J Glob Optim
DOI 10.1007/s10898-007-9270-x

An inexact proximal point method for solving generalized fractional programs

Jean-Jacques Strodiot · Jean-Pierre Crouzeix · Jacques A. Ferland · Van Hien Nguyen

Received: 24 October 2007 / Accepted: 10 December 2007

Abstract In this paper, we present several new implementable methods for solving a generalized fractional program with convex data. They are Dinkelbach-type methods where a prox-regularization term is added to avoid the numerical difficulties arising when the solution of the problem is not unique. In these methods, at each iteration a regularized parametric problem is solved inexactly to obtain an approximation of the optimal value of the problem. Since the parametric problem is nonsmooth and convex, we propose to solve it by using a classical bundle method where the parameter is updated after each 'serious step'. We mainly study two kinds of such steps, and we prove the convergence and the rate of convergence of each of the corresponding methods. Finally, we present some numerical experience to illustrate the behavior of the proposed algorithms, and we discuss the practical efficiency of each one.

Keywords Fractional programming · Dinkelbach algorithms · Proximal point methods · Bundle methods

AMS Classification 90C32 · 90C26

J.-J. Strodiot (✉) · V. H. Nguyen
Department of Mathematics, University of Namur (FUNDP), Namur, Belgium
e-mail: jean-jacques.strodiot@fundp.ac.be

V. H. Nguyen
e-mail: vhnguyen@fundp.ac.be

J.-P. Crouzeix
LIMOS, CNRS-UMR 6158, Université Blaise Pascal, Clermont-Ferrand, France
e-mail: jp.crouzeix@isima.fr

J. A. Ferland
DIRO, Université de Montréal, Montreal, QC, Canada
e-mail: ferland@iro.umontreol.ca

1 Introduction

Consider the generalized fractional programming problem

$$(P)\quad \lambda^* = \inf_{x\in X} \left\{ \lambda(x) = \max_{1\le i\le p} \left\{ \frac{f_i(x)}{g_i(x)} \right\} \right\},$$

where $X \subseteq I\!R^n$ is nonempty, $f_i, g_i : X \to I\!R$ are continuous for all $1 \le i \le p$ and $g_i(x) > 0$ for all $x \in X$ and $1 \le i \le p$. We do not assume that λ^* is finite nor that (P) has optimal solutions. Crouzeix et al. [4,5] proposed two Dinkelbach-type algorithms based on the idea of solving a sequence of auxiliary parametric problems having a simpler structure. So they obtained two sequences. The first one is converging to λ^* and the second one to a solution of (P) if (P) has at least some solution. More precisely they first consider the parametric problem

$$(P_\lambda)\qquad F(\lambda) = \inf_{x\in X} F(x,\lambda),$$

where λ is a real parameter and

$$F(x,\lambda) = \max_{1\le i\le p} \{f_i(x) - \lambda g_i(x)\}. \tag{1}$$

In particular, they prove that if (P) has a solution, then $F(\lambda^*) = 0$, and that if $F(\lambda^*) = 0$, then (P) and (P_{λ^*}) have the same set (possibly empty) of solutions. The corresponding algorithm is as follows: given $x^k \in X$, first find λ_k such that $F(x^k, \lambda_k) = 0$, and then find x^{k+1} a solution of problem (P_{λ_k}). It is easy to see that

$$F(x^k,\lambda_k) = 0 \quad \Leftrightarrow \quad \lambda_k = \max_{1\le i\le p} \frac{f_i(x^k)}{g_i(x^k)}.$$

In [4], it is proven that if (P) has a solution, if each subproblem (P_{λ_k}) has a solution, and if $g_i(x) \le \gamma$ for all $x \in X$ and $1 \le i \le p$, then the sequence $\{\lambda_k\}$ converges linearly to λ^*. Moreover, if X is compact, every limit point of $\{x^k\}$ is a solution of (P). To obtain the superlinear convergence of the sequence $\{\lambda_k\}$, the algorithm has been modified in [5] by introducing a normalization vector parameter. The function $F(x,\lambda)$ defined in (1) has been replaced by the function

$$F(x,w,\lambda) = \max_{1\le i\le p} \left\{ \frac{f_i(x) - \lambda g_i(x)}{w_i} \right\}, \tag{2}$$

where $w \in I\!R^p$, $w_i > 0$ for all i. The algorithm becomes: given $x^k \in X$ and $w_i^k > 0$, $i = 1,\ldots,p$, find λ_k such that $F(x^k, w^k, \lambda_k) = 0$. Then compute x^{k+1} as a solution of problem (P_{w^k,λ_k}) where

$$(P_{w^k,\lambda_k})\qquad F_{w^k}(\lambda_k) = \inf_{x\in X} F(x, w^k, \lambda_k).$$

In [5] the authors use the specific normalization $w_i^k = g_i(x^k), i = 1,\ldots,p$ at iteration k to prove that the sequence $\{\lambda_k\}$ converges superlinearly to λ^* when X is compact and the sequence $\{x^k\}$ is converging to a solution of (P).

If the solution of (P) is not unique (see an example in [8]), it can happen that the solution of problem (P_{w^k,λ_k}) is also not unique causing difficulties in the numerical solution of this problem. The case when X is not compact is also a source of numerical difficulties. On the other hand, the performances of these methods heavily depend on the effective solution of the auxiliary problems (P_{w^k,λ_k}). In general the functions $F(\cdot, w, \lambda)$ are nonsmooth and nonlinear.

To overcome the fact that the solution may not be unique, Gugat [8] and Roubi [17] proposed to add a prox-regularization term to the objective function of each subproblem (P_{w^k,λ_k}). So the problem (P_{w^k,λ_k}) is replaced by

$$(P_{w^k,\lambda_k,\alpha_k}) \qquad \inf_{x\in X} \left\{F(x, w^k, \lambda_k) + (1/2\alpha_k)\|x - x^k\|^2\right\},$$

where $\alpha_k > 0$. This way of doing is used in the well-known proximal point algorithm introduced by Martinet [13] and developed by Rockafellar [15] for solving nonsmooth convex minimization problems. This method is valid when X is a nonempty closed convex subset of $I\!R^n$ and when each function $f_i - \lambda g_i$ is convex for all $\lambda \geq \lambda^*$. In that case, problem $(P_{w^k,\lambda_k,\alpha_k})$ is a convex problem while problem (P) is only a pseudo-convex problem.

Concerning the nonsmoothness of problems (P_{w^k,λ_k}), it has been recently proposed in [1] to work with a smooth approximation of the max-function in (P_{w^k,λ_k}). This approximation is from above up to some $\epsilon > 0$.

In this paper, we use the prox-regularization principle to solve the problem (P_{w^k,λ_k}) and we approximate from below the convex nonsmooth objective function of this problem in order to make it easier to solve. Our lower approximation is a piecewise linear convex function built piece by piece until a criterion measuring the quality of the approximation is satisfied. This criterion is related to the serious steps used in the bundle methods. The approximation is suitable if it allows to obtain a sufficient decrease in the value of F. With this criterion, the method can be viewed as a classical bundle method where after each serious step the value of the parameter λ is updated. We refer the reader to [3] for more details on the bundle method in convex programming and to [10,14,18] for the bundle method in the framework of variational inequalities. However, to prove the superlinear convergence of the sequence $\{\lambda_k\}$, we have to consider a more restrictive criterion. The solution of every approximate problem must be an η_k-solution of problem (P_{w^k,λ_k}) with η_k related to $\|x^{k+1} - x^k\|^2$. Roubi also considers η_k-solutions but with unspecified $\eta_k > 0$. Unfortunately we cannot use Roubi's convergence theory [17] because his theory is based on the assumption that the series $\sum \sqrt{\eta_k \alpha_k}$ is convergent, what we were not able to prove in our situation. However thanks to the special form of η_k, we can modify some results of Roubi as well as some classical arguments in bundle methods to obtain the rate of convergence of the sequence $\{\lambda_k\}$ generated by our algorithm.

The paper is organized as follows: in Sect. 2 we introduce the inexact proximal point method tailored for solving generalized fractional programming problems and we study the convergence and the rate of convergence of the sequences $\{x^k\}$ and $\{\lambda_k\}$. Section 3 is devoted to the construction of the piecewise linear convex approximations of the function $F(\cdot, w^k, \lambda_k)$ and Sect. 4 to the report of some numerical results. Finally in a last section 'Conclusions and Perspectives' we give some arguments for dealing with the nonconvex case, i.e., the case where the functions $f_i - \lambda g_i$, $i = 1, \ldots, p$ are not necessarily convex.

2 An inexact proximal point method

From now on, we assume that X is a nonempty closed convex subset of $I\!R^n$ and that each function $f_i - \lambda g_i$ is convex for all $\lambda \in [\lambda^*, \lambda(x^0)]$ where $x^0 \in X$ is known (x^0 will be used as the starting point in our algorithms). This can be ensured in particular in each of the three following cases: (i) the functions f_i are convex and the functions g_i are affine, (ii) the functions f_i are convex, the functions g_i are concave and λ^* is known to be positive, (iii) the functions f_i are convex, the functions g_i are convex and $\lambda_0 = \lambda(x^0)$ is negative. In all these cases the function $F(\cdot, w^k, \lambda)$ is convex over X for all $\lambda \in [\lambda^*, \lambda_0]$. This assumption is justified because in our algorithms, the sequence $\{\lambda_k\}$ will be nonincreasing and bounded below by λ^*.

