成蹊大学理工学研究報告
J. Fac. Sci.Tech., Seikei Univ.
Vol.53 No.2 (2016)pp.15-20

# Worst case bounds on facial reduction for conic programming

Bruno Figueira Lourenço [*1], Masakazu Muramatsu [*2], Takashi Tsuchiya [*3]

**ABSTRACT** : Conic linear programming is a powerful modelling technique with many applications in engineering, planning, statistics and many others. Typically, a conic linear program (CLP) is expressed as the task of minimizing some linear function subject to linear equations and conic constraints. Sometimes, however, the CLPs can exhibit nasty theoretical behavior. This is where regularization techniques come to play. They fix ill-behaved problems and put them in a shape that solvers can successfully handle them. In this note, we present a brief account of Facial Reduction Algorithms and discuss worst case bounds for their termination.

**Keywords** : conic linear programming, facial reduction

(Received October 21, 2016)

## 1. Introduction

A conic linear program (CLP) corresponds to the following optimization problem:

$$\begin{aligned} &\textbf{minimize } \mathbf{c}^{\mathrm{T}}\boldsymbol{x} \qquad (\boldsymbol{P}) \\ &\text{subject to } \mathrm{A}x = \mathrm{b} \\ &x \in K, \end{aligned}$$

where A is a $m \times n$ real matrix, $\mathrm{b} \in \Re^{\mathrm{m}}, \mathrm{c} \in \Re^{n}$ and $K$ is a *closed convex cone.* By a "convex cone", we mean any subset of $\Re^{\mathrm{n}}$ such that for all $x, y \in K, \alpha, \beta \in \Re$ we have $\alpha x + \beta y \in K$. The "closed" part just means that $K$ is closed as set of $\Re^{\mathrm{n}}$ in the usual Euclidean topology.

Depending on the set $K$, we get different types of problems. For instance, when $K$ is the nonnegative orthant $\Re^{\mathrm{n}}_{+} = \{\mathrm{x} \in \Re^{\mathrm{n}} \mid x_i \geq 0, \forall i\}$, we have *Linear Programming* (LP). When $K$ is the set of $\mathrm{n} \times \mathrm{n}$ positive semidefinite symmetric matrices, we have *Semidefinite Programming (SDP).* Numerous applications of SDPs can be seen, for instance, in the survey by Todd [1].

Typically, a CLP is solved through the so-called *interior point methods (IPMs)* [2], which are algorithms that that have been developed in the 90s and continue to be actively researched.

For an IPM to succeeded in solving $(P)$, it is typically required that some special conditions be satisfied. In order to explain them, we need to introduce the *dual problem.* Recall that $(P)$ is called the *primal* problem and its dual counterpart is the following optimization problem:

$$\begin{aligned} &\textbf{maximize } \mathbf{b}^{\mathrm{T}}\boldsymbol{y} \qquad (\boldsymbol{D}) \\ &\text{subject to } \mathrm{c} - \mathrm{A}^{\mathrm{T}}y \in K^*, \end{aligned}$$

where $K^* = \{z \in \Re^n \mid z^T x \geq 0, \forall x \in K\}$ is the so-called *dual cone* of $K$ and $\mathrm{A}^{\mathrm{T}}$ denotes the adjoint of A.

Let $\theta_{\mathrm{P}}, \theta_{\mathrm{D}}$ denote the optimal values of (P) and (D) respectively, where it is understood that $\theta_{\mathrm{P}} = -\infty$ if (P) is unbounded and $\theta_{\mathrm{P}} = +\infty$ if (P) is infeasible, that is, it admits no feasible solution. Similarly, $\theta_{\mathrm{D}} = +\infty$ if (D) is unbounded and $\theta_{\mathrm{D}} = -\infty$ if (D) is infeasible.

Note that in this general context, even if $\theta_{\mathrm{P}}$ is finite we cannot take for granted the existence of a primal optimal solution. The same remark holds for (D).

With this notation, the *weak duality theorem* states that we always have $\theta_{\mathrm{P}} \geq \theta_D$. When, in fact, we have $\theta_{\mathrm{P}} = \theta_D$, we say that the *duality gap is zero.* From both theoretical and practical perspectives, it is desirable that a problem has zero duality gap, but unfortunately that is not always the case. In contrast to what happens in Linear Programming, we cannot take zero duality gap for granted.

For IPMs to work it is usually assumed that both (P) and (D)

*1 : Seikei University, Department of Computer and Information Science. (lourenco@st.seikei.ac.jp).

*2 : The University of Electro-Communications, Department of Computer Science.

