Mathematical Methods of Operations Research manuscript No.
(will be inserted by the editor)

Minimumization of a convex function on the convex hull of a point set

Nikolai D. Botkin[1⋆], **Josef Stoer**[2]

[1] Center of advanced european studies and research (caesar),
Ludwig-Erhard-Allee 2, 53175 Bonn, Germany (e-mail: botkin@caesar.de)
[2] Institute for applied mathematics and statistic of the University of Würzburg,
97074 Würzburg, Germany (e-mail: jstoer@mathematik.uni-wuerzburg.de)

Received: / Revised version:

Abstract An algorithm for the minimumization of a differentiable convex function defined on the convex hull of m points in R^n is proposed. Each iteration of the algorithm is implemented in barycentric coordinates, the number of which is equal to m. The method is based on a new procedure for finding the projection of the gradient of the objective function onto a simplicial cone in R^m, which is the tangent cone at the current point to the simplex defined by the usual constraints on barycentric coordinates. It is shown that this projection can be computed in $O(m \ln m)$ operations.

Key words: Convex functions on simplexes, Barycentric coordinates

1 Formulation of the Problem

Let $Z = \{z^1, z^2, ..., z^m\}$ be a finite set of points in the Euclidean space R^n. Let $f : R^n \to R$ be a differentiable convex function. We consider the following problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in \text{co } Z. \tag{1}$$

Obviously, problem (1) is equivalent to

$$\text{minimize} \quad g(\alpha) \quad \text{subject to} \quad \alpha \in \Lambda, \tag{2}$$

where

$$g(\alpha) := f\left(\sum_{i=1}^{m} \alpha_i z^i\right), \quad \Lambda = \left\{\alpha \in R^m : \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i = 1\right\}.$$

⋆ The author was supported by the Alexander von Humboldt-Foundation, Germany.

Denote by Λ^0 the set of minimizers of (2). For arbitrary $\alpha \in \Lambda$ define the set of admissible directions $\mathcal{P}(\alpha)$ as follows

$$\mathcal{P}(\alpha) = \left\{ p \in R^m : \sum_{i=1}^{m} p_i^2 \leq 1, \ \exists \lambda > 0 \ \text{ such that } \ \alpha + \lambda p \in \Lambda \right\}.$$

It is easily seen that $\mathcal{P}(\alpha)$ is given by the formula

$$\mathcal{P}(\alpha) = \left\{ p \in R^m : \sum_{i=1}^{m} p_i^2 \leq 1, \ \sum_{i=1}^{m} p_i = 0, p_j \geq 0, \ j \in I^0(\alpha) \right\},$$

where

$$I^0(\alpha) = \{ i \in \overline{1,m} : \alpha_i = 0 \}.$$

For $\alpha \in \Lambda$ and $h \in R^m$, we denote

$$\omega(\alpha, h) = \min_{p \in \mathcal{P}(\alpha)} \langle h, p \rangle, \quad p^0(\alpha, h) = \arg \min_{p \in \mathcal{P}(\alpha)} \langle h, p \rangle. \tag{3}$$

The vector $p^0(\alpha, h)$ is called the projection of h onto $\mathcal{P}(\alpha)$.

The following propositions can be easily obtained using properties of convex functions.

Proposition 1. *Let $\alpha \in \Lambda$ and $p \in \mathcal{P}(\alpha)$. If $\langle Dg(\alpha), p \rangle < 0$, then there exists $\lambda > 0$ such that $\alpha + \lambda p \in \Lambda$ and $g(\alpha + \lambda p) < g(\alpha)$.*

Proposition 2. *The following is true:*
a) $\alpha \in \Lambda \setminus \Lambda^0$ iff $\omega(\alpha, Dg(\alpha)) < 0$,
b) $\alpha \in \Lambda^0$ iff $\omega(\alpha, Dg(\alpha)) = 0$.

Thus, we arrive at the following gradient algorithm for finding a minimizing point $\alpha^* \in \Lambda^0$.

