A CONIC DUALITY FRANK–WOLFE TYPE THEOREM VIA EXACT PENALIZATION IN QUADRATIC OPTIMIZATION

WERNER SCHACHINGER† AND IMMANUEL M. BOMZE†

Abstract. The famous Frank–Wolfe theorem ensures attainability of the optimal value for quadratic objective functions over a (possibly unbounded) polyhedron if the feasible values are bounded. This theorem does not hold in general for conic programs where linear constraints are replaced by more general convex constraints like positive-semidefiniteness or copositivity conditions, despite the fact that the objective can be even linear. This paper studies exact penalizations of (classical) quadratic programs, i.e. optimization of quadratic functions over a polyhedron, and applies the results to establish a Frank–Wolfe type theorem for the primal-dual pair of a class of conic programs which frequently arises in applications. One result is that uniqueness of the solution of the primal ensures dual attainability, i.e., existence of the solution of the dual.

1. Introduction

The famous Frank–Wolfe theorem ensures attainability of the optimal value for quadratic objective functions over a (possibly unbounded) polyhedron if the feasible values are bounded. This theorem does not hold in general for conic programs where linear constraints are replaced by more general convex constraints like positive-semidefiniteness or copositivity conditions, despite the fact that the objective can be even linear. In some cases, even if the duality gap is zero and the primal problem has a compact feasible set, the primally optimal solution value need not be attained in the dual problem; see, e.g., Boyd and Vandenberghe [3]. Among the more recent extensions of the Frank-Wolfe theorem are Belousov and Klatte [1], Luo and Zhang [9], Ozdaglar and Tseng [11], and Pataki [12]. However, none of the results there seem to be directly applicable to our problem.

We start with a section motivating the study of a class of conic programs, showing that the phenomenon of non-attainability in such programs is not a purely academic construction but rather may emerge quite naturally in problems arising from applications. Section 3 then constitutes the main part of this note. Here we develop a characterization of dual attainability in terms of a zero-value optimality condition on a quadratic program parameterized by the dual variable, and establish simple conditions guaranteeing this property with the help of a seemingly new exact penalization result which may be of general interest. One consequence for the primal-dual pair of the above-mentioned conic program is that uniqueness of the solution of the primal ensures dual attainability, i.e., existence of the solution

Date: This version: February 28, 2007.
1991 *Mathematics Subject Classification.* Primary: 90C20; Secondary: 90C22, 90C30, 90C46.
Key words and phrases. Copositive Programming, Dual Attainability.
† University of Vienna, Austria.

of the dual. Dual attainability becomes important if (primal-)dual algorithms for conic programs are employed, and also to settle tightness questions of dual bounds.

2. Multi-Standard Quadratic Problems

Multi-Standard Quadratic Problems arise in diverse fields of applications, like relaxation labelling processes for speech and pattern recognition, see Hummel and Zucker [7] and Pelillo [13], and machine learning (support vector machines for classification). Recently two monotonely improving interior point methods have been proposed and also a conic reformulation has been established to obtain rigid yet relatively cheap bounds via semidefinite programming, see Bomze et al. [2] and Burer [4].

Formally speaking, given an arbitrary symmetric $M \times M$ matrix Q, we consider the following problem as a *multi-Standard Quadratic Problem (m-StQP) in maximization form*:

$$\max \left\{ z^\top Q z : z \in \Lambda \right\} , \tag{1}$$

where for $m \geq 2$ and $1 \leq i \leq m$ the sets $\Delta_i = \left\{ x \in \mathbb{R}^{n_i}_+ : (e_i)^\top x = 1 \right\}$ are standard simplices in $\mathbb{R}^{n_i}$ with $e_i = [1, \ldots, 1]^\top \in \mathbb{R}^{n_i}$, and

$$\Lambda = \bigotimes_{i=1}^{m} \Delta_i ,$$

which is a polytope in $\mathbb{R}^M$ with $M = \sum_{i=1}^m n_i$.

The conic reformulation of (1) from Bomze et al. [2] uses the convex, non-polyhedral cone of **completely positive** $M \times M$ matrices

$$\mathcal{C} = \left\{ X : X = FF^\top \text{ for some } M \times k \text{ matrix } F \text{ with no negative entries } \right\} .$$

It is well known that with respect to the Frobenius duality $X \bullet Y = \text{trace } (XY)$ of symmetric $M \times M$ matrices, this cone is the dual cone of the cone of **copositive** $M \times M$ matrices

$$\mathcal{C}^* = \left\{ Y = Y^\top : x^\top Y x \geq 0 \text{ for all } x \in \mathbb{R}^M_+ \right\} .$$

Consider the $m \times M$ matrix

$$H = \begin{bmatrix} e_1^\top & o^\top & \cdots & o^\top \\ o^\top & e_2^\top & \cdots & o^\top \\ \vdots & & \ddots & \vdots \\ o^\top & o^\top & \cdots & e_m^\top \end{bmatrix} . \tag{2}$$

Then the (possibly non-convex) m-StQP (1) is equivalent to a linear problem over the completely positive cone:

$$\begin{aligned} &\max \left\{ z^\top Q z : z \in \Lambda \right\} \\ &\quad = \max \left\{ Q \bullet Z : HZH^\top = E ,\ Z \in \mathcal{C} \right\} , \end{aligned} \tag{3}$$

where $E = ee^\top$ with $e = [1, \ldots, 1]^\top \in \mathbb{R}^m$. To be more precise, we have

$$\begin{aligned} &\text{argmax } \left\{ Q \bullet Z : HZH^\top = E ,\ Z \in \mathcal{C} \right\} \\ &\quad = \text{conv}\left\{ zz^\top : z \in \text{argmax } \left\{ z^\top Q z : z \in \Lambda \right\} \right\} , \end{aligned} \tag{4}$$

where for a set $\mathcal{S}$ we denote by $\operatorname{conv}\mathcal{S}$ the convex hull of $\mathcal{S}$. Thus the set of optimal solutions to (3) is compact, and Theorem 30.4 of Rockafellar [14] guarantees that the optimal value of (3) equals

$$\inf\left\{\sum_{i,j} X_{ij} = e^\top Xe : X = X^\top, H^\top XH - Q \in \mathcal{C}^*\right\}, \tag{5}$$

which represents a linear problem over the copositive cone.

