Journal of Convex Analysis
Volume 17 (2010), No. 3&4, 721–736

A Proximal Extension of the Column Generation Method to Nonconvex Conic Optimization Providing Bounds for the Duality Gap

Alfred Auslender

Institut Camille Jordan, Université Claude Bernard-Lyon I, Lyon, France; and: Département d'Economie, Ecole Polytechnique, Palaiseau, France

Dedicated to Hedy Attouch on the occasion of his 60th birthday.

Received: October 2, 2008
Revised manuscript received: April 5, 2009

In this paper we consider nonconvex conic optimization that covers Standard Nonlinear Programming, Semidefinite Programming, Second Order Cone Programming. To the dual Lagrangian problem, we associate a relaxed primal convex problem, and give bounds for the duality gap. Then we propose a proximal extension of the column generation method of Dantzig-Wolfe algorithm (PECGM) which provides these bounds if we suppose in addition Slater's condition. Finally new applications are given in order to make implementable the step of PECGM for which a nonconvex program is supposed to be solved numerically.

Keywords: Standard nonlinear programming, semidefinite programming, second order cone programming, duality gap, generation column algorithm, proximal method

1. Introduction

Let Q be a nonempty compact set in $\mathbb{R}^n$ and let $f, f_i : Q \to \mathbb{R},\ i = 1, \ldots, m$, be a collection of continuous functions on Q. Consider the standard nonlinear problem

$$(P_{st}) \quad v_{P_{st}} = \min\{f(x) \mid x \in Q \cap E\} \quad \text{with } E = \{x \mid f_i(x) \le 0,\ i = 1, \ldots, m\}.$$

To (P_{st}) we associate the usual dual Lagrangian

$$(D_{st}) \quad v_{D_{st}} = \sup\{\theta(u) \mid u \ge 0\},$$

where

$$L(x, u) = f(x) + \sum_{i=1}^{m} u_i f_i(x) \quad \text{and} \quad \theta(u) = \min\{L(x, u) \mid x \in Q\}$$

are respectively the usual Lagrangian and the associated dual functional.

The starting point of this paper is the article of Magnanti, Shapiro, Wagner [12] where the motivation of these authors is related to the fact that the optimal value $v_{D_{st}}$ ($\le v_{P_{st}}$) of the usual Lagrangian dual problem of (P_{st}) is useful for analyzing nonconvex problems, despite the duality gap. A large number of applications can be found in management science as multi-item production control, resource constrained network

ISSN 0944-6532 / $ 2.50

scheduling, cutting stock, multi-commodity flows. This motivation led the authors of [12] to provide a very important formula through the equality

$$v_{D_{st}} = v_c := \inf\{\xi \mid (\xi, 0) \in W\}, \tag{1}$$

where W is the convex hull of the union of the epigraphs of $f, f_i,\ i = 1, \dots, m$ over Q.

Analyzing this formula they showed that the extension of the column generation method of Dantzig-Wolfe [6] (in short **ECGM**) is a natural algorithm for providing a dual sequence with values converging to $v_{D_{st}}$, even in the non convex case. Let us briefly review **ECGM**.

At iteration k this algorithms solves the master linear program in the variables $\lambda_1, \lambda_2, \dots, \lambda_k$

$$(P_k) \quad \alpha_k = \min \left\{ \sum_{i=1}^{k} \lambda_j f(x_j) \;\middle|\; \sum_{i=1}^{k} \lambda_j f_i(x_j) \le 0,\ i = 1, \dots, m, \right.$$
$$\left. \sum_{i=1}^{k} \lambda_j = 1,\ \lambda_j \ge 0,\ j = 1, \dots k \right\},$$

where the points $x_j \in Q$ for $j = 1, \dots, k$ have been generated previously. At this step the algorithm provides also an optimal solution u^k of the usual dual linear program associated to (P_k) and proceeds by solving

$$x_{k+1} \in \operatorname{argmin}\{L(x, u^k) \mid x \in Q\}. \tag{2}$$

Then adding the column x_{k+1} to the points $x_j, j = 1, \dots, k$ the process is iterated by replacing k by $k + 1$.

The present paper opens several important questions. First, the fact that $v_{D_{st}} = v_c$ does not indicate clearly what are primal variables as usual in any duality theory. Also it is of interest to know if the infimum in formula (1) is attained and to know more if possible, about the primal optimal set. This will be studied in Section 2.2.

The second question is the following. Can we obtain anything better than formula (1), more precisely an upper estimate of the duality gap $v_{P_{st}} - v_{D_{st}}$? As an immediate consequence of the results obtained in Section 2.2, we will give a positive response to this question in Section 2.3 when Q is convex and when the functions $f_i,\ i = 1, \dots, m$ are convex. This will be obtained by using the notion of lack of convexity of an extended real valued function introduced by Aubin and Ekeland in [2]. After extending the definition of duality gap for nonconvex functions f_i as in [2], we will also derive as an immediate consequence a result similar to Theorem C in [2], but with other conditions on the data. Furthermore if we suppose in addition Slater's condition we will be able to obtain those theoretical results by using **ECGM** which provides a primal bounded sequence $\{y^k\}$ giving limit-points y_∞ satisfying these duality gap bounds. This will be the object of Section 3, and we will emphasize that in [12], there are no results concerning the convergence of a primal sequence generated by **ECGM**.

The third question concerns the computation of x_{k+1} in formula (2). Indeed this concerns a nonconvex problem, which is in general impossible to solve by an implementable algorithm. So the third question is: can we exhibit some large classes of nonconvex problems, for which **ECGM** or some variant can effectively compute a point x_{k+1} satisfying (2)? To answer to this question which was not investigated in [12], we will propose in Section 3, a proximal variant of **ECGM** called in short **PECGM** where (2) is replaced by

$$x_{k+1} \in \operatorname{argmin}\left\{L(x, u^k) + \frac{\gamma}{2}||x - y^k||^2 \mid x \in Q\right\}, \tag{3}$$

where y^k and $\gamma \geq 0$ are suitably chosen. In Section 3, Q will be supposed to be convex, and if the data f, f_i are convex on Q and $\gamma > 0$, we will show that the sequence y^k is bounded and that all limit points are optimal solutions of P_{st}. The advantage of using **PECGM** instead of **ECGM** is that the objective function in formula (3) is, as in the proximal method, strongly convex, which provides subroutines for computing x_{k+1} in principle more efficient. Furthermore in the nonconvex case we will enlarge the set of applications by considering semi-convex functions used in particular in numerical methods for other purposes by [9], [5] with some $\gamma > 0$ well chosen. This will be done in Section 4.