Algorithm 1

Step 0: Choose an initial point $\alpha^0 \in \Lambda$ and set $k := 0$.
Step 1: Find $p^k = p^0(\alpha^k, Dg(\alpha^k))$, $\omega^k = \omega(\alpha^k, Dg(\alpha^k))$.
If $\omega^k = 0$ then $\alpha^* := \alpha^k$ and stop.
Step 2: Find

$$\lambda_{max} = \max\{\lambda \geq 0 : \alpha^k + \lambda p^k \in \Lambda\} = \min_{p_i^k < 0} \left\{ -\frac{\alpha_i^k}{p_i^k} \right\}$$

($\lambda_{max} > 0$ because $p^k \in \mathcal{P}(\alpha^k)$).
Step 3: Perform a line search along p^k in order to find

$$\lambda^k = \arg \min_{\lambda \in [0, \lambda_{max}]} g(\alpha^k + \lambda p^k)$$

($\lambda^k > 0$ and $g(\alpha^k + \lambda^k p^k) < g(\alpha^k)$ due to Propositions 1,2).
Step 4: Let

$$\alpha^{k+1} := \alpha^k + \lambda^k p^k, \quad k := k + 1$$

and go to Step 1.

Algorithm 1 is the well known gradient projection method (compare, for example, with [1]–[3]). The purpose of this paper is to propose a fast procedure for updating the projection p^k of the gradient $Dg(\alpha^k)$ in order to gain an efficient realization of this method. The next section deals with this procedure.

2 A procedure for finding the projection

A recursive polynomial method for finding the projection onto the canonical simplex was proposed in [4]. This method also solves the problem of finding the projection $p^0(\alpha, h)$ defined by (3) for any fixed h. Nevertheless, we suggest a different procedure, which requires less arithmetic operations then the procedure of [4] ($O(m \ln m)$ instead of $O(m^2)$).

The following proposition establishes the uniqueness of the vector $p^0(\alpha, h)$.

Proposition 3. *If $\omega(\alpha, h) < 0$, then $p^0(\alpha, h)$ is unique.*

Proof. Assume that there are two minimizing vectors $p^0,\ q^0$ (the arguments α, h are omitted). Then

$$\langle h, p^0\rangle = \langle h, q^0\rangle = \omega(\alpha, h) < 0,$$

which implies, in particular, that $|p^0| = |q^0| = 1$. Obviously,

$$\hat{p} = \frac{\frac{1}{2}p^0 + \frac{1}{2}q^0}{|\frac{1}{2}p^0 + \frac{1}{2}q^0|} \in \mathcal{P}(\alpha)$$

and

$$\langle h, \hat{p}\rangle = \frac{\omega(\alpha, h)}{|\frac{1}{2}p^0 + \frac{1}{2}q^0|} = \frac{\omega(\alpha, h)}{(\frac{1}{2} + \frac{1}{2}\langle p^0, q^0\rangle)^{\frac{1}{2}}} < \omega(\alpha, h).$$

This contradiction proves Proposition 3.

Without any loss of generality we may assume that $I^0(\alpha) = \overline{1, r}$, i.e

$$\alpha = (0, ..., 0, \alpha_{r+1}, ...\alpha_m)^T,$$

where $\alpha_i > 0,\ i = r+1, ..., m$. For any $p \in \mathcal{P}(\alpha)$, we have $\sum_{i=1}^{m} p_i = 0$ and, therefore, $p_m = -\sum_{i=1}^{m-1} p_i$ (any index $l = r+1, ..., m$ could be chosen for such an elimination). Thus, we arrive at the following problem

$$\begin{aligned} &\text{minimize} \ \sum_{i=1}^{m-1} -d_i p_i \\ &\text{subject to} \ \ p \in \mathcal{P}', \end{aligned} \tag{4}$$

where

$$d_i = h_m - h_i, \quad i = 1, ..., m-1,$$

$$\mathcal{P}' = \left\{ p \in R^{m-1} : \left(\sum_{i=1}^{m-1} p_i \right)^2 + \sum_{i=1}^{m-1} p_i^2 \le 1, \quad p_i \ge 0, \quad i = 1, ..., r \right\}.$$

