Math. Program., Ser. B 88: 341–355 (2000)

Digital Object Identifier (DOI) 10.1007/s101070000173

Zhi-Quan Luo

New error bounds and their applications to convergence analysis of iterative algorithms*

Received: October 5, 1999 / Accepted: January 1, 2000
Published online July 20, 2000 –

Abstract. We present two new error bounds for optimization problems over a convex set whose objective function f is either semianalytic or γ-strictly convex, with $\gamma \geq 1$. We then apply these error bounds to analyze the rate of convergence of a wide class of iterative descent algorithms for the aforementioned optimization problem. Our analysis shows that the function sequence $\{f(x^k)\}$ converges at least at the sublinear rate of $k^{-\varepsilon}$ for some positive constant ε, where k is the iteration index. Moreover, the distances from the iterate sequence $\{x^k\}$ to the set of stationary points of the optimization problem converge to zero at least sublinearly.

1. Introduction

Error bounds for a given subset S of an Euclidean space is an inequality that bounds the distance from an arbitrary vector to S in terms of a residual function. The latter is usually an easily computable function satisfying the property that $r(x) \geq 0, \forall x \in \Re^n$ and $r(x) = 0$ iff $x \in S$. In recent years, there has been considerable interest in the study of error bounds in mathematical programming. This is partly motivated by the applications of error bounds which are rich and diverse. The application areas of error bounds include, among other things, the rate of convergence analysis of iterative algorithms for degenerate problems, the regularity and stability of inequality systems, weak sharp minima, accurate identification of active constraints, the study of stationary/minimizing sequences of convex programs, exact penalty functions, and more recently the analysis of interior point methods for linear conic programs. For an excellent summary of the theory and applications of error bounds, we refer the readers to the recent survey of Pang [38] and the references cited therein.

In this paper, we consider error bounds and iterative algorithms for a general constrained differentiable minimization problem of the form:

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to } x \in X, \end{aligned}$$

where X is a closed convex set. Let S denote the set of stationary points of the above optimization problem which by definition satisfy $x = [x - \nabla f(x)]_X^+$, where $[\cdot]_X^+$ is the

Z.-Q. Luo: Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4K1, e-mail: luozq@mcmail.cis.mcmaster.ca

* This work is based on research supported by the Natural Science and Engineering Research Council of Canada, Grant No. OPG0090391.

projection operator to X. We present some new projection-type error bounds of the form

$$\operatorname{dist}(x, S) \leq \sigma \| x - [x - \nabla f(x)]_X^+ \|^{1/\gamma},$$

for some positive constants σ and γ, where $\operatorname{dist}(x, S)$ denotes the distance function from x to the set S. These error bound results are based on either the semianalyticity or γ-strict convexity of the objective function f (see Sect. 2 for definitions), and certain analytic structure of X. Furthermore, we apply these new error bounds in the rate of convergence analysis of iterative descent algorithms for the minimization of f over X. It is shown that in general the convergence rate of $\{f(x^k)\}$ and $\{\operatorname{dist}(x^k, S)\}$ is sublinear (i.e., like $k^{-\varepsilon}$ for some $\varepsilon > 0$). These results extend the early work of [25]–[31] and [39,40] where certain form of strong convexity is assumed. Our analysis is fairly general as it is applicable to a wide class of iterative descent algorithms, including a gradient projection algorithm of Goldstein [11] and Levitin and Polyak [2]; a certain matrix splitting algorithm ([33, 36]), coordinate descent methods (see [1,5,7,22,35,42]); the extragradient method of Korpelevich [16]; the proximal minimization algorithm of Martinet [34].

The rest of the paper is organized as follows. Section 2 presents the notions of semianalyticity and derives the new error bound results. Section 3 introduces a general algorithmic framework of [28] and establishes the rate of convergence for the iterate sequence $\{x^k\}$ and the function sequence $\{f(x^k)\}$. The conclusions are given in Sect. 4.

2. New error bounds for optimization problems

Let us first introduce some definitions. We recall that a real-valued function f defined on open subset U of $\Re^n$ is *analytic* if it can be represented by a convergent power series in the neighborhood of any point of U; a vector-valued function F from the open set U into $\Re^m$ is *analytic* if each of its component functions is analytic. A subset X of $\Re^n$ is *semianalytic* if for each vector $a \in \Re^n$ there are a neighborhood U of a and a finite family of sets $X_{ij} \subseteq \Re^n$, $i = 1, \ldots, p$, and $j = 1, \ldots, q$, each of the form

$$X_{ij} \equiv \{x \in U : f_{ij}(x) = 0\} \quad \text{or} \quad \{x \in \Re^n : f_{ij}(x) < 0\}$$

for some real analytic function f_{ij} on U, such that

$$X \cap U = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} X_{ij}.$$

A subset X of $\Re^n$ is *subanalytic* if for each vector $a \in \Re^n$ there exists a neighborhood U of a such that $X \cap U$ is the projection of a bounded semianalytic set $A \subset \Re^{n+p}$ for some nonnegative integer p; i.e.,

$$X \cap U = \Pi(A),$$

where $\Pi : \Re^{n+p} \to \Re^n$ is given by $\Pi(x, y) = x$ for all $(x, y) \in \Re^{n+p}$. A vector-valued function is *semianalytic (subanalytic)* if its graph is semianalytic (subanalytic). The reference [6] provides an extensive study of semianalytic and subanalytic sets, and [8, Sect. 2] summarizes several important examples of these sets and functions. The following is a list of known subanalytic functions:

(a) piecewise analytic functions defined over a semianalytic partition are semianalytic (thus subanalytic);
(b) the pointwise supremum of a finite family of continuous subanalytic functions is subanalytic;
(c) the class of continuous subanalytic functions defined on a compact subanalytic set is closed under algebraic operations;
(d) the image of a bounded subanalytic set by a subanalytic function is subanalytic but this property is not valid if "subanalytic" is replaced by "semianalytic".

