

A multiplier method with a class of penalty functions for convex programming [‡]

Rómulo Castillo [§] Luiz C. Matioli [§] Clavel Quintana [¶]

January 20, 2016

Abstract

We consider a class of augmented Lagrangian methods for solving convex programming problems with inequality constraints. This class involves a family of penalty functions and specific values of parameters $p, q, \tilde{y} \in R$ and $c > 0$. The penalty family includes the classical modified barrier and the exponential function. The associated proximal method for solving the dual problem is also considered. Convergence results are shown, specifically we prove that any limit point of the primal and the dual sequence generated by the algorithms are optimal solutions of the primal and dual problem respectively.

Key words: Multiplier methods, proximal point methods, convex programming.

1 Introduction

Consider the convex programming problem given by

$$(P) \quad f^* = \inf\{f(x) : g_i(x) \leq 0 \quad i = 1, \dots, m\}$$

[‡]This work was partially supported by CDCHT-Venezuela, CNPq-Brazil (Prosul - 490045/2010-3, Universal - 477611/2013-3, Individual grants 308413/2009-1 and 307714/2011-0) and PNPd/CAPEs-Brazil.

[§]Department of Mathematics, Federal University of Paraná. Cx. Postal 19081, 81531-980. Curitiba, PR, Brazil. E-mail: romulo.castillo@ufsc.br, matioli@ufpr.br.

[¶]Department of Mathematics, Universidad Centro Occidental Lisandro Alvarado. E-mail: clcarlone@gmail.com.

where $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ for $i = 1, \dots, m$ are closed proper convex functions. Let $\ell : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be the usual Lagrangian function defined by

$$\ell(x, \mu) = f(x) + \sum_{i=1}^m \mu_i g_i(x). \quad (1)$$

The dual convex problem associated with (P) is defined as

$$(D) \quad d^* = \inf\{-d(\mu) : \mu \geq 0\}$$

where $d(\mu) = \inf\{\ell(x, \mu) : x \in \mathbb{R}^n\}$.

We suppose that the following conditions are satisfied:

- (A1) The set of optimal solutions of problem (P) is nonempty and compact.
- (A2) There exists $\hat{x} \in \text{dom } f$ such that $g_i(\hat{x}) < 0$ for $i = 1, \dots, m$ (Slater's condition).

Remark 1: Note that (A2) implies that the set of optimal solutions of problem (D) is nonempty and compact and $f^* = d^*$. Furthermore, for each $\beta > d^*$, the level set

$$\{\mu \in \mathbb{R}_+^m : -d(\mu) \leq \beta\}$$

is compact.

2 Primal method

In this section, we present a class of multiplier methods for solving the primal problem. The approach used in the augmented Lagrangian method coincides to that one considered in [1], but in this case with the convergence results for a particular class of penalty functions.

We consider the family \mathcal{F} of penalty functions $\theta : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ with $\text{dom } \theta = (-\infty, b)$ and $0 < b \leq +\infty$, that satisfy the following properties:

- (\theta1) θ is a proper twice differentiable strictly increasing convex function.
- (\theta2) $\lim_{t \rightarrow b^-} \theta'(t) = +\infty$.
- (\theta3) $\lim_{t \rightarrow -\infty} \theta'(t) = 0$.

- (\theta4) There exists $M > 0$ such that $\theta''(t) \geq \frac{1}{M}$ for all $t \in [0, b]$.

Consider $\theta \in \mathcal{F}$, $p, q \in \mathbb{R}$. The penalty function with shift $\hat{P}^{(p,q)} : \mathbb{R}^m \times \mathbb{R}_{++}^m \times \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined by

$$\hat{P}^{(p,q)}(y, \mu, r, c) = \sum_{i=1}^m P^{(p,q)}(y_i, \mu_i, r, c), \quad (2)$$

where $P^{(p,q)} : \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R} \cup \{+\infty\}$ is given by

$$P^{(p,q)}(y_i, \mu_i, r, c) = r \frac{\mu_i^{p-q}}{c} \left[\theta \left(\frac{y_i}{\mu_i^{p-1} r} + \tilde{y}_i \right) - \theta(\tilde{y}_i) \right] \quad (3)$$

with \tilde{y}_i satisfying $\theta'(\tilde{y}_i) = c \mu_i^q$, for all $i = 1, \dots, m$ and $0 < \underline{r} < r < \bar{r}$. By [9, Corollary 23.5.1],

$$\theta'(\tilde{y}_i) = c \mu_i^q \quad \Leftrightarrow \quad \tilde{y}_i = (\theta^*)'(c \mu_i^q), \quad (4)$$

where θ^* is the conjugate function of θ .

In almost all cases known we take $q = 1$ but the generalization allows us to introduce some variants in the approach.

Note that, for all $i = 1, \dots, m$, we have $P^{(p,q)}(0, \mu_i, r, c) = 0$ and

$$(P^{(p,q)})'_1(y_i, \mu_i, r, c) = \frac{\mu_i^{1-q}}{c} \theta' \left(\frac{y_i}{\mu_i^{p-1} r} + \tilde{y}_i \right),$$

where $(P^{(p,q)})'_1 = \frac{\partial P^{(p,q)}}{\partial y_i}$.

