Using Nemirovski's Mirror-Prox method as Basic Procedure in Chubanov's method for solving homogeneous feasibility problems

Zhang Wei (National University of Singapore)
Kees Roos (Delft University of Technology)

June 24, 2020

Abstract

We introduce a new variant of Chubanov's method for solving linear homogeneous systems with positive variables. In the Basic Procedure we use a recently introduced cut in combination with Nemirovski's Mirror-Prox method. We show that the cut requires at most $O(n^3)$ time, just as Chabonov's cut. In an earlier paper it was shown that the new cuts are at least as sharp as those of Chubanov. Our Modified Main Algorithm is in essence the same as Chubanov's Main Algorithm, except that it uses the new Basic Procedure as a subroutine. The new method has $O(n^{4.5}L)$ time complexity. Some computational results are presented in comparison with Gurobi.

Keywords: linear homogeneous systems, algorithm, polynomial-time, mirror-prox method

1 Introduction

We deal with the (primal) problem

$$\begin{array}{l} \text{find } x \in \mathbf{R}^n \\ \text{subject to } Ax = 0,\ x > 0, \end{array} \tag{1}$$

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

$$\begin{array}{l} \text{find } w \in \mathbf{R}^m \\ \text{subject to } A^T w \geq 0,\ A^T w \neq 0. \end{array} \tag{2}$$

According to a variant of Farkas' lemma, due to Stiemke [10], the systems (1) and (2) form an alternative pair in the sense that exactly one of them is feasible [7].

Recently Chubanov [3] proposed a polynomial-time algorithm to deal with this pair of problems. A key ingredient in his algorithm is the so-called Basic Procedure (BP). As a result of the BP one gets either

(i) a feasible solution of (1), or

(ii) a feasible solution for the dual problem (2) of (1), or

(iii) a cut for the feasible region of (1).

The cut in (iii) has the form $x_k \leq \frac{1}{2}$, for some index k, for all solutions of (1) in the unit cube. The cut is used to rescale the matrix A. The rescaling happens in Chubanov's Main Algorithm (MA), which sends the rescaled matrix to his Basic Procedure (BP) until the BP returns (i) or (ii). If the BP yields (ii) then (1) is infeasible.

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 of the matrix A. This follows from a classical result of Khachiyan [5] that gives a positive lower bound on the positive entries of a basic solution of a linear system of equations.[1] The BP in [3] needs at most $4n^3$ iterations per call and $O(n)$ time per iteration, in total $O(n^4)$ time per call. So the overall time complexity becomes $O(n^5L)$. By performing a more careful analysis Chubanov reduced this bound by a factor n to $O(n^4L)$ [3, Theorem 2.1].

In [8, 9] some improvements of Chubanov's method and its analysis were presented. One improvement was the introduction of a new type of cut. It was shown in [9] that the new cuts are at least as tight as the cuts used by Chubanov. An important feature of Chubanov's cut is that is generated in at most $4n^3$ iterations, whereby an iteration requires $O(n)$ time. For the new cuts it is much more difficult to understand that they also can be generated in at $O(n^3)$ iterations. In [9] this was claimed in Lemma 4.2, but unfortunately this lemma is wrong. The first result of this paper is in Section 4, where we prove a slightly modified version of this lemma. It implies that when the new cuts are used in Chubanov's BP at most n^3 iterations are needed to generate such a cut.

The second result is the use of the Mirror-Prox method introduced by Nemirovski [6] to devise a new BP. It generates a cut in only $2n\sqrt{n}$ iterations, with each iteration requiring $O(n^2)$ time. So each call of the new BP needs only $O(n^{3.5}L)$ time. As a result we obtain that the MA - which is in essence the same as in [3] – finds a solution of either (1) or (2) in $O(n^{4.5}L)$ time. This should be compared with the $O(n^5L)$ complexity of Chubanov that we mentioned above. As said before, Chubanov improved his complexity with a factor n by using an additional, ingenious argument. We leave it to future research to investigate if such an improvement is also possible in the current case.

The outline of the paper is as follows. In Section 2 we introduce the notions *inducing primal* and *inducing dual feasibility*, for any vector $y \in \mathbf{R}^n$. We recall in Section 3 how cuts were obtained in [9], as well some properties of these cuts. The main result in Section

[1]Without loss of generality we restrict the feasible set in (1) to the unit cube. So we assume $x \in [0,1]^n$. Chubanov's method maintains a vector d such that $x_i \leq d_i$ for every feasible solution x of $Ax = 0$ with $x \in [0,1]^n$, and for every i. If (1) has a solution it also has a positive solution $x \in [0,1]^n$. Then x is a convex combination of basic feasible solutions (shortly, bfs's) of the system $Ax = 0$, $x \in [0,1]^n$. Since $x > 0$, for each i there must exist a bfs $z^{(i)}$ such that $z_i^{(i)} > 0$, for each i. Then we must have $z_i^{(i)} \geq \tau$, where τ denotes Khachiyan's lower bound for positive entries in bfs's. Since $z^{(i)}$ is a solution of $Ax = 0$, $x \in [0,1]^n$, the definition of d implies $\tau \leq d_i$ for each i. Hence, all entries in d must be at least τ.

4 is Lemma 4.1; it guarantees that Chubanov's BP generates the new cuts in this $O(n^3)$ iterations.

As a preparation for Section 6, where we present the version of Nemirovski's Mirror-Prox method, we derive in Section 5 the saddle point problem that we want to solve. In Section 7 we present the new BP and its analysis, and in Section 8 the MA. Finally, Section 9 contains some computational results; we compare the method presented in this paper with Gurobi, which is considered to be one of the fastest solvers for linear optimization problems. We conclude with some comments in Section 10.

2 Preliminaries

As usual $\mathbf{R}$ denotes the set of real numbers and $\mathbf{R}_+$ the set of nonnegative real numbers. The all-one vector in $\mathbf{R}^n$ is denoted as $\mathbf{1}$ and the $n \times n$ identity matrix as I_n. The null space of matrix A is denoted as $\mathcal{L}$ and $\mathcal{L}^\perp$ denotes the row space of A. So

$$\mathcal{L} := \{x \in \mathbf{R}^n \; : \; Ax = 0\}, \quad \mathcal{L}^\perp := \left\{A^T u \; : \; u \in \mathbf{R}^m\right\}. \tag{3}$$

Since rank $(A) = m$, the inverse of AA^T exists. Hence the orthogonal projections P_A and Q_A of $\mathbf{R}^n$ onto $\mathcal{L}$ and $\mathcal{L}^\perp$ are respectively given by

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

For any $y \in \mathbf{R}^n$ we use the notation

$$y^{\mathcal{L}} = P_A y, \quad y^{\mathcal{L}^\perp} = Q_A y.$$

So $y^{\mathcal{L}}$ and $y^{\mathcal{L}^\perp}$ are the orthogonal components of y in the spaces $\mathcal{L}$ and $\mathcal{L}^\perp$ respectively:

$$y = y^{\mathcal{L}} + y^{\mathcal{L}^\perp}, \quad y^{\mathcal{L}} \in \mathcal{L}, \quad y^{\mathcal{L}^\perp} \in \mathcal{L}^\perp.$$

Lemma 2.1 *Let $y \in \mathbf{R}^n$. If $y^{\mathcal{L}} > 0$ then $y^{\mathcal{L}}$ solves the* primal *problem (1) and if $0 \neq y^{\mathcal{L}^\perp} \geq 0$ then $y^{\mathcal{L}^\perp}$ gives rise to a solution of the* dual *problem (2) in $O(n^3)$ time.*

Proof: Since $y^{\mathcal{L}}$ is the projection of y into the null space of A we have $Ay^{\mathcal{L}} = 0$. Hence the first statement in the lemma is obvious. The second statement follows by noting that $y^{\mathcal{L}^\perp} \in \mathcal{L}^\perp$ implies $y^{\mathcal{L}^\perp} = A^T w$ for some w, thus yielding a solution w of (2). Since A has full row rank, w is uniquely determined by $y^{\mathcal{L}^\perp}$ and can be computed from $y^{\mathcal{L}^\perp}$ in $O(n^3)$ time. □

Because of Lemma 2.1 it becomes natural to say that a vector y *induces primal feasibility* if $y^{\mathcal{L}} > 0$ and that y *induces dual feasibility* if $0 \neq y^{\mathcal{L}^\perp} \geq 0$. If y does not yield a solution of the primal or the dual problem, it is modified in Chubanov's BP until it induces primal or dual feasibility, or can be used to generate a cut. In the next section we discuss how cuts can be obtained from a vector y, as proposed in [9].

