On exploiting structure induced when modelling an intersection of cones in conic optimization

E. D. Andersen*

July 20, 2004

Abstract

Conic optimization is the problem of optimizing a linear function over an intersection of an affine linear manifold with the Cartesian product of convex cones. However, many real world conic models involves an intersection rather than the product of two or more cones. It is easy to deal with an intersection of one or more cones but unfortunately it leads to an expansion in the optimization problem size and hence to an increase in the computational complexity of solving the optimization problem. In this note we discuss how to handle the intersection of two or more cones. In particular we show that the important special case of the intersection of a linear and a quadratic cone can be handled in a computational efficient way.

1 Introduction

The conic optimization problem can formally be stated as

$$\begin{array}{lll} \text{minimize} & \sum_{k=1}^{r} (c^k)^T x^k & \\ \text{subject to} & \sum_{k=1}^{r} A^k x^k \quad = & b, \\ & x^k \in \mathcal{K}^k, & k = 1, \ldots, r, \end{array} \tag{1}$$

*MOSEK ApS, Fruebjergvej 3 Box 16, 2100 Copenhagen, Denmark Email: e.d.andersen@mosek.com.

where $c^k \in R^{n^k}$, $A^k \in R^{m \times n^k}$ and $b \in R^m$. $\mathcal{K}^k$ is assumed to be a convex cone.

It is well-known that any convex optimization problem can be stated in conic form [6] but until now the linear, quadratic, and semi-definite case have achieved the most attention because highly efficient primal-dual interior-point algorithms have been developed for this class of problems [7, 8]. Moreover, this special case has many practical applications [1].

Observe the conic constraints

$$x^k \in \mathcal{K}^k, \ k = 1, \ldots, r,$$

are the same as saying

$$\begin{pmatrix} x^1 \\ \vdots \\ x^r \end{pmatrix} \in \mathcal{K}^1 \times \cdots \times \mathcal{K}^r.$$

This seems to exclude the case

$$y \in K^1 \cap K^2 \tag{2}$$

where a variable y has to lie in the intersection of two cones. However, (2) is equivalent to

$$\begin{array}{rcl} y - z & = & 0, \\ y \in \mathcal{K}^1, & & \\ z \in \mathcal{K}^2. & & \end{array}$$

Therefore, by introducing some additional constraints and variables the intersection of two cones can easily be modelled within the framework of conic optimization. This unfortunately leads to an expansion in the problem dimension which makes the problem computationally more expensive to solve.

Observe the type of constraints introduced when modelling the intersection of two cones are very special and perhaps the special structure can be exploited in the solution algorithm. In this note we show that this is indeed the case when modelling the intersection of a linear and a quadratic cone.

2 Notation

The problem of interest is the conic optimization problem

$$\begin{array}{lll} \text{minimize} & \sum_{k=1}^{r} (c^k)^T x^k & \\ \text{subject to} & \sum_{k=1}^{r} A^k x^k = b, & \\ & x^k \in \mathcal{K}^k, & k = 1, \ldots, r, \end{array} \tag{3}$$

where $A^k \in R^{m \times n^k}$ and all other quantities have conforming dimensions. $\mathcal{K}^k$ is assumed to be a convex cone. Without loss of generality we will assume that the matrix

$$[\ A^1 \quad A^2 \quad \cdots \quad A^r\ ]$$

has full row rank.

We will restrict our attention to the case where $\mathcal{K}^k$ is a symmetric cone, i.e., $\mathcal{K}^k$ is either

- A linear cone:
$$\{x \in \mathcal{R} : \ x \geq 0\}.$$
- A quadratic cone:
$$\{x \in \mathcal{R}^n : \ x_1 \geq \|x_{2:n}\|\}.$$
- A semi-definite cone (we assume $n^k = p^2$):
$$\{vec(X) : X \in \mathcal{R}^{p \times p} \text{ is symmetric and positive semi-definite}\}.$$

Subsequently we will need the Sherman-Morrison-Woodbury formula

$$(H + VGV^T)^{-1} = H^{-1} - H^{-1}V(G^{-1} + V^T H^{-1} V)^{-1} V^T H^{-1} \tag{4}$$

where it is assumed all the required inverses exists [4, p. 243].

