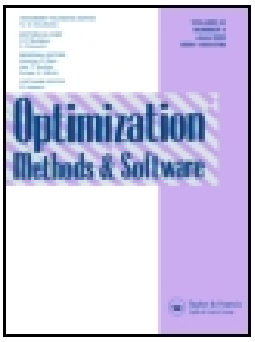

Taylor & Francis
Taylor & Francis Group

Optimization Methods and Software

ISSN: 1055-6788 (Print) 1029-4937 (Online) Journal homepage: http://www.tandfonline.com/loi/goms20

Convergence rate analysis and error bounds for projection algorithms in convex feasibility problems

Amir Beck & Marc Teboulle

To cite this article: Amir Beck & Marc Teboulle (2003) Convergence rate analysis and error bounds for projection algorithms in convex feasibility problems, Optimization Methods and Software, 18:4, 377-394, DOI: 10.1080/10556780310001604977

To link to this article: http://dx.doi.org/10.1080/10556780310001604977

Published online: 19 Feb 2010.

Optimization Methods and Software
Vol. 18, No. 4, August 2003, pp. 377–394

Taylor & Francis
Taylor & Francis Group

CONVERGENCE RATE ANALYSIS AND ERROR BOUNDS FOR PROJECTION ALGORITHMS IN CONVEX FEASIBILITY PROBLEMS

AMIR BECK and MARC TEBOULLE*

School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv 69978, Israel

(Received 31 October 2002; Revised 10 March 2003; In final form 28 April 2003)

We analyze the rate of convergence of three basic projections type algorithms for solving the convex feasibility problem (CFP). Error bounds are known to be central in establishing the rate of convergence of iterative methods. We study the interplay between Slater's hypothesis on CFP and a specific local error bound (LEB). We show that without Slater's hypothesis on CFP, projections type algorithms can in fact behave quite badly, i.e., with a rate of convergence which is not bounded. We derive a new and simple convex analytic proof showing that Slater's hypothesis on CFP implies LEB and hence linear convergence of projection algorithms is guaranteed. We then propose an alternative local error bound derived from the gradient projection algorithm for convex minimization which is proven to be weaker than LEB and used to derive further convergence rate results.

Keywords: Convex feasibility; Projection algorithms; Error bounds; Convergence rate; Convex minimization

1 INTRODUCTION

Given m closed convex sets $C_1, C_2, \ldots, C_m$ of $\Re^d$ such that $C \equiv \cap_{i=1}^m C_i \neq \emptyset$, the Convex Feasibility Problem (CFP) consists of finding a point in the intersection C. The convex feasibility problem arises in a wide variety of contexts and applications such as best approximation theory, image reconstruction (both discrete and continuous models) and subgradient methods. For a detailed survey on the CFP, see e.g., [3] and references therein. The algorithms discussed in this paper are projection algorithms. For any set $S \subseteq \Re^d$, let P_S denote the projection operator. At each iteration, a basic projection algorithm generates a point which is a convex combination of the projections of the previous point on the convex sets. We will consider three kinds of projection algorithms:

Mean Projection Algorithm (MPA):
First step: Take an arbitrary $x^0 \in \Re^d$
General step: $x_{n+1} = \sum_{i=1}^m \alpha_i P_{C_i}(x_n)$.

Here $\alpha_1, \ldots, \alpha_m$ are positive constants such that $\sum_{i=1}^m \alpha_i = 1$.

* Corresponding author. E-mail: teboulle@post.tau.ac.il

ISSN 1055-6788 print; ISSN 1029-4937 online
DOI: 10.1080/10556780310001604977

Cyclic Projection Algorithm (CPA):
First step: Take an arbitrary $x^0 \in \Re^d$
General step: $x_{n+1} = P_{C_{(n \bmod m)+1}}(x_n)$.

Maximum Distance Projection Algorithm (MDPA):
First step: Take an arbitrary $x^0 \in \Re^d$
General step: $x_{n+1} = P_{C_j}(x_n)$ where $j = \operatorname{argmax}_{1 \le i \le m} d(x_n, C_i)$.

Note that in MPA, there is no real reason to limit the discussion to algorithms which use the same convex combination of the projections of x_n to each of the m convex sets $C_1, C_2, \ldots, C_m$. Indeed, we can define the general step by:

$$x_{n+1} = \sum_{i=1}^{m} \alpha_i^n P_{C_i}(x_n)$$

where $\sum_{i=1}^m \alpha_i^n = 1$ provided that $\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n > 0$ are assumed bounded away from zero, i.e., there exists numbers $\beta_1, \ldots, \beta_m > 0$ such that $\alpha_i^n \ge \beta_i \ \ \forall_i = 1, \ldots, m, n = 0, 1, 2, \ldots$.

The cyclic projection algorithm goes back to von Neumann [15] who considered the case of two subspaces and the mean projection algorithm with equal weights (i.e. $\alpha_i = 1/m$) was proposed by Cimmino [6] who considered the case where each C_i is a halfspace. In this paper we investigate only these three schemes (MPA, MDPA, CPA). Many variant of these schemes can be found in the literature, see for example Ref. [3] and references therein.

The first natural question concerning these methods is to establish their convergence. Proofs of global convergence of projection methods can be found in the classical work of Ref. [7] and in the book of Auslender [1]. These results will be briefly reviewed in the next section. The second main question is to analyze the rate of convergence of these algorithms, and that will be the main purpose of this paper.

Error bounds are known to play a central role in the rate of convergence analysis of iterative algorithms. An error bound is a quantity, usually called a residual function, that becomes zero whenever a point is a solution of CFP. For an excellent survey on the theory and applications of error bounds, we refer the readers to the work of Pang [12] and the references therein. In this paper we investigate the residual function

$$T(x) = \max_{1 \le i \le m} d(x, C_i).$$

Indeed, in this case one obviously has $T(x) = 0$ if and only if $x \in C$. In the context of CFP, the residual T was first proposed by Gubin *et al.* [7] and leads to the following kinds of error bounds: Global Error Bound (GEB) and Local Error Bound (LEB).