In particular, semianalytic and subanalytic functions need not be smooth.

Below we state a local error bound result of Lojasiewicz [21] for a semianalytic inequality systems (see also [23, 24]):

Theorem 1. *For $i = 1, \ldots, m$, let $g_i(x)$ be a semianalytic function defined in $\Re^n$. Suppose that the set S defined by*

$$S \equiv \{ x \in \Re^n : g(x) \leq 0 \}. \tag{1}$$

is nonempty. Then for every scalar $\rho > 0$, there exist positive scalars c and γ such that

$$\mathit{dist}(x, S) \leq c \, \| g(x)_+ \|^{1/\gamma}, \quad \forall x \in \Re^n \text{ satisfying } \|x\| \leq \rho.$$

Now let us consider the following constrained minimization problem:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to } x \in X \end{aligned} \tag{2}$$

where f is a differentiable semianalytic function and X is a convex set defined by

$$X \equiv \{x : g_1(x) \leq 0, \ g_2(x) \leq 0, \ldots, g_m(x) \leq 0\}$$

with each g_i convex, differentiable and semianalytic. Notice that x^* is a stationary point of (2) if and only if

$$x^* = [x^* - \nabla f(x^*)]_X^+,$$

where $[\cdot]_X^+$ denotes the projection operator to the convex set X. Let S denote the set of stationary points of (2). We will use Theorem 1 to show a new error bound for the set S. In particular, for any $x \in \Re^n$, we show that it is possible to use the residual function

$$\|x - [x - \nabla f(x)]_X^+\|$$

to bound the distance $\mathrm{dist}(x, S)$.

Theorem 2. *Suppose f and $g_1, \ldots, g_m$ are differentiable semianalytic functions, and $g_1, \ldots, g_m$ are convex. Moreover, suppose the feasible set*

$$X \equiv \{x : g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}$$

satisfies a constraint qualification (say, Slater condition) so that the Lagrangian multipliers exist for (2). Let S be the set of stationary points of (2) satisfying $x = [x - \nabla f(x)]_X^+$. Then for each $\rho > 0$, there exists some constants $\gamma > 0$ and $\sigma > 0$ such that

$$\mathit{dist}(x, S) \leq \sigma \|x - [x - \nabla f(x)]_X^+\|^{1/\gamma}, \quad \forall x \in \Re^n \text{ satisfying } \|x\| \leq \rho. \tag{3}$$

Proof. Let $g(x) = (g_1(x), g_2(x), \ldots, g_m(x))^T$. Fix any $x \in \Re^n$ with $\|x\| \le \rho$. Let $z = [x - \nabla f(x)]_X^+$. By assumption, a constraint qualification (say, Slater condition) is satisfied for X, the Karush-Kuhn-Tucker condition for the projection operation $[\cdot]_X^+$ gives

$$z - x + \nabla f(x) - \nabla g(z)^T \mu = 0, \quad \mu \ge 0, \quad \mu^T (Az - b) = 0.$$

Since $x^* \in S$ is equivalent to $x^* = [x^* - \nabla f(x^*)]_X$, it follows that $x^* \in S$ if and only if there exist some z^* and λ^* such that

$$z^* - x^* + \nabla f(x^*) - \nabla g(z^*)^T \lambda^* = 0, \quad \lambda^* \ge 0, \quad (\lambda^*)^T (g(z^*) - b) = 0, \quad z^* = x^*. \tag{4}$$

Let $\bar{S}$ denote the solution set of (4), that is,

$$\bar{S} = \{(x^*, z^*, \lambda^*) \ := \{(x^*, z^*, \lambda^*) \text{ satisfies (4)}\}.$$

Since all the functions involved in the above system are semianalytic, we can apply Theorem 1 to the system (4) and obtain

$$\mathrm{dist}((x, z, \mu), \bar{S}) \le \sigma \|x - z\|^{1/\gamma} = \sigma \|x - [x - \nabla f(x)]_X^+\|^{1/\gamma},$$

where σ and γ are some positive constants depending of $f, g_1, \ldots, g_m$ and ρ only. Since $\mathrm{dist}(x, S) \le \mathrm{dist}((x, z, \mu), \bar{S})$, the above inequality easily implies the desired error bound (3).

□

One drawback of above theorem is that the exponent γ is in general difficult to determine and that the error bound holds only over a compact set. Our next result gives an global error bound with the exponent explicitly determined. To achieve this, we need to make additional assumptions on the objective function f and the constraint set X. In particular, we consider the following restricted version of problem (2):

$$\begin{aligned} &\text{minimize} \ \ f(x) \\ &\text{subject to } x \in X \equiv \{x \in \Re^n \ : \ Ax \ge b\} \end{aligned} \tag{5}$$

where A is a matrix and a is vector of matching dimension, $f : \Re^n \to \Re$ is a γ-strictly convex differentiable function in the sense,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \ge \sigma \|x - y\|^{1+\gamma}, \quad \text{for all } x, \ y \in \Re^n. \tag{6}$$

Here $\gamma \ge 1$ is a constant. Clearly, the case $\gamma = 1$ corresponds to the standard notion of strong convexity. Since f is γ-strictly convex, the minimization problem (5) has exactly one solution, say x^*, which must satisfy the following optimality condition

$$x^* = [x^* - \nabla f(x^*)]_X^+.$$

The next theorem states that, for any $x \in \Re^n$, we can use the residual function

$$\|x - [x - \nabla f(x)]_X^+\|$$

to bound the distance $\|x - x^*\|$.