*3 : National Graduate Institute for Policy Studies.

have what is called *relative interior feasible solutions.* First of all, denote by $\mathrm{span}(K)$ the *smallest subspace that contains* $K$. Then, the relative interior of $\mathrm{ri}(K)$ is the topological interior of $K$ in the Euclidean topology induced by $\mathrm{span}(K)$. That is: $\mathrm{ri}(K) = \{x \in K \mid \exists$ an open set $U \subseteq \Re^n$ such that $x \in U$ and $U \cap K \subseteq span(K)\}$.

More details on the properties of the relative interior and a general discussion on convex analysis can be seen in the classical book by Rockafellar [3].

A *primal relative interior feasible solution* is some vector $x$ satisfying the constraints in (P) and also satisfying $x \in \mathrm{ri}(K)$. When such a $x$ exists, we say that *Slater's condition* is satisfied for (P) or that (P) is *strongly feasible.* Similarly, a *dual relative interior feasible solution* is some y satisfying the constraints in (D) such that $c - A^T y \in \mathrm{ri}(K^*)$. When such an $y$ exists, we say that (D) is strongly feasible or that the *Slater's condition* is satisfied for (D).

Then, the *strong duality theorem for conic programs* states that if both (P) and (D) are strongly feasible, then we have $\theta_P = \theta_D$ and there are both primal and dual optimal solutions.

Unfortunately, as we remarked before, not all problems are strongly feasible. And this can cause a series of theoretical and practical problems [4] [5]. For instance, a solver might converge to a wrong answer or it might mislabel an infeasible problem as feasible.

This is where *Facial Reduction Algorithms (FRA)* comes into play. FRAs aim at reformulating (P) in such a way that strong feasibility is satisfied for (P). This is accomplished by finding a smaller part of $K$ that still contains all the feasible solutions of $(P)$. Then, a new but equivalent problem $(P^*)$ is obtained that has better theoretical properties and is more likely to be solved correctly by existing methods. Then, if necessary, we can also apply FRA to the corresponding dual problem of $(P^*)$ to fully regularize the problem.

## 2. Facial Reduction

Facial reduction was developed by Borwein and Wolkowicz [6] in the 80s for very general conic convex programs. However, it took some time before it became widely studied. It was in fact with the advent of CLPs and IPMs that researchers started to pay attention to it. Modern descriptions of the technique can be found in the work by Pataki [7] and also in the work by Waki and Muramatsu [8]. We will now present a brief overview of the technique.

Let $V = \{x \in \Re^n \mid Ax = b\}$. Note that $V$ is an *affine set*, which, as we recall, simply means that it is some subspace of $\Re^n$ translated by some vector. Note that (P) is not strongly feasible (i.e., (P) does not satisfy Slater's condition) if and only if $V \cap \mathrm{ri}(K) = \emptyset$. In this case, because $V$ is an affine set and K is, in particular, a convex set, by invoking one of the many separation theorems that exist [3], we can find some hyperplane H such that V and K belong to opposite half-spaces defined by H. Let us explain more precisely the meaning of that. First, since H is a hyperplane there are $d \in \Re^n$, $\alpha \in \Re$ such that $H = \{x \in \Re^n \mid x^T d = \alpha\}$. Then, H divides the whole space in two half-spaces $H^+ = \{x \in \Re^n \mid x^T d \geq \alpha\}$ and $H^- = \{x \in \Re^n \mid x^T d \leq \alpha\}$. The statement that V and K belong to opposite half-spaces is the same as saying that, for instance, $V \subseteq H^-, K \subseteq H^+$.

The statement that $V \subseteq H^-, K \subseteq H^+$ does not exclude the possibility that $V \subseteq H, K \subseteq H$, which is not very interesting since it does not give us much information about the relation between V and K. Still, due to some technicalities, we can in fact ensure that they are not both contained in H at the *same time.* Then, exploring the properties of V and K, we can always assume that $\alpha = 0$ and after some technical arguments we are left with two possibilities:

1. $V \subseteq H$, $K \subsetneq H$ and $d \in K^*$. (See Figure 1 below).
2. $V \cap H = \emptyset$ and $d \in K^*$, which implies that (P) is infeasible, since $V \subseteq H^-, K \subseteq H^+$