DEFINITION 1.1 (GEB) *m closed convex sets $C_1, \ldots, C_m$ are said to satisfy GEB if there exists $\theta > 0$ such that*:

$$\forall x \in \Re^d \quad d(x, C) \le \theta \max_{i=1,\ldots,m} \{d(x, C_i)\}.$$

DEFINITION 1.2 (LEB) *m closed convex sets $C_1, \ldots, C_m$ are said to satisfy LEB if for every bounded set B there exists $\theta_B > 0$ such that*:

$$\forall x \in B \quad d(x, C) \le \theta_B \max_{i=1,\ldots,m} \{d(x, C_i)\}.$$

Notice that in both definitions, while $d(x, C)$ is usually impossible to estimate (otherwise the original problem is trivial) the error bound $T(x)$ is in many cases easily calculated. Thus, GEB and LEB state that we can bound an unknown quantity by a computable quantity. The advantage of GEB over LEB relies on the fact that, when the former is satisfied, the rate of convergence of projection algorithms does not depend on the initial starting point of the algorithm, as opposed to LEB. However, GEB is satisfied only in rare cases. One of the cases that satisfies GEB is when all the C_i are polyhedral sets. This is the celebrated Hoffmann's Lemma [8]. For the non-polyhedral case, GEB is usually not satisfied and thus an important question that arises is when does LEB holds? Gubin *et al.* [7] were apparently the first to prove that if there exists a $1 \leq j \leq m$ such that $C_j \cap \operatorname{int}(\cap_{i \neq j} C_i) \neq \emptyset$ then LEB is satisfied, and as a consequence the linear rate of convergence was obtained for the sequences generated by CPA and MDPA. In a recent paper, Bauschke *et al.* [4] refined and extended this result by proving that the standard Slater's condition for CFP implies LEB (LEB is called there "bounded linear regularity"). However, to achieve this goal, a rather involved and complicated analysis was needed in Ref. [4].

The aim of this paper is to further analyze the rate of convergence of projections algorithms for CFP and in particular to clarify the important roles play on one hand by Slater's condition, and on the other, by the local error bound (LEB). After a brief review on known convergence results of projection algorithms and other preliminaries given in Section 2, our contributions can be summarized as follows:

- We give a new and simple proof, which relies on elementary convex analysis arguments, to show that Slater's condition implies LEB (and thus implies linear rate of convergence), see Section 3.
- In Section 4, we show that projection algorithms for CFP can be very slow if Slater's condition is not satisfied. Moreover, without Slater's condition it is shown in fact that we cannot bound the rate of convergence. More precisely, we exhibit an example where the sequence $\{x_n\}$ generated by CPA satisfies $d(x_n, C) \geq 1/n^{1/1000}$, which obviously implies a *very* slow rate of convergence.
- Projection algorithms for convex feasibility problems with two sets are revisited through the use of the gradient projection method applied to a convex minimization reformulation of CFP, see Section 5. This allows us to derive another local error bound that is proven to be weaker than LEB, and then used to derive further rate of convergence results.

2 CONVERGENCE OF PROJECTION ALGORITHMS FOR THE CONVEX FEASIBILITY PROBLEM

In this section we briefly recall the basic properties and convergence results of projection algorithms for solving CFP that will be needed in the rest of this paper. At this point, we would like to recall two concepts of linear rate of convergence, see e.g., Ref. [11]. A sequence $\{x_n\}$ converges with a *R-linear rate* to x^* if there exists $\gamma \in (0, 1)$, $A > 0$ such that $\|x_n - x^*\| \leq A\gamma^n$ for every n large enough. A sequence $\{x_n\}$ converges with a *Q-linear rate* to x^* is there exists $\gamma \in (0, 1)$ such that $\|x_{n+1} - x^*\| \leq \gamma\|x_n - x^*\|$ for every n large enough. Clearly, Q-linear convergence implies R-linear convergence. In the rest of this paper, we use the terminology "linear convergence" to refer to "R-linear convergence".

Let $\{x_n\}$ be a sequence generated by any one of the algorithms CPA, MDPA or MPA. The proof of convergence of these algorithms relies on the following three basic facts and relations which can be found in Auslender [1, p. 78]. Note that these relations are independent from the type of algorithms used to generate the sequence $\{x_n\}$ and are based on well known properties of the projection operator P_S, see e.g., [1].

$$\forall j = 1, \ldots, m \quad d^2(x_n, C_j) \le d^2(x_n, C) - \|P_{C_j}(x_n) - P_C(x_n)\|^2. \tag{1}$$

$$\sum_{j=1}^{m} \alpha_j d^2(x_n, C_j) \le d^2(x_n, C) - d^2(x_{n+1}, C). \tag{2}$$

$$\|x_{n+1} - y\| \le \|x_n - y\| \quad \forall y \in C \tag{3}$$

The later property is often referred as the Fejér monotonicity of the sequence with respect to C. Assuming that the set $C = \cap_{i=1}^{m} C_i \neq \emptyset$ we then have the following convergence result.

THEOREM 2.1 (Global Convergence of Projection Algorithms) [1,7] *Let $\{x_n\}$ be a sequence generated by any one of the algorithms MPA, MDPA or CPA. Then there is a point $c \in C$ such that*

$$x_n \to c.$$

Convergence of the algorithms is proved under the mild assumption that $C \neq \emptyset$. In order to derive their rate of convergence, and here, more precisely to prove linear convergence, the required additional assumption is precisely LEB (see Definition 1.2) introduced by Gubin *et al.* [7] who also proved the following results. For completeness, the proof is given in the Appendix.

THEOREM 2.2 (LEB implies linear rate of convergence of MPA and MDPA) *Assume that LEB is satisfied. Then the distances of sequence generated by MPA (MDPA) from C converge with a Q-linear rate to* 0. *More specifically,*

$$d(x_{n+1}, C) \le \gamma_B d(x_n, C), \tag{4}$$

with

$$\gamma_B = \sqrt{1 - \frac{\min_{j=1,\ldots,m}\{\alpha_j\}}{\theta_B^2}}, \quad \left(\gamma_B = \sqrt{1 - \frac{1}{\theta_B^2}}\right) \tag{5}$$

where $B = \{x\colon \|x - y\| \le \|x_0 - y\|\}$ and y is an arbitrary point in C.

Note that the above rate of convergence depends on the initial point x_0 (γ_B is dependent on x_0). This dependency can be removed if we assume instead of LEB the stronger global error bound GEB.

2.1 Linear Convergence of CPA for Two Sets

For two sets, it is easy to prove using the previous results that the distances of the sequence generated by CPA from C converge 0 with a Q-linear rate of convergence. However for the general problem of m sets we cannot prove a result like (4) for the simple reason that it is not true. Indeed, take for example the case where we have three closed convex sets C_1, C_2, C_3 such that $C_1 = C_2$ then obviously $x_{n+1} = x_n$ every three times thus Eq. (4) cannot hold for $\gamma_B < 1$.

COROLLARY 2.1 (LEB implies linear rate of convergence of CPA for two sets) *If LEB is satisfied then the distances of the sequence generated by CPA from C converges with a Q-linear rate to* 0. *More specifically*,

$$d(x_{n+1}, C) \leq \gamma_B d(x_n, C),$$

with

$$\gamma_B = \sqrt{1 - \frac{1}{\theta_B^2}}, \quad B = \{x \colon \|x - y\| \leq \|x_0 - y\|\}$$

where y is an arbitrary point in C.

Proof CPA for two sets is the same as MDPA for two sets and thus the result follows. ■

Until now, we have seen that under the assumption that $\cap_{i=1}^m C_i \neq \emptyset$ there is a $x^* \in C$ such that $x_n \to x^*$. Moreover, under the LEB assumption we have $d(x_n, C) \to 0$ with a linear rate of convergence. As an easy consequence of these results we can prove that $x_n \to x^*$ with a Q-linear rate of convergence.