3 A motivating example

We will use the motivating example

$$\begin{array}{lll} \text{minimize} & \|x\| + c^T x & \\ \text{subject to} & Ax = b, & \\ & x \geq 0. & \end{array} \tag{5}$$

where $A \in \mathcal{R}^{m\times n}$ and all other quantities have conforming dimensions. This problem can be reformulated as

$$\begin{array}{lcl} \text{minimize} & c^T x + t & \\ \text{subject to} & Ax & = \ b, \\ & (t;x) \in \{(t;x):\ t \geq \|x\|\} \cap \{(t;x):\ x \geq 0\}. & \end{array} \tag{6}$$

However, this problem is not in conic form because it involves an intersection of two cones but the reformulated problem

$$\begin{array}{lcl} \text{minimize} & c^T x + t & \\ \text{subject to} & Ax & = \ b, \\ & x - z & = \ 0, \\ & (t;z) \in \{(t;z):\ t \geq \|z\|\}, & \\ & x \in (x:\ x \geq 0\}. & \end{array} \tag{7}$$

is in conic form. Observer the reformulation has the obvious drawback that n new constraints and variables are introduced into the problem which makes it much bigger and more expensive to solve.

4 Intersection of linear and quadratic cones

In this section we restrict our attention to the case of linear and quadratic cones. For this restricted class of problems the primal-dual interior-point algorithm suggested in Nesterov and Todd (NT) [7, 8] is the most efficient solution algorithm both in theory and practice. A detailed discussion of the implementation of NT algorithm can be seen in [2].

The main computational work performed in each iteration of the NT algorithm is the solution of a system of linear equations of the form

$$\sum_{k=1}^{r} A^k (W^k)^{-1} A^k d_y = f \tag{8}$$

where W^k is a positive definite matrix, d_y is the unknowns, and f is an arbitrary right-hand side. W^k is called the NT scaling matrix. This implies that the coefficient matrix

$$\sum_{k=1}^{r} A^k (W^k)^{-1} A^k \tag{9}$$

is a symmetric positive definite matrix. Moreover, the matrix tends to be sparse which implies that the system linear equations (8) can be solved efficiently using a sparse Cholesky factorization. Therefore, this is the common approach employed in most implementations [1, p. 45].

The NT scaling matrices W^k all have the special form

$$W^k = \theta_k^2(-Q^k + 2w^k(w^k)^T).$$

where θ^k is a positive scalar, $w^k \in \mathcal{R}^{n^k}$ and if cone k is linear cone then Q^k has the form

$$Q^k = 1$$

whereas if cone k is quadratic then Q^k has the form

$$Q^k = diag(-1, 1, \ldots, 1).$$

Hence, for linear cones W^k is a positive scalar whereas for quadratic cones W^k is a diagonal matrix plus a rank-1 term. This is an important fact in practice because it implies (9) can be computed efficiently. We refer to [2] for further details about the NT scaling matrices. From [2] we recall the important fact

$$\begin{array}{rcl} (W^k)^{-1} & = & \theta_k^{-2}(-Q^k + 2(Q^k w^k)(Q^k w^k)^T) \\ & = & \theta_k^{-2}(I + 2(Q^k w^k (Q^k w^k)^T - e_1 e_1^T)). \end{array}$$

This demonstrates that $(W^k)^{-1}$ can be written as a sum of a positive definite diagonal matrix plus a rank-2 term.

Next we show the special structure arising when modelling the intersection of a linear and a quadratic cone can be exploited to reduce the computational cost when solving the system (8).

The problem of interest is

$$\begin{array}{lccl} \text{minimize} & (c^1)^T x^1 & & \\ \text{subject to} & A^{11} x^1 & = & b^1, \\ & A^{21} x^1 + A^{22} x^2 & = & b^2, \\ & x_1^1 \geq \|x_{2:n^1}^1\|, & & \\ & x^2 \geq 0 & & \end{array} \tag{10}$$

where $c^1 \in R^{n^1}$, $A^{11} \in R^{m^1 \times n^1}$, $A^{21} \in R^{m^2 \times n^1}$, $A^{22} \in R^{m^2 \times n^2}$, $b^1 \in R^{m^1}$ and $b^2 \in R^{m^2}$. Note this problem has one quadratic cone and the x^1 variables belong to this cone. We will make the following assumption

Assumption 4.1 *i) Each row of A^{21} has exactly one nonzero element.*

ii) A^{22} is a nonsingular diagonal matrix.