Strict feasibility, i.e., Slater's condition is not satisfied for the primal problem (3). To be more precise, we have the following properties.

Proposition 2.1. *For the primal-dual pair of problems (3) and (5)*

(a) Slater's condition for (3) is always violated;
(b) if Q has a representation $Q = H^\top \bar{Q} H$ for some symmetric $m\times m$ matrix $\bar{Q}$, then the set of minimizers of (5) is nonempty;
(c) if the set of minimizers of (5) is nonempty, it is always unbounded;
(d) however, the minimum of (5) need not be attained in general.

Proof. To show assertion (a), assume by contradiction that there is $Z \in \operatorname{int}\mathcal{C}$ satisfying $HZH^\top = E$. Then for some $\varepsilon > 0$ and the $m \times m$-matrix Y with $Y_{i,j} = 1$ if $\{i,j\} = \{1,2\}$, and $Y_{i,j} = 0$, else, we still have $Z' := Z + \varepsilon H^\top YH \in \mathcal{C}$, and therefore also $HZ'H^\top = E + \varepsilon n_1 n_2 Y$ is completely positive. The latter is now contradicted by looking at the 2×2 principal minor of $H^\top Z'H$, which has determinant $\det\begin{bmatrix} 1 & 1+\varepsilon n_1 n_2 \\ 1+\varepsilon n_1 n_2 & 1\end{bmatrix} < 0$.
(b) Take an arbitrary $z \in \Lambda \subset \mathbb{R}^M_+$. Then $Hz = e$, and $Q = H^\top \bar{Q} H$ implies

$$0 \le z^\top(H^\top XH - Q)z = z^\top H^\top (X - \bar{Q})Hz = e^\top (X-\bar{Q})e = e^\top Xe - e^\top \bar{Q} e$$

for any (5)-feasible X, which establishes that the infimum in (5) is not smaller than $e^\top \bar{Q} e$. On the other hand, $X = \bar{Q}$ obviously is (5)-feasible with exactly this objective value, and hence is a minimizer of (5). In fact,

$$H^\top CH \text{ is copositive if and only if } C \text{ is copositive.} \tag{6}$$

(c) If a matrix X is a minimizer of (5) then so is any matrix $X + t(mI - E)$ with $t > 0$, hence the set of minimizers is unbounded.
(d) We now give an example showing that the set of minimizers of (5) can be empty: Let $m = 2$, $X = \begin{bmatrix} a & b \\ b & c\end{bmatrix}$, $n_1 = 2$, $n_2 = 1$, $M = 3$ and $Q = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 0\end{bmatrix}$.
Denote by $M_{a,b,c} = H^\top XH - Q = \begin{bmatrix} a+2 & a & b-1 \\ a & a-2 & b+1 \\ b-1 & b+1 & c\end{bmatrix}$. Now we claim that

$$\mu := \inf\{a + 2b + c : M_{a,b,c} \in \mathcal{C}^*\}$$

is not attained. Let $z = [1,1,2]^\top$ and observe $z^\top M_{a,b,c} z = 4(a + 2b + c)$, thus $M_{a,b,c} \in \mathcal{C}^*$ implies $a + 2b + c \ge 0$ and therefore $\mu \ge 0$. Next we show that $M_{a,b,c} \in \mathcal{C}^*$ even implies $a + 2b + c > 0$. First note that any $M_{a,b,c} \in \mathcal{C}^*$ satisfies $a - 2 \ge 0$, $c \ge 0$, and $b - 1 \ge -\sqrt{(a+2)c}$, according to Hadeler [6]. Assuming that $a + 2b + c = 0$ holds, we have the inequality

$$0 = a + 2b + c \ge a + 2 - 2\sqrt{(a+2)c} + c = \left(\sqrt{a+2} - \sqrt{c}\right)^2 \ge 0\,,$$

therefore $c = a + 2$ and $b = -a - 1$. Now we derive the contradiction

$$M_a := \begin{bmatrix} a+2 & a & -a-2 \\ a & a-2 & -a \\ -a-2 & -a & a+2 \end{bmatrix} \notin \mathcal{C}^*, \text{ for } a \geq 2\,,$$

just taking $z = [0, a+1, a-1]^\top \in \mathbb{R}^3_+$ and computing $z^\top M_a z = -4a < 0$. Next we show that $A := aM_a + [1,1,1][1,1,1]^\top \in \mathcal{C}^*$, for $a \in \mathbb{R}$:

$$\begin{aligned} z^\top A z &= a^2(z_1+z_2-z_3)^2 + 2a(z_1+z_2-z_3)(z_1-z_2-z_3) + (z_1-z_2-z_3)^2 \\ &\quad - (z_1-z_2-z_3)^2 + (z_1+z_2+z_3)^2 \\ &= (az_1+az_2-az_3+z_1-z_2-z_3)^2 + 4z_1(z_2+z_3) \\ &\geq 0\,, \end{aligned}$$

for $z = [z_1, z_2, z_3]^\top \in \mathbb{R}^3_+$. Since $\frac{1}{a}A = M_{a+\frac{1}{a}, -a-1+\frac{1}{a}, a+2+\frac{1}{a}}$ we obtain $\mu \leq \frac{4}{a}$ for any $a > 0$, i.e., $\mu = 0$. □