So, using the definition of \tilde{y}_i , we obtain, for all $i = 1, \dots, m$,

$$(P^{(p,q)})'_1(0, \mu_i, r, c) = \frac{\mu_i^{1-q}}{c} \theta'(\tilde{y}_i) = \mu_i. \quad (5)$$

As was pointed out in [1], the shift given in (3) is a translation that allows us to write (5).

We define the augmented Lagrangian function $L^{(p,q)} : \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R} \cup \{+\infty\}$ by

$$L^{(p,q)}(x, \mu, r, c) = f(x) + \sum_{i=1}^m P^{(p,q)}(g_i(x), \mu_i, r, c),$$

with $P^{(p,q)}$ defined in (3). Next we state the algorithm that define the multiplier method for solving the problem (P).

Algorithm 1.

Data: $p, q \in \mathbb{R}$, $c > 0$, $0 < \underline{r} < \bar{r}$, $r^0 \in (\underline{r}, \bar{r})$, $\mu^0 \in \mathbb{R}_{++}^m$.
 Let $\tilde{y}_i^0 \in \mathbb{R}$ such that $\theta'(\tilde{y}_i^0) = c(\mu_i^0)^q$, for $i = 1, \dots, m$.
 $k = 0$
 REPEAT

$$\begin{aligned}
& \text{Compute } x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ L^{(p,q)}(x, \mu^k, r^k, c) \right\}. \\
& \mu_i^{k+1} = \frac{(\mu_i^k)^{1-q}}{c} \theta' \left(\frac{g_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \tilde{y}_i^k \right), \quad \text{for } i = 1, \dots, m. \\
& \tilde{y}_i^{k+1} = (\theta^*)' \left(c(\mu_i^{k+1})^q \right), \quad \text{for } i = 1, \dots, m. \\
& \text{Choose } r^{k+1} \in (\underline{r}, \bar{r}). \\
& k = k + 1.
\end{aligned}$$

Observe that, for all $p, q \in \mathbb{R}$, $c > 0$ and $k \geq 0$,

$$0 \in \partial_x L^{(p,q)}(x^{k+1}, \mu^k, r^k, c) \Leftrightarrow 0 \in \partial_x \ell(x^{k+1}, \mu^{k+1}) \quad (6)$$

where ℓ is the Lagrangian function defined in (1).

2.1 Examples

In this section we discuss the above approach by considering some particular penalty functions in the family \mathcal{F} .

Example 2.1.

Consider $\theta_1(t) = -\log(a - t)$, with $a > 0$. Substituting in (3), we have

$$P^{(p,q)}(y_i, \mu_i, r, c) = \frac{r\mu_i^{p-q}}{c} \log \left(\frac{a - \tilde{y}_i}{a - \tilde{y}_i - \frac{y_i}{\mu_i^{p-1} r}} \right).$$

From (4), $\theta_1'(\tilde{y}_i) = \frac{1}{a - \tilde{y}_i} = c\mu_i^q$. So, $a - \tilde{y}_i = \frac{1}{c\mu_i^q}$ and

$$\begin{aligned}
P^{(p,q)}(y_i, \mu_i, r, c) &= \frac{r\mu_i^{p-q}}{c} \log \left(\frac{\frac{1}{c\mu_i^q}}{\frac{\mu_i^{p-1} r - c\mu_i^q y_i}{c\mu_i^q \mu_i^{p-1} r}} \right) \\
&= \frac{r\mu_i^{p-q}}{c} \log \left(\frac{r}{r - c\mu_i^{q-p+1} y_i} \right) \\
&= -\frac{r\mu_i^{p-q}}{c} \log \left(1 - \frac{c y_i}{r\mu_i^{p-q-1}} \right). \quad (7)
\end{aligned}$$

The penalty function (7) represents a family of modified log-barrier penalty functions which yields to different multiplier methods.

Note that if we choose $p = 2, q = c = 1$, we get

$$P^{(2,1)}(y_i, \mu_i, r, 1) = -r\mu_i \log \left(1 - \frac{y_i}{r} \right),$$

which is the modified log-barrier penalty function considered by Polyak, [7].

If we choose $p = q = c = 1$ we get

$$P^{(1,1)}(y_i, \mu_i, r, 1) = r \log \left(1 - \frac{\mu_i y_i}{r} \right),$$

which corresponds to the M^2BF penalty function considered in [6]. They didn't show convergence results for the primal sequence generated by the multiplier method. A remarkable fact is that our convergence result include it.

Choosing $p = p, q = 0, c = 1$ we get

$$P^{(p,0)}(y_i, \mu_i, r, 1) = -r \mu_i^p \log \left(1 - \frac{y_i}{\mu_i^{p-1} r} \right),$$

considered in [2] for $p \geq 2$.

Note that from proposition 4.4, convergence results for new particular cases can be gotten considering $p + q \geq 2$. For example, with $p = 2.5, q = 2, c = 1$ we get

$$P^{(2.5,2)}(y_i, \mu_i, r, 1) = -r \sqrt{\mu_i} \log \left(1 - \frac{\sqrt{\mu_i} y_i}{r} \right)$$

Example 2.2.

