Noname manuscript No.
(will be inserted by the editor)

An improved projection and rescaling algorithm for conic feasibility problems

Wei Zhang · Kees Roos · Yanqin Bai

Received: date / Accepted: date

Abstract Motivated by Chubanov's projection-based method for linear feasibility problems [13], a projection and rescaling algorithm for the conic feasibility problem

$$find\ x \in L \bigcap \Omega$$

is proposed in [1], where L and Ω are respectively a linear subspace and the interior of a symmetric cone in a finitely dimensional vector space V. When V is the Euclidean space R^n, L is the null space of some matrix $A \in R^{m\times n}$ and Ω is R^n_{++}, the problem reduces to the linear feasibility problem. In their method for the general case, the condition measure $\delta(L \bigcap \Omega)$ is adopted to analyze the iteration complexity. In this paper, we utilize another condition measure $\delta_\infty(L \bigcap \Omega)$, which is based on the L_∞ norm. Besides, in the basic procedure, the stopping criterion is determined by the projection onto the space $L^\perp$ instead of L. Based on this, we can reduce the number of iterations for a von Neumann type basic procedure and a smoothed perceptron basic procedure by $O(r)$ and $O(r^{\frac{1}{2}})$ respectively, where r is the Jordan algebra rank of V. Furthermore, the prox-mirror method is also customized as a basic procedure, which can achieve the same iteration complexity of the smoothed perceptron method. Moreover, by carefully examining the fast and slow iterations, we can further reduce the total iteration complexity with the von Neumann type basic procedure. Similarly to the algorithm in [11], we can alternatively stop the algorithm when the possible solutions are near the boundary of the cone. The multi-block case is also considered.

Wei Zhang
National University of Singapore
Tel.: +65-93528088
E-mail: lindelfeel@gmail.com

Kees Roos
Delft University of Technology

Yanqin Bai
Shanghai University

Keywords Projection and Rescaling · Symmetric cones · Condition measure · Conic systems

Mathematics Subject Classification (2010) 90C25 · 90C60 · 52A20 · 65F35

1 Introduction

We consider the feasibility problem

$$\text{find } x \in L \bigcap \Omega, \tag{1}$$

where L and Ω are respectively a linear subspace and the interior of a symmetric cone in a finitely dimensional vector space V. When V is the Euclidean space R^n, L is the null space of some matrix $A \in R^{m\times n}$ and Ω is R^n_{++}, the problem reduces to the linear feasibility problem

$$\begin{aligned} &\text{find } \boldsymbol{x} \\ &s.t. A\boldsymbol{x} = \boldsymbol{0}, \\ &\boldsymbol{x} > \boldsymbol{0}. \end{aligned} \tag{2}$$

The traditional algorithms for solving problem (2) include the perceptron algorithm, the von Neumann algorithm and their variations. However, a drawback of them is that, only when the problem is strictly infeasible, the perceptron algorithm can stop in finite iterations [5], while the von Neumann algorithm can only stop in finite iterations if the problem is feasible [6, 7].

In [2][3][4] and [8], the perceptron algorithm is combined with some kind of rescaling. The perceptron algorithm is used as the basic procedure, while the rescaling stretches the matrix A along the direction of the output vector of the basic procedure. The progress of the rescaling is measured by the roundness of the feasible region with the condition number defined in [9]:

$$\rho(A) = \max_{\boldsymbol{y}\in im(A)\setminus \boldsymbol{0}} \min_{j\in[n]} \langle \hat{\boldsymbol{a}}_j, \hat{\boldsymbol{y}} \rangle,$$

where $\hat{\boldsymbol{y}} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$.

Problem (1) is equivalent to the problem

$$\text{find } y \in V \textit{ such that } P_L y \in \Omega, \tag{3}$$

where P_L denotes the orthogonal projection onto the linear subspace L. Denote the orthocomplement space of L as $M = L^\perp$ and P_M the orthogonal projection onto M. As $I = P_L + P_M$, $P_L P_M = 0$, we have $y = z + v$, where $z = P_L y$, $v = P_M y$, $\langle z, v \rangle = 0$. Then problem (1) has no solution if and only if the set

$$M \bigcap \overline{\Omega} \tag{4}$$

contains a nonzero element, which is the dual problem of (1).

The following lemma concludes an important role of this relation.

Lemma 1 *If $z \succ 0$ then z solves the primal problem (1) and if $0 \neq v \succeq 0$ then v solves the dual problem.*

As a result, in [13] the basic procedure is to find a solution of the problem

$$\boldsymbol{y} \in R^n, \ P_A \boldsymbol{y} > \boldsymbol{0} \tag{5}$$

instead of problem (2), where $P_A = I - A^\top (AA^\top)^{-1} A$ is the projection onto the null space of A. The von Neumann algorithm is used as the basic procedure to find a positive vector in $im(P_A)$, while the main procedure rescales the null space of A after a certain criterion is met by the output of the basic procedure. The rescaling in [13] is to shorten some column of the matrix A, meaning that the corresponding coordinate of the null space is stretched by a constant factor, which is different from that in [2].

According to a classical result of [10], the entries of some solution of the linear system (2) can be bounded below by a positive number dependent on the matrix A, if the largest entry is equal to 1. After a certain amount of rescalings, if no solution is found, we can conclude that problem (2) has no solution actually. As a result, by the algorithm of Chubanov, we can stop in a finite number of iterations no matter problem (2) is feasible or infeasible. Following Chubanov's algorithm, Roos [14] has demonstrated a new stopping criterion of the basic procedure and established an improved algorithm based on it. Meanwhile, different kinds of basic procedures have been adopted in [1], resulting in different complexity bounds.

Besides, this kind of projection and rescaling method is also generalized to the case of conic programming in [1], with the progress measured by a kind of condition measure $\delta(L \bigcap \Omega)$. In [11], the conic feasibility problem is also considered, where the rescaling is based on a half space containing the null space. After a certain number of rescalings, it can be stated that for every solution of problem (1), the ratio of the minimal eigenvalue to the maximal eigenvalue is smaller than a given bound.

In this paper, based on the projection and rescaling operations, we improve the algorithm in [1] by adopting another condition measure, another stopping criteria of the basic procedure and an analysis on the fast and slow iterations. On the other hand, similarly to [11], we can also state that after a certain number of rescalings, for every solution of (1), if it does exist, the minimal eigenvalue is smaller than ϵ, given that the maximal eigenvalue is not larger than 1.

2 Preliminaries

V is a linear space endowed with a bilinear operation $\circ$: $V \times V \to V$ and $e \in V$ is a particular element of V. The triple $(V, \circ, e)$ is an *Euclidean Jordan algebra* with identity element e if the following conditions hold:

- $x \circ y = y \circ x$ for all $x, y \in V$
- $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$ for all $x, y \in V$, where $x^2 = x \circ x$

– $x \circ e = x$ for all $x \in V$
– There exists an associative positive definite bilinear form on V.

An element $c \in V$ is idempotent if $c^2 = c$. An idempotent element of V is a primitive idempotent if it is not the sum of two other idempotents. The rank r of V is the smallest integer such that for all $x \in V$ the set $\{e, x, x^2, \dots, x^r\}$ is linearly dependent.

Every element $x \in V$ has a spectral decomposition

$$x = \sum_{i=1}^{r} \lambda_i(x) c_i, \tag{6}$$

where $\lambda_i(x) \in \mathbb{R}$ $(i = 1, \dots, r)$ are the eigenvalues of x and $\{c_1, \dots, c_r\}$ is a *Jordan Frame*, that is, a collection of non-zero primitive idempotents such that $c_i \circ c_j = 0$ for $i \neq j$, and $c_1 + \cdots + c_r = e$. Given the spectral decomposition (6), we have $x \circ c_i = \lambda_i(x) c_i$, $i = 1, \dots, r$.

The trace and determinant of $x \in V$ are respectively defined as $trace(x) = \sum_{i=1}^{r} \lambda_i(x)$ and $det(x) = \Pi_{i=1}^{r} \lambda_i(x)$. In this paper we assume that $(V, \circ, e)$ is an *Euclidean Jordan algebra* with identity and V is endowed with the following trace inner product:

$$\langle x, y \rangle \doteq trace(x \circ y). \tag{7}$$

We also assume that Ω is the interior of the cone of squares in V, that is, $\Omega = int(\{x^2 : x \in V\})$.

The Frobenius norm is $\|x\|_F \doteq \sqrt{\langle x, x \rangle} = \|\lambda(x)\|_2$, where $\lambda(x) \in \mathbb{R}^r$ denotes the vector of eigenvalues of $x \in V$. The operator norm $\|x\| = \|\lambda(x)\|_\infty$ is similar to the infinity norm of vectors in the Euclidean space. Let P_L be the projection mapping onto L relative to the inner product defined in (7).

Consider the following *condition measure* of the set $L \bigcap \Omega$:

$$\delta_\infty(L \bigcap \Omega) \doteq \max_x \{det(x) : x \in L \bigcap \Omega, \|\lambda(x)\|_\infty = 1\}.$$

$L \bigcap \Omega \neq \emptyset$ implies $\delta_\infty(L \bigcap \Omega) \in (0, 1]$, with the equality holding exactly when $e \in L \bigcap \Omega$. This condition measure is in the same spirit as that in [13] and [14] when $V = R^n$, but different from the one in [1]:

$$\delta(L \bigcap \Omega) \doteq \max_x \{det(x) : x \in L \bigcap \Omega, \|x\|_F^2 = r\}.$$

Generally $\delta_\infty(L \bigcap \Omega) \leq \delta(L \bigcap \Omega)$. Later we can see introducing $\delta_\infty(L \bigcap \Omega)$ helps to reduce the complexity of the basic procedure.