Therefore, the set of constraints

$$\begin{array}{rcl} A^{21}x^1 + A^{22}x^2 & = & b^2, \\ x^2 \geq 0 & & \end{array}$$

can be used to model simple bounds on the x^1 variables. For instance the bound constraint

$$x^1_j \geq b^2_k$$

can be modelled using the set of constraints

$$\begin{array}{rcl} e_j^T x^1 - x^2_k & = & b^2_k, \\ x^2_k \geq 0. & & \end{array}$$

It should be obvious that the problem (10) allows us to deal with the intersection of the linear and quadratic cone in a flexible way.

In the case of problem (10) the system (9) has the form

$$\left[\begin{array}{cc} A^{11} & 0 \\ A^{21} & A^{22} \end{array} \right] \left[\begin{array}{cc} W^{-1} & 0 \\ 0 & D^{-1} \end{array} \right] \left[\begin{array}{cc} A^{11} & 0 \\ A^{21} & A^{22} \end{array} \right]^T \left[\begin{array}{c} d^1_y \\ d^2_y \end{array} \right] = \left[\begin{array}{c} f^1 \\ f^2 \end{array} \right]$$

or equivalently

$$\left[\begin{array}{cc} A^{11}W^{-1}(A^{11})^T & A^{11}W^{-1}(A^{21})^T \\ A^{21}W^{-1}(A^{11})^T & A^{21}W^{-1}(A^{21})^T + A^{22}D^{-1}(A^{22})^T \end{array} \right] \left[\begin{array}{c} d^1_y \\ d^2_y \end{array} \right] = \left[\begin{array}{c} f^1 \\ f^2 \end{array} \right]. \tag{11}$$

W is the NT scaling matrix associated with the quadratic cone and D is the diagonal NT scaling matrix associated with the linear cones i.e. the $x^2 \geq 0$ constraint.

The solution to (11) can be obtained by solving

$$A^{11}\bar{W}^{-1}(A^{11})^T d^1_y = \bar{f}^1 \tag{12}$$

and

$$(A^{21}W^{-1}(A^{21})^T + A^{22}D^{-1}(A^{22})^T)d^2_y = f^2 - A^{21}W^{-1}(A^{11})^T d^1_y \tag{13}$$

where

$$\begin{aligned}\bar{W} &:= (W^{-1} - W^{-1}(A^{21})^T(A^{21}W^{-1}(A^{21})^T + A^{22}D^{-1}(A^{22})^T)^{-1}A^{21}W^{-1})^{-1}\\ \bar{f}^1 &:= f^1 - A^{11}W^{-1}(A^{21})^T(A^{21}W^{-1}(A^{21})^T + A^{22}D^{-1}(A^{22})^T)^{-1}f^2.\end{aligned}$$

Instead of solving the big system (8) we suggest to solve the two smaller systems (12) and (13). Subsequently we show that that the system (12) and (13) can be solved in $O((n^1)^3)$ and $O(n^1 + n^2)$ operations respectively. Even for moderate values of n^1 and n^2 this implies a dramatic reduction in the computational complexity because $O((n^1 + n^2)^3)$ operations is required to solve (8).

Observe the complexity of forming the right-hand side of (12) is the same as solving (13). Therefore, we will ignore this operation and first consider the solution of (12) for an arbitrary known right-hand side.

Lemma 4.1

$$\bar{W} = H^{-1} - H^{-1}V(G^{-1} + V^TH^{-1}V)^{-1}V^TH^{-1}$$

where

$$\begin{aligned}H &:= \theta^{-2}I + (A^{21})^T(A^{22}D^{-1}(A^{22})^T)^{-1}A^{21},\\ V &:= \sqrt{2}\theta^{-1}[\begin{array}{cc} w & e_1 \end{array}],\\ G &:= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.\end{aligned}$$