Similarly, if we consider $\theta_2(t) = \frac{1}{a-t} - a, a \geq 0$ with $\theta'_2(\tilde{y}_i) = \frac{1}{(a-\tilde{y}_i)^2} = c \mu_i^q$ we get

$$\begin{aligned} P^{(p,q)}(y_i, \mu_i, r, c) &= \frac{r \mu_i^{p-q}}{c} \left(\frac{1}{a - \frac{y_i}{\mu_i^{p-1} r} - \tilde{y}_i} - \frac{1}{a - \tilde{y}_i} \right) \\ &= r \mu_i \left(\frac{(y_i/r)}{1 - \sqrt{c} \mu_i^{\frac{q}{2}-p+1} (y_i/r)} \right) \end{aligned}$$

If $c = p = 1$ and $q = 0$ we get $P^{(1,0)}(y_i, \mu_i, r, 1) = r \mu_i \left(\frac{(y_i/r)}{1-(y_i/r)} \right)$, the modified inverse barrier multiplier method considered by Polyak [7]. In the same way, if we consider any translation of the inverse barrier function $\theta(t) = \frac{1}{1-t} - 1 \in \mathcal{F}$ given by θ_2 , our approach yields to the modified inverse barrier penalty function.

Example 2.3.

Consider $\theta_3(t) = e^t + a$, $a \geq 0$ with $\theta'_3(\tilde{y}_i) = e^{\tilde{y}_i} = c \mu_i^q$ we get

$$\begin{aligned} P^{(p,q)}(y_i, \mu_i, r, c) &= \frac{r \mu_i^{p-q}}{c} \left(e^{\frac{y_i}{\mu_i^{p-1} r} + \tilde{y}_i} - e^{\tilde{y}_i} \right) \\ &= \frac{r \mu_i^{p-q}}{c} \left(e^{\tilde{y}_i} \left(e^{\frac{y_i}{\mu_i^{p-1} r}} - 1 \right) \right) \\ &= r \mu_i^p \left(e^{\frac{y_i}{\mu_i^{p-1} r}} - 1 \right) \end{aligned}$$

If $p = 1$ we get $P^{(1,q)}(y_i, \mu_i, r, c) = r \mu_i \left(e^{\frac{y_i}{r}} - 1 \right)$ the classical exponential multiplier method.

3 Dual method

In this section we consider the proximal method associated with the multiplier one. The distance-like function to be used was introduced in [1], in this case, involving the conjugate functions of members in the family \mathcal{F}

The conjugate function of θ , namely $\varphi = \theta^*$, satisfies the properties given in next proposition:

Proposition 3.1. *Consider $\theta \in \mathcal{F}$. Then the function $\varphi = \theta^*$, which is the conjugate function of θ , verifies the following properties:*

- ($\varphi 1$) φ is a strictly convex and differentiable function on $\text{dom } \varphi = (0, +\infty)$.
- ($\varphi 2$) $\varphi(\kappa) = 0$, $\kappa > 0$.
- ($\varphi 3$) $\varphi'(\kappa) = 0$.
- ($\varphi 4$) $\lim_{s \rightarrow 0^+} \varphi'(s) = -\infty$.
- ($\varphi 5$) $\lim_{s \rightarrow +\infty} \varphi'(s) = b$.
- ($\varphi 6$) There exists $M > 0$ such that $\varphi''(s) \leq M$ for $s \geq \kappa$.

Proof: We show ($\varphi 4$), ($\varphi 5$) and ($\varphi 6$), for the rest see [8], Proposition 3.1.

From ($\theta 2$)

$$\lim_{s \rightarrow +\infty} \varphi'(s) = \lim_{s \rightarrow +\infty} (\theta')^{-1}(s) = b,$$

and so ($\varphi 5$) holds.

By other hand

$$\varphi''(s) = [(\theta')^{-1}]'(s) = \frac{1}{\theta''[(\theta')^{-1}(s)]} = \frac{1}{\theta''[\varphi'(s)]}. \quad (8)$$

From $(\varphi 1)$, $(\varphi 2)$ and $(\varphi 3)$, $\varphi'(s) \geq 0$ for $s \geq \kappa$, and by $(\varphi 5)$ and continuity of φ' we can consider $\varphi'(s)$ satisfying

$$s \geq \kappa \Rightarrow 0 \leq \varphi'(s) \leq +\infty. \quad (9)$$

Finally, from (8), (9) and $(\theta 6)$, we have $\varphi''(s) \leq \frac{1}{M}$ for $s \geq \kappa$. \square

Remark 2: From $(\varphi 1)$, $(\varphi 2)$ and $(\varphi 3)$ it follows φ is decreasing on $(0, \kappa)$ and increasing on $(\kappa, +\infty)$.

We call Φ to the class of functions that verify the properties in Proposition 3.1.

3.0.1 The distance-like function

We use the distance-like function introduced in [1] involving the function $\varphi \in \Phi$.