3 Cuts and cutting vectors

Since the system (1) is homogeneous in x it has a feasible solution if and only if the system

$$Ax = 0, \ \ x \in (0, 1]^n \tag{4}$$

is feasible (cf. [3, 9]). Moreover, if $d > 0$ is a vector such that $x \leq d$ holds for every feasible solution of (4) then $x' = x/d \leq \mathbf{1}$, where x/d denotes the entrywise quotient of x and d, so $x'_i = x_i/d_i$ for each i. Defining $D = \operatorname{diag}(d)$, it follows that $x' = D^{-1}x$ is feasible for the system

$$ADx' = 0, \ \ x' \in (0, 1]^n, \tag{5}$$

if and only if x is feasible for (4). Problem (5) is of the same type as problem (4): it arises from (4) by *rescaling* A to AD.

Every x satisfying (4) also satisfies $x \leq \mathbf{1}$. We therefore start the algorithm with $d = \mathbf{1}$. The BP successively generates cuts of the form $x_k \leq \frac{1}{2}$. This allows us to divide d_k by 2. While Chubanov used the vector $y^{\mathcal{L}}$ to construct cuts for (4), we showed in [9] that by using the vector $y^{\mathcal{L}^\perp}$ one gets cuts that are at least as tight as the cuts used by Chubanov. Next we recall how this goes.

We introduce the following notations. Let $v = y^{\mathcal{L}^\perp}$. The vector that arises from v by replacing all its negative entries by zero is denoted as v^+. So $v^+ = \max(v, 0)$. Similarly, $v^- = \min(v, 0)$. For each nonzero entry v_k of v we define

$$\sigma_k(v) := \mathbf{1}^T \left(\frac{v}{-v_k} \right)^+, \quad v_k \neq 0. \tag{6}$$

Obviously, $\sigma_k(v) \geq 0$ and $\sigma_k(v) = \sigma_k(-v)$. A useful property is that $\sigma_k(v)$ is homogeneous in v, i.e., $\sigma_k(\beta v) = \sigma_k(v)$ for each $\beta \neq 0$. Moreover, $\sigma_k(v) > 0$ holds if and only if v has entries with a sign opposite to the sign of v_k. Hence, $\sigma_k(v) = 0$ holds if and only if $v_k \neq 0$ and $v \geq 0$ or $v \leq 0$.

Without repeating its (simple) proof we recall the following important lemma [9, Lemma 2.2].

Lemma 3.1 *Let x be feasible for (4) and $v \in \mathcal{L}^\perp$. Then every nonzero entry v_k in v gives rise to an upper bound for x_k, according to*

$$x_k \leq \sigma_k(v). \tag{7}$$

If x is feasible for (4) then we already know that $x_k \leq 1$, for each k. Therefore we say that the cut (7) is void if $\sigma_k(v) \geq 1$. The following lemma slightly differs from Corollary 2.3 in [9], but is in the same spirit.

Lemma 3.2 *Let $0 < \sigma_k(v) \leq 1$ for some k. If $\mathbf{1}^T v \neq 0$ then $\mathbf{1}^T v$ and v_k have the same sign. Otherwise, when $\mathbf{1}^T v = 0$, v_k is the only negative entry in v or the only positive entry and then $\sigma_k(v) = 1$.*

Proof: We first note that the hypothesis in the lemma implies $v \neq 0$ and v has positive and negative entries. We deduce from $\sigma_k(v) \leq 1$ that $\mathbf{1}^T v^+ \leq -v_k$ if $v_k < 0$ and $-\mathbf{1}^T v^- \leq v_k$ if $v_k > 0$. Hence we may write

$$v_k < 0 \quad \Rightarrow \quad \mathbf{1}^T v = \mathbf{1}^T v^+ + \mathbf{1}^T v^- \leq \mathbf{1}^T v^+ + v_k \leq 0,$$

$$v_k > 0 \quad \Rightarrow \quad \mathbf{1}^T v = \mathbf{1}^T v^+ + \mathbf{1}^T v^- \geq v_k + \mathbf{1}^T v^- \geq 0.$$

If $\mathbf{1}^T v \neq 0$ these implications yield the first statement in the lemma. If $\mathbf{1}^T v = 0$, the inequalities in the above two implications hold with equality, whence we have respectively $v_k = \mathbf{1}^T v^-$ and $v_k = \mathbf{1}^T v^+$, which implies the second statement. □

Below it is always assumed that $v = y^{\mathcal{L}^\perp}$ for some y. We define

$$\sigma(v) := \min_k \sigma_k(v).$$

We call y a *cutting vector* if $\sigma(v) < 1$, and a *proper cutting vector* if $\sigma(v) \leq \frac{1}{2}$. By Lemma 3.2 then all v_k's such that $\sigma_k(v) = \sigma(v)$ have the same sign as $\mathbf{1}^T v$. This does not hold if $\sigma(v) \geq 1$. Take, e.g., $v = [1, 1, 1, 1, -2]$. Then $\sigma_k(v) = 2$ for each k. Also note that if $v \neq 0$ and $\sigma(v) = 0$ then $v \geq 0$ or $v \leq 0$. In that case problem (1) is infeasible, by Lemma 2.1.

If y induces primal feasibility then $y^{\mathcal{L}} > 0$. Since then $y^{\mathcal{L}} = P_A y^{\mathcal{L}} > 0$ we may restrict our search for a cutting vector to nonnegative vectors y. Due to homogeneity we may further assume $\mathbf{1}^T y = 1$. We conclude this section with a slight modification of a result from [9] that provides a sufficient condition for y being a cutting vector.

Lemma 3.3 *Let $y \geq 0$ satisfy $\mathbf{1}^T y = 1$ and*

$$\frac{1}{\|y^{\mathcal{L}}\|^2} \geq n^3.$$

Then y is a proper cutting vector.

As mentioned in the Introduction, the proof of the original lemma in [9] is wrong. However, as we show in the next section, the above lemma is correct. As is known each iteration of Chubanov's BP increases $1/\|y^{\mathcal{L}}\|^2$ with at least 1 [3, 8]. Therefore, Lemma 3.3 implies that when Chubanov's BP is equipped with the new cuts, it will require no more than n^3 iterations, despite the fact that the new cut is usually tighter than Chubanov's cut and never less tight [8, Section 2.2].

4 Sufficient condition for cutting vectors

In this section $\mathcal{L}$ denotes a fixed linear space in $\mathbf{R}^n$. We restate Lemma 3.3 as follows.

Lemma 4.1 *Let y be such that $y \geq 0$, $\mathbf{1}^T y = 1$. If $v = y^{\mathcal{L}^\perp}$, $z = y^{\mathcal{L}}$ and $\sigma(v) \geq \frac{1}{2}$, then*

$$\frac{1}{\|z\|^2} < n^3. \tag{8}$$

For the proof we need a couple of lemmas. As a preparation to these lemmas, let y and v be as above and $z = y^{\mathcal{L}}$. We then have

$$y \geq 0,\ \mathbf{1}^T y = 1,\ y = z + v,\ z^T v = 0, \tag{9}$$

with $\sigma(v) \geq \frac{1}{2}$. In order to derive a lower bound for $\|z\|$ we consider the problem

$$\min_{y,z} \left\{ \|z\| \ :\ y \geq 0,\ \mathbf{1}^T y = 1,\ y = z + \beta v,\ z^T v = 0,\ \beta \in \mathbf{R} \right\}, \tag{10}$$

where we assume for the moment that v is given. We introduced an additional variable β because if we require that $\beta = 1$, as in (9), then problem (10) may become infeasible.[2] On the other hand, if $\beta = 0$ then $y = z$, which implies $\mathbf{1}^T z = 1$ and hence $\|z\| \geq 1/\sqrt{n}$. So, if $\beta = 0$ the inequality in Lemma 4.1 certainly holds. We therefore assume below that $\beta \neq 0$. As a consequence, the optimal value of (10) will be positive. Because otherwise $z = 0$ would be optimal. But then $y = \beta v$. Since $0 \neq y \geq 0$ and $\beta \neq 0$ we would have either $v \geq 0$ or $v \leq 0$, which is equivalent to $\sigma(v) = 0$, contradicting the hypothesis in Lemma 4.1.

The proof of Lemma 4.1 uses that problem (10) can be solved analytically by exploiting duality. In the sequel I denotes a subset of the index set $\{1, 2, \ldots, n\}$ and J its complement. The restriction of $v \in \mathbf{R}^n$ to the coordinates in I is denoted as v_I.

Lemma 4.2 *Let v be such that $\sigma(v) > 0$ and let the optimal value of (10) be denoted as α^*. Then there exists a nonempty index set I such that*

$$\alpha^{*2} = \frac{\|v_J\|^2}{|I|\,\|v_J\|^2 + (\mathbf{1}_I^T v_I)^2}$$

Proof: The dual problem of (10) is given by

$$\max_{\lambda,\alpha} \left\{ \alpha \ :\ \lambda \geq \alpha \mathbf{1},\ \|\lambda\| \leq 1,\ \lambda^T v = 0 \right\}. \tag{11}$$

This problem is strictly feasible (take $\lambda = 0$, $\alpha = -1$). The dual objective value is bounded above ($\alpha < 1$, because if $\alpha \geq 1$ then $\lambda \geq \mathbf{1}$, whence $\|\lambda\| \geq \sqrt{n}$). These two properties imply that (10) is solvable (i.e., has an optimal solution) and that the optimal values of (10) and (11) coincide [1, Thm. 2.4.1]. In the same way it follows that the dual problem is solvable. Hence, there exist optimal solutions (y^*, z^*, β^*) and (λ^*, α^*) of (10) and (11), respectively. Moreover, $\|z^*\| = \alpha^* > 0$ and $\beta^* \neq 0$.

