Global Complexity Analysis of Inexact Successive Quadratic Approximation methods for Regularized Optimization under Mild Assumptions

Wei Peng · Hui Zhang · Xiaoya Zhang

Received: date / Accepted: date

Abstract Successive quadratic approximations (SQA) are numerically efficient for minimizing the sum of a smooth function and a convex function. The iteration complexity of inexact SQA methods has been analyzed recently. In this paper, we present an algorithmic framework of inexact SQA methods with four types of line searches, and analyze its global complexity under milder assumptions. First, we show its well-definedness and some decreasing properties. Second, under the quadratic growth condition and a uniform positive lower bound condition on stepsizes, we show that the function value sequence and the iterate sequence are linearly convergent. Moreover, we obtain a $o(1/k)$ complexity without the quadratic growth condition, improving existing $O(1/k)$ complexity results. At last, we show that a local gradient-Lipschitz-continuity condition could guarantee a uniform positive lower bound for the stepsizes.

Keywords Inexactness · Line search · Successive quadratic approximation · quadratic growth condition · Linear convergence

1 Introduction

A fundamental optimization model is ubiquitous in many fields such as machine learning, signal and image processing, and compressed sensing. Typi-
cally, the model is to minimize the sum of a smooth function $f$ and a convex regularizer $g$:

$$
\min_{x \in \mathcal{H}} F(x) := f(x) + g(x),
$$

(1)

where $\mathcal{H}$ is a real Hilbert space. In recent years, there has been a great deal of interest in developing algorithms to solve (1). A classic framework to solve it is the forward-backward splitting (FBS) method [8], which could be formulated as follows:

$$
x_{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau_k} \| x - x_k \|^2 + g(x) \right\}.
$$

(2)

Note that $p_k$ could be regarded as a second-order approximation to $f$ around $x_k$. Therefore, it is reasonable to further exploit the second-order information of $f$ like

$$
q_k(x) := f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k).
$$

Actually, this kind of approximation was used to design the proximal Newton method [12,16]. If $\nabla^2 f$ enjoys some special structure, the proximal Newton method can be very efficient, as shown in [9] for solving the $\ell_1$-regularized inverse covariance matrix estimation problem. However, in general cases and for large-scale problems, the storage and computation of $\nabla^2 f(x_k)$ could be prohibitive. Thus, one might pursue some approximations to $\nabla^2 f$. In practice, we usually seek positive definite matrices $H_k$ to approximate $\nabla^2 f(x_k)$. This motivates the proximal successive quadratic approximation (SQA) method:

$$
x_{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2}(x - x_k)^T H_k(x - x_k) + g(x) \right\}.
$$

(3)

The scheme above can be viewed as a generalization of the FBS and proximal Newton methods, as it reduces to them by setting $H_k = \frac{1}{\tau_k}I_n$ and $H_k = \nabla^2 f(x_k)$, respectively. In this sense, SQA is also known as the proximal quasi-Newton method or the variable metric forward-backward splitting method.

Numerically, one vital issue is how to solve the subproblems (3), whose closed-form solutions are usually hard to obtain. Therefore, iterative algorithms are needed to find inexact solutions to the subproblems. On this road, the authors of [6] proposed a proximal gradient-based inexactness condition for solving subproblems inexactly, along with a global convergence result. Later, the authors of [17] showed superlinear and $Q$-quadratic convergences (with respect to outer iterations) for a family of inexact SQA methods by a delicate parameter choosing strategy for the proximal gradient-based condition. It should be noted that they employed the Luo-Tseng error bound to replace the strong convexity near optimal points.
However, we observe that the inexactness condition might increase the inner iteration complexity as the outer iteration goes on, which is scarcely taken into account in previous works. This is the first motivation of our study.

Besides the gradient-based inexactness condition, which might make the number of inner iterations hard to estimate, some inexactness conditions based on the decrease of function values appeared. For example, the authors of [14] proposed an inexact SQA method such that the function values of $Q_k(·)$ (which are modifications of the objective functions in (3)) decrease to a given absolute error. With a proper parameter selection strategy and for arbitrary $\varepsilon > 0$, they showed that $O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$ inner iterations is enough to yield an approximate solution $\bar{x}^*$ satisfying

$$F(\bar{x}^*) - \inf F \leq \varepsilon.$$  \hspace{1cm} (4)

To put a uniform upper bound on the number of inner iterations, the authors of [10] presented another inexactness condition, which requires that the function values of $Q_k(·)$ decrease to some relative error. They showed linear convergence results under the optimal set strongly convexity (OSSC), which is weaker than the strong convexity assumption.

