alpha, \beta)$ for all $\ell = 1, \dots, \ell(k)$ and all $(\alpha, \beta) \in \Sigma$;
- (C3) $\psi^{\ell+1}(\alpha, \beta, d_k) \leq \mathcal{G}(\alpha_k^\ell, \beta_k^\ell, d_k) + \langle s_k^\ell, (\alpha, \beta) - (\alpha_k^\ell, \beta_k^\ell) \rangle + \varepsilon_k^\ell$ for all $\ell = 1, \dots, \ell(k)$ and all $(\alpha, \beta) \in \Sigma$, where s_k^ℓ is any ε_k^ℓ -subgradient of $-\mathcal{G}(\cdot, \cdot, d_k)$ at $(\alpha_k^\ell, \beta_k^\ell)$.

In the following we give some possible choices of $\psi^\ell(\cdot, \cdot, d_k)$.

Example 3.1. Consider the piecewise-affine model, defined for all $k \in \mathbb{N}$ and $\ell \in \mathbb{N}$, by

$$\psi^{\ell+1}(\alpha, \beta, d_k) = \min_{0 \leq q \leq \ell} \{ \mathcal{G}(\alpha_k^q, \beta_k^q, d_k) + \langle s_k^q, (\alpha, \beta) - (\alpha_k^q, \beta_k^q) \rangle + \varepsilon_k^q \}$$

for all $(\alpha, \beta) \in \Sigma$, where $(\alpha_k^0, \beta_k^0) = (\alpha_k, \beta_k)$.

Example 3.2. For all $k \in \mathbb{N}$ and $\ell \in \mathbb{N}$, we can choose, for all $(\alpha, \beta) \in \Sigma$,

$$\psi^{\ell+1}(\alpha, \beta, d_k) = \min \left\{ \mathcal{L}_{k,\ell}(\alpha, \beta), \mathcal{G}(\alpha_k^\ell, \beta_k^\ell, d_k) + \langle s_k^\ell, (\alpha, \beta) - (\alpha_k^\ell, \beta_k^\ell) \rangle + \varepsilon_k^\ell \right\},$$

where $(\alpha_k^0, \beta_k^0) = (\alpha_k, \beta_k)$.

Example 3.3. For all $k \in \mathbb{N}$, $\ell \in \mathbb{N}$, and $(\alpha, \beta) \in \Sigma$, let

$$\begin{aligned} \psi^{\ell+1}(\alpha, \beta, d_k) \\ := \min \left\{ \mathcal{L}_{k,\ell}(\alpha, \beta), \min_{0 \leq q \leq \ell} \left\{ \mathcal{G}(\alpha_k^q, \beta_k^q, d_k) + \langle s_k^q, (\alpha, \beta) - (\alpha_k^q, \beta_k^q) \rangle + \varepsilon_k^q \right\} \right\}, \end{aligned}$$

where $(\alpha_k^0, \beta_k^0) = (\alpha_k, \beta_k)$.

Below we describe a procedure to construct a strong c -approximation, at an iteration k , of the function $\mathcal{G}(\cdot, \cdot, d_k)$ at (α_k, β_k) . This also permits the construction of a c -approximation, following Remark 3.12.

Algorithm 3.2 Construction of the c -approximation.

0. Let $c \in]0, 1[$ be a given parameter, $(\alpha_k, \beta_k) \in \Sigma$, $\alpha_k \neq 0$, and $\ell = 1$.
1. Choose a piecewise linear concave function $\psi^\ell(\cdot, \cdot, d_k)$ satisfying (C1)–(C3). Determine $(\alpha_k^\ell, \beta_k^\ell) \in \Sigma$ as the maximizer of

$$\max_{(\alpha, \beta) \in \Sigma} \left\{ \psi^\ell(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}.$$

Let $x_k^\ell \in S$ and $\varepsilon_k^\ell \geq 0$ be such that

$$(\alpha_k^\ell)^\top [f(x_k^\ell) - d_k g(x_k^\ell)] + (\beta_k^\ell)^\top h(x_k^\ell) \leq \mathcal{G}(\alpha_k^\ell, \beta_k^\ell, d_k) + \varepsilon_k^\ell.$$

