Mathematical Programming 40 (1988) 1-14
North-Holland

A SUCCESSIVE PROJECTION METHOD

Shih-Ping HAN
Department of Mathematics, University of Illinois, Urbana-Champaign, IL 61801, USA

Received 8 August 1986
Revised manuscript received 19 January 1987

It is of both theoretical and practical interest to find the projection of a point to the intersection of a finite number of closed convex sets by a sequence of projections to the individual sets successively. In this paper we study such a method and analyze its convergence properties. A main feature of the method is its capability to decompose the projection problem into several small ones. For some structured sparse problems these small subproblems can be solved independently and the presented method has a potential use in parallel computation.

Key words: Minimization, successive projection.

1. Introduction

The problem to be considered in this paper is:

$$\text{(P)} \qquad \begin{aligned} &\min \quad f(x) := \tfrac{1}{2}\langle x-d, Q(x-d)\rangle \\ &\text{s.t.} \quad x \in C := C_1 \cap \cdots \cap C_m \end{aligned}$$

where d is a given point and C_i, $i = 1, \ldots, m$, are closed convex sets in R^n and Q is an n by n positive definite symmetric matrix. When the norm $\|x\|_Q := \langle x, Qx\rangle^{1/2}$ is used, the objective function $f(x)$ can be written as $\frac{1}{2}\|x-d\|_Q^2$ and the problem is simply to find the projection of the point d to the intersection of C_i's. It is of both practical and theoretical interest to study how to find such a projection by doing a sequence of projections to each individual C_i successively. In the paper we propose such a method and discuss some of its applications.

A straightforward successive projection method is to do projection to each set one after another. A simple example can show that this method does not work generally. In our method, an outer normal vector is added before each projection; in so doing, the generated points converge to the solution of the problem under very mild regularity conditions. Even though the proposed method bears a strong geometrical meaning for dealing with problem (P), it can also be viewed as a method for solving some dual problems. Indeed, the convergence properties of the method are actually analyzed through studying the monotone decreasing behavior of an objective function of a dual problem at a generated sequence of points.

This material is based on work supported in part by the National Science Foundation under Grant DMS-8602419 and by the Center for Supercomputing Research and Development, University of Illinois.

A main feature of the method is its capability to decompose a large problem into several small ones. Its efficiency depends on how much effort is needed to carry out each projection. With C_i subspaces or halfspaces, the successive projection method is extremely simple and is closely related to the SOR method [1, 2, 4]. Also, for some structured sparse problems the method can be implemented to do projections to several sets simultaneously and has a potential use in parallel computation.

The paper is organized as follows. In Section 2 we describe the method. For analyzing the method we study two dual problems of (P) in Section 3. Convergence analysis of the method is given in Section 4 and its applications are given in Section 5.

Most symbols and notation used in this paper are the same as those in [7]. Recall that $\delta(\cdot\,|\,C)$ is the indicator function of the set C defined as: $\delta(x\,|\,C)=0$ if x is in C, $\delta(x\,|\,C)=\infty$ otherwise; $\delta^*(\cdot\,|\,C)$ is the support functions of the set C defined as: $\delta^*(y\,|\,C)=\sup\{\langle y, x\rangle\,|\,x\in C\}$. We also use ri C to denote the relative interior of C. Unless specified, the symbol $\sum$ is to denote the summation from $i=1$ to m, where m is the number of sets C_i. It is also noted that for practical consideration we restrict ourselves to the space R^n, even though many results hold equally well for a general Hilbert space.

2. The method

To facilitate our discussion, we use $P_i(\cdot)$ to denote the projector to the set C_i with respect to the norm $\|\cdot\|_Q$; that is, for a given point a in R^n the point $\bar{x} := P_i(a)$ is the unique solution to the following problem:

$$\begin{aligned} &\min \quad \tfrac{1}{2}\langle x-a, Q(x-a)\rangle \\ &\text{s.t.} \quad x\in C_i. \end{aligned}$$

Equivalently, by the minimum principle, the point $\bar{x}$ satisfies the following conditions:

(a) $\bar{x}\in C_i$;
(b) $\langle z-\bar{x}, Q(a-\bar{x})\rangle \leq 0 \quad \forall z\in C_i$.

We now state our method:

The method. At the beginning, we set $y_1^{(0)}=\cdots=y_m^{(0)}=0$ and $x_m^{(0)}=d$, which is the given projected point. In each subsequent iteration, we compute $2m$ vectors $x_1^{(k)}, y_1^{(k)}, \ldots, x_m^{(k)}, y_m^{(k)}$ as follows: set $x_0^{(k)}=x_m^{(k-1)}$ and, for $i=1,\ldots,m$, calculate

$$z := x_{i-1}^{(k)} + y_i^{(k-1)},$$

$$x_i^{(k)} := P_i(z),$$

$$y_i^{(k)} := z - P_i(z).$$

It seems more natural to consider the following straightforward successive projection method: $x_m^{(0)} = d$ and for $k = 1, 2, \ldots$ set $x_0^{(k)} = x_m^{(k-1)}$ and also for $i = 1, \ldots, m$, set $x_i^{(k)} := P_i(x_{i-1}^{(k)})$. The method does not work generally. This can be seen from the following simple example in R^2. Let $C_1 := \{(\xi_1\ \xi_2) \mid \xi_2 \leq 0\}$ and $C_2 := \{(\xi_1\ \xi_2) \mid \xi_1 + \xi_2 \leq 0\}$, the straightforward projection method does not work for any point d outside C_1 and C_2 with $\xi_1 \neq 0$ and $\xi_1 \neq \xi_2$. In our method the outer normal vector $y_i^{(k)}$ of the set C_i at $x_i^{(k)}$ is calculated and we add the previous outer normal $y_i^{(k-1)}$ to $x_{i-1}^{(k)}$ before projecting it to the set C_i. Therefore, by each projection we replace an old outer normal vector by a new one and the sequence of normal vectors are intended to converge to a solution of a dual problem of (P).