Assuming $c \in V$ and $a > 0$ is a positive constant, let $L_v : V \to V$ be the linear mapping associated to $v = e + ac$, that is,

$$L_v x = v \circ x = x + ac \circ x.$$

In [1], as in Chubanov's algorithm [13], the basic procedure can be seen as a systematic search method for a nonzero vector $y \in L \bigcap \Omega$. The clever finding is that a rescaling procedure of the set $L \bigcap \Omega$ can be implemented when a vector y has been found such that

$$\|(P_L y)^+\|_F \leq \frac{1}{4r}\|y\|,$$

where r is the rank of the symmetric cone Ω, and x^+ denotes the vector arising from x by replacing all the negative $\lambda_i(x)$ by zero, or formally

$$x^+ = \sum_{i=1}^{r}(\max\{\lambda_i(x), 0\})c_i.$$

Similarly we denote x^-, so that $x^- = -(-x)^+$. In this paper, we adopt another criterion for the rescaling procedure, thus accelerating the basic procedures in [1]. By an analysis of the fast and slow iterations, we can furthermore reduce the total complexity of the algorithm. In section 3, the stopping criterion for a basic procedure is developed based on the projection onto M. Besides, a smoothed perceptron-von Neumann type, a primal-dual prox-mirror type and a modified von Neumann type basic procedure are raised respectively. In section 4, two kinds of main procedures are designed respectively for these three basic procedures. Finally, the iteration complexity is analyzed in section 5, where the complexity with the modified von Neumann type basic procedure can be reduced by exploring the fast and slow iterations carefully. Moreover, our algorithm can be extended to the multi block case and compared with the one designed in [11].

3 Basic Procedure

Given $\Omega = int(\{x^2 : x \in V\})$, the primal problem (1) is feasible if and only if the system

$$x \in L \bigcap \Omega, \lambda(x) \in (0, 1]^r \tag{8}$$

is feasible. As a result, we focus on this problem instead.

In the basic procedure, we try to find a proper $z \in L \bigcap \Omega$ by iterating the vector $y \in V$, with $z = P_L y$ and $v = P_M y = y - z$. The next lemma shows that for every solution x of (8), the value of $\lambda_k(x)$ is actually related to the vector $\lambda(v)$, which is a generalization of Lemma 4.1 in [14]:

Lemma 2 *Let x be feasible for (8), and $v = P_M y$ for some $y \in V$. Then every nonzero element $\lambda_k(v)$ gives rise to an upper bound for $\lambda_k(x)$, according to the relation*

$$\lambda_k(x) \leq \langle e, [\frac{v}{-\lambda_k(v)}]^+\rangle.$$

Proof Since $\langle x, v\rangle = 0$, $0 < \lambda_j(x) \leq 1, \forall j$, when $\lambda_k(v) < 0$, we have

$$-\lambda_k(v)\lambda_k(x) = \sum_{i \neq k}\lambda_i(v)\lambda_i(x) \leq \sum_{i,\lambda_i(v)>0}\lambda_i(v)\lambda_i(x) \leq \sum_{i,\lambda_i(v)>0}\lambda_i(v) = \langle e, v^+\rangle. \tag{9}$$

On the other hand, when $\lambda_k(v) > 0$, we have

$$\lambda_k(v)\lambda_k(x) = -\sum_{i \neq k} \lambda_i(v)\lambda_i(x) \leq -\sum_{i,\lambda_i(v)<0} \lambda_i(v)\lambda_i(x) \leq \sum_{i,\lambda_i(v)<0} -\lambda_i(v) = -\langle e, v^- \rangle. \tag{10}$$

The result can be implied directly from these two inequalities.

Unlike the cutting criterion in [1], this lemma builds up a relationship between a feasible solution in $L \bigcap \Omega$ and the vectors in the orthogonal subspace $M = L^\perp$. Adopting this criterion, we are able to reduce the time complexity of the basic procedure.

Denoting

$$bound_j(y) = \begin{cases} \frac{-\langle e, v^- \rangle}{\lambda_j(v)}, \ \lambda_j(v) > 0, \\ \frac{\langle e, v^+ \rangle}{-\lambda_j(v)}, \ \lambda_j(v) < 0, \end{cases}$$

and

$$bound(y) = \begin{cases} \frac{-\langle e, v^- \rangle}{\max(\lambda(v)^+)}, \ \langle e, v \rangle > 0, \\ \frac{\langle e, v^+ \rangle}{-\min(\lambda(v)^-)}, \ \langle e, v \rangle < 0, \end{cases}$$

we call y a σ-cutting vector if $bound(y) \leq \frac{1}{\sigma}$. This implies there exists some index k, such that $x_k \leq bound(y) \leq \frac{1}{\sigma}$, where x is any feasible solution of Problem (8). Actually, the vector v here does not need to be related to the vector y, but only needs to lie in the subspace $M = L^\perp$. However, as we are iterating a vector $y \in V$, the relationship between y and v helps us to determine the stopping criterion easily.

Define the *spectraplex* $\Delta(\Omega)$ as

$$\Delta(\Omega) = \{x \in \overline{\Omega} : \langle e, x \rangle = 1\}. \tag{11}$$

Actually, when V is the Euclidean space and $\overline{\Omega}$ denotes the nonnegative orthogonal cone, this reduces to the probability simplex.

Given $v \in V$, assume that it is induced by a non σ-cutting vector y satisfying $bound(y) > \frac{1}{\sigma}$. To bound the iteration complexity of the basic procedure, we need to find how small the value of $\|z\|_F$ can achieve under this condition. Consider

$$\min_{y,z,\beta} \{\|z\|_F : y \in \Delta(\Omega), y = z + \beta v, \langle z, v \rangle = 0\}. \tag{12}$$

When $\beta \neq 0$, $v \in V$ if and only if $\beta v \in V$. And by definition, the factor β will not change the value of $bound(y)$.

Lemma 3 *Let $r \geq 2$, $\sigma \geq 2$ and $v \in V$. If y and z satisfy (12), and y is not σ-cutting, then*

$$\frac{1}{\|z\|^2} \leq \frac{r^3\sigma^2}{5}.$$

Proof Consider the Lagrange dual of Problem (12):

$$\max_{\alpha,\lambda}\{\alpha : \lambda \succeq \alpha e, \|\lambda\|_F \leq 1, \langle \lambda, v\rangle = 0\}. \tag{13}$$

By the weak duality property, if (y, z, β) and (λ, α) are respectively feasible for the problems (12) and (13), we have $\|z\|_F \geq \alpha$.

For $v \in V$, define the index sets S and T as :

$$S = \{i : \lambda_i(v) \geq 0\},\ T = \{i : \lambda_i(v) < 0\}.$$

With $\tau \geq \alpha \geq 0$, consider the vector $\lambda \in \Omega$ defined by $\lambda = \alpha e_S + \tau e_T = \sum_{i\in S} \alpha c_i + \sum_{i\in T} \tau c_i$, where $v_S = \sum_{i\in S}\langle v, c_i\rangle c_i$ and $v_T = \sum_{j\in T}\langle v, c_j\rangle c_j$. Then $\lambda \succeq \alpha e$ is feasible for (13) with objective value α if $\langle \lambda, v\rangle = 0$ and $\|\lambda\|_F = 1$. In this case, α and τ should satisfy the equations

$$\alpha\langle e_S, v_S\rangle + \tau\langle e_T, v_T\rangle = 0, \tag{14}$$
$$\alpha^2|S| + \tau^2|T| = 1. \tag{15}$$
$$\tag{16}$$

The above system determines the relation between τ and α as follows:

$$\tau = \frac{\alpha\langle e_S, v_S\rangle}{-\langle e_T, v_T\rangle},\ \frac{1}{\alpha^2} = |S| + \frac{\langle e_S, v_S\rangle^2}{\langle e_T, v_T\rangle^2}|T|. \tag{17}$$

As $v_T = v^-$ and $bound(y) \geq \frac{1}{\sigma}$, implying

$$\langle e_S, v_S\rangle \leq |S| \max(\lambda(v)) \leq -|S|\sigma\langle e_T, v^-\rangle = |S|\sigma\langle e_T, -v_T\rangle,$$

we can obtain

$$\frac{1}{\alpha^2} \leq |S| + \sigma^2|S|^2|T|.$$

As $|S| + |T| = r$, it reduces to determine the maximal value of the function

$$f(s) = s(1 + \sigma^2 s(r - s)), 1 \leq s \leq r.$$

By a simple calculation we can obtain that

$$\frac{1}{\alpha^2} \leq \frac{r^3\sigma^2}{5} \Rightarrow \frac{1}{\|z\|_F^2} \leq \frac{r^3\sigma^2}{5}.$$

In this paper, we mainly consider the case $\sigma = 2$, where it reduces to

$$\frac{1}{\|z\|_F^2} \leq \frac{4r^3}{5}.$$

We can further show that this stopping criterion based on $bound(y)$ is stronger than the one used in [1], although we cannot prove any complexity reduction. Fixing k, consider the problem

$$\max\{\lambda_k(x) : x \in L \bigcap \overline{\Omega}, \|x\|_\infty \leq 1\}.$$

Noticing that the objective is just $\langle c_k, x\rangle$, the dual problem can be obtained:

$$\min\{\langle e, w\rangle : u + w \succeq c_k, u \in L^{\perp}, w \in \overline{\Omega}\} = \min\{\langle e, [c_k - u]^{+}\rangle : u \in L^{\perp}\}.$$

Choosing $u = \alpha v$, by the duality theorem we have

$$\lambda_k(x) \leq \langle e, [c_k - \alpha v]^{+}\rangle = [1 - \alpha\lambda_k(v)]^{+} + \sum_{i\neq k}[-\alpha\lambda_i(v)]^{+}, \forall\alpha \in \mathbb{R}.$$

Different values of α induce different upper bounds for $\lambda_k(x)$. When $\alpha = \frac{1}{\lambda_k(y)}$ for $\lambda_k(y) > 0$, one has

$$[c_k - \alpha v]^{+} = [c_k - \frac{v}{\lambda_k(y)}]^{+} = [c_k - \frac{y - z}{\lambda_k(y)}]^{+} \preccurlyeq [\frac{z}{\lambda_k(y)}]^{+}$$

as $c_k - \frac{y}{\lambda_k(y)} \preceq 0$. In this case we obtain

$$\langle e, [c_k - \frac{v}{\lambda_k(y)}]^{+}\rangle \leq \langle e, [\frac{z}{\lambda_k(y)}]^{+}\rangle = \frac{\langle e, z^{+}\rangle}{\lambda_k(y)} \leq \frac{\sqrt{r}\|z^{+}\|_F}{\lambda_k(y)} \leq \frac{\sqrt{r}\|z\|_F}{\lambda_k(y)}.$$

On the other hand, the minimal value of the piecewise linear function $[1 - \alpha\lambda_k(v)]^{+} + \sum_{i\neq k}[-\alpha\lambda_i(v)]^{+}$ is $\min(1, \sum_{i=1}^{r}[\frac{-\lambda_i(v)}{\lambda_k(v)}]^{+})$. As a result,

$$\min(1, \sum_{i=1}^{r}[\frac{-\lambda_i(v)}{\lambda_k(v)}]^{+}) \leq \frac{\sqrt{r}\|z\|_F}{\lambda_k(y)}.$$

When $\sum_{i=1}^{r}[\frac{-\lambda_i(v)}{\lambda_k(v)}]^{+} > \frac{1}{\sigma}$, we always have

$$\lambda_k(y) < \sigma\sqrt{r}\|z\|_F. \tag{18}$$

That is, when our criterion for the cut is not satisfied, the criterion used in [1] is also not satisfied.