Finally, if Standard Nonlinear Programming covers numerous applications, nevertheless there are real-life models that cannot be covered by P_{st}, in particular Semidefinite Programming (SDP) and Second Order Cone Programming (SOCP), which have been widely developped these last three decades. It appears that we can extend all the above questions, and responses as well **PECGM** in a more general framework called nonconvex conic optimization, containing SDP, SOCP, and Standard Nonlinear Programming. This will be done in the whole present paper and is new to the best of our knowledge.

2. Theoretical results

2.1. Preliminaries

In this paper we identify every Euclidean linear space of dimension m with $\mathbb{R}^m$ endowed with the usual inner product: $\langle x, y\rangle = \sum_{i=1}^m x_i y_i$.

For example, let S^m denotes the set of (m, m) symmetric matrices. In this case by identification the inner product is the Frobenius inner product

$$\langle x, y\rangle = tr(xy), \quad \text{where } tr(x) \text{ is the trace of } x.$$

Let K be a pointed closed convex cone in $\mathbb{R}^m$ with nonempty interior int(K). Such a cone defines a partial ordering as follows

$$a \preceq_K b \equiv b - a \in K, \qquad a \prec_K b \equiv b - a \in \operatorname{int}(K),$$

and we refer for the definition and the basic properties of partial ordering to [3], Chapter 2. We note also $a \succeq_K b$ if $-a \preceq_K -b$ and $a \succ_K b$ if $-a \prec_K -b$. The partial orderings we are especially interested in correspond to the three fundamental cones

$$K_1 := \mathbb{R}^m_+,$$

$$K_2 := S^m_+ = \{x \in S^m \text{ semidefinite positive}\},$$

$$K_3 := L^m = \left\{ x \in \mathbb{R}^m \mid x_m \geq \sqrt{x_1^2, + \cdots + x_{m-1}^2} \right\}, \quad \text{the second order cone.}$$

Indeed they are the pillars for the standard nonlinear programming, semidefinite programming (SDP), second order cone programming (SOCP).

Let $K_* = \{x \in \mathbb{R}^m \mid \langle x, y \rangle \geq 0 \; \forall y \in K\}$ the (non negative) polar cone of K. Since dual multipliers will belong to K_* it is important to have a characterization of its elements. Then it is worthwhile to note that for $K = K_i, \; i = 1, 2, 3$ we have

$$K = K_*. \tag{4}$$

However there are other cones which are pointed closed convex with nonempty interior and for which a characterization of K_* is available. This is the case of cones of non-negative polynomials on an infinite interval, or on a semi-infinite interval, or on a finite interval, or for polynomials on a finite interval and trigonometric polynomials where such properties and characterizations have been given by Nesterov ([13], Section 3). A particular interesting cone is the cone of non-negative trigonometric polynomials based on $\cos kt, \; k \in \mathcal{N}$ for $t \in [0, \pi]$, since it has been identified to be the cone of finite autocorrelation sequences (see [1]). As shown in [1] problems where some of the variables are constrained to be finite autocorrelation sequences arise in signal processing and communications with applications in filter design and system identification.

Finally since a set $K \subset \mathbb{R}^m$ is a pointed closed convex cone with nonempty interior if and only if the set K_* is so (see [3], Corollary 2.3.1) then K_* induces a partial ordering allowing the notation $a \succeq_{K_*} b$, for elements $a, b \in \mathbb{R}^m$.

For the rest of this paper K will be a pointed closed convex cone with nonempty interior, Q a nonempty compact subset of $\mathbb{R}^n$ and $F : Q \to \mathbb{R}^m$ a continuous function on Q, with $F(x) = (f_1(x), \ldots, f_m(x))$. For such K, if Q is supposed to be convex we recall that F is said to be K-convex on Q if

$$F(\lambda x_1 + (1-\lambda)x_2) - [\lambda F(x_1) + (1-\lambda)F(x_2)] \preceq_K 0 \quad \forall \lambda \in [0,1], \; \forall x_1, x_2 \in Q.$$

By definition of K_* and from [14], Theorem 13.1 this is equivalent to say that for each $u \in K_*$ the function $x \to \langle F(x), u \rangle$ is convex on Q.

When $K = \mathbb{R}^m_+$, this is equivalent to say that each component $f_i, \; i = 1, \ldots, m$ is convex on Q. When F is affine on $\mathbb{R}^n$ then obviously F is K-convex on Q. If $F(x) = \sum_{i=1}^{l} g_i(x) A_i$ with $A_i \in K$, $g_i : Q \to \mathbb{R}$ convex on Q, then F is K-convex on Q. When $K = S^m_+$ examples can be found in the literature, (in particular [3], Chapter 4) as X^2 defined on S^m or $X^t X$ defined for $X \in \mathbb{R}^{p \times q}$ or $-\sqrt{X}$ defined and K-convex on S^m_+.

2.2. Duality Results

Let $E := \{x \mid F(x) \preceq_K 0\}$, $C = Q \cap E$, and let $f : Q \to \mathbb{R}$ continuous on Q. In this paper we will consider the nonconvex conic problem

$$(P) \quad v_P = \min\{f(x) \mid x \in C\},$$

for which the standard problem P_{st} defined in the introduction is a particular case.

To (P) we associate the usual Lagrangian dual

$$(D) \quad v_D = \sup\{\theta(u) \mid u \succeq_{K_*} 0\},$$

where

$$L(x,u) = f(x) + \langle u, F(x)\rangle, \qquad \theta(u) = \min\{L(x,u) \mid x \in Q\}$$

are respectively the usual Lagrangian and the dual functional.