[2] If v is given, then (9) is feasible if and only if $\|v\|^2 \leq \max(v)$. For a proof we refer to [9]. It may be worth mentioning that this result is not used in this paper.

Defining $I = \{i : y_i^* > 0\}$ and J its complement, we are going to express the optimal value α^* in the entries of the given vector v and the partition (I, J) of the index set. For this we need to deal with the optimality conditions for (10) and (11).

The above duality properties of (10) and (11) imply $\|z\| \geq \alpha$ whenever (y, z, β) is feasible for (10) and (λ, α) for (11). By using the Cauchy-Schwarz inequality and the feasibility conditions for both problems, this also follows simply by writing

$$\|z\| \geq \|z\| \|\lambda\| \geq \lambda^T z = \lambda^T (y - \beta v) = \lambda^T y \geq \alpha \mathbf{1}^T y = \alpha. \tag{12}$$

As a consequence, at optimality the three inequalities in (12) hold with equality. Thus we obtain

$$\|z^*\| = \|z^*\| \|\lambda^*\|, \quad \|z^*\| \|\lambda^*\| = \lambda^{*T} z^*, \quad y^{*T}(\lambda^* - \alpha^* \mathbf{1}) = 0. \tag{13}$$

Next we draw some conclusions from these three equalities. Since $\|z^*\| > 0$, the first equality in (13) implies $\|\lambda^*\| = 1$ and then the second equality yields

$$\lambda^* = \frac{z^*}{\|z^*\|}.$$

Since $\lambda^* \geq \alpha^* \mathbf{1}$ and $\alpha^* = \|z^*\| > 0$, we get $\lambda^* > 0$, and hence also $z^* = \alpha^* \lambda^* > 0$. Since $y^* \geq 0$ and $\lambda^* - \alpha^* \mathbf{1} \geq 0$, the third equality in (13) implies $y_i^* (\lambda_i^* - \alpha^*) = 0$, for each i. Therefore, due to the definition of I and J, we get

$$\lambda_I^* = \alpha^* \mathbf{1}_I, \quad \lambda_J^* \geq \alpha^* \mathbf{1}_J,$$

where λ_I (λ_J) denotes the restriction of λ to the indices in I (J).

To simplify notation, from now on we omit the superscripts *. So, below it is assumed that (y, z, β) and (λ, α) are optimal solutions for (10) and (11), respectively. From $y_J = 0$ we deduce $z_J = -\beta v_J$. Also using $z = \alpha \lambda$, and $z_I = \alpha \lambda_I = \alpha^2 \mathbf{1}_I$, we get the following expressions:

$$v = \begin{bmatrix} v_I \\ v_J \end{bmatrix}, \quad y = \begin{bmatrix} \alpha^2 \mathbf{1}_I + \beta v_I \\ 0 \end{bmatrix}, \quad z = \begin{bmatrix} \alpha^2 \mathbf{1}_I \\ -\beta v_J \end{bmatrix}, \quad \lambda = \begin{bmatrix} \alpha \mathbf{1}_I \\ -\frac{\beta}{\alpha} v_J \end{bmatrix}. \tag{14}$$

Now the restrictions $\mathbf{1}^T y = 1$ and $z^T v = 0$ yield two linear relations between α^2 and β, namely

$$\begin{aligned} 1 &= \mathbf{1}^T y = \mathbf{1}_I^T y_I = \mathbf{1}_I^T \left(\alpha^2 \mathbf{1}_I + \beta v_I\right) = \alpha^2 |I| + \beta \mathbf{1}_I^T v_I \\ 0 &= z^T v = z_I^T v_I + z_J^T v_J = \alpha^2 \mathbf{1}_I^T v_I + (-\beta v_J)^T v_J = \alpha^2 \mathbf{1}_I^T v_I - \beta \|v_J\|^2. \end{aligned}$$

The determinant of the matrix of coefficients of this linear system of equations equals $(\mathbf{1}_I^T v_I)^2 + |I| \|v_j\|^2$. The last expression is positive[3] and hence the solution is unique. It is given by

$$\beta = \frac{\alpha^2 \mathbf{1}_I^T v_I}{\|v_J\|^2}, \quad \alpha^2 = \frac{\|v_J\|^2}{|I| \|v_J\|^2 + (\mathbf{1}_I^T v_I)^2}. \tag{15}$$

[3]Otherwise $\mathbf{1}_I^T v_I = 0$, whence $1 = \alpha^2 |I|$, which would imply $\alpha^2 = 1/|I| \geq 1/n$.

The last equality implies the lemma. □

It remains to show that $\alpha^2 \geq 1/n^3$ for every vector v that has entries with different signs. The expression for α depends highly on the optimal partition (I, J) of the index set. In order to proceed we need to know more about this partition.

To start with we point out that the common optimal value of (10) and (11) does not change if we replace v by $-v$. So we may assume $\mathbf{1}^T v \geq 0$. Next we establish that then $\beta > 0$. This goes as follows. We derive from $\|\lambda\| = 1$ and $\lambda \geq \alpha \mathbf{1}$ that $1 = \sum_{i=1}^n {\lambda_i}^2 \geq n\alpha^2$, whence $\alpha \leq 1/\sqrt{n}$. We also have $\mathbf{1}^T \lambda \leq \|\mathbf{1}\| \, \|\lambda\| = \sqrt{n}$. Now taking the inner product with $\mathbf{1}$ at both sides of $y = z + \beta v$ we get

$$\beta \mathbf{1}^T v = \mathbf{1}^T y - \mathbf{1}^T z = 1 - \mathbf{1}^T z = 1 - \alpha \mathbf{1}^T \lambda \geq 0.$$

This makes clear that $\beta \mathbf{1}^T v = 0$ can hold only if $\alpha = 1/\sqrt{n}$. In that case there is nothing to prove. So we only need to deal with the case where $\beta \mathbf{1}^T v > 0$. As we assumed $\mathbf{1}^T v \geq 0$ this implies $\mathbf{1}^T v > 0$ and $\beta > 0$. Therefore, due to (15), $\mathbf{1}_I^T v_I > 0$.

By (14) we have $y_I > 0$ if and only if $\alpha^2 + \beta v_i > 0$ for $i \in I$. Also by (14), $\lambda_J \geq \alpha \mathbf{1}_J$ holds if and only if $-\frac{\beta}{\alpha} v_j \geq \alpha$. Since $\alpha > 0$ this holds if and only if $\alpha^2 + \beta v_j \leq 0$ for $j \in J$. Thus we obtain, by also using the expression for β in (15),

$$v_i > -\frac{\|v_J\|^2}{\mathbf{1}_I^T v_I} \geq v_j, \qquad i \in I, \, j \in J. \tag{16}$$

We conclude from this that the entries in v_I are strictly larger than those in v_J, and they are separated by the number $-\|v_J\|^2 / \mathbf{1}_I^T v_I$. Since this number is negative, v_J contains only negative entries, whence all positive entries of v occur in v_I. Hence it becomes natural to order the entries of v in nonincreasing order. Then we have, for some p,

$$v_1 \geq v_2 \geq \ldots \geq v_p \geq 0 > v_{p+1} \geq \ldots \geq v_n,$$

and for some $q \geq p$, $I = \{1, \ldots, q\}$ and $J = \{q+1, \ldots, n\}$. Then (16) holds if and only if

$$v_q > -\frac{\|v_J\|^2}{\mathbf{1}_I^T v_I} \geq v_{q+1}. \tag{17}$$

Next we show not only that (17) determines q uniquely, and hence also I and J, but also that q can be found in $O(n)$ time. For that purpose we define, for each k ($1 \leq k \leq n$), index sets $I_k := \{1, \ldots, k\}$ and $J_k := \{k+1, \ldots, n\}$, and numbers τ_k, according to

$$\tau_k := \frac{\|v_{J_k}\|^2}{\mathbf{1}_{I_k}^T v_{I_k}} = \frac{\sum_{i=k+1}^n v_i^2}{\sum_{i=1}^k v_i}.$$

Then we have $q = k$ if and only if $k \geq p$ and

$$v_k > -\tau_k \geq v_{k+1}. \tag{18}$$

Note that $k \geq p$ implies $v_{k+1} < 0$.