The copositive programming formulation above differs from the representation recently introduced by Burer [4] in a more general set-up: the copositive and the completely positive matrices there have an additional row and column, compared to the ones in the primal-dual pair (3),(5). In the case of Multi-Standard Quadratic Problems, Burer's representation would read

$$\begin{aligned} &\max\left\{z^\top Q z : z \in \Lambda\right\} \\ &\quad = \max\left\{Q \bullet Z : \operatorname{diag}(HZH^\top) = e\,,\ Hz = e\,,\ \zeta = 1\,,\ \begin{bmatrix} \zeta & z^\top \\ z & Z \end{bmatrix} \in \widehat{\mathcal{C}}\right\}\,, \end{aligned}$$

with $\widehat{\mathcal{C}}$ denoting the cone of completely positive $(M+1)\times(M+1)$ matrices. Here we denote by $\operatorname{diag}(\cdot)$ the vector which is the diagonal of a given matrix, and by $\operatorname{Diag}(\cdot)$ the diagonal matrix whose diagonal is a given vector. By standard duality arguments we then obtain the following linear problem over the copositive cone:

$$\inf\left\{a + (2u^\top + v^\top)e : \begin{bmatrix} a & u^\top H \\ H^\top u & H^\top \operatorname{Diag}(v) H - Q \end{bmatrix} \in \widehat{\mathcal{C}^*}\right\}\,. \tag{7}$$

We can again use Theorem 30.4 of Rockafellar [14] to deduce that the duality gap is zero. However, also with this formulation, the same dual attainability problem occurs:

Proposition 2.2. *For the dual (7) of Burer's representation,*

(a) the minimum in (5) is attained if and only if the minimum in (7) is attained,
(b) if the set of minimizers of (7) is nonempty, it is unbounded.

Proof. In the following, a hat always shall designate a similar vector/matrix with an additional (first) row and/or column, which get the index 0 for notational convenience. So consider for instance $\widehat{E}$, the $(m+1)\times(m+1)$ all-ones matrix. Denote the infima in question by

$$\mu = \inf\left\{E \bullet X : X = X^\top, H^\top X H - Q \in \mathcal{C}^*\right\} \tag{8}$$

and by

$$\mu_B = \inf\left\{\widehat{E} \bullet \widehat{X} : \widehat{X} = \widehat{X}^\top, \widehat{X}_{i,j} = 0 \text{ for } 1 \leq i < j,\ \widehat{H}^\top \widehat{X} \widehat{H} - \widehat{Q} \in \widehat{\mathcal{C}^*}\right\}\,, \tag{9}$$

where $\widehat{H} = \begin{bmatrix} 1 & o^\top \\ o & H \end{bmatrix}$ and $\widehat{Q} = \begin{bmatrix} 0 & o^\top \\ o & Q \end{bmatrix}$. As both formulations have zero duality gap, we have $\mu = \mu_B$.

(a) If for some Q the minimum is attained in (9), then there is an arrowhead matrix $\widehat{X} = \begin{bmatrix} a & u^\top \\ u & \mathrm{Diag}\,(v) \end{bmatrix}$ such that $\widehat{E} \bullet \widehat{X} = \mu$ and $\widehat{Y} := \widehat{H}^\top \widehat{X} \widehat{H} - \widehat{Q} \in \widehat{\mathcal{C}}^*$. Then also $\widehat{L}\widehat{Y}\widehat{L}^\top \in \widehat{\mathcal{C}}^*$, if $\widehat{L} = I_{M+1} + \widehat{f}\,\widehat{e}_0^\top$, where I_k denotes the $k \times k$ identity matrix and $\widehat{f} = [0, (e_1)^\top, o^\top]^\top \in \mathbb{R}^{M+1}$ (remember $e_1 = [1, \dots, 1]^\top \in \mathbb{R}^{n_1}$) while $\widehat{e}_0 = [1, 0, \dots, 0]^\top \in \mathbb{R}^{M+1}$. Since $\widehat{Q}\widehat{e}_0 = o$ by construction, $\widehat{L}\widehat{Q}\widehat{L}^\top = \widehat{Q}$ results. Moreover, put $\widehat{R} = I_{m+1} + \widehat{f}_1 \widehat{f}_0^\top$ with $\widehat{f}_r \in \mathbb{R}^{m+1}$ having zero coordinates except the r-th which equals one. Then we have $\widehat{H}^\top \widehat{f}_1 = [0, (H^\top f_1)^\top]^\top = \widehat{f}$ and also $\widehat{e}_0^\top \widehat{H}^\top = (\widehat{H}\widehat{e}_0)^\top = \widehat{f}_0^\top$, so that we arrive at $\widehat{L}\widehat{H}^\top = \widehat{H}^\top + \widehat{f}\widehat{f}_0^\top = \widehat{H}^\top \widehat{R}$. Thus

$$\widehat{L}\widehat{Y}\widehat{L}^\top = \widehat{H}^\top \widehat{R}\widehat{X}\widehat{R}^\top \widehat{H} - \widehat{Q} \in \widehat{\mathcal{C}}^* .$$

By dropping the first row and column of the latter matrix, we arrive at a completely positive $M \times M$ matrix of the form $H^\top R\widehat{X}R^\top H - Q \in \mathcal{C}^*$, where $\widehat{R} = [\widehat{f}_0^\top, R^\top]^\top$ so that $R = [f_1, I_m]$ is an $m \times (m+1)$ matrix with $ER = [Ef_1, E] = [e, E]$ and thus $R^\top E R = \widehat{E}$. This shows that $X := R\widehat{X}R^\top$ is feasible for (8) with objective value

$$E \bullet X = E \bullet (R\widehat{X}R^\top) = (R^\top E R) \bullet \widehat{X} = \widehat{E} \bullet \widehat{X} = \mu .$$