The aim of this subsection is to answer to the first question asked in the introduction. Let us denote by

- C(Q), the Banach space of real-valued continuous functions on Q, equipped with the max norm $||f|| := \max\{|f(x)| \mid x \in Q\}$.
- M(Q), its topological dual linear space, i.e., the Banach space of finite signed Borel measures on $(Q, \mathcal{B})$, where $\mathcal{B}$ is the Borel sigma-algebra of Q equipped with the norm given by the total variation of the corresponding measure. Thus the linear spaces $C(Q)$ and $M(Q)$ are in duality with the duality bracket

$$[h, \sigma] := \int_Q h(x) d\sigma(x), \quad \forall h \in C(Q)\ \forall \sigma \in M(Q).$$

- $M_+(Q)$, the positive cone of $M(Q)$, i.e., the set of non negative finite Borel measures σ on $(Q, \mathcal{B})$ with $||\sigma|| = \sigma(Q)$.
- $w^* := \sigma(M(Q), C(Q))$ the associated *weak** topology, i.e., the coarsest topology on $M(Q)$ for which $\sigma \to [h, \sigma]$ is continuous for each $h \in C(Q)$.
- $\mathcal{P}(Q)$, the set of probability measures on $(Q, \mathcal{B})$, i.e.

$$\mathcal{P}(Q) = \{\sigma \in M_+(Q) \mid \sigma(Q) = 1\}.$$

- δ_x, the Dirac measure concentrated at $x \in Q$.
- $\mathcal{F}(Q) = co\{\delta_x \mid x \in Q\}$, i.e., the set of finite convex combinations of Dirac measures concentrated on points $x \in Q$.
- $\mathcal{F}_{m+2}(Q)$, the set of finite convex combinations of at most $m+2$ Dirac measures concentrated on points $x \in Q$.
- $[F, \sigma] = ([f_1, \sigma], \ldots, [f_m, \sigma])$.

We recall that the separability of $C(Q)$ entails that the w^* topology is metrizable and that the closed unit ball in $M(Q)$, $B^* = \{\sigma \in M(Q) : \ ||\sigma|| \leq 1\}$ is *weakly** compact and *weakly** sequentially compact.

We recall also the fundamental result of Rogosinsky [15], Lemma 6.3 [8].

Lemma 2.1. *Set*

$$R_1 := \{\eta \in \mathrm{I\!R}^m \mid \exists \sigma \in \mathcal{P}(Q) : \ \eta = [F, \sigma]\}, \tag{5}$$

$$R_2 := \{\eta \in \mathrm{I\!R}^m \mid \exists \sigma \in \mathcal{F}(Q) : \ \eta = [F, \sigma]\}. \tag{6}$$

Then $R_1 = R_2$.

Now let us consider the following minimization relaxed convex problem

$$(PR) \quad v_{PR} = \inf\{[f, \sigma] \mid \sigma \in \mathcal{P}(Q) : [F, \sigma] \preceq_K 0\}. \tag{7}$$

Lemma 2.2. *The infimum for* (PR) *is attained for some* $\sigma \in \mathcal{F}_{m+2}(Q)$ *and* $v_{PR} \leq v_P$.

Proof. From the properties of the ball B^* it follows that $\mathcal{P}(Q)$ is a *weakly** compact convex set, so that the feasible set of (PR) is *weakly** compact and convex. Since the objective function is *weakly** continuous, then the infimum is attained for some $\sigma \in \mathcal{P}(Q)$. Using Lemma 2.1 this implies that the infimum is attained for some $\nu \in \mathcal{F}(Q)$. Using Carathéodory's Theorem it follows that $\nu \in \mathcal{F}_{m+2}(Q)$. Furthermore since δ_x with $x \in Q \cap E$ is feasible it follows that $v_{PR} \leq v_P$. □

Now let us consider the dual (DR) of (PR). For this we introduce the associated Lagrangian defined on $M(Q) \times \mathbb{R}^m$ by

$$L_R(\sigma, u) := [f(\cdot), \sigma] + \langle [F(\cdot), \sigma], u\rangle = [f(\cdot), \sigma] + [\langle F(\cdot), u\rangle, \sigma] = [L(\cdot, u), \sigma],$$

and then

$$(DR) \quad v_{DR} = \sup\{\theta_R(u) \mid u \succeq_{K_*} 0\}, \quad \text{with } \theta_R(u) = \inf_{\sigma \in \mathcal{P}(Q)} L_R(\sigma, u).$$

Theorem 2.3. *For each* $u \succeq_{K_*} 0$, $\theta_R(u) = \theta(u)$, $v_{DR} = v_D$.

Proof. Since $L(., u) \in C(Q)$, then $S(u) := \operatorname{argmin}_{x \in Q} L(x, u)$ is nonempty. Then for each $x \in S(u)$ we have $\theta(u) = L(x, u) \leq L(y, u)\ \forall y \in Q$. As a consequence we get

$$\theta(u) \leq [L(\cdot, u), \sigma] \quad \forall \sigma \in \mathcal{P}(Q),$$

so that

$$\theta(u) \leq \theta_R(u) \leq [L(\cdot, u), \delta_x] = \theta(u),$$

which obviously ends the proof. □

Lemma 2.4. *Let*

$$T := \{(\xi, \eta) \in \mathbb{R}^{m+1} \mid \exists \sigma \in \mathcal{P}(Q) : \ \xi \geq [f, \sigma],\ \eta \succeq_K [F, \sigma]\}. \tag{8}$$

Then T *is a closed convex set.*

Proof. We note first that the maps $\sigma \to [F, \sigma]$, $\sigma \to [f, \sigma]$ are linear. As a consequence the convexity of T follows from the fact that $\mathcal{P}(Q)$ is convex, and since $\mathcal{P}(Q)$ is *weakly** sequentially compact it follows obviously that T is closed. □

The following theorem which says that $v_{PR} = v_{DR}$ can perhaps be obtained by using some general convex duality theory for infinite dimensional locally convex linear spaces, but in fact with Lemma 2.4, we do not need such a theory. The proof is very simple with only arguments in finite dimensional spaces. Indeed it is enough to follow the classical proof given in finite dimensional spaces when $K = \mathbb{R}^m_+$ by replacing only the nonconvex closed set

$$Z = \{(\xi, \eta) \in \mathbb{R}^{m+1} \mid \exists x \in Q : \ \xi \geq f(x),\ \eta \succeq_K F(x)\}, \tag{9}$$

by the closed convex set T.

