Iranian Journal of Operations Research
Vol. 9, No. 2, 2018, pp. 95-105

Using Nesterov's Excessive Gap Method as Basic Procedure in Chubanov's Method for Solving a Homogeneous Feasibility Problem

Z. Wei[1], C. Roos[2,*]

We deal with a recently proposed method of Chubanov [1], for solving linear homogeneous systems with positive variables. We use Nesterov's excessive gap method in the basic procedure. As a result, the iteration bound for the basic procedure is reduced by the factor $n\sqrt{n}$. The price for this improvement is that the iterations are more costly, namely $O(n^2)$ instead of $O(n)$. The overall gain in the complexity hence becomes a factor of $\sqrt{n}$.

Keywords: *Linear homogeneous systems, Algorithm, Polynomial-time.*

Manuscript was received on 01/07/2018 revised on 12/09/2018 and accepted for publication on 20/10/2018.

1. Introduction

We deal with the problem

$$\begin{aligned} &\text{find } x \in \mathbb{R}^n \\ &\text{subject to } Ax = 0, \qquad x > 0, \end{aligned} \tag{1}$$

where A is an integer (or rational) matrix of size $m \times n$ and $\operatorname{rank}(A) = m$.

Recently Chubanov [1] proposed a polynomial-time algorithm for solving this problem. He explored the fact that (1) is homogeneous as follows. If x is feasible for (1), then also $x' = x/\max(x)$ is feasible for (1), and this solution belongs to the unit cube, i.e., $x' \in [0,1]^n$. It follows that (1) is feasible if and only if the system

$$Ax = 0, \quad x \in (0,1]^n \tag{2}$$

is feasible. Moreover, if $d > 0$ is a vector such that $x \leq d$ holds for every feasible solution of (2), then $x'' = x/d \in (0,1]^n$, where x/d denotes the entry-wise quotient of x and d, and so $x_i'' = x_i/d_i$ for each i. This means that x'' is feasible for the system

$$ADx = 0, \quad x \in (0,1]^n, \tag{3}$$

* Corresponding Author.
1 National University of Singapore, Singapore, Email: lindelfeel@gmail.com.
2 Department of Electrical Engineering, Computer Science and Mathematics, Delf University of Technology, Delf, The Netherlands, Email: c.roos@tudelf.nl.

where $D = \text{diag}(d)$. Obviously, problem (3) is of the same type as problem (2), since it arises from (2) by rescaling A to AD. Moreover, if x solves (3), then Dx solves (2). The main algorithm starts with $d = \mathbf{1}$, with $\mathbf{1}$ denoting the all-one vector, and successively improves d.

A key ingredient in Chubanov's algorithm is the so-called Basic Procedure (BP). The BP generates one of the following three outputs:

case 1: a feasible solution of (1);
case 2: a certificate for the infeasibility of (1);
case 0: a cut for the feasible region of (2).

In case 0, the cut has the form $x_k \leq \frac{1}{2}$ for some index k and is used to update d by dividing d_k by 2. The rescaling happens in the main algorithm, which sends the rescaled matrix AD to the BP until the BP returns case 1 or case 2.

Since A has integer (or rational) entries, the number of calls of the BP is polynomially bounded by $O(nL)$, where L denotes the bit size A. This follows from a classical result of Khachiyan [2] that gives a positive lower bound on the positive entries of a solution of a linear system of equations.

The BP of [1] needs at most $O(n^3)$ iterations per call and $O(n)$ time per iteration. So, per call the BP needs $O(n^4)$ time and hence the overall time complexity becomes $O(n^5 L)$. By performing a more careful analysis, Chubanov reduced this bound by a factor n to $O(n^4 L)$ [1, Theorem 2.1].

Other BPs have been proposed in, e.g., [4, 5]. These BPs also need $O(n^3)$ iterations per call and $O(n)$ time per iteration, and so they also yield an overall time complexity of $O(n^5 L)$.

In [6], we proposed a BP based on the Mirror-Prox method of Nemirovski. It improves the iteration bound per call with a factor $n\sqrt{n}$ and leads to an overall time complexity of $O(n^{4.5} L)$, because it requires $O(n^2)$ time per iteration.