A common assumption made in these existing inexact SQA methods is that the gradient of the smooth part $f$ is Lipschitz continuous, which may fail in many problems. Therefore, it is natural to ask how inexact SQA methods behave without this assumption. This is the second motivation of our study. Before us, we note that a couple of exact gradient-type methods without this assumption have been investigated recently [2,4,3]. Among them, the author of [13] studied a class of exact SQA methods. The main tool employed in [13] is the quasi-Fejér monotone property and thus put a strong restriction on $\{H_k\}$. For inexact SQA methods, we find the line of thought in [13] is infeasible because the inexact solving of subproblems brings essential difficulties in analyzing iteration points. Besides, we want to drop the additional restriction on $\{H_k\}$. To this end, we develop new proof methods to study inexact SQA methods under weaker conditions than the gradient-Lipschitz-continuity assumption, and make the following contributions:

1. We present an algorithmic framework of inexact SQA methods with four types of backtracking line search strategies, and show its well-definedness without assuming the the gradient-Lipschitz-continuity property.
2. We relax OSSC [10] to the weaker quadratic growth condition to obtain $Q$-linear convergence of the function value sequence and $R$-linear convergence of the iterate sequence.
3. Without the quadratic growth property, we derive a $o(k^{-1})$ convergence of the function value sequence by non-trivially modifying [10, Lemma 6], which improves the $O(k^{-1})$ convergence in most existing related results. The author of [13] also presented a $o(k^{-1})$ convergence result but for exact SQA method and with stronger restrictions on $\{H_k\}_{k\geq 0}$. 
4. Finally, we show that a local gradient-Lipschitz-continuity condition could guarantee a uniform positive lower bound for stepsizes when performing backtracking line searches.

The paper is organized as follows. In Section 2, we introduce notations and assumptions. In Section 3, we present the algorithmic framework of inexact SQA methods with four types of backtracking line search strategies. In Section 4, we analyze the global complexity of the algorithmic framework under mild assumptions. In Section 5, we give a short summary of this paper, along with some discussion for future work.

2 Notations & Assumptions

For a nonempty closed set \( C \subset \mathbb{R}^n \), we denote the distance from \( x \) to \( C \) by \( \text{dist}(x,C) := \inf_{y \in C} \|x - y\| \). The domain of an extended-value function \( h : \mathbb{R}^n \to [-\infty, +\infty) \) is defined as \( \text{dom} h := \{x \in \mathbb{R}^n : h(x) < +\infty\} \). We say that \( h \) is proper if \( h(x) > -\infty \) for every \( x \) and \( \text{dom} h \neq \emptyset \). The gradient of a differentiable function \( f \) is denoted by \( \nabla f \). We say that \( \nabla f \) is \( L \)-Lipschitz continuous on a convex set \( S \) if
\[
\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \quad \forall x, y \in S.
\]
We say a sequence \( \{x_k\} \) \( R \)-linearly converges to \( x^* \) if
\[
\limsup_{k \to +\infty} \|x_k - x^*\|_k^{1/k} < 1.
\]

We modify the subproblem (3) into the following form:
\[
\min_{x \in H} Q_k(x) := \langle \nabla f(x_k), x - x_k \rangle + g(x) - g(x_k) + \frac{1}{2} \|x - x_k\|_k^2,
\]
and let \( Q_k^* \) be the minimum. The notation \( \| \cdot \|_k \) will be explained below. We denote by \( \bar{x}_{k+1} \) the \( \eta \)-approximate minimizer, which satisfies
\[
Q_k(\bar{x}_{k+1}) \leq \eta(Q_k^* - Q_k(x_k)) \quad \text{for given } \eta \in (0, 1].
\]

Now, we list main assumptions involved in this paper as follows:

A0 The set of minimizers of (1) is nonempty, denoted by \( X \). The objective function \( F \) attains its minimum \( F^* = \inf_{x \in X} F(x) \in \mathbb{R} \).

A1 \( f, g : H \to (-\infty, \infty] \) are two proper lower semi-continuous (lsc) convex functions with \( \text{dom} g \subset \text{int}(\text{dom} f) \) and thus \( \text{dom} F = \text{dom} g \).

