

Rounding of convex sets and efficient gradient methods for linear programming problems

Yu. Nesterov *

January, 2004

Abstract

In this paper we propose new efficient gradient schemes for two non-trivial classes of linear programming problems. These schemes are designed to compute approximate solutions with relative accuracy δ . We prove that the upper complexity bound for both schemes is $O(\frac{\sqrt{n \ln m}}{\delta} \ln n)$ iterations of a gradient-type method, where n and m , ($n < m$), are the sizes of the corresponding linear programming problems. The proposed schemes are based on preliminary computation of an ellipsoidal rounding for some polytopes in R^n . In both cases this computation can be performed very efficiently, in $O(n^2 m \ln m)$ operations at most.

Keywords: Nonlinear optimization, convex optimization, complexity bounds, relative accuracy, fully polynomial approximation schemes, gradient methods, optimal methods.

CORE DISCUSSION PAPER

2004/04

*Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: nesterov@core.ucl.ac.be.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the author.

The paper was produced during the visit of the author to the research centers IFOR and FIM at ETH, Zurich. Their support is kindly acknowledged.

1 Introduction

Motivation. Among the modern methods for solving linear programming problems (LP-problems, for short), the interior-point methods (IPM) are considered as the most efficient ones. However, these methods are based on a heavy machinery. For an LP-problem of the size $m \times n$, ($m > n$), in order to get an approximate solution with *absolute* accuracy ϵ , these methods need to perform

$$O(\sqrt{m} \ln \frac{m}{\epsilon})$$

iterations of a Newton method. Note that for a problem with a dense matrix of linear constraints each iteration can take up to $O(n^2 m)$ operations.

Clearly these bounds leave considerable room for competition from the gradient-type schemes, for which each iteration is much cheaper. However the main drawback of these schemes is their very slow convergence. In general the gradient schemes need $O(\frac{C_0}{\epsilon^2})$ iterations in order to find an ϵ -solution to the problem. In this estimate a heavy dependence on ϵ is coupled with the presence of some constant C_0 , which depends on a norm of the constraint matrix, size of the solution, etc, and which can be uncontrollably large. Hence, until recently, the gradient-type schemes only provided serious competition for IPM on very large transportation-type problems (see [4], [12], and [3]).

However, it was shown recently in [9] that it is possible to use the structure of LP-problems in order to design gradient-type schemes which converge in $O(\frac{C_1}{\epsilon})$ iterations. Moreover it was shown that for some LP-problems the constant C_1 can be found explicitly and that it is reasonably small. In [11] this result was extended to cover minimization schemes for finding an approximate solution with a certain *relative* accuracy. Specifically, it was shown that for some classes of LP-problems it is possible to compute an approximate solution of relative accuracy δ with $O(\frac{\sqrt{m}}{\delta})$ iterations of a gradient-type scheme. Note that for many applications the concept of relative accuracy is very attractive since it adapts automatically to *any* size of solution. So, there is no necessity to fight against big and unknown constants. For many problems in engineering and economics a level of relative accuracy of the order 0.5% – 0.05% is acceptable.

The approach of [11] is applicable to a special *conic* unconstrained minimization problem. That is the minimization of a non-negative homogeneous convex function $f(x)$, $x \in R^n$, on an affine subspace. In order to compute a solution to this problem with some relative accuracy, we need to know a rounding ellipsoid for the subdifferential of $f(x)$ at the origin. In [11] it was shown that for some LP-problems it is possible to use the structure of $f(x)$ in order to compute such an ellipsoid with a radius $O(\sqrt{m})$.

It is well known that there exists a rounding ellipsoid of radius $\sqrt{n}$ for any centrally symmetric set in R^n (see [5]). Moreover, a good approximation to such an ellipsoid can be easily computed (see [6], [1]). It appears that this ellipsoid provides us with a reasonable norm, which allows one to solve corresponding minimization problem up to a certain relative accuracy. Here we analyze two non-trivial classes of LP-problems and show that for both classes approximate solutions with relative accuracy δ can be computed in $O(\frac{\sqrt{n \ln m}}{\delta} \ln n)$ iterations of a gradient-type method.

Note that the preliminary computation of the rounding ellipsoids in both situations is fairly cheap: it takes $O(n^2 m \ln m)$ operations at most. Up to a logarithmic factor, this is the complexity of finding a projection onto a linear subspace in R^m defined by n

linear equations. Since each iteration of a gradient scheme takes $O(nm)$ operations, in the proposed methods the dependence in n and m of the complexity of the preliminary stage is of the higher order than that of optimization stage.

Contents. In Section 2 we present efficient algorithms for computing rounding ellipsoids for different types of convex sets: central symmetric sets (Section 2.1), general sets (Section 2.2), and sign-symmetric sets (Section 2.3). The results presented in Sections 2.1, 2.2 can be found in [6]; however, we provide the algorithms with simple geometrical proofs. In all situations the computation of a rounding ellipsoid of acceptable quality can be carried out in $O(n^2m \ln m)$ operations. It is important that for the sign-symmetric sets the rounding ellipsoid can be chosen to be diagonal. In Section 3 we consider the problem of minimizing the maximal absolute value of several linear functions over an affine subspace. LP-problems of this type arise in regression analysis, or in truss topology design [2]. We show that an appropriate Euclidean metric (defined by a rounding ellipsoid for the subdifferential of the objective function) ensures an $O(\frac{\sqrt{n \ln m}}{\delta} \ln n)$ bound for the number of iterations of a gradient scheme. In Section 4 we consider the problem of finding a solution to a bilinear matrix game, whose matrix has nonnegative coefficients (linear *packing* problem). We show that after appropriate diagonal preconditioning (which can be efficiently computed by the technique of Section 2.3), this problem can be solved in $O(\frac{\sqrt{n \ln m}}{\delta} \ln n)$ iterations of a gradient-type scheme.

Notation. Throughout the paper we work in a finite dimensional vector space E . We let $n \stackrel{\text{def}}{=} \dim E$. Sometimes it is convenient to identify E with R^n , which we treat as a real linear space of column vectors. We denote by E^* the space dual to E ; that is the space of linear functions on E . For $s \in E^*$ and $x \in E$, we denote the inner product $s(x) = x(s) = \langle s, x \rangle$. In coordinate form

$$\langle s, x \rangle = \sum_{i=1}^n s^{(i)} x^{(i)}.$$

A linear operator $G : E \rightarrow E^*$ is positive semidefinite ($G \succeq 0$) if

$$\langle Gx, x \rangle \geq 0, \quad \forall x \in E.$$

It is positive definite ($G \succ 0$) if the above inequality is strict for all nonzero vectors. Any $G \succ 0$ defines a norm on E :

$$\|x\|_G = \langle Gx, x \rangle^{1/2}, \quad x \in E.$$

The dual norm is defined in the usual way:

$$\|s\|_G^* = \sup_x \{ \langle s, x \rangle : \|x\|_G \leq 1 \} = \langle s, G^{-1}s \rangle^{1/2}, \quad s \in E^*.$$

For a closed convex bounded set $C \subset E^*$, $\xi_C(x)$ denotes its support function:

$$\xi_C(x) = \max_{s \in C} \langle s, x \rangle, \quad x \in E.$$

The notation $\partial f(x) \subset E^*$ is used for the subdifferential of convex function $f(\cdot)$ at point $x \in E$. Thus

$$\partial \xi_C(0) = C.$$

Finally, for $g \in E^*$, gg^* denotes the following linear operator:

$$(gg^*)(x) = \langle g, x \rangle \cdot g \in E^*, \quad x \in E.$$

In a coordinate representation, $D(a)$ denotes a diagonal $n \times n$ -matrix with the vector $a \in R^n$ on the diagonal. In this setting $e_k \in R^n$ denotes the k th basis vector, $\bar{e}_n \in R^n$ denotes the vector of all ones. Thus, $I_n \equiv D(\bar{e}_n)$. Notation R_+^n is used for the positive orthant and $\Delta_n \equiv \{x \in R_+^n : \langle \bar{e}_n, x \rangle = 1\}$ denotes the standard simplex.