Given (x^k, w^k, λ_k), the prox-regularization method consists in replacing the problem $\min_{x \in X} F(x, w^k, \lambda_k)$ by the problem

$$(P_{w^k,\lambda_k,\alpha_k}) \qquad \min_{x \in X} \left\{ F(x, w^k, \lambda_k) + \frac{1}{2\alpha_k} \|x - x^k\|^2 \right\},$$

where $\alpha_k > 0$.

In order to obtain an implementable algorithm, we only compute an approximate solution of this problem. Practically this will be done by approximating in problem $(P_{w^k,\lambda_k,\alpha_k})$ the nonsmooth convex function $F(\cdot, w^k, \lambda_k)$ by a convex function $\varphi(\cdot, w^k, \lambda_k)$ in such a way that the problem

$$(AP_{w^k,\lambda_k,\alpha_k}) \qquad \min_{x \in X} \left\{ \varphi(x, w^k, \lambda_k) + \frac{1}{2\alpha_k} \|x - x^k\|^2 \right\}$$

is easier to solve exactly. The form of this function and how to construct it will be the subject of the next section. Here we only define the properties that the approximation $\varphi(\cdot, w^k, \lambda_k)$ of $F(\cdot, w^k, \lambda_k)$ must satisfy so that the sequence $\{\lambda_k\}$ converges to λ^*, the optimal value of (P), and the sequence $\{x^k\}$ converges to some solution of (P) if such a solution exists. The following approximation is classical in bundle methods.

Definition 2.1 Let $c \in (0, 1)$ and let $w^k > 0$, $\lambda_k \geq \lambda^*$ and $x^k \in X$. A convex function $\varphi(\cdot, w^k, \lambda_k)$ is a c-approximation of $F(\cdot, w^k, \lambda_k)$ at x^k if $\varphi(x, w^k, \lambda_k) \leq F(x, w^k, \lambda_k)$ for all $x \in X$, and if

$$\varphi(x^{k+1}, w^k, \lambda_k) \geq \frac{1}{c} F(x^{k+1}, w^k, \lambda_k), \tag{3}$$

where x^{k+1} is the solution of problem $(AP_{w^k,\lambda_k,\alpha_k})$.

Observe that if $\varphi(\cdot, w^k, \lambda_k)$ is a c-approximation of $F(\cdot, w^k, \lambda_k)$ at x^k, then at x^{k+1}, we can write

$$\frac{1}{c} F(x^{k+1}, w^k, \lambda_k) \leq \varphi(x^{k+1}, w^k, \lambda_k) \leq F(x^{k+1}, w^k, \lambda_k). \tag{4}$$

In particular, since $c \in (0, 1)$, we have

$$\varphi(x^{k+1}, w^k, \lambda_k) \leq F(x^{k+1}, w^k, \lambda_k) \leq 0. \tag{5}$$

We can now summarize our general algorithm as follows:

Algorithm 2.1

0. Choose $x^0 \in X$, $w^0 > 0$, $\alpha_0 > 0$, $c \in (0, 1)$ and set $\lambda_0 = \lambda(x^0)$.
1. At step k, we have x^k, w^k, α_k and λ_k. Then, construct a c-approximation of $F(\cdot, w^k, \lambda_k)$ and find $x^{k+1} \in X$ the unique solution of problem $(AP_{w^k,\lambda_k,\alpha_k})$. Set $\lambda_{k+1} = \lambda(x^{k+1})$, choose $w^{k+1} > 0$, $\alpha_{k+1} > 0$, set $k \leftarrow k + 1$, and go back to 1.

First observe that w^k does not intervene in the computation of λ_k and that for any $w > 0$,

$$\lambda_{k+1} = \lambda(x^{k+1}) \Leftrightarrow F(x^{k+1}, w, \lambda_{k+1}) = 0.$$

Consequently at each iteration, we have that $F(x^k, w^k, \lambda_k) = 0$, and that a c-approximation of $F(\cdot, w^k, \lambda_k)$ at x^k satisfies the property:

$$F(x^k, w^k, \lambda_k) - F(x^{k+1}, w^k, \lambda_k) \geq c\,[\,F(x^k, w^k, \lambda_k) - \varphi(x^{k+1}, w^k, \lambda_k)\,]. \tag{6}$$

Observe that condition (6) can be interpreted as *the real decrease on F when passing from x^k to x^{k+1} (the left–hand side) is greater than a fraction of the decrease predicted by the model φ (the right–hand side)*. This kind of step, called a serious step, is used in the bundle methods for minimizing a nonsmooth convex function as well as in the trust region methods in nonlinear programming.

In order to study the convergence of the sequence $\{\lambda_k\}$, we introduce the following notations. For $x \in X$, $w > 0$, and λ, we define the sets

$$I(x) = \left\{ i \mid \frac{f_i(x)}{g_i(x)} = \lambda(x) \right\}, \quad J(x, \lambda) = \{\, j \mid f_j(x) - \lambda g_j(x) = F(x, \lambda)\},$$

$$J(x, w, \lambda) = \left\{ j \mid \frac{f_j(x) - \lambda g_j(x)}{w_j} = F(x, w, \lambda) \right\}. \tag{7}$$

Proposition 2.1 *Assume $c \in (0, 1)$. Then the following results hold:*

1. *the sequence $\{\lambda_k\}$ is nonincreasing and converges to some $\hat{\lambda} \geq \lambda^*$;*
2. *if $\lambda^* > -\infty$ and if $g_i(x^k) \leq \gamma$ and $w_i^k \geq \underline{w} > 0$ for all k and $1 \leq i \leq p$, then $F(x^{k+1}, w^k, \lambda_k) \to 0$.*

Proof

1. By definition of F, we have for all $1 \leq i \leq p$ that

$$F(x^{k+1}, w^k, \lambda_k) \geq (1/w_i^k)\,[f_i(x^{k+1}) - \lambda_k g_i(x^{k+1})]. \tag{8}$$

Since $f_{i^*}(x^{k+1}) = \lambda_{k+1} g_{i^*}(x^{k+1})$ for $i^* \in I(x^{k+1})$, we obtain, using (8) and (5), that

$$\frac{g_{i^*}(x^{k+1})}{w_{i^*}^k}\,[\lambda_{k+1} - \lambda_k] \leq F(x^{k+1}, w^k, \lambda_k) \leq 0. \tag{9}$$

Since $g_{i^*}(x^{k+1}) > 0$ and $w_{i^*}^k > 0$, it follows that $\lambda_{k+1} \leq \lambda_k$. So $\lambda_k \to \hat{\lambda} \geq \lambda^*$ because $\lambda_k \geq \lambda^*$ for all k.
2. If $\lambda^* > -\infty$, then $\hat{\lambda} > -\infty$ and $\lambda_{k+1} - \lambda_k \to 0$. Since, by assumption, there exist $\gamma > 0$ and $\underline{w} > 0$ such that for all $1 \leq i \leq p$ and all k, $g_i(x^k) \leq \gamma$ and $w_i^k \geq \underline{w}$, and since $\lambda_{k+1} - \lambda_k \leq 0$, it follows from (9) that

$$\frac{\gamma}{\underline{w}}(\lambda_{k+1} - \lambda_k) \leq F(x^{k+1}, w^k, \lambda_k) \leq 0. \tag{10}$$

Consequently, $F(x^{k+1}, w^k, \lambda_k) \to 0$ when $k \to \infty$. □

In order to prove that $\hat{\lambda} = \lambda^*$, i.e., that $\lambda_k \to \lambda^*$, we need the following lemma.

Lemma 2.1 *Let $\{(x^k, w^k, \lambda_k)\}$ be the sequence generated by Algorithm 2.1. Then the following properties hold:*

(i) for all k, one has $\varepsilon_k \equiv -\varphi(x^{k+1}, w^k, \lambda_k) - \alpha_k^{-1}\|x^{k+1} - x^k\|^2 \geq 0$;

(ii) for all $x \in X$,

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \|x^{k+1} - x^k\|^2 + 2\alpha_k\,[F(x, w^k, \lambda_k) + \varepsilon_k];$$

(iii) if, in addition, $\lambda^* > -\infty$, $g_i(x^k) \leq \gamma$ and $w_i^k \geq \underline{w} > 0$ for all k and $1 \leq i \leq p$, then the series $\sum_{k\geq 1}[\varepsilon_k + (2\alpha_k)^{-1}\|x^{k+1} - x^k\|^2]$ is convergent.

