Unconstrained convex minimization in relative scale

Yu. Nesterov*

November 2003

Abstract

In this paper we present a new approach to constructing schemes for unconstrained convex minimization, which compute approximate solutions with a certain relative accuracy. This approach is based on a special conic model of the unconstrained minimization problem. Using a structural model of the objective function we can employ the efficient smoothing technique. The fastest of our algorithms solves a linear programming problem with relative accuracy δ in at most $e \cdot \sqrt{m}\,(2 + \ln m) \cdot \left(1 + \frac{1}{\delta}\right)$ iterations of a gradient-type scheme, where m is the largest dimension of the problem and e is the Euler number.

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

*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. Usually, the theoretical justification of convex optimization methods requires problems with bounded feasible sets (see, for example, [1], [5], [6], [7], etc.). Besides the technical convenience, this assumption allows one to introduce a reasonable scale for measuring the absolute accuracy of the approximate solution. In the cases in which the initial problem does not possess this property, some algorithms even require one to bound the domain artificially ("big M" approach). This approach is, perhaps, acceptable for polynomial-time methods, in which the "big M" appears in the complexity estimates only inside a logarithm. However, it is clear that for gradient-type methods, such a strategy cannot work.

In fact it is an old question: Do problems with unbounded feasible sets really arise? And if so, then how should they be treated? In the experience of the author, there is one, very important class of such problems, namely, the problems obtained as Lagrangian relaxations of linear equality constraints. If there were some reasonable bounds on the dual variables for these constraints, then it would be natural to incorporate them into the primal problem also. Then, instead of linear equality constraints in the primal problem, one would have an additional term in the objective function.

Another difficulty lies in choosing the way of bounding unbounded feasible sets. It is hard ti believe that it is possible to find a reasonable localizator a-priori, without collecting additional information on the topology of the problem numerically.

In this paper we suggest an alternative way to treat some unconstrained minimization problems. Indeed, if we cannot say anything reasonable about the size of the solution, then may be it is worth to search for an approximation in *relative* scale? This suggestion may appear naive since in non-smooth convex optimization such algorithms are known only for a rather narrow class of transportation-type problems (so-called *fully-polynomial* approximation schemes, see [3], [10], and [2]). However, it appears that this idea works at least for a special class of *conic* unconstrained minimization problems. These are the problems of minimizing a homogeneous convex function over an affine subspace. In order to be able to compute an approximate solution to this problem with a certain *relative accuracy*, we need to know two John ellipsoids for subdifferential of the objective function evaluated at the origin.

It appears that in many cases all the necessary information about the objective function can be easily obtained from its structure. For example, using this approach, for problems of the type

$$\sum_{i=1}^{m} |\langle a_i, x\rangle + c^{(i)}| \quad \rightarrow \quad \min_{x \in R^n},$$

it is possible to find an approximate solution with relative accuracy δ in $O\left(\frac{\sqrt{m}}{\delta} \ln m\right)$ iterations of a gradient scheme. The most efficient algorithms of that type are based on a *smoothing technique* proposed recently in [8, 9].

Contents. In Section 2 we define the model of our problem, that is the conic form of unconstrained minimization problem. We give several examples of such problems and establish some bounds on the size of optimal solution in terms of a parameter $\alpha \in (0, 1)$, which measures the asphericity of the subdifferential of the objective function. In Section 3 we establish the efficiency estimates for two black-box subgradient schemes as applied to

the problem under consideration. It appears that they can find an approximate solution in $O\left(\frac{1}{\alpha^4\delta^2}\right)$ and $O\left(\frac{1}{\alpha^2\delta^2}\ln\frac{1}{\alpha}\right)$ iterations. The fastest scheme uses a recursive strategy for updating an estimate for the distance to the optimal solution. In Section 4 we introduce a two-level model of the objective function. On one hand, this model is used to define an appropriate Euclidean metric in the primal space. On the other hand, it allows us to apply the smoothing technique [8]. We present two algorithms of this type with complexity estimates $O\left(\frac{1}{\alpha^2\delta}\right)$ and $O\left(\frac{1}{\alpha\delta}\ln\frac{1}{\alpha}\right)$ iterations. The second scheme is based on the recursive strategy considered in Section 3. The last Section 5 is devoted to examples of applications. We specify the total complexity estimates of the fastest algorithm from Section 4 for a linear programming problem, a problem of minimizing the spectral radius of an affine family of symmetric matrices, and a truss topology design problem. We compare these bounds with those of a short-step path-following interior-point scheme. It appears, for example, that in the case of linear programming problem the estimate of the new gradient scheme is better if the required relative accuracy is not too small, namely, if $\delta \geq O\left(\frac{1}{\min\{n,m\}}\right)$.

Notation. We say that the value $f(\bar{x})$ approximates the minimal value $f^* > 0$ with *relative accuracy* δ if

$$f(\bar{x}) \leq (1+\delta)f^*.$$

For a norm $\|\cdot\|$ in R^n, the dual norm is defined in a standard way:

$$\|g\|^* = \max_{x\in R^n}\{\langle g, x\rangle : \|x\| \leq 1\}, \quad g \in R^n.$$

We use the following notation for the balls in $\|\cdot\|$:

$$B_{\|\cdot\|}(r) = \{x \in R^n : \|x\| \leq r\}.$$

Notation $\pi_{Q,\|\cdot\|}(x)$ is used for projection of the point x onto the set Q with respect to the norm $\|\cdot\|$. If no ambiguity arises, the indication of the norm is omitted.

In what follows I_n denotes the unit matrix in R^n, e_i denotes the ith coordinate vector, and $\bar{e}_n$ stands for the vector of all ones. For an $n \times n$ matrix X we denote by $\lambda_1(X), \ldots, \lambda_n(X)$ its spectrum of eigenvalues. Notation $X \succeq 0$ means that a symmetric matrix X is positive semidefinite and $X \succ 0$ means that it is positive definite. Finally, by $\partial f(x)$ we denote the subdifferential of convex function f at the point x.

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 Conic unconstrained minimization problem

The most general form of optimization problem considered in this paper is as follows:

$$\begin{array}{rcl} \text{Find} \quad f^* & = & \min\limits_x \{f(x) : \; x \in \mathcal{L}\}, \\ \mathcal{L} & = & \{x \in R^n : \; Cx = b\}, \end{array} \tag{2.1}$$

where $f(x)$ is a convex homogeneous function with degree of homogeneity equal to one, C is a $p \times n$-matrix, and $b \neq 0$. Without loss of generality we can assume that C has full row rank. Our main assumptions on problem (2.1) are that

$$\operatorname{dom} f \equiv R^n, \quad 0 \in \operatorname{int} \partial f(0). \tag{2.2}$$

It follows that $f^* > 0$ and the problem of finding an approximate solution to (2.1) with a certain *relative* accuracy is well posed. In what follows we call the setting (2.1), (2.2) the *conic* unconstrained minimization problem.