THEOREM 2.3 *Let* $\{x_n\}$ *be a sequence generated by MPA or by MDPA and assume LEB holds. Then, there is a* $x^* \in C$ *such that*

$$\|x_n - x^*\| \leq D\gamma_B^n,$$

where $D = d(x_0, C)/(1 - \gamma_B) > 0$ *and* γ_B *is defined by* (5).

Proof First, let us consider MPA. We have already proved that there is a $x^* \in C$ such that $x_n \to x^*$ and that $d(x_{n+1}, C) \leq \gamma_B d(x_n, C)$ for every $n \geq 0$. Now,

$$\begin{aligned}
\|x_{n+1} - x_n\| &= \left\| \sum_{i=1}^m \alpha_i P_{C_i}(x_n) - x_n \right\| = \left\| \sum_{i=1}^m \alpha_i (P_{C_i}(x_n) - x_n) \right\| \\
&\leq \sum_{i=1}^m \alpha_i \|P_{C_i}(x_n) - x_n\| = \sum_{i=1}^m \alpha_i d(x_n, C_i) \\
&\overset{C \subseteq C_i}{\leq} \sum_{i=1}^m \alpha_i d(x_n, C) = d(x_n, C) \leq t\gamma_B^n
\end{aligned}$$

where $t = d(x_0, C)$. Thus, for every $N > n$:

$$\|x_N - x_n\| \leq \sum_{j=n}^{N-1} \|x_{j+1} - x_j\| \leq \sum_{j=n}^{N-1} t\gamma_B^j = t\gamma_B^n \left(\frac{1 - \gamma_B^{N-n}}{1 - \gamma_B} \right).$$

Taking $N \to \infty$ we have:

$$\|x_n - x^*\| \leq \frac{t}{1 - \gamma_B} \gamma_B^n.$$

Substitute $D = t/(1 - \gamma_B)$ and obtain the result. The proof for MDPA is essentially the same, the only difference is that we do not use a constant convex combination like in the MPA case, but instead we use a different convex combination at each iteration. ■

3 SLATER'S CONDITION IMPLIES LEB

We begin by recalling Slater's condition for the convex feasibility problem (CFP).

Slater's Condition: Let $C_1, \ldots, C_m$ be m closed convex sets. Suppose that $C_1, \ldots, C_k (k \leq m)$ are polyhedral sets. Then, $C_1, \ldots, C_m$ are said to satisfy the *Slater's condition* if:

$$\left(\bigcap_{i=1}^{k} C_i\right) \bigcap \left(\bigcap_{i=k+1}^{m} \operatorname{ri}(C_i)\right) \neq \emptyset$$

The aim of this section is to prove that Slater's condition implies LEB. This result was recently derived in Ref. [4] through a quite long and rather complex analysis which appears to be unnecessary. Our objective is to give a new and simple proof which relies on elementary convex analysis arguments and Hoffmann's Lemma. For that purpose, we begin with the following key result from Gubin *et al.* [7], which seems to have been overlooked in the literature. For completeness we include a proof.

LEMMA 3.1 *Let $C_1, C_2 \subseteq \Re^d$ be two closed convex sets such that $C_1 \cap \operatorname{int}(C_2) \neq \emptyset$. Then LEB is satisfied i.e. for every bounded set B there is a $\theta_B > 0$ such that*

$$\forall x \in B \quad d(x, C_1 \bigcap C_2) \leq \theta_B \max\{d(x, C_1), d(x, C_2)\}.$$

Proof Let $x \in \Re^d$, denote $\eta = 2\max\{d(x, C_1), d(x, C_2)\}$. For every $z \in C_1 \cap C_2$ we have:

$$\begin{aligned} d(x, C_1 \bigcap C_2) &\leq \|x - z\| \\ &\leq \|x - P_{C_1}(x)\| + \|P_{C_1}(x) - z\| \\ &= d(x, C_1) + \|P_{C_1}(x) - z\| \\ &\leq \frac{\eta}{2} + \|P_{C_1}(x) - z\|. \end{aligned} \tag{6}$$

$d(\cdot, C_2)$ is Lipschitz with constant 1 and thus,

$$d(y, C_2) \leq \|y - x\| + d(x, C_2) \quad \forall x, y \in \Re^d.$$

Set $y = P_{C_1}(x)$ and obtain:

$$\begin{aligned} d(P_{C_1}(x), C_2) &\leq \|P_{C_1}(x) - x\| + d(x, C_2) \\ &= d(x, C_1) + d(x, C_2) \\ &\leq \eta \end{aligned} \tag{7}$$

Let $u \in C_1 \cap \operatorname{int}(C_2)$. $u \in \operatorname{int}(C_2)$ and thus there is $\varepsilon > 0$ such that:

$$\|u - v\| \leq \varepsilon \Rightarrow v \in C_2.$$

Let $v = u + \mu(P_{C_1}(x) - P_{C_2}(P_{C_1}(x)))$. Then

$$\|v - u\| = \mu\|P_{C_1}(x) - P_{C_2}(P_{C_1}(x))\| = \mu d(P_{C_1}(x), C_2) \overset{(7)}{\leq} \mu\eta.$$

So pick $\mu = \varepsilon/\eta$ and therefore $\|v - u\| \leq \varepsilon \Rightarrow v \in C_2$. Now, construct a specific z in $C_1 \cap C_2$:

$$z = \frac{1}{\mu+1}\underbrace{v}_{\in C_2} + \frac{\mu}{\mu+1}\underbrace{P_{C_2}(P_{C_1}(x))}_{\in C_2} \overset{C_2 \text{ is convex}}{\Longrightarrow} z \in C_2.$$

Using the definition of v we also have:

$$\begin{aligned} z &= \frac{1}{\mu+1}v + \frac{\mu}{\mu+1}P_{C_2}(P_{C_1}(x)) \\ &= \frac{1}{\mu+1}(u + \mu(P_{C_1}(x) - P_{C_2}(P_{C_1}(x)))) + \frac{\mu}{\mu+1}P_{C_2}(P_{C_1}(x)) \\ &= \frac{1}{\mu+1}\underbrace{u}_{\in C_1} + \frac{\mu}{\mu+1}\underbrace{P_{C_1}(x)}_{\in C_1} \end{aligned}$$

Thus $z \in C_1$ (and as a conclusion $z \in C_1 \cap C_2$). So now

$$\begin{aligned} \|z - P_{C_1}(x)\| &= \left\| \overbrace{\frac{1}{\mu+1}u + \frac{\mu}{\mu+1}P_{C_1}(x)}^{z} - P_{C_1}(x) \right\| \\ &= \frac{1}{\mu+1}\|u - P_{C_1}(x)\| \leq \frac{1}{\mu}\|u - P_{C_1}(x)\| \\ &= \frac{1}{\mu}\|P_{C_1}(u) - P_{C_1}(x)\| \leq \frac{1}{\mu}\|u - x\| = \frac{\eta}{\varepsilon}\|u - x\|. \end{aligned}$$