Proof:

$$\begin{aligned}\bar{W}^{-1} &= (W^{-1} - W^{-1}(A^{21})^T(A^{21}W^{-1}(A^{21})^T + A^{22}D^{-1}(A^{22})^T)^{-1}A^{21}W^{-1})\\ &= (W + (A^{21})^T(A^{22}D^{-1}(A^{22})^T)^{-1}A^{21})^{-1}\\ &= (\theta^{-2}(-Q + 2ww^T) + (A^{21})^T(A^{22}D^{-1}(A^{22})^T)^{-1}A^{21})^{-1}\\ &= (\theta^{-2}I + (A^{21})^T(A^{22}D^{-1}(A^{22})^T)^{-1}A^{21} + 2\theta^{-2}(ww^T - e_1e_1^T))^{-1}\\ &= (H + V^TGV)^{-1}\\ &= H^{-1} - H^{-1}V(G^{-1} + V^TH^{-1}V)^{-1}V^TH^{-1}\end{aligned}$$

The second and the sifth equality follow from (4). □

Corollary 4.1 *H is a positive definite diagonal matrix.*

Lemma 4.2

$$G^{-1} + V^TH^{-1}V = LDL^T$$

where

$$\begin{array}{rcl} D & := & 2\theta^{-2}\left[\begin{array}{cc} \frac{\theta^2}{2}+w^T H^{-1}w & 0 \\ 0 & -\frac{\theta^2}{2}+e_1^T H^{-1}e_1-\frac{(e_1^T H^{-1}w)^2}{\frac{\theta^2}{2}+w^T H^{-1}w} \end{array}\right], \\ L & := & \left[\begin{array}{cc} 1 & 0 \\ \frac{e_1^T H^{-1}w}{\frac{\theta^2}{2}+w^T H^{-1}w} & 1 \end{array}\right]. \end{array}$$

Proof:

$$\begin{array}{rl} & G^{-1}+V^T H^{-1}V \\ = & \left[\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array}\right]+2\theta^{-2}[\begin{array}{cc} w & e_1 \end{array}]^T H^{-1}[\begin{array}{cc} w & e_1 \end{array}] \\ = & 2\theta^{-2}\left[\begin{array}{cc} \frac{\theta^2}{2}+w^T H^{-1}w & w_1^T H^{-1}e_1 \\ e_1^T H^{-1}w & -\frac{\theta^2}{2}+e_1^T H^{-1}e_1 \end{array}\right] \\ = & LDL^T. \end{array}$$

□

Corollary 4.2 $\bar{W}^{-1}=H^{-1}-(H^{-1}VL^{-T})D^{-1}(H^{-1}VL^{-T})^T$.

Corollary 4.3 *Given a vector v then $H^{-1}v$ can be computed in $O(n^1+n^2)$ operations.*

The previous results demonstrate that $\bar{W}^{-1}$ is a positive definite diagonal matrix plus a rank-2 term and hence $A^{11}\bar{W}^{-1}(A^{11})^T$ can be computed efficiently. Furthermore, assuming no (lucky) cancellations then the sparsity pattern of

$$A^{11}\bar{W}^{-1}(A^{11})^T$$

is identical to the sparsity pattern of

$$A^{11}W^{-1}(A^{11})^T.$$

Hence, even when sparsity is considered then (12) should be cheaper to solve than (8).

After solving the system (12) the system (13) should be solved.

Lemma 4.3

$$\begin{array}{rl} & (A^{21}W^{-1}(A^{21})^T+A^{22}D^{-1}(A^{22})^T)^{-1} \\ = & \hat{H}^{-1}-\hat{H}^{-1}\hat{V}(G^{-1}+\hat{V}^T\hat{H}^{-1}\hat{V})^{-1}\hat{V}^T\hat{H}^{-1} \end{array}$$

where

$$\begin{array}{rcl} \hat{H} & := & A^{22}D^{-1}(A^{22})^T+\theta^{-2}A^{21}(A^{21})^T, \\ \hat{V} & := & \sqrt{2}\theta^{-1}A^{21}[\begin{array}{cc} Qw & e_1 \end{array}]. \end{array}$$