Note that any unconstrained minimization problem

$$\min_{y \in R^{n-1}} \phi(y)$$

with convex objective $\phi(\cdot)$ can be rewritten in the form (2.1) by simple homogenization:

$$x = (y, \tau) \in R^{n-1} \times R^1_+, \quad f(x) = \tau \phi(y/\tau), \quad Cx \equiv \tau, \quad b = 1,$$

(see, for example, [4] for justification). However in general we cannot guarantee that such a function satisfies (2.2). Let us look at the following examples.

Example 1 *Let our initial problem be to find approximately an unconstrained minimum of the function*

$$\phi_1(y) = \max_{1 \leq i \leq m} |\langle a_i, y \rangle + c^{(i)}|, \quad y \in R^{n-1}.$$

Let us introduce $x = \begin{pmatrix} y \\ \tau \end{pmatrix}$, *and* $\hat{a}_i = \begin{pmatrix} a_i \\ c^{(i)} \end{pmatrix}$, $i = 1, \ldots, m$. *Let*

$$A^T = (\hat{a}_1 \ldots, \hat{a}_m), \quad F_1(v) = \max_{1 \leq i \leq m} |v^{(i)}|,$$

$$p = 1, \quad C = (\underbrace{0, \ldots, 0}_{(n-1) \text{ times}}, 1), \quad b = 1.$$

Then for positive τ *we can define*

$$f(x) = \tau \phi_1(y/\tau) \equiv F_1(Ax).$$

Thus, the last description of $f(x)$ *can be extended onto the whole space. In a similar way, for the function*

$$\phi_\infty(y) = \sum_{i=1}^m |\langle a_i, y \rangle + c^{(i)}|, \quad y \in R^{n-1},$$

we can get a representation (2.1), which satisfies (2.2). In this case we use $f(x) = F_\infty(Ax)$ *with*

$$F_\infty(v) = \sum_{i=1}^{m} |v^{(i)}|.$$

However, for function

$$\phi(y) = \max_{1 \le i \le m} [\langle a_i, y \rangle + c^{(i)}], \quad y \in R^{n-1},$$

the above lifting cannot guarantee (2.2). □

Let us fix some norm $\|\cdot\|$ in R^n. Then we can rewrite our main assumption (2.2) in a quantitative form. Let $\gamma_0 \le \gamma_1$ be some positive values satisfying the following condition:

$$B_{\|\cdot\|^*}(\gamma_0) \subseteq \partial f(0) \subseteq B_{\|\cdot\|^*}(\gamma_1). \tag{2.3}$$

Thus, by (2.2) we just assume that such values are well defined. Note that these values depend on the choice of the norm $\|\cdot\|$. However, this choice will always be evident from the context.

Let

$$\alpha = \tfrac{\gamma_0}{\gamma_1} < 1.$$

As we see later, this parameter determines the complexity of finding approximate solutions to problem (2.1) with certain relative accuracy. It may look as if such complexity estimates are data dependent. However, this is not the case. First of all note that in many situations it is reasonable to choose $\|\cdot\|$ as an ellipsoidal norm. In view of the John theorem there always exists a norm such that

$$\alpha \ge \tfrac{1}{n}. \tag{2.4}$$

Moreover, if $\partial f(0)$ is symmetric, that is $f(x) = f(-x)$, $x \in R^n$, then the bound for the ellipsoidal norm is even better:

$$\alpha \ge \tfrac{1}{\sqrt{n}}. \tag{2.5}$$

Of course, it may be difficult to find a norm suitable for a particular objective function $f(x)$. However, in this case we can try to employ our knowledge of its structure.

For example, it may happen that we know a self-concordant barrier $\psi(v)$ for the convex set $\partial f(0)$ and $\psi'(0) = 0$. Then we can use

$$\|v\|^* = \langle v, \psi''(0)v \rangle^{1/2}, \quad \|x\| = \langle [\psi''(0)]^{-1}x, x \rangle^{1/2}.$$

In this case it is possible to choose

$$\gamma_0 = 1, \quad \gamma_1 = \nu + 2\sqrt{\nu},$$

where ν is the parameter of the barrier $\psi(\cdot)$ (see, for example, Theorem 4.2.6 in [7]).

It may happen that the subdifferential $\partial f(0)$ is polyhedral. For some problems the following result is useful.

Lemma 1 *Let* $f(x) = \max\limits_{1 \le i \le m} \langle a_i, x \rangle$. *Assume that the matrix* $A = (a_1, \ldots, a_m)$ *has full row rank and* $\sum\limits_{i=1}^{m} a_i = 0$*; thus* $m > n$. *Then the norm*

$$\|x\| = \left[\sum_{i=1}^{m} \langle a_i, x \rangle^2 \right]^{1/2}$$

is well defined and we can choose $\gamma_1 = 1$ *and* $\gamma_0 = \frac{1}{\sqrt{m(m-1)}}$.

Proof:
Note that the matrix $G = \sum\limits_{i=1}^{m} a_i a_i^T$ is non-degenerate. Then $\|v\|^* = \langle v, G^{-1} v \rangle^{1/2}$ and therefore

$$\begin{aligned} (\|a_i\|^*)^2 &= \langle a_i, G^{-1} a_i \rangle = \max_{x \in R^n} \{ 2\langle a_i, x \rangle - \langle Gx, x \rangle \} = \max_{x \in R^n} \left\{ 2\langle a_i, x \rangle - \sum_{k=1}^{m} \langle a_k, x \rangle^2 \right\} \\ &\le \max_{x \in R^n} \{ 2\langle a_i, x \rangle - \langle a_i, x \rangle^2 \} = 1. \end{aligned}$$

Since $\partial f(0) = \mathrm{Conv}\,\{a_i,\ i = 1, \ldots, m\}$, we can take $\gamma_1 = 1$.