Proof

(i) By definition of x^{k+1}, we have that $\alpha_k^{-1}\,(x^k - x^{k+1}) \in \partial[\varphi(\cdot, w^k, \lambda_k) + \psi_X]\,(x^{k+1})$ where ψ_X is the indicator function associated with X ($\psi_X = 0$ over X and $+\infty$ otherwise). Hence for all $x \in X$, we obtain

$$\varphi(x, w^k, \lambda_k) \geq \varphi(x^{k+1}, w^k, \lambda_k) + \alpha_k^{-1}\,\langle x^k - x^{k+1}, x - x^{k+1}\rangle. \tag{11}$$

Taking $x = x^k$ and using Definition 2.1, we deduce that

$$0 \geq \varphi(x^{k+1}, w^k, \lambda_k) + \alpha_k^{-1}\,\|x^{k+1} - x^k\|^2,$$

i.e., $\varepsilon_k \geq 0$.

(ii) Let $x \in X$. Since $\varphi(x, w^k, \lambda_k) \leq F(x, w^k, \lambda_k)$, it follows from (11) and from the definition of ε_k that

$$\begin{aligned} F(x, w^k, \lambda_k) &\geq \varphi(x^{k+1}, w^k, \lambda_k) + \alpha_k^{-1}\langle x^k - x^{k+1}, x - x^k + x^k - x^{k+1}\rangle \\ &= \varphi(x^{k+1}, w^k, \lambda_k) + \alpha_k^{-1}\|x^{k+1} - x^k\|^2 + \alpha_k^{-1}\langle x^k - x^{k+1}, x - x^k\rangle \\ &= -\varepsilon_k + \alpha_k^{-1}\langle x^k - x^{k+1}, x - x^k\rangle. \end{aligned}$$

Using this inequality in the last term of the following equality:

$$\|x^{k+1} - x\|^2 = \|x^{k+1} - x^k\|^2 + \|x^k - x\|^2 + 2\langle x^{k+1} - x^k, x^k - x\rangle,$$

we obtain that

$$\|x^{k+1} - x\|^2 \leq \|x^{k+1} - x^k\|^2 + \|x^k - x\|^2 + 2\alpha_k\,[F(x, w^k, \lambda_k) + \varepsilon_k].$$

(iii) By definition of ε_k and by (3), we have that

$$\varepsilon_k + \alpha_k^{-1}\|x^{k+1} - x^k\|^2 = -\varphi(x^{k+1}, w^k, \lambda_k) \leq -\frac{1}{c}\,F(x^{k+1}, w^k, \lambda_k).$$

From (10) with k replaced by $k-1$, we can write $0 \leq F(x^k, w^{k-1}, \lambda_{k-1}) + (\gamma/\underline{w})\,[\lambda_{k-1} - \lambda_k]$. Combining the last two inequalities, we obtain

$$\begin{aligned} \varepsilon_k + \alpha_k^{-1}\|x^{k+1} - x^k\|^2 &\leq \frac{1}{c}\,[F(x^k, w^{k-1}, \lambda_{k-1}) - F(x^{k+1}, w^k, \lambda_k)] \\ &+ \frac{\gamma}{\underline{w}c}\,[\lambda_{k-1} - \lambda_k]. \end{aligned}$$

Summing the previous inequalities from $k = 1$ to $k = q$ yields

$$\begin{aligned} \sum_{k=1}^{q}\left\{\varepsilon_k + \alpha_k^{-1}\|x^{k+1} - x^k\|^2\right\} &\leq \frac{1}{c}\,[F(x^1, w^0, \lambda_0) - F(x^{q+1}, w^q, \lambda_q)] \\ &\quad + \frac{\gamma}{\underline{w}c}\,[\lambda_0 - \lambda_q]. \end{aligned}$$

Since, by Proposition 2.1, $F(x^{q+1}, w^q, \lambda_q) \to 0$ and $\lambda_q \to \hat{\lambda}$, we have that the series $\sum_{k\geq 1}[\varepsilon_k + \alpha_k^{-1}\|x^{k+1} - x^k\|^2]$ is convergent. Hence the result. □

From this lemma, we can derive the convergence of the sequence $\{\lambda_k\}$ toward λ^*.

Theorem 2.1 *Let $c \in (0, 1)$. Assume $0 < \nu \le g_i(x^k) \le \gamma$ and $0 < \underline{w} \le w_i^k \le \overline{w}$ for all k and $1 \le i \le p$. Assume also that $\sum_{k\ge 0} \alpha_k = +\infty$ and that either $\alpha_k \le \overline{\alpha}$ for all k or $\alpha_k \le \alpha_{k+1}$ for all k. Then the sequence $\{\lambda_k\}$ generated by Algorithm 2.1 converges to λ^*, the optimal value of problem (P).*

Proof Since, by Proposition 2.1, the sequence $\{\lambda_k\}$ converges to $\hat{\lambda} \ge \lambda^*$, it remains to prove that $\hat{\lambda} = \lambda^*$. If $\hat{\lambda} = -\infty$, then $\hat{\lambda} = \lambda^*$. So we can suppose that $\hat{\lambda} > -\infty$. Now let $x \in X$. Then, for $j \in J(x, w^k, \lambda_k)$, we have $F(x, w^k, \lambda_k) = (f_j(x) - \lambda_k g_j(x))/w_j^k$ and since $\lambda(x) \ge f_j(x)/g_j(x)$, we obtain

$$F(x, w^k, \lambda_k) \le (\lambda(x) - \lambda_k) g_j(x)/w_j^k.$$

Then, by assumption,

$$F(x, w^k, \lambda_k) \le (\lambda(x) - \lambda_k)\nu/\overline{w} \quad \text{if } \lambda(x) - \lambda_k \le 0. \tag{12}$$

Now we prove that for all $x \in X$, we have

$$\limsup_{k\to\infty} \left(F(x, w^k, \lambda_k) \right) \ge 0. \tag{13}$$

Suppose, to get a contradiction, that (13) is not true. Then there exist $\varepsilon > 0$, $\tilde{x} \in X$ and k_ε such that

$$F(\tilde{x}, w^k, \lambda_k) < -\varepsilon \quad \text{for all } k \ge k_\varepsilon.$$

Then, for all $k \ge k_\varepsilon$, it follows from the second part of Lemma 2.1 with $x = \tilde{x}$ that

$$\|x^{k+1} - \tilde{x}\|^2 \le \|x^k - \tilde{x}\|^2 + 2\alpha_k \left[\varepsilon_k + (2\alpha_k)^{-1} \|x^{k+1} - x^k\|^2\right] - 2\alpha_k \varepsilon. \tag{14}$$

First assume that $\alpha_k \le \overline{\alpha}$. Since $2\alpha_k \varepsilon > 0$, we deduce from the previous inequality that

$$\|x^{k+1} - \tilde{x}\|^2 \le \|x^k - \tilde{x}\|^2 + 2\overline{\alpha} \left[\varepsilon_k + (2\alpha_k)^{-1} \|x^{k+1} - x^k\|^2\right].$$

But, by Lemma 2.1, the series $\sum_{k\ge 1} [\varepsilon_k + (2\alpha_k)^{-1} \|x^{k+1} - x^k\|^2]$ is convergent and thus the sequence $\{\|x^k - \tilde{x}\|^2\}$ converges to some $u \ge 0$. Summing the inequality (14) from $k = k_\varepsilon$ to $k = q$, and using $\alpha_k \le \overline{\alpha}$, we have

$$\|x^{q+1} - \tilde{x}\|^2 - \|x^{k_\varepsilon} - \tilde{x}\|^2 \le 2\overline{\alpha} \sum_{k=k_\varepsilon}^{q} \left[\varepsilon_k + (2\alpha_k)^{-1} \|x^{k+1} - x^k\|^2\right] - 2\varepsilon \sum_{k=k_\varepsilon}^{q} \alpha_k.$$

Taking the limit as $q \to \infty$ and using the assumption $\sum_{k\ge 0} \alpha_k = +\infty$, we obtain that $u - \|x^{k_\varepsilon} - \tilde{x}\|^2$ is less than $-\infty$, which is impossible. So (13) holds.
Assume now that $\alpha_k \le \alpha_{k+1}$ for all k. Then (14) implies that

$$(2\alpha_{k+1})^{-1} \|x^{k+1} - \tilde{x}\|^2 \le (2\alpha_k)^{-1} \|x^k - \tilde{x}\|^2 + \left[\varepsilon_k + (2\alpha_k)^{-1} \|x^{k+1} - x^k\|^2\right] - \varepsilon. \tag{15}$$

Since $\varepsilon > 0$ and the series $\sum_{k\ge 1} [\varepsilon_k + (2\alpha_k)^{-1} \|x^{k+1} - x^k\|^2]$ is convergent, it follows that the sequence $\{(2\alpha_k)^{-1} \|x^k - \tilde{x}\|^2\}$ converges to some $u \ge 0$. Summing the inequality (15) from $k = k_\varepsilon$ to $k = q$, we have

$$\begin{aligned} &(2\alpha_{q+1})^{-1} \|x^{q+1} - \tilde{x}\|^2 - (2\alpha_{k_\varepsilon})^{-1} \|x^{k_\varepsilon} - \tilde{x}\|^2 \\ &\quad \le \sum_{k=k_\varepsilon}^{q} \left[\varepsilon_k + (2\alpha_k)^{-1} \|x^{k+1} - x^k\|^2\right] - \varepsilon(q - k_\varepsilon). \end{aligned}$$

Taking the limit as $q \to \infty$, we obtain that $u - (2\alpha_{k_\varepsilon})^{-1}\|x^{k_\varepsilon} - \tilde{x}\|^2 \leq -\infty$, which is impossible. So (13) holds.