On the other hand, suppose for some Q the minimum is attained in (8), then there is $X = X^\top$ such that $E \bullet X = \mu$ and $H^\top X H - Q \in \mathcal{C}^*$. For $1 \le i < j$, let $\widehat{C}_{i,j}$ and $\widehat{D}_{i,j}$ be copositive $(m+1) \times (m+1)$-matrices of rank 1 defined by $\widehat{C}_{i,j} = (\widehat{f}_i - \widehat{f}_j)(\widehat{f}_i - \widehat{f}_j)^\top$ and $\widehat{D}_{i,j} = (2\widehat{f}_0 - \widehat{f}_i - \widehat{f}_j)(2\widehat{f}_0 - \widehat{f}_i - \widehat{f}_j)^\top$. These matrices satisfy $\widehat{E} \bullet \widehat{C}_{i,j} = \widehat{E} \bullet \widehat{D}_{i,j} = 0$. Now define

$$\widehat{X} := \begin{bmatrix} 0 & o^\top \\ o & X \end{bmatrix} + \sum_{1 \le i < j : \hat{X}_{i,j} > 0} X_{i,j} C_{i,j} - \sum_{1 \le i < j : \hat{X}_{i,j} < 0} X_{i,j} D_{i,j} \succeq \begin{bmatrix} 0 & o^\top \\ o & X \end{bmatrix},$$

so that

$$\widehat{H}^\top \widehat{X} \widehat{H} - \widehat{Q} \succeq \begin{bmatrix} 0 & o^\top \\ o & H^\top X H - Q \end{bmatrix} \in \widehat{\mathcal{C}}^* ,$$

where for symmetric matrices M_1, M_2 we put $M_1 \succeq M_2$ when $M_1 - M_2$ is copositive. Moreover, $\widehat{X}$ satisfies $(\widehat{X})_{i,j} = 0$ for $1 \le i < j$ by construction, so $\widehat{X}$ is feasible for (9), furthermore $\widehat{E} \bullet \widehat{X} = E \bullet X = \mu = \mu_B$, so the minimum in (9) is also attained for that Q.

(b) If a matrix $\widehat{X}$ is a minimizer of (9) then so is any matrix $\widehat{X} + t(\widehat{f}_0 - \widehat{f}_1)(\widehat{f}_0 - \widehat{f}_1)^\top$ with $t > 0$, hence the set of minimizers is unbounded. □

To summarize, we here face a situation in-between weak and strong duality: the duality gap is zero but the dual optimal value need not be attained. However, there are some conditions which guarantee also dual attainability, one is that the primal solution is unique (note that by (4), problem (3) has a unique solution if and only if problem (1) has a unique solution). To the best of our knowledge, this situation has not yet been analyzed in the literature, although there is abundant work on strong duality at all levels of sophistication and generality.

3. Exact penalization and dual attainability

We start with an observation relating dual attainability with the zero solution value property of a related quadratic optimization problem: for a fixed symmetric $m \times m$ matrix X, consider

$$\min\left\{z^\top\left(H^\top XH - Q\right)z : z \in \Delta\right\}, \tag{10}$$

where $\Delta = \left\{z \in \mathbb{R}^M_+ : \sum_i z_i = 1\right\}$ is the standard simplex in $\mathbb{R}^M$.

Lemma 3.1. *Let $\overline{z}$ be a solution to (1) and put $z^* = \frac{1}{m}\overline{z}$. Then X is a minimizer of (5) if and only if z^* is a minimizer of (10) attaining the minimum* 0.

Proof. ($\Rightarrow$) Let $Y = mI_m - E$. Then $z \in \Delta$ implies

$$\begin{aligned}\|Hz - \frac{1}{m}e\|^2 &= z^\top H^\top Hz - \frac{2}{m}e^\top Hz + \frac{1}{m} = z^\top H^\top Hz - \frac{1}{m}\\ &= z^\top H^\top Hz - \frac{1}{m}z^\top H^\top EHz = \frac{1}{m}z^\top H^\top YHz\,.\end{aligned} \tag{11}$$

Assume that X solves (5), so that $H^\top XH - Q \in \mathcal{C}^*$. Since H has no negative entries and Y is copositive, we also get $H^\top X'H - Q = \mu H^\top YH + H^\top XH - Q \in \mathcal{C}^*$ if $X' := X + \mu Y$ with $\mu > 0$. Hence

$$\min\left\{z^\top\left(H^\top X'H - Q\right)z : z \in \Delta\right\} \geq 0\,.$$

Suppose that $\min\left\{z^\top\left(H^\top X'H - Q\right)z : z \in \Delta\right\} > 0$.

Then for some $\varepsilon > 0$ and $X'' := X' - \varepsilon E$ we would have $H^\top X''H - Q \in \mathcal{C}^*$ and $e^\top X''e = e^\top Xe - \varepsilon m^2 < e^\top Xe$, contradicting minimality of X in (5). Hence we arrive at $\hat{z}^\top\left(H^\top X'H - Q\right)\hat{z} = 0$ for some $\hat{z} \in \Delta$. This implies of course both $\hat{z}^\top\left(H^\top XH - Q\right)\hat{z} = 0$ and, by (11), $\|H\hat{z} - \frac{1}{m}e\|^2 = \frac{1}{m}\hat{z}^\top H^\top YH\hat{z} = 0$. So $\hat{z}$ solves (10) with zero value, and we show that the same holds true for z^*. Indeed, both points belong to Δ and satisfy $Hz^* = H\hat{z} = \frac{1}{m}e$, as argued above, so that

$$\begin{aligned}(\hat{z})^\top(H^\top XH - Q)\hat{z} &= \tfrac{1}{m^2}\,e^\top Xe - (\hat{z})^\top Q\hat{z}\\ &\geq \tfrac{1}{m^2}\,e^\top Xe - (z^*)^\top Qz^* = (z^*)^\top(H^\top XH - Q)z^*\,,\end{aligned}$$

since $\overline{z} = mz^*$ solves (1), and $m\hat{z} \in \Lambda$. On the other hand,