Theorem 3. *Suppose f is a γ-strictly convex function satisfying* (6) *for some $\gamma \geq 1$. Let x^* be the unique optimal solution for* (5). *Then there holds*

$$\|x - x^*\| \leq \sigma^{-\frac{1}{\gamma}} \left\|x - [x - \nabla f(x)]_X^+\right\|^{\frac{1}{\gamma}}, \qquad \forall x \in \Re^n. \tag{7}$$

Proof. Fix any $x \in \Re^n$. Since the constraints are linear, the Karush-Kuhn-Tucker condition holds:

$$\nabla f(x^*) - A^T\lambda = 0, \quad \lambda \geq 0, \quad \lambda^T(Ax^* - b) = 0. \tag{8}$$

Let $z = [x - \nabla f(x)]_X^+$. Then the Karush-Kuhn-Tucker condition for the projection operation $[\cdot]_X^+$ gives

$$z - x + \nabla f(x) - A^T\mu = 0, \quad \mu \geq 0, \quad \mu^T(Ax - b) = 0. \tag{9}$$

Subtracting (8) from the above equation, we obtain

$$z - x + \nabla f(x) - \nabla f(x^*) - A^T(\mu - \lambda) = 0.$$

Multiplying both sides by $(x - x^*)$ and rearranging the terms yields

$$\begin{aligned}\langle x - z, x - x^*\rangle &= \langle \nabla f(x) - \nabla f(x^*), x - x^*\rangle + \langle x - x^*, A^T(\mu - \lambda)\rangle \\ &\geq \sigma\|x - x^*\|^{1+\gamma} - \langle x - x^*, A^T(\mu - \lambda)\rangle\end{aligned}$$

where $\sigma > 0$ is a constant due to the γ-strict convexity of f (cf. (6)). By further using (8)–(9), we see

$$\begin{aligned}\langle x - x^*, A^T(\mu - \lambda)\rangle &= \langle A(x - x^*), \mu - \lambda\rangle \\ &= \langle A(x - x^*), \mu\rangle + \langle A(x^* - x), \lambda\rangle \\ &= \langle b - Ax^*, \mu\rangle + \langle b - Ax, \lambda\rangle \\ &\leq 0.\end{aligned}$$

Combining this with the previous inequality, we obtain

$$\langle x - z, x - x^*\rangle \geq \sigma\|x - x^*\|^{1+\gamma}.$$

Using Cauchy-Schwartz inequality and simplifying yields

$$\|x - x^*\| \leq \sigma^{-\frac{1}{\gamma}} \|x - z\|^{\frac{1}{\gamma}} = \sigma^{-\frac{1}{\gamma}} \left\|x - [x - \nabla f(x)]_X^+\right\|^{\frac{1}{\gamma}}.$$

The proof is complete.

□

In the case of strong convexity (i.e., $\gamma = 1$), Theorem 3 is well known and was derived in the more general setting of strongly monotone variational inequalities [37, Theorem 3.1]. Simple (unconstrained) examples show that the error bound (7) in general does not hold without γ-strict convexity of f. One of the main difficulties to establish (7) is the non-uniqueness of optimal solutions in the absence of strict convexity. However, if f has a certain composite structure and convex, the above error bound still holds. This was shown in the work of Luo and Tseng [25].

3. Rate of convergence of iterative algorithms

In the convergence analysis of iterative descent algorithms for minimizing a convex function over a convex set, a key role played by error bounds is to establish the rate of convergence of the sequence of functions and iterates. Previously, such convergence analysis invariably assumes strong convexity in which case the optimal solution is unique and the proof of convergence becomes relatively easy. In particular, we only need to argue that the first order optimality condition will hold in the limit. Since there is a unique solution satisfying the first order optimality condition, the convergence of the algorithm follows. This line of argument is not applicable to problems without strong convexity or when the objective function is nonconvex. It turns out that error bounds can help to establish iterate convergence in this case. So far there are many such results obtained for a wide variety of algorithms; these include the Goldstein-Levitin-Polyak gradient projection algorithm for convex minimization problems [9,28] and an extension to monotone variational inequalities [3,16]; the Martinet-Rockafellar proximal algorithms for convex minimization and set-valued inclusion [12,17,20,28,32] and its extensions [15,4]; coordinate descent methods and the related dual relaxation methods [14,26,30]; matrix splitting methods for symmetric, monotone linear complementarity problems and variational inequalities [27,19]; and asynchronous versions of some of these methods [31, 39]. Several papers have presented a unified treatment of these convergence results; see e.g. [25,28,40].

3.1. A preliminary result

In the forthcoming analysis, we will need the following useful rate of convergence result.

Theorem 4. *Let $r : \Re^n \to \Re_+$ be a residual function for a subset S of $\Re^n$. Let there be a set $X \supset S$ and let $\{x^k\} \subset X$ be arbitrary sequence. Suppose there exist positive constants δ, c and γ such that all $x^k \in X$ with $r(x^k) \le \delta$ there hold*

(A1) $r(x^k) > 0$ *for all k;*
(A2)

$$r(x^{k+1}) - r(x^k) \;\le\; -c\, r(x^k)^\gamma. \tag{10}$$

Then the following statement holds.

1. *The sequence $\{r(x^k)\}$ converges to zero and $\gamma \ge 1$.*
2. *If $\gamma = 1$, then the sequence $\{r(x^k)\}$ converges to zero at least* **Q**-*linearly; that is, for all k sufficiently large there holds*

$$r(x^{k+1}) \;\le\; (1-c) r(x^k). \tag{11}$$

3. *If $\gamma > 1$, then there exist positive scalar σ such that for all k sufficiently large:*

$$r(x^k) \;\le\; \sigma k^{-\frac{1}{\gamma-1}}. \tag{12}$$

Proof. Notice that the assumptions (A1) and (A2) imply that the sequence $\{r(x^k)\}$ is monotonically decreasing. Therefore, the sequence $\{r_2(x^k)\}$ has a limit, say r^∞. Clearly, $r^\infty \geq 0$. Taking limit in (10) yields

$$r^\infty \leq r^\infty - c(r^\infty)^\gamma,$$

which further implies $r^\infty = 0$. Next we rewrite (10) to obtain

$$r(x^{k+1}) \leq r(x^k)\big(1 - cr(x^k)^{\gamma-1}\big), \qquad \forall k. \tag{13}$$

Since $r(x^k) \to 0$ and $r(x^k) > 0$, it follows that $\gamma \geq 1$.