3. Dual problems

As mentioned before, for analyzing the successive projection method we need to study some dual problems of (P). One relevant dual problem is:

(D) $\quad \min g(y) := f(y) + \delta^*(y \mid QC)$

where y is in R^n. Clearly, (D) is a slight modification of Fenchel's dual [7]. Similar to Moreau's result [6; 7, Theorem 31.5], the following theorem describes the relationship between problems (P) and (D).

Theorem 3.1. *Let $\bar{x}$, $\bar{y}$ be two points in R^n such that $\bar{x} + \bar{y} = d$ and let $C \neq \emptyset$, $C \neq R^n$. Then the following statements are equivalent*:

(a) *$\bar{x}$ solves* (P);
(b) *$\bar{y}$ solves* (D);
(c) *$\bar{x} \in C$ and $\delta^*(\bar{y} \mid QC) = \langle \bar{x}, Q\bar{y} \rangle$.*

Furthermore, the solutions of (P) *and* (D) *exist and are unique.*

Proof. In view of $\partial f(\bar{x}) = \{-Q\bar{y}\}$ and $\partial f(\bar{y}) = \{-Q\bar{x}\}$ we have

$$\begin{aligned}
\bar{x} \text{ solves (P)} &\Leftrightarrow 0 \in \partial f(\bar{x}) + \partial\delta(\bar{x} \mid C) \\
&\Leftrightarrow Q\bar{y} \in \partial\delta(\bar{x} \mid C) \\
&\Leftrightarrow \delta(\bar{x} \mid C) + \delta^*(Q\bar{y} \mid C) = \langle \bar{x}, Q\bar{y} \rangle \\
&\Leftrightarrow (C). \\
\bar{y} \text{ solves (D)} &\Leftrightarrow 0 \in \partial f(\bar{y}) + \partial\delta^*(\bar{y} \mid QC) \\
&\Leftrightarrow Q\bar{x} \in \partial\delta^*(\bar{y} \mid QC) \\
&\Leftrightarrow \delta(Q\bar{x} \mid QC) + \delta^*(\bar{y} \mid QC) = \langle \bar{x}, Q\bar{y} \rangle \\
&\Leftrightarrow \delta(\bar{x} \mid C) + \delta^*(Q\bar{y} \mid C) = \langle \bar{x}, Q\bar{y} \rangle \\
&\Leftrightarrow (C).
\end{aligned}$$

The last statement follows from the fact that a projection on a non-empty closed convex set exists. The uniqueness is a result of strict convexity. □

Another related dual problem is:

$$\text{(D}'\text{)} \qquad \min h(y_1, \ldots, y_m) := f(y_1 + \cdots + y_m) + \sum \delta^*(y_i \mid QC_i)$$

where y_i are in R^n and the minimization is over the R^{mn} space. We now discuss the connections between the three problems (P), (D) and (D′). First we give a necessary and sufficient condition for a solution to (D′).

Lemma 3.2. *A vector* $(\bar{y}_1, \ldots, \bar{y}_m)$ *in* R^{mn} *solves* (D′) *if and only if*

(a) $\bar{x} := d - \bar{y}_1 \cdots - \bar{y}_m$ *is in* C;
(b) *for each* i, $\langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(\bar{y}_i \mid QC_i)$.

Proof. By direct calculation, we have

$$\partial h(y_1, \ldots, y_m) = \begin{pmatrix} -Qx + \partial\delta^*(y_1 \mid QC_1) \\ \vdots \\ -Qx + \partial\delta^*(y_m \mid QC_m) \end{pmatrix}$$

where $x = d - y_1 \cdots - y_m$. Hence,

$$\begin{aligned} (\bar{y}_1, \ldots, \bar{y}_m) \text{ solves (D}'\text{)} &\Leftrightarrow 0 \in \partial h(\bar{y}_1, \ldots, \bar{y}_m) \\ &\Leftrightarrow Q\bar{x} \in \partial\delta^*(\bar{y}_i \mid QC_i), \quad i = 1 \ldots m \\ &\Leftrightarrow \langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(\bar{y}_i \mid QC_i) + \delta(\bar{x} \mid C_i), \quad i = 1 \ldots m \\ &\Leftrightarrow \langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(\bar{y}_i \mid QC_i), \bar{x} \in C_i, \quad i = 1 \ldots m. \qquad \square \end{aligned}$$

The following theorem is fundamental for our analysis of convergence in the next section.

Theorem 3.3. *Let* $C_1, \ldots, C_k$ *be polyhedral and* $C_{k+1}, \ldots, C_m$ *be closed convex sets such that* $C_1 \cap \cdots \cap C_k \cap \operatorname{ri} C_{k+1} \cap \cdots \cap \operatorname{ri} C_m \neq \emptyset$. *A point* $\bar{x}$ *solves* (P) *if and only if there exist points* $\bar{y}_1, \ldots, \bar{y}_m$ *such that* $\bar{x} + \bar{y}_1 + \cdots + \bar{y}_m = d$ *and* $(\bar{y}_1, \ldots, \bar{y}_m)$ *solves* (D′). *Furthermore, in this case, the vector* $\bar{y} := \bar{y}_1 + \cdots + \bar{y}_m$ *solves* (D) *and*

$$\inf h(y_1, \ldots, y_m) = \inf g(y).$$

Proof. "$\Rightarrow$" If $\bar{x}$ solves (P), let $\bar{y} := d - \bar{x}$ and it follows that $Q\bar{y} \in \partial\delta(\bar{x} \mid C)$. By the assumption, we have that $\partial\delta(\bar{x} \mid C) = \partial\delta(\bar{x} \mid C_1) + \cdots + \partial\delta(\bar{x} \mid C_m)$. Therefore, there exist $\bar{y}_1, \ldots, \bar{y}_m$ such that $\bar{y} = \bar{y}_1 + \cdots + \bar{y}_m$ and $Q\bar{y}_i \in \partial\delta(\bar{x} \mid C_i)$. This implies that $\langle \bar{x}, Q\bar{y}_i \rangle = \delta^*(Q\bar{y}_i \mid C_i) = \delta^*(\bar{y}_i \mid QC_i)$ and by Lemma 3.2 the point $(\bar{y}_1, \ldots, \bar{y}_m)$ solves (D′).