$$\hat{z}^\top(H^\top XH - Q)\hat{z} = 0 \leq (z^*)^\top(H^\top XH - Q)z^*\,,$$

since $H^\top XH - Q \in \mathcal{C}^*$. Hence the result.
($\Leftarrow$) We start with X such that $H^\top XH - Q \in \mathcal{C}^*$, and z^* such that $Hz^* = \frac{1}{m}e$ and $(z^*)^\top\left(H^\top XH - Q\right)z^* = 0$. For any X' such that $H^\top X'H - Q \in \mathcal{C}^*$ we then have

$$\begin{aligned}0 &\leq (z^*)^\top\left(H^\top X'H - Q\right)z^*\\ &= (z^*)^\top\left(H^\top XH - Q\right)z^* + (z^*)^\top H^\top(X' - X)Hz^*\\ &= \tfrac{1}{m^2}e^\top(X' - X)e,\end{aligned}$$

i.e., $e^\top Xe \leq e^\top X'e$, which says that X minimizes (5). □

Hence we will construct an X yielding a zero value solution of (10), to obtain a solution to (5). To this end we need several auxiliary results, some of which may be of more general interest. We start with a result relating KKT points of a scaled version of (1) to a relaxation of (10).

Lemma 3.2. *If z^* is a KKT point of*

$$\max\left\{z^\top Qz : z\in\mathbb{R}^M_+ : Hz=\tfrac{1}{m}e\right\} = \max\left\{z^\top Qz : z\in\tfrac{1}{m}\Lambda\right\}\,,$$

then there is a symmetric $m\times m$ matrix X' such that z^* is also a KKT point of

$$\min\left\{z^\top\left(H^\top X'H-Q\right)z : z\in\mathbb{R}^M_+\right\}, \tag{12}$$

with objective value 0*. This means*

$$(z-z^*)^\top(H^\top X'H-Q)z^*\geq 0=(z^*)^\top(H^\top X'H-Q)z^* \quad \text{for all } z\in\mathbb{R}^M_+\,. \tag{13}$$

Proof. Since z^* is a KKT point of $\max\left\{z^\top Qz : z\in\frac{1}{m}\Lambda\right\}$, there are vectors $v\in\mathbb{R}^m$ and $u\in\mathbb{R}^M_+$ such that

$$-Qz^*=H^\top v+u \quad\text{ and }\quad u^\top z^*=0\,.$$

Now for $X':=-m\,\mathrm{Diag}\,(v)$ we obtain

$$(H^\top X'H-Q)z^*=H^\top X'\tfrac{1}{m}e-Qz^*=-H^\top v+H^\top v+u=u\,,$$

so that z^* is indeed a KKT point of (12). Now for $z\in\mathbb{R}^M_+$ we have $z^\top u\geq 0=u^\top z^*$, thus also $(z-z^*)^\top u\geq 0$, wherefrom the result (13) follows immediately. □

We now establish an exact penalization result for quadratic problems under the condition that any solution to the restricted problem also is a solution to the *linearized* unrestricted problem, i.e., a KKT point of the original problem. Note that exact penalization is here not employed as usual, to simplify the problem by removing *difficult* constraints, as the constraints $\Pi\zeta=o$ would rather reduce problem dimension. Rather the result below allows us to reduce the study of (1) to that of (10). In contrast to the setup of Theorem 4.2 of Friedlander and Tseng [5] which treats the case of a convex objective function, our result does not hold without additional conditions, see Remark 3.1 below. Condition (14) is termed *ascent condition* for obvious reasons, and in some sense takes into account the non-convexity of the objective. In the sequel, the gradient $\nabla f(x)\in\mathbb{R}^n$ is always understood as a column vector.

Theorem 3.1. *Let $0<k<n$ be integers, and denote by $\Pi:\mathbb{R}^n\to\mathbb{R}^k$ the linear transformation that maps a vector to its first k coordinates. Let $f:\mathbb{R}^n\to\mathbb{R}$ be a quadratic function, let $P\subset\mathbb{R}^n$ be a convex polytope, and assume that the set $P_0:=\{\zeta\in P:\Pi\zeta=o\}$ is not empty. Define $P_0^*:=\mathrm{argmin}\,\{f(\zeta):\zeta\in P_0\}$ and assume that*

$$(\zeta-\zeta^*)^\top\nabla f(\zeta^*)\geq 0 \quad \text{for all } \zeta\in P\,,\ \zeta^*\in P_0^*\,. \tag{14}$$

Then, denoting by $\|\cdot\|$ the Euclidean norm, there is $\mu\geq 0$ such that

$$P_0^*=\mathrm{argmin}\,\left\{f(\zeta)+\mu\|\Pi\zeta\|^2:\zeta\in P\right\}.$$

Proof. Assume w.l.o.g. that $f(\zeta^*)=0$ for all $\zeta^*\in P_0^*$. Since P_0^* is a finite union of convex polytopes by Lemma 3.1 of Luo and Tseng [8], the convex hull of P_0^* is a polytope, say

$$\mathrm{conv}(P_0^*)=\mathrm{conv}(\{\zeta_1^*,\dots,\zeta_\ell^*\})$$

with $\zeta_k^*\in P_0^*$ for $1\leq k\leq\ell$, and ℓ minimal. For any $z\in\mathbb{R}^n$ we define the set

$$C_0^*(z):=\left\{z+\sum_{i=1}^{\ell}\alpha_i(z-\zeta_i^*):\alpha_i\geq 0, 1\leq i\leq\ell\right\}.$$