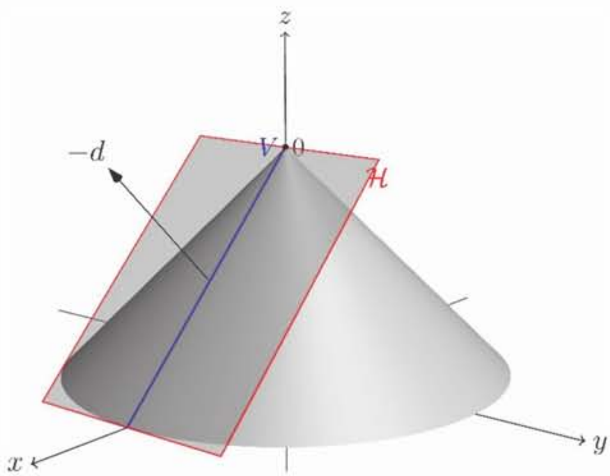


Figure 1 – In this picture we can see a piece of the cone $K$ (grey). In our context, a cone is an object that extend infinitely in the directions it contains, so this only shows a part of it. We see that V (blue) intersects $K$ at the boundary only, so there are no relative interior points. Because of that, we can find the hyperplane H, which contains V but not K. In this case, $K \subseteq H^+$.

If 2. holds, we stop. Note that if 1. holds, it is not necessarily the case (P) is feasible, we simply do not know this information at this point. Nevertheless, if 1. holds, we let $F = K \cap H$. Then, F has a few special properties:

1. F is a face of $K$.
2. F contains all feasible solutions of (P).
3. F is strictly smaller than $K$, that is $F \subsetneq K$. Not only that, but the dimension of F is smaller.

A *face* of $K$ is a convex cone F contained in $K$ with the property that $x, y \in K$ and $x + y \in F$ imply that $x, y \in F$. For many useful cones, we can describe its faces in a rather complete and comprehensive manner.

Property 2. ensures that if we substitute $K$ for F in (P) we will get an equivalent problem in the sense that all feasible solutions will stay the same. In addition, the optimal value and optimal solutions will not change. However, since F is smaller than $K$, the corresponding dual problem can gain new feasible solutions, since $F^*$ is potentially larger than $K^*$. So consider the following new problem

$$\begin{aligned} &\textbf{minimize } \mathbf{c}^{\mathrm{T}}\boldsymbol{x} \qquad (\boldsymbol{P_1}) \\ &\text{subject to } \mathrm{A}x = \mathrm{b} \\ &x \in F. \end{aligned}$$

If $(P_1)$ is strongly feasible, then we are done. Otherwise, by the same principle, we can find some hyperplane $H_2$ that separates F and V and either find out that that the problem is, in fact, infeasible or find a smaller face of F that contains all feasible solutions to (P).

This is the essence behind facial reduction. As long as the problem is *not* strongly infeasible we can keep replacing the cones by smaller and smaller faces. Property 3 listed above ensures that this process will eventually come to an end since the dimension of the faces is getting smaller and smaller.

In fact, it can be shown that this process ends at the so-called *minimal face* of (P), which is defined as the smallest face of $K$ that contains all the feasible solutions of (P). We denote the minimal face by $F_{min}^{P}$. We can then state a FRA algorithm as follows.

**Algorithm 1 - Facial Reduction**

**Input:** (P)

**Output:** $F_{min}^{P}$. ($F_{min}^{P} = \emptyset$ if (P) is infeasible)

1 If $K$ is strongly feasible, let $F_{min}^{P} \leftarrow K$ and stop.

2 If $K$ is not strongly infeasible let $H = \{x \in \Re^n \mid x^T d = 0\}$ be a hyperplane such that $V \subseteq H^-, K \subseteq H^+$ together with either:

A) $V \subseteq H$, $K \subsetneq H$, or

B) $V \cap H = \emptyset$,

3 If A) holds, we let $K \leftarrow K \cap H$ and return to 1. If B) holds, we let $F_{min}^{P} \leftarrow \emptyset$ and stop.

After applying Facial Reduction, we finally obtain the following regularized version of (P).

$$\begin{aligned} &\textbf{minimize } \mathbf{c}^{\mathrm{T}}\boldsymbol{x} \qquad (\boldsymbol{P^*}) \\ &\text{subject to } \mathrm{A}x = \mathrm{b} \\ &x \in F_{min}^{P}, \end{aligned}$$

If $F_{min}^{P}$ is not empty, it can be shown that the problem $(P^*)$ has the following properties: it is strongly feasible and given some $x \in \Re^n$ we have that $x$ is feasible for (P) if and only if it is feasible for $(P^*)$. In particular, we have $\theta_P = \theta_{P^*}$.