A2 \( f \) is Fréchet differentiable on an open set containing \( \text{dom} g \). Its gradient \( \nabla f \) is continuous on \( \text{dom} g \).

A3 \( \{\langle \cdot, \cdot \rangle_k\} \) is a sequence of inner products on \( H \), with induced norms \( \{\| \cdot \|_k\} \) and associated positive operators \( \{H_k\} \), i.e.,
\[
\forall k \geq 0, \quad H_k : H \to H, \quad \langle \cdot, \cdot \rangle_k = \langle \cdot, H_k \cdot \rangle.
\]
There exist positive constants \( M, m \) such that
\[
\forall k \geq 0, \quad m\| \cdot \|_k^2 \leq \| \cdot \|_k^2 \leq M\| \cdot \|_k^2.
\]
A4 There exists an linearly convergent algorithm for the subproblem (6) with a uniform parameter $\sigma$ for all $k \geq 0$ such that

$$Q_k(y^{(k)}) - Q_k^* \leq -(1 - \sigma)^l Q_k^*, \forall l \geq 0$$

where $0 < \sigma < 1, y^{(k)}_0 := x_k$.

The assumptions A0 and A1 are standard. The assumption A2 is weaker than the standard assumption that supposes $\nabla f$ to be Lipschitz continuous. We do not assume any special structures on $H_k$ in A3. Note that the auxiliary function $Q_k$ is a regularized strongly convex function. The standard proximal gradient method, as shown in [15, Theorem 2.1], could satisfy A4. This assumption is used to guarantee that each subproblem could be solved to satisfy the inexactness condition in a fixed number of iterations. Hence, the complexity of solving subproblems could not increase as $k \to +\infty$.

3 The algorithm

3.1 Line Search

For a directional line search method, we fix the direction $\bar{x}_{k+1} - x_k$, along which we search for a stepsize as large as possible. The main advantage of this kind of line search compared to [10 Algorithm 2] is that we only need to find an $\eta$-approximate minimizer of (8) in each iteration. And then we determine the next iterate by $x_{k+1} := x_k + \alpha_k(\bar{x}_{k+1} - x_k)$. Below, we give several line search strategies to determine $\alpha_k$. For simplicity, we define

$$\Delta_k(x) := \langle \nabla f(x), x - x_k \rangle + g(x) - g(x_k).$$

LS1. Let $\beta, \gamma, \in (0, 1), \bar{\alpha} \in (0, 1]$ and $\forall k \geq 0$,

$$\alpha_k = \max \{ \alpha > 0 \mid \exists i \geq 0, \alpha = \bar{\alpha}\beta^i, F(x_k + \alpha(\bar{x}_{k+1} - x_k)) - F(x_k) \leq \gamma \alpha \{ \langle \nabla f(x_k), \bar{x}_{k+1} - x_k \rangle + g(\bar{x}_{k+1}) - g(x_k) \} \}.$$  \hfill (10)

LS2. Let $\bar{\beta} \in (0, 1), \gamma \in (0, m/2), \bar{\alpha} \in (0, 1]$ and $\forall k \geq 0$,

$$\alpha_k = \max \{ \alpha > 0 \mid \exists i \geq 0, \alpha = \bar{\alpha}\beta^i, \nabla f(x_k + \alpha(\bar{x}_{k+1} - x_k)) - \nabla f(x_k) \| \leq \gamma \| x_{k+1} - x_k \| \}.$$ \hfill (11)

LS3. Let $\beta, \gamma, \in (0, 1), \bar{\alpha} \in (0, 1]$ and $\forall k \geq 0$,

$$\alpha_k = \max \{ \alpha > 0 \mid \exists i \geq 0, \alpha = \bar{\alpha}\beta^i, F(x_k + \alpha(\bar{x}_{k+1} - x_k)) - F(x_k) \leq \alpha \{ \langle \nabla f(x_k), \bar{x}_{k+1} - x_k \rangle + g(\bar{x}_{k+1}) - g(x_k) + \frac{\gamma}{2} \| \bar{x}_{k+1} - x_k \|_2^2 \} \}.$$ \hfill (12)

LS4. Let $\beta, \gamma, \in (0, 1), \bar{\alpha} \in (0, 1]$ and $\forall k \geq 0$,

$$\alpha_k = \max \{ \alpha > 0 \mid \exists i \geq 0, \alpha = \bar{\alpha}\beta^i, f(x_k + \alpha(\bar{x}_{k+1} - x_k)) - f(x_k) \leq \alpha \{ \langle \nabla f(x_k), \bar{x}_{k+1} - x_k \rangle + \frac{\gamma}{2} \| \bar{x}_{k+1} - x_k \|_2^2 \} \}.$$ \hfill (13)
3.2 Algorithmic Framework

Now, we present the promised algorithmic framework of inexact SQA methods.