Fix μ for the moment, define $f_\mu(z) := f(z) + \mu\|\Pi z\|^2$ and assume there is $z \in P$ such that $f_\mu(z) < 0$. We claim that then for any $\zeta = z + \sum_{i=1}^{\ell} \alpha_i(z - \zeta_i^*) \in P \cap C_0^*(z)$ we have $f_\mu(\zeta) < 0$. This claim is proved by using an alternative representation for ζ,

$$\zeta = \prod_{i=1}^{\ell}(1+\beta_i)z - \sum_{j=1}^{\ell}\beta_j \prod_{i=j+1}^{\ell}(1+\beta_i)\zeta_j^*\,,$$

where $\beta_i = \frac{\alpha_i}{1+\sum_{j>i}\alpha_j} \geq 0$ for $1 \leq i \leq \ell$. Denoting moreover

$$\zeta_k := \prod_{i=1}^{k}(1+\beta_i)z - \sum_{j=1}^{k}\beta_j \prod_{i=j+1}^{k}(1+\beta_i)\zeta_j^*,$$

we have (by the usual default conventions) $\zeta_0 = z$ and also $\zeta_\ell = \zeta$, as well as $\zeta_{k+1} = (1+\beta_{k+1})\zeta_k - \beta_{k+1}\zeta_{k+1}^*$ for $0 \leq k < \ell$. Hence $\zeta_k \in \mathrm{conv}(\zeta_{k+1}, \zeta_{k+1}^*)$, so that $\zeta_{k+1} \in P$ implies $\zeta_k \in P$ for $0 \leq k < \ell$, which by assumption on $\zeta = \zeta_\ell$ establishes $\{z = \zeta_0, \ldots, \zeta = \zeta_\ell\} \subset P$. Now observe that $f_\mu(\zeta_k)$ is decreasing in k for $0 \leq k < \ell$: to this end, consider the univariate quadratic function $g_k(\alpha) = f_\mu(\zeta_k^* + \alpha(\zeta_{k-1} - \zeta_k^*))$. Since $g_k(0) = f_\mu(\zeta_k^*) = f(\zeta_k^*) = 0$ as well as $g_k'(0) = (\zeta_{k-1} - \zeta_k^*)^\top \nabla f(\zeta_k^*) \geq 0$ by assumption, the property $g_k(1) = f_\mu(\zeta_{k-1}) < 0$ implies that $g_k(\alpha)$ is strictly decreasing for $\alpha \geq 1$, and thus

$$f_\mu(\zeta_k) = g_k(1+\beta_k) < g_k(1) = f_\mu(\zeta_{k-1}) < 0\,.$$

Clearly $f_\mu(\zeta_0) < 0$ by assumption, which settles the claim by induction. So we have established a subset $P \cap C_0^*(z)$ of feasible points with negative f_μ-values. We now study points within this set $P \cap C_0^*(z)$ farthest away from P_0^*. To this end define the system of sets $\mathcal{P} := \{F : F \text{ is a face of } P\}$ and $\mathcal{F} := \{F \in \mathcal{P} : F \cap P_0^* = \emptyset\}$, as well as $E_0^* := \bigcup(F : F \in \mathcal{F})$. Then clearly

$$\mathrm{dist}(E_0^*, P_0^*) = \min\{\mathrm{dist}(F, P_0^*) : F \in \mathcal{F}\} =: \delta > 0\,,$$

where we denote by $\mathrm{dist}(\mathcal{S}_1, \mathcal{S}_2) := \inf\{\|s_1 - s_2\| : s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2\}$ the distance of the subsets $\mathcal{S}_1, \mathcal{S}_2$ of some euclidian space. Now we claim that $C_0^*(z) \cap E_0^* \neq \emptyset$ if $f_\mu(z) < 0$ and $z \in P$. Indeed, if for some vertex v of P we have $v \in C_0^*(z)$, then also $\{v\} \in \mathcal{F}$, since $C_0^*(z) \cap P_0^* = \emptyset$ as $f_\mu(\zeta) = 0$ on P_0^* by assumption, and we are done. Else let $j \geq 1$ be the smallest number such that $C_0^*(z) \cap F^* \neq \emptyset$ for some $F^* \in \mathcal{P}$ of dimension j, but $C_0^*(z) \cap F = \emptyset$ for all $F \in \mathcal{P}$ of dimension strictly less than j. This also means that $C_0^*(z) \cap \partial F^* = \emptyset$ holds. If $P = \mathrm{conv}(p_1, \ldots, p_r)$ we w.l.o.g. may and do assume that $F^* = \mathrm{conv}(p_1, \ldots, p_s)$ for some s, $1 < s \leq r$. Then there is a supporting hyperplane of P with normal vector q containing F^* such that

$$q^\top(p_j - p_1) = 0 < q^\top(p_k - p_1) \quad \text{for all } j,k \text{ with } 1 \leq j \leq s < k \leq r\,. \tag{15}$$

We have to show that $F^* \in \mathcal{F}$. Assume this is not the case. Then we could pick a $\zeta \in F^* \cap P_0^* \subseteq F^* \cap \mathrm{conv}P_0^* \subseteq P$, and hence

$$0 = q^\top(\zeta - p_1) = \sum_{i=1}^{\ell}\lambda_i q^\top(\zeta_i^* - p_1)\,, \text{ with } \lambda_i \geq 0 \text{ for } 1 \leq i \leq \ell \text{ and } \sum_{i=1}^{\ell}\lambda_i = 1\,.$$

But as $\zeta_i^* \in P$, we always have $q^\top(\zeta_i^* - p_1) \geq 0$, hence $\lambda_i > 0$ implies $q^\top(\zeta_i^* - p_1) = 0$ (at least one such i exists, of course). Now $\zeta_i^* = \sum_{j=1}^{r}\mu_j p_j$ with $\mu_j \geq 0$ and