Let $x \in X$. First if $\lambda(x) \geq \lambda_k$ for an infinite set of indices k, then $\lambda(x) \geq \hat{\lambda}$. Otherwise, $\lambda(x) < \lambda_k$ is true for all k greater than some k_0. But then from (12), we have $F(x, w^k, \lambda_k) \leq (\lambda(x) - \lambda_k)\nu/\overline{w}$ for all $k \geq k_0$. Taking the superior limit of both members and using (13), we deduce that $\lambda(x) \geq \hat{\lambda}$. So in both cases, we obtain that $\lambda(x) \geq \hat{\lambda}$. Since x is arbitrary, we have that $\lambda^* \geq \hat{\lambda}$ and thus that $\lambda^* = \hat{\lambda}$. □

Observe that it is not supposed that problem (P) has a solution to get the convergence of the sequence $\{\lambda_k\}$. Moreover, the assumption $\sum_{k\geq 0} \alpha_k = +\infty$ is usual in the convergence theorems concerning the proximal point algorithms (see, for example, [15]). Here we impose, in addition, that either the sequence $\{\alpha_k\}$ is bounded above or nondecreasing. In particular, we can choose for $\{\alpha_k\}$ a constant sequence or a nondecreasing sequence converging to $+\infty$. In the next theorem, we prove the convergence of the sequence $\{x^k\}$, but this time under the assumption that (P) has a solution. However to prove this result, we need the following lemma (see e.g., [3]).

Lemma 2.2 *Let $\bar{z}$ be a limit point of a sequence $\{z^k\}$ satisfying*

$$\|z^{k+1} - \bar{z}\|^2 \leq \|z^k - \bar{z}\|^2 + \delta_k$$

where $\{\delta_k\}$ is a sequence of nonnegative numbers such that $\sum_{k\geq 0} \delta_k < +\infty$. Then the whole sequence $\{z^k\}$ converges to $\bar{z}$.

Theorem 2.2 *Assume that the assumptions of Theorem 2.1 are satisfied. Then*

(i) any limit point of the sequence $\{x^k\}$ is a solution of (P);
(ii) if $\alpha_k \leq \overline{\alpha}$ for all k and the solution set of problem (P) is nonempty, then the sequence $\{x^k\}$ converges to some solution of (P).

Proof

(i) Let x^* be a limit point of the sequence $\{x^k\}$. Then $x^{n_k} \to x^*$ and since $\lambda(x)$ is a continuous function, $\lambda(x^{n_k}) \to \lambda(x^*)$. But $\lambda(x^{n_k}) = \lambda_{n_k} \to \lambda^*$ (by Theorem 2.1). So $\lambda(x^*) = \lambda^*$ and x^* is a solution of problem (P).

(ii) First we prove that the sequence $\{x^k\}$ is bounded. In that purpose, let $\bar{x}$ be a solution of problem (P). Then $F(\bar{x}, w^k, \lambda_k) \leq 0$. Indeed, since $\lambda_k \geq \lambda^* = \lambda(\bar{x}) = \max_i f_i(\bar{x})/g_i(\bar{x})$, we have that $\max_i\{f_i(\bar{x}) - \lambda_k g_i(\bar{x})\} \leq 0$ and thus that $F(\bar{x}, w^k, \lambda_k) \leq 0$. Now using the second part of Lemma 2.1 with $x = \bar{x}$, we obtain

$$\begin{aligned} \|x^{k+1} - \bar{x}\|^2 &\leq \|x^k - \bar{x}\|^2 + \|x^{k+1} - x^k\|^2 + 2\alpha_k\,[F(\bar{x}, w^k, \lambda_k) + \varepsilon_k] \\ &\leq \|x^k - \bar{x}\|^2 + 2\overline{\alpha}\,[\,\varepsilon_k + (2\alpha_k)^{-1}\,\|x^{k+1} - x^k\|^2\,]. \end{aligned}$$

Since the series $\sum_{k\geq 1}[\varepsilon_k + (2\alpha_k)^{-1}\|x^{k+1} - x^k\|^2]$ is convergent, it follows that the sequence $\{\|x^k - \bar{x}\|\}$ is convergent and thus that the sequence $\{x^k\}$ is bounded. Let x^* be a limit point of the sequence $\{x^k\}$. By (i), x^* is a solution of (P). Using again the second part of Lemma 2.1, but this time with $x = x^*$, we obtain

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + 2\overline{\alpha}\,[\,\varepsilon_k + (2\alpha_k)^{-1}\|x^{k+1} - x^k\|^2\,].$$

Since x^* is a limit point of $\{x^k\}$ and since $\sum_{k\geq 1}[\varepsilon_k + (2\alpha_k)^{-1}\|x^{k+1} - x^k\|^2]$ is convergent, it follows from Lemma 2.2, that the whole sequence $\{x^k\}$ converges to x^*. □

Observe that the assumptions on the sequence $\{w^k\}$, i.e., $\overline{w} \geq w_i^k \geq \underline{w} > 0$ for all k and $1 \leq i \leq p$, are satisfied when $w_i^k = 1$ for all k and $1 \leq i \leq p$ as well as when $w_i^k = g_i(x^k)$ for all k and $1 \leq i \leq p$ because in the second case, it is a consequence of the condition $\gamma \geq g_i(x) \geq \nu$ for all $x \in X$ and $1 \leq i \leq p$ imposed on the functions g_i.

In order to analyze the rate of convergence of Algorithm 2.1, we consider the following assumption denoted by (H):

(H) there exist $\delta > 0$, $\kappa > 0$ such that
$F(x, \lambda^*) \geq \kappa \mathrm{dist}(x, X^*)^2$ for all $x \in B(X^*, \delta) \cap X$,

where X^* is the solution set of (P), $B(X^*, \delta) = \cup_{\bar{x} \in X^*} B(x, \delta)$, $B(x, \delta) = \{y \in I\!R^n \mid \|y - x\| < \delta\}$ and $\mathrm{dist}(x, X^*) = \inf_{\bar{x} \in X^*} \|x - \bar{x}\|$. This assumption has been introduced by Roubi [17] to prove the linear rate of convergence of his prox-regularization method. In particular, it is satisfied when the function $F(\cdot, \lambda^*)$ is polyhedral (for example when the functions f_i are polyhedral and the functions g_i are affine) and X is polyhedral. This assumption is also satisfied when $F(\cdot, \lambda^*)$ is strongly convex.

Theorem 2.3 *Assume that the solution set X^* of problem (P) is nonempty and that the function $F(\cdot, \lambda^*)$ satisfies assumption (H). Assume also that $\tau = \inf_{x \in X^*} \min_i g_i(x) > 0$ and that the sequence $\{x^k\}$ converges to some solution of (P). Then for α_k sufficiently large, the sequence $\{\lambda_k\}$ converges linearly to λ^*.*

Proof Let $\tilde{x}^k \in X^*$ such that $\|x^k - \tilde{x}^k\| = \mathrm{dist}(x^k, X^*)$. From the definition of x^{k+1} and the inequality $\|x^{k+1} - x^k\|^2 \geq 0$, we obtain that

$$\varphi(\tilde{x}^k, w^k, \lambda_k) + (2\alpha_k)^{-1} \|\tilde{x}^k - x^k\|^2 \geq \varphi(x^{k+1}, w^k, \lambda_k),$$

and thus, by definition of a c-approximation and from (10), we can deduce that

$$F(\tilde{x}^k, w^k, \lambda_k) + (2\alpha_k)^{-1} \|\tilde{x}^k - x^k\|^2 \geq \frac{1}{c} F(x^{k+1}, w^k, \lambda_k) \geq \frac{\gamma}{\underline{w} c}(\lambda_{k+1} - \lambda_k). \tag{16}$$

Since the sequence $\{x^k\}$ converges to some $x^* \in X^*$, we have that $x^k \in B(X^*, \delta) \cap X$ for k large enough and thus, by assumption (H)

$$F(x^k, \lambda^*) \geq \kappa \|x^k - \tilde{x}^k\|^2 \quad \text{for } k \text{ large enough.} \tag{17}$$