Here, we analyze a BP based on the Excessive Gap technique of Nesterov [3]. The outline of the remainder of the paper is as follows. We present some preliminary results in Section 2. In Section 3 we describe the new BP and prove the iteration bound of $O\left(n\sqrt{n}\right)$. Since the time complexity per iteration is $O(n^2)$, the overall time complexity is the same as the one given in [6].

2. Preliminaries

Let $\mathcal{N}_A$ denote the null space of the $m \times n$ matrix A and $\mathcal{R}_A$ denote its row space, that is,

$$\mathcal{N}_A := \{x \in \mathbb{R}^n : \ Ax = 0\}, \qquad \mathcal{R}_A := \{A^T u : u \in \mathbb{R}^m\}.$$

We denote the orthogonal projections of $\mathbb{R}^n$ onto $\mathcal{N}_A$ and $\mathcal{R}_A$ as P_A and Q_A, respectively:

$$P_A := I - A^T (AA^T)^{-1} A, \qquad Q_A := A^T (AA^T)^{-1} A.$$

Our assumption $\text{rank}(A) = m$ implies that the inverse of AA^T exists. Obviously, we have

$$I = P_A + Q_A, \qquad P_A Q_A = 0, \qquad AP_A = 0, \qquad AQ_A = A.$$

Now, let $y \in \mathbb{R}^n$. In the sequel, we use the notation

$$z = P_A y, \qquad v = Q_A y.$$

So, z and v are the orthogonal components of y in the spaces $\mathcal{N}_A$ and $\mathcal{R}_A$, respectively:

$$y = z + v, \qquad z \in \mathcal{N}_A, \qquad v \in \mathcal{R}_A.$$

These vectors play a crucial role in our approach. This is due to the following lemma.

Lemma 2.1. (Lemma 2.1 of [5]) If $z > 0$, then z solves the *primal* problem (1) and if $0 \neq v \geq 0$, then v provides a certificate for the infeasibility of (1).

As usual, we always assume $y \in \Delta$, where Δ denotes the unit simplex in $\mathbb{R}^n$. So,

$$\Delta = \{u : \mathbf{1}^T u = 1, \; u \geq 0\}.$$

In the literature we nowadays have several ways to derive from y, z and v an upper bound for the k-th coordinate of every x that is feasible for (3). For example,

$$x_k \leq \begin{cases} \dfrac{\sqrt{n}\|z\|}{y_k}, & \text{in [1].} \\ \dfrac{\mathbf{1}^T z^+}{y_k}, & \text{in [4,5].} \\ \mathbf{1}^T \left(\dfrac{v}{-v_k}\right)^+, & \text{in [5,6].} \end{cases}$$

Here, we are only interested in the so-called proper cuts, where the upper bound is smaller than or equal to $\frac{1}{2}$. If $2n\sqrt{n}\|z\| \leq 1$, then the first two cuts are proper for at least one k. This follows for the first bound simply by taking k such that $y_k \geq 1/n$, and for the second bound by also using $\mathbf{1}^T z^+ \leq \sqrt{n}\|z^+\| \leq \sqrt{n}\|z\|$. For the third bound, it seems far from trivial that we have the same property; for a proof we refer to the Appendix in [6][3].

It may also be mentioned that the third cut is always at least as tight as the other two cuts; this is shown in [5]. In the rest of the paper, we use this cut, denoting the upper bound as $\sigma_k(y)$ and defining

$$\sigma(y) = \min_k \sigma_k(y).$$

Next, the BP based on Nesterov's Excessive Gap method is described as in Algorithm 1.

[3] There is also a 'proof' of this statement in [5], but unfortunately there is a gap in that proof that has been overlooked.