On the other hand, for any $x \in R^n$ we have $\sum\limits_{i=1}^{m} \langle a_i, x \rangle = 0$. Therefore

$$\langle Gx, x \rangle = \sum_{i=1}^{m} \langle a_i, x \rangle^2 \le \max_{s \in R^m} \left\{ \sum_{i=1}^{m} (s^{(i)})^2 :\ \sum_{i=1}^{m} s^{(i)} = 0,\ s^{(i)} \le f(x),\ i = 1, \ldots, m \right\}.$$

The extremum in the above maximization problem is attained, for example, at

$$\hat{s} = f(x) \cdot (\bar{e}_m - m e_1).$$

This means that $\langle Gx, x \rangle \le m(m-1) f^2(x)$. That is $f(x) \ge \frac{\|x\|}{\sqrt{m(m-1}}$, which justifies the choice $\gamma_0 = \frac{1}{\sqrt{m(m-1)}}$. □

The possibility of employing another structural representation of problem (2.1) is discussed in Section 4.

Let us conclude this section with a statement which determines that one can solve the problem (2.1) with a certain relative accuracy. Denote by x_0 the projection of the origin onto the affine subspace $\mathcal{L}$ with respect to the norm $\|\cdot\|$:

$$\|x_0\| = \min_x \{ \|x\| :\ Cx = b \}.$$

Theorem 1
1. For any $x \in R^n$ *we have*

$$\gamma_0 \cdot \|x\| \le f(x) \le \gamma_1 \cdot \|x\|. \tag{2.6}$$

Therefore the function $f(x)$ *is Lipschitz continuous on* R^n *in the norm* $\|\cdot\|$ *with the Lipschitz constant* γ_1. *Moreover,*

$$\alpha f(x_0) \le \gamma_0 \cdot \|x_0\| \le f^* \le f(x_0) \le \gamma_1 \cdot \|x_0\|. \tag{2.7}$$

2. For any optimal solution x^* to (2.1) we have

$$\|x_0 - x^*\| \leq \tfrac{2}{\gamma_0} f^* \leq \tfrac{2}{\gamma_0} f(x_0). \tag{2.8}$$

If the norm $\|\cdot\|$ is Euclidean, then this inequality can be strengthened to:

$$\|x_0 - x^*\| \leq \tfrac{1}{\gamma_0} f^* \leq \tfrac{1}{\gamma_0} f(x_0). \tag{2.9}$$

Proof:
For any $x \in R^n$ we have

$$f(x) \;=\; \max_v \{\langle v, x\rangle : \; v \in \partial f(0)\} \geq \max_u \{\langle v, x\rangle : \; v \in B_{\|\cdot\|^*}(\gamma_0)\} = \gamma_0 \cdot \|x\|,$$

$$f(x) \;=\; \max_v \{\langle v, x\rangle : \; v \in \partial f(0)\} \leq \max_u \{\langle v, x\rangle : \; v \in B_{\|\cdot\|^*}(\gamma_1)\} = \gamma_1 \cdot \|x\|.$$

Therefore for any x and $h \in R^n$ we have

$$f(x+h) \leq f(x) + f(h) \leq f(x) + \gamma_1 \cdot \|h\|.$$

Moreover,

$$f^* \;=\; \min_x \{f(x) : \; Cx = b\} \geq \min_x \{\gamma_0 \|x\| : \; Cx = b\} = \gamma_0 \cdot \rho.$$

Hence, in view of (2.6) we have

$$f^* \;\geq\; \gamma_0 \cdot \|x_0\| \geq \alpha f(x_0),$$

$$f^* \;\leq\; f(x_0) \leq \gamma_1 \cdot \|x_0\|.$$

In order to prove the second statement note that in view of the first item of the theorem we have

$$\|x_0 - x^*\| \leq \|x_0\| + \|x^*\| \leq \tfrac{2}{\gamma_0} \cdot f^*.$$

For the Euclidean norm this bound can be strengthened, since in this case

$$\|x_0 - x^*\|^2 = \|x^*\|^2 - \|x_0\|^2 < \|x^*\|^2.$$

□

3 Subgradient approximation scheme

Let us discuss now different possibilities for constructing fully polynomial approximation schemes for problem (2.1). For the sake of simplicity we assume that the norm $\|\cdot\|$ is Euclidean.

The first scheme we consider is based on a standard subgradient method for minimizing non-smooth convex functions. Denote by $g(x)$ an arbitrary subgradient of the function

$f(x)$ at point x. Consider the simplest variant of the subgradient method as applied to the problem (2.1).

Subgradient method $G_N(R)$

for $k := 0$ **to** N **do**
begin
 	
Compute $f(x_k)$ and $g(x_k)$.
 	
$x_{k+1} := \pi_{\mathcal{L}}\left(x_k - \frac{R}{\sqrt{N+1}} \cdot \frac{g(x_k)}{\|g(x_k)\|^*}\right)$.
 	
end.

Output: $\bar{x} = \arg\min\{f(x) : \ x = x_0, \ldots, x_N\}$.

(3.1)

In what follows the output of this process $\bar{x}$ is denoted by $G_N(R)$. In view of Theorem 3.2.2 [7], the rate of convergence of this method is as follows:

$$f(G_N(R)) - f^* \leq \frac{\gamma_1}{\sqrt{N+1}} \cdot \frac{\|x_0 - x^*\|^2 + R^2}{2R}. \tag{3.2}$$

Thus, in order to be efficient, the subgradient method needs a good estimate for the distance between the starting point x_0 and the solution x^*. In our case, this estimate can be obtained from inequality (2.9). However, since in our problem the value f^* is not known in advance, let us use its second part:

$$\hat{\rho} \stackrel{\text{def}}{=} \frac{1}{\gamma_0} f(x_0). \tag{3.3}$$

The efficiency of such a choice is described in the following statement.

Theorem 2 *For a fixed δ from $(0, 1)$, let us choose*

$$N = \left\lfloor \frac{1}{\alpha^4 \delta^2} \right\rfloor. \tag{3.4}$$

Then $f(G_N(\hat{\rho})) \leq (1 + \delta) \cdot f^*$.

Proof:
In view of inequality (3.2), the choice (3.4) and inequalities (2.9), (2.7), we have

$$f(G_N(\hat{\rho})) - f^* \leq \alpha^2 \delta \gamma_1 \cdot \frac{\|x_0 - x^*\|^2 + \hat{\rho}^2}{2\hat{\rho}} \leq \alpha^2 \delta \gamma_1 \hat{\rho} = \alpha \delta f(x_0) \leq \delta \cdot f^*.$$

□

Note that we pay quite a high price for the bad estimate of the initial distance. If we were able to use the first part of inequality (2.9), then the corresponding complexity

bound would be much better. Let us show that a better bound for the distance to the optimal solution can be derived from the trivial observation that $f^* \leq f(x)$ for any point x from $\mathcal{L}$.