Acknowledgements. The author would like to thank Hans-Jakob Luthi and Fabian Chudak for numerous stimulating discussions, and Laurence Wolsey for careful reading and suggestions on the improvement of presentation of the results.

2 Computing the rounding ellipsoids

In this section we analyze efficient algorithms for constructing the rounding ellipsoids for different types of convex sets. Throughout the paper we represent an ellipsoid $W_r(v, G) \subset E^*$ in the following form:

$$W_r(v, G) = \{s \in E^* : \|s - v\|_G^* \equiv \langle s, G^{-1}s \rangle^{1/2} \leq r\},$$

where $G \succ 0$ is an operator from E to E^* . If $v = 0$, we often use the notation $W_r(G)$. We say that an ellipsoid $W_1(v, G)$ is a β -rounding for a convex set $C \subset E^*$, ($\beta \geq 1$), if

$$W_1(v, G) \subseteq C \subseteq W_\beta(v, G).$$

We call β the *radius* of the ellipsoidal rounding.

2.1 Convex sets with central symmetry

For an arbitrary $g \in E^*$ consider the set $C_{\pm g}(G) = \text{Conv}\{W_1(G), \pm g\}$. For $\alpha \in [0, 1]$ denote

$$G(\alpha) = (1 - \alpha)G + \alpha gg^*.$$

Lemma 1 *For any $\alpha \in [0, 1]$ the following inclusion holds:*

$$W_1(G(\alpha)) \subset C_{\pm g}(G). \tag{2.1}$$

If $\sigma \stackrel{\text{def}}{=} \frac{1}{n}(\|g\|_G^*)^2 - 1 > 0$, then the function

$$V(\alpha) \stackrel{\text{def}}{=} \ln \frac{\det G(\alpha)}{\det G(0)} = \ln(1 + \alpha(n(1 + \sigma) - 1)) + (n - 1) \ln(1 - \alpha),$$

is maximized at $\alpha^* = \frac{\sigma}{n(1 + \sigma) - 1}$. Moreover,

$$V(\alpha^*) = \ln(1 + \sigma) + (n - 1) \ln \frac{(n-1)(1+\sigma)}{n(1+\sigma)-1} \geq \ln(1 + \sigma) - \frac{\sigma}{1+\sigma} \geq \frac{\sigma^2}{2(1+\sigma)^2}. \tag{2.2}$$

Proof:

For any $x \in E$, we have

$$\begin{aligned}
\xi_{W_1(G(\alpha))}(x) &= \langle G(\alpha)x, x \rangle^{1/2} = [(1-\alpha)\langle Gx, x \rangle + \alpha\langle g, x \rangle^2]^{1/2} \\
&\leq \max\{\langle Gx, x \rangle^{1/2}, |\langle g, x \rangle|\} \\
&= \max\{\xi_{W_1(G)}(x), \xi_{\text{Conv}\{\pm g\}}(x)\} = \xi_{C_{\pm g}(G)}(x).
\end{aligned}$$

Hence inclusion (2.1) is proved.

Furthermore

$$\begin{aligned}
V(\alpha) &= \ln \det(G^{-1/2}G(\alpha)G^{-1/2}) \\
&= \ln \det\left((1-\alpha)I_n + \alpha G^{-1/2}gg^*G^{-1/2}\right) \\
&= \ln(1-\alpha + \alpha(\|g\|_G^*)^2) + (n-1)\ln(1-\alpha) \\
&= \ln(1 + \alpha(n(1+\sigma) - 1)) + (n-1)\ln(1-\alpha).
\end{aligned}$$

Hence the first order optimality condition for function $V(\alpha)$ is as follows:

$$\frac{n-1}{1-\alpha} = \frac{n(1+\sigma)-1}{1+\alpha(n(1+\sigma)-1)}.$$

The optimal solution of this equation is $\alpha^* = \frac{\sigma}{n(1+\sigma)-1}$. Note that

$$\begin{aligned}
V(\alpha^*) &= \ln(1+\sigma) + (n-1)\ln \frac{(n-1)(1+\sigma)}{n(1+\sigma)-1} \\
&= \ln(1+\sigma) - (n-1)\ln\left(1 + \frac{\sigma}{(n-1)(1+\sigma)}\right) \\
&\geq \ln(1+\sigma) - \frac{\sigma}{1+\sigma} \geq \frac{\sigma^2}{2(1+\sigma)^2},
\end{aligned}$$

which is the inequality (2.2). $\square$

In this section we are interested in solving the following problem. Let C be a convex centrally symmetric body, i.e. $\text{int } C \neq \emptyset$, and $x \in C \Leftrightarrow -x \in C$. For a given $\gamma > 1$ we need to find an ellipsoidal rounding for C of radius $\gamma\sqrt{n}$. An initial approximation to the solution of our problem is given by a matrix $G_0 \succ 0$ such that $W_1(G_0) \subseteq C$, and $C \subseteq W_R(G_0)$ for a certain $R \geq 1$.

Let us give an example of such a problem.

Example 1 Consider a set of vectors $a_i \in E^*$, $i = 1, \dots, m$, which span the whole dual space. Let C be the set:

$$C = \text{Conv}\{\pm a_i, i = 1, \dots, m\}. \quad (2.3)$$

Define $G_0 = \frac{1}{m} \sum_{i=1}^m a_i a_i^*$. Note that for any $x \in E$ we have $\xi_C(x) = \max_{1 \leq i \leq m} |\langle a_i, x \rangle|$. Therefore

$$\xi_{W_1(G_0)}(x) = \left[\frac{1}{m} \sum_{i=1}^m \langle a_i, x \rangle^2 \right]^{1/2} \leq \xi_C(x),$$

$$\xi_{W_{m^{1/2}}(G_0)}(x) = m^{1/2} \left[\frac{1}{m} \sum_{i=1}^m \langle a_i, x \rangle^2 \right]^{1/2} \geq \xi_C(x).$$

Thus, $W_1(G_0) \subseteq C \subseteq W_{m^{1/2}}(G_0)$. $\square$

Let us analyze the following algorithmic scheme.

For $k \geq 0$ iterate:

1. Compute $g_k \in C : \|g_k\|_{G_k}^* = r_k \stackrel{\text{def}}{=} \max_g \{\|g\|_{G_k}^* : g \in C\}$.

2. If $r_k \leq \gamma n^{1/2}$ then Stop else (2.4)

$$\alpha_k := \frac{1}{n} \cdot \frac{r_k^2 - n}{r_k^2 - 1}, \quad G_{k+1} := (1 - \alpha_k)G_k + \alpha_k g_k g_k^*.$$

end.

A complexity bound for this scheme is given in the following statement.

Theorem 1 *Let $W_1(G_0) \subseteq C \subseteq W_R(G_0)$ for some $R \geq 1$. Then scheme (2.4) terminates after at most*

$$\frac{2n \ln R}{2 \ln \gamma - 1 + \gamma^{-2}} \quad (2.5)$$

iterations.