Apply the Kuhn–Tucker theorem by writing the Lagrangian of (4) as

$$\mathcal{L}(p, \mu, \mu_0) = \sum_{i=1}^{m-1} -d_i p_i + \mu_0 \left(\left(\sum_{i=1}^{m-1} p_i \right)^2 + \sum_{i=1}^{m-1} p_i^2 - 1 \right) - \sum_{i=1}^{r} \mu_i p_i. \quad (5)$$

Assume that $\mu_0 = 1/2$. The arguments which justify this assumption will be given later.

From the equation

$$\frac{\partial \mathcal{L}(p, \mu, \mu_0)}{\partial p} = 0,$$

one can easily obtain explicit formulas for p:

$$p_i = d_i - \frac{1}{m} \sum_{j=1}^{m-1} d_j + \mu_i - \frac{1}{m} \sum_{j=1}^{r} \mu_j, \quad i = 1, ..., r,$$

$$p_i = d_i - \frac{1}{m} \sum_{j=1}^{m-1} d_j - \frac{1}{m} \sum_{j=1}^{r} \mu_j, \quad i = r+1, ..., m-1.$$

Denote

$$a_i = d_i - \frac{1}{m} \sum_{j=1}^{m-1} d_j$$

and assume without any loss of generality that $a_1 \le a_2 \le ... \le a_r$. Then, for finding nonnegative $\mu_1, \mu_2, ..., \mu_r$ satisfying the Kuhn–Tucker conditions, we have the following system

$$\begin{aligned}
&p_1 = a_1 + \mu_1 - \frac{1}{m} \sum_{j=1}^{r} \mu_j \ge 0, \quad \mu_1 \ge 0 \\
&p_2 = a_2 + \mu_2 - \frac{1}{m} \sum_{j=1}^{r} \mu_j \ge 0, \quad \mu_2 \ge 0 \\
&... \\
&p_r = a_r + \mu_r - \frac{1}{m} \sum_{j=1}^{r} \mu_j \ge 0, \quad \mu_r \ge 0 \qquad (6) \\
&p_{r+1} = a_{r+1} - \frac{1}{m} \sum_{j=1}^{r} \mu_j \\
&... \\
&p_{m-1} = a_{m-1} - \frac{1}{m} \sum_{j=1}^{r} \mu_j
\end{aligned}$$

with the complementarity conditions

$$\mu_i p_i = 0, \quad i = 1, ..., r. \tag{7}$$

It will be shown that solving this system requires $O(m)$ operations. Taking into account that the optimal procedure for ordering a_i requires $O(m \ln m)$ operations implies that the number of operations for finding the projection is $O(m \ln m)$.

Denote

$$\begin{aligned} S_0 &= 0, \\ S_i &= -\frac{m}{m-i} \sum_{j=1}^{i} a_j, \quad i = 1, ..., r. \end{aligned} \tag{8}$$

Let the index ℓ be defined as follows

$$\ell = \begin{cases} 0 & \text{if } a_1 \geq 0, \\ \max\{k \in \overline{1,r} : \forall i \in \overline{1,k} \ \ (a_i - \frac{1}{m} S_{i-1} < 0)\} & \text{if } a_1 < 0. \end{cases} \tag{9}$$

Observe that ℓ can be found in $O(m)$ operations with the use of the following recurrence relation

$$S_i = \frac{m-i+1}{m-i} S_{i-1} - \frac{m}{m-i} a_i, \quad i = 1, ..., r.$$

Proposition 4. *System (6) with complementarity conditions (7) has a unique solution given by the following formulas*

$$\begin{aligned} \mu_i &= -a_i + \tfrac{1}{m} S_\ell, \ i = 1, ..., \ell, \\ \mu_i &= 0, \qquad\qquad\ \ i = \ell + 1, ..., r. \end{aligned} \tag{10}$$