We now estimate the rate of convergence for the sequence $\{r(x^k)\}$ when $\gamma = 1$. In this case, the inequality (13) implies

$$r(x^{k+1}) \leq r(x^k)\,(1-c)\,, \qquad \forall k,$$

so the sequence $\{r(x^k)\}$ converges to zero **Q**-linearly.

It remains to estimate the rate of convergence for the sequence $r(x^k)$ for the case $\gamma > 1$. In this case, we show inductively that (12) holds, or in other words, the sequence $\{r(x^k)\}$ converges to zero at least like $\sigma/k^{\frac{1}{1-\gamma}}$, for some σ and ϵ. Suppose (12) holds for some k. Let us consider the case $k+1$. First we notice the following elementary inequality from calculus:

$$y\big(1 - cy^{\gamma-1}\big) \leq \bar{y}\big(1 - c\bar{y}^{\gamma-1}\big), \quad \forall y \in [0, \bar{y}], \text{ when } \bar{y} \text{ is sufficiently small.} \tag{14}$$

$$\frac{1}{k^{\gamma-1}}\left(1 - \frac{\bar{c}}{k}\right) \leq \frac{1}{(k+1)^{\gamma-1}}, \qquad \text{for all large } k, \text{ where } \bar{c} > (\gamma-1). \tag{15}$$

The inequality (14) is due to the fact that the derivative of the function $y\left(1 - cy^{\gamma-1}\right)$ at $y = 0$ is equal to 1, which is positive. The second inequality (15) can be inferred as follows. Let $\epsilon = \bar{c} - (\gamma - 1)$, which is positive under the assumption $\bar{c} > \gamma - 1$. Since $\gamma > 1$, it follows from Taylor expansion that

$$\begin{aligned}
\frac{k^{\gamma-1}}{(k+1)^{\gamma-1}} &= \left(1 - \frac{1}{k+1}\right)^{\gamma-1} \\
&\geq 1 - \frac{\gamma - 1 + 0.5\epsilon}{k+1} \qquad \text{(for all large } k\text{)} \\
&\geq 1 - \frac{\bar{c} - 0.5\epsilon}{k+1} \qquad (\text{use } \gamma - 1 + 0.5\epsilon = \bar{c} - 0.5\epsilon) \\
&\leq 1 - \frac{\bar{c}}{k},
\end{aligned}$$

where the last step follows from $\epsilon > 0$. This shows the validity of (15).

Using the inequality (14) and the inductive hypothesis $r(x^k) \leq \sigma/k^{\gamma-1}$, we obtain from (13) that

$$\begin{aligned}
r(x^{k+1}) &\leq r(x^k)\big(1 - cr(x^k)^{\gamma-1}\big) \\
&\leq \frac{\sigma}{k^\epsilon}\left(1 - \frac{c\,\sigma^{\gamma-1}}{k^{(\gamma-1)\epsilon}}\right), \qquad \text{for all sufficiently large } k.
\end{aligned}$$

Therefore, to complete the induction, we only need to argue that there exists some σ such that

$$\frac{\sigma}{k^{\gamma-1}}\left(1-\frac{c\,\sigma^{\gamma-1}}{k}\right)\leq\frac{\sigma}{(k+1)^{\gamma-1}}. \tag{16}$$

Let us select a $\sigma > 0$ such that

$$c\,\sigma^{\gamma-1} > \gamma - 1.$$

Then the previous inequality (16) follows directly from (15) by identifying $\bar{c} = c\,\sigma^{\gamma-1}$. This completes the proof of (12).

$\square$

Theorem 4 shows that the exponent γ in (10) characterizes the rate of convergence completely. It is seen that in general the sequence $\{r(x^k)\}$ converges to zero sublinearly, with the rate becoming faster as γ gets closer to 1. When $\gamma = 1$, the fastest rate is achieved: the sequence $\{r(x^k)\}$ converges to zero **Q**-linearly. In Sect. 3.3, we shall specialize Theorem 4 to some specific iterative algorithms and perform the rate of convergence analysis.

3.2. A general algorithmic framework

The algorithmic framework illustrated below follows from the work of Luo and Tseng [28]. Suppose $f : \Re^n \to \Re$ is a continuously differentiable function whose gradient is Lipschitz continuous on some nonempty closed convex set X in $\Re^n$, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \quad \forall x \in X,\; y \in X, \tag{17}$$

where L is a positive scalar and $\|\cdot\|$ denotes the usual Euclidean norm. The set X may be specified by linear equalities and/or convex inequalities. We are interested in finding a stationary point of f over X, i.e., a point $x \in \Re^n$ satisfying

$$x = [x - \nabla f(x)]_X^+, \tag{18}$$

where $[\cdot]_X^+$ denotes the orthogonal projection on to X. We assume that $\inf_{x\in X} f(x) > -\infty$ and that the set of stationary points, denoted by S, is nonempty. Well known examples of this problem include linear programs and quadratic programs, among others. We consider the class of feasible descent methods and, in particular, methods that update the iterates according to the formula

$$x^{k+1} := [x^k - \alpha^k \nabla f(x^k) + e^k]_X^+, \qquad k = 0, 1, \ldots \tag{19}$$

where α^k is a positive scalar and e^k is a sufficiently small “error” vector depending on x^k (see (19)). Here we assume the following error estimates holds for all $k > 0$:

$$\|e^k\| \leq \kappa_1\|x^k - x^{k+1}\|, \qquad \text{for some } \kappa_1 > 0. \tag{20}$$

and

$$f(x^{k+1}) - f(x^k) \leq -\kappa_2\|x^k - x^{k+1}\|^2 \qquad \text{for some } \kappa_2 > 0, \tag{21}$$

The condition (21) ensures sufficient descent at each iteration.