In Lemma 3, it is shown that before reaching a σ-cutting vector y, the value of $\frac{1}{\|P_L y\|_F}$ cannot be greater than $\frac{r^3\sigma^2}{5}$. So after a certain number of iterations, we shall be able to rescale the initial problem. In what follows, the systematic search method for the vector y are stated. In [1], four types of basic procedures are listed. Here we only discuss two of them, a modified von Neumann algorithm and the smooth perceptron-von Neumann algorithm, as they induce two different iteration complexities. Moreover, the primal-dual prox mirror algorithm is also customized as a basic procedure, which induces the same iteration complexity as the smooth perceptron-von Neumann algorithm.

Algorithm 1 $[y, u, z, J, case]$ = Modified smooth perceptron algorithm(P_L, y, σ)

initialize: $t = 0, u_0 = \overline{u} = \frac{1}{r}e, \overline{y} = 0, \mu_0 = 2, case = 0, y_0 = y_{\mu_0}(P_L u_0)$,
while $bound(y) > \frac{1}{\sigma}$ and case=0 **do**
  if $P_L u_t \succ 0$ **then**
    case=1 (u is primal feasible); return
  else
    if $v = P_M u \succeq 0$ and $v \neq 0$ **then**
      case=2 (u is dual feasible); return
    else
      $\overline{y} = y_t$
      $\theta_t = \frac{2}{t+3}$
      $u_{t+1} = (1-\theta_t)(u_t + \theta_t y_t) + \theta_t^2 y_{\mu_t}(P_L u_t)$
      $\mu_{t+1} = (1-\theta_t)\mu_t$
      $y_{t+1} = (1-\theta_t)y_t + \theta_t y_{\mu_{t+1}}(P_L u_{t+1})$
      $t = t+1$
    end if
  end if
end while
if case=0 **then**
  find a nonempty set J such that $J \subseteq \{j : bound_j(y) \leq \frac{1}{\sigma}\}$
end if

3.1 The Smooth Perceptron-Von Neumann Algorithm

First we shall have a look at the smooth perceptron-von Neumann algorithm [15]. Defining a mapping based on the fixed $\overline{u} = \frac{1}{r}e$ as

$$y_\mu : V \to \Delta(\Omega)$$

$$v \mapsto \arg\min_{u\in\Delta(\Omega)} \{\langle u, v\rangle + \frac{\mu}{2}\|u - \overline{u}\|_F^2\},$$

the smooth perceptron-von Neumann algorithm is described in Algorithm 1.

As we know, the aim of the basic procedure is to find a vector y such that one of the following three cases occurs:

1. $z = P_L y$ is feasible for (8);
2. $v = y - z \succcurlyeq 0$, meaning that v satisfies (4);
3. y is a σ-cutting vector.

As case 1 holds if and only if $\langle P_L y, u\rangle > 0, \forall u \in \Omega$, it suffices to show that

$$\psi(y) = \min_{u\in\Delta(\Omega)} \langle P_L y, u\rangle > 0$$

to make sure the vector $P_L y$ is feasible for (8). Denoting

$$\varphi(y) = -\frac{1}{2}\|P_L y\|_F^2 + \min_{u\in\Delta(\Omega)} \langle P_L y, u\rangle = -\frac{1}{2}\|P_L y\|_F^2 + \psi(y),$$

as $\psi(y) \geq \varphi(y)$, we have $\psi(y) > 0$ given $\varphi(y) > 0$. On the other hand, if $\psi(y) < 0$, then $\varphi(y) < 0$. The reason that we construct such a function $\varphi(y)$ is that the excessive gap technique of Nesterov [12] can be adopted to maximize $\varphi(y)$.

In [12], a specific optimization problem $\min_{x\in Q_1} f(x)$ is considered, with a structural objective function

$$f(x) = \hat{f}(x) + \max_{u\in Q_2}\{\langle Ax, u\rangle - \hat{\phi}(u)\},$$

while the adjoint problem of which is $max_{u\in Q_2}\phi(u)$,

$$\phi(u) = -\hat{\phi}(u) + \min_{x\in Q_1}\{\langle Ax, u\rangle + \hat{f}(x)\}.$$

Here $\hat{f}(x)$ is a continuous convex function on Q_1, $\hat{\phi}(x)$ is a continuous convex function on Q_2, Q_1 and Q_2 both bounded closed convex sets in finite-dimensional real vector spaces, A a linear transformation. The classical subgradient method ensures a convergence rate of order $O(\frac{1}{\sqrt{k}})$, where k is the number of iterations. However, introducing a pair of smoothed functions

$$f_{\mu_2}(x) = \hat{f}(x) + \max_{u\in Q_2}\{\langle Ax, u\rangle - \hat{\phi}(u) - \mu_2 d_2(u)\}$$

and

$$\phi_{\mu_1}(u) = -\hat{\phi}(u) + \min_{x\in Q_1}\{\langle Ax, u\rangle + \hat{f}(x) + \mu_1 d_1(x)\},$$

with a pair of smoothness parameters μ_1, μ_2 and continuous strongly convex functions $d_1(x), d_2(u)$, the problem can be solved by a primal dual symmetric technique ensuring a convergence rate $O(\frac{1}{k})$. The idea is to iteratively approximate both the primal and dual problem by a family of parameters μ_1, μ_2. As the gap between the two approximation functions grows smaller, the distance to the optimum will also grow smaller.

In our problem, define the approximating function as

$$\varphi_\mu(y) = -\frac{1}{2}\|P_L y\|_F^2 + \min_{u\in\Delta(\Omega)}\{\langle u, P_L y\rangle + \frac{\mu}{2}\|u - \overline{u}\|_F^2\}.$$

This is a smooth function, i.e., with a Lipschitz continuous gradient. A basic property is that the approximating function is not so far away from the original function.

Lemma 4 $0 \le \varphi_\mu(y) - \varphi(y) \le \mu$.

Proof As

$$0 \le \min_{u\in\Delta(\Omega)}\{\langle u, P_L y\rangle + \frac{\mu}{2}\|u-\overline{u}\|_F^2\} - \min_{u\in\Delta(\Omega)}\langle u, P_L y\rangle \le \max_{u\in\Delta(\Omega)}\frac{\mu}{2}\|u-\overline{u}\|_F^2 \le \mu,$$

we have $0 \le \varphi_\mu(y) - \varphi(y) \le \mu$.

The main property is the following.

Lemma 5 $\frac{1}{2}\|P_L y_t\|_F^2 \le \varphi_{\mu_t}(u_t)$.

Proof We prove this by induction. First, when $t = 0$, by the definition of y_0,

$$
\begin{aligned}
\frac{1}{2}\|P_L y_0\|_F^2 &= \frac{1}{2}\|P_L(y_0 - \overline{u}) + P_L\overline{u}\|_F^2 \\
&= \frac{1}{2}\|P_L(y_0 - \overline{u})\|_F^2 + \frac{1}{2}\|P_L\overline{u}\|_F^2 + \langle P_L\overline{u}, P_L(y_0 - \overline{u})\rangle \\
&= -\frac{1}{2}\|P_L\overline{u}\|_F^2 + \langle P_L\overline{u}, P_L y_0\rangle + \frac{1}{2}\|P_L(y_0 - \overline{u})\|_F^2 \\
&\leq -\frac{1}{2}\|P_L\overline{u}\|_F^2 + \langle P_L\overline{u}, y_0\rangle + \|y_0 - \overline{u}\|_F^2 \\
&= \varphi_{\mu_0}(\overline{u}) = \varphi_{\mu_0}(u_0).
\end{aligned}
\tag{19}
$$