Using (6) we have:

$$d\left(x, C_1 \bigcap C_2\right) \leq \frac{\eta}{2} + \frac{\eta}{\varepsilon}\|u - x\| \quad \forall x \in \Re^d, \quad u \in C_1 \bigcap C_2.$$

Assuming $x \in B$ we have from the boundedness of B that there is $M > 0$ such that $\|x\| \leq M$ and thus:

$$\begin{aligned} d\left(x, C_1 \bigcap C_2\right) &\leq \frac{\eta}{2} + \frac{\eta}{\varepsilon}\|u - x\| \\ &\leq \frac{\eta}{2} + \frac{\eta}{\varepsilon}(\|u\| + \|x\|) \\ &\leq \frac{\eta}{2} + \frac{\eta}{\varepsilon}\overbrace{M'}^{M+\|u\|} \\ &= 2\left(\frac{1}{2} + \frac{M'}{\varepsilon}\right)\max\{d(x, C_1), d(x, C_2)\} \\ &= \theta_B \max\{d(x, C_1), d(x, C_2)\}. \end{aligned}$$

where $\theta_B = 1 + 2M'/\varepsilon > 0$. ■

COROLLARY 3.1 *Let $D_1, \ldots, D_m \subseteq \Re^d$ be m closed convex sets. If $\cap_{i=1}^m \mathrm{int}(D_i) \neq \emptyset$ then LEB is satisfied i.e. for every bounded set B there is $\theta_B > 0$ such that*:

$$\forall x \in B \quad d\left(x, \bigcap_{i=1}^m D_i\right) \leq \theta_B \max_{i=1,\ldots,m} \{d(x, D_i)\}.$$

Proof Define:

$$C_1 = \left\{ \underbrace{(x, x, \ldots, x)}_{m \text{ times}} : x \in \Re^d \right\}$$

$$C_2 = D_1 \times D_2 \times \cdots \times D_m$$

Now,

$$C_1 \bigcap \mathrm{int}(C_2) = \left\{ (x, x, \ldots, x) \colon x \in \bigcap_{i=1}^m \mathrm{int}(D_i) \right\}.$$

By the assumption that $\bigcap_{i=1}^m \mathrm{int}(D_i) \neq \emptyset$ we have that $C_1 \cap \mathrm{int}(C_2) \neq \emptyset$. Thus, by Lemma 3.1 there is $\theta_B > 0$ such that:

$$\forall y \in B^m \bigcap (\Re^d)^m \quad d(y, C_1 \bigcap C_2) \leq \theta_B \max\{d(y, C_1), d(y, C_2)\}, \tag{8}$$

where $B^m = \underbrace{B \times \cdots \times B}_{m \text{ times}}$. Let $y = (x, x, \ldots, x)$ then:

$$d(y, C_1 \bigcap C_2) = \sqrt{m} \cdot d\left(x, \bigcap_{i=1}^m D_i\right),$$

$$d(y, C_1) = 0,$$

$$d(y, C_2) = \sqrt{\sum_{j=1}^m d^2(x, D_j)} \leq \sqrt{m} \max_{j=1,\ldots,m} \{d(x, D_j)\}.$$

Substituting these equations in (8) we obtain:

$$\forall x \in B \quad d\left(x, \bigcap_{i=1}^m D_i\right) \leq \theta_B \max_{i=1,\ldots,m} \{d(x, D_i)\}.$$

■

The next simple convex analysis fact will allow us to relax the interior condition to a relative interior condition.

LEMMA 3.2 *Let C be a closed convex set in $\Re^d$. Then there exists a closed convex sets $\tilde{C}$ such that $C \subseteq \tilde{C}$ and the following is satisfied*:

$$\mathrm{aff}(C) \bigcap \mathrm{int}(\tilde{C}) = \mathrm{ri}(C),$$

$$\mathrm{aff}(C) \bigcap \tilde{C} = C.$$

Proof Take $\tilde{C}$ to be:

$$\tilde{C} = C + M,$$

where M is the orthogonal complement to the linear subspace parallel to aff(C). First, we prove that $\mathrm{aff}(C) \cap \tilde{C} = C$. It is obvious that $C \subseteq \mathrm{aff}(C)$, $C \subseteq \tilde{C}$ and thus $C \subseteq \mathrm{aff}(C) \subset \tilde{C}$. Now, we will prove the second direction: $\mathrm{aff}(C) \cap \tilde{C} \subseteq C$. Let $x \in \mathrm{aff}(C) \cap \tilde{C}$. $x \in \tilde{C}$ so there axe y, z such that:

$$x = y + z \quad y \in C, \quad z \in M.$$

Thus,

$$x - y = z \in M.$$

Since $x, y \in \mathrm{aff}(C)$ it follows that $x - y \in M^{\perp}$ which yields that $z = 0$. As a conclusion $x = y \in C$ which proves that $\mathrm{aff}(C) \cap \tilde{C} = C$. Now,

$$\mathrm{ri}(C) = \mathrm{ri}(\mathrm{aff}(C) \bigcap \tilde{C}) = \mathrm{ri}(\mathrm{aff}(C)) \bigcap \mathrm{ri}(\tilde{C}) = \mathrm{aff}(C) \bigcap \mathrm{ri}(\tilde{C}).$$

All that is left to be proved is that $\mathrm{aff}(\tilde{C}) = \Re^d$ and indeed:

$$\mathrm{aff}(\tilde{C}) = \mathrm{aff}(C + M) = \mathrm{aff}(C) + M = \Re^d$$

■

We are now in position to prove the main result of this section, i.e., that Slater's condition implies LEB.

THEOREM 3.1 (Slater Implies LEB) *Let $C_1, \ldots, C_k$ be polyhedral sets and let $D_1, \ldots, D_m$ be closed convex sets. If Slater's condition is satisfied i.e.,*

$$\left(\bigcap_{i=1}^{k} C_i\right) \bigcap \left(\bigcap_{j=1}^{m} \mathrm{ri}(D_j)\right) \neq \emptyset, \tag{9}$$

then LEB is satisfied, i.e. for every bounded set B there is $\theta_B > 0$ such that,

$$\forall x \in B \quad d\left(x, \left(\bigcap_{i=1}^{k} C_i\right) \bigcap \left(\bigcap_{j=1}^{m} D_j\right)\right) \leq \theta_B \max_{i=1,\ldots,k,\, j=1,\ldots,m} \{d(x, C_i), d(x, D_j)\}.$$

Proof Let $\tilde{D}_1, \ldots, \tilde{D}_m$ be defined as in Lemma 3.2 i.e.

$$\mathrm{ri}(D_j) = \mathrm{aff}(D_j) \bigcap \mathrm{int}(\tilde{D}_j)$$
$$D_j = \mathrm{aff}(D_j) \bigcap \tilde{D}_j.$$