Proof:

Note that the coefficient α_k in Step 2 of (2.4) is chosen in accordance with Lemma 1. Since the method runs as long as $\sigma_k \stackrel{\text{def}}{=} \frac{1}{n} r_k^2 - 1 \geq \gamma^2 - 1$, in view of inequality (2.2), we have at each step $k \geq 0$

$$\ln \det G_{k+1} \geq \ln \det G_k + \gamma^{-2} - 1 + 2 \ln \gamma. \quad (2.6)$$

Note that for any $k \geq 0$ we have

$$\begin{aligned} \det(G_k)^{1/2} \cdot \text{vol}_n(W_1(I_n)) &= \text{vol}_n(W_1(G_k)) \leq \text{vol}_n(C) \leq \text{vol}_n(W_R(G_0)) \\ &= R^n \cdot \det(G_0)^{1/2} \cdot \text{vol}_n(W_1(I_n)). \end{aligned}$$

Hence, $\ln \det G_k - \ln \det G_0 \leq 2n \ln R$, and we get bound (2.5) by summing up (2.6) over k . $\square$

Let us estimate the total arithmetical complexity of scheme (2.4) as applied to the particular symmetric convex set (2.3). It appears that in this situation it is reasonable to update the inverse matrices $H_k \stackrel{\text{def}}{=} G_k^{-1}$ recursively, same as the set of values

$$\nu_k^{(i)} = \langle a_i, H_k a_i \rangle, \quad i = 1, \dots, m,$$

which we treat as a vector $\nu_k \in R^m$. The modified variant of scheme (2.4) looks as follows.

A. Compute $H_0 = \left[\frac{1}{m} \sum_{i=1}^m a_i a_i^* \right]^{-1}$ and the vector $\nu_0 \in R^m$.	
B. For $k \geq 0$ iterate: 1. Find i_k such that $\nu_k^{(i_k)} = \max_{1 \leq i \leq m} \nu_k^{(i)}$. Set $r_k = [\nu^{(i_k)}]^{1/2}$. 2. If $r_k \leq \gamma n^{1/2}$ then Stop else 2.1. Set $\sigma_k = \frac{1}{n} r_k^2 - 1$, $\alpha_k = \frac{\sigma_k}{r_k^2 - 1}$, and compute $x_k = H_k a_{i_k}$. 2.2. Update $H_{k+1} := \frac{1}{1 - \alpha_k} \left[H_k - \frac{\alpha_k}{1 + \sigma_k} \cdot x_k x_k^* \right]$. 2.3. Update $\nu_{k+1}^{(i)} := \frac{1}{1 - \alpha_k} \left[\nu_k^{(i)} - \frac{\alpha_k}{1 + \sigma_k} \cdot \langle a_i, x_k \rangle^2 \right]$, $i = 1, \dots, m$. end.	(2.7)

Let us estimate the arithmetical complexity of this scheme. For simplicity we assume that the matrix $A = (a_1, \dots, a_m)$ is dense. We write down only the leading polynomial terms of the complexity of corresponding computations, in which we count only multiplications.

- **Phase A** takes $\frac{mn^2}{2}$ operations to compute the matrix G_0 , plus $\frac{n^3}{6}$ operations to compute its inverse, and $\frac{mn^2}{2}$ more operations to compute the vector ν_0 .
- **Step 2.1** takes n^2 operations.
- **Step 2.2** takes $\frac{n^2}{2}$ operations.
- **Step 2.3** takes mn operations.

Using now the estimate (2.5) with $R = \sqrt{m}$ (see Example 1), we conclude that for $\gamma > 1$ and the centrally symmetric set (2.3), the scheme (2.7) can find a $\gamma\sqrt{n}$ -rounding in

$$\frac{n^2}{6} (n + 6m) + \frac{n^2(2m+3n) \ln m}{2[2 \ln \gamma - 1 + \gamma^{-2}]}$$

arithmetic operations. Note that for a sparse matrix A the complexity of **Phase A** and **Step 2.3** will be much lower.

2.2 General convex sets

For an arbitrary g from E^* , consider the set $C_g(G) = \text{Conv}\{W_1(G), g\}$. Note that the support function of this set is as follows:

$$\xi_{C_g(G)}(x) = \max\{\|x\|_G, \langle g, x \rangle\}, \quad x \in E.$$

Let $r = \|g\|_G^*$ and

$$G(\alpha) = (1 - \alpha)G + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2}\right)^2 \cdot \left(\frac{\alpha}{r}\right)^2\right) \cdot gg^*.$$

Lemma 2 *For all $\alpha \in [0, 1)$, the ellipsoid*

$$E_\alpha = \{s \in E^* : \|s - \frac{r-1}{2r} \cdot \alpha g\|_{G(\alpha)}^* \leq 1\}$$

belongs to the set $C_g(G)$. If $r \geq n$, then the function

$$V(\alpha) \stackrel{\text{def}}{=} \ln \frac{\det G(\alpha)}{\det G(0)} = 2 \ln \left(1 + \alpha \cdot \frac{r-1}{2}\right) + (n-1) \ln(1 - \alpha),$$

is maximized at $\alpha^ = \frac{2}{n+1} \cdot \frac{r-n}{r-1}$. Moreover,*

$$\begin{aligned} V(\alpha^*) &= 2 \ln \frac{r-1}{n+1} + (n-1) \ln \frac{(n-1)(r+1)}{(n+1)(r-1)} \\ &\geq 2 \left[\ln(1 + \sigma) - \frac{\sigma}{1+\sigma} \right] \geq \left(\frac{\sigma}{1+\sigma} \right)^2, \end{aligned} \tag{2.8}$$

where $\sigma = \frac{r-n}{n+1}$.

Proof:

We need to prove that for all $x \in E$

$$\begin{aligned} \xi_{E_\alpha}(x) &\equiv \alpha \cdot \frac{r-1}{2r} \cdot \langle g, x \rangle + \left[(1 - \alpha) \|x\|_G^2 + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2} \right)^2 \cdot \left(\frac{\alpha}{r} \right)^2 \right) \langle g, x \rangle^2 \right]^{1/2} \\ &\leq \xi_{C_g(G)}(x) = \max\{\|x\|_G, \langle g, x \rangle\}. \end{aligned}$$

First, if $\|x\|_G \leq \langle g, x \rangle$, then

$$\xi_{E_\alpha}(x) \leq \alpha \cdot \frac{r-1}{2r} \cdot \langle g, x \rangle + |1 - \alpha \cdot \frac{r-1}{2r}| \cdot \langle g, x \rangle = \langle g, x \rangle.$$

Otherwise, we have $-r\|x\|_G \leq \langle g, x \rangle \leq \|x\|_G$. Note that the value $\xi_{E_\alpha}(x)$ depends on $\langle g, x \rangle$ in a convex way. Therefore its maximum is achieved at the end points of the feasible interval for $\langle g, x \rangle$. At the end point $\langle g, x \rangle = \|x\|_G$ we have already proved that $\xi_{E_\alpha}(x) = \|x\|_G$. Consider now the case $\langle g, x \rangle = -r\|x\|_G$. Then

$$\xi_{E_\alpha}(x) = -\alpha \cdot \frac{r-1}{2} \cdot \|x\|_G + \left[(1 - \alpha) \|x\|_G^2 + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2} \right)^2 \cdot \left(\frac{\alpha}{r} \right)^2 \right) r^2 \|x\|_G^2 \right]^{1/2} = \|x\|_G.$$

Thus, we have proved that $E_\alpha \subseteq C_g(G)$ for any $\alpha \in [0, 1)$.