Then, assuming that the inequality holds when $t \leq k$, we should show that it also holds when $t = k+1$. For short, denote $y = y_t, \mu = \mu_t, u = u_t, y_+ = y_{t+1}, \mu_+ = \mu_{t+1}, u_+ = u_{t+1}, \theta = \theta_t$, and

$$\hat{y} = (1-\theta)y + \theta y_\mu(P_L u).$$

Thus $u_+ = (1-\theta)u + \theta\hat{y}$, $y_+ = (1-\theta)y + \theta y_{\mu_+}(P_L u_+)$, and

$$y_+ - \hat{y} = \theta(y_{\mu_+}(P_L u_+) - y_\mu(P_L u)).$$

For $\varphi_{\mu_+}(u_+)$, as the function $\|\cdot\|_F^2$ is convex, we have

$$
\begin{aligned}
\varphi_{\mu_+}(u_+) &= -\frac{1}{2}\|P_L u_+\|_F^2 + \min_{u\in\Delta(\Omega)}\{\langle u, P_L u_+\rangle + \frac{\mu_+}{2}\|u - \overline{u}\|_F^2\} \\
&= -\frac{1}{2}\|(1-\theta)P_L u + \theta P_L\hat{y}\|_F^2 + \frac{\mu_+}{2}\|y_{\mu_+}(P_L u_+) - \overline{u}\|_F^2 + \langle y_{\mu_+}(P_L u_+), P_L u_+\rangle \\
&\geq -\frac{1}{2}(1-\theta)\|P_L u\|_F^2 - \frac{1}{2}\theta\|P_L\hat{y}\|_F^2 + \theta\langle y_{\mu_+}(P_L u_+), P_L\hat{y}\rangle + (1-\theta)\langle y_{\mu_+}(P_L u_+), P_L u\rangle \\
&\quad + \frac{\mu}{2}(1-\theta)\|y_{\mu_+}(P_L u_+) - \overline{u}\|_F^2 \\
&= (1-\theta)[-\frac{1}{2}\|P_L u\|_F^2 + \langle y_{\mu_+}(P_L u_+), P_L u\rangle + \frac{\mu}{2}\|y_{\mu_+}(P_L u_+) - \overline{u}\|_F^2]_1 \\
&\quad + \theta[-\frac{1}{2}\|P_L\hat{y}\|_F^2 + \langle y_{\mu_+}(P_L u_+), P_L\hat{y}\rangle]_2.
\end{aligned}
\tag{20}
$$

In the first bracket we have

$$
\begin{aligned}
&-\frac{1}{2}\|P_L u\|_F^2 + \langle y_{\mu_+}(P_L u_+), P_L u\rangle + \frac{\mu}{2}\|y_{\mu_+}(P_L u_+) - \overline{u}\|_F^2 \\
&= \varphi_\mu(u) + \langle y_{\mu_+}(P_L u_+) - y_\mu(P_L u), P_L u\rangle + \frac{\mu}{2}(\|y_{\mu_+}(P_L u_+) - \overline{u}\|_F^2 - \|y_\mu(P_L u) - \overline{u}\|_F^2) \\
&= \varphi_\mu(u) + \langle P_L u + \mu(y_\mu(P_L u) - \overline{u}), y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\rangle + \frac{\mu}{2}\|y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\|_F^2 \\
&\geq \varphi_\mu(u) + \frac{\mu}{2}(\|y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\|_F^2 \\
&\geq \frac{1}{2}\|P_L y\|_F^2 + \frac{\mu}{2}(\|y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\|_F^2 \\
&\geq \frac{1}{2}\|P_L \hat{y}\|_F^2 + \langle P_L \hat{y}, P_L(y - \hat{y})\rangle + \frac{\mu}{2}(\|y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\|_F^2,
\end{aligned}
\tag{21}
$$

where the second equality is by the three point equality

$$
\|y_{\mu_+}(P_L u_+) - \overline{u}\|_F^2 = \|y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\|_F^2 + \|y_\mu(P_L u) - \overline{u}\|_F^2 + 2\langle y_{\mu_+}(P_L u_+) - y_\mu(P_L u), y_\mu(P_L u) - \overline{u}\rangle,
$$

and the first inequality is by the optimality condition at $y_\mu(P_L u)$:

$$
\langle P_L u + \mu(y_\mu(P_L u) - \overline{u}), y - y_\mu(P_L u)\rangle \geqslant 0, \forall y \in \Delta(\Omega).
$$

In the second term

$$
-\frac{1}{2}\|P_L \hat{y}\|_F^2 + \langle y_{\mu_+}(P_L u_+), P_L \hat{y}\rangle = \frac{1}{2}\|P_L \hat{y}\|_F^2 + \langle P_L y_{\mu_+}(P_L u_+) - P_L \hat{y}, P_L \hat{y}\rangle,
$$

thus we can imply that

$$
\begin{aligned}
&\varphi_{\mu_+}(u_+) \\
&\geq \frac{1}{2}\|P_L \hat{y}\|_F^2 + \langle \theta P_L y_{\mu_+}(P_L u_+) - \theta P_L \hat{y} + (1-\theta) P_L(y - \hat{y}), P_L \hat{y}\rangle + \frac{\mu(1-\theta)}{2}\|y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\|_F^2 \\
&= \frac{1}{2}\|P_L \hat{y}\|_F^2 + \langle P_L y_+ - P_L \hat{y}, P_L \hat{y}\rangle + \frac{\mu(1-\theta)}{2}\|y_{\mu_+}(P_L u_+) - y_\mu(P_L u)\|_F^2 \\
&\geq \frac{1}{2}\|P_L \hat{y}\|_F^2 + \langle P_L y_+ - P_L \hat{y}, P_L \hat{y}\rangle + \frac{1}{2}\|y_+ - \hat{y}\|_F^2 \\
&\geq \frac{1}{2}\|P_L \hat{y}\|_F^2 + \langle P_L y_+ - P_L \hat{y}, P_L \hat{y}\rangle + \frac{1}{2}\|P_L y_+ - P_L \hat{y}\|_F^2 = \frac{1}{2}\|P_L y_+\|_F^2,
\end{aligned}
\tag{22}
$$

where the second inequality is because

$$
\frac{(1-\theta)\mu}{\theta^2} = \frac{4}{(t+2)(t+3)} \cdot \frac{(t+3)^2}{4} > 1.
$$

Based on these two lemmas, we can obtain the decreasing rate of the value of $\|P_L y\|_F$:

Proposition 1 *If Algorithm 1 has not halted after $t \geq 1$ iterations, then*

$$\|P_L y_t\|_F^2 \leq \frac{8}{(t+1)^2}.$$

Proof Since the algorithm has not halted after t iterations, we have $P_L u_t \notin \Omega$ and consequently $\varphi(u_t) \leq 0$. Thus Lemma 5 yields

$$\|P_L y_t\|_F^2 \leq 2\varphi_{\mu_t}(u_t) \leq 2(\mu_t + \varphi(u_t)) \leq 2\mu_t.$$

$\mu_t = \frac{4}{(t+1)(t+2)} \leq \frac{4}{(t+1)^2}$ concludes the proof.

Combined with the bound in Lemma 3, the following proposition can be derived for the iteration complexity of the smooth perceptron-von Neumann basic procedure.

Proposition 2 *After at most $O(\sigma r^{\frac{3}{2}})$ iterations the basic procedure either yields a vector u such that $P_L u$ is feasible for the problem (8) or yields a vector u with $P_M u$ feasible for the problem (4) or indicates a vector y with a set of indicies j such that $bound_j(y) \leq \frac{1}{\sigma}$.*

3.2 A Primal Dual Proximal Mirror Algorithm

Noting that actually we want to find some $u \in V$ satisfying $P_L u > 0$, or equivalently

$$\psi(u) = \min_{y \in \Delta(\Omega)} \langle P_L u, y \rangle > 0, \tag{23}$$

it suffices to show that

$$\max_{u \in \mathscr{B}_V} \psi(u) = \max_{u \in \mathscr{B}_V} \min_{y \in \Delta(\Omega)} \langle P_L u, y \rangle > 0, \tag{24}$$

where $\mathscr{B}_V$ denotes the closed unit ball in V centered at 0:

$$\mathscr{B}_V = \{v \in V, \|v\|_F \leq 1\}.$$

When this inequality holds, we can find a specific u meeting our demand. On the other hand, if this does not hold, there is no such u. Thus the primal problem is infeasible. In this subsection, we are going to solve this saddle point problem (24). For a pair of vectors $u_t \in \mathscr{B}_V$ and $y_t \in \Delta(\Omega)$, we use

$$\begin{aligned} &\epsilon_{sad}(u_t, y_t) \\ =& \max_{u \in \mathscr{B}_V} \langle P_L u, y_t \rangle - \min_{y \in \Delta(\Omega)} \langle P_L u_t, y \rangle \\ =& [\max_{u \in \mathscr{B}_V} \langle P_L u, y_t \rangle - \min_{y \in \Delta(\Omega)} \max_{u \in \mathscr{B}_V} \langle P_L u, y \rangle] + [\max_{u \in \mathscr{B}_V} \min_{y \in \Delta(\Omega)} \langle P_L u, y \rangle - \min_{y \in \Delta(\Omega)} \langle P_L u_t, y \rangle] \end{aligned}$$

to characterize the approximation accuracy of the iteration points. Clearly, $\epsilon(u_t, y_t) = 0$ if and only if the pair $[u_t, y_t]$ is the solution of the saddle point problem (24).

Formally, we have the following lemma, which has been presented in [16].

Lemma 6 *Let $u \in \mathscr{B}_V$ and $y \in \Delta(\Omega)$. We have*

1. *If $\psi(u) > 0$, then $P_L u > 0$; otherwise*
2. *If $\epsilon_{sad}(u,y) \le \epsilon$, then $\|P_L y\|_F \le \epsilon$.*

Proof For the first case, as $\psi(u) = \min_{y\in\Delta(\Omega)} \langle P_L u, y\rangle > 0$, it is easy to see that $P_L u > 0$. For the second case, we have

$$\|P_L y\|_F = \max_{\|u\|_F\le 1} \langle P_L y, u\rangle = \max_{\|u\|_F\le 1} \langle y, P_L u\rangle \le \epsilon_{sad}(u,y) + \psi(u) \le \epsilon.$$

In the saddle point problem (24), the function $\langle P_L u, y\rangle$ is convex and concave (actually both linear) for y and u separately. In [17], a kind of proximal mirror descent is designed to solve this kind of problems

$$\min_{x\in X}\max_{y\in Y} \phi(x,y),$$

and the associated variational inequality

$$\text{find } z_* = (x_*, y_*) \in X\times Y : \langle \Phi(z_*), z - z_*\rangle \ge 0, \forall z \in X \times Y$$

where

$$\Phi(x,y) = \begin{pmatrix} \frac{\partial}{\partial x}\phi(x,y) \\ -\frac{\partial}{\partial y}\phi(x,y) \end{pmatrix}.$$

This method has been adopted in [16] to solve Problem (2). Here we will customize this method as a basic procedure of our algorithm, with some modifications of the implementation.