"$\Leftarrow$" If $(\bar{y}_1, \ldots, \bar{y}_m)$ solves (D′), $\bar{y} = \bar{y}_1 + \cdots + \bar{y}_m$ and $\bar{x} + \bar{y} = d$, then it follows from Lemma 3.2 again that $\langle \bar{x}, Q\bar{y}\rangle = \sum \langle \bar{x}, Q\bar{y}_i\rangle = \sum \delta^*(\bar{y}_i \mid QC_i)$ and $\bar{x} \in C$. By the assumption and Corollary 16.4.1 of [7], we have

$$\begin{aligned}\langle \bar{x}, Q\bar{y}\rangle &= \sum \delta^*(\bar{y}_i \mid QC_i)\\ &\geqslant \delta^*(\bar{y} \mid QC).\end{aligned}$$

On the other hand, it follows from $\bar{x} \in C$ that $\delta^*(\bar{y} \mid QC) \geqslant \langle \bar{x}, Q\bar{y}\rangle$. Therefore, we have that $\langle \bar{x}, Q\bar{y}\rangle = \delta^*(\bar{y} \mid QC)$ and, by Theorem 3.1, $\bar{x}$ solves (P) and $\bar{y}$ solves (D).

The last statement follows from $\inf h = h(\bar{y}_1, \ldots, \bar{y}_m) = g(\bar{y}) = \inf g$. □

4. Convergence analysis

The successive projection method presented here actually produces three sequences $\{(y_1^{(k)}, \ldots, y_m^{(k)})\}$, $\{y^{(k)}\}$ and $\{x^{(k)}\}$, where $y^{(k)} = y_1^{(k)} + \cdots + y_m^{(k)}$ and $x^{(k)} = x_m^{(k)}$. The three sequences will be shown to converge to the solutions of (D′), (D) and (P) respectively. The analysis of their convergence is based on establishing the monotone decrease of the values $\{h(y_1^{(k)}, \ldots, y_m^{(k)})\}$. For convenience, in the following discussion we assume $C \neq R^n$ and $C \neq \emptyset$ and use $\bar{x}$ and $\bar{y}$ to denote the unique solutions of (P) and (D), respectively.

Lemma 4.1. *For each $i = 1, \ldots, m$ and for each $k = 1, 2, \ldots$ the following statements hold:*

(a) $x_i^{(k)} \in C_i$.
(b) $x^{(k)} + y^{(k)} = d$ *where* $x^{(k)} = x_m^{(k)}$ *and* $y^{(k)} = y_1^{(k)} + \cdots + y_m^{(k)}$.
(c) $\langle x_i^{(k)}, Qy_i^{(k)}\rangle = \delta^*(y_i^{(k)} \mid QC_i)$.

Proof. Statement (a) is obviously true because $x_i^{(k)}$ is the projection of some point onto C_i. To prove (b), we have

$$y_i^{(k)} = x_{i-1}^{(k)} + y_i^{(k-1)} - x_i^{(k)}.$$

By adding the above equations from $i = 1$ to m, we have that

$$\sum y_i^{(k)} = x_0^{(k)} - x_m^{(k)} + \sum y_i^{(k-1)}$$

which, in conjuction with $x_0^{(k)} = x_m^{(k-1)}$, implies

$$x^{(k)} + y^{(k)} = x^{(k-1)} + y^{(k-1)}.$$

Then, statement (b) follows from $y^{(0)} = y_1^{(0)} + \cdots + y_m^{(0)} = 0$ and $x^{(0)} = x_m^{(0)} = d$.

Since $x_i^{(k)}$ is the projection of $x_{i-1}^{(k)} + y_i^{(k-1)}$ onto the closed convex set C_i, by the minimum principle we have

$$\langle z - x_i^{(k)}, Q(x_{i-1}^{(k)} + y_i^{(k-1)} - x_i^{(k)})\rangle \leqslant 0 \quad \forall z \in C_i.$$

By $y_i^{(k)} = x_{i-1}^{(k)} + y_i^{(k-1)} - x_i^{(k)}$ and the definition of the support function $\delta^*(\cdot \mid QC_i)$, we get (c). □

Lemma 4.2. $h(y_1^{(k-1)}, \ldots, y_m^{(k-1)}) \geq h(y_1^{(k)}, \ldots, y_m^{(k)}) + \frac{1}{2}\sum \| y_i^{(k)} - y_i^{(k-1)} \|_Q^2$.

Proof. By the way we calculate $y_i^{(k)}$, we have

$$x_{i-1}^{(k)} = x_i^{(k)} + y_i^{(k)} - y_i^{(k-1)}.$$

Hence,

$$\| x_{i-1}^{(k)} \|_Q^2 = \| x_i^{(k)} \|_Q^2 + \| y_i^{(k)} - y_i^{(k-1)} \|_Q^2 + 2\langle x_i^{(k)}, Q(y_i^{(k)} - y_i^{(k-1)})\rangle.$$