Further,

$$\begin{aligned}
V(\alpha) &= \ln \det(G^{-1/2}G(\alpha)G^{-1/2}) \\
&= \ln \det \left((1-\alpha)I_n + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2} \right)^2 \cdot \left(\frac{\alpha}{r} \right)^2 \right) G^{-1/2}gg^*G^{-1/2} \right) \\
&= \ln \left(1 - \alpha + \left(\frac{\alpha}{r} + \left(\frac{r-1}{2} \right)^2 \cdot \left(\frac{\alpha}{r} \right)^2 \right) \cdot r^2 \right) + (n-1) \ln(1-\alpha) \\
&= 2 \ln \left(1 + \alpha \cdot \frac{r-1}{2} \right) + (n-1) \ln(1-\alpha).
\end{aligned}$$

Hence, the first order optimality condition for function $V(\alpha)$ is as follows

$$\frac{n-1}{1-\alpha} = \frac{r-1}{1+\alpha \cdot \frac{r-1}{2}}.$$

Thus, the maximum is attained at $\alpha^* = \frac{2}{n+1} \cdot \frac{r-n}{r-1}$. Letting $\sigma = \frac{r-n}{n+1}$, we get

$$\begin{aligned}
V(\alpha^*) &= 2 \ln \left(1 + \alpha^* \cdot \frac{r-1}{2} \right) + (n-1) \ln(1-\alpha^*) \\
&= 2 \ln(1+\sigma) - (n-1) \ln \left(1 + \frac{2(r-n)}{(n-1)(r+1)} \right) \\
&\geq 2 \ln(1+\sigma) - \frac{2(r-n)}{r+1} = 2 \left[\ln(1+\sigma) - \frac{\sigma}{1+\sigma} \right].
\end{aligned}$$

□

In this section we are interested in solving the following problem. Let $C \subset E^*$ be a convex set with nonempty interior. For a given $\gamma > 1$ we need to find a γn -rounding for C . An initial approximation to the solution of this problem is given by a point v_0 and a matrix $G_0 \succ 0$ such that $W_1(v_0, G_0) \subseteq Q \subseteq W_R(v_0, G_0)$ for certain $R \geq 1$. We assume that $n \equiv \dim E \geq 2$.

Let us analyze the following algorithmic scheme.

For $k \geq 0$ iterate:

1. Compute $g_k \in C$: $\|g_k - v_k\|_{G_k}^* = r_k \stackrel{\text{def}}{=} \max_g \{\|g - v_k\|_{G_k}^* : g \in C\}$.

2. If $r_k \leq \gamma n$ then Stop else

$$\alpha_k := \frac{2}{n+1} \cdot \frac{r_k-n}{r_k-1}, \quad v_{k+1} := v_k + \alpha_k \frac{r_k-1}{2r_k} (g_k - v_k),$$

$$G_{k+1} := (1 - \alpha_k)G_k + \left(\frac{\alpha_k}{r_k} + \left(\frac{r_k-1}{2} \right)^2 \cdot \left(\frac{\alpha_k}{r_k} \right)^2 \right) \cdot (g_k - v_k)(g_k - v_k)^*.$$

end.

(2.9)

A complexity bound for this scheme is given in the following statement.

Theorem 2 *Let $W_1(v_0, G_0) \subseteq C \subseteq W_R(v_0, G_0)$ for some $R \geq 1$. Then scheme (2.9) terminates after at most*

$$\frac{(1+2\gamma)^2}{2(\gamma-1)^2} \cdot n \ln R \quad (2.10)$$

iterations.

Proof:

Note that the coefficient α_k , the vector v_{k+1} and the matrix G_{k+1} in Step 2 of (2.9) are chosen in accordance with Lemma 2. Since the method runs as long as

$$\sigma_k \stackrel{\text{def}}{=} \frac{r_k - n}{n+1} \geq \frac{n}{n+1}(\gamma - 1) \geq \frac{2}{3}(\gamma - 1),$$

in view of inequality (2.8) at each step $k \geq 0$ we have

$$\ln \det G_{k+1} \geq \ln \det G_k + \left(\frac{\sigma_k}{1+\sigma_k} \right)^2 \geq \ln \det G_k + \left(\frac{2(\gamma-1)}{1+2\gamma} \right)^2. \quad (2.11)$$

Note that for any $k \geq 0$ we have

$$\begin{aligned} \det(G_k)^{1/2} \cdot \text{vol}_n(W_1(I_n)) &= \text{vol}_n(W_1(G_k)) \leq \text{vol}_n(C) \leq \text{vol}_n(W_R(G_0)) \\ &= R^n \cdot \det(G_0)^{1/2} \cdot \text{vol}_n(W_1(I_n)). \end{aligned}$$

Hence, $\ln \det G_k - \ln \det G_0 \leq 2n \ln R$, and we get bound (2.10) by summing up the inequalities (2.11). $\square$

Note that in the case $C = \text{Conv}\{a_i, i = 1, \dots, m\}$ scheme (2.9) can be implemented efficiently in the style of (2.7). We leave the derivation of this modification and its complexity analysis as an exercise for the reader. The starting rounding ellipsoid for such a set C can be chosen as follows.

Lemma 3 *Assume that the set $C = \text{Conv}\{a_i, i = 1, \dots, m\}$ has nonempty interior. Define*

$$\hat{a} = \frac{1}{m} \sum_{i=1}^m a_i, \quad G = \frac{1}{R^2} \sum_{i=1}^m (a_i - \hat{a})(a_i - \hat{a})^*,$$

where $R = \sqrt{m(m-1)}$. Then $W_1(\hat{a}, G) \subset C \subset W_R(\hat{a}, G)$.

Proof:

For any $x \in E$ and $r > 0$, we have

$$\xi_{W_r(\hat{a}, G)}(x) = \langle \hat{a}, x \rangle + r \|x\|_G = \langle \hat{a}, x \rangle + \frac{r}{R} \left[\sum_{i=1}^m \langle a_i - \hat{a}, x \rangle^2 \right]^{1/2}.$$

Thus we have $\xi_{W_R(\hat{a}, G)}(x) \geq \max_{1 \leq i \leq m} \langle a_i, x \rangle = \xi_C(x)$; hence $W_R(\hat{a}, G) \supset C$. Further, let

$$\tau_i = \langle a_i - \hat{a}, x \rangle, \quad i = 1, \dots, m, \quad \text{and}$$

$$\hat{\tau} = \max_{1 \leq i \leq m} \langle a_i, x \rangle - \langle \hat{a}, x \rangle \geq 0.$$

Note that $\sum_{i=1}^m \tau_i = 0$ and $\tau_i \leq \hat{\tau}$ for all i . Therefore

$$\begin{aligned} \xi_{W_1(\hat{a}, G)}(x) - \langle \hat{a}, x \rangle &\leq \frac{1}{R} \max_{\tau_i} \left\{ \left[\sum_{i=1}^m \tau_i^2 \right]^{1/2} : \sum_{i=1}^m \tau_i = 0, \tau_i \leq \hat{\tau}, i = 1, \dots, m \right\} \\ &= \frac{\hat{\tau}}{R} \sqrt{m(m-1)} = \max_{1 \leq i \leq m} \langle a_i, x \rangle - \langle \hat{a}, x \rangle = \xi_C(x) - \langle \hat{a}, x \rangle. \end{aligned}$$

Thus $W_1(\hat{a}, G) \subset C$. $\square$

2.3 Sign-invariant convex sets

We call a set $C \subset R^n$ *sign-invariant* if, for any point g from C , an arbitrary change of signs of its entries leaves the point inside C . In other words, for any $g \in C \cap R_+^n$, we have

$$B(g) \equiv \{s \in R^n : -g \leq s \leq g\} \subseteq C.$$

Example of such sets are given by unit balls of l_p -norms or of Euclidean norms generated by diagonal matrices.

Clearly any sign-invariant set is centrally symmetric. Thus, in view of Lemma 1, there exists an ellipsoidal rounding of the radius $\sqrt{n}$ for such a set (that is John Theorem [5]). We will see that the important additional feature of sign-invariant sets is that the matrix of the corresponding quadratic form can be chosen to be diagonal.