Respectively,

$$\begin{aligned} p_i &= 0, \qquad\qquad\ \ i = 1, ..., \ell, \\ p_i &= a_i - \tfrac{1}{m} S_\ell, \ i = \ell + 1, ..., m - 1. \end{aligned} \tag{11}$$

Proof. By substituting (10) in (6) one can prove that (11) holds. From the definition of ℓ, it follows that $p_i \geq 0, \quad i = \ell + 1, ..., r$. Obviously, (7) holds for p's and μ's given by (10) and (11). Thus, it remains to prove that $\mu_i > 0, \quad i = 1, ..., \ell$. In fact, from the definition of ℓ, we have

$$a_\ell - \frac{1}{m} S_{\ell-1} < 0. \tag{12}$$

Relations (8) imply

$$a_\ell = -\frac{m-\ell}{m} S_\ell + \frac{m-\ell+1}{m} S_{\ell-1},$$

which with (12) yields

$$-\frac{m-\ell}{m} S_\ell + \frac{m-\ell+1}{m} S_{\ell-1} - \frac{1}{m} S_{\ell-1} < 0,$$

This gives

$$S_\ell > S_{\ell-1}.$$

The last inequality and (12) imply

$$a_\ell - \frac{1}{m} S_\ell < 0. \tag{13}$$

Since $a_1 \le a_2 \le ... \le a_\ell$, we have from (13)

$$a_i - \frac{1}{m} S_\ell < 0, \quad i = 1, ..., \ell.$$

Taking into account (10) implies that $\mu_i > 0, \quad i = 1, ..., \ell$.

Prove that the problem (6) and (7) has only one solution (which is given by (10)). Assume there is another solution $\bar\mu_1, \bar\mu_2, ..., \bar\mu_r$. If $\bar\mu_i = 0, \quad i = 1, ..., r$, then deduce from (6) that $a_1 \ge 0$ and, therefore, $\ell = 0$ by (9). Hence by (10), $\mu_i = 0, \quad i = 1, ..., r$, i.e $\mu_i = \bar\mu_i, \quad i = 1, ..., r$. Suppose now that $\bar\mu_j > 0$ for some $j \in \overline{1, r}$. Set

$$\ell' = \max\{i \in \overline{1, r} : \bar\mu_i > 0\}.$$

Obviously, $\bar\mu_{\ell'} > 0$ and we have

$$\begin{aligned}
\bar p_1 &= a_1 + \bar\mu_1 - \frac{1}{m} \sum_{j=1}^{\ell'} \bar\mu_j \ge 0 \\
\bar p_2 &= a_2 + \bar\mu_2 - \frac{1}{m} \sum_{j=1}^{\ell'} \bar\mu_j \ge 0 \\
&\dots \\
\bar p_{\ell'-1} &= a_{\ell'-1} + \bar\mu_{\ell'-1} - \frac{1}{m} \sum_{j=1}^{\ell'} \bar\mu_j \ge 0 \\
\bar p_{\ell'} &= a_{\ell'} + \bar\mu_{\ell'} - \frac{1}{m} \sum_{j=1}^{\ell'} \bar\mu_j = 0 \\
\bar p_{\ell'+1} &= a_{\ell'+1} - \frac{1}{m} \sum_{j=1}^{\ell'} \bar\mu_j \ge 0 \\
&\dots \\
\bar p_r &= a_r - \frac{1}{m} \sum_{j=1}^{\ell'} \bar\mu_j \ge 0
\end{aligned} \tag{14}$$

due to (6) and (7). Since $a_1 \le a_2 \le ... \le a_{\ell'}$ the first $\ell' - 1$ inequalities of (14) can be satisfied iff