Then it follows from (a) and (c) of Lemma 4.1 and the above equation that

$$\begin{aligned} \tfrac{1}{2}\| x_{i-1}^{(k)} \|_Q^2 + \delta^*(y_i^{(k-1)} \mid QC_i) &\geq \tfrac{1}{2}\| x_{i-1}^{(k)} \|_Q^2 + \langle x_i^{(k)}, Qy_i^{(k-1)}\rangle \\ &= \tfrac{1}{2}\| x_i^{(k)} \|_Q^2 + \langle x_i^{(k)}, Qy_i^{(k)}\rangle + \tfrac{1}{2}\| y_i^{(k)} - y_i^{(k-1)} \|_Q^2 \\ &= \tfrac{1}{2}\| x_i^{(k)} \|_Q^2 + \delta^*(y_i^{(k)} \mid QC_i) + \tfrac{1}{2}\| y_i^{(k)} - y_i^{(k-1)} \|_Q^2. \end{aligned}$$

By adding the above equations from $i = 1$ to m, we get

$$\begin{aligned} &\tfrac{1}{2}\| x_0^{(k)} \|_Q^2 + \textstyle\sum \delta^*(y_i^{(k-1)} \mid QC_i) \\ &\quad \geq \tfrac{1}{2}\| x_m^{(k)} \|_Q^2 + \textstyle\sum \delta^*(y_i^{(k)} \mid QC_i) + \tfrac{1}{2}\sum \| y_i^{(k)} - y_i^{(k-1)} \|_Q^2. \end{aligned}$$

Therefore, by $x_0^{(k)} = x_m^{(k-1)} = d - y_1^{(k-1)} - \cdots - y_m^{(k-1)}$ and $x_m^{(k)} = d - y_1^{(k)} - \cdots - y_m^{(k)}$, the above inequality is just

$$h(y_1^{(k-1)}, \ldots, y_m^{(k-1)}) \geq h(y_1^{(k)}, \ldots, y_m^{(k)}) + \tfrac{1}{2}\sum \| y_i^{(k)} - y_i^{(k-1)} \|_Q^2. \qquad \square$$

Corollary 4.3. *For $i = 1, \ldots, m$ the following statements hold:*

(a) $\lim \| y_i^{(k)} - y_i^{(k-1)} \|_Q = 0$;
(b) $\lim \| x_{i-1}^{(k)} - x_i^{(k)} \|_Q = 0$.

Consequently, $\lim \| y^{(k)} - y^{(k-1)} \|_Q = 0$ *and* $\lim \| x^{(k)} - x^{(k-1)} \|_Q = 0$.

Proof. By adding the inequalities in Lemma 4.2 from 0 to k, we get

$$h(y_1^{(0)}, \ldots, y_m^{(0)}) - h(y_1^{(k)}, \ldots, y_m^{(k)}) \geq \tfrac{1}{2}\sum_{j=1}^{k} \left(\sum \| y_i^{(j)} - y_i^{(j-1)} \|_Q^2\right).$$

Noticing that for any $(y_1, \ldots, y_m)$, we have

$$\begin{aligned} h(y_1, \ldots, y_m) &= f(y_1 + \cdots + y_m) + \textstyle\sum \delta^*(y_i \mid QC_i) \\ &\geq g(y) \\ &\geq g(\bar{y}) \end{aligned}$$

where $y = y_1 + \cdots + y_m$ and $\bar{y}$ is the solution to (D). Therefore, h is bounded below and for any k

$$h(y_1^{(0)}, \ldots, y_m^{(0)}) - g(\bar{y}) \geq \tfrac{1}{2}\sum_{j=1}^{k} \left(\sum \| y_i^{(j)} - y_i^{(j-1)} \|_Q^2\right).$$

This implies that $\lim \| y_i^{(k)} - y_i^{(k-1)} \|_2 = 0$. Statement (b) also follows because

$$x_{i-1}^{(k)} - x_i^{(k)} = y_i^{(k)} - y_i^{(k-1)}.$$

The last statement is an obvious consequence of (a) and (b). □

Proposition 4.4. *Any accumulation point of* $\{(y_1^{(k)}, \ldots, y_m^{(k)})\}$ *solves* (D′).

Proof. Without loss of generality, we may assume $(y_1^{(k)}, \ldots, y_m^{(k)}) \to (\tilde{y}_1, \ldots, \tilde{y}_m)$. Let $\tilde{x} := d - \tilde{y}_1 - \cdots - \tilde{y}_m$, then it follows from Corollary 4.3 that for each i, we have $\lim x_i^{(k)} = \tilde{x}$. Therefore, $\tilde{x} \in C$ because $x_i^{(k)} \in C_i$ for each i. Hence, by the lower semi-continuity of δ^*, we have that

$$\begin{aligned}\langle \tilde{x}, Q\tilde{y}_i \rangle &\leq \delta^*(\tilde{y}_i \mid QC_i) \\ &\leq \liminf \delta^*(y_i^{(k)} \mid QC_i) \\ &= \liminf \langle x_i^{(k)}, Qy_i^{(k)} \rangle \\ &= \langle \tilde{x}, Q\tilde{y}_i \rangle\end{aligned}$$

which proves $\langle \tilde{x}, Q\tilde{y}_i \rangle = \delta^*(\tilde{y}_i \mid QC_i)$. It now follows from and Lemma 3.2 that $(\tilde{y}_1, \ldots, \tilde{y}_m)$ solves (D′). □

Theorem 4.5. *Let* $C_1, \ldots, C_k$ *be polyhedral and* $C_{k+1}, \ldots, C_m$ *be closed convex sets such that* $C_1 \cap \cdots \cap C_k \cap \operatorname{ri} C_{k+1} \cap \cdots \cap \operatorname{ri} C_m \neq \emptyset$. *If* $\{(y_1^{(k)}, \ldots, y_m^{(k)})\}$ *has an accumulation point then the sequences* $\{x^{(k)}\}$ *and* $\{y^{(k)}\}$ *converge to the solutions* (P) *and* (D) *respectively.*

Proof. By the assumption and Theorem 3.3 we have $\inf h = \inf g$. On the other hand, by Proposition 4.4 and the monotone decrease of $\{h(y_1^{(k)}, \ldots, y_m^{(k)})\}$, we also have

$$\lim h(y_1^{(k)}, \ldots, y_m^{(k)}) = \inf h.$$

Therefore, it follows from $h(y_1^{(k)}, \ldots, y_m^{(k)}) \geq g(y^{(k)})$ that

$$\lim g(y^{(k)}) = \inf g(y).$$

Because g is a closed convex function and $\bar{y}$ is the only solution of $\inf g$, we have $y^{(k)} \to \bar{y}$. From $x^{(k)} + y^{(k)} = \bar{x} + \bar{y} = d$ we also have $x^{(k)} \to \bar{x}$. □

For the sequence $\{(y_1^{(k)}, \ldots, y_m^{(k)})\}$ to have accumulation points, we need some conditions to ensure its boundedness.