Next $F(x^k, \lambda^*) = f_j(x^k) - \lambda^* g_j(x^k)$ for some $j \in J(x^k, \lambda^*)$. Hence, since $f_j(x^k) \leq \lambda_k g_j(x^k)$, we have

$$F(x^k, \lambda^*) \leq (\lambda_k - \lambda^*) g_j(x^k) \leq \gamma(\lambda_k - \lambda^*) \tag{18}$$

because $\lambda_k \geq \lambda^*$. Combining (17) and (18), we obtain that for k large enough

$$\kappa \|x^k - \tilde{x}^k\|^2 \leq \gamma(\lambda_k - \lambda^*). \tag{19}$$

Similarly for $j \in J(\tilde{x}^k, w^k, \lambda_k)$, we have $F(\tilde{x}^k, w^k, \lambda_k) = (1/w_j^k)[f_j(\tilde{x}^k) - \lambda_k g_j(\tilde{x}^k)]$. Since $f_j(\tilde{x}^k) \leq \lambda(\tilde{x}^k) g_j(\tilde{x}^k) = \lambda^* g_j(\tilde{x}^k)$, we obtain

$$F(\tilde{x}^k, w^k, \lambda_k) \leq (\lambda^* - \lambda_k) \frac{g_j(\tilde{x}^k)}{w_j^k} \leq \frac{\tau}{\overline{w}}(\lambda^* - \lambda_k) \tag{20}$$

because $\lambda_k \geq \lambda^*$. So, using successively (20), (19), and (16) yields for k large enough

$$\frac{\tau}{\overline{w}}(\lambda^* - \lambda_k) + (2\alpha_k)^{-1} \|\tilde{x}^k - x^k\|^2 \geq F(\tilde{x}^k, w^k, \lambda_k) + (2\alpha_k)^{-1} \|\tilde{x}^k - x^k\|^2$$
$$\frac{\tau}{\overline{w}}(\lambda^* - \lambda_k) + (2\alpha_k)^{-1} \frac{\gamma}{\kappa}(\lambda_k - \lambda^*) \geq \frac{\gamma}{\underline{w} c}(\lambda_{k+1} - \lambda_k).$$

Writing $\lambda_{k+1} - \lambda_k = \lambda_{k+1} - \lambda^* + \lambda^* - \lambda_k$, we deduce from the previous inequality, after division by $\gamma/\underline{w}c$, that

$$\left[1 + c\left(\frac{\underline{w}}{2\alpha_k \kappa} - \frac{\tau \underline{w}}{\gamma \overline{w}}\right)\right](\lambda_k - \lambda^*) \geq (\lambda_{k+1} - \lambda^*).$$

Since $\tau \underline{w} \leq \gamma \overline{w}$, the coefficient of $(\lambda_k - \lambda^*)$ in the left-hand side is positive. It is strictly less than 1 when $\liminf_{k\to\infty} \alpha_k > (\gamma \overline{w})/(2\kappa\tau)$. Hence the linear convergence of the sequence $\{\lambda_k\}$ when α_k is sufficiently large. □

Theoretically, since $c \in (0, 1)$ and since $\frac{\underline{w}}{2\alpha_k \kappa} - \frac{\tau \underline{w}}{\gamma \overline{w}} < 0$, the best linear rate of convergence is attained when c is near 1. But the more c is close to 1, the more accurate is the c-approximation (it is exact when $c = 1$) and the more difficult it is to compute it.

Now to obtain the superlinear convergence of the sequence $\{\lambda_k\}$, we have to impose that the regularization parameter α_k tends to $+\infty$. We also have to assume a stronger condition on the c-approximating function φ. This is the subject of the next definition.

Definition 2.2 Let $c \in (0, 1)$ and let $w^k > 0$, $\lambda_k \geq \lambda^*$ and $x^k \in X$. A convex function $\varphi(\cdot, w^k, \lambda_k)$ is a strong c-approximation of $F(\cdot, w^k, \lambda_k)$ at x^k if $\varphi(x, w^k, \lambda_k) \leq F(x, w^k, \lambda_k)$ for all $x \in X$ and if

$$F(x^{k+1}, w^k, \lambda_k) - \varphi(x^{k+1}, w^k, \lambda_k) \leq \frac{1-c}{\alpha_k} \|x^{k+1} - x^k\|^2, \tag{21}$$

where x^{k+1} is the solution of problem $(AP_{w^k,\lambda_k,\alpha_k})$.

This definition is justified by the next proposition.

Proposition 2.2 *Let $c \in (0, 1)$. A strong c-approximation of $F(\cdot, w^k, \lambda_k)$ at x^k is also a c-approximation of $F(\cdot, w^k, \lambda_k)$ at x^k.*

Proof By definition of x^{k+1}, we have that $\alpha_k^{-1}(x^k - x^{k+1}) \in \partial[\varphi(\cdot, w^k, \lambda_k) + \psi_X](x^{k+1})$. So

$$\varphi(x^k, w^k, \lambda_k) - \varphi(x^{k+1}, w^k, \lambda_k) \geq \alpha_k^{-1} \|x^{k+1} - x^k\|^2. \tag{22}$$

Since, by assumption, $\varphi(x^k, w^k, \lambda_k) \leq F(x^k, w^k, \lambda_k) = 0$, we obtain from the previous inequality and from (21) that

$$-\varphi(x^{k+1}, w^k, \lambda_k) \geq \frac{1}{1-c}[F(x^{k+1}, w^k, \lambda_k) - \varphi(x^{k+1}, w^k, \lambda_k)],$$

i.e.,

$$\frac{c}{1-c}\varphi(x^{k+1}, w^k, \lambda_k) \geq \frac{1}{1-c} F(x^{k+1}, w^k, \lambda_k).$$

But this inequality is equivalent to (3). □

Theorem 2.4 *Assume that the solution set X^* of problem (P) is nonempty, that the function $F(\cdot, \lambda^*)$ satisfies assumption (H), and that $\tau = \inf_{x \in X^*} \min_i \ g_i(x) > 0$. Assume also that at each iteration, $\varphi(\cdot, w^k, \lambda_k)$ is a strong c-approximation of $F(\cdot, w^k, \lambda_k)$ with $c > 1/2$ and that the sequence $\{x^k\}$ converges to some solution of (P). Then the sequence $\{\lambda_k\}$ converges superlinearly to λ^* if α_k tends to $+\infty$ when $k \to \infty$ provided that at each iteration, w_i^k is chosen equal to $\beta g_i(x^k)$ with $\beta > 0$.*

Proof By definition of F and λ_{k+1}, we have for $i \in I(x^{k+1}, \lambda_{k+1})$ that

$$F(x^{k+1}, w^k, \lambda_k) \geq \frac{f_i(x^{k+1}) - \lambda_k g_i(x^{k+1})}{w_i^k} = (\lambda_{k+1} - \lambda_k)\, \frac{g_i(x^{k+1})}{w_i^k}.$$

Since $\lambda_k \geq \lambda^*$ for all k, we can deduce from the previous inequality that

$$F(x^{k+1}, w^k, \lambda_k) \geq (\lambda_{k+1} - \lambda^*) \min_i \frac{g_i(x^{k+1})}{w_i^k} - (\lambda_k - \lambda^*) \max_i \frac{g_i(x^{k+1})}{w_i^k} \tag{23}$$

Let $\tilde{x}^k \in X^*$ such that $\|x^k - \tilde{x}^k\|^2 = d(x^k, X^*)$. By definition of x^{k+1}, we have

$$\varphi(\tilde{x}^k, w^k, \lambda_k) + (2\alpha_k)^{-1}\, \|\tilde{x}^k - x^k\|^2 \geq \varphi(x^{k+1}, w^k, \lambda_k) + (2\alpha_k)^{-1}\, \|x^{k+1} - x^k\|^2.$$

Since, by assumption, $\varphi(\tilde{x}^k, w^k, \lambda_k) \leq F(\tilde{x}^k, w^k, \lambda_k)$, we obtain from the previous inequality and from (21) that

$$F(\tilde{x}^k, w^k, \lambda_k) + (2\alpha_k)^{-1}\, \|\tilde{x}^k - x^k\|^2 \geq F(x^{k+1}, w^k, \lambda_k) + \frac{2c-1}{2\alpha_k}\, \|x^{k+1} - x^k\|^2,$$

and thus, since $c > 1/2$, that

$$F(\tilde{x}^k, w^k, \lambda_k) + (2\alpha_k)^{-1}\, \|\tilde{x}^k - x^k\|^2 \geq F(x^{k+1}, w^k, \lambda_k). \tag{24}$$

Combining (24) and (23) yields

$$\begin{aligned} &F(\tilde{x}^k, w^k, \lambda_k) + (2\alpha_k)^{-1}\, \|\tilde{x}^k - x^k\|^2 \\ &\geq (\lambda_{k+1} - \lambda^*) \min_i \frac{g_i(x^{k+1})}{w_i^k} - (\lambda_k - \lambda^*) \max_i \frac{g_i(x^{k+1})}{w_i^k}. \end{aligned}$$

Then, using this inequality and the inequalities (19) and (20), we obtain

$$\left[\max_i \frac{g_i(x^{k+1})}{w_i^k} - \min_i \frac{g_i(\tilde{x}^k)}{w_i^k} + \frac{\gamma}{2\alpha_k \kappa} \right] (\lambda_k - \lambda^*) \geq (\lambda_{k+1} - \lambda^*) \min_i \frac{g_i(x^{k+1})}{w_i^k}.$$

Thanks to (19), the sequences $\{x^k\}$ and $\{\tilde{x}^k\}$ converge to the same limit. Combining this with the choice of w: $w_i^k = \beta g_i(x^k)$ for all k and $1 \leq i \leq p$, we have

$$\max_i \frac{g_i(x^{k+1})}{w_i^k} \to 1/\beta, \quad \min_i \frac{g_i(\tilde{x}^k)}{w_i^k} \to 1/\beta \quad \text{and} \quad \min_i \frac{g_i(x^{k+1})}{w_i^k} \to 1/\beta$$

as $k \to \infty$. Hence $(\lambda_{k+1} - \lambda^*)/(\lambda_k - \lambda^*) \to 0$ as $k \to \infty$ because $\alpha_k \to +\infty$ as $k \to \infty$. □

This theorem must be compared with Theorem 2.2 of [5] where the superlinear convergence of the sequence $\{\lambda_k\}$ is also obtained. In this theorem, there are no regularization terms, i.e., $\alpha_k = +\infty$, and all the components w_i^k of w are equal to $g_i(x^k)$.