Since $\mathbf{1}^T v > 0$ and the entries in v_J are negative, we have $\sum_{i=1}^k v_i > 0$ for every k, and hence $\tau_k > 0$. An important observation is that τ_{k+1} can be written as an affine combination of τ_k and v_{k+1}^2, as follows:

$$\tau_{k+1} = \frac{\sum_{i=k+2}^n v_i^2}{\sum_{i=1}^{k+1} v_i} = \frac{-v_{k+1}^2 + \sum_{i=k+1}^n v_i^2}{\sum_{i=1}^{k+1} v_i} = \frac{-v_{k+1}^2 + \tau_k \sum_{i=1}^k v_i}{\sum_{i=1}^{k+1} v_i}. \tag{19}$$

The last expression can be further reduced:

$$\tau_{k+1} = \frac{-v_{k+1}^2 + \tau_k \left(-v_{k+1} + \sum_{i=1}^{k+1} v_i\right)}{\sum_{i=1}^{k+1} v_i} = \tau_k + \frac{-v_{k+1}\left(\tau_k + v_{k+1}\right)}{\sum_{i=1}^{k+1} v_i}.$$

If $k < p$ then $v_{k+1} \geq 0$. With $\tau_k > 0$ we get $\tau_{k+1} \leq \tau_k$. So, if $k < p$ then τ_k is nonincreasing at k. On the other hand, if $p \leq k < n$ then $v_{k+1} < 0$, and hence we have

$$\tau_{k+1} > \tau_k \qquad \Leftrightarrow \qquad \tau_k + v_{k+1} > 0, \qquad p \leq k < n. \tag{20}$$

We claim that one also has

$$\tau_{k+1} > \tau_k \qquad \Leftrightarrow \qquad \tau_{k+1} + v_{k+1} > 0, \qquad p \leq k < n. \tag{21}$$

This follows by writing

$$\begin{aligned}
\tau_{k+1} > \tau_k \qquad &\Leftrightarrow \quad \frac{\sum_{i=k+2}^n v_i^2}{\sum_{i=1}^{k+1} v_i} > \frac{\sum_{i=k+1}^n v_i^2}{\sum_{i=1}^{k} v_i} \\
&\Leftrightarrow \quad \left(\sum_{i=k+2}^n v_i^2\right) \sum_{i=1}^k v_i > \left(\sum_{i=k+1}^n v_i^2\right) \sum_{i=1}^{k+1} v_i \\
&\Leftrightarrow \quad \left(\sum_{i=k+2}^n v_i^2\right) \left(-v_{k+1} + \sum_{i=1}^{k+1} v_i\right) > \left(v_{k+1}^2 + \sum_{i=k+2}^n v_i^2\right) \sum_{i=1}^{k+1} v_i \\
&\Leftrightarrow \quad \left(\sum_{i=k+2}^n v_i^2\right) (-v_{k+1}) > v_{k+1}^2 \sum_{i=1}^{k+1} v_i \\
&\Leftrightarrow \quad \tau_{k+1} > -v_{k+1}
\end{aligned}$$

The following lemma now easily follows from (20) and (21).

Lemma 4.3 *One has*

$$\begin{aligned}
\tau_{k+1} < \tau_k \quad &\text{if} \quad k < p, && (22)\\
\tau_{k+1} > \tau_k \quad &\text{if} \quad p \leq k < q, && (23)\\
\tau_{k+1} \leq \tau_k \quad &\text{if} \quad k \geq q. && (24)
\end{aligned}$$

Proof: We first deal with the case where $k = q$. Recall that $q \geq p$. By (18) we then have

$$\tau_q + \upsilon_q > 0, \qquad \tau_q + \upsilon_{q+1} \leq 0.$$

The first inequality means that $\tau_q > \tau_{q-1}$, by (21), and the second inequality is equivalent to $\tau_{q+1} \leq \tau_q$, by (20). Thus we see that τ_k is increasing at $k = q - 1$ and nonincreasing at $k = q$. Now let k be an arbitrary index such that $k \geq p$. Suppose that τ_k is nonincreasing at k. Then $\tau_{k+1} \leq \tau_k$. Due to (21) we then have $\tau_{k+1} + \upsilon_{k+1} \leq 0$. Since $\upsilon_{k+2} \leq \upsilon_{k+1}$ it follows that $\tau_{k+1} + \upsilon_{k+2} \leq 0$, which means that $\tau_{k+2} \leq \tau_{k+1}$, by (20). So, if τ_k is nonincreasing at $k \geq p$, then τ_k remains nonincreasing if k increases. As we proved that τ_k is nonincreasing at $k = q$, it follows that τ_k is nonincreasing at all $k \geq q$. If τ_k were nonincreasing at some $k < q$, it would yield the contradiction that τ_k is nonincreasing at $k = q$. Hence τ_k must be increasing at all $k < q$. □

We conclude from Lemma 4.3 that q is the first index such that $q \geq p$ and $\tau_q \geq \tau_k$ for all $k \geq p$.

Example 4.4 *By way of example we consider the case where*

$$\upsilon = [5\,;\ 4\,;\ 3\,;\ 2\,;\ 1\,;\ -1\,;\ -2\,;\ -2\,;\ -3\,;\ -4]\,.$$

Table 1: Computation of q.

k	υ_k	τ_k	$\tau_k + \upsilon_{k+1}$	$\tau_{k+1} + \upsilon_{k+1}$
1	5.0000	12.8000	16.8000	9.3333
2	4.0000	5.3333	8.3333	6.2500
3	3.0000	3.2500	5.2500	4.5000
4	2.0000	2.5000	3.5000	3.2667
5	1.0000	2.2667	1.2667	1.3571
6	-1.0000	2.3571	0.3571	0.4167
7	-2.0000	2.4167	0.4167	0.5000
8	-2.0000	2.5000	-0.5000	-0.7143
9	-3.0000	2.2857	-1.7143	-4.0000
10	-4.0000	0.0000	-	-

In this case $p = 5$. For $k \geq p$ the largest value of τ_k occours at $k = 8$. So $q = 8$. In agreement with (20) and (21) the table demonstrates that $\tau_{k+1} + \upsilon_{k+1}$ and $\tau_k + \upsilon_{k+1}$ have the same sign, for each k. Observe that q is the first index for which these expressions are negative.

Lemma 4.5 *With v, $\alpha = \alpha^*$ and I as above, one has*[4]

$$\alpha^2 \geq \frac{\|v^-\|^2}{q\,\|v^-\|^2 + \left(\mathbf{1}^T v^+\right)^2}.$$

Proof: As before, $I = \{1, \ldots, q\}$. Because of Lemma 4.2, the inequality in the lemma holds if and only if

$$\frac{\|v_J\|^2}{|I|\,\|v_J\|^2 + \left(\mathbf{1}_I^T v_I\right)^2} \geq \frac{\|v^-\|^2}{q\,\|v^-\|^2 + \left(\mathbf{1}^T v^+\right)^2}.$$

Since $|I| = q$ this is equivalent to

$$\frac{\|v_J\|^2}{\left(\mathbf{1}_I^T v_I\right)^2} \geq \frac{\|v^-\|^2}{\left(\mathbf{1}^T v^+\right)^2},$$

which can be written as

$$\left(\mathbf{1}^T v^+\right)\tau_q \geq \left(\mathbf{1}_I^T v_I\right)\tau_p.$$

By Lemma 4.3 we have $\tau_q \geq \tau_p > 0$. Since $\mathbf{1}^T v^+ \geq \mathbf{1}_I^T v_I > 0$ the lemma follows. □

Now we are ready to prove Lemma 4.1. According to Lemma 4.5 we have

$$\frac{1}{\alpha^2} \leq q + \frac{\left(\mathbf{1}^T v^+\right)^2}{\|v^-\|^2}.$$

The largest element in v is v_1 and $v_1 > 0$. So $\mathbf{1}^T v^+ \leq p v_1$. By the Cauchy-Schwarz inequality, $\|v^-\|^2 \geq \frac{\left(\mathbf{1}^T v^-\right)^2}{n-p}$ Also using the definition of $\sigma(v)$ we may write

$$\frac{\left(\mathbf{1}^T v^+\right)^2}{\|v^-\|^2} \leq (n-p)\frac{\left(p v_1\right)^2}{\left(\mathbf{1}^T v^-\right)^2} = (n-p)\frac{p^2}{\sigma(v)^2} \leq 4(n-p)p^2,$$

where the last inequality follows from $\sigma(v) \geq \frac{1}{2}$. Also using $q \leq n$, substitution yields

$$\frac{1}{\alpha^2} \leq n + 4(n-p)p^2.$$

Putting $f(p) := n + 4(n-p)p^2$, one easily verifies that $f(p)$ is increasing for $p < 2n/3$ and decreasing for larger values of p. Hence $f(p) \leq n + 16n^3/27$. So $f(p) \leq n^3$ holds if and only if $n + 16n^3/27 \leq n^3$, which is equivalent to $11n^2 \geq 27$. Since $n \geq 2$, because $\sigma(v) > 0$, this completes the proof of Lemma 4.1. □

[4] At this stage the mistake in [9] becomes visible. There we incorrectly assumed that $q = p$, or, more precisely, that v_I consists of the positive entries in v. This, however, yields a dual feasible solution that in general is not optimal.