The above scheme (19) is a broad class that includes

(a) a gradient projection algorithm of Goldstein [11] and Levitin and Polyak [2];
(b) a certain matrix splitting algorithm ([33,36]), coordinate descent methods (see [1, 5,7,22,35,42]);
(c) the extragradient method of Korpelevich [16];
(d) the proximal minimization algorithm of Martinet [34].

The reference [28] provides a detailed justification of why the above listed algorithms can be cast in the framework (19) and satisfy the conditions (20)–(21).

3.3. Convergence analysis

We proceed to establish the linear convergence of iterative descent algorithms described by (19)–(21). We need to make two assumptions on the function f.

Assumption 1 (Local error bound). *For every* $\upsilon \geq \inf_{x\in X} f(x)$ *there exist scalars* $\gamma > 0$, γ *and* $\tau > 0$ *such that*

$$dist(x, S) \leq \tau \|x - [x - \nabla f(x)]_X^+\|^{1/\gamma}, \tag{22}$$

for all $x \in X$ *with* $f(x) \leq \upsilon$ *and* $\|x - [x - \nabla f(x)]_X^+\| \leq \delta$.

Assumption 2 (Proper separation of isocost surfaces). *There exists a scalar* $\epsilon > 0$ *such that*

$$x \in S, \quad y \in S, \quad f(x) \neq f(y), \quad \Rightarrow \quad \|x - y\| \geq \epsilon.$$

Notice that Assumption 2 holds automatically if f is convex in which case S consists of only one connected convex piece. There are many problems that satisfy the above two assumptions. For example, it is known from the early error bound researches that Assumptions 1 and 2 hold if any of the following conditions hold:

(a) (Semianalytic case; see Theorem 2) f is differentiable and semianalytic with compact level sets. $X = \{x \;:\; g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}$ satisfies a constraint qualification such as Slater condition. Each g_i is convex differentiable and semianalytic. In this case Assumption 1 holds with some $\gamma > 0$.
(b) (γ-Strictly convex case; see Theorem 3). f is γ-strictly convex satisfying (6) and X is polyhedral.
(c) (Quadratic case; see Lemma 3.1 of [27]). f is quadratic (possibly nonconvex). X is a polyhedral set. In this case Assumption 1 holds with $\gamma = 1$.
(d) (Composite case; see [25]).

$$f(x) = \langle q, x\rangle + g(Ex), \quad \forall x,$$

where E is an $m \times n$ matrix with no zero column, q is a vector in $\Re^n$, and g is a strongly convex differentiable function in $\Re^m$ with ∇g Lipschitz continuous in $\Re^m$. X is a polyhedral set. In this case Assumption 1 holds with $\gamma = 1$.

(e) (Dual functional case; see Theorem 4.1 in [30]).

$$f(x) = \langle q, x\rangle + \max_{y\in Y}\{\langle Ex, y\rangle - g(y)\} \quad \forall x,$$

where Y is a polyhedral set in $\Re^m$, E is an $m \times n$ matrix with no zero column, q is a vector in $\Re^n$, and g is a strongly convex differentiable function in $\Re^m$ with ∇g Lipschitz continuous in $\Re^m$. X is a polyhedral set. In this case Assumption 1 holds with $\gamma = 1$.

Based on Assumptions 1 and 2 and using Theorem 4, we can establish the following rate of convergence results.

Theorem 5. *Let Assumptions 1 and 2 hold. Let $x^0, x^1, x^2, \ldots$ be any sequence which, together with some sequence of scalars $\{\alpha^k\}$ satisfying $\liminf_r \alpha^k > 0$ and some sequence $\{e^k\}$ in $\Re^n$, satisfies (19)–(21). Then*

(a) *If error bound exponent γ in (22) is equal to 1, then $\{f(x^k)\}$ converges at least **Q**-linearly and $\{x^k\}$ converges at least **R**-linearly to an element of S.*
(b) *If $\gamma > 1$, then $\{f(x^k)\}$ converges at least sublinearly at the rate $k^{1-\gamma}$.*

Proof. If $r(x^k) = 0$ for some $k = \bar{k}$, then $r(x^k) = 0$ for all $k \geq \bar{k}$ and the conclusions of the theorem hold trivially. So, in what follows, we assume $r(x^k) > 0$.

By (19), we have

$$x^{k+1} = [x^k - \alpha^k \nabla f(x^k) + e^k]^+_X$$

This together with the nonexpansive property of the projection operator $[\cdot]^+_X$ implies

$$\begin{aligned}\left\|x^k - [x^k - \alpha^k \nabla f(x^k)]^+_X\right\| &\leq \|x^k - x^{k+1}\| + \left\|x^{k+1} - [x^k - \alpha^k \nabla f(x^k)]^+_X\right\| \\ &\leq \|x^k - x^{k+1}\| + \|e^k\|.\end{aligned}$$

It is known that, for any $x \in X$ and d in $\Re^n$, $\|x - [x - \alpha d]^+_X\|/\alpha$ is monotonically decreasing with $\alpha > 0$ (see [10, Lemma 1]), so the left-hand side of the above relation is bounded below by $\min\{1, \alpha^k\}\|x^k - [x^k - \nabla f(x^k)]^+_X\|$. Also using (20) to bound the right-hand side of the above relation, we obtain

$$\underline{\alpha}\left\|x^k - [x^k - \nabla f(x^k)]^+_X\right\| \leq (1 + \kappa_1)\|x^k - x^{k+1}\|,$$

where $\underline{\alpha} = \min\{1, \liminf_k \alpha^k\} > 0$. By (21), we have $f(x^k) \leq f(x^0)$ for all k and $x^k - x^{k+1} \to 0$. The above relation implies $x^k - [x^k - \nabla f(x^k)]^+_X \to 0$ so, by Assumption 1, there is an index $\bar{k}$ and a scalar $\tau > 0$ such that, for all $k \geq \bar{k}$, (22) holds with $x = x^k$. This implies