3 Building *c*-approximations

In order to obtain an implementable algorithm, we have now to indicate how to construct a c-approximation of $F(\cdot, w^k, \lambda_k)$ at x^k such that the subproblem $(AP_{w^k,\lambda_k,\alpha_k})$ is easier to solve than problem $(P_{w^k,\lambda_k,\alpha_k})$. For the sake of simplicity, we denote the function $F(\cdot, w^k, \lambda_k)$ by F^k and a (strong) c-approximation of F^k at x^k by φ^k. When X is described by linear equalities and/or inequalities and when φ^k is a piecewise linear convex function, it is very easy to see that problem $(AP_{w^k,\lambda_k,\alpha_k})$ is equivalent to a convex quadratic programming problem. Indeed, let $\varphi^k(x) = \max_{1\le q\le m}\{\langle a_q, x\rangle + b_q\}$ where $a_q \in I\!R^n$ and $b_q \in I\!R$ for all $1 \le q \le m$. Then problem $(AP_{w^k,\lambda_k,\alpha_k})$ is equivalent to the problem

$$\begin{cases} \min \ v + (1/2\alpha_k)\|x - x^k\|^2 \\ \text{s.t.} \quad v \ge \langle a_q, x\rangle + b_q, \ q = 1, \ldots, m \\ \qquad\ \ x \in X \end{cases}$$

This problem is a convex quadratic problem and efficient numerical methods exist for solving it. The construction of φ^k is based on the following classical assumption in nonsmooth convex programming [3]:

Assumption: At each point y of X, one subgradient of F^k at y is available. This subgradient is denoted by $s(y)$.

A natural strategy to obtain a piecewise linear convex function for the function φ^k is to construct it piece by piece by generating successive models

$$\varphi_j^k, \ j = 1, 2, \ldots$$

until (if possible) $\varphi_{j_k}^k$ is a (strong) c-approximation of F^k at x^k for some $j_k \ge 1$. For $j = 1, 2, \ldots,$ we denote by y_j^k the unique solution of the problem

$$(P_j^k) \quad \min_{y\in X}\{\varphi_j^k(y) + (1/2\alpha_k)\|y - x^k\|^2\},$$

and we set $\varphi^k = \varphi_{j_k}^k$ and $x^{k+1} = y_{j_k}^k$.

In order to obtain a (strong) c-approximation $\varphi_{j_k}^k$ of F^k at x^k, we have to impose some conditions on the successive models $\varphi_j^k, \ j = 1, 2, \ldots$. However, before presenting them, we need to define the affine functions $l_j^k, \ j = 1, 2, \ldots$ by

$$l_j^k(y) = \varphi_j^k(y_j^k) + \langle \gamma_j^k, y - y_j^k\rangle \quad \forall y \in I\!R^n,$$

where $\gamma_j^k = \alpha_k^{-1}(x^k - y_j^k)$. By optimality of y_j^k, we have

$$\gamma_j^k \in \partial[\varphi_j^k + \psi_X](y_j^k) \tag{25}$$

where ψ_X is the indicator function associated with X. It is then easy to observe that

$$l_j^k(y_j^k) = \varphi_j^k(y_j^k) \quad \text{and} \quad l_j^k(y) \le \varphi_j^k(y) \quad \text{for all} \quad y \in X. \tag{26}$$

Now, we assume that the following conditions introduced in ([3], p. 269), are satisfied by the convex approximating functions φ_j^k,

(C1) $\varphi_j^k \le F^k$ on X for $j = 1, 2, \ldots$
(C2) $\varphi_{j+1}^k \ge F^k(y_j^k) + \langle s(y_j^k), \cdot - y_j^k\rangle$ on X for $j = 1, 2, \ldots$
(C3) $\varphi_{j+1}^k \ge l_j^k$ on X for $j = 1, 2, \ldots,$

where $s(y_j^k)$ denotes the subgradient of F^k available at y_j^k.

Several models fulfill these conditions. For example, for the first model φ_1^k, we can take the linear function

$$\varphi_1^k(y) = F^k(x^k) + \langle s(x^k), y - x^k\rangle \quad \forall y \in I\!R^n.$$

Since $s(x^k) \in \partial F^k(x^k)$, condition $(C1)$ is satisfied for $j = 1$. For the next models φ_j^k, $j = 2, \ldots$, there exist several possibilities. A first example is to take for $j = 1, 2, \ldots$

$$\varphi_{j+1}^k(y) = \max \{l_j^k(y), F^k(y_j^k) + \langle s(y_j^k), y - y_j^k \rangle\} \quad \forall y \in I\!R^n. \tag{27}$$

Conditions $(C2)$, $(C3)$ are obviously satisfied and condition $(C1)$ is also satisfied because each linear piece of these functions are below F^k.

Another example is to take for $j = 1, 2, \ldots$

$$\varphi_{j+1}^k(y) = \max_{0 \le q \le j} \{F^k(y_q^k) + \langle s(y_q^k), y - y_q^k \rangle\} \quad \forall y \in I\!R^n, \tag{28}$$

where $y_0^k = x^k$. Since $s(y_q^k) \in \partial F^k(y_q^k)$ for $q = 0, \ldots, j$ and since $\varphi_{j+1}^k \ge \varphi_j^k \ge l_j^k$, it is easy to see that conditions $(C1)$–$(C3)$ are satisfied.

Comparing (27) and (28), we can say that l_j^k plays the same role as the j linear functions $F^k(y_q^k) + \langle s(y_q^k), y - y_q^k \rangle$, $q = 0, \ldots, j-1$. It is the reason why this function l_j^k is called the aggregate affine function (see, e.g., [3]).

Now the algorithm to construct a strong c-approximation of F^k at x^k as well as the next iterate x^{k+1} can be expressed as follows:

Algorithm 3.1 Let $x^k \in X$ and $c \in (1/2, 1)$. Set $j = 1$.

Step 1. *Step 1.* Choose φ_j^k a convex piecewise linear function that satisfies $(C1)$–$(C3)$ and solve problem (P_j^k) to get y_j^k.

Step 2. *Step 2.* If $F^k(y_j^k) - \varphi_j^k(y_j^k) \le (1-c)\alpha_k^{-1}\|y_j^k - x^k\|^2$, then set $x^{k+1} = y_j^k$, $j_k = j$ and STOP; the function $\varphi_{j_k}^k$ is a strong c-approximation of F^k at x^k and x^{k+1} is the next iterate.

Step 3. *Step 3.* Increase j by 1 and go to Step 1.

A c-approximation can also be obtained by replacing in Step 2 the inequality by $\varphi_j^k(y_j^k) \ge c^{-1}F^k(y_j^k)$. As a strong c-approximation is also a c-approximation, it is immediate that if a strong c-approximation is obtained after finitely many iterations, then the same holds for the c-approximation. So we only consider strong c-approximation in the next theorem. Furthermore, the fact that φ_j^k satisfies $(C1)$–$(C3)$ means that φ_j^k satisfies $(C1)$ and, if $j \ge 2$, φ_j^k satisfies $(C2)$ and $(C3)$ with $j+1$ replaced by j.

Our aim is now to prove that if x^k is not a minimum of F^k and if the models φ_j^k, $j = 1, \ldots$ satisfy $(C1)$–$(C3)$, then there exists $j_k \in I\!N_0$ such that $\varphi_{j_k}^k$ is a strong c-approximation of F^k at x^k, i.e., that the procedure stops in Step 2 after finitely many iterations.

Theorem 3.1 *Suppose that the models φ_j^k, $j = 1, 2, \ldots$ satisfy conditions $(C1)$–$(C3)$, and let, for each j, y_j^k be the unique solution of problem (P_j^k). Let also $\bar{x}^k$ be the unique solution of problem $(P_{w^k,\lambda_k,\alpha_k})$. Then*

(1) $F^k(y_j^k) - \varphi_j^k(y_j^k) \to 0$ and $y_j^k \to \bar{x}^k$ when $j \to +\infty$.