Compute the ε_k^ℓ -subgradient of $-\mathcal{G}(\cdot, \cdot, d_k)$ at $(\alpha_k^\ell, \beta_k^\ell)$,

$$s_k^\ell = - \begin{pmatrix} f(x_k^\ell) - d_k g(x_k^\ell) \\ h(x_k^\ell) \end{pmatrix}$$

2. If

$$\mathcal{G}(\alpha_k^\ell, \beta_k^\ell, d_k) - \psi^\ell(\alpha_k^\ell, \beta_k^\ell, d_k) \geq 2(c-1)\eta_k \|(\alpha_k^\ell, \beta_k^\ell) - (\alpha_k, \beta_k)\|^2,$$

then stop, set $\ell(k) = \ell$, and $\psi_k^{\ell(k)}(\cdot, \cdot, d_k)$ is a strong c -approximation of $\mathcal{G}(\cdot, \cdot, d_k)$ at (α_k, β_k) . Otherwise, increase ℓ by 1 and go back to step 1.

Now we present results to show the finite termination of Algorithm 3.2.

THEOREM 3.13. *Suppose that the models ψ^ℓ satisfy conditions (C1)–(C3). Let*

$$(\bar{\alpha}_k^\ell, \bar{\beta}_k^\ell) := \operatorname{argmax}_{(\alpha, \beta) \in \Sigma} \left\{ \psi^\ell(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}$$

and

$$(\bar{\alpha}_k, \bar{\beta}_k) := \operatorname{argmax}_{(\alpha, \beta) \in \Sigma} \left\{ \mathcal{G}(\alpha, \beta, d_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}.$$

If $\ell(k) = \infty$ and $\varepsilon_k^\ell \rightarrow 0$ when $\ell \rightarrow \infty$, then

1. $\psi^\ell(\alpha_k^\ell, \beta_k^\ell, d_k) \rightarrow 0$ as $\ell \rightarrow \infty$;
2. $\mathcal{G}(\alpha_k^\ell, \beta_k^\ell, d_k) \rightarrow 0$ as $\ell \rightarrow \infty$;
3. $(\alpha_k^\ell, \beta_k^\ell) \rightarrow (\bar{\alpha}_k, \bar{\beta}_k)$ as $\ell \rightarrow \infty$ and $(\bar{\alpha}_k, \bar{\beta}_k) = (\alpha_k, \beta_k)$;
4. (α_k, β_k) is an optimal solution of (D) and $d_k = \bar{\lambda}$, the optimal value of (P).

Therefore, if $(\bar{\alpha}_k, \bar{\beta}_k) \neq (\alpha_k, \beta_k)$, then $\ell(k) < \infty$.

Proof. The results may be checked by adapting some standard results and techniques from, e.g., [12, 2, 8, 9]. \square

At this stage, we can describe the complete proximal bundle algorithm by inserting the procedure to construct (strong) c -approximations, detailed in Algorithm 3.2, in the general scheme described in Algorithm 3.1. But before doing so, recall that we need to choose a set S as in (2.1), accordingly define the functions \mathcal{G} and d , and to check Assumptions 2.3 and 2.6 and the hypotheses on δ and Δ .

Now we summarize our proximal bundle dual algorithm.

Algorithm 3.3 Proximal bundle dual algorithm.

1. Let $(\alpha_0, \beta_0) \in \Sigma$, and compute $\lambda_0 = d(y_0, z_0)$. Let $x_0 \in S$ and $\varepsilon_0 \geq 0$ be such that

$$\alpha_0^\top [f(x_0) - \lambda_0 g(x_0)] + \beta_0^\top h(x) \leq \mathcal{G}(\alpha_0, \beta_0, \lambda_0) + \varepsilon_0.$$

Set

$$s_0^0 := - \begin{pmatrix} f(x_0) - \lambda_0 g(x_0) \\ h(x_0) \end{pmatrix},$$

$k = 0$, and $\ell = 1$.