Let $\delta \in (0,1)$ be the desired relative accuracy. Let

$$\hat{N} = \left\lfloor \tfrac{e}{\alpha^2} \cdot \left(1 + \tfrac{1}{\delta}\right)^2 \right\rfloor,$$

where e is the Euler number. Consider the following process. Set $\hat{x}_0 = x_0$, and for $t \geq 1$ iterate

$$\boxed{\begin{array}{l} \hat{x}_t := G_{\hat{N}}\left(\tfrac{1}{\gamma_0} f(\hat{x}_{t-1})\right); \\ \\ \textbf{if } f(\hat{x}_t) \geq \tfrac{1}{\sqrt{e}} f(\hat{x}_{t-1}) \textbf{ then } T := t \textbf{ and Stop.} \end{array}} \tag{3.5}$$

Theorem 3 *The number of points generated by the process (3.5) is bounded:*

$$T \leq 1 + 2 \ln \tfrac{1}{\alpha}. \tag{3.6}$$

The last point generated satisfies the inequality $f(\hat{x}_T) \leq (1+\delta) f^*$. The total number of lower-level gradient steps in the process (3.5) does not exceed

$$\tfrac{e}{\alpha^2} \cdot \left(1 + \tfrac{1}{\delta}\right)^2 \cdot \left(1 + 2 \ln \tfrac{1}{\alpha}\right). \tag{3.7}$$

Proof:
By simple induction it is easy to prove that at the beginning of stage t in (3.5) the following inequality holds:

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

Thus, in view inequality (2.7), at the last stage T of the process we have

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

This leads to inequality (3.6).

In view of (2.9) we have $\|x_0 - x^*\| \leq \tfrac{1}{\gamma_0} f^* \leq \tfrac{1}{\gamma_0} f(\hat{x}_{T-1})$. Therefore, at the last stage of the process, using (3.2) and the termination rule in (3.5) we get

$$f(\hat{x}_T) - f^* \;\leq\; \tfrac{\gamma_1}{\sqrt{\hat{N}+1}} \cdot \tfrac{1}{\gamma_0} \cdot f(\hat{x}_{T-1}) \leq \tfrac{\sqrt{e}}{\alpha\sqrt{\hat{N}+1}} \cdot f(\hat{x}_T) \leq \tfrac{\delta}{1+\delta} \cdot f(\hat{x}_T).$$

□

4 Structural optimization

In Section 3 we have shown that the outer and inner ellipsoidal approximations of the set $\partial f(0)$ are the key ingredients of minimization schemes for computing an approximate solution to (2.1) in relative scale. However, in order to find an ellipsoidal norm, which is good for our problem, we need to somehow employ its structure. In this section we introduce a model of problem (2.1), which is suitable both for explicit indication of such a norm and for applying the smoothing technique [8]. Note that the efficiency of the latter approach significantly dominates that of the subgradient method.

Since the objective function $f(x)$ in problem (2.1) is homogeneous, the simplest possible structure of such an object might be as follows. Let us assume that $f(x)$ is a composition of two functions, a linear function $A(x)$ and a *simple* nonlinear convex homogeneous function $F(v)$; that is $f(x) = F(A(x))$. Let us introduce this object in a formal way. In this section we switch to the notation of [8].

Let Q_2 be a closed bounded convex set in R^m containing the origin in its interior. Define a convex homogeneous function $F(v)$ as follows:

$$F(v) = \max_u \{\langle v, u\rangle : \ u \in Q_2\}. \tag{4.1}$$

Further, let A be an $m \times n$-matrix, which has a full column rank (thus, $m \geq n$). Define the objective function

$$f(x) = F(Ax), \quad x \in R^n. \tag{4.2}$$

Clearly, $f(x)$ is a convex function with degree of homogeneity one. The problem of interest is still (2.1), which we repeat here:

$$\text{Find } f^* = \min_{x \in R^n} \{f(x) : \ Cx = b\}. \tag{4.3}$$

Since $\partial F(0) \equiv Q_2$, we have $\partial f(0) = A^T Q_2$. Thus problem (4.3) satisfies the main assumption (2.2).

Let $\|\cdot\|_2$ denote the standard Euclidean norm in R^m:

$$\|u\|_2 = \left[\sum_{i=1}^m \left(u^{(i)}\right)^2\right]^{1/2}, \quad u \in R^m.$$

Let us introduce the following characteristics of the function F:

$$\begin{aligned}
\gamma_0(F) &= \max_{r>0}\{r : \ B_{\|\cdot\|_2}(r) \subseteq \partial F(0)\}, \\
\gamma_1(F) &= \min_{r>0}\{r : \ B_{\|\cdot\|_2}(r) \supseteq \partial F(0)\}, \\
\alpha(F) &= \tfrac{\gamma_1(F)}{\gamma_0(F)} \geq 1.
\end{aligned}$$

For the sets from Example 1 these values are as follows:

$$\begin{aligned}
&\gamma_0(F_1) = \tfrac{1}{\sqrt{m}}, \quad \gamma_1(F_1) = 1, \qquad \alpha(F_1) = \sqrt{m}, \\
&\gamma_0(F_\infty) = 1, \qquad \gamma_1(F_\infty) = \sqrt{m}, \quad \alpha(F_\infty) = \sqrt{m}.
\end{aligned} \tag{4.4}$$

Let us define now the following Euclidean norm in the primal space:

$$\|x\|_1 = \|Ax\|_2^*, \quad x \in R^n.$$

Since A is non-degenerate, this norm is well defined. Setting $G = A^T A$, we get the following representations:

$$\begin{aligned} \|x\|_1 &= \langle Gx, x\rangle^{1/2} = \left[\sum_{i=1}^m \langle a_i, x\rangle^2\right]^{1/2}, \\ \|g\|_1^* &= \langle g, G^{-1} g\rangle^{1/2}, \end{aligned} \tag{4.5}$$

where a_i, $i = 1, \ldots, m$, denote columns of matrix A^T.

Lemma 2 *For the norm $\|\cdot\|_1$, the condition (2.3) holds with*

$$\gamma_0 = \gamma_0(F), \quad \gamma_1 = \gamma_1(F).$$

Thus, we can take $\alpha = \alpha(F) = \frac{\gamma_0(F)}{\gamma_1(F)}$.