$$\bar\mu_i \ge \bar\mu_{\ell'} > 0, \quad i = 1, ..., \ell' - 1. \tag{15}$$

Therefore, from (7) we have

$$
\begin{aligned}
&\bar{p}_1 = a_1 + \bar{\mu}_1 - \frac{1}{m}\sum_{j=1}^{\ell'} \bar{\mu}_j = 0 \\
&\bar{p}_2 = a_2 + \bar{\mu}_2 - \frac{1}{m}\sum_{j=1}^{\ell'} \bar{\mu}_j = 0 \\
&\ldots \\
&\bar{p}_{\ell'-1} = a_{\ell'-1} + \bar{\mu}_{\ell'-1} - \frac{1}{m}\sum_{j=1}^{\ell'} \bar{\mu}_j = 0 \\
&\bar{p}_{\ell'} = a_{\ell'} + \bar{\mu}_{\ell'} - \frac{1}{m}\sum_{j=1}^{\ell'} \bar{\mu}_j = 0 \\
&\bar{p}_{\ell'+1} = a_{\ell'+1} - \frac{1}{m}\sum_{j=1}^{\ell'} \bar{\mu}_j \geq 0 \\
&\ldots \\
&\bar{p}_r = a_r - \frac{1}{m}\sum_{j=1}^{\ell'} \bar{\mu}_j \geq 0.
\end{aligned}
\tag{16}
$$

Taking the sum of the first ℓ' equalities of (16), one can easily obtain that

$$
\sum_{j=1}^{\ell'} \bar{\mu}_j = S_{\ell'}
$$

and

$$
\bar{\mu}_i = -a_i + \frac{1}{m} S_{\ell'}, \quad i = 1, \ldots, \ell' \tag{17}
$$

Thus, if $\ell' = \ell$ then $\bar{\mu}_i = \mu_i, \quad i = 1, \ldots, r$. If $\ell' < \ell$, then (16) and the definition of ℓ yield

$$
\bar{p}_{\ell'+1} = a_{\ell'+1} - \frac{1}{m} S_{\ell'} < 0,
$$

which is a contradiction. If $\ell' > \ell$, then the definition of ℓ and monotonicity of a_i in i imply

$$
\begin{aligned}
&a_{\ell+1} - \frac{1}{m} S_\ell \geq 0 \\
&a_{\ell+2} - \frac{1}{m} S_\ell \geq 0 \\
&\ldots \\
&a_{\ell'} - \frac{1}{m} S_\ell \geq 0.
\end{aligned}
\tag{18}
$$

Taking the sum of these inequalities yields

$$
\sum_{i=\ell+1}^{\ell'} a_i - \frac{\ell' - \ell}{m} S_\ell \geq 0,
$$

which can be represented using the definition of S_ℓ as

$$\sum_{i=\ell+1}^{\ell'} a_i + \frac{\ell'-\ell}{m-\ell}\sum_{i=1}^{\ell} a_i \geq 0. \tag{19}$$

On the other hand, we have from (17)

$$\bar{\mu}_{\ell+1} = -a_{\ell+1} + \frac{1}{m}S_{\ell'} = -(a_{\ell+1} - \frac{1}{m}S_\ell) - (\frac{1}{m}S_\ell - \frac{1}{m}S_{\ell'}). \tag{20}$$

By the definition of ℓ,

$$a_{\ell+1} - \frac{1}{m}S_\ell \geq 0. \tag{21}$$

From (8) and (19),

$$\begin{aligned}
\frac{1}{m}S_\ell - \frac{1}{m}S_{\ell'} &= -\frac{1}{m-\ell}\sum_{i=1}^{\ell} a_i + \frac{1}{m-\ell'}\sum_{i=1}^{\ell'} a_i \\
&= -\frac{1}{m-\ell}\sum_{i=1}^{\ell} a_i + \frac{1}{m-\ell'}\sum_{i=1}^{\ell} a_i + \frac{1}{m-\ell'}\sum_{i=\ell+1}^{\ell'} a_i \\
&= \frac{\ell'-\ell}{(m-\ell')(m-\ell)}\sum_{i=1}^{\ell} a_i + \frac{1}{m-\ell'}\sum_{i=\ell+1}^{\ell'} a_i \qquad (22) \\
&= \frac{1}{m-\ell'}\left(\sum_{i=\ell+1}^{\ell'} a_i + \frac{\ell'-\ell}{m-\ell}\sum_{i=1}^{\ell} a_i\right) \geq 0.
\end{aligned}$$

Using (20)-(22), we obtain that $\bar{\mu}_{\ell+1} \leq 0$, which is a contradiction with (15). Thus, Proposition 4 is proved.