Given $p, q \in \mathbb{R}$, the distance-like function $\hat{d}_\varphi^{(p,q)} : \mathbb{R}_+^m \times \mathbb{R}_{++}^m \rightarrow \{+\infty\}$ is defined by

$$\hat{d}_\varphi^{(p,q)}(s, \mu) = \sum_{i=1}^m d_\varphi^{(p,q)}(s_i, \mu_i) \quad (10)$$

where for $i = 1, \dots, m$

$$d_\varphi^{(p,q)}(s_i, \mu_i) = \frac{\mu_i^{p-q}}{c} \varphi\left(\frac{cs_i}{\mu_i^{1-q}}\right) - \frac{\mu_i^{p-q}}{c} \varphi(c\mu_i^q) - \mu_i^{p-1}(\varphi)'(c\mu_i^q)(s_i - \mu_i). \quad (11)$$

Note

$$\left(d_\varphi^{(p,q)}\right)'_1(s_i, \mu_i) = \frac{\partial d(s_i, \mu_i)}{\partial s_i} = \mu_i^{p-1} \left[(\varphi)' \left(\frac{cs_i}{\mu_i^{1-q}} \right) - (\varphi)'(c\mu_i^q) \right]. \quad (12)$$

It was proved in [1] that 10 is a divergence measure [5, Def. 2.1] for a general class of functions that includes the functions in our families. So we consider (10) with $\varphi \in \Phi$.

3.0.2 The proximal point method

The proximal point method for solving the dual problem (D) using the distance-like function (10) generates a sequence $\{\mu^k\}$ such that $\mu^0 \in \mathbb{R}_{++}^m$ and

$$\mu^{k+1} = \operatorname{argmin}\{-d(\mu) + r^k \hat{d}_\varphi^{(p,q)}(\mu, \mu^k)\}, \quad (13)$$

where $0 < \underline{r} < r^k < \bar{r}$ and $\varphi \in \Phi$.

Remark 3:

Note from (6) that, x^{k+1} minimize $L^{(p,q)}(x, \mu^k, r^k, c)$ se e somente se x^{k+1} minimize $l(x, \mu^k, r^k, c)$.

On the other side, we have

$$\begin{aligned}
 d(\mu) &= \inf\{l(x, \mu) : x \in \mathbb{R}^n\} \\
 &= \inf\{f(x) + \sum_{i=1}^m \mu_i g_i(x)\} \\
 &\leq f(x^{k+1}) + \sum_{i=1}^m \mu_i g_i(x^{k+1}) \\
 &= f(x^{k+1}) + \sum_{i=1}^m \mu_i g_i(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} g_i(x^{k+1}) - \sum_{i=1}^m \mu_i^{k+1} g_i(x^{k+1}) \\
 &= f(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} g_i(x^{k+1}) + \sum_{i=1}^m (\mu_i - \mu_i^{k+1}) g_i(x^{k+1}) \\
 &= d(\mu_i^{k+1}) + \sum_{i=1}^m (\mu_i - \mu_i^{k+1}) g_i(x^{k+1}) \\
 &= d(\mu_i^{k+1}) + (f_1(x^{k+1}), \dots, f_m(x^{k+1}))^t (\mu - \mu^{k+1}).
 \end{aligned}$$

Therefore

$$(f_1(x^{k+1}), \dots, f_m(x^{k+1}))^T \in \partial d(\mu^{k+1}). \quad (14)$$

By updating formulae of the multiplier, for $i = 1, \dots, m$

$$\mu_i^{k+1} = \frac{(\mu_i^k)^{1-q}}{c} \theta' \left(\frac{g_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \tilde{y}_i^k \right)$$

and using $(\theta')^{-1} = (\varphi)'$ we have, for $i = 1, \dots, m$

$$\begin{aligned}
 \frac{c\mu_i^{k+1}}{(\mu_i^k)^{1-q}} &= \theta' \left(\frac{g_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \tilde{y}_i^k \right), \\
 (\theta')^{-1} \left(\frac{c\mu_i^{k+1}}{(\mu_i^k)^{1-q}} \right) &= \frac{g_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \tilde{y}_i^k, \\
 (\theta')^{-1} \left(\frac{c\mu_i^{k+1}}{(\mu_i^k)^{1-q}} \right) &= \frac{g_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + (\varphi)'(c(\mu^k)^q).
 \end{aligned}$$

Then, for $i = 1, \dots, m$,

$$g_i(x^{k+1}) = (\mu_i^k)^{p-1} r^k \left[(\theta')^{-1} \left(\frac{c\mu_i^{k+1}}{(\mu_i^k)^{1-q}} \right) - (\varphi)'(c(\mu^k)^q) \right] \quad (15)$$

and so from (14) and (15),

$$(\mu_i^k)^{p-1} r^k \left[(\theta')^{-1} \left(\frac{c\mu_i^{k+1}}{(\mu_i^k)^{1-q}} \right) - (\varphi)'(c(\mu^k)^q) \right] \in \partial d(\mu^{k+1})$$

which is the optimality condition for (13).

So next proposition ensures that the sequence (μ^k) generated by the Algorithm 1 coincides with the sequence given in (13).