Lemma 4.6. *If* $\operatorname{int} C \neq \emptyset$ *then the sequence* $\{(y_1^{(k)}, \ldots, y_m^{(k)})\}$ *is bounded.*

Proof. Let $z \in \operatorname{int} C$. Then we have

$$\begin{aligned}h(y_1, \ldots, y_m) &= \tfrac{1}{2} \| y_1 + \cdots + y_m - d \|_Q^2 + \textstyle\sum \delta^*(y_i \mid QC_i) \\ &= \tfrac{1}{2} \| y_1 + \cdots + y_m - d + z \|_Q^2 + \textstyle\sum \delta^*(y_i \mid QC_i) - \sum \langle y_i, Qz \rangle \\ &\quad + \langle d - \tfrac{1}{2} z, Qz \rangle \\ &= \tfrac{1}{2} \| y_1 + \cdots + y_m - d + z \|_Q^2 + \textstyle\sum \delta^*(y_i \mid Q(C_i - z)) \\ &\quad + \langle d - \tfrac{1}{2} z, Qz \rangle.\end{aligned}$$

It follows from Lemma 4.2 and the above equality that

$$\begin{aligned}\tfrac{1}{2}\| d-z \|_Q^2 &= h(y_1^{(0)},\ldots,y_m^{(0)})+\tfrac{1}{2}\langle z, Qz\rangle-\langle d, Qz\rangle\\ &\geq h(y_1^{(k)},\ldots,y_m^{(k)})+\tfrac{1}{2}\langle z, Qz\rangle-\langle d, Qz\rangle\\ &\geq \sum \delta^*(y_i^{(k)}\,|\,Q(C_i-z))\end{aligned}$$

Because $0\in Q(C_i-z)$, we have $\delta^*(y_i^{(k)}\,|\,Q(C_i-z))\geq 0$. Hence, it follows that, for each i,

$$y_i^{(k)}\in S_i:=\{y\,|\,\delta^*(y\,|\,Q(C_i-z))\leq\tfrac{1}{2}\| d-z \|_Q^2\}.$$

The sets S_i are bounded because

$$0\in \operatorname{int}(Q(C_i-z))=\operatorname{int}(\operatorname{dom}(\delta(\cdot\,|\,Q(C_i-z))). \qquad \square$$

Theorem 4.7. *If* int $C\neq\emptyset$ *then the sequences* $\{x^{(k)}\}$ *and* $\{y^{(k)}\}$ *converge to the solutions of* (P) *and* (D) *respectively.*

Proof. This follows directly from the previous lemma and Theorem 4.5. $\square$

When some sets C_i are polyhedral, a stronger result can be obtained.

Theorem 4.8. *Let* $C_1,\ldots,C_p$ *be polyhedral and* $C_{p+1},\ldots,C_m$ *be closed convex sets such that* $C_1\cap\cdots C_p\cap\operatorname{int} C_{p+1}\cap\cdots\cap\operatorname{int} C_m\neq\emptyset$, *then the sequences* $\{x^{(k)}\}$ *and* $\{y^{(k)}\}$ *converge to the solutions of* (P) *and* (D), *respectively.*

Proof. This is proven by showing that the sequence $\{y^{(k)}\}$ is bounded and the solution $\bar{y}$ of (D) is its only accumulation point. We have

$$g(y^{(k)})\leq h(y_1^{(k)},\ldots,y_m^{(k)})\leq h(y_1^{(0)},\ldots,y_m^{(0)})=\tfrac{1}{2}\| d \|_Q^2.$$

Therefore, the points $y^{(k)}$ are in the level set $\{y\,|\,g(y)\leq\tfrac{1}{2}\| d\|_Q^2\}$ which is bounded because inf g is attained at one point. By the argument used in Lemma 4.5, we also have the boundedness of the sequence $\{(y_{p+1}^{(k)},\ldots,y_m^{(k)})\}$.

Let $\tilde{y}$ be an accumulation point of $\{y^{(k)}\}$, then we can extract convergent subsequences $\{y^{(k_j)}\}$ and $\{(y_{p+1}^{(k_j)},\ldots,y_m^{(k_j)})\}$ such that $\{y^{(k_j)}\}$ converges to $\tilde{y}$ and $\{(y_{p+1}^{(k_j)},\ldots,y_m^{(k_j)})\}$ converges to a point, $(\tilde{y}_{p+1},\ldots,\tilde{y}_m)$ say. Then, $x^{(k_j)}\to\tilde{x}:=d-\tilde{y}$. It follows from Corollary 4.3 that for each i, $x_i^{(k_j)}\to\tilde{x}$ which, in conjunction with $x_i^{(k_j)}\in C_i$, implies $\tilde{x}\in C$. From Theorem 24.4 of [7] we also have $Q\tilde{y}_i\in\partial\delta(\tilde{x}\,|\,C_i)$ for $i=p+1,\ldots,m$. On the other hand, it follows from the polyhedrality of C_i that, for i from 1 to p and for sufficiently large k_j,

$$\partial\delta(x_i^{(k_j)}\,|\,C_i)\subset\partial\delta(\tilde{x}\,|\,C_i).$$

Hence, for sufficiently large k_j,

$$Q(y^{(k_j)}-y_{p+1}^{(k_j)}-\cdots-y_m^{(k_j)})\in\partial\delta(\tilde{x}\,|\,C_1)+\cdots+\partial\delta(\tilde{x}\,|\,C_p)$$

which implies that

$$Q(\tilde{y}-\tilde{y}_{p+1}-\cdots-\tilde{y}_m)\in\partial\delta(\tilde{x}\,|\,C_1)+\cdots+\partial\delta(\tilde{x}\,|\,C_p)$$

and thus

$$Q\tilde{y}\in\sum\partial\delta(\tilde{x}\,|\,C_i)\subset\partial\delta(\tilde{x}\,|\,C).$$