Theorem 2.5. $v_{PR} = v_{DR}$.

Proof. Obviously $v_{PR} \geq v_{DR}$, and we have only to prove that for any $\epsilon > 0$ there exists $u^\epsilon \succeq_K 0$ such that $v_{PR} \leq \theta(u^\epsilon) + \epsilon$. Thus let $\epsilon > 0$, and set $v_\epsilon := v_{PR} - \epsilon$. Since $(v_\epsilon, 0) \notin T$, it follows from Lemma 2.4 that there exists an hyperplane separating strongly $(v_\epsilon, 0)$ from T, i.e., there exist $(u_0^\epsilon, u^\epsilon) \in \mathbb{R}^{m+1}$, not equal to 0, and $\alpha \in \mathbb{R}$ such that

$$u_0^\epsilon v_\epsilon < \alpha \leq u_0^\epsilon \xi + \langle u^\epsilon, \eta \rangle \quad \forall (\xi, \eta) \in T. \tag{10}$$

Let us fix $(\xi, \eta) \in T$ associated with σ in formula (8), and let us prove that

$$(u_0^\epsilon, u^\epsilon) \in \mathbb{R}_+ \times K_*. \tag{11}$$

Suppose the contrary, then there exists some $v = (v_0, v_*) \in \mathbb{R}_+ \times K$ such that

$$v_0 u_0^\epsilon + \langle u^\epsilon, v_* \rangle < 0. \tag{12}$$

For each $\lambda \geq 0$, since $(\xi + \lambda v_0, \eta + \lambda v_*) \in T$, it follows from (10) that

$$u_0^\epsilon v_\epsilon < \alpha \leq u_0^\epsilon \xi + \langle u^\epsilon, \eta \rangle + \lambda [u_0^\epsilon v_0 + \langle u^\epsilon, v_* \rangle].$$

Then using (12), with $\lambda \to \infty$ we get a contradiction. Since for each $x \in Q$, $(f(x), F(x)) \in T$, then from (10) we obtain

$$u_0^\epsilon v_\epsilon < \alpha \leq {u_0}^\epsilon f(x) + \langle u^\epsilon, F(x) \rangle, \quad \forall x \in Q. \tag{13}$$

Let us prove now that $u_0^\epsilon \neq 0$. Indeed in the contrary case it follows from (13) that

$$0 < k\alpha \leq \langle k u^\epsilon, F(x) \rangle, \quad \forall x \in Q, \ \forall k > 0.$$

Thus if we set $u_k := k u^\epsilon$, then $L(x, u_k) \geq f(x) + k\alpha$ for all $x \in Q$, so that using Theorem 2.3 we get

$$v_{PR} \geq v_{DR} \geq \theta(u_k) \geq f(y) + k\alpha \quad \text{where } y = \operatorname*{argmin}_{x \in Q} f(x).$$

Passing to the limit with $k \to \infty$ we get a contradiction.

Now we may suppose without loss of generality that $u_0^\epsilon = 1$. Then using (13) we get

$$v_{PR} - \epsilon := v_\epsilon \leq \min_{x \in Q} L(x, u^\epsilon) = \theta(u^\epsilon),$$

and thus the theorem is proved. □

Let $\mathrm{co}(Z)$ be the convex hull of Z and let us consider the optimization problem

$$(P_c) \quad v_c = \inf\{\xi \mid (\xi, 0) \in \mathrm{co}(Z)\}. \tag{14}$$

A main result proven by Magnanti, Shapiro, Wagner ([12], Lemma 2.2) was that $v_D = v_c$ for the standard case. As we shall see now, this result can be obtained for the general case with something more, as a direct consequence of Theorem 2.5 and Lemma 2.2.

Corollary 2.6. $v_c = v_D = v_{PR}$. *Furthermore the infimum in* (14) *is attained.*

Proof. From Lemma 2.2, there exists $\nu_* \in \mathcal{F}_{m+2}(Q)$ an optimal solution of (PR), i.e., there exist

$$x_i \in Q,\ \lambda_i \geq 0,\ i = 1, \dots m+2 \ \text{ with } \sum_{i=1}^{m+2} \lambda_i = 1,$$

such that

$$\nu_* = \sum_{i=1}^{m+2} \lambda_i \delta_{x_i},$$

and

$$\xi^* := [f, \nu_*] = \sum_{i=1}^{m+2} \lambda_i f(x_i) = v_{PR}, \qquad \eta^* := [F, \nu_*] = \sum_{i=1}^{m+2} \lambda_i F(x_i) \preceq_K 0.$$

Thus $(\xi^*, 0) \in co(Z)$ so that $v_c \leq v_{PR}$. But since $\{(\xi, 0) \in co(Z)\} \subset \{(\xi, 0) \in T\}$ it follows that $v_{PR} \leq v_c$, so that $v_{PR} = v_c$, and the infimum in (14) is attained at ξ^*. $\square$

2.3. Bounds for the duality gap

We suppose now for the rest of the paper, that Q is convex. When $K = \mathbb{R}^m_+$ bounds for the duality gap were obtained by Aubin and Ekeland [2]. For this purpose the authors introduced the notion of the lack of convexity:

Definition 2.7. Let f be a real valued function defined and continuous on Q, the lack of convexity of f (on Q) is the number

$$\rho(f) = \sup \left\{ f\left(\sum_{i=1}^{q} \lambda_i x_i\right) - \sum_{i=1}^{q} \lambda_i f(x_i) \;\middle|\; q \in \mathcal{N},\ x_i \in Q,\ \lambda_i \geq 0, \right. \\ \left. i = 1, \dots q,\ \sum_{i=1}^{q} \lambda_i = 1 \right\}.$$

Obviously we have

$$0 \leq \rho(f) < +\infty, \qquad \rho(f) = 0 \Leftrightarrow f \text{ is convex on } Q, \qquad \rho(f+g) \leq \rho(f) + \rho(g).$$

Clearly such a notion is not appropriate for defining the lake of convexity of F related to a general partial ordering defined by K except if $K = \mathbb{R}^m_+$. Then we will restrict for the following our study to nonconvex conic problems where

$$K = \prod_{i=1}^{l} K_i, \qquad F(x) = (F_1(x), \dots, F_l(x)) \ \text{ with } F_i(x) \in K_i,\ i = 1, \dots, l.$$