2. At step k , we have $\eta_k > 0$, (α_k, β_k) and λ_k . Choose $\psi^\ell(\cdot, \cdot, \lambda_k)$ a concave piecewise linear function that satisfies (C1)–(C3) and solve the problem

$$\max_{(\alpha, \beta) \in \Sigma} \left\{ \psi^\ell(\alpha, \beta, \lambda_k) - \eta_k \|(\alpha, \beta) - (\alpha_k, \beta_k)\|^2 \right\}$$

to get $(\alpha_k^\ell, \beta_k^\ell)$. Let x_k^ℓ and $\varepsilon_k^\ell \geq 0$ be such that

$$(\alpha_k^\ell)^\top [f(x_k^\ell) - \lambda_k g(x_k^\ell)] + (\beta_k^\ell)^\top h(x_k^\ell) \leq \mathcal{G}(\alpha_k^\ell, \beta_k^\ell, \lambda_k) + \varepsilon_k^\ell,$$

and compute

$$s_k^\ell := - \begin{pmatrix} f(x_k^\ell) - \lambda_k g(x_k^\ell) \\ h(x_k^\ell) \end{pmatrix}.$$

3. If

$$\psi^\ell(\alpha_k^\ell, \beta_k^\ell, \lambda_k) \leq \frac{1}{c} \mathcal{G}(\alpha_k^\ell, \beta_k^\ell, \lambda_k),$$

then set $(\alpha_{k+1}, \beta_{k+1}) = (\alpha_k^\ell, \beta_k^\ell)$, $\ell(k) = \ell$, and $(\alpha_{k+1}^0, \beta_{k+1}^0) = (\alpha_{k+1}, \beta_{k+1})$. Compute $\lambda_{k+1} = d(\alpha_{k+1}, \beta_{k+1})$, increase k by 1 and set $\ell = 0$.

4. Increase ℓ by 1 and go to step 2.
-

4. Conclusions. We have proposed a new dual, and related duality results, for generalized fractional programs that are not necessarily convex. For this dual problem, we have associated dual parametric programs. By approximating the dual functions of these dual programs, by means of c -approximation functions, we have proposed a dual inexact proximal point algorithm, which acts as a general approximating scheme, for solving this dual program.

As an application of this general scheme, we have constructed bundle methods by using the convexity property of the dual functions of the parametric problems. The most important effort at this stage is the calculation of the values and subgradients of these dual functions. We have suggested (and this is well known in the recent literature dealing with the subject) calculating only approximately the values and the subgradients, up to $\varepsilon_k^\ell \geq 0$. As shown by Theorem 3.13, the only requirement is that $\varepsilon_k^\ell \rightarrow 0$ as $\ell \rightarrow \infty$ for each $k \in \mathbb{N}$. Furthermore, if an optimal solution is not reached at an iteration k , a c -approximation is found after a finite number of iterations $\ell(k)$. Thus, the construction of a c -approximation at each iteration k is performed by solving a finite number of problems like (3.13) that are equivalent to quadratic programs with simple constraints. The latter can be solved efficiently by appropriate methods, such as the one given in [34].

Acknowledgments. The authors would like to thank the Associate Editor and the two anonymous referees for their careful reading of the manuscript and their valuable remarks.