Let $D \succ 0$ be a diagonal matrix. Let us choose an arbitrary vector $g \in R_+^n \subset E^*$. Let

$$C = \text{Conv} \{W_1(D), B(g)\}, \quad \text{and}$$

$$G(\alpha) = (1 - \alpha)D + \alpha D^2(g).$$

Clearly the set C is sign-invariant. Let

$$V(\alpha) = \ln \frac{\det G(0)}{\det G(\alpha)} = - \sum_{i=1}^n \ln(1 + \alpha(\tau_i - 1)), \quad \alpha \in [0, 1),$$

where $\tau_i = \frac{(g^{(i)})^2}{D^{(i)}}$, $i = 1, \dots, n$. Note that $V(\alpha)$ is a self-concordant function (see, for example, Chapter 4.1 in [8]). For our analysis it is important that

$$V'(0) = n - \sum_{i=1}^n \tau_i = n - (\|g\|_D^*)^2, \quad \text{and} \tag{2.12}$$

$$V''(0) = \sum_{i=1}^n (\tau_i - 1)^2.$$

Lemma 4 *For any $\alpha \in [0, 1)$, $W_1(G(\alpha)) \subseteq C$. Moreover, if $(\|g\|_D^*)^2 \geq n + \sqrt{n}$, then the step*

$$\alpha^* \stackrel{\text{def}}{=} \frac{-V'(0)}{V''(0) + |V'(0)| \cdot [V''(0)]^{1/2}}$$

belongs to $(0, 1)$ and for any $\gamma \in [1, \frac{1}{\sqrt{n}} \|g\|_D^]$,*

$$V(\alpha^*) \leq \ln \left(1 + \frac{\gamma^2 - 1}{\gamma^2} \right) - \frac{\gamma^2 - 1}{\gamma^2} < 0. \tag{2.13}$$

Proof:

For any $x \in R^n \equiv E$, we have

$$\begin{aligned} [\xi_{W_1(G(\alpha))}(x)]^2 &= (1 - \alpha)\langle Dx, x \rangle + \alpha \sum_{i=1}^n (g^{(i)} x^{(i)})^2 \\ &\leq (1 - \alpha)\langle Dx, x \rangle + \alpha \left(\sum_{i=1}^n g^{(i)} \cdot |x^{(i)}| \right)^2 \\ &\leq \left[\max\{\xi_{W_1(D)}(x), \xi_{B(g)}(x)\} \right]^2 = [\xi_C(x)]^2. \end{aligned}$$

Further, let $S = \sum_{i=1}^n \tau_i = (\|g\|_D^*)^2$. By assumption, $S \geq n + \sqrt{n}$. Therefore $V''(0) = \sum_{i=1}^n (\tau_i - 1)^2 \geq n \left(\frac{S}{n} - 1 \right)^2 \geq 1$ and we conclude that $\alpha^* < 1$. Finally, let us estimate from below the Newton decrement of the self-concordant function $V(\cdot)$ at zero. Note that

$$\begin{aligned} V''(0) &\leq \max_{\tau} \left\{ \sum_{i=1}^n (\tau_i - 1)^2 : \sum_{i=1}^n \tau_i = S, \tau_i \geq 0, i = 1 \dots n \right\} \\ &= (S - 1)^2 + n - 1. \end{aligned}$$

Therefore

$$\lambda(0)^2 \stackrel{\text{def}}{=} \frac{(V'(0))^2}{V''(0)} \geq \frac{(S-n)^2}{(S-1)^2+n-1} \geq \frac{n^2(\gamma^2-1)^2}{n^2\gamma^4-2n\gamma^2+n} \geq \left(\frac{\gamma^2-1}{\gamma^2} \right)^2.$$

It remains to note that the step size α^* is exactly the step size of the damped Newton method:

$$\alpha^* = \frac{1}{1+\lambda(0)} \cdot \frac{-V'(0)}{V''(0)}.$$

Thus, $V(\alpha_*) \leq -[\lambda(0) - \ln(1 + \lambda(0))]$ in view of Theorem 4.1.4 [8], and (2.13) follows. $\square$

Corollary 1 *For any sign-symmetric set $C \subset R^n$ with nonempty interior, there exists a diagonal matrix $D \succ 0$ such that*

$$W_1(D) \subseteq C \subseteq W_{\sqrt{n}}(D).$$

Proof:

Note that for R big enough the set $\{D \succeq 0 : W_1(D) \subseteq C \subseteq W_R(D)\}$ is nonempty, closed, and bounded. Therefore the existence follows from the first statement of Lemma 4 and expression (2.12) for $V'(0)$. $\square$

The above corollary is important for us because of the following statement.

Lemma 5 *Let all vectors $a_i \in R^n$, $i = 1, \dots, m$, have nonnegative coefficients. Assume that there exists a diagonal matrix $D \succ 0$ such that*

$$W_1(D) \subseteq \text{Conv} \{B(a_i), i = 1, \dots, m\} \subseteq W_{\gamma\sqrt{n}}(D)$$

for certain $\gamma \geq 1$. Then the function $f(x) = \max_{1 \leq i \leq m} \langle a_i, x \rangle$ satisfies the inequalities

$$\|x\|_D \leq f(x) \leq \gamma\sqrt{n} \cdot \|x\|_D \quad \forall x \in R_+^n. \quad (2.14)$$

Proof:

Consider the function: $\hat{f}(x) = \max_{1 \leq i \leq m} \sum_{j=1}^n a_i^{(j)} |x^{(j)}|$. Note that its subdifferential can be expressed as follows:

$$\partial \hat{f}(0) = \text{Conv} \{B(a_i), i = 1, \dots, m\}.$$

Thus, for any $x \in R^n$ we have

$$\begin{aligned} \|x\|_D &= \max_s \{\langle s, x \rangle : s \in W_1(D)\} \leq \max_s \{\langle s, x \rangle : s \in \partial \hat{f}(0)\} \equiv \hat{f}(x) \\ &\leq \max_s \{\langle s, x \rangle : s \in W_{\gamma\sqrt{n}}(D)\} = \gamma\sqrt{n} \cdot \|x\|_D. \end{aligned}$$

It remains to note that $\hat{f}(x) \equiv f(x)$ for all $x \in R_+^n$. □

Corollary 2 *Let $a_i \in R_+^n$, $i = 1, \dots, m$. Consider the set*

$$\mathcal{F} = \{x \in R_+^n : \langle a_i, x \rangle \leq b_i, i = 1, \dots, m\}$$

with $b_i > 0$, $i = 1, \dots, m$. Then there exists a diagonal matrix $D \succ 0$ such that

$$W_1(D) \cap R_+^n \subset \mathcal{F} \subset W_{\sqrt{n}}(D) \cap R_+^n. \quad (2.15)$$

Proof:

Consider $f(x) = \max_{1 \leq i \leq m} \frac{1}{b_i} \langle a_i, x \rangle$. Then $\mathcal{F} = \{x \in R_+^n : f(x) \leq 1\}$, and the inclusions (2.15) follow from inequalities (2.14). □

In this section we are interested in finding a diagonal ellipsoidal rounding for the following sign-symmetric set:

$$C = \text{Conv} \{B(a_i), i = 1, \dots, m\},$$

where $a_i \in R_+^n \setminus \{0\}$, $i = 1, \dots, m$. Our main assumption on these vectors is as follows:

$$\hat{a} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^m a_i > 0.$$

Let $\hat{D} = D^2(\hat{a})$.

Lemma 6 $W_1(\hat{D}) \subset C \subset W_{m\sqrt{n}}(\hat{D})$.

Proof:

As $\hat{a} \in C$, $W_1(\hat{D}) \subset B(\hat{a}) \subseteq C$. On the other hand,

$$C \subseteq B(m\hat{a}) \subset \left\{x \in R^n : \sum_{i=1}^n \left(\frac{x^{(i)}}{m\hat{a}^{(i)}}\right)^2 \leq n\right\} = W_{m\sqrt{n}}(\hat{D}).$$

□

For the above sign-symmetric set $C \subset R^n$, consider the following algorithmic scheme which finds a diagonal rounding of radius $\gamma\sqrt{n}$ with $\gamma > \left[1 + \frac{1}{\sqrt{n}}\right]^{1/2}$.