Then (9) is equivalent to:

$$\underbrace{\left(\bigcap_{i=1}^{k} C_i\right) \bigcap \left(\bigcap_{j=1}^{m} \mathrm{aff}(D_j)\right)}_{E} \bigcap \mathrm{int}\left(\underbrace{\bigcap_{j=1}^{m} \tilde{D}_j}_{F}\right) \neq \emptyset.$$

Since $E \cap \text{int}(F) \neq \emptyset$ we have from Lemma 3.1 that there is $\theta_B > 0$ such that:

$$\forall x \in B \quad d(x, E \bigcap F) \leq \theta_B \max\{d(x, E), d(x, F)\}$$

By Corollary 3.1 there is $\gamma_B > 0$ such that:

$$\forall x \in B \quad d(x, F) \leq \gamma_B \max_{j=1,\dots,m} \{d(x, \tilde{D}_j)\}.$$

Noticing that the set E is a nonempty polyhedral set (since it contains $E \cap \text{int}(F)$), we can apply Hoffman's Lemma [8], namely, there is a $\delta_B > 0$ such that:

$$\begin{aligned}\forall x \in B \quad d(x, E) &= d\left(x, \left(\bigcap_{i=1}^{k} C_i\right) \bigcap \left(\bigcap_{j=1}^{m} \text{aff}(D_j)\right)\right) \\ &\leq \delta_B \max_{i=1,\dots,k, j=1,\dots,m} \{d(x, C_i), d(x, \text{aff}(D_j))\}.\end{aligned}$$

Therefore,

$$\begin{aligned}\forall x \in B \quad d(x, E \bigcap F) &\leq \theta_B \max\{\gamma_B, \delta_B\} \\ &\quad \times \max_{i=1,\dots,k, j=1,\dots,m} \{d(x, \tilde{D}_j), d(x, C_i), d(x, \text{aff}(D_j))\} \\ &\leq \theta_B \max\{\gamma_B, \delta_B\} \max_{i=1,\dots,k, j=1,\dots,m} \{d(x, C_i), d(x, D_j)\}.\end{aligned}$$

The last inequality is true because $D_i \subseteq \text{aff}(D_i)$, $D_i \subseteq \tilde{D}_i$ and thus,

$$d(x, \tilde{D}_i), d(x, \text{aff}(D_i)) \leq d(x, D_i).$$

■

4 THE RATE OF CONVERGENCE OF PROJECTION ALGORITHMS CAN BE VERY SLOW

We concentrate only on the CPA for CFP. Similar results can be established for the other projection algorithms. We have already proven in Section 2 that the sequence $\{x_n\}$ generated by CPA converges to a point $x^* \in C_1 \cap C_2$, with C_1, C_2 as defined there. We will now see that is Slater's condition is not valid then for every even p we can find an example of closed convex sets C_1, C_2 such that:

$$\frac{1}{(an + b)^{1/(2p-2)}} \leq \|x_n - x^*\| \leq \frac{1}{(cn + d)^{1/(2p-2)}},$$

where a, b, c, d are some positive numbers. That is, the rate of convergence can be very slow. The two closed convex sets that we will consider are:

$$C_1 = \{(s, 0): s \in \Re\},$$

$$C_2 = \{(s, t): s^p - t \leq 0\}.$$

Notice that $C_1 \cap C_2 = \{(0, 0)\}$ and thus Slater's condition is not satisfied. Also, since we will prove that the sequence generated by CPA does not have a linear rate of convergence we necessarily obtain that LEB is not satisfied for C_1, C_2 (otherwise, by our previous results we would have linear convergence). The sequence generated by CPA is:

$$(s_0, t_0) = (1, 0), \tag{10}$$

$$(s_{n+1}, 0) = P_{C_1}(P_{C_2}(s_n, 0)). \tag{11}$$

The following lemma states that the sequence $\{s_n\}$ satisfies an implicit recursive relation.

LEMMA 4.1 *Let p be an even integer. The sequence $\{s_n\}$ generated by CPA satisfies the following relation.*

$$\begin{aligned} s_0 &= 1 \\ s_n &= s_{n+1} + p s_{n+1}^{2p-1} \end{aligned} \tag{12}$$

Proof Denote $P_{C_2}(s_n, 0) = (s_{n+1}, t_{n+1})$. Then, (s_{n+1}, t_{n+1}) is the solution of the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & (s - s_n)^2 + t^2 \\ \text{s.t.} \quad & s^p - t \leq 0 \end{aligned}$$

By the KKT condition we have that:

$$\begin{cases} 2t_{n+1} - \lambda = 0 \\ 2(s_{n+1} - s_n) + p\lambda s_{n+1}^{p-1} = 0 \end{cases}$$

Also, $(s_{n+1}, t_{n+1}) \in bd(C_2)$ and thus $t_{n+1} = s_{n+1}^p$. To conclude, we have:

$$2(s_{n+1} - s_n) = -p\lambda s_{n+1}^{p-1} = -2pt_{n+1}s_{n+1}^{p-1} = -2ps_{n+1}^{2p-1}.$$

■

The next lemma bounds the value of s_{n+1} with respect to the value of s_n. These bounds will play a crucial role in establishing the convergence rate of the sequence.

LEMMA 4.2 *There exists $0 < \gamma < 1$ such that for every n:*

$$\gamma s_n < s_{n+1} < s_n.$$

Proof $s_{n+1} < s_n$ by the definition of the sequence. Also,

$$s_{n+1} = s_n - ps_{n+1}^{2p-1} \overset{s_{n+1} < s_n}{>} s_n - ps_n^{2p-1}.$$

On the other hand, by the convergence of the sequence generated by CPA we have that $s_n \to 0$ and thus there exists a natural N such that for every $n > N$ we have $x_n < 1/2$. Thus, for every $n > N$,

$$s_{n+1} = s_n - ps_{n+1}^{2p-1} > s_n - p\left(\frac{1}{2}\right)^{2p-2} s_n = s_n\left(1 - p\left(\frac{1}{2}\right)^{2p-2}\right).$$

Define γ to be smaller than $\max\{1 - p(1/2)^{2p-2}, s_1/s_0, s_2/s_1, \ldots, s_{N+1}/s_N\}$ and the lemma is proved. ■

We are now ready to prove the main result of this section: "*CPA can converge as slow as one wishes*".