Proof:
Since $\partial f(0) = A^T Q_2$, we have the following representation for the support function of this set:

$$\xi(x) \stackrel{\text{def}}{=} \max_{s \in \partial f(0)} \langle s, x\rangle = \max_{u \in Q_2} \langle A^T u, x\rangle = \max_{u \in Q_2} \langle Ax, u\rangle.$$

Thus,

$$\xi(x) \leq \max_{\|u\|_2 \leq \gamma_1(F)} \langle Ax, u\rangle = \gamma_1(F)\|Ax\|_2^* = \gamma_1(F)\|x\|_1,$$

$$\xi(x) \geq \max_{\|u\|_2 \leq \gamma_0(F)} \langle Ax, u\rangle = \gamma_0(F)\|Ax\|_2^* = \gamma_0(F)\|x\|_1.$$

Hence, $\partial f(0) \subseteq B_{\|\cdot\|_1^*}(\gamma_1(F))$, and $\partial f(0) \supseteq B_{\|\cdot\|_1^*}(\gamma_0(F))$. □

Note that for many simple sets Q_2 the parameters $\gamma_1(F)$ and $\gamma_0(F)$ are easily available (see, for example, (4.4)). Therefore the metric (4.5) can be used to find an approximate solution to the corresponding problems by the subgradient scheme (3.5). However, the main advantage of the representation (4.2) lies in the possibility of employing the smoothing technique [8]. Let us show how this can be done.

Compared with the problem setting in [8], there is only one difference in (4.3), namely that we have an unbounded primal feasible set. Thus, the straightforward application of the efficient technique [8, 9] to (4.3) is impossible. However, we can introduce an artificial bound by using inequality (2.9). Let

$$Q_1(R) = \{x \in R^n : \; Cx = b, \; \|x - x_0\|_1 \leq R\}.$$

In view of (2.9), we have $x^* \in Q_1(\hat\rho)$ for $\hat\rho = \frac{1}{\gamma_0(F)} f(x_0)$. Thus, problem (4.3) is equivalent to the following:

$$\begin{aligned} \text{Find } f^* &= \min_{x \in R^n} \{f(x) : \; x \in Q_1(\hat\rho)\} \\ &= \min_{x \in Q_1(\hat\rho)} \max_{u \in Q_2} \langle Ax, u\rangle \\ &= \max_{u \in R^m} \{\phi_{\hat\rho}(u) : \; u \in Q_2\}, \end{aligned} \tag{4.6}$$

where $\phi_R(u) = \min\limits_{x \in Q_1(R)} \langle Ax, u \rangle$. Thus, we have managed to represent our problem in the form required by [8].

Let us introduce the objects necessary to apply the smoothing technique. In the primal space we choose the prox-function $d_1(x) = \frac{1}{2}\|x - x_0\|_1^2$. This function has convexity parameter $\sigma_1 = 1$. Its maximum on the feasible set $Q_1(\hat{\rho})$ is equal to $D_1 = \frac{1}{2}\hat{\rho}^2$.

Similarly, for the dual feasible set we choose $d_2(u) = \frac{1}{2}\|u\|_2^2$. Then $\sigma_2 = 1$, and the maximum of this function on the dual feasible set Q_2 does not exceed $D_2 = \frac{1}{2}\gamma_1^2(F)$. It remains to note that

$$\begin{aligned} \|A\|_{1,2} &= \max_{x,u}\{\langle Ax, u\rangle : \ \|x\|_1 \le 1, \ \|u\|_2 \le 1\} = \max_x\{\|Ax\|_2^* : \ \|x\|_1 \le 1\} \\ &= \max_x\{\|x\|_1 : \ \|x\|_1 \le 1\} = 1. \end{aligned} \tag{4.7}$$

For the reader's convenience we present here the algorithm (3.11) from [8] adopted for our needs. This method is applied to a smooth approximation of the objective function $f(x)$:

$$f_\mu(x) = \max_u\{\langle Ax, u\rangle - \mu d_2(u) : \ u \in Q_2\}. \tag{4.8}$$

Note that this function has Lipschitz continuous gradient

$$\nabla f_\mu(x) = A^T u_\mu(x),$$

where $u_\mu(x)$ is a unique solution to optimization problem in (4.8). In view of (4.7), the Lipschitz constant for the gradient is equal to $\frac{1}{\mu}$ (see, for example, Theorem 1 in [8]).

Method $S_N(R)$

Set $\mu = \frac{2R}{\gamma_1(F)\cdot(N+1)}$.

for $k := 0$ **to** N **do**
begin

$$u_\mu(x_k) := \arg\max_u\{\langle Ax_k, u\rangle - \tfrac{\mu}{2}\|u\|_2^2 : \ u \in Q_2\},$$

$$y_k := \arg\min_x\{\langle A(x - x_k), u_\mu(x_k)\rangle + \tfrac{1}{2\mu}\|x - x_k\|_1^2 : \ x \in Q_1(R)\},$$

$$z_k := \arg\min_x\left\{\tfrac{1}{2\mu}\|x - x_0\|_1^2 + \langle A(x - x_0), \sum_{i=0}^{k} \tfrac{i+1}{2} u_\mu(x_i)\rangle : \ x \in Q_1(R)\right\},$$

$$x_{k+1} := \tfrac{2}{k+3} z_k + \tfrac{k+1}{k+3} y_k.$$

end. (4.9)

Output: $\bar{x} := y_N$.

In what follows we denote the output $\bar{x}$ of this process by $S_N(R)$. It is easy to check that all conditions of Theorem 3 [8] are satisfied. Thus, if $\|x_0 - x^*\|_1 \leq R$, then the output of this process satisfies the inequality

$$f(S_N(R)) - f^* \leq \frac{2\gamma_1(F)R}{N+1}. \tag{4.10}$$

This observation has an important corollary.

Theorem 4 *For $\delta \in (0,1)$, let*

$$N = \left\lfloor \frac{2}{\alpha^2(F)\,\delta} \right\rfloor. \tag{4.11}$$

Then $f\left(S_N\left(\frac{1}{\gamma_0(F)} f(x_0)\right)\right) \leq (1+\delta) f^*$.

Proof:
Since by (2.9) $\|x_0 - x^*\|_1 \leq \frac{1}{\gamma_0(F)} f(x_0)$, and by (4.11) $N + 1 \geq \frac{2}{\alpha^2(F)\,\delta}$, from (4.10) and (2.7) we have

$$f(S_N(R)) - f^* \leq \delta \cdot \alpha(F) f(x_0) \leq \delta \cdot f^*.$$

□

Note that the complexity (4.11) of the scheme (4.9) is lower than that of the subgradient scheme (3.5) with the recursively updated estimate for the distance to the optimum. Let us show that a similar updating strategy can also accelerate scheme (4.9).