Algorithm 1: $[y, v, z, J, \text{case}] = \text{EXCESSIVE GAP BP}(P_A)$

1: INITIALIZE: $k = 0$; $\bar{u} = \frac{1}{n}\mathbf{1}$; case $= 0$; $J = \emptyset$; $u_0 = \bar{u}$; $\mu_0 = 2$; $y_0 = y_{\mu_0}(u_0)$
2: **while** $\sigma(y_k) > \frac{1}{2}$ **and** $\sigma(u_k) > \frac{1}{2}$ **and** case $= 0$ **do**
3: $\quad z_k = P_A y_k$
4: $\quad$ **if** $z_k > 0$ **then**
5: $\quad\quad$ case $= 1$ (z_k is primal feasible); **return**
6: $\quad$ **else**
7: $\quad\quad v_k = y_k - z_k$
8: $\quad\quad$ **if** $v_k \geq 0$ **then**
9: $\quad\quad\quad$ case $= 2$ (u_k is dual feasible); **return**
10: $\quad\quad$ **else**
11: $\quad\quad\quad \theta_k = \frac{2}{k+3}$
12: $\quad\quad\quad u_{k+1} = (1-\theta_k)u_k + \theta_k\left((1-\theta)y_k + \theta_k y_{\mu_k}(y_k)\right)$
13: $\quad\quad\quad \mu_{k+1} = (1-\theta_k)\mu_k$
14: $\quad\quad\quad y_{k+1} = (1-\theta_k)y_k + \theta_k y_{\mu_{k+1}}(u_{k+1})$
15: $\quad\quad\quad k = k+1$
16: $\quad\quad$ **end**
17: $\quad$ **end**
18: **end**
19: **if** case $= 0$ **then**
20: $\quad$ find a nonempty set J such that $J \subseteq \{j : \text{bound}_j(y_k) \leq \frac{1}{2}\} \bigcup \{j : \text{bound}_j(u_k) \leq \frac{1}{2}\}$
21: **end**

In Algorithm 1, k serves as the iteration counter. We also use the following notation:

$$y_\mu(v) := \text{argmin}_{u\in\Delta}\left\{u^T P_A v + \frac{\mu}{2}\|u - \bar{u}\|^2\right\}, \qquad v \in \Delta, \tag{4}$$

where $\bar{u} = \mathbf{1}/n$. Note that in each iteration two problems of this type need to be solved, in line 12 and line 14, respectively. In Section 4 we show that if $P_A v$ is known, then problem (4) can be solved in O(n) time. But, we first show in the next section that the number of iterations of Algorithm 1 never exceeds $O(n\sqrt{n})$.

3. Iteration Bound

Recall that y yields a solution of problem (2) if $z = P_A y > 0$. If $u \in \Delta$ then $u^T z \geq \min z$, for each $z \in \mathbb{R}^n$ and $\min_{u\in\Delta} u^T z = \min z$. Hence, $P_A y > 0$ holds if and only if $\psi(y) > 0$, where

$$\psi(y) := \min_{u\in\Delta} u^T P_A y.$$

This certainly holds if y solves the problem

$$\max_{y\in\Delta}\psi(y)=\max_{y\in\Delta}\min_{u\in\Delta}u^TP_Ay>0.$$

In order to deal with this problem, we use an adapted version of the excessive gap technique of Nesterov [3] by considering a smoothed version of the above problem. For decreasing values of the parameter μ, we consider instead the problem of maximizing the function $\phi_\mu(y)$, where

$$\phi_\mu(y)=-\frac{1}{2}\|P_Ay\|^2+\min_{u\in\Delta}\left\{u^TP_Ay+\frac{\mu}{2}\|u-\bar{u}\|^2\right\},$$

with $\bar{u}=\mathbf{1}/n$ and $\mu\geq 0$.

In this section, we show that the algorithm needs at most $O\left(n\sqrt{n}\right)$ iterations to generate a vector $y\in\Delta$ such that $2n\sqrt{n}\|z\|\leq 1$. For the proof, we consider a run of the BP during which z has always a nonpositive entry and v a negative entry. So, the BP does not halt in line 5 or line 9. In that case, the algorithm stops after at most $2n\sqrt{2n}$ iterations, as we show below. We start with a relatively simple lemma.

Lemma 3.1. $0\leq\phi_\mu(y)-\phi_0(y)\leq\mu$.