Let us come back to the assumption $\mu_0 = 1/2$. If (11) gives a nonzero vector $p \in R^{m-1}$, then the vector $\hat{p} = p/\xi$, where $\xi = \left(\sum_{i=1}^{m-1} p_i\right)^2 + \sum_{i=1}^{m-1} p_i^2$, satisfies all conditions of the Kuhn - Tucker theorem with $\mu_0 = \xi/2$ and μ_i given by (10). Hence $\hat{p}$ is a minimizing vector of (4).

Consider the case where (11) yields $p = 0$. If $\hat{p}$ is a minimizing vector of (4), which satisfies the conditions of the Kuhn - Tucker theorem with $\mu_0 > 0$, then the vector $2\mu_0\hat{p}$ satisfies the system (6) and the complementarity conditions (7). Because of the uniqueness of the solution of (6),(7) we conclude that $\hat{p} = 0$, which is a contradiction with $\mu_0 > 0$. Therefore, $\hat{p}$ satisfies the conditions of the Kuhn - Tucker theorem with $\mu_0 = 0$. In this case one can easily obtain from (5) that $d_i = 0, i = r+1, ..., m-1$ (because $\mathcal{L}(\cdot, \mu, \mu_0)$ attains its minimum at $\hat{p}$), $d_i \leq 0, i = 1, ..., r$ (because there should be $d_i + \mu_i = 0, \mu_i \geq 0$) and $d_i = 0$ whenever $\hat{p}_i > 0, i = 1, ..., r$ (complementarity conditions). This implies $\sum_{i=1}^{m-1} -d_i\hat{p}_i = 0$. On the other hand, the vector $p = 0$ obtained using (11), i.e. under the assumption $\mu_0 = 1/2$, gives the same value.

Remark 1. It is easily seen that $a_1 \leq a_2 \leq ... \leq a_r$ iff $h_1 \geq h_2 \geq ... \geq h_r$. Some calculations show that

$$a_i - \frac{1}{m} S_{i-1} = \frac{1}{m-i+1} \sum_{j=i}^{m} h_j - h_i.$$

Hence,

$$\ell = \begin{cases} 0 \;\text{ if }\; a_1 = \frac{1}{m}\sum_{j=1}^{m} h_j - h_1 \geq 0, \\ \\ \max\{k \in \overline{1,r} : \forall i \in \overline{1,k} \;\; (\frac{1}{m-i+1}\sum_{j=i}^{m} h_j - h_i < 0)\} \;\text{ if }\; a_1 < 0. \end{cases}$$

From (11),

$$p_i \; = \; \frac{1}{m-\ell} \sum_{j=\ell+1}^{m} h_j - h_i, \quad i = \ell+1, ..., m-1,$$

$$p_m \stackrel{def}{=} -\sum_{j=1}^{m-1} p_j = \frac{1}{m-\ell} \sum_{j=\ell+1}^{m} h_j - h_m.$$

Therefore, the procedure of finding $p^0(\alpha, h)$ is formulated in terms of the components of h. One can also check that

$$\omega(\alpha, h) = -\left(\sum_{i=\ell+1}^{m} h_i^2 - \frac{1}{m-\ell} \left(\sum_{i=\ell+1}^{m} h_i \right)^2 \right)^{\frac{1}{2}}.$$

3 Modifications of Algorithm 1

The following algorithm uses the idea to avoid updating the gradient $Dg(\alpha)$ at each iteration (see [3]).