Each iteration in the proximal mirror descent method is a kind of prox-mapping

$$P_z(\xi) = \arg\min_{w\in Z}\{\omega(w) + \langle w, \gamma\xi - \omega'(z)\rangle\} = \arg\min_{w\in Z}\{\langle w, \gamma\xi\rangle + V_z(w)\},$$

where Z is the domain, $\omega(\cdot)$ is a strongly convex function (distance generating function), $V_z(w) = \omega(w) - \langle\omega'(z), w - z\rangle - \omega(z)$ is the corresponding Bregman distance function, and γ is the so called step size. When the function $\Phi(z)$ is Lipschitz continuous with a coefficient L, a proper step size γ can result in an elegant convergence analysis of this algorithm.

To implement this method as a basic procedure, we shall first define the domain $Z = \mathscr{B}_V \times \Delta(\Omega)$, $\omega(u,y) = \frac{1}{2}\|u\|_F^2 + \frac{1}{2}\|y - \frac{1}{r}e\|_F^2$. In [16] and [17], the distance generating function on $\Delta(\Omega)$ is set to be the regularized entropy, as the associated prox-mapping admits a closed form expression. However, with $\omega(y) = \|y - \frac{1}{r}e\|_F^2$, the iteration complexity can be reduced by a factor $\log r$, while in many cases the number of arithmetic operations will not increase. With this distance generating function, the Bregman distance is $V_z(w) = \|w - z\|_F^2$.

The minimum point y_0 and u_0 of the functions $\|y - \frac{1}{r}e\|_F^2$ and $\|u\|_F^2$ are respectively $y_0 = \frac{1}{r}e$ and $u_0 = 0$. As

$$\Theta(u_0) = \max_{u\in\Delta(\Omega)} \|u - \frac{1}{r}e\|_F^2 \le 1,$$

Algorithm 2 $[y, u, z, J, case]$ = Prox-mirror algorithm(P_L, y, σ)

initialize: $t = 0, v_0 = 0, v_0 = 0, y_0 = s_0 = \frac{1}{r}e, \gamma = \frac{1}{2}, case = 0,$
while $bound(y_t) > \frac{1}{\sigma}$ and case=0 **do**
  if $P_L u_t \succ 0$ **then**
    case=1 (u_t is primal feasible); return
    if $P_L y_t \succ 0$ **then**
      case=1 (y_t is primal feasible); return
    else
      if $v_t = P_M y_t \succeq 0$ and $v_t \neq 0$ **then**
        case=2 (y is dual feasible); return
      else
        $w_t = \arg\min_{\|w\|_F \le 1}\{-\langle \gamma P_L s_t, w\rangle + \frac{1}{2}\|w - v_t\|_F^2\}$
        $q_t = \arg\min_{q \in \Delta(\Omega)}\{\langle \gamma P_L v_t, q\rangle + \frac{1}{2}\|q - s_t\|_F^2\}$
        $v_{t+1} = \arg\min_{\|v\|_F \le 1}\{-\langle \gamma P_L q_t, v\rangle + \frac{1}{2}\|v - v_t\|_F^2\}$
        $s_{t+1} = \arg\min_{s \in \Delta(\Omega)}\{\langle \gamma P_L w_t, s\rangle + \frac{1}{2}\|s - s_t\|_F^2\}$
        $u_{t+1} = \frac{\sum_{j=0}^t w_j}{t+1}$
        $y_{t+1} = \frac{\sum_{j=0}^t q_j}{t+1}$
        $t = t + 1$
      end if
    end if
  end if
end while
if case=0 **then**
  find a nonempty set J such that $J \subseteq \{j : bound_j(y_{t+1}) \le \frac{1}{\sigma}\}$
end if

$$\Theta(y_0) = \max_{\|y\|_F \le 1} \|y\|_F^2 \le 1,$$

we have

$$\Theta(y_0, u_0) = \max_{u \in \Delta(\Omega)} \frac{1}{2}\|u - \frac{1}{r}e\|_F^2 + \frac{1}{2}\|y\|_F^2 \le 1.$$

Then we are going to prove the convergence rate of this algorithm. First we need to show that the proximal projection in general is actually Lipschitz continuous.

Lemma 7 *Let a nonempty set $U \subset V$ be convex and closed, and let $v \in V$. Consider the points*

$$w = \arg\min_{y \in U}\{\langle \gamma\xi, y\rangle + V_v(y)\},$$

$$v_+ = \arg\min_{y \in U}\{\langle \gamma\eta, y\rangle + V_v(y)\},$$

where ξ and η are two points in V, $\gamma > 0$ is the step size. Then for all $v \in V$ one has

$$\|w - v_+\|_F \le \frac{\gamma}{\alpha}\|\xi - \eta\|_F,$$

where α is a constant satisfying $V_v(y) \ge \frac{\alpha}{2}\|v - y\|_F^2$.

Proof By the optimality condition, we have

$$\langle \gamma\xi + \omega'(w) - \omega'(v), u - w \rangle \geq 0, \forall u \in U,$$

and

$$\langle \gamma\eta + \omega'(v_+) - \omega'(v), u - v_+ \rangle \geq 0, \forall u \in U.$$

Setting $u = v_+$ in the first inequality and $u = w$ in the second, these become

$$\langle \gamma\xi, w - v_+ \rangle \leq \langle \omega'(v) - \omega'(w), w - v_+ \rangle,$$

and

$$\langle \gamma\eta, v_+ - w \rangle \leq \langle \omega'(v_+) - \omega'(v), w - v_+ \rangle.$$

Summing them up, it can be obtained that

$$\langle \gamma(\xi - \eta), w - v_+ \rangle \leq \langle \omega'(v_+) - \omega'(w), w - v_+ \rangle \leq -\alpha\|w - v_+\|_F^2.$$

Thus we have

$$-\gamma\|\xi - \eta\|_F\|w - v_+\|_F \leq \gamma\langle(\xi - \eta), w - v_+\rangle \leq -\alpha\|w - v_+\|_F^2$$

$$\Rightarrow \|w - v_+\|_F \leq \frac{\gamma}{\alpha}\|\xi - \eta\|_F.$$

In our algorithm, the convex sets U are $\mathscr{B}_V$ for u and $\Delta(\Omega)$ for y respectively. The Bregman distance $V_v(y)$ are always $\frac{1}{2}\|y - v\|_F^2$, thus $\alpha = 1$. Based on this and the contracting property of the projection P_L, we have the following result for each iteration.

Lemma 8

$$\langle \gamma P_L w_t, q_t - s_{t+1} \rangle - \langle \gamma P_L q_t, w_t - v_{t+1} \rangle - \frac{1}{2}\|v_{t+1} - v_t\|_F^2 - \frac{1}{2}\|s_{t+1} - s_t\|_F^2 \leq 0,$$

given the step size $\gamma \leq \frac{\sqrt{2}}{2}$.

Proof First we need the optimality conditions at w_t and q_t:

$$\langle -\gamma P_L s_t + w_t - v_t, u - w_t \rangle \geq 0, \forall u \in \mathscr{B}_V,$$

$$\langle \gamma P_L v_t + q_t - s_t, y - q_t \rangle \geq 0, \forall y \in \Delta(\Omega).$$

Setting $u = v_{t+1}$ and $y = s_{t+1}$, it indicates

$$\langle -\gamma P_L s_t, w_t - v_{t+1} \rangle \leq \langle w_t - v_t, v_{t+1} - w_t \rangle$$

and

$$\langle \gamma P_L v_t, q_t - s_{t+1} \rangle \leq \langle q_t - s_t, s_{t+1} - q_t \rangle.$$

Thus we have

$$\begin{aligned}
&\langle \gamma P_L w_t, q_t - s_{t+1}\rangle - \frac{1}{2}\|s_{t+1} - s_t\|_F^2 \\
&= \langle \gamma P_L w_t - \gamma P_L v_t, q_t - s_{t+1}\rangle + \langle \gamma P_L v_t, q_t - s_{t+1}\rangle - \frac{1}{2}\|s_{t+1} - s_t\|_F^2 \\
&\leq \langle \gamma P_L w_t - \gamma P_L v_t, q_t - s_{t+1}\rangle + \langle s_{t+1} - q_t, q_t - s_t\rangle - \frac{1}{2}\|s_{t+1} - s_t\|_F^2 \\
&= \langle \gamma P_L w_t - \gamma P_L v_t, q_t - s_{t+1}\rangle - \frac{1}{2}(\|s_{t+1} - q_t\|_F^2 + \|q_t - s_t\|_F^2) \\
&\leq \gamma^2\|v_t - w_t\|_F^2 - \frac{1}{2}(\|s_{t+1} - q_t\|_F^2 + \|q_t - s_t\|_F^2) \\
&\leq \gamma^2\|v_t - w_t\|_F^2 - \frac{1}{2}\|q_t - s_t\|_F^2.
\end{aligned}$$

Similarly,

$$\langle -\gamma P_L q_t, w_t - v_{t+1}\rangle - \frac{1}{2}\|v_{t+1} - v_t\|_F^2 \leq \gamma^2\|s_t - q_t\|_F^2 - \frac{1}{2}\|w_t - v_t\|_F^2.$$

Summing these two inequalities up, it indicates that

$$\langle \gamma P_L w_t, q_t - s_{t+1}\rangle - \langle \gamma P_L q_t, w_t - v_{t+1}\rangle - \frac{1}{2}\|v_{t+1} - v_t\|_F^2 - \frac{1}{2}\|s_{t+1} - s_t\|_F^2 \leq 0$$

as $\gamma \leq \frac{\sqrt{2}}{2}$.

Based on these lemmas, we can show the convergence rate of this algorithm.