5 A bilinear saddle point problem

In this section we derive a simple saddle point problem that arises when searching for a cutting vector. In the next section we describe Nemirovski's mirror-prox method for solving that saddle point problem.

According to (8) y is a cutting vector if it satisfies $y \geq 0$, $\mathbf{1}^T y = 1$ and $\|y^{\mathcal{L}}\| \leq 1/n\sqrt{n}$, where $\mathcal{L}$ denotes the null space of A. So, $y^{\mathcal{L}} = P_A y$. We therefore consider the (primal) problem

$$\min\left\{\|P_A y\| \;:\; y \geq 0,\, \mathbf{1}^T y = 1\right\}. \tag{25}$$

This is a so-called second-order cone problem. Its dual problem is given by

$$\max\left\{\alpha \;:\; P_A u \geq \alpha \mathbf{1},\, \|u\| \leq 1\right\}. \tag{26}$$

Note that for a given u the best value of α is equal to the smallest entry in $P_A u$. We therefore assume below that α always satisfies $\alpha = \min(P_A u)$. Then the feasible regions of (25) and (26) are respectively the unit simplex Δ and the unit sphere $\mathcal{B}$, as defined by

$$\Delta = \left\{y \in \mathbf{R}^n \;:\; \mathbf{1}^T y = 1,\, y \geq 0\right\}, \quad \mathcal{B} = \left\{u \in \mathbf{R}^n \;:\; \|u\| \leq 1\right\}.$$

If $y \in \Delta$ and $u \in \mathcal{B}$ then we may write

$$\|P_A y\| \geq \|P_A y\|\,\|u\| \geq y^T P_A u \geq y^T(\alpha \mathbf{1}) = \alpha \mathbf{1}^T y = \alpha = \min(P_A u), \tag{27}$$

proving weak duality. We denote the *duality gap* at any pair (y, u) with $y \in \Delta$ and $u \in \mathcal{B}$ as

$$\text{gap}(y, u) := \|P_A y\| - \min(P_A u), \tag{28}$$

Obviously $\text{gap}(y, u) \geq 0$, and $\text{gap}(y, u) = 0$ if and only if y and u are optimal solutions. Since the problems (25) and (26) are strictly feasible with bounded domain, it follows that both problems are solvable and their optimal values are equal (cf. [1]). Hence, at optimality the duality gap vanishes.

Now let y^* and u^* denote optimal solutions of (25) and (26), respectively, and $\alpha^* = \min(P_A u^*)$. Since the duality gap vanishes at optimality it follows that the three inequalities in (27) then hold with equality. In other words,

$$\|P_A y^*\| = \|P_A y^*\|\,\|u^*\| = {y^*}^T P_A u^* = \alpha^* = \min(P_A u^*). \tag{29}$$

Putting $y = y^*$ in (27) it follows that ${y^*}^T P_A u \leq \|P_A y^*\|$ for each $u \in \mathcal{B}$. Similarly, putting $u = u^*$ in (27) we get $y^T P_A u^* \geq \min(P_A u^*)$ for each $y \in \Delta$. Also using (29) we obtain

$${y^*}^T P_A u \leq {y^*}^T P_A u^* \leq y^T P_A u^*, \qquad \forall y \in \Delta, \forall u \in \mathcal{B}. \tag{30}$$

This reveals that (y^*, u^*) is a saddle point for the bilinear function $y^T P_A u$. We conclude this section with an interesting lemma. It makes clear that the saddle point in fact solves both (1) and (2).

Lemma 5.1 *Problem (1) has a solution if and only if* $\alpha^* > 0$*, and then* u^* *solves (1). Otherwise* y^* *uniquely determines a solution of (2).*

Proof: Suppose that (1) has a solution x. Then $x > 0$ and $x \in \mathcal{L}$. The latter implies $P_A x = x$. Defining $u = x/\|x\|$, we have $P_A u = u$ and $\|u\| \leq 1$. Hence the pair (u, α) with $\alpha = \min(u)$ is feasible for (26). Since $u > 0$, also $\alpha > 0$, so this yields a feasible solution for (26) with positive objective value. We conclude that the common optimal value of (25) and (26) is certainly positive. This proves the 'only if'-part of the first statement in the lemma.

To prove the converse part, let us assume that y^* and u^* are optimal solutions and the common optimal value is positive. So we have $\|P_A y^*\| = \min(P_A u^*) = \alpha^* > 0$. Now the first equality in (29) implies $\|u^*\| = 1$. Then the second equality in (29) implies

$$u^* = \frac{P_A y^*}{\|P_A y^*\|}. \tag{31}$$

Multiplication from the left with P_A at both sides yields $P_A u^* = u^*$, since $P_A^2 = P_A$. Since $\min(P_A u^*) = \alpha^* > 0$ we get $P_A u^* > 0$, whence $u^* > 0$. Since $Au^* = AP_A u^* = 0$ it follows that u^* solves (1). This proves the first part of the lemma.

Finally, if the common optimal value equals zero, then we deduce from $P_A y^* = 0$ that $y^* \in \mathcal{L}^\perp$. Since $y^* \in \Delta$, we have $y^* \geq 0$ and $y^* \neq 0$. Due to Lemma 2.1 this completes the proof of the lemma. □

6 Nemirovski's Mirror-Prox method

6.1 Definition of the method

To simplfy notation, from now on we denote P_A simply as P. Following [6], we define the vector field $F(z)$, where $z = (y, u) \in \Delta \times \mathcal{B}$, by

$$F(z) = \begin{bmatrix} \frac{\partial}{\partial y} y^T P_A u \\ -\frac{\partial}{\partial u} y^T P_A u \end{bmatrix} = \begin{bmatrix} Pu \\ -Py \end{bmatrix}.$$

The mirror-prox method can now be stated as follows. It is initialized with

$$z_1 = (y_1, u_1) \in \Delta \times \mathcal{B},$$

and uses the following update in each iteration:

$$\hat{z}_k = \text{argmin}_{z \in \Delta \times \mathcal{B}} \left\{ \gamma F(z_k)^T z + \tfrac{1}{2} \|z - z_k\|^2 \right\} \tag{32}$$

$$z_{k+1} = \text{argmin}_{z \in \Delta \times \mathcal{B}} \left\{ \gamma F(\hat{z}_k)^T z + \tfrac{1}{2} \|z - z_k\|^2 \right\}, \tag{33}$$

where γ is a fixed positive 'step size', with $\gamma \in (0, 1]$.

Below we present the analysis of this method. Not surprisingly, due to the linearity in y and u the analysis goes easier than in [6], but the resulting iteration bound is the same.

6.2 Analysis of the method

Lemma 6.1 *For any $z, \bar{z} \in \Delta \times \mathcal{B}$ we have*

(i) $F(z)^T \bar{z} = -F(\bar{z})^T z$*;*

(ii) $(F(z) - F(\bar{z}))^T (z - \bar{z}) = 0$*;*

(iii) $\|F(z) - F(\bar{z})\|_2 \leq \|z - \bar{z}\|_2$*;*

Proof: With $z = (y; u)$ and $\bar{z} = (\bar{y}; \bar{u})$ we have

$$F(z)^T \bar{z} = \begin{bmatrix} Pu \\ -Py \end{bmatrix}^T \begin{bmatrix} \bar{y} \\ \bar{u} \end{bmatrix} = \bar{y}^T Pu - \bar{u}^T Py = \begin{bmatrix} -P\bar{u} \\ P\bar{y} \end{bmatrix}^T \begin{bmatrix} y \\ u \end{bmatrix} = -F(\bar{z})^T z,$$

proving (i). Taking $\bar{z} = z$ in (i) we get $F(z)^T z = 0$ for each z. As a consequence we have

$$(F(z) - F(\bar{z}))^T (z - \bar{z}) = F(z)^T z - F(z)^T \bar{z} - F(\bar{z})^T z + F(\bar{z})^T \bar{z} = 0,$$

proving (ii). Since P is an orthogonal projection matrix, we may write

$$\|F(z) - F(\bar{z})\|_2 = \left\| \begin{bmatrix} P(u - \bar{u}) \\ -P(y - \bar{y}) \end{bmatrix} \right\|_2 \leq \left\| \begin{bmatrix} u - \bar{u} \\ y - \bar{y} \end{bmatrix} \right\|_2 = \|z - \bar{z}\|_2, \tag{34}$$

proving (iii). □

To measure the distance between z and $\bar{z}$ we introduce the notation

$$\delta(z, \bar{z}) := \tfrac{1}{2} \|z - \bar{z}\|^2. \tag{35}$$

Lemma 6.2 *Let $u, v, w \in \Delta \times \mathcal{B}$. Then*

$$(F(u) - F(v))^T (u - w) \leq \delta(u, v) + \delta(u, w).$$

Proof: By the Cauchy-Schwarz inequality and Lemma 6.1 we get

$$(F(u) - F(v))^T (u - w) \quad \leq \quad \|F(u) - F(v)\| \, \|u - w\| \leq \|u - v\| \, \|u - w\|.$$

The last expression does not exceed $\frac{1}{2} \|u - v\|^2 + \frac{1}{2} \|u - w\|^2$. Hence, due to (35) the lemma follows. □

The following identity is known as the three-point relation of Bregman. It holds for any three vectors u, v and w of the same dimension; its proof is elementary.[5]

$$(u - v)^T (u - w) = \delta(u, v) + \delta(u, w) - \delta(v, w). \tag{36}$$

For the remaining analysis the following result is crucial (cf. [4, Theorem 18.2]).