Proof. Let $u\in\Delta$. Then $\|u\|^2=\sum_{i=1}^n u_i^2\leq\sum_{i=1}^n u_i=1$. Similarly, $\|\bar{u}\|^2\leq 1$. Hence,

$$\frac{1}{2}\|u-\bar{u}\|^2=\frac{1}{2}(\|u\|^2+\|\bar{u}\|^2-2u^T\bar{u}\leq 1,$$

where we also used $u\geq 0$ and $\bar{u}\geq 0$. Using this, we write

$$\phi_\mu(y)\leq-\frac{1}{2}\|P_Ay\|^2+\min_{u\in\Delta}u^TP_Ay+\mu=\mu+\phi_0(y),$$

It remains to show that $\phi_\mu(y)\geq\phi_0(y)$. This follows since $\phi_\mu(y)$ is increasing in μ. Hence the proof is complete. ■

Lemma 3.2. $\frac{1}{2}\|P_Ay_k\|^2\leq\phi_{\mu_k}(u_k)$.

Proof. We start with the case where $k=0$. Then, we have $u_0=\bar{u}=\frac{\mathbf{1}}{n}$, $\mu_0=2$, $y_0=y_{\mu_0}(u_0)$. We simplify the notation by denoting P_A simply as P. Then, we may write

$$\begin{aligned}\frac{1}{2}\|P_Ay_0\|^2&=\frac{1}{2}\|P(y_0-\bar{u})+P(\bar{u})\|^2\\&=\frac{1}{2}\|P(y_0-\bar{u})\|^2+\frac{1}{2}\|P\bar{u}\|^2+(Py_0)^TP\bar{u}-\|P\bar{u}\|^2\\&\leq-\frac{1}{2}\|P\bar{u}\|^2+y_0^TP\bar{u}+\frac{1}{2}\|y_0-\bar{u}\|^2\\&\leq-\frac{1}{2}\|P\bar{u}\|^2+y_0^TP\bar{u}+\frac{\mu_0}{2}\|y_0-\bar{u}\|^2\\&=-\frac{1}{2}\|P\bar{u}\|^2+\min_{u\in\Delta}\left\{u^TPu_0+\frac{\mu_0}{2}\|u-\bar{u}\|^2\right\}=\phi_{\mu_0}(u_0),\end{aligned}$$

where the last but one equality is due to the definition of y_0. We proceed with induction on k. To simplify the notation further, we denote $y = y_k$, $\mu = \mu_k$, $u = u_k$, $y' = y_{k+1}$, $\mu' = \mu_{k+1}$, $u' = u_{k+1}$ and

$$\hat{y} = (1-\theta)y + \theta y_\mu(y). \tag{5}$$

Then, we have

$$\begin{aligned} u' &= (1-\theta)(u+\theta y) + \theta^2 y_\mu(y) \\ &= (1-\theta)u + \theta\big[(1-\theta)y + \theta y_\mu(y)\big] \\ &= (1-\theta)u + \theta\hat{y}. \end{aligned}$$

Moreover,

$$\mu' = (1-\theta)\mu,$$

and

$$y' = (1-\theta)y + \theta y_{\mu'}(u'). \tag{6}$$

Under the assumption that $\frac{1}{2}\|Py\|^2 \le \phi_\mu(u)$ we need to show that $\frac{1}{2}\|Py'\|^2 \le \phi_{\mu'}(u')$.

We have

$$\begin{aligned} \phi_{\mu'}(u') &= -\frac{1}{2}\|Pu'\|^2 + \min_{u\in\Delta}\left\{u^T P u' + \frac{\mu'}{2}\|u-\bar{u}\|^2\right\} \\ &= -\frac{1}{2}\|Pu'\|^2 + y_{\mu'}(u')^T P u' + \frac{\mu'}{2}\left\|y_{\mu'}(u') - \bar{u}\right\|^2. \end{aligned}$$

Due to the definition of u' and since $\|z\|^2$ is convex in z, we get

$$\|Pu'\|^2 = \|(1-\theta)Pu + \theta P\hat{y}\|^2 \le (1-\theta)\|Pu\|^2 + \theta\|P\hat{y}\|^2.$$