Theorem 1 *If the algorithm has not halted after N steps and $\gamma \leqslant \frac{\sqrt{2}}{2}$, then*

$$\epsilon_{sad}(u_{N+1}, y_{N+1}) \leq \frac{\Theta(v_0, s_0)}{N\gamma}.$$

Proof By the optimality conditions at v_{t+1} and s_{t+1}, we have

$$\langle -\gamma P_L q_t + v_{t+1} - v_t, u - v_{t+1}\rangle \geq 0, \forall u \in \mathscr{B}_V,$$

$$\langle \gamma P_L w_t + s_{t+1} - s_t, y - s_{t+1}\rangle \geq 0, \forall y \in \Delta(\Omega).$$

The first inequality indicates that

$$\begin{aligned}
&-\langle \gamma P_L q_t, w_t - u\rangle \\
&\leq \langle v_{t+1} - v_t, u - v_{t+1}\rangle - \langle \gamma P_L q_t, w_t - v_{t+1}\rangle \\
&= \langle v_{t+1} - v_t, u - v_t\rangle - \|v_{t+1} - v_t\|_F^2 - \langle \gamma P_L q_t, w_t - v_{t+1}\rangle,
\end{aligned}$$

while the second inequality indicates

$$\langle \gamma P_L w_t, q_t - y\rangle \leq \langle s_{t+1} - s_t, y - s_t\rangle - \|s_{t+1} - s_t\|_F^2 + \langle \gamma P_L w_t, q_t - s_{t+1}\rangle.$$

Summing up these two and applying Lemma 8 indicate that

$$\begin{aligned}
&\langle \gamma P_L w_t, q_t - y\rangle - \langle \gamma P_L q_t, w_t - u\rangle \\
&\leq \langle s_{t+1} - s_t, y - s_t\rangle - \|s_{t+1} - s_t\|_F^2 + \langle \gamma P_L w_t, q_t - s_{t+1}\rangle + \langle v_{t+1} - v_t, u - v_t\rangle - \|v_{t+1} - v_t\|_F^2 - \langle \gamma P_L q_t, w_t - v_{t+1}\rangle \\
&\leq \langle s_{t+1} - s_t, y - s_t\rangle - \frac{1}{2}\|s_{t+1} - s_t\|_F^2 + \langle v_{t+1} - v_t, u - v_t\rangle - \frac{1}{2}\|v_{t+1} - v_t\|_F^2 \\
&= \frac{1}{2}(\|y - s_t\|_F^2 - \|y - s_{t+1}\|_F^2) + \frac{1}{2}(\|u - v_t\|_F^2 - \|u - v_{t+1}\|_F^2).
\end{aligned}$$

Summing this from $t = 0$ to $t = N - 1$, we can obtain

$$\begin{aligned}
&\sum_{t=0}^{N-1}(\langle \gamma P_L q_t, u\rangle - \langle \gamma P_L w_t, y\rangle) \\
&\leq \sum_{t=0}^{N-1}[\frac{1}{2}(\|y - s_t\|_F^2 - \|y - s_{t+1}\|_F^2) + \frac{1}{2}(\|u - v_t\|_F^2 - \|u - v_{t+1}\|_F^2)] \\
&= \frac{1}{2}(\|y - s_0\|_F^2 - \|y - s_N\|_F^2) + \frac{1}{2}(\|u - v_0\|_F^2 - \|u - v_N\|_F^2) \\
&\leq \frac{1}{2}\|y - s_0\|_F^2 + \frac{1}{2}\|u - v_0\|_F^2 \\
&\leq \Theta(v_0, s_0).
\end{aligned}$$

As $u_N = \frac{\sum_{j=0}^{N-1} w_j}{N}$ and $y_N = \frac{\sum_{j=0}^{N-1} q_j}{N}$, this implies

$$N(\langle \gamma P_L u, y_N\rangle - \langle \gamma P_L u_N, y\rangle) \leq \Theta(v_0, s_0), \forall u \in \mathscr{B}_V, y \in \Delta(\Omega)$$

$$\Rightarrow \epsilon_{sad}(u_N, y_N) = \max_{\|u\|_F \leq 1, y \in \Delta(\Omega)}(\langle P_L u, y_N\rangle - \langle P_L u_N, y\rangle) \leq \frac{\Theta(s_0, v_0)}{N\gamma}.$$

Notice that in our algorithm, $\Theta(s_0, v_0) = 1$, and the step size is chosen as $\gamma = \frac{1}{2} \leq \frac{\sqrt{2}}{2}$, which can ensure the convergence of the algorithm. Then the following result can be obtained:

Proposition 3 *If Algorithm 2 has not halted after* $t \geq 1$ *iterations, then*

$$\|P_L y_t\|_F^2 \leq \frac{4}{t^2}.$$

So the decreasing rate of the value of $\|P_L y_t\|_F^2$ is of the same order as that in the smooth perceptron-von Neumann algorithm. And we can obtain the same result for the iteration complexity as in Proposition (2).

Algorithm 3 $[y, \overline{y}, z, J, case]$ = Modified Von Neumann Algorithm(P_L, y, σ)

initialize: $z = P_L y, u_0 = \overline{u} = \frac{1}{r}e, \overline{y} = 0, case = 0$,,
while $bound(y) > \frac{1}{\sigma}$ and case=0 **do**
  if $P_L y \succ 0$ **then**
    case=1 (y is primal feasible); return
  else
    if $P_M y \succeq 0$ and $P_M y \neq 0$ **then**
      case=2 (y is dual feasible); return
    else
      find $K \neq \emptyset$ such that $\sum_{k\in K} \langle z, c_k \rangle \leq 0$
      $\overline{y} = y$
      $\alpha = \langle p_K, p_K - z \rangle / \|p_K - z\|_F^2$
      $y = \alpha y + (1-\alpha) e_K$
      $z = \alpha z + (1-\alpha) p_K$
    end if
  end if
end while
if case=0 **then**
  find a nonempty set J such that $J \subseteq \{j : bound_j(y) \leq \frac{1}{\sigma}\}$
end if

3.3 A Modified Von Neumann Algorithm

In [14], a type of von Neumann algorithm is adopted for the basic procedure. It iterates not only in one coordinate, but maybe several at the same time. Although the iteration complexity may be larger than the smoothed version, the arithmetic complexity may be smaller in each iteration.

We extend this algorithm for the conic feasibility problem in Algorithm 3. Instead of finding one coordinate $\lambda_i(z) < 0$ as in the classical von Neumann algorithm, a set K of indices is found such that

$$\sum_{k\in K} \lambda_k(z) \leq 0.$$

Denote $p_k = P_L c_k$ and $p_K = \frac{1}{|K|}\sum_{k\in K} P_L c_k$. As $\langle z, p_k \rangle = \langle z, P_L c_k \rangle = \langle z, c_k \rangle = \lambda_k(z)$, it can be indicated that

$$\langle z, p_K \rangle = \frac{1}{|K|}\sum_{k\in K} \langle z, p_k \rangle = \frac{1}{|K|}\sum_{k\in K} \lambda_k(z) \leq 0.$$

Thus $\|p_K - z\|_F^2 = (\|z\|_F^2 - \langle z, p_k \rangle) + (\|p_K\|_F^2 - \langle z, p_k \rangle) > \langle z, z - p_k \rangle$, implying that $\alpha \in (0, 1)$. This shows that the value of α in the modified von Neumann algorithm is valid.

In order to bound the number of iterations, we shall first measure the progress in each iteration. The following lemma is just an extension of Lemma 4.1 in [14].

Lemma 9 *Let $z \neq 0$ and let K be such that $\sum_{k\in K} \langle c_k, z \rangle \leq 0$ and $p_K \neq 0$. One has*

$$\frac{1}{\|z_+\|_F^2} \geq \frac{1}{\|z\|_F^2} + |K|,$$

where $z_+ = P_L y_+$, $y_+ = \alpha y + (1-\alpha)e_K$, $e_K = \frac{1}{|K|}\sum_{k\in K} c_k$ *and* $\alpha = \frac{\langle p_K, p_K - z\rangle}{\|p_K - z\|_F^2}$.

Proof

$$z_+ = P_L y_+ = \alpha z + (1-\alpha)p_K = p_K + \alpha(z - p_K).$$

As the value of α exactly minimizes the expression of z_+, if follows that

$$\|z_+\|_F^2 = \|p_K\|_F^2 - \frac{\langle p_K, z - p_K\rangle^2}{\|z - p_K\|_F^2} = \frac{\|p_K\|_F^2\|z\|_F^2 - \langle p_K, z\rangle^2}{\|p_K\|_F^2 + \|z\|_F^2 - 2\langle p_K, z\rangle} \leqslant \frac{\|p_K\|_F^2\|z\|_F^2}{\|p_K\|_F^2 + \|z\|_F^2}.$$

As

$$\|p_K\|_F^2 = \|P_L c_K\|_F^2 \leq \|c_K\|_F^2 = \frac{1}{|K|},$$

we have the result

$$\frac{1}{\|z_+\|_F^2} \geq \frac{1}{\|z\|_F^2} + \frac{1}{\|p_k\|_F^2} \geq \frac{1}{\|z\|_F^2} + |K|.$$

4 Main Procedure

In the basic procedure, the aim is to find the vector y that either induces a primal feasible vector z or a dual feasible vector v. If not, after a certain number of iterations, a rescaling of $L \cap \Omega$ can be found, which is done in the main procedure. We have the following proposition about the main procedure. Compared with the proposition in [1], it is more similar to the ones in [13] and [14].

Proposition 4 *Assume* $y \in \overline{\Omega} \setminus \{0\}$ *is such that* $bound(y) \leq \epsilon$. *Let* K *be the set of indices* k *such that* $bound_k(y) \leq \epsilon$, *and let* $L_u : V \to V$ *be the linear mapping associated to* $u = e + \sum_{k\in K} ac_k$ *for some constant* a *with* $\frac{1}{\epsilon} - 1 \geq a > 0$. *Then*

$$\delta_\infty(L_u(L)\bigcap\Omega) \geq (1+a)^{|K|} \cdot \delta_\infty(L\bigcap\Omega).$$

In particular, if $\epsilon = \frac{1}{2}$ *and* $a = 1$, *we have*

$$\delta_\infty(L_u(L)\bigcap\Omega) \geq 2^{|K|}\delta_\infty(L\bigcap\Omega).$$

Proof It suffices to show that for $x \in L\bigcap\Omega$ with $\|x\|_\infty = 1$, we have $det(L_u x) = (1+a)^{|K|}det(x)$ and $\|L_u x\|_\infty \leq 1$ under the conditions.