REFERENCES

- [1] A. ADDOU AND A. ROUBI, *Proximal-type methods with generalized Bregman functions and applications to generalized fractional programming*, Optimization, 59 (2010), pp. 1085–1105, <https://doi.org/10.1080/02331930903395857>.
- [2] S. ADDOUINE, K. BOUFI, AND A. ROUBI, *Proximal bundle algorithms for nonlinearly constrained convex minimax fractional programs*, J. Optim. Theory Appl., 179 (2018), pp. 212–239, <https://doi.org/10.1007/s10957-018-1342-1>.
- [3] S. ADDOUINE, M. EL HAFFARI, AND A. ROUBI, *A proximal point algorithm for generalized fractional programs*, Optimization, 66 (2017), pp. 1495–1517, <https://doi.org/10.1080/02331934.2017.1338698>.
- [4] A. I. BARROS, J. B. G. FRENK, S. SCHAIBLE, AND S. ZHANG, *A new algorithm for generalized fractional programs*, Math. Program., 72 (1996), pp. 147–175, <https://doi.org/10.1007/BF02592087>.
- [5] A. I. BARROS, J. B. G. FRENK, S. SCHAIBLE, AND S. ZHANG, *Using duality to solve generalized fractional programming problems*, J. Global Optim., 8 (1996), pp. 139–170, <https://doi.org/10.1007/BF00138690>.
- [6] C. R. BECTOR, S. CHANDRA, AND M. K. BECTOR, *Generalized fractional programming duality: A parametric approach*, J. Optim. Theory Appl., 60 (1989), pp. 243–260, <https://doi.org/10.1007/BF00940006>.
- [7] J. C. BERNARD AND J. A. FERLAND, *Convergence of interval-type algorithms for generalized fractional programming*, Math. Program., 43 (1989), pp. 349–363, <https://doi.org/10.1007/BF01582298>.
- [8] H. BOUALAM AND A. ROUBI, *Dual algorithms based on the proximal bundle method for solving convex minimax fractional programs*, J. Ind. Manag. Optim., 13 (2017), <https://doi.org/10.3934/jimo.2018128>.
- [9] H. BOUALAM AND A. ROUBI, *Proximal bundle methods based on approximate subgradients for solving Lagrangian duals of minimax fractional programs*, J. Global Optim. (2019), <https://doi.org/10.1007/s10898-019-00757-2>.
- [10] K. BOUFI AND A. ROUBI, *Prox-regularization of the dual method of centers for generalized fractional programs*, Optim. Methods Softw., 34 (2019), pp. 515–545, <https://doi.org/10.1080/10556788.2017.1392520>.
- [11] K. BOUFI AND A. ROUBI, *Dual method of centers for solving generalized fractional programs*, J. Global Optim., 69 (2017), pp. 387–426, <https://doi.org/10.1007/s10898-017-0523-z>.