Hence,

$$\begin{aligned} \phi_{\mu'}(u') &\ge -\frac{1}{2}(1-\theta)\|Pu\|^2 - \frac{1}{2}\theta\|P\hat{y}\|^2 + y_{\mu'}(u')^T P u' + \frac{\mu'}{2}\left\|y_{\mu'}(u') - \bar{u}\right\|^2 \\ &= -\frac{1}{2}(1-\theta)\|Pu\|^2 - \frac{1}{2}\theta\|P\hat{y}\|^2 + y_{\mu'}(u')^T P\big((1-\theta)u + \theta\hat{y}\big) + \frac{\mu'}{2}\left\|y_{\mu'}(u') - \bar{u}\right\|^2 \\ &= (1-\theta)\left[-\frac{1}{2}\|Pu\|^2 + y_{\mu'}(u')^T P u + \frac{\mu}{2}\|y_{\mu'}(u') - \bar{u}\}\|^2\right] + \theta\left[-\frac{1}{2}\|P\hat{y}\|^2 + y_{\mu'}(u')^T P\hat{y}\right]. \end{aligned}$$

Let us denote the two bracketed expressions shortly by T_1 and T_2, respectively. We proceed by evaluating T_1, the first bracketed expression. This can be reduced as follows:

$$T_1 = -\frac{1}{2}\|Pu\|^2 + y_{\mu'}(u')^T P u + \frac{\mu}{2}\left\|y_{\mu'}(u') - \bar{u}\right\|^2$$

$$= \left(\phi_\mu(u) - y_\mu(u)^T Pu - \frac{\mu}{2}\left\|y_\mu(u) - \bar{u}\right\|^2\right) + \; y_{\mu'}(u')^T Pu \; + \frac{\mu}{2}\left\|y_{\mu'}(u') - \bar{u}\right\|^2$$
$$= \phi_\mu(u) + (Pu)^T\left(y_{\mu'}(u') - y_\mu(u)\right) + \frac{\mu}{2}\left(\left\|y_{\mu'}(u') - \bar{u}\right\|^2 - \left\|y_\mu(u) - \bar{u}\right\|^2\right).$$

Putting $a = y_{\mu'}(u')$ and $b = y_\mu(u)$, we have

$$\begin{aligned}\|a - \bar{u}\|^2 - \|b - \bar{u}\|^2 &= \|a\|^2 - \|b\|^2 - 2a^T\bar{u} + 2b^T\bar{u} \\ &= \|a - b\|^2 - 2\|b\|^2 + 2a^T b - 2a^T\bar{u} + 2b^T\bar{u} \\ &= \|a - b\|^2 + 2(b - \bar{u})^T(a - b) \qquad (7)\end{aligned}$$

Using this, we obtain

$$T_1 = \phi_\mu(u) + \left(Pu + \mu\left(y_\mu(u) - \bar{u}\right)\right)^T\left(y_{\mu'}(u') - y_\mu(u)\right) \; + \frac{\mu}{2}\left\|y_{\mu'}(u') - y_\mu(u)\right\|^2.$$

From (5) and (6), we deduce

$$\theta\left(y_{\mu'}(u') - y_\mu(u)\right) = y' - \hat{y}.$$

We also use that the definition of $y_\mu(u)$ implies that this vector minimizes $y^T Pu + \frac{\mu}{2}\|y - \bar{u}\|^2$ over all $y \in \Delta$. Hence, at $y = y_\mu(u)$ the vector $\nabla_y\left(y^T Pu + \frac{\mu}{2}\|y - \bar{u}\|^2\right)$ has nonnegative inner product with $u - y_\mu(u)$, for all $u \in \Delta$. Since $y_{\mu'}(u') \in \Delta$, we get

$$\left(Pu + \mu\left(y_\mu(\,u) - \bar{u}\right)\right)^T\left(y_{\mu'}(u') - y_\mu(u)\right) \geq 0.$$

Therefore, by using the induction hypothesis, we obtain

$$T_1 \geq \phi_\mu(u) + \frac{\mu}{2\theta^2}\|y' - \;\hat{y}\|^2 \geq \frac{1}{2}\|Py\|^2 + \frac{\mu}{2\theta^2}\|y' - \hat{y}\|^2.$$