THEOREM 4.1 *Let* $\{s_n\}$ *be the sequence generated by CPA as described by* (10), (11). *Then there exists positive numbers* a, b, c, d *such that*:

$$\frac{1}{(an+b)^{1/(2p-2)}} \leq s_n \leq \frac{1}{(cn+d)^{1/(2p-2)}}.$$

Proof Notice that:

$$\begin{aligned}
\frac{1}{s_{n+1}^{2p-2}} - \frac{1}{s_n^{2p-2}} &= \frac{s_n^{2p-2} - s_{n+1}^{2p-2}}{s_{n+1}^{2p-2} s_n^{2p-2}} \\
&= \frac{(s_n - s_{n+1})\left(\sum_{k=0}^{2p-3} s_n^k s_{n+1}^{2p-3-k}\right)}{s_{n+1}^{2p-2} s_n^{2p-2}} \\
&\overset{(12)}{=} \frac{p s_{n+1}^{2p-1}\left(\sum_{k=0}^{2p-3} s_n^k s_{n+1}^{2p-3-k}\right)}{s_{n+1}^{2p-2} s_n^{2p-2}} \\
&= \frac{p s_{n+1}\left(\sum_{k=0}^{2p-3} s_n^k s_{n+1}^{2p-3-k}\right)}{s_n^{2p-2}}
\end{aligned}$$

Invoking Lemma 4.2, we can thus bound the expression $1/s_{n+1}^{2p-2} - 1/s_n^{2p-2}$ from above:

$$\frac{1}{s_{n+1}^{2p-2}} - \frac{1}{s_n^{2p-2}} = \frac{p s_{n+1}\left(\sum_{k=0}^{2p-3} s_n^k s_{n+1}^{2p-3-k}\right)}{s_n^{2p-2}} \overset{s_{n+1}<s_n}{<} p(2p-2),$$

and from below:

$$\frac{1}{s_{n+1}^{2p-2}} - \frac{1}{s_n^{2p-2}} = \frac{p s_{n+1}\left(\sum_{k=0}^{2p-3} s_n^k s_{n+1}^{2p-3-k}\right)}{s_n^{2p-2}} \overset{s_{n+1}>\gamma s_n}{>} p(2p-2)\gamma^{2p-2},$$

Summing this inequalities we obtain:

$$\frac{1}{s_n^{2p-2}} - 1 = \sum_{k=0}^{n-1}\left(\frac{1}{s_{k+1}^{2p-2}} - \frac{1}{s_k^{2p-2}}\right) < p(2p-2)n,$$

$$\frac{1}{s_n^{2p-2}} - 1 = \sum_{k=0}^{n-1}\left(\frac{1}{s_{k+1}^{2p-2}} - \frac{1}{s_k^{2p-2}}\right) > \gamma^{2p-2} p(2p-2)n.$$

Therefore,

$$\frac{1}{(p(2p-2)n+1)^{1/(2p-2)}} < s_n < \frac{1}{(\gamma^{2p-2} p(2p-2)n+1)^{1/(2p-2)}},$$

and the theorem is proved with $a := p(2p-2), b := 1, c := \gamma^{2p-2}p(2p-2)$, and $d = 1$. ■

5 TWO POINT PROJECTION ALGORITHMS

5.1 Definition of TPA

In this section we consider convex feasibility problems with two sets. Note that such a model of the CFP includes conic feasibility problems, i.e., find $x \in \{x: Ax = b\} \cap K$ where A is a linear map, K is a closed convex cone, and which arise in several applications, see e.g. Ref. [14]. We begin by recalling the gradient projection algorithm (see e.g., Refs. [5,9]) to solve the following optimization problem:

$$\text{(OP)} \quad \min_{x \in S} f(x).$$

Here we assume that S is a closed convex set and f is a convex differentiable function with Lipschitz gradient with Lipschitz constant L, i.e.,:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in S.$$

We assume that the optimal set of (OP), denoted by X^*, is nonempty and denote the optimal value of (OP) by f^*. The Gradient Projection Algorithm (in short, GPA) is defined as follows:

Gradient Projection Algorithm (GPA):
First step: Take an arbitrary $x_0 \in \Re^d$
General step: $x_{n+1} = P_S(x_n - \alpha \nabla f(x_n))$.

Here $\alpha > 0$ is the step size. One of the main results about the gradient projection algorithm is the sublinear rate of convergence of the function values, see Ref. [9] for a proof, and the Appendix.

THEOREM 5.1 *Let $\{x_n\}$ be the sequence generated by GPA with constant step size* $0 < \alpha < 2/L$. *Then for every $n \geq 1$ the following is satisfied*:

$$f(x_n) - f^* \leq \frac{c}{n},$$

for some constant $c > 0$. Furthermore, the sequence $\{x_n\}$ converges to an optimal solution of (OP).

There is a simple connection between the projection algorithms previously defined (CPA, MPA) and the gradient projection algorithm. Consider the convex feasibility problem with two closed convex sets C_1, C_2. Then, the feasibility problem is equivalent to the solution of the following optimization problem:

$$\text{(P)} \quad \min \left\{ \frac{1}{2}\|x - y\|^2 : x \in C_1, y \in C_2 \right\}.$$

The optimal set of (P) is $(C_1 \cap C_2) \times (C_1 \cap C_2)$ and the optimal value $f^* = 0$ (under the assumption that $C_1 \cap C_2 \neq \emptyset$). The gradient projection algorithm applied to (P) uses two points in each iteration (one for each set). Thus the algorithm will be called TPA (Two Points projection Algorithm) and takes the following form:

Two Points Projection Algorithm (TPA):
Initial step: Take an arbitrary $x_0 \in C_1$, $y_0 \in C_2$
General step: $x_{n+1} = P_{C_1}((1-\alpha)x_n + \alpha y_n)$, $y_{n+1} = P_{C_2}((1-\alpha)y_n + \alpha x_n)$

Here $\alpha > 0$ is the step size and thus by Theorem 5.1 the convergence is guaranteed when $0 < \alpha < 2/L$ where L is the Lipschitz constant of the gradient of the objective function. It is easy to see that here $L = 2$ and thus the algorithm will converge for every $0 < \alpha < 1$.

Remark 5.1 It is interesting to note that:

1. In the case $\alpha = 1/2$, TPA reduces to MPA with equal weights ($x_{n+1} = [P_{C_1}(x_n) + P_{C_2}(x_n)]/2$). The convergence of TPA is then guaranteed by Theorem 2.1 and also from the known results about the gradient projection algorithm given in Theorem 5.1.
2. In the case $\alpha = 1$, convergence of TPA is not anymore guaranteed by Theorem 5.1 which requires $\alpha \in (0, 1)$. But when $\alpha = 1$ TPA reduces to CPA so convergence is guaranteed by Theorem 2.1.

5.2 The Rate of Convergence of TPA

As just noticed, TPA reduces to MPA with equal weights when $\alpha = 1/2$. Thus, as already indicated in Section 4, this algorithm can converge very slowly. However, in this case, and as shown below, we can still bound the rate of convergence of the residual: $\max\{d(x_n, C_2), d(y_n, C_1)\}$ (Recall that by the definition of TPA $x_n \in C_1$, $y_n \in C_2$).