- [12] R. CORREA AND C. LEMARÉCHAL, *Convergence of some algorithms for convex minimization*, Math. Program., 62 (1993), pp. 261–275, <https://doi.org/10.1007/BF01585170>.
- [13] J.-P. CROUZEIX AND J. A. FERLAND, *Algorithms for generalized fractional programming*, Math. Program., 52 (1991), pp. 191–207, <https://doi.org/10.1007/BF01582887>.
- [14] J.-P. CROUZEIX, J. A. FERLAND, AND H. V. NGUYEN, *Revisiting Dinkelbach-type algorithms for generalized fractional programs*, Opsearch, 45 (2008), pp. 97–110, <https://doi.org/10.1007/BF03398807>.
- [15] J.-P. CROUZEIX, J. A. FERLAND, AND S. SCHAIBLE, *Duality in generalized linear fractional programming*, Math. Program., 27 (1983), pp. 342–354, <https://doi.org/10.1007/BF02591908>.
- [16] J.-P. CROUZEIX, J. A. FERLAND, AND S. SCHAIBLE, *An algorithm for generalized fractional programs*, J. Optim. Theory Appl., 47 (1985), pp. 35–49, <https://doi.org/10.1007/BF00941314>.
- [17] J.-P. CROUZEIX, J. A. FERLAND, AND S. SCHAIBLE, *A note on an algorithm for generalized fractional programs*, J. Optim. Theory Appl., 50 (1986), pp. 183–187, <https://doi.org/10.1007/BF00938484>.
- [18] I. EKELAND AND R. TEMAM, *Analyse Convexe et Problèmes Variationnels*, Gauthier-Villars, Paris, 1974.
- [19] M. EL HAFFARI AND A. ROUBI, *Convergence of a proximal algorithm for solving the dual of a generalized fractional program*, RAIRO Oper. Res., 51 (2017), pp. 985–1004, <https://doi.org/10.1051/ro/2017004>.
- [20] M. EL HAFFARI AND A. ROUBI, *Prox-dual regularization algorithm for generalized fractional programs*, J. Ind. Manag. Optim., 13 (2017), pp. 1991–2013, <https://doi.org/10.3934/jimo.2017028>.
- [21] J. E. FALK, *Maximization of signal-to-noise ratio in an optical filter*, SIAM J. Appl. Math., 17 (1969), pp. 582–592.
- [22] K. FAN, *Minimax theorems*, Proc. Nat. Acad. Sci. USA, 39 (1953), pp. 42–47.
- [23] J. B. G. FRENK, G. KASSAY, AND J. KOLUMBÁN, *On equivalent results in minimax theory*, European J. Oper. Res., 157 (2004), pp. 46–58, <https://doi.org/10.1016/j.ejor.2003.08.013>.
- [24] J. B. G. FRENK AND S. SCHAIBLE, *Fractional Programming*, ERIM Report Series ERS-2004-074-LIS, 2004; available at <https://ssrn.com/abstract=595012>.
- [25] M. FUKUSHIMA, *A descent algorithm for nonsmooth convex optimization*, Math. Program., 30 (1984), pp. 163–175, <https://doi.org/10.1007/BF02591883>.
- [26] M. GUGAT, *Prox-regularization methods for generalized fractional programming*, J. Optim. Theory Appl., 99 (1998), pp. 691–722, <https://doi.org/10.1023/A:1021759318653>.
- [27] O. GÜLER, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., 29 (1991), pp. 403–419, <https://doi.org/10.1137/0329022>.
- [28] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms I*, Springer, New York, 1993.
- [29] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms II*, Springer, New York, 1993.
- [30] R. JAGANNATHAN AND S. SCHAIBLE, *Duality in generalized fractional programming via Farkas' lemma*, J. Optim. Theory Appl., 41 (1983), pp. 417–424, <https://doi.org/10.1007/BF00935361>.
- [31] G. KASSAY, *A simple proof of König's minimax theorems*, Acta Math. Hungar., 63 (1994), pp. 371–374.
- [32] K. C. KIWIĘL, *An aggregate subgradient method for nonsmooth convex minimization*, Math. Program., 27 (1983), pp. 320–341, <https://doi.org/10.1007/BF02591907>.
- [33] K. C. KIWIĘL, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Math. 1133, Springer, Berlin, 1985.
- [34] K. C. KIWIĘL, *A method for solving certain quadratic programming problems arising in nonsmooth optimization*, IMA J. Numer. Anal., 6 (1986), pp. 137–152, <https://doi.org/10.1093/imanum/6.2.137>.
- [35] K. C. KIWIĘL, *Proximity control in bundle methods for convex nondifferentiable minimization*, Math. Program., 46 (1990), pp. 105–122, <https://doi.org/10.1007/BF01585731>.
- [36] C. LEMARÉCHAL, *Bundle methods in nonsmooth optimization*, in *Nonsmooth Optimization*, C. Lemaréchal and R. Mifflin, eds., Pergamon Press, Oxford, 1978, pp. 79–102.
- [37] C. LEMARÉCHAL, *Constructing bundle methods for convex optimization*, in *Fermat Days 85: Mathematics for Optimization*, J. B. Hiriart-Urruty, ed., North-Holland, Amsterdam, 1986, pp. 201–240.
- [38] B.-L. LIN AND C.-Z. CHENG, *A minimax theorems involving weakly downward functions*, Acta Math. Hungar., 87 (2000), pp. 287–293, <https://doi.org/10.1023/A:1006721718184>.
- [39] M. MÄKELÄ, *Survey of bundle methods for nonsmooth optimization*, Optim. Methods Softw.,