Algorithm 2

Step 0: Choose an initial point $\alpha^0 \in \Lambda$ and set $k := 0$.
Step 1: Update the gradient $h^k = Dg(\alpha^k)$, set $\alpha^{k,0} := \alpha^k$ and $l := 0$.
Step 2: Find $p^{k,l} = p^0(\alpha^{k,l}, h^k)$, $\omega^{k,l} = \omega(\alpha^{k,l}, h^k)$.
If $\omega^{k,l} = 0$ and $l = 0$, then $\alpha^* := \alpha^k$ and stop.
If $\omega^{k,l} = 0$ and $l \neq 0$, then $\alpha^{k+1} := \alpha^{k,l}$, $k := k+1$ and go to Step 1.
Step 3: Find

$$\lambda_{max} = \max\{\lambda \geq 0 : \alpha^{k,l} + \lambda p^{k,l} \in \Lambda\} = \min_{p_i^{k,l} < 0} \left\{ -\frac{\alpha_i^{k,l}}{p_i^{k,l}} \right\}$$

($\lambda_{max} > 0$ because $p^{k,l} \in \mathcal{P}(\alpha^{k,l})$).

Step 4: Make a line search along $p^{k,l}$ in order to find

$$\lambda' = \arg \min_{\lambda \in [0, \lambda_{max}]} g(\alpha^{k,l} + \lambda p^{k,l})$$

($\lambda' > 0$ and $g(\alpha^{k,l} + \lambda' p^{k,l}) < g(\alpha^{k,l})$ due to Propositions 1, 2).
Step 5: Let

$$\alpha^{k,l+1} := \alpha^{k,l} + \lambda' p^{k,l}, \quad l := l + 1$$

and go to Step 2.

The next modification uses the idea to observe the subspace in which the process is running and to minimize $g(\cdot)$ over this subspace. For arbitrary $I \subset \overline{1,m}$, denote $L(I) = \{\alpha \in R^m : \alpha_i = 0 \text{ for } i \in I, \ \sum_{i=1}^m a_i = 1\}$.

Algorithm 3

Step 0: Choose an initial point $\alpha^0 \in \Lambda$ and set $k := 0$.
Step 1: Update the gradient $h^k = Dg(\alpha^k)$, set $\alpha^{k,0} := \alpha^k$ and $l := 0$.
Step 2: Find $p^{k,l} = p^0(\alpha^{k,l}, h^k)$, $\omega^{k,l} = \omega(\alpha^{k,l}, h^k)$.
If $\omega^{k,l} = 0$ and $l = 0$, then $\alpha^* := \alpha^k$ and stop.
If $\omega^{k,l} = 0$ and $l \neq 0$, then

(a) Let $\alpha^{k+1} := \alpha^{k,l}$ and find $I^0(\alpha^{k+1})$.
If $I^0(\alpha^{k+1}) \neq I^0(\alpha^k)$, then set $k := k + 1$ and go to Step 1.
(b) Find $\alpha' = \arg \min_{\alpha \in L(I^0(\alpha^k))} g(\alpha)$, set

$$\begin{aligned} \rho' &= \max\{\rho : \alpha^k + \rho(\alpha' - \alpha^{k+1}) \in \Lambda\}, \\ \rho'' &= \min\{1, \rho'\}, \\ \alpha^{k+1} &= \alpha^{k+1} + \rho''(\alpha' - \alpha^{k+1}), \\ k &:= k + 1 \end{aligned}$$

and go to Step 1.
Step 3: Find

$$\lambda_{max} = \max\{\lambda \geq 0 : \alpha^{k,l} + \lambda p^{k,l} \in \Lambda\} = \min_{p_i^{k,l} < 0} \left\{ -\frac{\alpha_i^{k,l}}{p_i^{k,l}} \right\}$$

($\lambda_{max} > 0$ because $p^{k,l} \in \mathcal{P}(\alpha^{k,l})$).
Step 4: Perform a line search along $p^{k,l}$ in order to find

$$\lambda' = \arg \min_{\lambda \in [0, \lambda_{max}]} g(\alpha^{k,l} + \lambda p^{k,l})$$

($\lambda' > 0$ and $g(\alpha^{k,l} + \lambda' p^{k,l}) < g(\alpha^{k,l})$ due to Propositions 1,2).
Step 5: Let

$$\alpha^{k,l+1} := \alpha^{k,l} + \lambda' p^{k,l}, \quad l := l + 1$$

and go to Step 2.