Due to (7), with $\bar{u} = 0$, we get

$$\|a\|^2 \geq \|b\|^2 + 2b^T(a - b), \qquad (8)$$

where a and b are arbitrary vectors. Using this and $P^2 = P$, we obtain

$$\frac{1}{2}\|Py\|^2 \geq \frac{1}{2}\|P\hat{y}\|^2 + (P\hat{y})^T P(y - \hat{y}) = \frac{1}{2}\|P\hat{y}\|^2 + (y - \hat{y})^T P\hat{y}.$$

It follows that

$$T_1 \geq \frac{1}{2}\|P\hat{y}\|^2 + (y - \hat{y})^T P\hat{y} + \frac{\mu}{2\theta^2}\|y' - \hat{y}\|^2.$$

For the second bracketed term we write

$$T_2 = -\frac{1}{2}\|P\hat{y}\|^2 + y_{\mu'}(u')^T P\hat{y} = \frac{1}{2}\|P\hat{y}\|^2 + \left(y_{\mu'}(u') - \hat{y}\right)^T P\hat{y}.$$

Substitution yields, while also using $(1-\theta)\mu = \mu'$,

$$\begin{aligned}
\phi_{\mu'}(u') &\geq (1-\theta)T_1 + \theta T_2 \\
&\geq \frac{1}{2}\|P\hat{y}\|^2 + (1-\theta)(y-\hat{y})^T P\hat{y} + \theta\left[y_{\mu'}(u') - \hat{y}\right]^T P\hat{y} + \frac{\mu'}{2\theta^2}\|y' - \hat{y}\|^2 \\
&= \frac{1}{2}\|P\hat{y}\|^2 + \left[(1-\theta)(y-\hat{y})^T + \theta\left(y_{\mu'}(u') - \hat{y}\right)\right]^T P\hat{y} + \frac{\mu'}{2\theta^2}\|y' - \hat{y}\|^2 \\
&= \frac{1}{2}\|P\hat{y}\|^2 + \left[-\hat{y} + \left[(1-\theta)y + \theta y_{\mu'}(u')\right]\right]^T P\hat{y} + \frac{\mu'}{2\theta^2}\|y' - \hat{y}\|^2 \\
&= \frac{1}{2}\|P\hat{y}\|^2 + (y' - \hat{y})^T P\hat{y} + \frac{\mu'}{2\theta^2}\|y' - \hat{y}\|^2.
\end{aligned}$$

According to the definition of k in Algorithm 1, the iteration number is given by $k+1$. We claim that

$$\mu_k = \frac{4}{(k+1)(k+2)}. \tag{9}$$

This is true if $k = 0$, because $\mu_0 = 2$. We proceed with induction on k. Suppose that the claim holds for some $k \geq 0$. Since $\theta_k = 2/(k+3)$, we get

$$\mu_{k+1} = (1-\theta_k)\mu_k = \left(1 - \frac{2}{k+3}\right)\mu_k = \frac{k+1}{k+3}\frac{4}{(k+1)(k+2)} = \frac{4}{(k+2)(k+3)},$$

as desired. As a consequence, we have

$$\frac{\mu'}{\theta^2} = \frac{\mu_{k+1}}{\theta_k^2} = \frac{\frac{4}{(k+2)(k+3)}}{\frac{4}{(k+3)^2}} = \frac{k+3}{k+2} > 1.$$

By also using that P is a projection matrix, we obtain

$$\phi_{\mu'}(u') \geq \frac{1}{2}\|P\hat{y}\|^2 + (y' - \hat{y})^T P\hat{y} + \frac{1}{2}\|P(y' - \hat{y})\|^2 = \frac{1}{2}\|Py'\|^2.$$

Hence the proof of the lemma is complete. ■

Lemma 3.3. If Algorithm 1 does not halt after $k \geq 1$ iterations, then

$$\|Py_k\|^2 \leq \frac{8}{(k+1)^2} - \frac{1}{n^3}.$$

Proof. Since the algorithm does not halt after k iterations, we have $\|P_A y_k\|^2 \leq 2\phi_{\mu_k}(u_k)$ by Lemma 3.2 and $\|P_A u_k\|^2 \geq \frac{1}{n^3}$ by the Appendix in [6]. Also, using Lemma 3.1, we get

$$\|P_A y_k\|^2 \leq 2\phi_{\mu_k}(u_k) \leq 2(\phi_0(u_k) + \mu_k) \leq 2\mu_k - \frac{1}{n^3},$$