$\sum_{j=1}^{r} \mu_j = 1$ gives via (15) and

$$0 = q^\top(\zeta_i^* - p_1) = \sum_{j=1}^{r} \mu_j q^\top(p_j - p_1) = \sum_{k=s+1}^{r} \mu_k q^\top(p_k - p_1)$$

finally $\mu_k = 0$ for all $k > s$, or, equivalently, $\zeta_i^* = \sum_{j=1}^{s} \mu_j p_j \in F^*$. But then, for some $\alpha \geq 0$ we would have $z + \alpha(z - \zeta_i^*) \in C_0^*(z) \cap \partial F^*$, which is absurd by construction. Note that if $F^* = P$ (i.e., if $s = r$), we can assess $\zeta_i^* \in P = F^*$ directly, avoiding (15), so that the argument also remains valid in this case. Hence we have established that $f_\mu(z) < 0$ for some $z \in P$ implies that there is also a $\zeta \in P$ with $f_\mu(\zeta) < 0$ but $\operatorname{dist}(\zeta, P_0^*) \geq \delta$. The remainder of the proof is now straightforward: Assume that there were no $\mu > 0$ with $P_0^* = \operatorname{argmin}\{f_\mu(\zeta) : \zeta \in P\}$. Then there were a sequence $(\zeta_\nu)_{\nu \in \mathbb{N}} \subset P \setminus P_0$ such that $f(\zeta_\nu) + \nu \|\Pi\zeta_\nu\|^2 < 0$. By the above reasoning, we may and do assume that $\operatorname{dist}(\zeta_\nu, P_0^*) \geq \delta$. Let now $\zeta_\infty \in P$ be an accumulation point of $(\zeta_\nu)_{\nu \in \mathbb{N}}$. Passing to a suitable subsequence we also may assume $\zeta_\nu \to \zeta_\infty$. We obtain $\|\Pi\zeta_\nu\|^2 < -\frac{1}{\nu} f(\zeta_\nu) \leq \frac{1}{\nu} \max\{-f(\zeta) : \zeta \in P\}$ and thus $\|\Pi\zeta_\infty\| = 0$, i.e., $\zeta_\infty \in P_0$. Moreover $\zeta_\infty \in P_0 \setminus P_0^*$, since $\operatorname{dist}(\zeta_\infty, P_0^*) \geq \delta$. On the other hand, by continuity of f we have $f(\zeta_\infty) \leq 0$ implying $\zeta_\infty \in P_0^*$, which is a contradiction. □

Remark 3.1. *The above penalization result does not hold in general, e.g., if*

- *the ascent condition (14) over the minimizers P_0^* is violated:*
  Example: $k = 1, n = 2, f(x, y) = xy, P = \{-1 \leq x, y \leq 1\}$.
- *P is (convex, but) not a polytope:*
  Example: $k = 1, n = 2, f(x, y) = xy, P = \{-1 \leq y \leq 1, y^4 \leq x \leq 1\}$.
- *f is not a quadratic function:*
  Example: $k = 1, n = 2, f(x, y) = xy + y^4, P = \{-1 \leq x, y \leq 1\}$.

We can however allow P to be a finite union of convex polytopes, without need to change the proof. More generally we can allow P to be any bounded set such that for some $\varepsilon > 0$ and a finite subset $\{z_1, \ldots, z_N\} \in P_0^*$ the sets $C_i := z_i + [-\varepsilon, \varepsilon]^M$ satisfy: $C_i \cap P$ are convex polytopes for $1 \leq i \leq N$, and we have $P_0^* \in \operatorname{int} \bigcup_{i=1}^{N} C_i$. Denoting $C_\varepsilon := \bigcup_{i=1}^{N} C_i \cap P$, we have $\operatorname{dist}(P_0^*, P \setminus C_\varepsilon) \geq \varepsilon$. On C_ε we can use the proof given above, yielding some $\mu_1 \geq 0$. Furthermore f is lower bounded by $c_1 - c_2\|\Pi\zeta\|$ on $P \setminus C_\varepsilon$, with $c_1 > 0, c_2 \in \mathbb{R}$, and clearly for suitable μ_2 we have $c_1 - c_2\|\Pi\zeta\| + \mu\|\Pi\zeta\|^2 > 0$ on $P \setminus C_\varepsilon$. Finally $\mu := \max(\mu_1, \mu_2)$ will work for the whole feasible domain P.

The condition on the gradient below includes trivially the case of a unique solution, but also, e.g., the case of a constant or linear objective. Notice that over Λ, any such function can be represented by a suitable quadratic form. However, the matrix Q used in the example for Proposition 2.1(d) is *not* suitable in that respect: the objective $z^\top Qz$ indeed is constantly zero over Λ whereas the gradient Qz varies with $z \in \Lambda$. This also shows that some properties of the conic reformulation may crucially depend upon the choice of Q among value-equivalent alternatives. Note that for single StQPs, i.e., for (1) with $m = 1$, there are no such value-equivalent alternatives as $\{x^\top Qx : x \in \Delta\}$ determines Q uniquely.

Theorem 3.2. *If the gradient function Qx is constant over the set of optimal solutions to (1), then the minimum of (5) is attained.*

Proof. Let $\bar{z}$ denote a solution to (1). Then $z^* := \frac{1}{m}\bar{z} \in \Delta$. Now, with X' from Lemma 3.2, let us abbreviate $q_{X'}(z) = z^\top(H^\top X'H - Q)z$. Since $z^\top H^\top X'Hz$ is constant over

$$\tfrac{1}{m}\Lambda = \left\{z \in \Delta : Hz = \tfrac{1}{m}e\right\} = \left\{z \in \mathbb{R}^M_+ : Hz = \tfrac{1}{m}e\right\} ,$$

the point z^* is not only a KKT point of (12) as stated in Lemma 3.2 (and thus also a KKT point of $\min\{q_{X'}(z) : z \in \Delta\}$,) but also a solution to

$$\min\left\{q_{X'}(z) : z \in \Delta\, ,\, Hz = \tfrac{1}{m}e\right\} .$$

This, after an affine change of variables, is exactly the situation met in Theorem 3.1. To be precise, we let $U = [H^\top|R]^\top$ be some orthogonal $M \times M$ matrix completion of H (the first m rows of U constitute H), and define ζ via $z - z^* = U^\top\zeta$ or $\zeta = U(z - z^*)$. If $\Pi\zeta = [\zeta_1, \dots, \zeta_m]^\top$, then $\Pi\zeta = Hz - Hz^* = Hz - \frac{1}{m}e$. Put $P := \{U(z - z^*) : z \in \Delta\}$ and

$$P_0 := \{U(z - z^*) : z \in \Delta\, ,\, \Pi\zeta = o\} = \left\{U(z - z^*) : z \in \Delta\, ,\, Hz = \tfrac{1}{m}e\right\} .$$