$$\|x^k - \bar{x}^k\| \leq \tau\left\|x^k - [x^k - \nabla f(x^k)]^+_X\right\|^{1/\gamma} \leq \tau\left(\frac{1+\kappa_1}{\underline{\alpha}}\right)^{1/\gamma}\|x^k - x^{k+1}\|^{1/\gamma} \;\; \forall k \geq \bar{k}, \tag{23}$$

where $\bar{x}^k$ denotes an element of S for which $\|x^k - \bar{x}^k\| = \operatorname{dist}(x^k, S)$. Combining (23) with $x^k - x^{k+1} \to 0$ gives

$$x^k - \bar{x}^k \to 0, \tag{24}$$

so $\bar{x}^k - \bar{x}^{k+1} \to 0$. Then, Assumption 2 implies that $\bar{x}^k$ eventually settles down at some isocost surface of f, i.e., there exist an index $\hat{k} \geq \bar{k}$ and a scalar $\bar{\upsilon}$ such that

$$f(\bar{x}^k) = \bar{\upsilon} \quad \forall k \geq \hat{k}. \tag{25}$$

Fix any index $k \geq \hat{k}$. Since $x^k \in X$ and $\bar{x}^k$ is a stationary point of f over X, we have $\langle \nabla f(\bar{x}^k), x^k - \bar{x}^k \rangle \geq 0$ and from the Mean Value Theorem that $f(\bar{x}^k) - f(x^k) = \langle \nabla f(\psi^k), \bar{x}^k - x^k \rangle$, for some $\psi^k \in \Re^n$ lying on the line segment joining $\bar{x}^k$ with x^k. Upon summing these two relations and using (25), we obtain

$$\begin{aligned} \bar{\upsilon} - f(x^k) &\leq \langle \nabla f(\psi^k) - \nabla f(\bar{x}^k), \bar{x}^k - x^k \rangle \\ &\leq \|\nabla f(\psi^k) - \nabla f(\bar{x}^k)\| \|\bar{x}^k - x^k\| \\ &\leq L \|\bar{x}^k - x^k\|^2, \end{aligned}$$

where the last inequality follows from the Lipschitz condition (17) and $\|\psi^k - \bar{x}^k\| \leq \|x^k - \bar{x}^k\|$. This together with (24) yields

$$\liminf_{r \to \infty} f(x^k) \geq \bar{\upsilon}. \tag{26}$$

Fix any index $r \geq \hat{r}$. Since x^{k+1} is obtained by projecting $x^k - \alpha^k \nabla f(x^k) + e^k$ onto X (cf. (19)) and $\bar{x}^k \in X$, we have

$$\langle x^k - \alpha^k \nabla f(x^k) + e^k - x^{k+1}, x^{k+1} - \bar{x}^k \rangle \geq 0.$$

We also have from the Mean Value Theorem that

$$f(x^{k+1}) - f(\bar{x}^k) = \langle \nabla f(\xi^k), x^{k+1} - \bar{x}^k \rangle,$$

for some vector $\xi^k \in \Re^n$ lying on the line segment joining $\bar{x}^k$ with x^{k+1}. Combining these two relations and using (25), we obtain

$$\begin{aligned} f(x^{k+1}) - \bar{\upsilon} &= \langle \nabla f(\xi^k) - \nabla f(x^k), x^{k+1} - \bar{x}^k \rangle + \langle \nabla f(x^k), x^{k+1} - \bar{x}^k \rangle \\ &\leq \|\nabla f(\xi^k) - \nabla f(x^k)\| \|x^{k+1} - \bar{x}^k\| + \frac{1}{\alpha^k} \langle x^k - x^{k+1} + e^k, x^{k+1} - \bar{x}^k \rangle \\ &\leq \left(L \|\xi^k - x^k\| + \frac{1}{\underline{\alpha}} (\|x^k - x^{k+1}\| + \|e^k\|) \right) \|x^{k+1} - \bar{x}^k\|. \end{aligned}$$

Using the inequalities $\|\xi^k - x^k\| \leq \|x^{k+1} - x^k\| + \|\bar{x}^k - x^k\|$ and $\|x^{k+1} - \bar{x}^k\| \leq \|x^{k+1} - x^k\| + \|x^k - \bar{x}^k\|$ and (20), we deduce that there exists some constant κ_3 (depending on $L, \ \underline{\alpha}, \ \kappa_1, \ \tau$) such that

$$f(x^{k+1}) - \bar{\upsilon} \leq \kappa_3 \big(\|\bar{x}^k - x^k\| + \|x^k - x^{k+1}\| \big)^2.$$

By using the estimate (23) to bound the term $\|\bar{x}^k - x^k\|$, we can further simplify the above relation to obtain

$$f(x^{k+1}) - \bar{\upsilon} \leq \kappa_4 \|x^k - x^{k+1}\|^{2/\bar{\gamma}} \quad \forall k \geq \hat{k},$$

where $\bar{\gamma} = \max\{1, \gamma\}$ and $\kappa_4 > 0$ is a constant (depending on $\underline{\alpha}$, κ_1, κ_3). Combining this with (21) yields

$$f(x^{k+1}) - \bar{\upsilon} \leq \frac{\kappa_4}{(\kappa_2)^{1/\bar{\gamma}}}(f(x^k) - f(x^{k+1}))^{1/\bar{\gamma}} \quad \forall k \geq \hat{k}.$$