Set $D_0 = \hat{D}$.

For $k \geq 0$ **iterate:**

1. Compute $i_k : \|a_{i_k}\|_{D_k}^* = r_k \stackrel{\text{def}}{=} \max_{1 \leq i \leq m} \|a_i\|_{D_k}^*$.

2. If $r_k \leq \gamma\sqrt{n}$ **then Stop else** (2.16)

$$\beta_k := \sum_{j=1}^n \left(\frac{(a_{i_k}^{(j)})^2}{D_k^{(j)}} - 1 \right)^2, \quad \alpha_k := \frac{r_k^2 - n}{\beta_k + (r_k^2 - n)\beta_k^{1/2}},$$

$$D_{k+1} := (1 - \alpha_k)D_k + \alpha_k D^2(a_{i_k}).$$

end.

Note that this scheme applies the rules described in Lemma 4 using notation β_k for $V''(0)$. Therefore, exactly as in Theorem 1 and Theorem 2, we get the following statement.

Theorem 3 For $\gamma \geq \left[1 + \frac{1}{\sqrt{n}}\right]^{1/2}$ the scheme (2.16) terminates at most after

$$\left[\frac{\gamma^2 - 1}{\gamma^2} - \ln \left(1 + \frac{\gamma^2 - 1}{\gamma^2} \right) \right]^{-1} \cdot n(\ln n + 2 \ln m)$$

iterations.

Note that the number of operations during each iteration of the scheme (2.16) is proportional to the number of nonzero elements in the matrix $A = (a_1, \dots, a_m)$.

3 Minimizing the maximum of absolute values of linear functions

Consider the following linear programming problem:

$$\max_{1 \leq i \leq m} |\langle \bar{a}_i, y \rangle - c_i| \quad \rightarrow \quad \min : y \in R^{n-1}. \quad (3.1)$$

Letting $a_i = (\bar{a}_i^T, -c_i)^T$, $i = 1, \dots, m$, $x = (y^T, \tau)^T \in R^n \equiv E$ and $d = e_n$, we can rewrite this problem as a conic unconstrained minimization problem [11]:

$$\text{Find } f^* = \min_x \left\{ f(x) \stackrel{\text{def}}{=} \max_{1 \leq i \leq m} |\langle a_i, x \rangle| : \langle d, x \rangle = 1 \right\}. \quad (3.2)$$

Another application, which can be rewritten in form (3.2), is the Truss Topology Design problem (see, for example, [2] or [11], Section 5.3).

In [11], in order to construct an ellipsoidal rounding for $\partial f(0)$, the use the composite structure of function $f(x)$ was suggested. However, the radius of this rounding was quite large, of the order $O(\sqrt{m})$. Now by (2.4) we can efficiently compute a rounding for this set with radius proportional to $O(\sqrt{n})$. Let us show that this leads to a much more efficient minimization scheme.

Let us fix some $\gamma > 1$. Assume that using the process (2.4) we managed to construct an ellipsoidal rounding for the centrally symmetric set $\partial f(0)$ of radius $\gamma\sqrt{n}$:

$$W_1(G) \subseteq \partial f(0) \equiv \text{Conv} \{\pm a_i, i = 1, \dots, m\} \subseteq W_{\gamma\sqrt{n}}(G).$$

The immediate consequences are:

$$\|x\|_G \leq f(x) \equiv \sup_s \{ \langle s, x \rangle : s \in \partial f(0) \} \leq \gamma\sqrt{n} \cdot \|x\|_G, \quad (3.3)$$

$$\|a_i\|_G^* \leq \gamma\sqrt{n}, \quad i = 1, \dots, m. \quad (3.4)$$

Let us fix now a smoothing parameter $\mu > 0$. Consider the following approximation of the function $f(x)$:

$$f_\mu(x) = \mu \ln \left(\sum_{i=1}^m \left[e^{\langle a_i, x \rangle / \mu} + e^{-\langle a_i, x \rangle / \mu} \right] \right).$$

Clearly $f_\mu(x)$ is convex and infinitely times continuously differentiable on E . Moreover,

$$f(x) \leq f_\mu(x) \leq f(x) + \mu \ln(2m), \quad \forall x \in R^n. \quad (3.5)$$

Finally note that for any point x and any direction h from E we have

$$\begin{aligned} \langle \nabla f_\mu(x), h \rangle &= \sum_{i=1}^m \lambda_\mu^{(i)}(x) \cdot \langle a_i, h \rangle, \\ \lambda_\mu^{(i)}(x) &= \frac{1}{\omega_\mu(x)} \cdot \left(e^{\langle a_i, x \rangle / \mu} - e^{-\langle a_i, x \rangle / \mu} \right), \quad i = 1, \dots, m, \\ \omega_\mu(x) &= \sum_{i=1}^m \left(e^{\langle a_i, x \rangle / \mu} + e^{-\langle a_i, x \rangle / \mu} \right). \end{aligned}$$

Therefore the expression for the Hessian is as follows:

$$\langle \nabla^2 f_\mu(x) h, h \rangle = \frac{1}{\mu} \sum_{i=1}^m \frac{e^{\langle a_i, x \rangle / \mu} + e^{-\langle a_i, x \rangle / \mu}}{\omega_\mu(x)} \cdot \langle a_i, h \rangle^2 - \frac{1}{\mu} \left(\sum_{i=1}^m \lambda_\mu^{(i)}(x) \cdot \langle a_i, h \rangle \right)^2.$$

In view of (3.4), we have

$$\langle \nabla^2 f_\mu(x) h, h \rangle \leq \frac{1}{\mu} \left(\max_{1 \leq i \leq m} \|a_i\|_G^* \right)^2 \cdot \|h\|_G^2 \leq \frac{\gamma^2 n}{\mu} \cdot \|h\|_G^2.$$

This implies that the gradient of function $f_\mu(\cdot)$ is Lipschitz continuous in the metric $\|\cdot\|_G$ with Lipschitz constant $L_\mu = \frac{\gamma^2 n}{\mu}$:

$$\|\nabla f_\mu(x) - \nabla f_\mu(y)\|_G^* \leq L_\mu \|x - y\|_G \quad \forall x, y \in E.$$

Our approach is very similar to that of [11]. Consider the problem

$$\min_x \{\phi(x); x \in Q\}, \quad (3.6)$$

where Q is a closed convex set and a differentiable convex function $\phi(x)$ has a gradient, which is Lipschitz continuous in the Euclidean norm $\|\cdot\|_G$ with constant L . Let us write down an efficient method for solving (3.6) (that is scheme (3.11) in [9]).

Method $S(\phi, L, Q, G, x_0, N)$	
For $k = 0, \dots, N$ do 1. Compute $\nabla\phi(x_k)$. 2. $y_k := \arg \min_{y \in Q} \left[\langle \nabla\phi(x_k), y - x_k \rangle + \frac{L}{2} \ y - x_k\ _G^2 \right]$. 3. $z_k := \arg \min_{z \in Q} \left[\langle \sum_{i=0}^k \frac{i+1}{2} \nabla\phi(x_i), z - x_0 \rangle + \frac{L}{2} \ z - x_0\ _G^2 \right]$. 4. $x_{k+1} := \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$.	(3.7)
Return: $S(\phi, L, Q, G, x_0, N) \equiv y_N$.	

In accordance with Theorem 2 [9], the output of this scheme y_N satisfies the inequality

$$\phi(y_N) - \phi(x_\phi^*) \leq \frac{2L\|x_0 - x_\phi^*\|_G^2}{(N+1)^2}, \quad (3.8)$$

where x_ϕ^* is an optimal solution to problem (3.6).