Let $f(\zeta) = q_{X'}(U^\top\zeta + z^*)$. Then $\zeta = o$ belongs to

$$P_0^* = \operatorname{argmin}\{f(\zeta) : \zeta \in P_0\} .$$

Now consider any solution $\hat{\zeta} \in P_0^*$ and put $\hat{z} = U^\top\hat{\zeta} + z^*$. Then $f(\hat{\zeta}) = f(o) = q_{X'}(z^*) = q_{X'}(\hat{z})$ by construction, so that also $\hat{z}$ solves (1), but the transport from X' depending on z^* to the corresponding counterpart emerging from $\hat{z}$ is not obvious. So next we establish the ascent condition (14) directly, employing the gradient condition in the theorem, which gives $(H^\top X'H - Q)(\hat{z} - z^*) = o$, noting $H\hat{z} = \frac{1}{m}e = Hz^*$. Now for any $\zeta \in P$ we get $z = z^* + U^\top\zeta \in \Delta \subset \mathbb{R}^M_+$ by construction, so that

$$\begin{aligned}
\frac{1}{2}(\zeta - \hat{\zeta})^\top\nabla f(\hat{\zeta}) &= (z - \hat{z})^\top(H^\top X'H - Q)\hat{z} \\
&= z^\top(H^\top X'H - Q)z^* + z^\top(H^\top X'H - Q)(\hat{z} - z^*) - q_{X'}(\hat{z}) \\
&= (z - z^*)^\top(H^\top X'H - Q)z^* + q_{X'}(z^*) - q_{X'}(\hat{z}) \\
&= (z - z^*)^\top(H^\top X'H - Q)z^* \\
&\geq 0
\end{aligned}$$

due to (13). So Theorem 3.1 yields a $\mu \geq 0$ such that z^* solves

$$\min\left\{z^\top\left(H^\top X'H - Q\right)z + \mu\|Hz - \tfrac{1}{m}e\|^2 : z \in \Delta\right\}$$

with zero objective value. Finally, putting $X = X' + \frac{\mu}{m}Y$ with Y as in (11), Lemma 3.1 yields the desired result. □

Notice that, again in the example for Proposition 2.1(d), we do in fact have $(z - z^*)^\top Q(\hat{z} - z^*) = 0$ for all $z \in \frac{1}{m}\Lambda$, but for the attainability proof to work as above, we need $(z - z^*)^\top Q(\hat{z} - z^*) = 0$ also for all $z \in \Delta \setminus \frac{1}{m}\Lambda$, and this condition is violated. This example also shows that a condition similar to that appearing in Lemma 4.5 of Monteiro and Wang [10], namely that $Hd = o$ implies $d^\top Q(z - z') = 0$ for all solutions z, z' to (1), is not sufficient to guarantee dual attainability here.

Acknowledgment

The authors are indebted to Paul Tseng, for valuable hints and thoughtful remarks on an earlier version of this paper.

References

[1] E. G. Belousov and D. Klatte, *A Frank-Wolfe Type Theorem for Convex Polynomial Programs*, Comput. Optim. Appl. **22** (2002), 37–48.

[2] I. M. Bomze, L. Palagi and W. Schachinger, *Multi-Standard Quadratic Optimization Problems*, presented at **TE-29**, EURO XXI, Reykjavik, Iceland, July 2006, see also http://www.euro2006.org/EuroFinalProgramme.pdf, last accessed Dec.06 (2006).

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, UK, 2004.

[4] S. Burer, *On the copositive representation of binary and continuous nonconvex quadratic programs*, Preprint, Univ. of Iowa, available at http://www.optimization-online.org/DB_HTML/2006/10/1501.html, last accessed Dec.06 (2006).

[5] M. P. Friedlander and P. Tseng, *Exact regularization of convex programs*, Department of Computer Science Tech. Report TR-2006-26, November 2006, University of British Columbia, available at http://www.optimization-online.org/DB_HTML/2006/11/1531.html, last accessed Dec.06 (2006).

[6] K. P. Hadeler, *On copositive matrices*, Linear Algebra Appl. **49** (1983), 79–89.

[7] R. A. Hummel and S. W. Zucker, *On the foundations of relaxation labeling processes*, IEEE Trans. Pattern Anal. Machine Intell. **5** (1983), 267-287.

[8] Z.-Q. Luo and P. Tseng, *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem*, SIAM J. Optim. **2** (1992), 43-54.

[9] Z.-Q. Luo and Sh. Zhang, *On Extensions of the Frank–Wolfe Theorems*, Comput. Optim. Appl. **13** (1999), 87–110.

[10] R. D. C. Monteiro and Y. Wang, *Trust region affine scaling algorithms for linearly constrained convex and concave programs*, Math. Prog. **80** (1998), 283–313.

[11] A. Ozdaglar and P. Tseng, *Existence of Global Minima for Constrained Optimization*, J. Optim. Theory Appl. **128**, 523–546 (2006).

[12] G. Pataki, *On the closedness of the linear image of a closed convex cone*, Preprint, Univ. of North Carolina, available at http://www.optimization-online.org/DB_HTML/2006/12/1556.html, last accessed Feb.07 (2006).

[13] M. Pelillo, *The dynamics of nonlinear relaxation labeling processes*, Journal of Mathematical Imaging and Vision **7** (1997), 309-323.

[14] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

Corresponding author: Werner Schachinger, Dept. of Statistics and Decision Support Systems, University of Vienna, Brünnerstr. 72, A-1210 Wien, Austria

E-mail address: `Werner.Schachinger@univie.ac.at`