Each K_i is supposed to be a pointed closed convex cone of $\mathbb{R}^{m_i}$ with nonempty interior and $\sum_{i=1}^{l} m_i = m$. Obviously $K_* = \prod_{i=1}^{l} (K_i)_*$ and

$$a = (a_1, \dots, a_l) \preceq_K b = (b_1, \dots, b_l) \ \text{ iff } a_i \preceq_{K_i} b_i \ \forall i = 1, \dots, l.$$

In order to define the lack of convexity of F related to Kwe will use the following assumption

$$(H) \quad \text{for each } i\ F_i \text{ is } K_i - \text{convex on Q when } K_i \text{ is not equal to } \mathbb{R}^{m_i}_+.$$

With this assumption we are leading to the following definition:

Definition 2.8. Let 0_{m_i} be the vector zero in $\mathbb{R}^{m_i}$. Set $\rho(F_i) = 0_{m_i}$ if K_i is not equal to $\mathbb{R}^{m_i}_+$ and $\rho(F_i) = (\rho(f^i_1), \dots, \rho(f^i_{m_i}))$ otherwise ($f^i_j,\ j = 1, \dots, m_i$ being the components of F_i). Then we define the lack of convexity of F (related to K) as $\rho(F) := (\rho(F_1), \dots, \rho(F_l))$

We are now able to give bounds for the duality gap:

Proposition 2.9. *Suppose that Assumption (H) holds. Then there exists $x^* \in Q$ such that*

$$f(x^*) \leq v_D + \rho(f), \qquad F(x^*) \preceq_K \rho(F). \tag{15}$$

Proof. From Corollary 2.6 there exist

$$x_i \in Q,\ \lambda_i \geq 0,\ i = 1, \dots m+2, \quad \text{with} \sum_{i=1}^{m+2} \lambda_i = 1,$$

such that

$$\sum_{i=1}^{m+2} \lambda_i f(x_i) = v_D, \qquad \sum_{i=1}^{m+2} \lambda_i F(x_i) \preceq_K 0. \tag{16}$$

Let $x^* := \sum_{i=1}^{m+2} \lambda_i x_i$. Then $x^* \in Q$ and from Definitions 2.7 and 2.8 and using componentwise Assumption (H) we get

$$\sum_{i=1}^{m+2} \lambda_i f(x_i) = f(x^*) + \left[\sum_{i=1}^{m+2} \lambda_i f(x_i) - f(x^*)\right] \geq f(x^*) - \rho(f),$$

and

$$\sum_{i=1}^{m+2} \lambda_i F(x_i) = F(x^*) + \left[\sum_{i=1}^{m+2} \lambda_i F(x_i) - F(x^*)\right] \succeq_K F(x^*) - \rho(F).$$

As a consequence, using relations (16), inequalities (15) follow. $\square$

Now in the following section we will show how such an x^* can be obtained by the proximal extension of the algorithm under the supplementary assumption that Slater's condition holds.

3. A proximal regularization version of the column generation method

For the rest of this paper we suppose that Slater's condition is satisfied, i.e.,

$$\exists \tilde{x} \in Q : \ F(\tilde{x}) \prec_K 0.$$

We present now our algorithm denoted by **PECGM** which is a proximal extension of the column generation method.

PECGM: Let $\gamma \geq 0$, and let $\{\epsilon_k\}$ be a sequence of non negative reals such that $\lim_{k\to\infty} \epsilon_k = 0$. Start with $x_1 = \tilde{x}$ and compute the sequence $\{x_k, y^k, u^k\}$ inductively as follows.

Suppose calculated $x_1, \ldots, x_k$, then

Step 1: Compute $\Lambda^k = (\lambda_1^k, \ldots, \lambda_k^k)$ optimal solution of the linear conic program

$$(P_k): \quad \alpha_k = \min\left\{\sum_{j=1}^{k} \lambda_j f(x_j) \mid \Lambda = (\lambda_1, \ldots, \lambda_k) \in T_k \cap S_k\right\}$$

with

$$T_k = \left\{\Lambda \mid \sum_{j=1}^{k} \lambda_j F(x_j) \preceq_K 0\right\}, \qquad S_k = \left\{\Lambda \mid \sum_{j=1}^{k} \lambda_j = 1,\ \lambda_j \geq 0,\ j = 1, \ldots, k\right\}.$$

Set

$$y^k = \sum_{j=1}^{k} \lambda_j^k x_j.$$

Let

$$L_k(\Lambda, u) = \sum_{j=1}^{k} \lambda_j L(x_j, u), \qquad \theta_k(u) = \min\{L_k(\Lambda, u) \mid \Lambda \in S_k\}.$$

Compute an optimal solution u^k of the dual (D_k) of (P_k):

$$(D_k): \quad \beta_k = \max\{\theta_k(u) \mid u \in K_*\}.$$

Step 2: Compute x_{k+1} by the relation

$$x_{k+1} \in Q, \quad L(x_{k+1}, u^k) + \frac{\gamma}{2}||x_{k+1} - y^k||^2 \leq L(x, u^k) + \frac{\gamma}{2}||x - y^k||^2 + \epsilon_k, \ \ \forall x \in Q. \quad (17)$$

Remark 3.1. The algorithm is well defined. Indeed since the feasible set of (P_k) is compact then Λ^k exists. Furthermore since x_1 satisfies Slater'condition then $\Lambda = (1, 0, \ldots, 0)$ satisfies Slater's condition for the linear conic program (P_k), and it follows from classical duality results that u^k exists and $\alpha_k = \beta_k$. Finally since Q is compact, $\operatorname{argmin}_{x\in Q} L(x, u^k) + \frac{\gamma}{2}||x - y^k||^2$ is nonempty and there exists at least a point x_{k+1} satisfying (17).