Proposition 3.2. *Let $\{\widehat{\mu}^k\}$ be the sequence generated by (13) for solving the dual problem (D) and let $\{x^k\}$ and $\{\mu^k\}$ be the sequences generated by the Algorithm 1 for solving the primal problem (P). If $\mu^0 = \widehat{\mu}^0$, then $\mu^k = \widehat{\mu}^k$ for all $k \geq 0$.*

Proof. Follows from Thm 7.1 in [4]. □

4 Convergence results

This section is inspired by the convergence results presented in [10] and [8]. We show a convergence study of the sequences generated by (13) and by the Algorithm 1.

Proposition 4.1. *The sequence $\{-d(\mu^k)\}$ is non-increasing and bounded, so it converges.*

Proof: : From (13), since $\widehat{d}_\varphi^{(p,q)}(\mu, \mu) \geq 0$ and $r^k > 0$, we have, for all $\mu \in \mathbb{R}_{++}^m$

$$-d(\mu^{k+1}) \leq -d(\mu^{k+1}) + r^k \widehat{d}_\varphi^{(p,q)}(\mu^{k+1}, \mu^k) \leq -d(\mu^k) + r^k \widehat{d}_\varphi^p(\mu^k, \mu^k),$$

but $\widehat{d}_\varphi^{(p,q)}(\mu^k, \mu^k) = 0$ and so $-d(\mu^{k+1}) \leq -d(\mu^k)$, then $\{-d(\mu^k)\}$ is non-increasing. Furthermore, by weak duality, $-d(\mu^k) \geq -f^*$, and consequently $\{-d(\mu^k)\}$ is convergent. □

Proposition 4.2. *The sequence $\{\mu^k\}$ is bounded.*

Proof: By (A2), the set of optimal Lagrange multipliers is nonempty and compact, so one level set of $-d$ is compact. Since $-d$ is a closed proper convex function then all its level sets are compact, in particular $\Lambda = \{\mu \in \mathbb{R}_+^m : -d(\mu) \leq -d(\mu^0)\}$. But by proposition 4.1, $\mu^k \in \Lambda$ for all k , so $\{\mu^k\}$ is bounded. \square

Lemma 4.1. *Let θ be a strictly convex function, $t, q \in R$, $c > 0$ and \tilde{y} as defined in (4) such that $t + \tilde{y} \in \text{dom } \theta$, then $t \frac{\theta'(t + \tilde{y})}{c} \geq t\mu^q$.*

Proof: If $t > 0$ then $t + \tilde{y} > \tilde{y}$, since θ is an strictly convex function θ' is an increasing function, $\theta'(t + \tilde{y}) > \theta'(\tilde{y}) = c\mu^q$, then

$$\frac{\theta'(t + \tilde{y})}{c} > \mu^q \text{ and finally } t \frac{\theta'(t + \tilde{y})}{c} > t\mu^q.$$

In a similar way for $t < 0$. \square

Proposition 4.3. *Let v, w be positive numbers.*

- 1) *We have $v > w$ if and only if $\left(d_\varphi^{(p,q)}\right)'_1(v, w) > 0$.*
- 2) *If $v > w$, then*

$$d_\varphi^{(p,q)}(v, w) \geq \frac{\left[\left(d_\varphi^{(p,q)}\right)'_1(v, w)\right]^2}{2cMw^{p+q-2}}.$$

Proof: 1) Since φ is an increasing function, we have

$$\begin{aligned} v > w &\Leftrightarrow \frac{cv}{w^{1-q}} > cw^q \\ &\Leftrightarrow (\varphi)' \left(\frac{cv}{w^{1-q}} \right) > (\varphi)'(cw^q) \\ &\Leftrightarrow (\varphi)' \left(\frac{cv}{w^{1-q}} \right) - (\varphi)'(cw^q) > 0 \\ &\Leftrightarrow w^{p-1} \left((\varphi)' \left(\frac{cv}{w^{1-q}} \right) - (\varphi)'(cw^q) \right) > 0 \\ &\Leftrightarrow \left(d_\varphi^{(p,q)}\right)'_1(v, w) > 0. \end{aligned}$$

2) Consider the quadratic function

$$q(t) = q(v) + (t - v) \left(d_\varphi^{(p,q)}\right)'_1(v, w) + \frac{1}{2}(t - v)^2 cMw^{p+q-2},$$

so

$$q'(t) = \left(d_\varphi^{(p,q)}\right)'_1(v, w) + (t - v)cMw^{p+q-2}$$

and

$$q'(t) = 0 \Leftrightarrow \left(d_{\varphi}^{(p,q)}\right)'_1(v, w) + (t - v)cMw^{p+q-2} = 0$$

then

$$t^* = \frac{-\left(d_{\varphi}^{(p,q)}\right)'_1(v, w)}{cMw^{p+q-2}} + v \quad (16)$$

is a minimizer of $q(\cdot)$.