[5]The rule in (36) can be recognized as a generalization of the cosine rule in Euclidean geometry.

Lemma 6.3

$$\gamma F(\hat{z}_k)^T(\hat{z}_k - z) \le \delta(z, z_k) - \delta(z, z_{k+1}), \quad \forall z \in \Delta \times \mathcal{B}.$$

Proof: We may write

$$\begin{aligned} &\gamma F(\hat{z}_k)^T(\hat{z}_k - z) = \gamma F(\hat{z}_k)^T(\hat{z}_k - z_{k+1}) + \gamma F(\hat{z}_k)^T(z_{k+1} - z) \\ &= \gamma\left(F(\hat{z}_k) - F(z_k)\right)^T(\hat{z}_k - z_{k+1}) + \gamma F(z_k)^T(\hat{z}_k - z_{k+1}) + \gamma F(\hat{z}_k)^T(z_{k+1} - z). \end{aligned}$$

Since $\gamma \le 1$, the first term in the last expression is bounded above by Lemma 6.2:

$$\gamma\left(F(\hat{z}_k) - F(z_k)\right)^T(\hat{z}_k - z_{k+1}) \le \delta(\hat{z}_k, z_k) + \delta(\hat{z}_k, z_{k+1}). \tag{37}$$

Next we derive upper bounds for the other two terms in the above expression for $\gamma F(\hat{z}_k)^T(\hat{z}_k - z)$. The gradient with respect to z of $\gamma F(z_k)^T z + \delta(z, z_k)$ is equal to $\gamma F(z_k) + z - z_k$. Hence optimality of $\hat{z}_k$ in (32) implies

$$\left(\gamma F(z_k) + \hat{z}_k - z_k\right)^T(z - \hat{z}_k) \ge 0, \quad \forall z \in \Delta \times \mathcal{B}.$$

Taking $z = z_{k+1}$ and using (36) this gives an upper bound for the second term:

$$\gamma F(z_k)^T(\hat{z}_k - z_{k+1}) \le \delta(z_{k+1}, z_k) - \delta(z_{k+1}, \hat{z}_k) - \delta(\hat{z}_k, z_k). \tag{38}$$

Finally, the optimality of z_{k+1} in (33) yields an upper bound for the third term:

$$\gamma F(\hat{z}_k)^T(z_{k+1} - z) \le \delta(z, z_k) - \delta(z, z_{k+1}) - \delta(z_k, z_{k+1}). \tag{39}$$

By taking the sum of the upper bounds in (37), (38) and (39) we get the inequality in the lemma. □

For $K = 1, 2, \ldots$ we define

$$\tilde{z}_K := (\tilde{y}_K; \tilde{u}_K) = \frac{1}{K}\sum_{k=1}^{K} \hat{z}_k.$$

So $\tilde{z}_K$ is simply the avarage of all iterates $\hat{z}_k$ with $1 \le k \le K$. The next theorem is the main result in this section. It says that the duality gap at $\tilde{z}_K$ is inversely proportional to the iteration number K.

Theorem 6.4 *By taking $y_1 = \mathbf{1}/n$ and $u_1 = 0$ we obtain*

$$\mathrm{gap}(\tilde{y}_K, \tilde{u}_K) = \|P\tilde{y}_K\| - \min(P\tilde{u}_K) \le \frac{1}{K\gamma}, \quad K \ge 1.$$

Proof: First we observe that

$$F(\hat{z}_k)^T(\hat{z}_k - z) = -F(\hat{z}_k)^T z = \begin{bmatrix} -P\hat{u}_k \\ P\hat{y}_k \end{bmatrix}^T \begin{bmatrix} y \\ u \end{bmatrix} = -y^T P\hat{u}_k + \hat{y}_k^T Pu.$$

Hence we obtain from Lemma 6.3 that for each $k = 1, 2, \ldots, K$,

$$\gamma \left(-y^T P\hat{u}_k + \hat{y}_k^T Pu\right) \leq \delta(z, z_k) - \delta(z, z_{k+1}), \quad \forall z = (y, u) \in \Delta \times \mathcal{B}.$$

By taking the sum of these inequalities for $k = 1, 2, \ldots, K$, and then dividing by $K\gamma$ we get

$$\frac{1}{K\gamma} \sum_{k=1}^{K} \gamma \left(\hat{y}_k^T Pu - y^T P\hat{u}_k\right) \leq \frac{1}{K\gamma} \sum_{k=1}^{K} \left(\delta(z, z_k) - \delta(z, z_{k+1})\right) \leq \frac{\delta(z, z_1)}{K\gamma}.$$

Since $y^T Pu$ is linear in y and linear in u this implies

$$\tilde{y}_K^T Pu - y^T P\tilde{u}_K \leq \frac{\delta(z, z_1)}{K\gamma}, \quad \forall z = (y; u) \in \Delta \times \mathcal{B}.$$

Taking the maximum over all $y \in \Delta$ and $u \in \mathcal{B}$, we obtain

$$\|P\tilde{y}_K\| - \min(P\tilde{u}_K) \leq \max_{z \in \Delta \times \mathcal{B}} \frac{\delta(z, z_1)}{K\gamma}.$$

Finally, since $y_1 = \mathbf{1}/n$ and $u_1 = 0$ we have

$$\max_{z \in \Delta \times \mathcal{B}} \delta(z, z_1) = \tfrac{1}{2} \left(\max_{y \in \Delta} \|y - y_1\|^2 + \max_{u \in \mathcal{B}} \|u - u_1\|^2 \right) \leq \tfrac{1}{2} \left(\frac{n-1}{n} + 1 \right) < 1.$$

This completes the proof. □

7 The Mirror-Prox Basic Procedure

In Section 1 we mentioned that the BP either yields a feasible solution of the primal problem (1) or a feasible solution of the dual problem (2) or a proper cut for (4). We can now state this as follows: the BP generates a vector y such that one of the following three cases occurs:

(i) $y^{\mathcal{L}} > 0$;

(ii) $0 \neq y^{\mathcal{L}^\perp} \geq 0$;

(iii) $\sigma(v) \leq \frac{1}{2}$, where $v = y^{\mathcal{L}^\perp}$.

In the first two cases the status of (1) is clear: in case *(i)* we have a solution of (4), and in case *(ii)* a certificate for its infeasibility. In case *(iii)* y induces for at least one index k an inequality $x_k \leq \frac{1}{2}$ for all solutions of (4). Obviously such an inequality cuts off half of the feasible region of (4). It justifies the update of the current vector d by dividing its k-th entry by 2.

We call the new BP Mirror-Prox Basic Procedure and refer to it with the abbreviation MPBP. Each iteration of the MPBP requires the computation of $\hat{z}_k$ and z_{k+1}, as given by (32) and (33). We claim that this can be done in $O(n^2)$ time. To understand this let $z = (y; u)$, $z_k = (y_k; u_k)$ and similar for $\hat{z}_k$. Since $F(z_k) = (Pu_k; -Py_k)$ we may rewrite (32) as follows

$$\begin{aligned} \hat{z}_k &= \operatorname{argmin}_{z \in \Delta \times \mathcal{B}} \left\{ \gamma F(z_k)^T z + \tfrac{1}{2} \|z - z_k\|^2 \right\} \\ &= \operatorname{argmin}_{y \in \Delta,\, u \in \mathcal{B}} \left\{ \gamma \left(y^T P u_k - u^T P y_k \right) + \tfrac{1}{2} \|y - y_k\|^2 + \tfrac{1}{2} \|u - u_k\|^2 \right\}. \end{aligned}$$

Obviously the computation of $\hat{z}_k$ can be separated into the computation of its components $\hat{y}_k$ and $\hat{u}_k$, as follows:

$$\hat{y}_k = \operatorname{argmin}_{y \in \Delta} \left\{ \gamma y^T P u_k + \tfrac{1}{2} \|y - y_k\|^2 \right\} \tag{40}$$

$$\hat{u}_k = \operatorname{argmin}_{u \in \mathcal{B}} \left\{ -\gamma u^T P y_k + \tfrac{1}{2} \|u - u_k\|^2 \right\}. \tag{41}$$

In a similar way problem (33) can be split into the simpler problems

$$y_{k+1} = \operatorname{argmin}_{y \in \Delta} \left\{ \gamma y^T P \hat{u}_k + \tfrac{1}{2} \|y - y_k\|^2 \right\} \tag{42}$$

$$u_{k+1} = \operatorname{argmin}_{u \in \mathcal{B}} \left\{ -\gamma u^T P \hat{y}_k + \tfrac{1}{2} \|u - u_k\|^2 \right\}. \tag{43}$$

Each of these four problems can be solved in $O(n^2)$ time. Let us demonstrate this for the first problem, which computes $\hat{y}_k$. We first need to compute the objective vector Pu_k. This requires $O(n^2)$ time. After this the solution of the resulting problem can be found in only $O(n)$ time. This is easy to understand for (41) and (43); for (40) and (42) it can be realized by using the approach used in [2].