Remark 3.2. Obviously we have

$$\theta_k(u) = \min\{L(x_j, u) \mid j = 1, \ldots, k\} \geq \theta_{k+1}(u) \geq \theta(u). \quad (18)$$

Theorem 3.3. *i)* *The sequence $\{x_k, y^k, u^k\}$ is bounded.*

ii) *Suppose that* $\gamma = 0$*. Let* $\sigma_k := \sum_{i=1}^k \lambda_i^k \delta_{x_i}$ *where* δ_{x_i} *is the Dirac measure concentrated at* x_i*. Then there exists at least a weak* limit point of the sequence* $\{\sigma_k\}$ *and each weak* limit point of this sequence is an optimal solution of the relaxed problem* (PR)*, while the limit points of the dual sequence* $\{u^k\}$ *are optimal solutions of* (D)*. Furthermore if Assumption* (H) *holds, then each limit point* y_∞ *of the sequence* $\{y^k\}$ *satisfies*

$$y_\infty \in Q, \quad F(y_\infty) \preceq_K \rho(F), \quad f(y_\infty) \leq v_D + \rho(f). \tag{19}$$

If in addition f *is convex on* Q *and* F K*-convex on* Q*, then each limit point of the sequence* $\{y^k\}$ *is an optimal solution of* (P)*.*

iii) *Suppose that* $\gamma > 0$ *and that Assumption* (H) *holds. Then each limit point* y_∞ *of the sequence* $\{y^k\}$ *satisfies*

$$y_\infty \in Q, \quad F(y_\infty) \preceq_K \rho(F), \quad j = 1, \ldots, m, \tag{20}$$

and

$$f(y_\infty) \leq f(x) + \frac{\gamma}{2}||x - y_\infty||^2 + \rho(f), \quad \forall x \in C. \tag{21}$$

Furthermore if we set $\nu(C) := \max_{x,z\in C} ||x - z||^2$*, then when* F *is* K*-convex on* Q *we have*

$$y_\infty \in C, \quad f(y_\infty) \leq v_P + \frac{\gamma}{2}\nu(C) + \rho(f). \tag{22}$$

Finally, if f *is convex on* Q *and* F K*-convex on* Q*, each limit point of the sequence* $\{y^k\}$ *is an optimal solution of* (P)*.*

Proof. *i)* Since $\alpha_k = \beta_k$ it follows from (18) that

$$\theta(u^k) \leq \sum_{i=1}^k \lambda_i^k f(x_i) = \beta_k \leq L(x_j, u^k), \quad j = 1, \ldots, k. \tag{23}$$

Using the definition of the lake of convexity we get from (23)

$$f(y^k) \leq L(x_j, u^k) + \rho(f), \quad j = 1, \ldots, k. \tag{24}$$

Since $y^k \in Q$, it follows that the sequence $\{y^k\}$ is bounded and there exists a constant L such that $|f(y^k)| \leq L$, $\forall k$. Using relation (24) for $j = 1$ we get

$$\langle -F(x_1), u^k\rangle \leq f(x_1) + \rho(f) + L. \tag{25}$$

Let us prove now that the sequence $\{u^k\}$ is bounded. In the contrary case there would exists a subsequence such that

$$\lim_{k\to\infty, k\in\mathcal{K}} ||u^k|| = +\infty, \qquad \lim_{k\to\infty, k\in\mathcal{K}} \frac{u^k}{||u^k||} = u \succeq_{K_*} 0 \text{ with } ||u|| = 1. \tag{26}$$

Dividing both members of (25) by $||u^k||$ and taking the limit we get

$$\langle F(x_1), u\rangle \geq 0.$$

But since x_1 satisfies Slater condition, it follows from (26) and Proposition 8.29(a) [14] that $\langle F(x_1), u\rangle < 0$, a contradiction.

Finally the sequence $\{x_k\}$ is also bounded since Q is compact and the first statement is proved.

Let now y_∞ be a limit point of the sequence $\{y^k\}$ and let $\{y^k\}_{k\in\mathcal{K}}$, $\{u^k\}_{k\in\mathcal{K}}$, $\{x_{k+1}\}_{k\in\mathcal{K}}$ be subsequences converging to $y_\infty \in Q,\ u_\infty \succeq_{K_*} 0,\ x_\infty \in Q$. Let j be fixed in (24). Then passing to the limit in (24) we have

$$f(y_\infty) \le L(x_j, u_\infty) + \rho(f), \quad \forall j. \tag{27}$$

Passing to the limit in (27) and in (17) we get

$$f(y_\infty) \le L(x_\infty, u_\infty) + \rho(f), \qquad L(x_\infty, u_\infty) \le L(x, u_\infty) + \frac{\gamma}{2}||x - y_\infty||^2 \ \ \forall x \in Q, \tag{28}$$

so that

$$f(y_\infty) \le L(x, u_\infty) + \frac{\gamma}{2}||x - y_\infty||^2 + \rho(f), \quad \forall x \in Q, \tag{29}$$

and inequality (21) follows obviously.

ii) Suppose now that $\gamma = 0$. Let j be fixed in (23). Since θ is continuous, passing to the limit in (23) with $k \to \infty$ it follows that

$$\theta(u_\infty) \le \lim_{k\to\infty} \beta_k \le L(x_j, u_\infty), \quad \forall j.$$

Passing to the limit with $j = k+1,\ k \in \mathcal{K},\ k \to \infty$ it follows that $\theta(u_\infty) \le \lim_{k\to\infty} \beta_k \le L(x_\infty, u_\infty)$, and with the second inequality in (28) we obtain

$$\theta(u_\infty) \le L(x_\infty, u_\infty) = \theta(u_\infty) \le v_D, \quad u_\infty \in K_*,$$

so that

$$L(x_\infty, u_\infty) = \theta(u_\infty) = \lim_{k\to\infty} \beta_k, \quad u_\infty \in K_*. \tag{30}$$

Using the first inequality in (28) it follows that

$$f(y_\infty) \le v_D + \rho(f). \tag{31}$$

Now since $\sigma_k \in B^*$ and since B^* is weakly* sequentially compact, there exists at least a weak* limit point of the sequence $\{\sigma_k\}$. Let σ be an arbitrary weak* limit point of this sequence, then obviously $\sigma \in \mathcal{P}(Q)$. Furthermore (23) is equivalent to

$$\theta(u^k) \le [f, \sigma_k] = \beta_k \le L(x_j, u^k) \ \ \forall j = 1, \dots, k, \tag{32}$$

while by definition of the set T_k in the algorithm we have

$$[F, \sigma_k] = \sum_{j=1}^{k} \lambda_j^k F(x_j) \preceq_K 0. \tag{33}$$

Passing to the weak* limit in (33), it follows that σ is a feasible solution of (PR). Using (30) and passing to the weak* limit in (32) we get

$$[\sigma, f] = \lim_{k\to\infty} \beta_k = \theta(u_\infty), \quad u_\infty \in K_*,$$

which implies that σ is an optimal solution of (PR) and u_∞ and optimal solution of (D). The fact that each limit point of $\{u^k\}$ is an optimal dual solution follows obviously.