Let $\delta \in (0,1)$ be the desired relative accuracy. Let

$$\tilde{N} = \left\lfloor \frac{2e}{\alpha(F)} \cdot \left(1 + \frac{1}{\delta}\right) \right\rfloor.$$

Consider the following process. Set $\hat{x}_0 = x_0$. For $t \geq 1$ iterate

$$\boxed{\begin{array}{l} \hat{x}_t := S_{\tilde{N}}\left(\frac{1}{\gamma_0(F)} f(\hat{x}_{t-1})\right); \\ \\ \textbf{if } f(\hat{x}_t) \geq \frac{1}{e} f(\hat{x}_{t-1}) \textbf{ then } T := t \textbf{ and } \text{Stop}. \end{array}} \tag{4.12}$$

Theorem 5 *The number of points T generated by scheme (4.12) is bounded by*

$$T \leq 1 + \ln \frac{1}{\alpha(F)}. \tag{4.13}$$

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 (4.12) does not exceed

$$\frac{2e}{\alpha(F)} \cdot \left(1 + \frac{1}{\delta}\right) \cdot \left(1 + \ln \frac{1}{\alpha(F)}\right). \tag{4.14}$$

Proof:
By simple induction it is easy to prove that at beginning of the stage t the following inequality holds:

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

Thus, in view of Item 1 of Theorem 1, at the last stage T of the process we have

$$\left(\tfrac{1}{e}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \alpha(F) f(x_0).$$

This leads to inequality (4.13).

Note that $\|x_0 - x^*\| \leq \frac{1}{\gamma_0(F)} f^* \leq \frac{1}{\gamma_0(F)} f(\hat{x}_{T-1})$. Therefore, at the last stage of the process in view of inequality (4.10) and the termination rule in (4.12) we have

$$f(\hat{x}_T) - f^* \quad \leq \quad \tfrac{2\gamma_1(F)}{\tilde{N}+1} \cdot \tfrac{1}{\gamma_0(F)} \cdot f(\hat{x}_{T-1}) \leq \tfrac{2e}{\alpha(F)\cdot(\tilde{N}+1)} \cdot f(\hat{x}_T) \leq \tfrac{\delta}{1+\delta} \cdot f(\hat{x}_T).$$

□

5 Application examples

In this section we discuss the complexity of implementation of the schemes presented in Section 4 as applied to different structural classes of optimization problems.

5.1 Linear programming

Let $\hat{A}$ be an $m \times (n-1)$-matrix, $m \geq n$, which has a full column rank. For a given vector $c \in R^m$ consider the following optimization problem:

$$\text{Find } f^* = \max_{u \in R^m} \{\langle c, u \rangle : \ \hat{A}^T u = 0, \ |u^{(i)}| \leq 1, \ i = 1, \ldots, m\}. \tag{5.1}$$

This problem is not trivial only if the column rank of the matrix $A = (\hat{A}, c)$ is equal to n, which we assume to be true.

Problem (5.1) can be rewritten in dual form. Define

$$\phi_\infty(y) = \max_{u \in R^m} \{\langle c, u \rangle : + \langle y, \hat{A}^T u \rangle, \ |u^{(i)}| \leq 1, \ i = 1, \ldots, m\} = \sum_{i=1}^m |\langle a_i, y \rangle + c_i|,$$

where a_i are the columns of the matrix $\hat{A}^T$. Then

$$f^* = \min_{y \in R^{n-1}} \ \phi_\infty(y).$$

In Example 1 we have already seen that the latter minimization problem can be represented in the form (4.2)-(4.3) with $x = (y^T, \tau)^T$, and $F_\infty(v) = \sum\limits_{i=1}^m |v^{(i)}|$. Thus,

$$Q_2 = \{u \in R^m : \ |u^{(i)}| \leq 1, \ i = 1, \ldots, m\}.$$

Choosing $\|u\|_2 = \left[\sum\limits_{i=1}^m (u^{(i)})^2\right]^{1/2}$, we get

$$\gamma_0(F_\infty) = 1, \quad \gamma_1(F_\infty) = \sqrt{m}, \quad \alpha(F_\infty) = \tfrac{1}{\sqrt{m}}.$$

Therefore, in view of Theorem 5, in order to estimate f^* with relative accuracy $\delta \in (0,1)$ we need at most

$$2e \cdot m^{1/2} \cdot \left(1 + \tfrac{1}{2} \ln m\right) \cdot \left(1 + \tfrac{1}{\delta}\right)$$

iterations of the scheme $S_N(R)$.

Note that to implement this method, we need to compute and invert the matrix $G = A^T A$. If A is dense, this takes $O(n^2 m)$ operations. Since each iteration of the scheme $S_N(R)$ requires $O(nm)$ operation, the total complexity becomes

$$O\left(n^2 m + \tfrac{1}{\delta} \cdot nm^{1.5} \ln m\right) \tag{5.2}$$

operations. The first ingredient of this estimate is dominant when $\delta > \frac{\sqrt{m}}{n} \ln m$.

Note that for problem (5.1) we can apply a standard short-step path-following scheme (see, for example, [7]). Each iteration of this scheme needs $O(n^2 m)$ operations. Therefore its worst-case efficiency estimate looks as follows:

$$O\left(n^2 m^{1.5} \ln \tfrac{m}{\delta}\right) \tag{5.3}$$

Another possibility is to solve this problem by the ellipsoid method [5]. In this case the total complexity of this scheme is

$$O\left(n^3 m \ln \tfrac{m}{\delta}\right). \tag{5.4}$$

Comparing the bounds (5.2), (5.3), and (5.4), we conclude that the scheme (4.12) is best when δ is not too small, say

$$\delta > O\left(\tfrac{1}{n} \max\left\{1, \tfrac{\sqrt{m}}{n}\right\}\right).$$

5.2 Minimization of spectral radius

Denote by S^n the space of symmetric $n \times n$-matrices. For $X \in S^n$ we can define its spectral radius:

$$\rho(X) = \max_{1 \le i \le n} |\lambda_i(X)|.$$

Note that this function is convex on S^n. For a vector of decision variables $x \in R^p$, let us introduce a linear operator $A(x)$:

$$A(x) = \sum_{i=1}^{p} x^{(i)} A_i \in S^n.$$

Now we can define the following objective function in problem (4.3):

$$f(x) = \rho(A(x)). \tag{5.5}$$

Assume also that the linear constraints in the problem (4.3), (5.5) are very simple; for example, that could be $x^{(1)} = 1$.