By item 1) since $v > w > 0$ we have $\left(d_{\varphi}^{(p,q)}\right)'_1(v, w) > 0$ and so from (16) we have $t^* < v$, using the fact that $d_{\varphi}^{(p,q)}(w, w) = 0$ and the mean value theorem, there exists $\widehat{w} \in [w, v]$ such that

$$\begin{aligned} \left(d_{\varphi}^{(p,q)}\right)'_1(v, w) &= \left(d_{\varphi}^{(p,q)}\right)'_1(v, w) - \left(d_{\varphi}^{(p,q)}\right)'_1(w, w) \\ &= (v - w) \left(d_{\varphi}^{(p,q)}\right)''_1(\widehat{w}, w) \\ &= (v - w)w^{p+q-2}(\varphi)''\left(\frac{c\widehat{w}}{w^{1-q}}\right)c \\ &\leq (v - w)w^{p+q-2}cM. \end{aligned}$$

Hence $\frac{\left(d_{\varphi}^{(p,q)}\right)'_1(v, w)}{cMy^{p+q-2}} \leq v - w$ and $w \leq v - \frac{\left(d_{\varphi}^{(p,q)}\right)'_1(v, w)}{cMy^{p+q-2}} = t^*$, so $w \leq t^* < v$.

Using again the mean value theorem, for all $t \in [w, v]$, there exists $\widehat{t} \in (t, v)$ such that

$$\begin{aligned} \left(d_{\varphi}^{(p,q)}\right)'_1(v, w) - \left(d_{\varphi}^{(p,q)}\right)'_1(t, w) &= (v - t) \left(d_{\varphi}^{(p,q)}\right)''_1(\widehat{t}, w) \\ &= (v - t)w^{p+q-2}(\varphi)''\left(\frac{c\widehat{t}}{w^{1-q}}\right)c \\ &\leq (v - w)w^{p+q-2}cM, \end{aligned}$$

So, for $t \in [w, v]$

$$\begin{aligned} \left(d_{\varphi}^{(p,q)}\right)'_1(v, w) &\leq \left(d_{\varphi}^{(p,q)}\right)'_1(t, w) + (v - t)w^{p+q-2}cM \text{ and so} \\ \left(d_{\varphi}^{(p,q)}\right)'_1(t, w) &\geq \left(d_{\varphi}^{(p,q)}\right)'_1(v, w) + (t - v)w^{p+q-2}cM = q'(t), \end{aligned}$$

then

$$\left(d_{\varphi}^{(p,q)}\right)'_1(t, w) \geq q'(t) \text{ for all } w \leq t \leq v.$$

Since $w \leq t^* < v$, integrating from t^* to v we have

$$d_\varphi^{(p,q)}(v, w) - d_\varphi^{(p,q)}(t^*, w) \geq q(v) - q(t^*).$$

So $d_\varphi^{(p,q)}(v, w) \geq q(v) - q(t^*) + d_\varphi^{(p,q)}(t^*, w) \geq q(v) - q(t^*)$ and

$$d_\varphi^{(p,q)}(v, w) \geq q(v) - q(v) - (t^* - v) \left(d_\varphi^{(p,q)} \right)'_1(v, w) - \frac{1}{2}(t^* - v)^2 c M w^{p+q-2}. \quad (17)$$

From (16) and (17) we have

$$d_\varphi^{(p,q)}(v, w) \geq \frac{\left[\left(d_\varphi^{(p,q)} \right)'_1(v, w) \right]^2}{2c M w^{p+q-2}}.$$

□

Proposition 4.4. Consider $p + q - 2 \geq 0$ and the sequences $\{x^k\}, \{\mu^k\}$ generated by the algorithm 1, then

- 1) $\{[g_i(x^k)]_+\} \rightarrow 0$ for $i = 1, \dots, m$.
- 2) $\mu_i^k g_i(x^k) \rightarrow 0$ for $i = 1, \dots, m$.
- 3) $f(x^k) \rightarrow f^*$. Moreover, the sequences $\{x^k\}$ and $\{\mu^k\}$ are bounded and each of their limit points are optimal solutions of problems P and D respectively.

Proof: 1) Suppose by contradiction that there exists an infinite set of indices $\{k_j\}$ and $\epsilon > 0$ such that $[f_l(x^{k_j})]_+ > \epsilon$ for some $l \in \{1, \dots, m\}$.

From (15) and (12) we have

$$f_l(x^{k_j+1}) = r^{k_j} \left(d_\varphi^{(p,q)} \right)'_1(\mu_l^{k_j+1}, \mu_l^{k_j}) \text{ for } l \in \{1, \dots, m\},$$

since $0 < \underline{r} < r^k < \bar{r}$ for all k , we have $\frac{1}{r^k} > \frac{1}{\bar{r}}$, so

$$\left(d_\varphi^{(p,q)} \right)'_1(\mu_l^{k_j+1}, \mu_l^{k_j}) = \frac{f_l(x^{k_j+1})}{r^{k_j}} > \frac{\epsilon}{\bar{r}} > 0. \quad (18)$$

By lemma 4.3 item 1) we have $\mu_l^{k_j+1} > \mu_l^{k_j}$.