Therefore, $\tilde{y}$ solves (D), which implies $\tilde{y}=\bar{y}$ because $\bar{y}$ is the only solution of (D). Therefore, $\bar{y}$ is the only accumulation point of the bounded sequence $\{y^{(k)}\}$ and, consequently, $\{y^{(k)}\}$ converges to $\bar{y}$. It also follows that $x^{(k)}\to\bar{x}$ because $x^{(k)}+y^{(k)}=\bar{x}+\bar{y}=d$. $\square$

Even though it is satisfactory from a theoretical point of view to establish the convergence of the sequence $\{x^{(k)}\}$; but the convergence of the sequence $\{(y_1^{(k)},\ldots,y_m^{(k)})\}$ is more important to practical considerations because those vectors will be really computed in practice. But the sequence does not necessarily converge even when the sequences $\{x^{(k)}\}$ and $\{y^{(k)}\}$ converge well to the solutions of (P) and (D) respectively. This can be seen from a simple example. As in the following picture, let C_1 and C_2 be two disks of the same size with one point $\bar{x}$ in common. Let the point d be a point on the line which is tangent to both sets at $\bar{x}$. It is easy to see that $(y_1^{(k)},y_2^{(k)})\to\infty$ but $y_1^{(k)}+y_2^{(k)}\to\bar{y}$, which is $d-\bar{x}$.

For such a case the method has little practical use. Therefore, we are interested in the situations when the convergence of the sequence $\{(y_1^{(k)},\ldots,y_m^{(k)})\}$ is ensured. We consider such conditions in the following lemma and theorem.

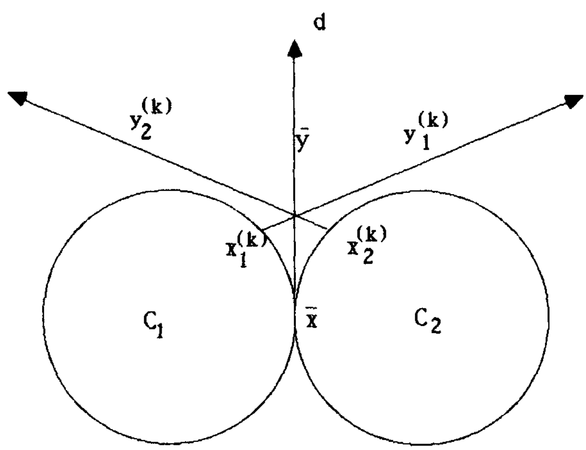

Fig. 1

Lemma 4.9. *If* (D′) *has an unique solution then the sequence* $\{(y_1^{(k)},\ldots,y_m^{(k)})\}$ *converges to the solution of* (D′).

Proof. By the assumption, the level sets of h are bounded and, hence, the sequence $\{(y_1^{(k)},\ldots,y_m^{(k)})\}$ is bounded. On the other hand, it follows from Proposition 4.4

that any accumulation point of the sequence solves (D′). Therefore, the lemma follows directly from the uniqueness of the solution. □

Theorem 4.10. *Let $C_1, \ldots, C_p$ be polyhedral and $C_{p+1}, \ldots, C_m$ be closed convex sets such that $C_1 \cap \cdots \cap C_p \cap \operatorname{ri} C_{p+1} \cap \cdots \cap \operatorname{ri} C_m \neq \emptyset$. If the subspaces $\operatorname{aff}(\partial\delta(\bar{x} \mid C_i))$ are linearly independent in the sense that*

$$\sum z_i = 0 \ \text{ and } \ z_i \in \operatorname{aff}(\partial\delta(\bar{x} \mid C_i)) \ \Rightarrow \ z_1 = \cdots = z_m = 0,$$

then the sequences $\{x^{(k)}\}$, $\{y^{(k)}\}$ and $\{(y_1^{(k)}, \ldots, y_m^{(k)})\}$ converge to the solutions of (P), (D) *and* (D′) *respectively.*

Proof. By Theorem 3.3 problem (D′) has solutions. If $(\tilde{y}_1, \ldots, \tilde{y}_m)$ and $(\bar{y}_1, \ldots, \bar{y}_m)$ are two solutions of (D′) then, by Theorem 3.3, $\bar{y} = \tilde{y}_1 + \cdots + \tilde{y}_m = \bar{y}_1 + \cdots + \bar{y}_m$ is the only solution of (D). Therefore, with $z_i := \tilde{y}_i - \bar{y}_i$, we have $\sum z_i = 0$ and $z_i \in \operatorname{aff}(\partial\delta(\bar{x} \mid C_i))$. It follows from the linear independence condition that $(\tilde{y}_1, \ldots, \tilde{y}_m) = (\bar{y}_1, \ldots, \bar{y}_m)$. Then the theorem follows immediately from Lemma 4.9 and Theorem 3.3. □

5. Applications

In this section we study some applications of the successive projection method. We note that the efficiency of the method depends very much on how much work is involved in carrying out each projection. So the method can be very simple when the sets are polyhedral or even ellipsoidal. We first consider the definite quadratic programming case.