Algorithm 1: $[\tilde{y}, \tilde{u}, \mathcal{J}, \text{case}] = \text{Mirror-Prox BP}(P_A, y, u)$

1: Initialize: $k = 0$; $y_0 = y$; $\tilde{y} = y$; $u_0 = u$; $\tilde{u} = u$; case $= 0$; $\mathcal{J} = \emptyset$;
2: **while** $\sigma(\tilde{v}) > \frac{1}{2}$ **and** case $= 0$ **do**
3:     **if** $\tilde{y}^{\mathcal{L}} > 0$ **or** $\tilde{u}^{\mathcal{L}} > 0$ **then**
4:         case $= 1$ (problem (1) is feasible)
5:     **else**
6:         **if** $0 \neq \tilde{y}^{\mathcal{L}^\perp} \geq 0$ **or** $0 \neq \tilde{u}^{\mathcal{L}^\perp} \geq 0$ **then**
7:             case $= 2$ (problem (2) is feasible)
8:         **else**
9:
$$\begin{aligned} \hat{y}_k &= \operatorname{argmin}_{y \in \Delta} \left\{ \gamma y^T u_k^{\mathcal{L}} + \tfrac{1}{2} \|y - y_k\|^2 \right\} \\ \hat{u}_k &= \operatorname{argmin}_{u \in \mathcal{B}} \left\{ -\gamma u^T y_k^{\mathcal{L}} + \tfrac{1}{2} \|u - u_k\|^2 \right\} \\ y_{k+1} &= \operatorname{argmin}_{y \in \Delta} \left\{ \gamma y^T \hat{u}_k^{\mathcal{L}} + \tfrac{1}{2} \|y - y_k\|^2 \right\} \\ u_{k+1} &= \operatorname{argmin}_{u \in \mathcal{B}} \left\{ -\gamma u^T \hat{y}_k^{\mathcal{L}} + \tfrac{1}{2} \|u - u_k\|^2 \right\} \end{aligned}$$
10:
$$\begin{aligned} \tilde{y} &= \tfrac{1}{k+1} (k\tilde{y} + \hat{y}_k) \\ \tilde{u} &= \tfrac{1}{k+1} (k\tilde{u} + \hat{u}_k) \\ k &= k + 1 \end{aligned}$$
11:         **end**
12:     **end**
13: **end**
14: **if** case $= 0$ **then**
15:     find a nonempty index set $\mathcal{J}$ such that $\mathcal{J} \subseteq \left\{ j \,:\, \sigma_j(\tilde{v}) \leq \frac{1}{2} \right\}$
16: **end**

The MPBP is presented in Algorithm 1. As usual, we use an iteration counter k as subscript of all relevant vectors. The input is the matrix P_A and vectors $y \in \Delta$ and $u \in \mathcal{B}$. The MPBP is initialized with $\tilde{y} = y_0 = y$ and $\tilde{u} = u_0 = u$. At the end of the k-th iteration $\tilde{y}$ denotes the average of the iterates y_1 to y_k, and similar for $\tilde{u}$.

At the start of the MPBP we first check if $\tilde{y}$ is a cutting vector (line 2). If so we compute in line 15 the proper cuts induced by $\tilde{y}$ and define $\mathcal{J}$ as the set of the corresponding indices and then return to the MA. Otherwise, we check if $\tilde{y}$ is inducing either primal or dual feasibility. If so we return to the MA with case $= 1$ or case $= 2$, respectively. Otherwise we perform an iteration of Nemirovski's mirror-prox method in lines 9-10, and then update the average vectors $\tilde{y}$ and $\tilde{u}$. Then k is increased by 1, and we enter the while loop again.

If $\tilde{y}$ is inducing primal or dual feasibility during the execution of a while loop, the MPBP returns to the MA with case > 0. If this does not happen the MPBP will stop only if $\tilde{y}$ is a cutting vector. If $\gamma = \frac{1}{2}$ this yields the following upper bound for the number of iterations.

Lemma 7.1 *The MPBP stops after at most* $2n\sqrt{n}$ *iterations.*

Proof: After the k-th iteration we have, by Theorem 6.4,

$$\|P\tilde{y}\| - \min(P\tilde{u}) \leq \frac{1}{k\gamma}, \quad k \geq 1.$$

Since $P\tilde{u}$ is not primal feasible, $\min(P\tilde{u}) \leq 0$. Hence we obtain

$$\|P\tilde{y}\| \leq \frac{1}{k\gamma}.$$

According to Lemma 3.3 the vector $\tilde{y}$ is cutting if $\|P\tilde{y}\| < \frac{1}{n\sqrt{n}}$. This certainly holds if

$$\frac{1}{k\gamma} < \frac{1}{n\sqrt{n}}.$$

Since $\gamma = \frac{1}{2}$, it follows that if $k \geq 2n\sqrt{n}$ the MPBP will have stopped. Hence the lemma follows. □

Theorem 7.2 *Each execution of the MPBP needs at most $O(n^{3.5})$ arithmetic operations.*

Proof: We established before that each iteration of the MPBP requires $O(n^2)$ time. The number of iterations being at most $2n\sqrt{n}$, the theorem follows. □

The above theorem makes clear that we have achieved an improvement over the original BP's in [3] and [9] which require $O(n^4)$ arithmetic operations.

8 Modified Main Algorithm

In order to solve (1) one needs to call the MPBP several times by another algorithm, named the Modified Main Algorithm (MMA), a modified version of Chubanov's Main Algorithm [3]. In essence it is the same as in [9]. For the sake of completeness the MMA is decribed in Algorithm 2. The input consists of the matrix A and a positive number τ such that positive entries in basic feasible solutions of (5) have values $\geq \tau$. Due to Khachiyan [5] there exists such a τ satisfying $1/\tau = O(2^L)$, where L denotes the bit size of the matrix A.

Algorithm 2: $[x, \text{case}] = $ MODIFIED MAIN ALGORITHM(A, τ)

1: INITIALIZE: $d = \mathbf{1}$; $y = \mathbf{1}/n$; $u = 0$; $x = 0$; case $= 0$;
2: **while** case $= 0$ **do**
3:     $P_A = I - A^T(AA^T)^{-1}A$
4:     $[y, u, \mathcal{J}, \text{case}] = $ Mirror-Prox BP(P_A, y, u)
5:     **if** case $= 0$ **then**
6:         $d_{\mathcal{J}} = d_{\mathcal{J}}/2$
7:         **if** $\min(d) < \tau$ **then**
8:             case $= 2$
9:         **else**
10:             **if** $\max(d) < 1$ **then**
11:                 case $= 2$
12:             **else**
13:                 $A_{\mathcal{J}} = A_{\mathcal{J}}/2$
14:                 $y_{\mathcal{J}} = y_{\mathcal{J}}/2$
15:             **end**
16:         **end**
17:     **end**
18: **end**
19: **if** case $= 1$ **then**
20:     $D = \text{diag}(d)$
21:     $x = Dy^{\mathcal{L}}$ **or** $x = Du^{\mathcal{L}}$
22: **end**

The MMA is initialized with $d = \mathbf{1}$, $x = 0$, case $= 0$, whereas y is the center of Δ and u the center of $\mathcal{B}$. At the start of the while loop the projection matrix P_A onto the null space of A is computed. Then the MPBP is called with P_A, y and u as input. The MPBP returns to the MMA with $y, u, \mathcal{J}, \text{case}$ as output. If case $= 1$ it has found a y or u that induces primal feasibility and the resulting solution x of the primal problem is constructed from y or u in lines 21-22; if case $= 2$ it halts with $x = 0$, thereby indicating that (1) has no solution.

If case $= 0$ the MPBP has found proper cuts, indexed by the set $\mathcal{J}$. Then the entries in $d_{\mathcal{J}}$ are divided by 2. If then $\min(d) < \tau$ or $\max(d) < 1$, problem (1) must be infeasible, and we put case $= 2$. Otherwise we rescale A and y as in lines 13-16, and restart the while loop again. Hence the MPBP will be called again, and so on.