Since $\mu^{k+1} = \operatorname{argmin}\{-d(\mu) + \tilde{d}_\varphi^{(p,q)}(\mu, \mu^k)\}$, we have

$$-d(\mu^{k_j+1}) + \tilde{d}_\varphi^{(p,q)}(\mu^{k_j+1}, \mu^{k_j}) \leq -d(\mu^{k_j}) + \tilde{d}_\varphi^{(p,q)}(\mu^{k_j}, \mu^{k_j}) = -d(\mu^{k_j}),$$

$$\text{then } d(\mu^{k_j+1}) - d(\mu^{k_j}) \geq \tilde{d}_\varphi^{(p,q)}(\mu^{k_j+1}, \mu^{k_j}) = \sum_{i=1}^m d_\varphi^{(p,q)}(\mu_i^{k_j+1}, \mu_i^{k_j}) \geq d_\varphi^{(p,q)}(\mu_l^{k_j+1}, \mu_l^{k_j}).$$

By proposition 4.2 $\{\mu^k\}$ is bounded, that is, there exists $N > 0$ such that

$$\mu_l^k \leq \|\mu^k\| \leq N \text{ for all } k, \text{ and so } \frac{1}{\mu_l^k} \geq \frac{1}{N}. \quad (19)$$

By item 2) of lemma 4.3 and using 18, we have

$$\begin{aligned} d(\mu^{k_j+1}) - d(\mu^{k_j}) &\geq \frac{1}{2cM} \frac{\left[\left(d_\varphi^{(p,q)} \right)'_1 (\mu_l^{k_j+1}, \mu_l^{k_j}) \right]^2}{(\mu_l^{k_j})^{p+q-2}} \\ &\geq \frac{\epsilon^2}{2cMN^{p+q-2}} = \delta > 0. \end{aligned}$$

So $d(\mu^{k_j+1}) \geq d(\mu^{k_j}) + \delta$ which is a contradiction because d is bounded above.

We now proof 2).

Suppose by contradiction that there exist $l \in \{1, \dots, m\}$, $\epsilon > 0$ and an infinite set of indices $\{k_j\}$ such that

$$|\mu_l^{k_j+1} g_l(x^{k_j+1})| \geq \epsilon, \text{ so using (19) we have } |g_l(x^{k_j+1})| \geq \frac{\epsilon}{N}. \quad (20)$$

But $\{[g_i(x^k)]_+\}$ converges to 0 for all $i = 1, \dots, m$, hence $g_l(x^{k_j+1}) \geq \frac{\epsilon}{N}$ is true only for a finite set of index k_j , so we can consider without lost of generality

$$g_l(x^{k_j+1}) \leq -\frac{\epsilon}{N} \text{ for all } j. \quad (21)$$

Since $(g_1(x^{k+1}), \dots, g_m(x^{k+1}))^t \in \partial d(\mu^{k+1})$, and d is a concave function, then

$$\sum_{i=1}^m g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \leq d(\mu^{k_j+1}) - d(\mu^{k_j}). \quad (22)$$

Since $\mu_i^{k+1} = \frac{(\mu_i^k)^{1-q}}{c} \theta' \left(\frac{g_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \tilde{y}^k \right)$ for $i = 1, \dots, m$ we have

$$\begin{aligned} g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) &= \mu_i^{k_j} g_i(x^{k_j+1}) \left(\frac{\mu_i^{k_j+1}}{\mu_i^{k_j}} - 1 \right) \\ &= \mu_i^{k_j} g_i(x^{k_j+1}) \left(\frac{1}{c(\mu_i^{k_j})^q} \theta' \left(\frac{g_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1} r^{k_j}} + \tilde{y}^{k_j} \right) - 1 \right) \end{aligned} \quad (23)$$

$$= \frac{g_i(x^{k_j+1})}{c(\mu_i^{k_j})^{q-1}} \theta' \left(\frac{g_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1} r^{k_j}} + \tilde{y}^{k_j} \right) - \mu_i^{k_j} g_i(x^{k_j+1}). \quad (24)$$

Since θ is a strictly convex function, θ' is increasing, using (24) and lemma 4.1 we have

$$\begin{aligned} g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) &= \frac{r^{k_j}(\mu_i^{k_j})^{p-q}}{c} \left[\frac{g_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1}r^{k_j}} \theta' \left(\frac{g_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1}r^{k_j}} + \tilde{y}^{k_j} \right) \right] - \mu_i^{k_j} g_i(x^{k_j+1}) \\ &\geq \mu_i^{k_j} g_i(x^{k_j+1}) - \mu_i^{k_j} g_i(x^{k_j+1}) \\ &= 0, \end{aligned}$$

hence, by (22)

$$0 \leq \sum_{i=1}^m g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \leq d(\mu^{k_j+1}) - d(\mu^{k_j}),$$

by proposition 4.1 $\{d(\mu^k)\}$ is convergent, so $\lim_{j \rightarrow +\infty} [d(\mu^{k_j+1}) - d(\mu^{k_j})] = 0$.

Then

$$\lim_{j \rightarrow +\infty} \sum_{i=1}^m g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = 0.$$

Since $g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \geq 0$ for all $i = 1, \dots, m$

we have $\lim_{j \rightarrow +\infty} g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = 0$ for all $i = 1, \dots, m$.