In order to treat the problem (4.3), (5.5) we need to represent the upper-level function $\rho(X)$ in a special form (4.1). Let

$$Q_2 = \{X \in S^n : \sum_{i=1}^{n} |\lambda_i(X)| \le 1\}.$$

Let us endow the space S^n with the standard Frobenius norm:

$$\|X\|_2 = \langle X, X\rangle^{1/2} = \left[\sum_{i,j=1}^{n} \left(X^{(i,j)}\right)^2\right]^{1/2}.$$

Lemma 3 *The set Q_2 is a closed convex set such that*

$$B_{\|\cdot\|_2}(\tfrac{1}{\sqrt{n}}) \subset Q_2 \subset B_{\|\cdot\|_2}(1). \tag{5.6}$$

Moreover, $\rho(X) = \max\limits_{U \in S^n}\{\langle X, U\rangle : \ U \in Q_2\}$.

Proof:
For any $X \in S^n$ we have:

$$\begin{aligned}
\rho(X) &= \min_{\tau \in R^1}\{\tau : \ \tau I_n \succeq X, \ \tau I_n \succeq -X\} \\
&= \min_{\tau \in R^1} \max_{Y_1, Y_2 \succeq 0} [\tau + \langle X - \tau I_n, Y_1\rangle - \langle X + \tau I_n, Y_2]\\
&= \max_{Y_1, Y_2 \succeq 0}\{\langle X, Y_1 - Y_2\rangle : \ \langle I_n, Y_1 + Y_2\rangle = 1\}.
\end{aligned}$$

Let $U = Y_1 - Y_2$ and $V = Y_1 + Y_2$. Then

$$\rho(X) = \max_{U \in S^n}\{\langle X, U\rangle, \ U \in \hat{Q}\},$$

where $\hat{Q} = \{U : \ \exists V \succeq \pm U, \ \langle I_n, V\rangle = 1\}$. It is clear that the set $\hat{Q}$ is closed, convex and bounded. Let us prove that $\hat{Q} = Q_2$.

Indeed, we can always represent U by its orthogonal basis of eigenvectors:

$$U = B\Lambda B^T, \quad BB^T = I_n,$$

where Λ is a diagonal matrix. Assume that $U \in Q_2$. Define a diagonal matrix $\hat{\Lambda}$ with the following diagonal entries:

$$\hat{\Lambda}^{(i,i)} = |\Lambda^{(i,i)}|/[\sum_{j=1}^{n} |\Lambda^{(j,j)}|], \quad i = 1, \ldots, n.$$

Then $V = B\hat{\Lambda}B^T \succeq \pm U$ and $\langle I_n, V\rangle = 1$. Thus $Q_2 \subset \hat{Q}$.

Conversely, if $U \in \hat{Q}$, then there exists $V \in S^n$ such that $B^T V B \succeq \pm\Lambda$. Therefore

$$\langle Vb_i, b_i\rangle \geq |\Lambda^{(i,i)}|, \quad i = 1, \ldots, n,$$

where b_i are the columns of the matrix B. Hence,

$$1 = \langle I_n, V\rangle = \langle BB^T, V\rangle = \langle I_n, B^T V B\rangle = \sum_{i=1}^{n} \langle Vb_i, b_i\rangle \geq \sum_{i=1}^{n} |\lambda_i(U)|.$$

Thus, $\hat{U} \subseteq Q_2$ and we conclude that $\hat{U} = Q_2$.

It remains to prove inclusion (5.6). Indeed, if $\|U\|_2^2 \leq \frac{1}{n}$, that is $\sum\limits_{i=1}^{n} \lambda_i^2(U) \leq \frac{1}{n}$, then

$$\sum_{i=1}^{n} |\lambda_i(U)| \leq \sqrt{n} \cdot \left[\sum_{i=1}^{n} \lambda_i^2(U)\right]^{1/2} \leq 1.$$

Conversely, if $\sum\limits_{i=1}^{n} |\lambda_i(U)| \leq 1$, then $\sum\limits_{i=1}^{n} \lambda_i^2(U) \leq \left[\sum\limits_{i=1}^{n} |\lambda_i(U)|\right]^2 \leq 1$. □

Thus, in view of inclusion (5.6) we have

$$\gamma_0(\rho) = \tfrac{1}{\sqrt{n}}, \quad \gamma_1(\rho) = 1, \quad \alpha(\rho) = \tfrac{1}{\sqrt{n}}.$$

Hence, in view of Theorem 5, the total number of iterations of the method $S_N(R)$ does not exceed

$$2e\sqrt{n}\left(1 + \tfrac{1}{2}\ln n\right) \cdot \left(1 + \tfrac{1}{\delta}\right).$$

In order to apply this approach we need to compute and to invert the matrix G. In our situation G is the matrix of the following quadratic form:

$$\langle Gx, x\rangle = \langle A(x), A(x)\rangle.$$

Thus, $G^{(i,j)} = \langle A_i, A_j\rangle$, $i, j = 1, \ldots, p$. If the matrices A_i are dense, the computation of this matrix takes $O(p^2n^2)$ arithmetic operations and the inversion takes $O(p^3)$ operations. Since we assume $p < \frac{n(n+1)}{2}$, the total cost of preliminary computation is of the order $O(p^2n^2)$ operations.

Further, the most costly operations at each step of the method $S_N(R)$ are as follows.

- Computation of the value of bilinear form $\langle A(x), U\rangle$ and its gradients takes $O(pn^2)$ operations.
- Finding a projection of a point X onto the set Q_2 with respect to the standard Frobenius norm. The most expensive part of this operation consists in solving an eigenvalue problem for matrix X. That can be done in $O(n^3)$ operations.
- The total amount of operations in the space R^p does not exceed $O(p^2)$.

Thus, the complexity of each iteration of $S_N(R)$ is of the order $O(n^2(n+p))$ operations. Hence, in total, the method (4.12) requires

$$O\left(n^2p^2 + \tfrac{1}{\delta} \cdot n^{2.5}(p+n)\ln n\right) \tag{5.7}$$

arithmetic operations.