Proof:

$$
\begin{aligned}
& (A^{21}W^{-1}(A^{21})^T + A^{22}D^{-1}(A^{22})^T)^{-1} \\
= \; & (A^{22}D^{-1}(A^{22})^T + \theta^{-2}A^{21}(I + 2((Qw)(Qw)^T - e_1e_1^T))(A^{21})^T)^{-1} \\
= \; & (A^{22}D^{-1}(A^{22})^T + \theta^{-2}A^{21}(A^{21})^T) + 2\theta^{-2}A^{21}(Qw(Qw)^T - e_1(e_1)^T)(A^{21})^T)^{-1} \\
= \; & (\hat{H} + \hat{V}^T G \hat{V})^{-1} \\
= \; & \hat{H}^{-1} - \hat{H}^{-1}\hat{V}(G^{-1} + \hat{V}^T\hat{H}^{-1}\hat{V})^{-1}\hat{V}^T\hat{H}^{-1}.
\end{aligned}
$$

□

Corollary 4.4 *The matrix $\hat{H}$ is a positive definite diagonal matrix.*

Corollary 4.5 *Given a vector v then $(A^{21}W^{-1}(A^{21})^T + A^{22}D^{-1}(A^{22})^T)^{-1}v$ can be computed in $O(n^1 + n^2)$ operations.*

We have now demonstrated that when a quadratic cone is intersected with several linear cones then this give rise to a special problem structure which can be exploited. Moreover, when the special structure is exploited the computational costs can be reduced significantly.

Finally, it should be mentioned that the technique presented above is also applicable if the problem contains multiple quadratic cones because the cones are independent and can each be treated in the same way.

5 Discussion

In summary we have shown that in the case where a linear cone is intersected with a quadratic cone then a special structure appears which can be exploited to reduce the computational costs. An important topic for further research is whether the idea presented in this paper can be generalized.

The paper [5] presents an application which requires solution of a problem of the form

$$
\begin{array}{lrcl}
\text{min} & \sum_j t_j + c^T x & & \\
\text{subject to} & Ax & = & b, \\
& \|B^j x\|^2 & \leq & t_j, \quad j = 1, \ldots, m, \\
& 0 \leq x \leq u. & &
\end{array}
\tag{14}
$$

where A has very few rows and B^j is close to full rank. The conic reformulation of this problem is

$$\begin{array}{lccll} \min & \sum_j t_j + c^T x & & & \\ \text{subject to} & Ax & = & b, & \\ & B^j x - z^j & = & 0, & j = 1, \ldots, m, \\ & v_j & = & 0.5, & j = 1, \ldots, m, \\ & \|z^j\|^2 \leq 2 v_j t_j, & & & j = 1, \ldots, m, \\ & 0 \leq x \leq u, 0 \leq t_j. & & & \end{array} \tag{15}$$

It is obvious that the conic reformulation has many more constraints and variables than the nonconic formulation. In practice this leads to slow solution time of the conic problem. However, the conic reformulation has the intersection of cones structure.

Another interesting intersection cone is

$$\{vec(X) : X \in \mathcal{R}^{p \times p} \text{ is symmetric and positive semi-definite}\} \\ \cap \{X \in \mathcal{R}^{p \times p} : x_{ij} \geq 0\}$$

which appears in [3].

6 Conclusions

We conclude by observing that there exist important applications where modelling the intersection of one or more cones are required. Moreover, we have showed that this give rise to a special structure which in one important case can be exploited in the computations. Finally, we think the idea presented in this note can be generalized to other cases as well.

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Programming*, 95(1):3–51, 2003.

[2] E. D. Andersen, C. Roos, and T. Terlaky. On implementing a primal-dual interior-point method for conic quadratic optimization. *Math. Programming*, 95(2), February 2003.

[3] K. M. Anstreicher and S. Burer. D.C. versus copistive bounds for standard QP. Technical report, Department of Management Sciences, The University of Iowa, May 2003.

[4] I. S. Duff, A. M. Erisman, and J. K. Reid. *Direct methods for sparse matrices*. Oxford University Press, New York, 1989.

[5] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M.I. Jordan. Learning the kernel matrix with semi-definite programming. Technical Report Technical Report CSD-02-1206, Division of Computer Science, University of California, Berkeley, 2002.

[6] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, PA, 1 edition, 1994.

[7] Y. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22(1):1–42, February 1997.

[8] Yu. E Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. on Optim.*, 8:324–364, 1998.