Algorithm 1: Inexact Successive Quadratic Approximation with Line-search

Initialization:
Given initial iterate $x_0 \in H$, $\eta \in (0, 1]$;
Choose Line-search($\cdot$) from LS1-4 with proper parameters;
for $i = 0, 1, 2, \cdots$ do
    Choose a symmetric $H_k$;
    Solving the subproblem inexactly:
    $\bar{x}_{k+1}$ := Linesearch$(x_k, \bar{x}_{k+1})$;
    $x_{k+1} := x_k + \alpha_k (\bar{x}_{k+1} - x_k)$
end

First of all, we state that the algorithmic framework is well defined. Its proof can be found in Appendix.

Lemma 1 The stepsize $\alpha_k$ of LS1-4 exists.

The result above indicates that LS1-4 could find $\alpha_k$ by initializing $\alpha := \bar{\alpha}$ and updating $\alpha := \beta \alpha$ in finite algorithmic steps.

Next, we show a sufficient decrease property of the algorithmic framework. Actually, we can derive that

$$Q_k(\bar{x}_{k+1}) = g(\bar{x}_{k+1}) - g(x_k) + \langle \nabla f(x_k), \bar{x}_{k+1} - x_k \rangle + \frac{1}{2} \| x_k - \bar{x}_{k+1} \|^2_k$$

(14)

$$\leq \eta \left( g(J_k) - g(x_k) + \langle \nabla f(x_k), J_k - x_k \rangle + \frac{1}{2} \| x_k - J_k \|^2_k \right)$$

$$\leq \eta \left( g(y) - g(x_k) + \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2} \| x_k - y \|^2_k - \frac{1}{2} \| y - J_k \|^2_k \right)$$

$$= \eta \left( g(x_k + \lambda(x - x_k)) - g(x_k) + \langle \nabla f(x_k), \lambda(x - x_k) \rangle \right.$$

$$\left. + \frac{1}{2} \| \lambda(x - x_k) \|^2_k - \frac{1}{2} \| x_k + \lambda(x - x_k) - J_k \|^2_k \right)$$

$$\leq \eta \left( \lambda(F(x) - F(x_k)) + \frac{\lambda^2}{2} \| x - x_k \|^2_k \right)$$

(15)

for any $\lambda \in [0, 1], x \in H$ and $y := x_k + \lambda(x - x_k)$, where the first inequality is due to that $\bar{x}_{k+1}$ is an $\eta$-approximate minimizer satisfying (7), the second inequality follows from the strong convexity of $Q_k(\cdot)$, and the last inequality from the convexity of $F$. With this deduction, we have the following two results, whose proofs can be found in Appendix.

Lemma 2 With the proper parameters selected in LS1-4, we have

(i) $F(x_{k+1}) - F(x_k) \leq \gamma \alpha_k Q_k(\bar{x}_{k+1})$ for LS1 and
(ii) $F(x_{k+1}) - F(x_k) \leq \alpha_k Q_k(\bar{x}_{k+1})$ for LS2-4.
Lemma 3 For LSI-4, we have the sufficient decreasing property for all $k \geq 0$:
\begin{equation}
F(x_{k+1}) - F(x_k) \leq -\alpha_k c_1 \|x_{k+1} - x_k\|^2_k,
\end{equation}
where $c_1$ is some positive constant. Thus, $\{F(x_k)\}$ is monotone decreasing and $\sum_{k=0}^{+\infty} \alpha_k \|x_{k+1} - x_k\|^2 < +\infty$.

4 Complexity Analysis

In this section, we will analyze the global complexity of the proposed algorithmic framework under mild assumptions. All proofs can be found in Appendix.