(1) *Definite quadratic program*

If we want to solve the following definite quadratic progamming problem:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\langle x - d, Q(x - d)\rangle \\ \text{s.t.} \quad & \mu_i \leq \langle a_i, x\rangle \leq \alpha_i, \quad i = 1, \ldots, m, \end{aligned}$$

where $-\infty \leq \mu_i \leq \alpha_i \leq \infty$, then we may let $C_i = \{x \mid \mu_i \leq \langle a_i, x\rangle \leq \alpha_i\}$. In this case, to carry out a projection to the set C_i and compute $x_i^{(k)}$ and $y_i^{(k)}$, we need to solve the following minimization problem with a simple bound constraint:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\langle x - z, Q(x - z)\rangle \\ \text{s.t.} \quad & \mu_i \leq \langle a_i, x\rangle \leq \alpha_i \end{aligned}$$

where $z = x_{i-1}^{(k)} + y_i^{(k-1)}$. This subproblem can be easily solved as follows. If $\mu_i \leq \langle a_i, z\rangle \leq \alpha_i$ then $x_i^{(k)} = z$ and $y_i^{(k)} = 0$. When $\langle a_i, z\rangle > \alpha_i$ then by the Karush–Kuhn–Tucker condition there exists a Lagrange multiplier $\lambda_i^{(k)}$ such that

(a) $Q(x_i^{(k)} - z) + \lambda_i^{(k)} a_i = 0$,
(b) $\langle a_i, x_i^{(k)}\rangle = \alpha_i$.

It follows from $y_i^{(k)}=z-x_i^{(k)}$ that $y_i^{(k)}=\lambda_i^{(k)}Q^{-1}a_i$; therefore, using $y_i^{(k-1)}=\lambda_i^{(k-1)}Q^{-1}a_i$ and (b) above, we have that

$$\begin{aligned}\langle a_i, Q^{-1}a_i\rangle\lambda_i^{(k)} &= \langle a_i, z\rangle-\alpha_i\\ &= \langle a_i, Q^{-1}a_i\rangle\lambda_i^{(k-1)}+\langle a_i, x_{i-1}^{(k)}\rangle-\alpha_i.\end{aligned}$$

Let $\hat{a}_i=\langle a_i, Q^{-1}a_i\rangle^{-1}a_i$ and $\beta_i=\langle a_i, Q^{-1}a_i\rangle^{-1}\alpha_i$; then we have that

$$\lambda_i^{(k)}=\lambda_i^{(k-1)}+\langle\hat{a}_i, x_{i-1}^{(k)}\rangle-\beta_i.$$

For the case that $\langle a_i, z\rangle<\mu_i$, we can similarly get

$$\lambda_i^{(k)}=\lambda_i^{(k-1)}+\langle\hat{a}_i, x_{i-1}^{(k)}\rangle-\gamma_i,$$

where $\gamma_i=\langle a_i, Q^{-1}a_i\rangle^{-1}\mu_i$. Therefore, we can restate the successive projection method for solving the above definite quadratic programming problem as follows:

For $i=1,\ldots,m$, compute $b_i=Q^{-1}a_i$, $\hat{a}_i=\langle a_i, b_i\rangle^{-1}a_i$, $\beta_i=\langle a_i, b_i\rangle^{-1}\alpha_i$ and $\gamma_i=\langle a_i, b_i\rangle^{-1}\mu_i$. Initially, set $\lambda_1^{(0)}=\cdots=\lambda_m^{(0)}=0$ and $x_m^{(0)}=d$. At the k-th iteration, set $x_0^{(k)}=x_m^{(k-1)}$ and for $i=1,\ldots,m$, compute

$$\lambda_i^{(k)}=\begin{cases}\lambda_i^{(k-1)}+\langle\hat{a}_i, x_{i-1}^{(k)}\rangle-\beta_i & \text{if } \lambda_i^{(k-1)}+\langle\hat{a}_i, x_{i-1}^{(k)}\rangle>\beta_i,\\ \lambda_i^{(k-1)}+\langle\hat{a}_i, x_{i-1}^{(k)}\rangle-\gamma_i & \text{if } \lambda_i^{(k-1)}+\langle\hat{a}_i, x_{i-1}^{(k)}\rangle>\gamma_i,\\ 0 & \text{otherwise,}\end{cases}$$

$$x_i^{(k)}=x_{i-1}^{(k)}+(\lambda_i^{(k-1)}-\lambda_i^{(k)})b_i.$$

The vectors b_i and $\hat{a}_i$ are computed only once at the beginning and the computation involved in each iteration is extremely simple. It is even more so when Q is diagonal and the constraint vectors a_i are sparse. It is noted that the method in this form is closely related to the SOR method [1, 2, 4].

There is a great flexibility in choosing the sets C_i. The performance of the method is certainly not the same when the sets are chosen differently. For instance, the set C_i may consist of many constraints in case that the corresponding subproblem can be effectively solved.

(2) *Problem with ellipsoidal constraints*

We now consider the problem with ellipsoidal constraints:

$$\begin{aligned}&\min\quad \tfrac{1}{2}\langle x-d, Q(x-d)\rangle\\ &\text{s.t.}\quad \tfrac{1}{2}\langle x-a_i, A_i(x-a_i)\rangle\leq\alpha_i,\qquad i=1,\ldots,m\end{aligned}$$

where A_i are positive semi-definite symmetric matrices. Let $C_i=\{x\,|\,\tfrac{1}{2}\langle x-a_i, A_i(x-a_i)\rangle\leq\alpha_i\}$. Here, to find $x_i^{(k)}$ and $y_i^{(k)}$ we solve the following problem with one constraint:

$$\begin{aligned}&\min\quad \tfrac{1}{2}\langle x-z, Q(x-z)\rangle\\ &\text{s.t.}\quad \tfrac{1}{2}\langle x-a_i, A_i(x-a_i)\rangle\leq\alpha_i\end{aligned}$$

where, as before, $z = x_{i-1}^{(k)} + y_i^{(k-1)}$. If z is in C_i then clearly $x_i^{(k)} = z$ and $y_i^{(k)} = 0$; otherwise, the constraint must be active and we solve the following system of equations for $x_i^{(k)}$ and $\lambda_i^{(k)}$

$$Q(x-z) + \lambda A_i(x - a_i) = 0,$$

$$\tfrac{1}{2}\langle x - a_i, A_i(x - a_i)\rangle = \alpha_i.$$

With x eliminated from the second equation, the number $\lambda_i^{(k)}$ becomes a root of the nonlinear function

$$g_i^{(k)}(\lambda) := \tfrac{1}{2}\langle z - a_i, Q(Q + \lambda A_i)^{-1} A_i (Q + \lambda A_i)^{-1} Q(z - a_i)\rangle - \alpha_i.$$