Let us compare this estimate with the worst-case complexity of a short-step path-following scheme as applied to the problem (4.3)-(5.5). For this method the most expensive computations at each iteration are the computations of the elements of the Hessian of barrier function. These are the values

$$\langle X^{-1}A_iX^{-1}, A_j\rangle, \quad i, j = 1, \ldots, p.$$

Such a computation needs $O(pn^2(p+n))$ operations. Thus, the total complexity of the interior-point method is of the order

$$O\left(pn^{2.5}(p+n)\ln\frac{n}{\delta}\right)$$

operations. Comparing this estimate with (5.7) we see that the gradient method is better if the required relative accuracy is not too small:

$$\delta \geq O\left(\tfrac{1}{p}\right).$$

5.3 Truss topology design problem

In this problem (see, for example, [1]), we have a set of points

$$x_i \in R^2, \quad i = 1, \ldots, n+p,$$

connected by a set of arcs (i_k, j_k), $k = 1, \ldots, m$. We always assume that $j_k > i_k$. Each arc has a nonnegative weight $t^{(k)}$, and the sum of the weights is one. The nodes $x_{n+1}, \ldots, x_{n+p}$ are fixed. To all other nodes we apply external forces

$$f_i \in R^2, \quad i = 1, \ldots, n, \quad f \stackrel{\text{def}}{=} (f_1, \ldots, f_n)^T \in R^{2n}.$$

The goal is to find an optimal design vector

$$t \stackrel{\text{def}}{=} (t^{(1)}, \ldots, t^{(m)})^T \in \Delta_m \equiv \left\{ t \in R^m_+ : \sum_{i=1}^m t^{(i)} = 1 \right\},$$

which minimizes the total *stiffness* $\psi(t)$ of the system.

In order to define the stiffness, let us assume first that $i_k < n$, $k = 1, \ldots, m$. For each arc k, let

$$d_k = \frac{x_{i_k} - x_{j_k}}{\|x_{i_k} - x_{j_k}\|^2} \in R^2, \quad k = 1, \ldots, m,$$

where $\|\cdot\|$ is the standard Euclidean norm in R^2. Now we can define the constraint vector $a_k = (a_{k,1}, \ldots, a_{k,n})^T \in R^{2n}$, which is composed of the following two-dimensional vectors:

$$a_{k,q} = \begin{cases} d_k, & \text{if } q = i_k, \\ -d_k, & \text{if } q = j_k \text{ and } j_k \leq n, \\ 0, & \text{otherwise.} \end{cases} \qquad q = 1, \ldots, n.$$

Let $B(t) = \sum\limits_{k=1}^m t^{(k)} a_k a_k^T$. Then the *truss topology design* problem can be written as follows

$$\text{Find } \psi^* = \inf_t \{ \langle [B(t)]^{-1} f, f \rangle : \ t \in \operatorname{rint} \Delta_m \}. \tag{5.8}$$

This problem is well defined if and only if the matrix $G \stackrel{\text{def}}{=} B(\bar{e}_m)$ is positive definite.

Let us show how this problem can be rewritten in the form (4.2)-(4.3).

$$\begin{aligned}
\psi^* &= \inf_{t \in \mathrm{rint}\,\Delta_m} \langle [B(t)]^{-1} f, f \rangle \\
&= \inf_{t \in \mathrm{rint}\,\Delta_m} \max_{x \in R^{2n}} [2\langle f, x \rangle - \langle B(t)x, x \rangle] \\
&= \max_{x \in R^{2n}} \inf_{t \in \mathrm{rint}\,\Delta_m} \left[2\langle f, x \rangle - \sum_{k=1}^{m} t^{(k)} \langle a_k, x \rangle^2 \right] \\
&= \max_{x \in R^{2n}} \left[\langle f, x \rangle^2 \cdot \left(\max_{1 \le k \le m} |\langle a_k, x \rangle| \right)^{-2} \right]
\end{aligned}$$

(in the last step of the transformations we perform a minimization of the objective function along direction x).

Thus, we can consider the problem

$$\text{Find } f^* = \min_{x \in R^{2n}} \{ f(x) \stackrel{\text{def}}{=} \max_{1 \le k \le m} |\langle a_k, x \rangle| : \ \langle f, x \rangle = 1 \}, \tag{5.9}$$

which is exactly in the desired form (4.2)-(4.3). Let A be an $m \times (2n)$-matrix with the rows a_k^T. Then, using notation of Example 1, the objective function of this problem can be written as

$$f(x) = F_1(Ax).$$

In view of (4.4) we have $\alpha(F_1) = \frac{1}{\sqrt{m}}$. Therefore, in order to find an approximate solution to (5.9) with relative accuracy δ, the method (4.12) needs at most

$$2e\sqrt{m} \left(1 + \tfrac{1}{2} \ln m \right) \cdot \left(1 + \tfrac{1}{\delta} \right) \tag{5.10}$$

iterations of the scheme $S_N(R)$. The most expensive operations of each iteration of the latter scheme are as follows.

- Computation of the value and the gradients of the bilinear form $\langle Ax, u \rangle$ needs $O(m)$ operations (recall that A is sparse).
- Euclidean projection on $Q_2 \subset R^m$ needs $O(m \ln m)$ operations.
- All steps in the primal space need $O(n^2)$ operations.

Note that the preliminary computation of the matrix G needs $O(m + n^2)$ operations, but its inversion needs $O(n^3)$. Since $m \le \frac{n(n+1)}{2}$, we come to the following upper bound for the total computational effort of the method (4.12):

$$O \left(n^3 + \tfrac{1}{\delta} \cdot (n^2 + m \ln m) \cdot \sqrt{m} \ln m \right) \tag{5.11}$$

arithmetic operations. For a dense truss with $m = O(n^2)$ this estimates becomes

$$O \left(\tfrac{n^3}{\delta} \ln^2 n \right).$$

References

[1] A. Ben-Tal, A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms; Engineering Applications*. SIAM-MPS Series in Optimization, 2001.

[2] D. Bienstock. *Potential Function Methods for Approximately Solving Linear Programming Problems. Theory and Practice*. Kluwer Academic Publishers, Boston, 2002.

[3] 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.

[4] J.-B. Hiriart-Urruty, C. Lemarechal. *Convex Analysis and Minimization Algorithms. Part 1.* A Series of Comprehensive Studies in Mathematics, Springer Verlag, 1993.

[5] A.S. Nemirovskij, D.B. Yudin. *Problem complexity and method efficiency in optimization*. Wiley-Interscience Series in Discrete Mathematics. A Wiley-Interscience Publication. John Wiley & Sons, 1983.

[6] Yu. Nesterov, A. Nemirovskii. *Interior point polynomial methods in convex programming: Theory and Applications*, SIAM, Philadelphia, 1994.

[7] Yu. Nesterov. *Introductory lectures on convex optimization. A basic course*. Kluwer, Boston, 2003.

[8] Yu. Nesterov. Smooth minimization of non-smooth functions. CORE DP #2003/12, February 2003. Accepted by *Mathematcal Programming*.

[9] Yu. Nesterov. Excessive gap technique in non-smooth convex minimizarion. CORE DP #2003/35, May 2003. Accepted by *SIAM J. Optimization*.

[10] 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.