As in [11], we are going to use the scheme (3.7) in order to compute an approximate solution to (3.2) with a certain relative accuracy $\delta > 0$. Let

$$Q(r) = \{x \in R^n : \langle d, x \rangle = 1, \|x\|_G \leq r\},$$

$$x_0 = \frac{G^{-1}d}{\langle d, G^{-1}d \rangle},$$

$$\tilde{N} = \left\lceil 2e\gamma\sqrt{2n\ln(2m)} \left(1 + \frac{1}{\delta}\right) \right\rceil.$$

Consider the following method.

Set $\hat{x}_0 = x_0$.

For $t \geq 1$ **iterate:**

$$\mu_t := \frac{\delta f(\hat{x}_{t-1})}{2e(1+\delta)\ln(2m)}; \quad L_{\mu_t} := \frac{\gamma^2 n}{\mu_t}; \quad (3.9)$$

$$\hat{x}_t := S\left(f_{\mu_t}, L_{\mu_t}, Q(f(\hat{x}_{t-1})), G, x_0, \tilde{N}\right);$$

if $f(\hat{x}_t) \geq \frac{1}{e}f(\hat{x}_{t-1})$ **then** $T := t$ **and Stop.**

Theorem 4 *The number of points generated by method (3.9) is bounded as follows:*

$$T \leq 1 + \ln(\gamma\sqrt{n}). \quad (3.10)$$

The last point generated satisfies the inequality $f(\hat{x}_T) \leq (1 + \delta)f^$. The total number of lower-level steps in the process (3.9) does not exceed*

$$2\gamma e(1 + \ln(\gamma\sqrt{n}))\sqrt{2n\ln(2m)}\left(1 + \frac{1}{\delta}\right). \quad (3.11)$$

Proof:

Let x^* be an optimal solution to the problem (3.2). Note that all points $\hat{x}_t$ generated by (3.9) are feasible for (3.2). Therefore in view of (3.3)

$$f(\hat{x}_t) \geq f^* \geq \|x^*\|_G.$$

Thus, $x^* \in Q(f(\hat{x}_t))$ for any $t \geq 0$. Let

$$f_t^* = f_{\mu_t}(x_t^*) = \min_x \{f_{\mu_t}(x) : x \in Q(f(\hat{x}_{t-1}))\}.$$

Since $x^* \in Q(f(\hat{x}_t))$, we have in view of (3.5)

$$f_t^* \leq f_{\mu_t}(x^*) \leq f^* + \mu_t \ln(2m).$$

By the first part of (3.5), $f(\hat{x}_t) \leq f_{\mu_t}(\hat{x}_t)$. Note that

$$\|x_0 - x_t^*\|_G \leq \|x_t^*\|_G \leq f(\hat{x}_{t-1}), \quad t \geq 1.$$

In view of (3.8), we have at the last iteration T :

$$\begin{aligned} f(\hat{x}_T) - f^* &\leq f_{\mu_T}(\hat{x}_T) - f_T^* + \mu_T \ln(2m) \\ &\leq \frac{2L_{\mu_T}f^2(\hat{x}_{T-1})}{(N+1)^2} + \mu_T \ln(2m) = \frac{2\gamma^2 n f^2(\hat{x}_{T-1})}{\mu_T(N+1)^2} + \mu_T \ln(2m) \\ &\leq \frac{f^2(\hat{x}_{T-1})\delta^2}{4\mu_T e^2 \ln(2m)(1+\delta)^2} + \mu_T \ln(2m) = 2\mu_T \ln(2m). \end{aligned}$$

Further, in view of the choice of μ_t and the stopping criterion, we have

$$2\mu_T \ln(2m) = \frac{\delta f(\hat{x}_{T-1})}{e(1+\delta)} \leq \frac{\delta f(\hat{x}_T)}{1+\delta}.$$

Thus $f(\hat{x}_T) \leq (1+\delta)f^*$.

It remains to prove the estimate (3.10) for the number of steps of the process. Indeed, by simple induction it is easy to prove that at the beginning of stage t the following inequality holds:

$$\left(\frac{1}{e}\right)^{t-1} f(x_0) \geq f(\hat{x}_{t-1}), \quad t \geq 1.$$

Note that x_0 is the projection of the origin on the hyperplane $\langle d, x \rangle = 1$. Therefore, in view of inequalities (3.3) we have

$$f^* \geq \|x^*\|_G \geq \|x_0\|_G \geq \frac{1}{\gamma\sqrt{n}} f(x_0).$$

Thus at the final step of the scheme we have

$$\left(\frac{1}{e}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \frac{1}{\gamma\sqrt{n}} f(x_0).$$

This leads to the bound (3.10). $\square$

Recall that the preliminary stage of the method (3.9), that is the computation of $\gamma\sqrt{n}$ -rounding for $\partial f(0)$ with relative accuracy $\gamma > 1$, can be performed by procedure (2.4) in

$$\frac{n^2}{6}(n+6m) + \frac{n^2(2m+3n)\ln m}{2[2\ln\gamma-1+\gamma^{-2}]} = O(n^2(n+m)\ln m)$$

arithmetic operations. Since each step of method (3.7) takes $O(mn)$ operations, the complexity of the preliminary stage is dominant if δ is not too small, say $\delta > \frac{1}{\sqrt{n}}$.

4 Bilinear matrix games with non-negative coefficients

Let $A = (a_1, \dots, a_m)$ be an $n \times m$ -matrix with nonnegative coefficients. Consider the problem

$$\text{Find } f^* = \min_x \{f(x) \stackrel{\text{def}}{=} \max_{1 \leq i \leq m} \langle a_i, x \rangle : x \in \Delta_n\}. \quad (4.1)$$

Note that this format fits different standard problem settings. Consider, for example, the *linear packing problem*

$$\text{Find } \psi^* = \max_{y \geq 0 \in R^n} \{\langle c, y \rangle : \langle a_i, y \rangle \leq b^{(i)}, i = 1, \dots, m\},$$

where all entries of vectors a_i are non-negative, $b > 0 \in R^m$, and $c > 0 \in R^n$. Then

$$\begin{aligned} \psi^* &= \max_{y \geq 0 \in R^n} \{\langle c, y \rangle : \max_{1 \leq i \leq m} \frac{1}{b^{(i)}} \langle a_i, y \rangle \leq 1\} = \max_{y \geq 0 \in R^n} \frac{\langle c, y \rangle}{\max_{1 \leq i \leq m} \frac{1}{b^{(i)}} \langle a_i, y \rangle} \\ &= \left[\min_{y \geq 0 \in R^n} \left\{ \max_{1 \leq i \leq m} \frac{1}{b^{(i)}} \langle a_i, y \rangle : \langle c, y \rangle = 1 \right\} \right]^{-1} \\ &= \left[\min_{x \in R^n} \left\{ \max_{1 \leq i \leq m} \frac{1}{b^{(i)}} \langle D^{-1}(c) a_i, x \rangle : x \in \Delta_n \right\} \right]^{-1}. \end{aligned}$$

As usual, we can approximate the objective function $f(x)$ in (4.1) by the following smooth function:

$$f_\mu(x) = \mu \ln \left(\sum_{i=1}^m e^{\langle a_i, x \rangle / \mu} \right).$$

In this case the following relations hold:

$$f(x) \leq f_\mu(x) \leq f(x) + \mu \cdot \ln m, \quad \forall x \in R^n. \quad (4.2)$$

Let

$$\hat{f}(x) = \max_{1 \leq i \leq m} \sum_{j=1}^n a_i^{(j)} |x^{(j)}|.$$

Note that the subdifferential of the homogeneous function $\hat{f}(x)$ at the origin is as follows:

$$\partial f(0) = \text{Conv} \{B(a_i), i = 1, \dots, m\}.$$

In Section 2.3 we have shown that it is possible to compute efficiently a diagonal matrix $D \succ 0$ such that

$$W_1(D) \subseteq \partial \hat{f}(0) \subseteq W_{2\sqrt{n}}(D),$$

(this corresponds to the choice $\gamma = 2$ in scheme (2.16)). In view of Lemma 5, using this matrix we can define a Euclidean norm $\|\cdot\|_D$ such that

$$\|x\|_D \leq f(x) \leq 2\sqrt{n} \cdot \|x\|_D, \quad \forall x \in R_+^n. \quad (4.3)$$

Moreover, in this norm the sizes of all a_i are bounded by $2\sqrt{n}$.