4.1 Linear Convergence Results

In this subsection, we focus on convergence analysis under the quadratic growth condition. First, we introduce the optimal set strongly convexity condition (OSSC), which is presented in [10] to get linear convergence. We say that a function $F$ satisfies OSSC if there exists $\mu > 0$ such that for any $x \in \text{dom } F$ and any $\lambda \in [0,1]$, it holds
\begin{equation}
F(\lambda x + (1 - \lambda)P_X(x)) \leq \lambda F(x) + (1 - \lambda)F^* - \frac{\mu \lambda (1 - \lambda)}{2} \|x - P_X(x)\|^2,
\end{equation}
where $P_X(x) := \arg\min_{y \in X} \|x - y\|$. Note that $X$ is nonempty, convex and closed and hence $P_X(x)$ is well-defined[7, Theorem 1.2.3].

Below, we recall the quadratic growth condition.

Definition 1 ([19][18]) We say the function $F$ satisfies the $\mu$-quadratic growth (QG) condition if there exists $\mu > 0$ such that
\begin{equation}
F(x) - F^* \geq \frac{\mu}{2} \|x - P_X(x)\|^2, \forall x \in \text{dom } F.
\end{equation}
Here, we claim that QG is strictly weaker than OSSC. For example, consider the function
\begin{equation}
F(x) = \begin{cases}
|x| & \text{if } |x| < 1 \\
x^2 & \text{else}
\end{cases},
\end{equation}
where $F(x) - F(0) \geq x^2$. It satisfies QG but not OSSC. Moreover, we observe that OSSC is sufficient for a nonsmooth extension of quasi strongly convexity, which is strictly stronger than QG; for details please refer to [11].

Now, we present the main result of this part.

Theorem 1 If $F$ satisfies the $\mu$-quadratic growth condition and $\inf_{k \geq 0} \alpha_k \geq \alpha$ for some $\alpha > 0$. Then

(i) The function value sequence $\{F(x_k)\}$ is Q-linearly convergent to $F^*$.
(ii) The iterate sequence \( \{x_k\} \) R-linearly converges to an optimal point \( x^* \).

Though the inexactness condition (7) is hard to verify, we could use a fixed number of iterations \( N_{\text{inner}} \). With the assumption A4, \( N_{\text{inner}} \) iterations achieve (7) with \( \eta = 1 - (1 - \sigma)^{N_{\text{inner}}} \). Then we immediately have the following corollary.

**Corollary 1** Assume that the conditions in Theorem 1 holds. Fixing inner iteration number \( N_{\text{inner}} \) of the algorithm that satisfies A4, Algorithm 1 attains a solver \( \tilde{x} \) such that

\[
F(\tilde{x}) - F^* \leq \varepsilon
\]

with \( O(\log(1/\varepsilon)) \) inner iterations in total.

### 4.2 Sublinear Convergence Results

In this subsection, we drop the QG assumption of \( F \). In order to illustrate the convergence of \( \{F(x_k)\} \), we modify lemma in [10] as follows

**Lemma 4** Assume we have three non-negative sequences \( \{\delta_k\}_{k \geq 0}, \{\lambda_k\}_{k \geq 0}, \{A_k\} \) and a positive constant \( c \in (0, 1] \) such that

\[
\delta_{k+1} \leq \delta_k + c \left( -\lambda_k \delta_k + \frac{A_k}{2} \lambda_k^2 \right), \quad \forall k \geq 0, \lambda_k \in [0, 1].
\]

(i) If \( A_k \leq \bar{A} \) for \( k \geq 0 \), where \( \bar{A} \) is a positive constant, we have

\[
\delta_k \sim O(k^{-1}). \quad (18)
\]

(ii) If \( \lim_{k \to 0} A_k = 0 \), we have

\[
\delta_k \sim o(k^{-1}). \quad (19)
\]

The following result improves the existing convergence rate of \( \{F(x_k)\}_{k \geq 0} \) from \( O(k^{-1}) \) to \( o(k^{-1}) \). Denote

\[
R_0 := \sup_{x: F(x) \leq F(x_0)} \|x - P_{\mathcal{X}}(x)\|
\]

and assume \( R_0 \) to be finite. Since \( F(x_k) \leq F(x_0) \) holds for \( k \geq 0 \), we have

\[
\text{dist}(x_k, \mathcal{X}) \leq R_0, \quad (k \geq 0). \quad (20)
\]

**Theorem 2** Suppose there exists \( \underline{a} > 0 \) such that \( \alpha_k \geq \underline{a} > 0 \) for all \( k \in \mathbb{N} \) and \( R_0 \) is finite. Then, \( \{F(x_k)\} \) converges to \( F^* \) sublinearly in the sense that

\[
F(x_k) - F^* \sim O(k^{-1}).
\]