Theorem 8.1 *The execution of the MMA needs at most $O(n^{4.5}L)$ time.*

Proof: From [3, 9] we recall that the number of iterations of the MMA is $O(nL)$. As a consequence the MPBP is called $O(nL)$ times. Apart from the time spent in the MPBP, the most time consuming part in the MMA is the computation of P_A, which requires $O(n^3)$ time per iteration. By Theorem 7.2 each execution of the MPBP needs at most $O(n^{3.5})$

time. So the time per MMA-iteration is at most $O(n^{3.5})$. Hence in total the MMA will require at most $O(n^{4.5}L)$ time. $\square$

9 Some computational results

We compared the computational performance of the method presented in this paper with Gurobi, which is one of the fastest solvers nowadays, if not the fastest. Like any solver for LO problems, Gurobi cannot handle strict inequalities. As in [9] we used Gurobi with as input the following LO problem, which is equivalent to the homogeneous problem that we want to solve:

$$\min \left\{ 0^T x \ : \ Ax = 0,\, x \geq e \right\}.$$

For each $m \in \{5, 25, 125, 625, 3125\}$ we generated 100 dense integer matrices A of size $m \times n$, with $n = 2m$. The entries in these matrices are randomly generated integers in the interval $[-100, 100]$, uniformly distributed. The 100 problems in each of the 5 classes were first solved by Gurobi, to establish feasibility or infeasibility for each instance. The results are presented in Table 2. As in all subsequent experiments we used $\tau = 10^{-6}$ in Algoritm 2. This value was small enough to distinguish feasible from infeasible problems in all cases. Each line in the table corresponds to a class of 100 problems, with the size of

Table 2: Numerical results of Gurobi.

size(A)		feasible problems				infeasible problems				all
m	n	#	it	barit	time	#	it	barit	time	av. time
5	10	51	4.75	0.00	0.00	49	2.67	0.00	0.00	0.0009
25	50	49	31.29	0.00	0.00	51	22.22	0.00	0.00	0.0017
125	250	51	203.90	0.00	0.03	49	165.90	0.00	0.02	0.0226
625	1250	55	1253.00	0.00	3.08	45	0.00	12.76	1.58	2.4028
3125	6250	47	3128.00	0.00	217.50	53	0.00	11.45	207.21	212.0435

the matrices in that class in the first two columns. We present the computational results separately for feasible and infeasible problems. Gurobi uses simplex iterations and barrier iterations. The next four columns give respectively the number of feasible problems, the average number of simplex iterations, the average number of barrier iterations, and the computational time (in seconds). The next four columns do the same for the infeasible problems. Finally, the last column gives the average solution time of all problems in the class.

After this we solved all problems with the method described in this paper. It should be mentioned that in our implementation we changed line 15 of the MPBP. Instead of the index set $\mathcal{J}$ we used a vector σ with $\sigma_j = \min\{\sigma_j(\tilde{v}),\, 1\}$, for $j = 1,\, 2 \ldots,\, n$. Moreover, in line 6 of the MA we redefined the vector d as follows: $d_j = d_j \sigma_j$ for each index j. So we used not only proper cuts, but also weak cuts. We discarded weak cuts with cut value > 0.9, however. These changes do not affect the theoretical behaviour, but significantly helped to improve the computational performance. The results are in Table 3. This table

Table 3: Results for the new method with threshold value 0.5 for $\sigma(\tilde{y})$.

size(A)		feasible problems					infeasible problems					all
m	n	#	itMA	itBP	$\vert\mathcal{J}\vert$	time	#	itMA	itBP	$\vert\mathcal{J}\vert$	time	av. time
5	10	51	1.90	3.53	0.86	0.00	49	2.47	0.37	8.22	0.00	0.0002
25	50	49	2.31	19.18	2.11	0.00	51	6.18	6.00	31.55	0.00	0.0017
125	250	51	2.57	72.71	6.50	0.02	49	8.24	24.69	120.29	0.02	0.0197
625	1250	55	1.07	36.65	1.18	0.60	45	9.62	62.82	565.82	1.45	0.9826
3125	6250	47	1.00	12.17	0.00	7.47	53	5.00	21.19	3513.88	26.69	17.6554

gives the averages of the numbers of MA iterations, BP iterations, number of cuts, and the solution time, separately for feasible and infeasible problems. Note that the number of feasible and infeasible problems are the same as in Table 2, as it should.

Two more tables are below. In Table 4 we changed the threshold value $\frac{1}{2}$ for $\sigma(\tilde{y})$ in line 2 of the MPBP to 0.1. This yielded the best results we obtained. Obviously, the smaller

Table 4: New method with threshold value 0.1 for $\sigma(\tilde{y})$.

size(A)		feasible problems					infeasible problems					all
m	n	#	itMA	itBP	$\vert\mathcal{J}\vert$	time	#	itMA	itBP	$\vert\mathcal{J}\vert$	time	av. time
5	10	51	1.18	5.71	0.35	0.00	49	1.96	1.88	9.09	0.00	0.0003
25	50	49	1.08	24.14	0.71	0.00	51	2.84	10.06	40.45	0.00	0.0017
125	250	51	1.14	88.31	5.23	0.03	49	3.43	44.37	181.88	0.02	0.0232
625	1250	55	1.00	36.02	0.00	0.60	45	3.56	77.07	881.07	1.41	0.9627
3125	6250	47	1.00	12.17	0.00	7.70	53	2.09	21.49	4761.37	16.03	12.1179

the value of this threshold, the 'deeper' the cuts that are generated, thereby decreasing the number of MA iterations, as can be seen in this table.

We also tried smaller threshold values, but that did not yield a positive effect on the solution times. Obviously by making the threshold value very small, e.g. smaller than the number τ, no cuts can be generated at all; the method will then behave as a pure Mirror-Prox method. Therefore we removed in our implementation all lines related to cuts and rescaling, thereby yielding an implementation of the pure Mirror-Prox method. The results are in Table 5.

Table 5: Pure Mirror-Prox method.

size(A)		feasible problems			infeasible problems			all
m	n	#	it	time	#	it	time	av. time
5	10	51	8.55	0.00	49	113.92	0.01	0.0037
25	50	49	28.71	0.00	51	81.06	0.01	0.0047
125	250	51	123.69	0.04	49	151.69	0.05	0.0422
625	1250	55	35.02	0.61	45	135.96	2.14	1.2996
3125	6250	47	11.17	7.49	53	49.81	22.55	15.4744

10 Conclusions

Taking into account that our implementation was in Matlab, and rather straightforward, it seems promising that the new approach competes with Gurobi. It must be admitted, however, that our experiments were conducted on a limited class of dense randomly generated problems and not on sparse problems. Besides this, being not so familiar with Gurobi, maybe a different setting of parameters might improve Gurobi's behavior. Any suggestion in this direction will be more than welcome. But at the moment the conclusion is that the variant of Chubanov's method considered in this paper does quite well. Our computational results also indicate that Nemirovski's Mirror-Prox method for solving the bilinear saddle point problem associated to the problem considered in this paper may profit from cut-generation and rescaling, as follows by comparing the results in Table 4 and Table 5.

References

[1] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Analysis, algorithms, and engineering applications.

[2] P. Brucker. An $O(n)$ algorithm for quadratic knapsack problems. *Oper. Res. Lett.*, 3(3):163–166, 1984.

[3] S. Chubanov. A polynomial projection algorithm for linear feasibility problems. *Mathematical Programming*, 153:687-713, 2015. DOI: 10.1007/s10107-014-0823-8.

[4] N. He. IE 598 Big Data Optimization. University Of Illinois At Urbana-Champaign. Lecture Notes, Fall October 2016.

[5] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, 244:1093–1096, 1979. Translated into English in *Soviet Mathematics Doklady* 20, 191–194.

[6] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.*, 15(1):229–251, 2004.

[7] C. Roos. Linear optimization: Theorems of the alternative. In C.A. Floudas and P.M. Pardalos, editors, *Encyclopedia of Optimization*, volume III, pages 181–184. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001. ISBN 0-19-512594-0.

[8] C. Roos. On Chubanovs method for solving a homogeneous inequality system. In Mehiddin Al-Baali, Lucio Grandinetti, and Anton Purnama, editors, *Numerical Analysis and Optimization. NAO-III, Muscat, Oman, January 2014*, pages 319 – 338, Switzerland, 2015. Springer Proceedings in Mathematics & Statistics. ISBN 978-3-319-17688-8.

[9] C. Roos. An improved version of Chubanov's method for solving a homogeneous feasibility problem. *Optimization Methods & Software*, 33(1):26–44, 2018. DOI: 10.1080/10556788.2017.1368509.

[10] E. Stiemke. Über positive Lösungen homogener linearer Gleichungen. *Mathematische Annalen*, 76:340–342, 1915.