Now, using the same reasoning as in Section 3, we can show that for any x and h from R^n

$$\langle \nabla^2 f_\mu(x) h, h \rangle \leq \frac{4n}{\mu} \cdot \|h\|_D^2.$$

Hence the gradient of this function is Lipschitz continuous with respect to the norm $\|\cdot\|_D$ with Lipschitz constant $\frac{4n}{\mu}$. This implies that function $f_\mu(x)$ can be minimized by the efficient method (3.7).

Let us fix a relative accuracy $\delta > 0$. Let

$$Q(r) = \{x \in \Delta_n : \|x\|_D \leq r\},$$

$$x_0 = \frac{D^{-1} \bar{e}_n}{\langle \bar{e}_n, D^{-1} \bar{e}_n \rangle},$$

$$\tilde{N} = \left\lceil 4e\sqrt{2n \ln m} \left(1 + \frac{1}{\delta}\right) \right\rceil.$$

Consider the following method.

Set $\hat{x}_0 = x_0$.

For $t \geq 1$ **iterate:**

$$\mu_t := \frac{\delta f(\hat{x}_{t-1})}{2e(1+\delta)\ln m}; \quad L_{\mu_t} := \frac{4n}{\mu_t}; \quad (4.4)$$

$$\hat{x}_t := S\left(f_{\mu_t}, L_{\mu_t}, Q(f(\hat{x}_{t-1})), D, x_0, \tilde{N}\right);$$

if $f(\hat{x}_t) \geq \frac{1}{e}f(\hat{x}_{t-1})$ **then** $T := t$ **and Stop.**

The justification of this scheme is very similar to that of (3.9).

Theorem 5 *The number of points generated by method (3.9) is bounded as follows:*

$$T \leq 1 + \ln(2\sqrt{n}). \quad (4.5)$$

The last point generated satisfies the inequality $f(\hat{x}_T) \leq (1 + \delta)f^$. The total number of lower-level steps in the process (3.9) does not exceed*

$$4e(1 + \ln(2\sqrt{n}))\sqrt{2n \ln m} \left(1 + \frac{1}{\delta}\right). \quad (4.6)$$

Proof:

Let x^* be an optimal solution to the problem (4.1). Note that all points $\hat{x}_t$ generated by (4.4) are feasible. Therefore in view of (4.3)

$$f(\hat{x}_t) \geq f^* \geq \|x^*\|_D.$$

Thus, $x^* \in Q(f(\hat{x}_t))$ for any $t \geq 0$. Let

$$f_t^* = f_{\mu_t}(x_t^*) = \min_x \{f_{\mu_t}(x) : x \in Q(f(\hat{x}_{t-1}))\}.$$

Since $x^* \in Q(f(\hat{x}_t))$, in view of (4.2), we have

$$f_t^* \leq f_{\mu_t}(x^*) \leq f^* + \mu_t \ln m.$$

By the first part of (4.2) $f(\hat{x}_t) \leq f_{\mu_t}(\hat{x}_t)$. Note that $\|x_0 - x_t^*\|_D \leq \|x_t^*\|_D \leq f(\hat{x}_{t-1})$ for $t \geq 1$. Thus in view of (3.8), at the last iteration T , we have:

$$\begin{aligned} f(\hat{x}_T) - f^* &\leq f_{\mu_T}(\hat{x}_T) - f_T^* + \mu_T \ln m \leq \frac{2L_{\mu_T}f^2(\hat{x}_{T-1})}{(N+1)^2} + \mu_T \ln m \\ &= \frac{8nf^2(\hat{x}_{T-1})}{\mu_T(N+1)^2} + \mu_T \ln m \leq \frac{f^2(\hat{x}_{T-1})\delta^2}{4\mu_T e^2 \ln m(1+\delta)^2} + \mu_T \ln m = 2\mu_T \ln m. \end{aligned}$$

Further, in view of the choice of μ_t and the stopping criterion, we have

$$2\mu_T \ln m = \frac{\delta f(\hat{x}_{T-1})}{e(1+\delta)} \leq \frac{\delta f(\hat{x}_T)}{1+\delta}.$$

Thus, $f(\hat{x}_T) \leq (1 + \delta)f^*$.

It remains to prove the estimate (4.5) for the number of steps of the process. Indeed, by simple induction it is easy to prove that at the beginning of stage t the following inequality holds:

$$\left(\frac{1}{e}\right)^{t-1} f(x_0) \geq f(\hat{x}_{t-1}), \quad t \geq 1.$$

Note that x_0 is the projection of the origin at the hyperplane $\langle \bar{e}_n, x \rangle = 1$. Therefore in view of inequalities (4.3), we have

$$f^* \geq \|x^*\|_D \geq \|x_0\|_D \geq \frac{1}{2\sqrt{n}} f(x_0).$$

Thus at the final step of the scheme we have

$$\left(\frac{1}{e}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \frac{1}{2\sqrt{n}} f(x_0).$$

This leads to the bound (4.5). □

Thus, we have seen that the scheme (4.4) needs $O\left(\frac{\sqrt{n \ln m}}{\delta} \ln n\right)$ iterations of the gradient scheme (3.7). Since the matrix D is diagonal, each iteration of this scheme is very cheap. Its complexity is proportional to the number of nonzero elements in the matrix A . Note also that in Steps 2 and 3 of scheme (3.7) it is necessary to compute projections onto the set $Q(r)$, which is the intersection of a simplex and a diagonal ellipsoid. However, since D is a diagonal matrix, this can be done in $O(n \ln n)$ operations.

References

- [1] K.M. Anstreicher. Ellipsoidal approximations of convex sets based on the volumetric barrier. CORE Discussion Paper 9745, 1997.
- [2] A. Ben-Tal, A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms; Engineering Applications*. SIAM-MPS Series in Optimization, 2001.
- [3] D. Bienstock. *Potential Function Methods for Approximately Solving Linear Programming Problems. Theory and Practice*. Kluwer Academic Publishers, Boston, 2002.
- [4] M.D. Grigoriadis, L.G. Khachiyan. Fast approximation schemes for convex programs with many blocks and coupling constraints. *SIAM J. Optimization*, 4 (1994), 86–107.
- [5] F. John. Extremum problems with inequalities as subsidiary conditions. In: *Studies and Essays, Presented to R. Courant on his 60th Birthday January 8, 1948*. Wiley Interscience, New York, 187-204.
- [6] L.G. Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, vol.21, No.2, 307-320, 1996.
- [7] Yu. Nesterov, A. Nemirovskii. *Interior point polynomial methods in convex programming: Theory and Applications*, SIAM, Philadelphia, 1994.
- [8] Yu. Nesterov. *Introductory lectures on convex optimization. A basic course*. Kluwer, Boston, 2003.
- [9] Yu. Nesterov. Smooth minimization of non-smooth functions. CORE DP #2003/12, February 2003. Accepted by *Mathematical Programming*.
- [10] Yu. Nesterov. Excessive gap technique in non-smooth convex minimization. CORE DP #2003/35, May 2003. Accepted by *SIAM J. Optimization*.
- [11] Yu. Nesterov. Unconstrained convex minimization in relative scale. CORE DP 2003/96, November 2003.
- [12] S.A. Plotkin, D.B. Shmoys and E. Tardos. Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, 20(1995), 257–301.