THEOREM 5.2 *Let C_1, C_2 be two closed convex sets with nonempty intersection and let $0 < \alpha < 1$. Then the sequence $\{(x_n, y_n)\}_{n=0}^{\infty}$ generated by TPA converge to a point in $(C_1 \cap C_2) \times (C_1 \cap C_2)$ and the following is satisfied: there exists a constant $A > 0$ such that:*

$$\max\{d(x_n, C_2), d(y_n, C_1)\} \leq \frac{A}{\sqrt{n}}.$$

Proof Denote $f(x, y) = (1/2)\|x - y\|^2$. Then by Theorem 5.1 we have that there exists $c > 0$ such that:

$$\|x_n - y_n\|^2 \leq \frac{c}{n},$$

or equivalently,

$$\|x_n - y_n\| \leq \frac{\sqrt{c}}{\sqrt{n}}.$$

Since $x_n \in C_1$, $y_n \in C_2$ this implies the following two inequalities:

$$d(x_n, C_2) \leq \|x_n - y_n\| \leq \frac{\sqrt{c}}{\sqrt{n}},$$

$$d(y_n, C_1) \leq \|x_n - y_n\| \leq \frac{\sqrt{c}}{\sqrt{n}}.$$

Define $A = \sqrt{c}$ and the theorem is proved. ■

5.3 A Relation Between Error Bounds

Error bounds can also be defined for optimization problems such as (OP). A classical residual function is given by $R(x) = \|x - P_S(x - \alpha\nabla f(x))\|$. Obviously, $R(x) \geq 0$ for any $x \in \Re^d$ and $R(x) = 0$ if and only if $x \in X^*$, see e.g., Ref. [12]. We consider now the following Gradient Error Bound (for short GREB), which is essentially a modification of an error bound suggested by Luo [10], and was recently studied in Ref. [2].

DEFINITION 5.1 (GREB) (OP) *is said to satisfy GREB with parameter* $0 < \alpha < 2/L$ *if for every closed bounded set B there exists* $\sigma_B > 0$ *such that*:

$$\forall x \in B \bigcap S \quad d(x, X^*) \leq \sigma_B \|x - P_S(x - \alpha\nabla f(x))\|,$$

where X^* *is the optimal set.*

The connection between error bound assumptions and linear rate of convergence has already been studied for some cases in the literature, see for example Luo [10] and references therein, who proves that a condition similar to GREB implies asymptotic linear rate of convergence of the function values of the sequence generated by the gradient projection method for solving OP even in the case where f is nonconvex. When f is also assumed convex GREB implies nonasymptotic linear rate of convergence of the sequence generated by GPA, (see Ref. [2]). Writing GREB for the optimization problem (P) induced by the feasibility problem we obtain: For every bounded set B, there exists $\sigma_B > 0$ such that:

$$\begin{aligned} d((x, y), (C_1 \bigcap C_2) \times (C_1 \bigcap C_2)) \leq \sigma_B \|(x - P_{C_1}((1-\alpha)x + \alpha y),\\ y - P_{C_2}((1-\alpha)y + \alpha x))\|. \end{aligned}$$

Using the equivalence of norms in finite dimensional spaces we than obtain the following new error bound for the feasibility problem, that will be called TPEB (Two Points Error Bound).

DEFINITION 5.2 (TPEB) *Two sets* C_1, C_2 *are said to satisfy TPEB with parameter* $0 < \alpha \leq 1$ *if for every bounded set B there exists* $\sigma_B > 0$ *such that*:

$$\begin{aligned} \max\{d(x, C_1 \bigcap C_2), d(y, C_1 \bigcap C_2)\} \leq \sigma_B \max\{\|x - P_{C_1}((1-\alpha)x + \alpha y)\|,\\ \|y - P_{C_2}((1-\alpha)y + \alpha x)\|\}, \end{aligned}$$

for every $x \in B \cap C_1$, $y \in B \cap C_2$.

Using the results just mentioned above for the GPA algorithm, we can thus conclude that TPEB implies linear convergence of the sequences generated by TPA. We record this in the following theorem.

THEOREM 5.3 (Linear Rate of Convergence of the sequence) *Let* C_1, C_2 *be two close convex sets with nonempty intersection. Suppose that TPEB is satisfied. Let* $\{(x_n, y_n)\}$ *be a sequence generated by TPA with constant* $0 < \alpha < 1$. *Then there is a* $0 < \eta < 1$ *such that*,

$$d(x_{n+1}, C_1 \bigcap C_2) \leq \eta d(x_n, C_1 \bigcap C_2),$$

$$d(y_{n+1}, C_1 \bigcap C_2) \leq \eta d(y_n, C_1 \bigcap C_2)$$

Both conditions: LEB, TPEB imply linear convergence of the sequence generated by their associated algorithms (MPA, TPA respectively). As already noted, TPA with $\alpha = 1/2$ is in fact MPA with equal weights. Thus, both LEB and TPEB imply the linear convergence of MPA with equal weights. The question that naturally arises is: *what is the weaker condition?* The next theorem answers this question and it turns out that LEB implies TPEB.

THEOREM 5.4 (LEB $\Rightarrow$ TPEB) *Let C_1, C_2 be two closed convex sets with nonempty intersection that satisfy LEB. Then, for every $0 < \alpha < 1$, TPEB holds.*

Proof Let B be a bounded set. Then, by LEB we have:

$$\forall x \in B \quad d(x, C_1 \bigcap C_2) \leq \theta_B \max\{d(x, C_1), d(x, C_2)\}.$$

Thus, for every $x \in C_1 \cap B$, $y \in C_2 \cap B$:

$$\begin{aligned} d(x, C_1 \bigcap C_2) &\leq \theta_B \max\{d(x, C_1), d(x, C_2)\} = \theta_B d(x, C_2) \\ &\leq \theta_B d(x, P_{C_2}((1-\alpha)x + \alpha y)). \end{aligned}$$

By taking y instead of x we obtain the inequality:

$$d(y, C_1 \bigcap C_2) \leq \theta_B d(y, P_{C_1}((1-\alpha)y + \alpha x)).$$

Combining these two inequalities we obtain TPEB. ■

References

[1] A. Auslender (1976). *Optimization: Methodes Numeriques.* Masson, Paris.
[2] A. Beck (2002). *Convergence Rate Analysis of Gradient-Based Algorithms*. PhD thesis, Tel-Aviv University, Israel.
[3] H.H. Bauschke and J.M. Borwein (1996). On projection algorithms for solving convex feasibility problems. *SIAM Review*, **38**, 367–426.
[4] H.H. Bauschke, J.M. Borwein and Wu Li (1999). On the strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and applications to convex optimization. *Math. Program.*, **86**, 135–160.
[5] D. Bertsekas (1999). *Nonlinear Programming*, 2nd ed. Athena Scientific, Belmont Massachussets.
[6] G. Cimmino (1938). Calcolo apppresimato per le soluzioni de sistemi di equazioni lineari. *La Ricera Scientifica Roma*, **1**, 326–333.
[7] L.C. Gubin, B.T. Polyak and E.V. Raik (1967). The method of projections for finding the common point of convex sets. *USSR Compututational Math. Math. Phys.*, **7**, 1–24.
[8] A.J. Hoffmann (1952). On approximate solutions of systems of linear inequalities. *J. Research Nat. Bur. Standards*, **49**, 263–265.
[9] E.S. Levitin and B.T. Polyak (1966). Constrained minimization methods. *USSR Computational Math. Math. Phys.*, **6**, 1–50.
[10] Z.Q. Luo (2000). New error bounds and their applications to convergence analysis of iterative algorithms. *Math. Program.*, **88**, 341–355.
[11] J.M. Ortega and W.C. Rheinbolt (1970). *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press Inc., London.
[12] J.S. Pang (1997). Error bounds in mathematical Programing. *Math. Program.*, **79**, 299–332.
[13] R.T. Rockafellar (1970). *Convert Analysis.* Princeton University Press, Princeton, NJ.
[14] H. Stark (1987). *Image Recovery: Theory and Applications.* Academic Press.
[15] J. von Neumann (1950). *Functional Operators. II. The Geometry of Orthogonal Spaces*, Annals of Math. Studies. no. 22. Princeton University Press, Princeton, NJ.