where we also used

$$\phi_0(u_k) = -\frac{1}{2}\|Pu_k\|^2 + \min_{u\in\Delta} u^T P u_k \leq -\frac{1}{2}\|Pu_k\|^2 \leq -\frac{1}{2n^3},$$

since Pu_k has at least one entry less than or equal to zero (otherwise, u_k would solve the problem and the algorithm would halt with case 1). Due to (9) it follows that

$$\|P_A y_k\|^2 \leq \frac{8}{(k+1)(k+2)} - \frac{1}{n^3} \leq \frac{8}{(k+1)^2} - \frac{1}{n^3},$$

proving the lemma. ■

Lemma 3.4. Algorithm 1 requires at most $2n\sqrt{n}$ iterations.

Proof. As we established in Section 2, y_k gives rise to a proper cut if $n^3\|P_A y_k\|^2 \leq 1$. This certainly holds if $4n^3 \leq (k+1)^2$, which is equivalent to $k+1 \geq 2n\sqrt{n}$. Hence, the proof is complete. ■

4. Time Complexity per Iteration

In this section, we prove that problem (4) can be solved in $O(n)$ time, provided that $z = P_A y$ has been computed. The problem can then be restated as

$$\min_u \left\{u^T z + \frac{\mu}{2}\|u - \bar{u}\|^2 :\ \mathbf{1}^T u = 1, u \geq 0\right\}. \tag{10}$$

The Lagrange dual of this problem can be simplified to

$$\max_{v,\xi} \left\{\xi - \frac{\mu}{2}\|v\|^2 : \mu v - \xi\mathbf{1} \geq w\right\}, \tag{11}$$

where

$$w = \frac{\mu}{2n}\mathbf{1} - z.$$

Indeed, as we next show we have weak duality. Let u be feasible for (10) and the pair (v, ξ) for (11). Then, the duality gap, i.e., the primal objective value minus the dual objective value, can be reduced as follows:

$$\begin{aligned} u^T z + \frac{\mu}{2}\|u - \bar{u}\|^2 - \left(\xi - \frac{\mu}{2}\|v\|^2\right) \\ &= u^T\left(\frac{\mu}{2n}\mathbf{1} - w\right) + \frac{\mu}{2}\|u\|^2 + \frac{\mu}{2}\|\bar{u}\|^2 - \mu\, u^T\bar{u} - \left(\xi - \frac{\mu}{2}\|v\|^2\right) \\ &= \frac{\mu}{2n} - u^T w + \frac{\mu}{2}\|u\|^2 + \frac{\mu}{2n} - \frac{\mu}{n}u^T\mathbf{1} - \left(\xi - \frac{\mu}{2}\|v\|^2\right) \end{aligned}$$

$$
\begin{aligned}
&= -\, u^T w + \frac{\mu}{2}\|u\|^2 - \left(\xi - \frac{\mu}{2}\|v\|^2\right) \\
&\geq u^T(\xi\mathbf{1} - \mu v) + \frac{\mu}{2}\|u\|^2 - \left(\xi - \frac{\mu}{2}\|v\|^2\right) \\
&= -\mu u^T v + \frac{\mu}{2}\|u\|^2 + \frac{\mu}{2}\|v\|^2 \\
&= \frac{\mu}{2}\|u - v\|^2 \geq 0.
\end{aligned}
$$

This makes clear that the duality gap vanishes if and only if

$$v = u, \qquad u^T(\mu v - \xi\mathbf{1} - w) = 0. \tag{12}$$

Using this, the optimality conditions for $u \in \Delta$ can be expressed in u alone as follows:

$$\mu u - \xi\mathbf{1} \geq w, \qquad u^T(\mu u - \xi\mathbf{1} - w) = 0, \tag{13}$$

for some ξ. Now, let $I := \{i : u_i > 0\}$. Since $u \geq 0$ and $\mu u - \xi\mathbf{1} - w \geq 0$, we deduce from $u^T(\mu u - \xi\mathbf{1} - w) = 0$ that

$$i \in I \;\; \Rightarrow \;\; \mu u_i - \xi = w_i.$$

If $j \notin I$, then $u_j = 0$, whence $\mu u - \xi\mathbf{1} \geq w$ implies $-\xi \geq w_j$. It follows that if $i \in I$ and $j \notin I$, then

$$w_i = \mu u_i - \xi \geq \mu u_i + w_j > w_j, \qquad \forall\, i \in I, \qquad \forall j \notin I. \tag{14}$$