## 3. Worst case bounds.

The facial reduction algorithm (FRA) starts at the original cone $K$ and progresses to the minimal face $F_{min}^{P}$. More precisely, the algorithm produces a chain of faces $F_1 \supsetneq F_2 \supsetneq \ldots \supsetneq F_\ell$ such that $F_1 = K$ and $F_\ell = F_{min}^{P}$.

From a computational point of view, FRA is expensive because the hyperplanes in Step 2 must be found by solving auxiliary conic linear programs. Fortunately, these auxiliary CLPs have nice theoretical properties so they do not suffer from the same ill-behavior that (P) might have, no matter how unfavorable the theoretical properties of (P) are.

Still, if possible, we would like to solve as few auxiliary CLPs as possible since our major goal is to solve (P). So, usually, we measure the performance of FRA by the number of hyperplanes H found throughout the algorithm. The length of the chain of faces $F_1 \supsetneq F_2 \supsetneq \ldots \supsetneq F_\ell$ is defined to be $\ell$. With that, the number of hyperplanes found is $\ell - 1$.

Note that the hyperplane H in Step 2 is not unique so for fixed $K$, A, b, we have some degree of freedom in the choice of H. A good H that cuts deep into the boundary of K will produce a smaller chain of faces, which is desirable. A bad H that produces a shallow cut only enough to ensure $K \subsetneq H$ and nothing more is more likely to induce a long chain of faces.

We define the *singularity degree* of (P) as the minimal number of hyperplanes needed to ensure that $F_\ell = F_{min}^{P}$. The singularity degree of (P) is denoted by $d(P)$ and is a well-defined quantity.

It is natural to consider how large can $d(P)$ be. In fact, it is possible to show that there is a bound for $d(P)$ that does not depend on A or b. Namely, we always have

$$d(P) \leq \ell_K - 1, \qquad (1)$$

where $\ell_K$ denotes the *longest chain strict descending chain of faces of K.*

As $K$ is contained in the finite dimensional space $\Re^n$, it also has finite dimension. In addition, if $F \subsetneq K$ is a face of $K$ then it must have dimension strictly smaller than $K$. Therefore, $\ell_K$ is finite, which tells us, in particular, that there is no risk of Facial Reduction running forever.

Recently, we noticed that the bound (1) is not very tight. That is, for fixed $K$, the difference between the worst possible d(P) and the quantity $\ell_K - 1$ can be very large.

In [9] we proved a better bound for d(P), which we now describe briefly. The motivation for it came from observing what happens when the cone $K$ is *polyhedral.* We recall that a set is said to be polyhedral if it can be described as the set of solutions of a finite number of linear equations and linear inequalities. In particular, all affine sets are polyhedral but the converse does not hold.

When $K$ is polyhedral, although it can have a very long chain of faces, it is possible to show that d(P) is at most one. This stems from the Goldman-Tucker Theorem for Linear Programming, which asserts the existence of the so-called *strict complementary optimal solutions* for LPs.

So the first observation is that when we are doing Facial Reduction as soon as we reach some polyhedral face, we can jump straight to $F^P_{min}$. This gives the bound

$$d(P) \leq 1 + \ell_{poly}(K), \qquad (2)$$

where $\ell_{poly}(K)$ is the length *minus one* of the longest chain of faces of $K$ that starts with $K$ and descends to a polyhedral face in such a way that all intermediate faces are nonpolyhedral. We call $\ell_{poly}(K)$ the *distance to polyhedrality* of $K$.

The inequality (2) is already an improvement over (1), but it is possible to be sharper. For many problems, the cone $K$ is actually a direct product of other cones. That is,

$$K = K^1 \times \ldots \times K^r, \qquad (3)$$

where the $K^i$ are themselves closed convex cones. In this case, all faces of $K$ are also direct products of faces of the $K^i$. That is, if F is a face of $K$ then

$$F = F^1 \times \ldots \times F^r \qquad (3)$$

where for all i, we have that $F^i$ is a face of $K^i$. Now, suppose that F is a face obtained in the intermediate steps of Algorithm 1. Under these circumstances, we proved in [9] that the following condition is enough to jump to $F^P_{min}$. If for every i, we have either

A) $F^i = (F^P_{min})^i$ or

B) $F^i$ is polyhedral,

then there exists a hyperplane H as in Step 2 of Algorithm 1 such that $F^P_{min} = F \cap H$. In other words, if an intermediate face is such that either a block is polyhedral or is already a part of the minimal face, then it is possible to jump to the minimal face $F^P_{min}$ in a single step. In the end, we get the following bound

$$d(P) \leq 1 + \sum_{i=1}^{r} \ell_{poly}(K^i). \qquad (4)$$

Also in [9] we show that the quantity $1 + \sum_{i=1}^{r} \ell_{poly}(K^i)$ is strictly smaller than $\ell_K - 1$ if we have the product of at least two cones that are not subspaces. In other words, not only the bound (4) is not worse than (1) but in most practical cases, it is strictly better.