APPENDIX: PROOF OF THEOREM 2.2 AND THEOREM 5.1

Proof of Theorem 2.2 First, we prove that $x_n \in B$ for every $n \geq 0$. By (3) we have that for all $y \in C$:

$$\|x_n - y\| \leq \|x_{n-1} - y\| \leq \cdots \leq \|x_0 - y\|,$$

and thus $x_n \in B$. By LEB there is a $\theta_B > 0$ such that

$$d(x_n, C) \leq \theta_B \max_{i=1,\ldots,m} \{d(x_n, C_i)\} \quad \forall n \geq 0.$$

Now, for all $n \geq 0$:

$$\begin{aligned}
d^2(x_n, C) &\leq \theta_B^2 \max_{i=1,\ldots,m} \{d^2(x_n, C_i)\} \\
&\leq \frac{\theta_B^2}{\min_{j=1,\ldots,m}\{\alpha_j\}} \sum_{i=1}^{m} \alpha_i d^2(x_n, C_i) \\
&\overset{(2)}{\leq} \frac{\theta_B^2}{\min_{j=1,\ldots,m}\{\alpha_j\}} (d^2(x_n, C) - d^2(x_{n+1}, C)).
\end{aligned}$$

Define $A = \theta_B^2/(\min_{j=1,\ldots,m}\{\alpha_j\})$ and we obtain that:

$$d^2(x_{n+1}, C) \leq \left(1 - \frac{1}{A}\right) d^2(x_n, C)$$

proving the desired result. ∎

Proof of Theorem 5.1 Let x^* be an arbitrary point in X^*. By the gradient inequality for convex functions we obtain

$$0 \geq f(x^*) - f(x_n) \geq \langle \nabla f(x_n), x^* - x_n \rangle. \tag{13}$$

From the basic properties of the projection operator we have that:

$$\langle x_n - \alpha \nabla f(x_n) - x_{n+1}, x^* - x_{n+1} \rangle \leq 0,$$

which is equivalent to the following inequality:

$$\langle x_n - x_{n+1}, x^* - x_{n+1} \rangle \leq \alpha \langle \nabla f(x_n), x^* - x_{n+1} \rangle. \tag{14}$$

Moreover,

$$\begin{aligned}
\|\nabla f(x_n)\|^2 &= \|\nabla f(x_n) - \nabla f(x_0) + \nabla f(x_0)\|^2 \\
&\overset{(a+b)^2 \leq 2a^2+2b^2}{\leq} 2\|\nabla f(x_n) - \nabla f(x_0)\|^2 + 2\|\nabla f(x_0)\|^2 \\
&\leq 2L^2\|x_n - x_0\|^2 + 2\|\nabla f(x_0)\|^2 \\
&= 2L^2\|x_n - x^* + x^* - x_0\|^2 + 2\|\nabla f(x_0)\|^2 \\
&\overset{(a+b)^2 \leq 2a^2+2b^2}{\leq} 4L^2\|x_n - x^*\|^2 + 4L^2\|x_0 - x^*\|^2 + 2\|\nabla f(x_0)\|^2 \\
&\leq 8L^2\|x^* - x_0\|^2 + 2\|\nabla f(x_0)\|^2 \qquad (15)
\end{aligned}$$

As a result,

$$
\begin{aligned}
(f(x_n) - f^*)^2 &\overset{(13)}{\leq} \langle \nabla f(x_n), x^* - x_n \rangle^2 \\
&= (\langle \nabla f(x_n), x^* - x_{n+1} \rangle + \langle \nabla f(x_n), x_{n+1} - x_n \rangle)^2 \\
&\overset{(a+b)^2 \leq 2(a^2+b^2)}{\leq} 2\langle \nabla f(x_n), x^* - x_{n+1} \rangle^2 + 2\langle \nabla f(x_n), x_{n+1} - x_n \rangle^2 \\
&\overset{(14)}{\leq} 2\left(\frac{1}{\alpha^2} \langle x_n - x_{n+1}, x^* - x_{n+1} \rangle^2 + \langle \nabla f(x_n), x_{n+1} - x_n \rangle^2 \right) \\
&\leq 2\left(\frac{1}{\alpha^2} \|x_n - x_{n+1}\|^2 \|x^* - x_{n+1}\|^2 + \|\nabla f(x_n)\|^2 \|x_{n+1} - x_n\|^2 \right) \\
&\leq 2\left(\frac{1}{\alpha^2} \|x_n - x_{n+1}\|^2 \|x^* - x_0\|^2 + \|\nabla f(x_n)\|^2 \|x_{n+1} - x_n\|^2 \right) \\
&= 2\|x_n - x_{n+1}\|^2 \left(\frac{1}{\alpha^2} \|x^* - x_0\|^2 + \|\nabla f(x_n)\|^2 \right) \\
&\leq \frac{1}{1/\alpha - L/2} \left(\frac{1}{\alpha^2} \|x^* - x_0\|^2 + \|\nabla f(x_n)\|^2 \right) (f(x_n) - f(x_{n+1})) \\
&\overset{(15)}{\leq} \frac{2}{1/\alpha - L/2} \left(\frac{1}{\alpha^2} \|x^* - x_0\|^2 + 8L^2 \|x^* - x_0\|^2 + 2\|\nabla f(x_0)\|^2 \right) \\
&\quad \times (f(x_n) - f(x_{n+1})) \\
&= \frac{2}{1/\alpha - L/2} \left(\left(\frac{1}{\alpha^2} + 8L^2 \right) \|x^* - x_0\|^2 + 2\|\nabla f(x_0)\|^2 \right) \\
&\quad \times (f(x_n) - f(x_{n+1})) \\
&= A((f(x_n) - f^*) - (f(x_{n+1}) - f^*)),
\end{aligned}
$$

where $A = 2/(1/\alpha - L/2)((1/\alpha^2 + 8L^2)\|x^* - x_0\|^2 + 2\|\nabla f(x_0)\|^2)$. The result then follows. ∎