Note that the last variant of the algorithm uses a recursive call to update the minimum of $g(\cdot)$ over the set $L(I^0(\alpha^k)) \cap \Lambda$.

Algorithm 4. All steps are the same as for Algorithm 3, but substeps (a) and (b) look as follows:

(a) Set $\alpha^{k+1} := \alpha^{k,l}$ and find $I^0(\alpha^{k+1})$.
If $I^0(\alpha^{k+1}) \neq I^0(\alpha^k)$ or $|I^0(\alpha^k)| = m$, then $k := k+1$ and go to Step 1.
(b) call Algorithm 4 to find $\alpha^{k+1} = \arg\min_{\alpha \in \Lambda'} g(\alpha)$,
where $\Lambda' = L(I^0(\alpha^k)) \cap \Lambda$,
set $k := k+1$ and go to Step 1.

4 Numerical examples

Let us consider some numerical examples characterizing the comparative speed of the algorithms proposed. These examples were implemented with a QuickBASIC program on a PC (80486, 98 MHz).

Example 1. $Z = \{z : z = (\delta_1, \delta_2, ..., \delta_n)^T,\ \delta_j = \pm 1, j = 1, ..., n\}$ (vertices of the unit cube in R^n). The number of vertices is $m = 2^n$. The objective function is $f(x) = |x - z^c|^2$, where $z^c = (10, 0.7, 0, ..., 0)^T$. Thus $g(\alpha) = |\sum_{j=1}^m \alpha_j (z^j - z^c)|^2$. The programs realizing Algorithms 1-4 were run with the initial point $\alpha^0 = (1, 0, ..., 0)^T$ (this point corresponds to the point $z^1 = (1, ..., 1)^T$). The results obtained for the dimension $n = 7$ are presented in the following table.

Algorithm	Iterations	Time(s)
1	28	40.01
2	53	43.61
3	49	116.16
4	30	8.68

Example 2. $Z = \{z^i : z^i = (\xi_1, ..., \xi_n)^T, i = 1, ...m\}$, where ξ_j are random values uniformly distributed on the interval $[-1, 1]$. The objective function and initial point α^0 were the same as in Example 1 except for $z^c = (10, 0, ..., 0)^T$. The results obtained for Algorithm 1-4 are presented in the following table. For Algorithm 4 the summarized number of iteration, including all recursive calls, is given. Although this number is greater then corresponding numbers for other algorithms, most of iterations are running in the space of low dimension and, as a rule, Algorithm 4 wins.

n=20	Iterations			
Alg.	m=100	m=200	m=300	m=500
1	74	83	72	190
2	79	94	91	228
3	14	24	39	58
4	118	95	92	554

n=20	Time(s)			
Alg.	m=100	m=200	m=300	m=500
1	48.43	154.71	267.54	1630.07
2	43.77	127.16	195.75	1413.00
3	5.00	19.28	52.61	205.25
4	8.12	19.34	28.01	172.85

References

1. Rosen J.B. (1960) The gradient Projection Method of Nonlinear Programming, Part 1, Linear Constraints. SIAM Journal on Applied Mathematics, Vol. 8 : 181-217.
2. Rosen J.B. (1961) The gradient Projection Method of Nonlinear Programming, Part 2, Nonlinear Constraints. SIAM Journal on Applied Mathematics, Vol. 9 : 514-532.
3. Bertsekas D.P. (1982) Projected Newton methods for optimization problems with simple constraints. Siam J. Control Optim. 20: 221-246.
4. Michelot C. (1986) A finite Algorithm for Finding the Projection of a Point onto the Canonical Simplex of R^n. JOTA, Vol. 50, No.1 : 195-200.