Suppose now that assumption (H) holds, then since $\sum_{j=1}^k \lambda_j^k F(x_j) \preceq_K 0$, it follows from Definitions 2.7 and 2.8 and Assumption (H) that

$$F(y^k) \preceq_K \rho(F).$$

Then passing to the limit it follows with (31) that (19) holds.

If we suppose now that f is convex on Q and that F is K-convex on Q, then from the above inequality it follows that $y_\infty \in C$ and then from (29) with $\gamma = 0$ it follows obviously that $y_\infty \in \operatorname{argmin}\{f(x) \mid x \in C\}$, which ends the proof of part *ii*).

Suppose now that $\gamma > 0$ and that Assumption (H) holds. Since $\sum_{j=1}^k \lambda_j^k F(x_j) \preceq_K 0$, it follows from Definition 2.7, Definition 2.8 and Assumption (H) that

$$F(y^k) \preceq_K \rho(F).$$

Then passing to the limit it follows that (20) holds.

Suppose now that F is K-convex on Q, then $y_\infty \in C$ and (22) follows obviously from (21). If in addition we suppose that f is convex on Q, it follows that

$$y_\infty \in \operatorname{argmin}\left\{f(x) + \frac{\gamma}{2}||x - y_\infty||^2 \mid x \in C\right\}.$$

Without loss of generality we may suppose that f is lower semicontinuous on the whole space, so that the above inclusion is equivalent to say that $0 \in \partial[f + i_C](y_\infty)$, the subdifferential at y_∞ of the sum of f and the indicator function of C. As a consequence y_∞ is an optimal solution of (P). □

4. Applications and Comments

First we will emphasize that **PECGM** concerns general conic optimization, that is not the case for **ECGM** [6], [12] which concerns only Standard Nonlinear Programming.

When $K = \mathbb{R}_+^m$ and when $f, f_i,\ i = 1, \dots, m$ are convex on Q, the algorithm coincides with the extension of the column generation method of Dantzig-Wolfe [6] given in [12], with $\gamma = 0$. However with $\gamma > 0$, the master sub-program described in *Step 2*, is a prox-regularization which consists of minimizing a strongly convex function on a compact convex set ("simple"). From a numerical point of view, this has well known advantages that are not shared for $\gamma = 0$, and this variant is new.

Let us now consider the nonconvex case. In order to use Theorem 3.3 we must estimate upper bounds of $\rho(f)$ and $\rho(F)$. Since for $K_i \neq \mathbb{R}_+^{m_i}$, $\rho(F_i) = 0$, we have to consider only bounds like $\rho(f)$.

Then let SC(Q) be the set of functions $g : Q \to \mathbb{R}$ continuous on Q such that there exists $\lambda(g) \geq 0$ for which the function $x \to h(x) := g(x) + \frac{\lambda(g)}{2}||x||^2$ is convex on Q. Such a class has been considered locally for theoretical results in nonlinear analysis (see for example [14] and references therein) and more recently for numerical methods (construction or/and convergence analysis [9], [5]). For such functions we have

$$\rho(g) \leq \rho(h) + \frac{\lambda(g)}{2}\rho(-||x||^2).$$

If we set

$$\Delta(Q) := \max_{x \in Q} ||x||^2 - \min_{x \in Q} ||x||^2,$$

then we get

$$\rho(g) \leq \frac{\lambda(g)}{2} \Delta(Q).$$

The class $C^2(Q)$ of twice continuously differentiable functions on Q belongs to SC(Q). Indeed for $f \in SC(Q)$, set $\tau^*(f) = \min\{\lambda_{min}(\nabla^2 f(x)) \mid x \in Q)\}$, where $\lambda_{min}(\nabla^2 f(x))$ is the minimum eigenvalue of the Hessian of f at x and let

$$\tau(f) = \max(0, -\tau^*(f)),$$

then $f \in SC(Q)$ with

$$\lambda(f) = \tau(f).$$

More generally the set of functions $g = h + f$ where h is convex on Q and $f \in C^2(Q)$ belongs to SC(Q) with

$$\rho(g) \leq \frac{\tau(f)}{2} \Delta(Q).$$

In particular if f is quadratic i.e., $f(x) = \frac{\langle Ax+b,x \rangle}{2} + c$, with A symmetric and $\lambda_{min}(A) < 0$ then we get

$$\rho(f+h) \leq \frac{|\lambda_{min}(A)|}{2} \Delta(Q).$$

Now let us give examples where we can see that **PECGM** can be effectively implementable, a point which was not addressed in [12].

It is worthwhile to note first that the computation of (Λ^k, u^k) in *Step 1* as solutions of linear conic optimization problems can be done by efficient algorithms. If k and m are not too large, this can be done in polynomial time by interior point methods and we refer in particular to the book of Ben-Tal and Nemirovski [3]. Otherwise this can be done by variants of the accelerated gradient method of Nesterov that are particularly efficient in this case. For more on this subject see [11].

In fact the main difficulty of the algorithm, is to compute x_{k+1} at *Step 2* of iteration k. Indeed the Lagrangian function is nonconvex in x on Q in general. However an ϵ_k optimal solution x_{k+1} can be computed with a finite number of calculations in the three following interesting cases.

Case 1. $Q := \{x : \langle Mx, x \rangle \leq r\}$ with M a symmetric positive definite matrix, $K = \mathbb{R}^m_+$ and all the functions f, f_i are quadratic $(f_i(x) = \langle A_i x + b_i, x \rangle + c_i,\ f(x) = \langle Ax+b, x\rangle + c)$. In this case the objective function in *Step 2* is quadratic, and the quadratic problem can be solved by the trust region algorithm (see for example Sorensen [16]). In fact in this case Ben-Tal and Teboulle [4] have shown that the subproblem is equivalent to a convex problem which can be solved also efficiently in a different way.