- 17 (2002), pp. 1–29, <https://doi.org/10.1080/10556780290027828>.
- [40] B. MARTINET, *Régularisation d'Inéquations Variationnelles par Approximation Successives*, Rev. Fr. Inform. Rech. O., 4 (1970), pp. 154–158.
 - [41] R. MIFFLIN, *An algorithm for constrained optimization with semismooth functions*, Math. Oper. Res., 2 (1977), pp. 191–207, <https://doi.org/10.1287/moor.2.2.191>.
 - [42] R. MIFFLIN, *A modification and extension of Lemaréchal's algorithm for nonsmooth minimization*, Math. Program. Stud., 17 (1982), pp. 77–90.
 - [43] J.-J. MOREAU, *Proximité et Dualité dans un Espace Hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
 - [44] A. NAGIH AND G. PLATEAU, *Problèmes Fractionnaires: Tour d'Horizon sur les Applications et Méthodes de Résolution*, RAIRO Oper. Res., 33 (1999), pp. 383–419, <https://doi.org/10.1051/ro:1999118>.
 - [45] B. T. POLYAK, *Introduction to Optimization*, Transl. Ser. Math. Engrg., Optimization Software, New York, 1987.
 - [46] R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898, <https://doi.org/10.1137/0314056>.
 - [47] A. ROUBI, *Method of centers for generalized fractional programming*, J. Optim. Theory Appl., 107 (2000), pp. 123–143, <https://doi.org/10.1023/A:1004660917684>.
 - [48] A. ROUBI, *Convergence of prox-regularization methods for generalized fractional programming*, RAIRO Oper. Res., 36 (2002), pp. 73–94, <https://doi.org/10.1051/ro:2002006>.
 - [49] S. SCHABIEBLE, *Fractional programming*, in *Handbook of Global Optimization*, R. Horst and P. M. Pardalos, eds., Kluwer, Dordrecht, 1995, pp. 495–608.
 - [50] H. SCHRAMM AND J. ZOWE, *A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results*, SIAM J. Optim., 2 (1992), pp. 121–152, <https://doi.org/10.1137/0802008>.
 - [51] S. SIMONS, *An upward-downward minimax theorem*, Arch. Math. (Basel), 55 (1990), pp. 275–279, <https://doi.org/10.1007/BF01191168>.
 - [52] S. SIMONS, *Minimax theorems and their proofs*, in *Minimax and Applications, Nonconvex Optim. Appl.* 4, D. Z. Du and P. M. Pardalos, eds., Springer, Boston, MA, 1995, pp. 1–23.
 - [53] M. SION, *On general minimax theorems*, Pac. J. Optim., 8 (1958), pp. 171–176, <https://doi.org/10.2140/pjm.1958.8.171>.
 - [54] A. M. STANCU, *Mathematical Programming with Type-I Functions*, MatrixRom, Romania, Bucharest, 2013.
 - [55] I. M. STANCU-MINASIAN, *Fractional Programming: Theory, Methods and Applications*, Kluwer Academic, Dordrecht, 1997.
 - [56] I. M. STANCU-MINASIAN, *A sixth bibliography of fractional programming*, Optimization, 55 (2006), pp. 405–428, <https://doi.org/10.1080/02331930600819613>.
 - [57] I. M. STANCU-MINASIAN, *A seventh bibliography of fractional programming*, Adv. Model. Optim., 15 (2013), pp. 309–386.
 - [58] I. M. STANCU-MINASIAN, *A eighth bibliography of fractional programming*, Optimization, 66 (2017), pp. 439–470, <https://doi.org/10.1080/02331934.2016.1276179>.
 - [59] J.-J. STRODIOT, J.-P. CROUZEIX, J. A. FERLAND, AND V. H. NGUYEN, *An inexact proximal point method for solving generalized fractional programs*, J. Global Optim., 42 (2008), pp. 121–138, <https://doi.org/10.1007/s10898-007-9270-x>.
 - [60] Z. XU, *Duality in generalized nonlinear fractional programming*, J. Math. Anal. Appl., 169 (1992), pp. 1–9, [https://doi.org/10.1016/0022-247X\(92\)90099-Y](https://doi.org/10.1016/0022-247X(92)90099-Y).
 - [61] G. ZALMAI, *Duality for generalized fractional programs involving n-set functions*, J. Math. Anal. Appl., 149 (1990), pp. 339–350, [https://doi.org/10.1016/0022-247X\(90\)90046-I](https://doi.org/10.1016/0022-247X(90)90046-I).