Let us cast this inequality into the form (10) by defining $r(x^k) := f(x^k) - \bar{\upsilon}$ so that Theorem 4 becomes applicable. By (26) and the fact $\{f(x^k)\}$ is a monotonically decreasing sequence, it follows that $r(x^k) \geq 0$ for all k, Rewriting the above inequality yields

$$r(x^{k+1}) \leq \frac{\kappa_4}{(\kappa_2)^{1/\bar{\gamma}}}(r(x^k) - r(x^{k+1}))^{1/\bar{\gamma}}, \quad \forall k \geq \hat{k},$$

which further implies

$$r(x^k) \leq \frac{\kappa_4}{(\kappa_2)^{1/\bar{\gamma}}}(r(x^k) - r(x^{k+1}))^{1/\bar{\gamma}} + r(x^k) - r(x^{k+1}), \quad \forall k \geq \hat{k}.$$

Since $\bar{\gamma} \geq 1$ and $r(x^k) - r(x^{k+1}) \downarrow 0$ (recall $x^k - x^{k+1} \to 0$), it follows

$$(r(x^k) - r(x^{k+1}))^{1/\bar{\gamma}} \geq r(x^k) - r(x^{k+1}).$$

Thus, we obtain from the above inequality that

$$r(x^k) \leq \left(1 + \frac{\kappa_4}{(\kappa_2)^{1/\bar{\gamma}}}\right)(r(x^k) - r(x^{k+1}))^{1/\bar{\gamma}}, \quad \forall k \geq \hat{k}.$$

Then, upon simplifying and rearranging terms, the above inequality becomes

$$r(x^{k+1}) - r(x^k) \leq -\left(\frac{(\kappa_2)^{1/\bar{\gamma}}}{(\kappa_2)^{1/\bar{\gamma}} + \kappa_4}\right)^{\bar{\gamma}} r(x^k)^{\bar{\gamma}}, \qquad \forall k \geq \hat{k}.$$

This shows that Assumption (A2) of Theorem 4 holds. Since $r(x^k) > 0$ for all k, Assumption (A1) of Theorem 4 also holds and we can conclude:

(a) If $\gamma = 1$, then $\bar{\gamma} = 1$ and $\{f(x^k)\}$ converges at least **Q**-linearly to $\bar{\upsilon}$. Since $\|x^k - x^{k+1}\|^2$ is of the order $f(x^k) - f(x^{k+1})$ (cf. (21)), this implies that $\{x^k\}$ converges at least **R**-linearly in this case. Since $\mathrm{dist}(x^k, S) \to 0$ (cf. (24)), then the point to which $\{x^k\}$ converges is an element of S.

(b) If $\gamma > 1$, then $\bar{\gamma} = \gamma > 1$. In this case, there exists some $\sigma > 0$ such that $r(x^k) \leq \sigma\, k^{1-\gamma}$, for all large k. Since $r(x^k) = f(x^k) - \bar{\upsilon}$, this further implies that the function sequence $\{f(x^k)\}$ converges to $\bar{\upsilon}$ sublinearly at the rate $k^{1-\gamma}$. By combining (21) with (23), we have

$$\mathrm{dist}(x^k, S) = O(\|x^k - x^{k+1}\|^{1/\gamma}) = O\big((f(x^k) - f(x^{k+1}))^{1/2\gamma}\big), \qquad \text{for large } k.$$

This shows that the sequence $\{\mathrm{dist}(x^k, S)\}$ converges to zero at the sublinear rate of $k^{(1-\gamma)/2\gamma}$.

This completes the proof of the theorem.

□

Theorem 5 extends the rate of convergence results reported in [25–27] for the case where X is a polyhedral set and for the case $\gamma = 1$. By the early discussions, the gradient projection method, the extra-gradient method, the proximal point minimization method and the matrix splitting method all satisfy the conditions (19)–(21). Therefore these methods all generate a sequence which converge to a stationary point, as long as the problem to be solved satisfies Assumptions 1 and 2. Of course, the rate of convergence depends on the error bound exponent γ. The aforementioned methods have been studied extensively. Unfortunately, the existing analysis typically requires some nondegeneracy assumption on the problem (such as uniqueness of the solution) which does not hold for many "real-world" problems or problems transformed in a way (e.g., by introducing artificial variables) so to bring about structures suitable for decomposition.

Notice that in the case $\gamma > 1$, Theorem 5 does not make any claim on the convergence of the iterate sequence $\{x^k\}$; it only states the (sublinear rate of) convergence of the function sequence $\{f(x^k)\}$ and the distance sequence $\{\mathrm{dist}(x^k, S)\}$. It remains to be seen if the iterate sequence converges or not in this case.

4. Concluding remarks

In this paper we have derived two new projection type error bounds and applied them the (rate of) convergence analysis of iterative descent algorithms. It has been shown that the error exponents of the associated projection type error bounds are closely related to the rate of convergence of functional and iterate sequences. In particular, we have shown that in general the functional sequence converges sublinearly at a rate $k^{-\epsilon}$, where k is the iteration index and ϵ is a positive constant related to the error bound exponent. Our convergence analysis is applicable to a wide class of well known iterative descent algorithms and to a fairly general class of constrained minimization problems (convex or non-convex).

In closing, we point out a possible direction to further extend the current work. Specifically, the error bound provided by Theorem 3 assumes that the objective function f is a γ-strictly convex function. It remains to be seen if this error bound result can be extended to the case where f has a composite structure

$$f(x) = \langle q, x\rangle + g(Ex), \quad \forall x,$$

where E is a matrix with no zero column, q is a vector, and g is a γ-strictly convex differentiable function with ∇g Lipschitz continuous. The above composite structure was considered by Luo and Tseng [25] where they established a local error bound with exponent $\gamma = 1$ under the assumption that g is strongly convex.