In addition with $\gamma = 0$, if set $\nu = \frac{r}{\lambda_{min}(M)}$, we get from Theorem 3.3 the bounds

$$y_\infty \in Q, \quad f_j(y_\infty) \leq |\min(\lambda_{min}(A_j), 0)|\nu,\ j = 1, \ldots, m,$$
$$f(y_\infty) \leq v_D + |\min(\lambda_{min}(A), 0)|\nu.$$

Case 2. F is K-convex on Q, and $f \in SC(Q)$. Then for $\gamma = \lambda(f)$, the objective function $x \to L(x, u^k) + \frac{\gamma}{2}||x - y^k||^2$ is convex and *Step 2* is implementable. In this case we have

$$y_\infty \in Q \cap E, \quad f(y_\infty) \leq v_P + \frac{\lambda(f)}{2}(\nu(C) + \Delta(Q)).$$

Case 3: Separable data. We can use now the fact that the algorithm is a decomposition method.

Let $K = \mathbb{R}^m_+$, set $\gamma = 0$, let $x = (v, w) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, and suppose that

$$f(x) = f_0(x), \quad f_j(x) = g_j(v) + h_j(w), \quad j = 0, 1, \ldots, m, \quad Q = Q_1 \times Q_2,$$

where $Q_i,\ i = 1, 2$ is compact convex and where for each $j,\ g_j : Q_1 \to \mathbb{R}$ is convex continuous on Q_1 while $h_j : Q_2 \to \mathbb{R}$ is continuous on Q_2. Suppose furthermore that for each $u \in \mathbb{R}^m_+$ we have an implementable algorithm for solving the nonconvex minimization problem

$$R(u): \quad \alpha(u) = \min_{w \in Q_2} \left(h_0 + \sum_{i=1}^m u_i h_i \right)(w).$$

Then *Step 2* is implementable and we have

$$x^{k+1} = (v^{k+1}, w^{k+1}), \quad \text{with } v^{k+1} \in \epsilon_k - \operatorname*{argmin}_{v \in Q_1} \left(g_0 + \sum_{i=1}^m u_i^k g_i \right)(v),$$
$$w^{k+1} \in \epsilon_k - \operatorname*{argmin}_{w \in Q_2} \left(h_0 + \sum_{i=1}^m u_i^k h_i \right)(w).$$

Let us give now two examples for which we are in such a situation.

Example 4.1. The functions h_j are polynomial and Q_2 is a ball or a box. In this case we can use, when n_2 is small the polynomial time algorithms proposed in the field of global polynomial optimization introduced by Lasserre (see for example [10] and references therein).

Example 4.2. Suppose that $Q_2 = \{w : Aw \leq b\}$ where A is some matrix (p, n_2), and that $h_j,\ j = 0, 1, \ldots, m$ are concave. Then $R(u)$ consists of minimizing a concave function on a polyhedral set. Actually, this is implementable by Tuy's type methods [17], and also by the Falk and Hoffman's method [17] but obviously in any cases only when n_2 is very small ($n_2 \leq 10$).

Acknowledgements. I would like to express my gratitude to an anonymous referee for his careful reading the manuscript and for his many useful comments.

References

[1] B. Alkire, L. Vandenberghe: Convex optimization problems involving autocorrelation sequences, Math. Program., Ser. A 93 (2002) 331–359.

[2] J. P. Aubin, I. Ekeland: Estimates of the duality gap in nonconvex optimization, Math. Oper. Res. 1(3) (1976) 225–245.

[3] A. Ben-Tal, A. Nemirovski: Lectures on Modern Convex Optimization. Analysis, Algorithms, and Engineering Applications, MPS/SIAM Series on Optimization 2, SIAM, Philadelphia (2001).

[4] A. Ben-Tal, M. Teboulle: Hidden convexity in some quadratically constrained quadratic programming, Math Program., Ser. A 72(1) (1996) 51–63.

[5] J. Bolte, A. Daniilidis, O. Ley, L. Mazet: Characterizations of Lojasiewicz inequalities and applications (2008), to appear.

[6] G. B. Dantzig: Linear Programming and Extensions, Princeton University Press, Princeton (1963).

[7] J. E. Falk, K. R. Hoffman: A successive underestimation method for concave minimization, Math. Oper. Res. 1(3) (1976) 251–259.

[8] R. Hettich, K. O. Kortanek: Semi-infinite programming: theory, methods, and applications, SIAM Rev. 35(3) (1993) 380–429.

[9] W. Hare, C. Sagastizabal: Computing points of nonconvex functions, Math. Program., Ser. B 116 (2009) 221–258.

[10] J. B. Lasserre: Global optimization with polynomials and the problem of moments, SIAM J. Optim. 11 (2001) 796–817.

[11] G. Lan, Z. Lu, R. D. C. Monteiro: Primal-dual first order methods with $0(\frac{1}{\epsilon})$ iterations for conic progamming, Technical Report, School of Industrial and Systems Enginnering, Georgia Institute of Technology, Atlanta (2006).

[12] T. L. Magnanti, J. F. Shapiro, M. H. Wagner: Generalized linear programming solves the dual, Manage. Sci. 22(11) (1976) 1195–1203.

[13] Yu. Nesterov: Squared functional systems and optimization problems, in: High Performance Optimization, H. Frenk et al. (ed.), Appl. Optim 33, Kluwer, Dordrecht (2000) 405–440.

[14] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Grundlehren der Mathematischen Wissenschaften 317, Springer, Berlin (1998)

[15] W. W. Rogosinsky: Moments of non-negative mass, Proc. R. Soc. Lond., Ser. A 245 (1958) 1–27.

[16] D. C. Sorensen: Minimization of a large-scale quadratic function subject to a spherical constraint, SIAM J. Optim. 7(1) (1997) 141–161.

[17] H. Tuy: Concave programming under linear constraints, Sov. Math., Dokl. 5 (1964) 1437–1440.