Assume $x \in L\bigcap\Omega$ is fixed. Since

$$L_u x = (e + \sum_{k\in K} ac_k) \circ x = x + \sum_{k\in K} a\lambda_k(x)c_k,$$

it can be implied that

$$det(L_u x) = (1+a)^{|K|}det(x).$$

On the other hand, as $\lambda_k(x) \leq \epsilon$ and $(1+a)\epsilon \leq 1$, we have

$$\|L_u x\|_\infty \leq 1.$$

Algorithm 4 $[y, z, case]$ =Projection and Rescaling Algorithm(P_L, y)

initialize: $v = e, y = \frac{1}{n}e, x = 0, case = 0, \mathcal{L} = L_e$,
while case=0 **do**
    $\sigma = \frac{1}{2}$;
    $[y, u, z, J, case]$ =Modified smooth perceptron algorithm(P_L, y, σ) or Prox-mirror algorithm(P_L, y);
    if $case = 0$ **then**
        $v = e + a\sum_{j\in J} c_j$
        Let L_v be the linear mapping associated to v;
        if $\overline{y} \neq 0$ **then**
            $y = \overline{y}$;
        end if
        $L = L_v(L)$;
        $\mathcal{L} = L_v\mathcal{L}$;
    end if
end while
if case=1 **then**
    return $x = \mathcal{L}^{-1}(z) \in L \cap \Omega$
end if

We shall notice that in different basic procedures, the values of $\frac{1}{\|z\|^2}$ have different tendencies. In the modified smooth perceptron-von Neumann algorithm, we have adopted a version of Nesterov's smoothing and accelerating method, where the values of $\frac{1}{\|z_t\|^2}$ are not promised to be monotonically increasing. So the basic procedures can start from any given point each time, as shown in Algorithm 4. On the other hand, in the modified von Neumann algorithm, the values of $\frac{1}{\|z_t\|^2}$ are monotonically increasing. As a result, we can utilize the previous z in the current basic procedure, which is formally stated in the following Algorithm 5.

Algorithm 5 $[x, y, d, case]$ = Modified Main Algorithm(P_L, τ)

initialize: $v = e, y = \frac{1}{n}e, x = 0, case = 0, \mathcal{L} = L_e$,
while case=0 **do**
    $\mathcal{L} = L_u\mathcal{L}$;
    $\sigma = \frac{1}{2}$;
    $[y, \overline{y}, z, J, case]$ =Modified Von Neumann Algorithm(P_L, y, σ)
    if $case = 0$ **then**
        $v = e + a\sum_{j\in J} c_j$
        Let L_v be the linear mapping associated to v;
        if $\overline{y} \neq 0$ **then**
            $y = \overline{y}$
        end if
        $y = \frac{L_v^{-1}(y)}{\langle e, L_v^{-1}y\rangle}$
    end if
end while
if case=1 **then**
    return $x = \mathcal{L}^{-1}(z) \in L \cap \Omega$
end if

5 Complexity Analysis

With different basic procedures, there shall be different complexities for the algorithm. By direct calculations, we can obtain an $O(n^3 \log_2(\frac{1}{\delta_\infty(L\bigcap \Omega)}))$ bound with the modified von Neumann basic procedure, and $O(n^{\frac{3}{2}} \log_2(\frac{1}{\delta_\infty(L\bigcap \Omega)}))$ with the smooth perceptron-von Neumann and the primal-dual proximal mirror basic procedures. However, with a more careful analysis as in [13] and [14], the iteration complexity for the modified main procedure Algorithm 5 can be reduced. First, we need the following lemma, which generalizes Lemma 6.1 in [14].

Lemma 10 *Let $\overline{y}$ be noncutting with respect to the linear subspace L, i.e. it does not induce a $\frac{1}{2}$-cutting in L with some u in Algorithm 5. Moreover, let $y = \frac{L_u^{-1}\overline{y}}{\langle e, L_u^{-1}\overline{y}\rangle}$. If $\overline{z} = P_L\overline{y}$ and $z = P_{L_u(L)}y$, then*

$$\frac{1}{\|\overline{z}\|_F^2} - \frac{1}{\|z\|_F^2} < 2r^2|K|,$$

where $K = \{k : bound_k(\overline{y}) < \frac{1}{\sigma}\}$ is the set returned by the basic procedure.

Proof Since $\overline{v} = \overline{y} - \overline{z} \in L^\perp$, and L_u is a self-adjoint operator, we have $L_u^{-1}(\overline{y} - \overline{z}) \in (L_u(L))^\perp$, which indicates that

$$P_{L_u(L)}(L_u^{-1}(\overline{y}-\overline{z})) = 0 \Rightarrow \|P_{L_u(L)}L_u^{-1}(\overline{y})\|_F = \|P_{L_u(L)}L_u^{-1}(\overline{z})\|_F \le \|L_u^{-1}(\overline{z})\|_F \le \|\overline{z}\|_F.$$

By the definitions of y and z it follows that

$$\|z\|_F = \|P_{L_u(L)}y\|_F = \|P_{L_u(L)}\frac{L_u^{-1}\overline{y}}{\langle e, L_u^{-1}\overline{y}\rangle}\|_F \le \frac{1}{\langle e, L_u^{-1}\overline{y}\rangle}\|\overline{z}\|_F.$$

As

$$\langle e, L_u^{-1}\overline{y}\rangle \ge \sum_{i\notin K}\lambda_i(\overline{y}) = 1 - \sum_{i\in K}\lambda_i(\overline{y}) \ge 0,$$

we can get that

$$\frac{1}{\|z\|_F^2} \ge \frac{\langle e, L_u^{-1}\overline{y}\rangle^2}{\|\overline{z}\|_F^2} \ge \frac{[1-\sum_{i\in K}\lambda_i(\overline{y})]^2}{\|\overline{z}\|_F^2} \ge \frac{1}{\|\overline{z}\|_F^2} - \frac{2\sum_{i\in K}\lambda_i(\overline{y})}{\|\overline{z}\|_F^2}.$$

Since $\overline{y}$ is noncutting, by (18) we have

$$\lambda_i(\overline{y}) \le 2\sqrt{r}\|\overline{z}\|_F.$$

Using $\frac{1}{\|\overline{z}\|_F} \le r^{\frac{3}{2}}$, it can be obtained that

$$\frac{1}{\|\overline{z}\|_F^2} - \frac{1}{\|z\|_F^2} \le \frac{2\sum_{i\in K}\lambda_i(\overline{y})}{\|\overline{z}\|_F^2} \le \frac{2\sqrt{r}|K|\|\overline{z}\|_F}{\|\overline{z}\|_F^2} = \frac{2\sqrt{r}|K|}{\|\overline{z}\|_F} \le 2|K|r^2,$$

which completes the proof.

In the following we will discuss two types of iterations in a modified von Neumann type basic procedure: iterations where y is changed and iterations that leave y unchanged. Following [13] and [14], we call these basic procedure iterations slow and fast respectively. Only in the case of slow iterations, the vector y is modified. The number of fast basic iterations is denoted as F and the number of slow iterations as S. So the total number of iterations in all the basic procedures counts as $N = F + S$.

A fast iteration happens if and only if at the start of the while loop, y is primal feasible or dual feasible or a cutting vector. Number the basic procedure iterations from 1 to N. Denote the vectors y and z at the jth iteration as y_j and z_j respectively. Furthermore, the indices $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ are defined to indicate the fast and slow iterations. From the a_ith iteration to the b_ith iteration are the slow iterations, while from the b_ith to the a_{i+1}th are the fast iterations. Formally we have $a_i \leq b_i < a_{i+1}$ for $1 \leq i < k$. If $a_i \leq j \leq b_i$ for some i then the iteration j is slow. Otherwise it is fast. As all the iterations between the iterations a_i and b_i are slow, and in each such iteration, the decrease of $\frac{1}{\|z\|_F^2}$ is at least 1, we can bound the number of them by the decrease of $\frac{1}{\|z\|_F^2}$.

Lemma 11 *For each i such that $1 \leq i \leq k$ one has*

$$b_i - a_i \leq \frac{1}{\|z_{b_i}\|_F^2} - \frac{1}{\|z_{a_i}\|_F^2}.$$

Proof

$$b_i - a_i = \sum_{j=a_i}^{b_i-1} 1 \leq \sum_{j=a_i}^{b_i-1} \left(\frac{1}{\|z_{j+1}\|_F^2} - \frac{1}{\|z_j\|_F^2}\right) = \frac{1}{\|z_{b_i}\|_F^2} - \frac{1}{\|z_{a_i}\|_F^2}.$$

So the total number of slow iterations should satisfy

$$\begin{aligned} S &\leq \sum_{i=1}^{k}\left(\frac{1}{\|z_{b_i}\|_F^2} - \frac{1}{\|z_{a_i}\|_F^2}\right) \\ &= \frac{1}{\|z_{b_k}\|_F^2} - \frac{1}{\|z_{a_1}\|_F^2} + \sum_{i=1}^{k-1}\left(\frac{1}{\|z_{b_i}\|_F^2} - \frac{1}{\|z_{a_{i+1}}\|_F^2}\right) \\ &\leq r^3 + \sum_{i=1}^{k-1}\left(\frac{1}{\|z_{b_i}\|_F^2} - \frac{1}{\|z_{a_{i+1}}\|_F^2}\right). \end{aligned}$$

It remains to analyze the value of $\frac{1}{\|z_{b_i}\|_F^2} - \frac{1}{\|z_{a_{i+1}}\|_F^2}$.