References

1. Auslender, A. (1976): Optimisation: Méthodes Numériques. Masson, Paris
2. Bertsekas, D.P. (1976): On the Goldstein–Levitin–Polyak gradient projection method. IEEE Trans. Auto. Contr. **21**, 174–184
3. Bertsekas, D.P., Gafni, E. (1982): Projection methods for variational inequalities with application to the traffic assignment problem. Math. Program. Study **12**, 139–159

4. Bertsekas, D.P., Tseng, P. (1994): Partial proximal minimization algorithms for convex programming. SIAM J. Optim. **4**, 551–572
5. Bertsekas, D.P., Tsitsiklis, J.N. (1989): Parallel and Distributed Computation: Numerical Methods. Prentice–Hall, Englewood Cliffs, NJ
6. Bierstone, E., Milman, P.D. (1988): Semianalytic and subanalytic sets. Institut des Hautes Etudes Scientifiques, Publications Mathématiques **67**, 5–42
7. D'Esopo, D. (1959): A convex programming procedure. Naval Logist. Quart. **6**, 33–42
8. Dedieu, J.P. (1992): Penalty functions in subanalytic optimization. Optimization **26**, 27–32
9. Dunn, J.C. (1981): Global and asymptotic convergence rate estimates for a class of projected gradient processes. SIAM J. Control Optim. **19**, 368–400
10. Gafni, E.M., Bertsekas, D.P. (1984): Two–metric projection methods for constrained optimization. SIAM J. Control Optim. **22**, 936–964
11. Goldstein, A.A. (1964): Convex programming in Hilbert space. Bull. Am. Math. Soc. **70**, 709–710
12. Güler, O. (1991): On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419
13. Hildreth, C. (1957): A quadratic programming procedure. Naval Res. Logist. Quart. **4**, 79–85 see also Erratum, Naval Res. Logist. Quart. **4**, 361
14. Iusem, A.N. (1991): On dual convergence and the rate of primal convergence of Bregman's convex programming method. SIAM J. Optim. **1**, 401–423
15. Iusem, A.N., Teboulle, M. (1995): Convergence rate analysis of nonquadratic proximal methods for convex and linear programming. Math. Oper. Res. **20**, 657–677
16. Korpelevich, G.M. (1976): The extragradient method for finding saddle points and other problems. Matecon **12**, 747–756
17. Kort, B.W., Bertsekas, D.P. (1976): Combined primal-dual and penalty methods for convex programming. SIAM J. Control Optim. **14**, 268–294
18. Levitin, E.S., Polyak, B.T. (1966): Convergence of minimizing sequences in conditional extremum problems. Soviet Math. Doklady **7**, 764–767
19. Li, W. (1993): Remarks on convergence of the matrix splitting algorithm for the symmetric linear complementarity problem. SIAM J. Optim. **3**, 155–164
20. Li, W. (1995): Error bounds for piecewise convex quadratic programs and applications. SIAM J. Control Optim. **33**, 1510–1529
21. Lojasiewicz, M.S. (1959): Sur le problème de la division. Studia Math. **18**, 87–136
22. Luenberger, D.G. (1984): Linear and Nonlinear Programming. Addison-Wesley, Reading, MA
23. Luo, X.D., Luo, Z.-Q. (1994): Extensions of Hoffman's error bound to polynomial systems. SIAM J. Optim. **4**, 383–392
24. Luo, Z.-Q., Pang, J.S. (1994): Error bounds for analytic systems and their applications. Math. Program. **67**, 1–28
25. Luo, Z.-Q., Tseng, P. (1992): On the linear convergence of descent methods for convex essentially smooth minimization. SIAM J. Control Optim. **30**, 408–425
26. Luo, Z.-Q., Tseng, P. (1992): On the convergence of coordinate descent method for convex differentiable minimization. J. Optim. Theory Appl. **72**, 7–35
27. Luo, Z.-Q., Tseng, P. (1992): Error bound and the convergence analysis of matrix splitting algorithms for the affine variational inequality problem. SIAM J. Optim. **2**, 43–54
28. Luo, Z.-Q., Tseng, P. (1993): Error bounds and the convergence analysis of feasible descent methods: a general approach. Ann. Oper. Res. **46**, 157–178
29. Luo, Z.-Q., Tseng, P. (1993): Error bound and reduced gradient projection algorithms for convex minimization over a polyhedral set. SIAM J. Optim. **3**, 43–59
30. Luo, Z.-Q., Tseng, P. (1993): On the convergence rate of dual ascent methods for linearly constrained convex minimization. Math. Oper. Res. **18**, 846–867
31. Luo, Z.-Q., Tseng, P. (1994): On the rate of convergence of a distributed asynchronous routing algorithm. IEEE Trans. Auto. Control **39**, 1123–1129
32. Luque, F.J. (1984): Asymptotic convergence analysis of the proximal point algorithm. SIAM J. Control Optim. **22**, 277–293
33. Mangasarian, O.L. (1977): Solution of symmetric linear complementarity problems by iterative methods. J. Optim. Theory Appl. **22**, 465–485
34. Martinet, B. (1970): Regularization d'inéquations variationelles par approximations successives. Rev. Française d'Auto. Inform. Rech. Opér. **4**, 154–159
35. Ortega, J.M., Rheinboldt, W.C. (1970): Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York
36. Pang, J.S. (1982): On the convergence of a basic iterative method for the implicit complementarity problem. J. Optim. Theory Appl. **37**, 149–162

37. Pang, J.S. (1987): A posteriori error bound for the linearly-constrained variational inequality problem. Math. Oper. Res. **12**, 474–484
38. Pang, J.S. (1997): Error bounds in mathemtical programming. Math. Program. **79**, 299–332
39. Tseng, P. (1991): On the rate of convergence of a partially asynchronous gradient projection algorithm. SIAM J. Optim. **1**, 603–319
40. Tseng, P. (1995): On linear convergence of iterative methods for the variational inequality problem. J. Comput. Appl. Math. **60**, 237–252
41. Walkup, D.W., Wets, R.J.B. (1969): A Lipschitzian characterization of convex polyhedra. Proceedings of the American Mathematical Society **20**, 167–173
42. Warga, J. (1963): Minimizing certain convex functions. J. Soc. Indust. Appl. Math. **11**, 588–593