(2) If $x^k \ne \bar{x}^k$, then the Algorithm 3.1 stops after finitely many iterations j_k with $\varphi_{j_k}^k$ a strong c-approximation of F^k at x^k and with $x^{k+1} = y_{j_k}^k$.

(3) If $x^k = \bar{x}^k$, then $\lambda_k = \lambda^*$ and x^k is a solution to problem (P).

Proof The proof of the first part is classical and can be found in ([3], Proposition 4.3). The second part is straightforward because the left-hand side of the inequality in Step 2. tends to zero while the right-hand side converges to the positive number $(1-c)\alpha_k^{-1}\|\bar{x}^k - x^k\|^2$.

Finally if $x^k = \bar{x}^k$, then $F(x^k, \lambda_k) = 0$ and the conclusion follows from Theorem 2.1 in [5]. □

Inserting Algorithm 3.1 in Step 1 of Algorithm 2.1, we obtain the following algorithm.

Bundle Algorithm
Choose $x^0 \in X$, $w^0 > 0$, $\alpha_0 > 0$, $c \in (1/2, 1)$ and set $\lambda_0 = \lambda(x^0)$, $y_0^0 = x^0$ and $k = 0$, $j = 1$.
Step 1. Choose a piecewise linear convex function φ_j^k satisfying $(C1)$–$(C3)$ and solve

$$(P_j^k) \quad \min_{y \in X} \left\{ \varphi_j^k(y) + \frac{1}{2\alpha_k} \|y - x^k\|^2 \right\},$$

to obtain the unique optimal solution y_j^k.
Step 2. If $F^k(y_j^k) - \varphi_j^k(y_j^k) \leq (1 - c)\alpha_k^{-1} \|y_j^k - x^k\|^2$, then set $x^{k+1} = y_j^k$, $y_0^{k+1} = x^{k+1}$, $\lambda_{k+1} = \lambda(x^{k+1})$, choose $w^{k+1} > 0$, $\alpha_{k+1} > 0$, increase k by 1 and set $j = 0$.
Step 3. Increase j by 1 and go to Step 1.

Another bundle algorithm is obtained by replacing in Step 2 the first inequality corresponding to a strong c-approximation by the inequality $\varphi_j^k(y_j^k) \geq c^{-1} F^k(y_j^k)$ corresponding to a c-approximation. To distinguish the two algorithms in the next section, we denote by B1 and B2 the bundle algorithms using the c-approximations and the strong c-approximations, respectively.

4 Numerical results

The computational experience reported here is performed with the software MATLAB. The purpose is to compare the numerical behavior of the two new bundle methods B1 and B2 introduced in Sect. 3 with the prox-regularization method (denoted M) introduced in Sect. 1 where each parametric subproblem $(P_{w^k, \lambda_k, \alpha_k})$ is solved using a nonsmooth exact minimization procedure before updating the value of λ_k. Numerical results for method M are reported in [8] by Gugat. For this comparison, we consider a first set of test problems proposed in [7] (see also [1], p. 21).

Problem 4.1

$$\min_{x \in X} \max \left\{ \frac{4x_1^3 + 11x_2}{16x_1 + 4x_2}, \frac{4x_1^2 - x_1}{3x_1 + x_2} \right\}$$

where

$$X = \{x \in I\!R^2 \mid x_1 + x_2 \geq 1,\ 2x_1 + x_2 \leq 4,\ x_1, x_2 \geq 0\}$$

and the initial point is $x_0 = (1, 1)^T$.

Problem 4.2

$$\min_{x \in X} \max \left\{ \left| \frac{3x_1 - 2x_2}{4x_1 + x_2} \right|, \left| \frac{x_1}{3x_1 + x_2} \right| \right\}$$

where

$$X = \{x \in I\!R^2 \mid x_1 + x_2 \geq 1,\ 2x_1 + x_2 \leq 4,\ x_1, x_2 \geq 0\}$$

and the initial point is $x_0 = (1, 1)^T$.

Table 1 Comparison of methods B1, B2, B3 and M on three test problems

Problem	4.1 ($n=2, p=2$)				4.2 ($n=2, p=4$)				4.3 ($n=4, p=18$)		
Method	B1	B2	B3	M	B1	B2	B3	M	B1	B2	B3
iter	4	3	7	3	9	9	9	9	41	37	56
QP	19	21	26	31	18	18	18	34	184	143	223
cpu	.22	.26	.32	.33	.27	.26	.33	.45	2.44	2.42	2.93
Sol	$\lambda^* = 0.432494$				$\lambda^* = 0.196152$				$\lambda^* = 0.074179$		

Problem 4.3

$$\min_{x \in X} \max_{i=0,\ldots,9} \left| \frac{8^4 x_1 + 8i^3 x_2 - i^4 x_3 - 8^3 i x_4}{8^4 x_4 + 8i^3 x_3} \right|$$

where

$$X = \left\{ x \in I\!R^4 \mid -1000 \le x_1, x_2 \le 1000,\ 1 \le \frac{i^3 x_3 + 8^3 x_4}{8^3} \le 1000,\ i = 0, \ldots, 9 \right\}$$

and the initial point is $x_0 = (0.5, 0, 0, 1)^T$. The results are summarized in Table 1 where iter, QP, cpu, n, p denote the number of iterations, the total number of quadratic problems solved, the cpu time in seconds, the number of variables, and the number of ratios, respectively. At each iteration the number of subgradient evaluations is equal to the number of quadratic problems solved. The numerical results indicate that the number of subgradient evaluations is quite constant at each iteration.

In all the tests, the parameters c and α_k are chosen as $c = 0.9$, and $\alpha_k = 50$ for all k, and the vector w^k as $w_i^k = g_i(x^k)$ for all k and $1 \le i \le p$. For all the methods the solution λ^* mentioned in [1] is achieved for the three problems and the two ratios in the objective function are active at the solution for the first two problems while four ratios out of 10 are active for the third problem. Although the dimension of these test problems is small, one can observe that methods B1 and B2 give better results than method M both in terms of the number of quadratic problems solved and of the cpu time. In general, less iterations are necessary in method M, but each iteration is more expensive in terms of the number of quadratic problems to solve. For problem 4.3, we do not report results for the method M because the cpu time exceeds the maximum time allowed in the tests. Note also that this problem has the particularity of having several optimal solutions.

To summarize, it seems more efficient to update the value of λ_k in problem $(P_{w^k,\lambda_k,\alpha_k})$ after completing a serious step as in methods B1 and B2 before the exact solution of this problem is obtained. Based on this observation, we consider another strategy in the numerical tests: the value of λ_k is updated as soon as, in the computation of a c-approximation, a negative value of F^k is reached. This method is denoted B3 in the following. Although no convergence proof has been established for this method, we observe that it converges for all problems solved. It follows that one iteration of method B3 should require solving a smaller number of quadratic problems at each iteration, but more iterations should be necessary to reach the solution. Note that we implement B3 using the c-approximation of F^k as in method B1.

To compare numerically the methods B1, B2 and B3, a second set of larger test problems is randomly generated. As suggested in [1], the ratios in these problems consist of quadratic functions $f_i(x) = (1/2)x^T G_i x + a_i^T x + b_i$ in the numerators, and linear functions

Table 2 Comparison of the three methods on randomly generated problems

Problem	$n = 15$ $p = 20$			$n = 20$ $p = 20$			$n = 50$ $p = 50$		
Method	B1	B2	B3	B1	B2	B3	B1	B2	B3
iter	6	6	8	7	7	9	7	7	13
QP	45	86	48	54	112	49	114	267	153
cpu	1.45	3.03	1.88	2.31	4.99	2.30	17.02	38.17	22.74
Sol	$\lambda^* = -0.534110$			$\lambda^* = -0.325440$			$\lambda^* = -0.092863$		
Problem	$n = 50$ $p = 100$			$n = 100$ $p = 100$			$n = 100$ $p = 150$		
Method	B1	B2	B3	B1	B2	B3	B1	B2	B3
iter	8	6	22	7	6	20	7	7	25
QP	124	229	223	124	234	195	146	301	301
cpu	16.7	30.2	32.7	69.6	133.6	115.2	86.1	179.8	190.4
Sol	$\lambda^* = -0.097989$			$\lambda^* = -0.083710$			$\lambda^* = -0.096492$		

$g_i(x) = c_i^T x + d_i$ in the denominators. The parameters of these functions are generated as follows:

1. The Hessian matrix G_i is given by $G_i = L_i D_i L_i^T$ where L_i is a unit lower triangular matrix with components randomly generated in $[-2.5, 2.5]$ and D_i is a positive diagonal matrix with components randomly generated in $[0.1, 1.6]$. In order to generate a positive semidefinite Hessian, the first element of D_i is set to zero.
2. The components of the vectors a_i and c_i are randomly generated in $[-15, 45]$ and $[0, 10]$, respectively.
3. The real numbers b_i and d_i are also randomly generated in $[-30, 0]$ and $[1, 5]$, respectively.