Lemma 12 *For each i such that $1 \leq i < k$ one has*

$$\frac{1}{\|z_{b_i}\|_F^2} - \frac{1}{\|z_{a_{i+1}}\|_F^2} \leq 2r^2|F_i|,$$

where $F_i = a_{i+1} - b_i - 1 \geq 1$.

Proof Since the iteration b_i is slow and the iteration b_i+1 is fast, the iteration b_i+1 starts with a cutting vector y generated in the iteration b_i and yields an index set K for the cuts induced by the vector y, without changing the non-cutting vector $\overline{y}$ that was the input of the iteration b_i. Following this iteration are F_i-1 fast iterations without changing the vector $\overline{y}$ until the iteration a_{i+1}. After each fast iteration $O(1)$ eigenvalues $\lambda_i(y)$ are changed. By Lemma 9,

$$\frac{1}{\|z_{b_i}\|_F^2}-\frac{1}{\|z_{a_{i+1}}\|_F^2}=\frac{1}{\|z_{b_i}\|_F^2}-\frac{1}{\|z_{b_i+F_i+1}\|_F^2}<2r^2O(F_i),$$

which completes the proof.

Based on this lemma, we can obtain that

$$S\leq r^3+2r^2\sum_{i=1}^{k-1}O(F_i)\leq r^3+2r^2O(F).$$

As $F=T$, where T is the number calling basic procedures, the total number of iterations in all the basic procedures amounts to

$$N=S+F\leq r^3+2r^2O(F)+F=r^3+O(r^2\log\frac{1}{\delta_\infty(L\bigcap\Omega)}).$$

Then we can conclude our results by the following theorem:

Theorem 2 *The total iteration complexities of Algorithm 4 and 5 are* $O(\max\{r^3,r^2\log\frac{1}{\delta_\infty(L\bigcap\Omega)}\})$ *and* $O(r^{\frac{3}{2}}\log\frac{1}{\delta_\infty(L\bigcap\Omega)})$, *respectively.*

Moreover, as the operations of different basic procedures differ, we also include the analysis for arithmetic complexities here. In all the basic procedures mentioned in this paper, $bound_j(y)$ is derived from all the components of P_Ly. So the bulk of the computation in each iteration is the projection P_Ly and the complete eigenvalue decomposition, which takes $O(mn)$ and $O(r^3)$ arithmetic operations respectively in the case of symmetric matrix space, as discussed in [1]. Hence, for the smooth perceptron-von Neumann and primal dual proximal mirror basic procedures, the number of arithmetic operations needed is bounded above by

$$O(\max(mn,r^3)\cdot r^{3/2}),$$

where m and n are respectively the dimensions of the space L and V. For the modified von Neumann basic procedure, it is

$$O(\max(mn,r^3)\cdot r^3).$$

For the other cases, the bounds can be computed similarly, except that the complete eigenvalue decomposition may not take so much extra time.

5.1 Comparisons with the Algorithm of Lourenço et al. [11]

In [11], the normalized conic feasibility problem (8) is also considered, with a different rescaling scheme regarding to the minimum eigenvalue of a feasible solution x. If the minimum eigenvalue of x is larger than ϵ, it is called an ϵ-feasible solution. In the algorithm designed by Lourenço et al., either a primal or a dual solution is found, or it is stated that no ϵ-feasible solution exists. Indeed, with our rescaling scheme, it is also possible to show that no ϵ-feasible solution exists after a certain number of iterations.

Theorem 3 *After at most $r \log_2 \frac{1}{\epsilon}+1$ times of rescaling in the main procedure, if no primal solution is found, there exists no ϵ-feasible solution for problem* (8).

Proof In each rescaling, the eigenvalues of all the solutions corresponding to the indices $k \in K$ are doubled. After at most $r \log_2 \frac{1}{\epsilon} + 1$ times of rescaling, there exists at least one index j which has been rescaled for at least $\log_2 \frac{1}{\epsilon} + 1$ times. After rescaling, the maximum eigenvalue still cannot exceed the value of 1, indicating that no ϵ-feasible solution exists for problem (8).

Moreover, the multi-block case is considered in [11], i.e. when the symmetric cone can be written as a direct product of symmetric cones: $\Omega = \Omega_1 \times \Omega_2 \cdots \times \Omega_l$, where each Ω_i is a simple symmetric cone contained in some finitely dimensional space V_i. Denoting the dimension and rank of each V_i as d_i and r_i respectively, we can show that the complexity of our algorithm depends only on $r_{max} = \max\{r_1, \cdots, r_l\}$ and l, which is similar to the result in [11]. In the basic procedure, as there are l blocks in total, and $\|y\|_1 = 1$, the maximum among the 1-norms of the blocks

$$\|y\|_{1,\infty} = \max\{\|y_1\|_1, \cdots, \|y_l\|_1\} = \max\{\langle y_1, e_1\rangle, \cdots, \langle y_l, e_l\rangle\}$$

satisfies $\|y\|_{1,\infty} \geq \frac{1}{l}$. By Lemma 3, $\frac{1}{\|y\|_F^2}$ cannot exceed r^3 in the basic procedure, otherwise we can obtain a cutting vector. As a result, for the block y_j achieving the value of $\|y\|_{1,\infty}$, when $\frac{1}{\|y\|_F^2} > l^2 r_{max}^3$, obviously $\frac{\|y_j\|_1^2}{\|y_j\|_F^2} > r_{max}^3$, indicating that a cut can be found in the jth block with the cutting vector $\frac{y_j}{\|y_j\|_1}$. Thus the iteration complexity of the modified von Neumann basic procedure can be bounded by $l^2 r_{max}^3$. This may be smaller than the bound r^3 when we consider y as a whole, and also can be better than $l^3 r_{max}^2$ in some conditions, which is the bound for the algorithm in [11], although not always.

On the other hand, considering the condition measure

$$\delta_{1,\infty}(L \bigcap \Omega) \doteq \max_x \{det(x) : x \in L \bigcap \Omega, \|\lambda(x)\|_{1,\infty} = 1\},$$

another bound can be obtained for the multi-block case. Based on the analysis above, we still can rescale the space L with some cutting vector. Similarly to Proposition 4, the condition measure $\delta_{1,\infty}(L \bigcap \Omega)$ can be doubled after each rescaling. As $\delta_{1,\infty}(L \bigcap \Omega) \leq 1$, and a cutting vector can be found in at most

$l^2 r_{max}^3$ iterations, the bound for iteration complexity in multi-block case can be written as

$$O(l^2 r_{max}^3 \log_2 \frac{1}{\delta_{1,\infty}(L \bigcap \Omega)}),$$

with von Neumann basic procedure.

6 Conclusion

The projection and rescaling scheme, first raised in [13], is a novel method for the linear feasibility problems, thus also for the general linear programming, as all the linear programming problems can be transformed into the linear feasibility systems. In this paper we explore its generalization to the conic feasibility problems, following the ideas of [1] and [11]. It is shown that our algorithm is an improvement based on the algorithms in [1] in iteration complexity when the problem is feasible. Besides, we show that our algorithm can also adopt the same stopping criterion of [11], with a comparable complexity. So that the algorithm can stop when no ϵ-feasible solution exists. This indicates that the projection and rescaling scheme is a potential framework. We may find more improvements by exploring the basic and main procedures further.

References

1. Pena J, Soheili N. Solving conic systems via projection and rescaling[J]. Mathematical Programming, 2016: 1-25.
2. Dunagan J, Vempala S. A simple polynomial-time rescaling algorithm for solving linear programs[J]. Mathematical Programming, 2008, 114(1): 101-114.
3. Belloni A, Freund R M, Vempala S. An efficient rescaled perceptron algorithm for conic systems[J]. Mathematics of Operations Research, 2009, 34(3): 621-641.
4. Betke U. Relaxation, new combinatorial and polynomial algorithms for the linear feasibility problem[J]. Discrete & Computational Geometry, 2004, 32(3): 317-338.
5. Minsky M, Papert S A. Perceptrons: an introduction to computational geometry[M]. MIT press, 2017.
6. Dantzig G B. Converting a converging algorithm into a polynomially bounded algorithm[R]. STANFORD UNIV CA SYSTEMS OPTIMIZATION LAB, 1991.
7. Dantzig G B. Bracketing to speed convergence illustrated on the von Neumann algorithm for finding a feasible solution to a linear program with a convexity constraint. Technical report[R]. Stanford Univ., CA (United States). Systems Optimization Lab., 1992.
8. Dadush D, Végh L A, Zambelli G. Rescaling algorithms for linear programming-part I: conic feasibility[J]. arXiv preprint arXiv:1611.06427, 2016.
9. Goffin J L. The relaxation method for solving systems of linear inequalities[J]. Mathematics of Operations Research, 1980, 5(3): 388-414.
10. Khachian L G. A polynomial algorithm in linear programming[J]. Doklady Akademii Nauks SSR, 1979, 244(5): 1093-1096.
11. Lourenço B F, Kitahara T, Muramatsu M, et al. An extension of Chubanov's algorithm to symmetric cones[J]. Mathematical Programming, 2016: 1-33.
12. Nesterov Y. Excessive gap technique in nonsmooth convex minimization[J]. SIAM Journal on Optimization, 2005, 16(1): 235-249.
13. Chubanov S. A polynomial projection algorithm for linear feasibility problems[J]. Mathematical Programming, 2015, 153(2): 687-713.

14. Roos K. An improved version of Chubanov's method for solving a homogeneous feasibility problem[J]. Optimization Methods and Software, 2017: 1-19.
15. Soheili N, Pena J. A primal–dual smooth perceptron–von Neumann algorithm[M]//Discrete Geometry and Optimization. Springer International Publishing, 2013: 303-320.
16. Yu A W, Kilinc-Karzan F, Carbonell J. Saddle points and accelerated perceptron algorithms[C]//International Conference on Machine Learning. 2014: 1827-1835.
17. Nemirovski A. Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems[J]. SIAM Journal on Optimization, 2004, 15(1): 229-251.