We can use a Newton-like method to sove $g_i^{(k)}(\lambda) = 0$ for $\lambda_i^{(k)}$ and then calculate $x_i^{(k)}$ by solving the linear system

$$(Q + \lambda_i^{(k)} A_i)x = Qz + \lambda_i^{(k)} A_i a_i.$$

The vector $y_i^{(k)}$ can be calculated by $y_i^{(k)} = \lambda_i^{(k)} Q^{-1} A_i (x_i^{(k)} - a_i)$.

We note that a similar technique for minimizing a definite quadratic function with one ellipsoidal constraint was used with success by Moré [5] in his version of a trust-region method for nonlinear least-squares problems, where an ellipsoidal constraint is used to restrict the size of a search direction.

(3) *Parallel computation*

Another interesting application of the successive projection method is in the area of parallel computation. Consider the problem

$$\begin{aligned} \min \quad & \tfrac{1}{2}\langle x - d, Q(x - d)\rangle \\ \text{s.t.} \quad & g_i(x) \leq 0, \qquad i = 1, \ldots, m, \end{aligned}$$

with Q diagonal and each g_i containing very few variables. In this case, by choosing the sets C_i properly, we can use the successive projection method to solve several independent small problems simultaneously. Specifically, we let $S_1, \ldots, S_r$ be a partition of the index set $\{1, \ldots, m\}$ such that for any i and for any two different constraints g_j and g_k with j and k in S_i contain no common variables. Then, we can let $C_i := \{x \mid g_j(x) \leq 0, j \in S_i\}$ and, in the successive projection method, each subproblem

$$\begin{aligned} \min \quad & \tfrac{1}{2}\langle x - z, Q(x - z)\rangle \\ \text{s.t.} \quad & x \in C_i \end{aligned}$$

reduces to several independent smaller problems with only one constraint each because of the separability of the constraints g_j in S_i. Therefore, these independent small problems can be solved parallelly. To illustrate this point, we given the following example:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\langle x-d, x-d\rangle \\ \text{s.t.} \quad & g_1(x_1, x_2, x_3) \leqslant 0, \\ & g_2(x_2, x_3, x_4) \leqslant 0, \\ & \vdots \\ & g_{n-2}(x_{n-2}, x_{n-1}, x_n) \leqslant 0. \end{aligned}$$

where each constraint contains only three variables. For simplicity, we assume $n=3m+2$. Then, we partition $\{1,\ldots,m\}$ into $S_1:=\{1,4,\ldots,3m\}$, $S_2:=\{2,5,\ldots,3m+1\}$ and $S_3:=\{3,6,\ldots,3m+2)$. For $i=1,2,3$, we set $C_i:=\{x \mid g_j(x)\leqslant 0;\ j\in S_i\}$. Then a projection to the set C_1, for example, reduces to the following m small problems:

$$\begin{aligned} \min \quad & (x_1-z_1)^2+(x_2-z_2)^2+(x_3-z_3)^2 \\ \text{s.t.} \quad & g_1(x_1, x_2, x_3) \leqslant 0, \\ \min \quad & (x_4-z_4)^2+(x_5-z_5)^2+(x_6-z_6)^2 \\ \text{s.t.} \quad & g_4(x_4, x_5, x_6) \leqslant 0, \\ & \vdots \\ \min \quad & (x_{3m}-z_{3m})^2+(x_{3m+1}-z_{3m+1})^2+(x_{3m+2}-z_{3m+2})^2 \\ \text{s.t.} \quad & g_{3m}(x_{3m}, x_{3m+1}, x_{3m+2}) \leqslant 0. \end{aligned}$$

The above m problems have three variables and one constraint each and, more importantly, they are independent of each other and can be solved simultaneously. We note here that a modification of the method for handling a general problem without any special sparse structure has been recently developed and is given in [3].

Acknowledgements

The paper was presented at the 1985 Optimization Symposium at Cambridge, England. The author would like to express his thanks to Prof. M.J.D. Powell for his valuable suggestions on a draft of the paper and to Dr. J.J. Moré for his constructive comments concerning an application of the method.

References

[1] C.W. Cryer, "The solution of a quadratic programming problem using systematic overrelaxation," *SIAM Journal on Control* 22 (1971) 385–392.

[2] U. Eckhardt, "Quadratic programming by successive overrelaxation," Kernforschungsanlage Jülich, Technical Report No. Jül-1064-MA, 1974.

[3] S.-P. Han and G. Lou, "A parallel algorithm for a class of convex programs," to appear in *SIAM Journal on Control and Optimization* (Oct., 1986).

[4] O.L. Mangasarian, "Solution of symmetric linear complementarity problem by iterative methods," *Journal of Optimization Theory and its Applications* 22 (1977) 465–485.

[5] J.J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," in: G.A. Watson, ed., *Proceedings of the 1978 Dundee Conference on Numerical Analysis*, Lecture Notes in Mathematics 630 (Springer-Verlag, Berlin, 1978) pp. 105–116.

[6] J.-J. Moreau, "Proximité et sous-gradient d'une fonctionelle," *Bulletin de la Societe Mathematique de France* 93 (1965) 273–299.

[7] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).