In fact, as far as we know, the bound in (4) is the best general bound available depending only on $K$. But, of course, there is no guarantee that the Algorithm 1, as we stated in this paper, will not end up finding more directions before reaching the minimal face. To account for that, in [9] we show how to design a facial reduction algorithm that is guaranteed to not perform worse than the bound in (4). This includes a detailed on discussion on how to find good hyperplanes through auxiliary conic linear programs.

## 4. Conclusion

In this paper, we presented an overview of facial reduction together with a discussion of worst case bounds. There are many interesting topics connected to facial reduction that we did not mention. For instance, Facial Reduction can be used to give generalized versions of the classical Farkas' Lemma that appears in Linear Programming [10] [11]. It is also a tool for obtaining the so-called "extended duals" which are substitutes for (D) that do not suffer from the same theoretical issues [12] [13] [7].

Moreover, FRAs can be used to study several different types of ill-behavior in CLPs such as weak infeasibility and unattained optimal values [14] [15] [16] [17]. There are also recent works discussing practical issues in the the implementation of facial reduction software [18] [19].

## References

[1] M. J. Todd, "Semidefinite optimization," *Acta Numerica,* Vol. 10, pp. 515-560, May 2001.

[2] Y. Nesterov , A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming, Society for Industrial and Applied Mathematics, 1994.

[3] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1997.

[4] H. Waki, M. Nakata , M. Muramatsu, "Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization," *Computational Optimization and Applications,* Vol. 53, pp. 823-844, December 2012.

[5] H. Waki, "How to generate weakly infeasible semidefinite programs via Lasserre's relaxations for polynomial optimization," *Optimization Letters,* Vol. 6, pp. 1883-1896, December 2012.

[6] J. M. Borwein , H. Wolkowicz, "Regularizing the abstract convex program," *Journal of Mathematical Analysis and Applications* , Vol. 83, pp. 495-530, 1981.

[7] G. Pataki, "Strong Duality in Conic Linear Programming: Facial Reduction and Extended Duals,"in: *Computational and Analytical Mathematics*, Vol. 50, Springer New York, 2013, pp. 613-634.

[8] H. Waki , M. Muramatsu, "Facial Reduction Algorithms for Conic Optimization Problems," *Journal of Optimization Theory and Applications,* Vol. 158, pp. 188-215, 2013.

[9] B. F. Lourenço, M. Muramatsu , T. Tsuchiya, "Facial Reduction and Partial Polyhedrality," *arXiv e-prints,* December 2015.

[10] M. Liu , G. Pataki, "Exact Duality in Semidefinite Programming Based on Elementary Reformulations," *SIAM Journal on Optimization,* Vol. 25, pp. 1441-1454, 2015.

[11] M. Liu , G. Pataki, "Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming," 2015.

[12] M. V. Ramana, "An Exact duality Theory for Semidefinite Programming and its Complexity Implications," *Mathematical Programming,* Vol. 77, 1995.

[13] M. V. Ramana, L. Tunçel , H. Wolkowicz, "Strong Duality for Semidefinite Programming," *SIAM Journal on Optimization,* Vol. 7, pp. 641-662, #aug# 1997.

[14] F. Permenter, H. A. Friberg , E. D. Andersen, "Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach," *Optimization Online,* Sep 2015.

[15] B. F. Lourenço, M. Muramatsu , T. Tsuchiya, "A structural geometrical analysis of weakly infeasible SDPs," *Journal of the Operations Research Society of Japan,* Vol. 59, pp. 241-257, 2016.

[16] B. F. Lourenço, M. Muramatsu , T. Tsuchiya, "Weak Infeasibility in Second Order Cone Programming," *To Appear in Optimization Letters,* 2015.

[17] B. F. Lourenço, M. Muramatsu , T. Tsuchiya, "Solving SDP Completely with an Interior Point Oracle," *arXiv e-prints,* Jul 2015.

[18] Y.-L. Cheung, S. Schurr , H. Wolkowicz, "Preprocessing and Regularization for Degenerate Semidefinite Programs," in: *Computational and Analytical Mathematics*, Vol. 50, Springer New York, 2013, pp. 251-303.

[19] F. Permenter , P. Parrilo, "Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone," *ArXiv e-prints,* 2014.