Moreover, the following feasible set is used for all the test problems:

$$X = \left\{ x \in I\!R^n \mid \sum_{j=1}^{n} x_j \leq 1,\ 0 \leq x_j \leq 1,\ j = 1, \ldots, n \right\},$$

and the initial feasible point is $x_0 = (1/n, \ldots, 1/n)$. Finally, the parameters $c = 0.9$, $\alpha_k = 50$, and $w_i^k = g_i(x^k)$ for all k and $1 \leq i \leq p$ are used. The results are summarized in Table 2. For each problem the three methods give the same optimal value for λ^*.

For these larger problems, we observe the same behavior as previously for B1 and B2. When the method B3 is compared with the method B1, we note that as expected, more iterations are required. Furthermore, even if the number of quadratic problems solved at each iteration of B3 is smaller than B1 in general, nevertheless the total number of quadratic problems solved in B3 (except for problem $n = 20$, $p = 20$) is larger inducing that the cpu time is also larger than B1. Now comparing methods B1 and B2, the results indicate that the first one is faster than the second one. Although the number of iterations for B2 is less than or equal to the number of iterations of B1, the cpu time used by B2 is larger since more quadratic problems have to be solved to reach an inexact solution for a strong c-approximation of F^k. In conclusion, the numerical results indicate that the method B1 seems to be the fastest amongst the methods studied in this paper for solving problem (P).

5 Conclusions and perspectives

We have shown that generalized fractional programs with convex data can be solved more efficiently by using an inexact proximal point method rather than an exact one. The strategy is to add a regularization term to the parametric function $F(\cdot, w^k, \lambda_k)$ and to introduce implementable criterions to decide when stopping the minimization of this regularized function in order to update the parameter λ_k. Since the subproblems are nonsmooth convex problems, we propose to use a classical bundle method where after each "serious step", the parameter is updated. Two sequences are obtained, $\{\lambda_k\}$ and $\{x^k\}$, converging to the optimal value and to some solution of (P), respectively. This procedure is particularly interesting when several solutions exist for problem (P). Finally, some numerical tests on randomly generated fractional problems indicate that the method using the c-approximations of $F(\cdot, w^k, \lambda_k)$ seems to be the most efficient.

We conjecture that the efficiency of the method could be improved by using the information at the end of an iteration to obtain a good starting c-approximation at the next iteration. This should be the subject of a future investigation.

In this paper we have assumed that the functions $f_i - \lambda g_i$ are convex for all $\lambda \in [\lambda^*, \lambda_0]$. But for several practical problems, this assumption is not true and the max-function $F(\cdot, w, \lambda)$ is no more convex. So there is a need to consider the nonconvex case. For taking this situation into account, several approaches have been proposed in the literature. One of them is to approximate the nonsmooth function by using an entropic regularization method (see, for example, [1,19]). Another way to deal with this difficulty is to adapt the proximal point method developed in this paper to the nonconvex case. In that direction, recent researches on proximal point methods (see, for example, [11]) have shown that for solving nonconvex optimization problems with this method, it is crucial, in order to get convergence of the iterates to a stationary point, that the proximal subproblems remain convex. In our situation, the subproblem $(P_{w,\lambda,\alpha})$ may remain convex even if the function $F(\cdot, w, \lambda)$ is nonconvex. For example, when $F(\cdot, w, \lambda)$ is a lower-C^2 function ([16], Def. 10.29, p. 447), it is possible to add to this function a quadratic term of the form $(1/2\alpha)\|\cdot\|^2$ such that the resulting function becomes convex ([16], Theorem 10.33, p. 450). In our setting, if all the functions f_i and g_i, $i = 1, \ldots, p$, are differentiable and if for each i the gradient $\nabla f_i - \lambda \nabla g_i$ is Lipschitz continuous with a constant L_i, then the function $F(\cdot, w, \lambda) + (1/2\alpha)\|\cdot\|^2$ is strongly convex for $\alpha < [\max_i \{L_i/w_i\}]^{-1}$ ([11], Proposition 1). In other words, if α is sufficiently small the problem $(P_{w,\lambda,\alpha})$ is strongly convex and consequently has a unique solution.

Another crucial issue is the design of an efficient method for computing the solution of problem $(P_{w,\lambda,\alpha})$ when $F(\cdot, w, \lambda)$ is nonconvex. Several bundle methods have been proposed in the literature for solving this problem when the function $F(\cdot, w, \lambda)$ is locally Lipschitz (see, for example, [6,12,20]). In these methods, the function $F(\cdot, w, \lambda)$ is approximated by a piecewise linear convex function (to obtain again a convex quadratic subproblem) built step by step by using the Clarke generalized gradient [2] instead of the usual subdifferential. However, due to the nonconvexity of $F(\cdot, w, \lambda)$, these approximations are only appropriate in a neighborhood of the current point x^k with the consequence that either a linesearch or a trust region strategy must be applied for finding the next point x^{k+1}. More recently, by means of variational analysis [16], Hare et al. [9] presented a new methodology for solving the subproblems based on the computation of proximal points of piecewise linear models of the nonconvex function. Convergence of the method is proven for the class of nonconvex functions that are prox-bounded and lower-C^2. From all these comments concerning nonconvex optimization problems, it follows that it is reasonable to think that the proximal point method studied in this paper could be adapted to the nonconvex case. It will be the subject of a future research.

Acknowledgements The authors are particularly grateful to an anonymous referee whose comments, questions, and suggestions greatly improved the paper. The work of the first author was supported in part by a grant from the Belgian Fonds National de la Recherche Scientifique (FNRS: B8/5-CB/SP-8.733).

References

1. Birbil, S.I., Frenk, J.B., Zhang, S.: Generalized fractional programming with user interaction. ERIM Report Series Research in Management ERS-2004-033-LIS (2004)
2. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
3. Correa, R., Lemaréchal, C.: Convergence of some algorithms for convex minimization. Math. Program. **62**, 261–275 (1993)
4. Crouzeix, J.P., Ferland, J.A., Schaible, S.: An algorithm for generalized fractional programs. J. Optim. Theory Appl. **47**, 35–49 (1985)
5. Crouzeix, J.P., Ferland, J.A., Schaible, S.: A note on an algorithm for generalized fractional programs. J. Optim. Theory Appl. **50**, 183–187 (1986)
6. Fuduli, A., Gaudioso, M., Giallombardo, G.: Minimizing nonconvex nonsmooth functions via cutting planes and proximity control. SIAM J. Optim. **14**, 743–756 (2004)
7. Gugat, M.: A fast algorithm for a class of generalized fractional programs. Manage. Sci. **42**, 1493–1499 (1996)
8. Gugat, M.: Prox-regularization methods for generalized fractional programming. J. Optim. Theory Appl. **99**, 691–722 (1998)
9. Hare, W., Sagastizábal, C.: Computing proximal points of nonconvex functions. Math. Program. Ser. B (2007). doi: 10.1007/s10107-007-0124-6
10. Hue, T.T., Strodiot, J.J., Nguyen, V.H.: Convergence of the approximate auxiliary problem method for solving generalized variational inequalities. J. Optim. Theory Appl. **121**, 119–145 (2004)
11. Kaplan, A., Tichatschke, R.: Proximal point methods and nonconvex optimization. J. Global Optim. **13**, 389–406 (1998)
12. Lemaréchal, C., Strodiot, J.J., Bihain, A.: On a bundle method for nonsmooth optimization. In: Mangasarian, O., Meyer, R., Robinson, S. (eds.) Nonlinear Programming, vol. 4, pp. 245–282. Academic Press, New York (1981)
13. Martinet, B.: Régularisation d'inéquations variationelles par approximations successives. RAIRO Rech. Oper. **4**, 154–159 (1970)
14. Nguyen, T.T.V., Strodiot, J.J., Nguyen, V.H.: A bundle method for solving equilibrium problems. Math. Program. Ser. B (2007). doi: 10.1007/s10107-007-0112-x
15. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
16. Rockafellar, R.T., Wets, R.B.: Variational Analysis. No. 317 in Grundlehren der Mathematischen Wissenschaften. Springer, Heidelberg (1998)
17. Roubi, A.: Convergence of prox-regularization methods for generalized fractional programming. RAIRO Rech. Oper. **36**, 73–94 (2002)
18. Salmon, G., Strodiot, J.J., Nguyen, V.H.: A bundle method for solving variational inequalities. SIAM J. Optim. **14**, 869–893 (2004)
19. Sheu, R.L., Lin, J.Y.: Solving continuous min-max problems by an iterative entropic regularization method. J. Optim. Theory Appl. **121**, 597–612 (2004)
20. Strodiot, J.J., Nguyen, V.H. : On the numerical treatment of the inclusion $0 \in \partial f(x)$. In: Moreau, J.J., Panagiotopoulos, P.D., Strang, G. (eds.) Topics in Nonsmooth Mechanics, Birkäuser Verlag, Basel (1988)