Then from (23),

$$\lim_{j \rightarrow +\infty} \mu_i^{k_j} g_i(x^{k_j+1}) \left[\frac{1}{c(\mu_i^{k_j})^q} \theta' \left(\frac{g_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1}r^{k_j}} + \tilde{y}^{k_j} \right) - 1 \right] = 0 \text{ for all } i = 1, \dots, m. \quad (25)$$

By other hand, using (21)

$$\frac{f_l(x^{k_j+1})}{r^{k_j}(\mu_l^{k_j})^{p-1}} + \tilde{y}^{k_j} \leq \frac{-\epsilon}{r^{k_j}(\mu_l^{k_j})^{p-1}N} + \tilde{y}^{k_j} < \tilde{y}^{k_j},$$

since θ' is increasing

$$\theta' \left(\frac{f_l(x^{k_j+1})}{r^{k_j}(\mu_l^{k_j})^{p-1}} + \tilde{y}^{k_j} \right) < \theta'(\tilde{y}^{k_j}) = c(\mu_l^{k_j})^q,$$

so

$$\frac{1}{c(\mu_l^{k_j})^q} \theta' \left(\frac{f_l(x^{k_j+1})}{r^{k_j}(\mu_l^{k_j})^{p-1}} + \tilde{y}^{k_j} \right) < 1. \quad (26)$$

From (25) and (26)

$$\lim_{j \rightarrow +\infty} \mu_l^{k_j} g_l(x^{k_j+1}) = 0.$$

Since

$$\lim_{j \rightarrow +\infty} g_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = 0$$

in particular for the indice l , $\lim_{j \rightarrow +\infty} \mu_l^{k_j+1} g_l(x^{k_j+1}) = 0$, which is a contradiction.

Consequently

$$\mu_i^k g_i(x^k) \rightarrow 0$$

para todo $i = 1, \dots, m$.

3) By 1) $\{x^k\}$ is asymptotically feasible, so that given $\epsilon > 0$ and k sufficiently large.

$$f(x^k) \geq f^* - \epsilon \quad (27)$$

Note by Remark 1

$$f^* = d^* \geq d(\mu^k) = \inf_x \{l(x, \mu^k)\} = l(x^k, \mu^k) = f(x^k) + \sum_{i=1}^m \mu_i^k g_i(x^k). \quad (28)$$

Using item 2), (27) and (28) we have that for all $\epsilon > 0$

$$f^* - \epsilon \leq f(x^k) \leq f^* - \sum_{i=1}^m \mu_i^k g_i(x^k) < f^* + \epsilon$$

for k sufficiently large.

Therefore

$$\lim_{k \rightarrow +\infty} f(x^k) = f^*. \quad (29)$$

From items 1) and 3), given $\epsilon > 0$, for $i = 1, \dots, m$ and k sufficiently large

$$f(x^k) < f^* + \epsilon, \text{ and } g_i(x^k) < \epsilon. \quad (30)$$

From (A1), for any α, β , the set

$$\{x \in R^n : g_i(x) < \alpha, f(x) < \beta \text{ for } i = 1, \dots, m\}$$

is compact, see Corollary 20 in [3].

Taking $\alpha = \epsilon$ and $\beta = f^* + \epsilon$ we have that the set

$$\Gamma = \{x \in \mathbb{R}^n : g_i(x) < \epsilon, f(x) < f^* + \epsilon, \text{ for } i = 1, \dots, m\}$$

is compact and so bounded. Using (30), $x^k \in \Gamma$ for k sufficiently large, therefore the sequence $\{x^k\}$ is bounded.

Note that by proposition 4.2 the sequence $\{\mu^k\}$ is bounded.

Let x^* and μ^* be limit points of the sequences $\{x^k\}$ and $\{\mu^k\}$ respectively. From item 1) and (29) we have x^* is an optimal solution of problem (P). This joint with item 2) and (28), we get $d(\mu^*) = f_0^* = d^*$. Then μ^* is an optimal solution of problem (D).

□

References

- [1] R. Castillo, C.C. Gonzaga, E.W. Karas, and L.C. Matioli. A unified approach to multiplier and proximal methods. Technical report, Federal University of Paraná, 2014.
- [2] R. Castillo, E. Hernández, and J. Campos. A p-order homogeneous like-distance in convex programming. Technical report, Department of Mathematics, Centroccidental Lisandro Alvarado University, 2014.
- [3] A. Fiacco and G. McCormick. *Nonlinear Programming, Sequential Unconstrained Minimization Techniques, Classics in Applied Mathematics*. SIAM, 1990.
- [4] A. Iusem. Augmented Lagrangian methods and proximal point methods for convex optimization. *Investigación Operativa*, 8(1):11–50, 1997.
- [5] A. Iusem, B. Svaiter, and M. Teboulle. Entropy-like proximal methods in convex programming. *Mathematics of Operations Research*, 19(4):790–814, 1994.
- [6] L. C. Matioli and C. C. Gonzaga. A new family of penalty for augmented Lagrangian methods. *Numerical Linear Algebra*, 15:925–944, 2008.
- [7] B. T. Polyak. Modified barrier functions(theory and methods). *Mathematical Programming*, 54:177–222, 1992.
- [8] R. Polyak and M. Teboulle. Nolinear rescaling and proximal-like methods in convex programming. *Mathematical Programming*, 76:667–739, 1997.
- [9] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, New Jersey, 1970.

- [10] A. Ben Tal and M. Zibulevsky. Penalty/barrier multiplier methods for convex programming problems. *SIAM Journal on Optimization*, 7:347–366, 1997.