We conclude from this that w_I consists of the $|I|$ largest entries of w and the elements outside I are strictly smaller than those in I. For the moment, assume that w is ordered in nonincreasing order, so that

$$w_1 \geq w_2 \geq \cdots \geq w_n. \tag{15}$$

It then follows that I has the form $I = \{1, \dots, k\}$, for some k, and $w_j < w_k$, for each $j > k$. Now, using $\mathbf{1}^T u = 1$ and $u_j = 0$, for $j > k$, we may write

$$1 = \mathbf{1}^T u = \sum_{i=1}^{k} u_i = \sum_{i=1}^{k} \frac{w_i + \xi}{\mu} = \frac{1}{\mu}\left(k\xi + \sum_{i=1}^{k} w_i\right).$$

From this, we obtain an expression for the optimal value of ξ, namely,

$$\xi = \frac{1}{k}\left(\mu - \sum_{i=1}^{k} w_i\right), \tag{16}$$

and then

$$u_i = \begin{cases} \frac{1}{\mu}(w_i + \xi), & i \leq k \\ 0, & i > k. \end{cases} \tag{17}$$

If $k < n$, then the domain of the primal problem (10) is given by

$$\{u \in \Delta : u_{k+1} = \cdots = u_n = 0\},$$

which expands if k increases. Hence, the optimal objective value occurs if k is maximal. One easily verifies that the vector u determined by (16) and (17) belongs to Δ only if

$$\mu + k w_k > \sum_{i=1}^{k} w_i. \tag{18}$$

Obviously, this holds for $k = 1$, because $\mu > 0$. A crucial observation is that if (18) does not hold for some k, then it does also not hold for larger values of k. Moreover, if it holds for some k, then testing (18) for $k + 1$ amounts to a comparison of $\mu + (k+1)w_{k+1}$ and $\sum_{i=1}^{k} w_i + w_{k+1}$, which requires $O(1)$ operations. Hence, the largest k satisfying (18) can be found in $O(k)$ time. We then know the index set I and hence we can compute ξ and then u_i, for $i \leq k$. We conclude that if w is ordered as in (15), then the solution of (10) requires only $O(n)$ time.

The above reasoning uses the fact that the vector w is already ordered in nonincreasing order; to get w ordered in this way, takes $O(n \log n)$ time. Thus, it follows that problem (11), and also (10), can be solved in $O(n \log n)$ time. The computation of z requires $O(n^2)$ time, which dominates the time for ordering w. Hence, solving problem (4) requires $O(n^2)$ time. As a consequence, the overall time complexity of BP becomes $O\left(nL \cdot n\sqrt{n} \cdot n^2\right) = O(n^{4.5}L)$time.

References

[1] Chubanov, S. (2015), A polynomial projection algorithm for linear feasibility problems, *Mathematical Programming*, 153(2), 687-713.

[2] Khachiyan, L. G. (1979), A polynomial algorithm in linear programming, *Doklady Academii Nauk SSSR*, 244:1093-1096. Translated into English in Soviet Mathematics Doklady 20:191-194.

[3] Nesterov, Y. (2005), Excessive gap technique in nonsmooth convex minimization, *SIAM Journal on Optimization*, 16(1), 235-249.

[4] Peña, J., and Soheili, N. (2016), A deterministic rescaled perceptron algorithm, *Mathematical Programming*, 155(1-2), 497-510.

[5] Roos, K. (2018), An improved version of Chubanov's method for solving a homogeneous feasibility problem, *Optimization Methods and Software*, 33(1), 26-44.

[6] Wei, Z. and Roos, K. (2019), Using Nemirovski's Mirror-Prox method as Basic Procedure in Chubanov's method for solving homogeneous feasibility problems, Manuscript (http://www.optimization-online.org/DB HTML/2018/04/6559.html).