Furthermore, if \( \text{dist}(x_k, \mathcal{X}) \to 0 \), then

\[
F(x_k) - F^* \sim o(k^{-1}).
\]
If the statement $\forall \{x_k\}, F(x_k) \downarrow F^* \Rightarrow \text{dist}(x_k, \mathcal{X}) \to 0$ holds, the condition of the second conclusion in Theorem 2 will be automatically satisfied. Unfortunately, it is not true in general as the following counterexample illustrates:

Consider the function $F : \mathbb{R}^2 \to (-\infty, \infty]$ satisfying lsc.,

$$F(x, y) = \begin{cases} x + \sqrt{x^2 + y^2} & \text{if } x + y^2 \leq 1, \\ \infty & \text{otherwise}. \end{cases} \quad (21)$$

$F$ is convex since it is the sum of two convex functions. The optimal set is $\{(x, 0) \in \mathbb{R}^2 | x \leq 0\}$ and the minimum is 0. Consider the sequence $\{z_k\} \subset \text{dom} F$ where for every $k \geq 0$, $z_k = (-\sum_{i=0}^{k} 1/(i+1), 1)$. It is obvious that $F(z_k) \downarrow 0$ but $\text{dist}(z_k, \mathcal{X}) \equiv 1$.

Below, we propose several mild conditions, under which $F(x_k) \downarrow 0$ implies $\text{dist}(x_k, \mathcal{X}) \to 0$.

**Proposition 1** If one of the following statements holds, then, any $\{x_k\} \subset \text{dom} F$ satisfying $F(x_k) \downarrow F^*$ implies $\text{dist}(x_k, \mathcal{X}) \to 0$.

(i) The level set $C_0 := \{x \in \mathcal{H} | F(x) \leq F(x_0)\}$ is compact.

(ii) $F$ is defined on $\mathbb{R}^n$ and its lineality space is equal to its recession cone (see definitions in [5]), i.e.,

$$L_F = R_F.$$

(iii) In particular, $F$ defined on $\mathbb{R}^n$ is level bounded, which implies

$$L_F = R_F = \emptyset.$$

Note that for a globally $L$-smooth function $f$, there exists a positive number $\alpha$ such that $\liminf_{k \to +\infty} \alpha_k = 0$. Therefore, using Theorem 2, we have a slightly stronger convergence rate compared with [10, Theorem 3].

Similar to Corollary 1, we could use a fixed number of inner iterations in practice for general convex cases as well.

**Corollary 2** Suppose there exists $\alpha > 0$ such that $\alpha_k \geq \alpha > 0$ for all $k \in \mathbb{N}$ and $A_4$ holds. With a fixed number of inner iterations replacing the stopping criterion [4], Algorithm 5 attains a solver $\tilde{x}$ satisfying

$$F(\tilde{x}) - F^* \leq \varepsilon$$

with $O(\varepsilon^{-1})$ inner iterations in total. Furthermore, the number of iterations is reduced to $o(\varepsilon^{-1})$ if $F$ is level bounded.
4.3 Lower Bound for Stepsizes

In this subsection, under a local gradient-Lipschitz-continuity condition, we prove that the stepsizes have a uniform positive lower bound, which guarantees that complexity of the line searches do not increase.

**Proposition 2** If \( \text{dist}(x_k, \mathcal{X}) \to 0 \), \( F(x_k) \downarrow F^* \) and \( \nabla f \) is \( L \)-Lipschitz continuous on \( \mathbb{B}_\varepsilon(\mathcal{X}) \cap \text{dom } F \) with \( L > 0, \varepsilon > 0 \) where
\[
\mathbb{B}_\varepsilon(\mathcal{X}) := \{ x \in \mathcal{H} | \text{dist}(x, \mathcal{X}) \leq \varepsilon \}.
\]
(22)

Denote \( d_k := \bar{x}_{k+1} - x_k \). Then
\[
\lim_{k \to +\infty} \| d_k \| = 0.
\]

Equipped with the result above, the following lemma illustrates that stepsizes must have a uniform positive lower bound.

**Theorem 3** Under the same conditions with Proposition 2, we have

(i) for LS1
\[
\liminf_{k \to +\infty} \alpha_k \geq \min \left\{ 1, \beta(1 - \gamma) \frac{m(\eta + 1 + \sqrt{1 - \eta})}{L(1 + \sqrt{1 - \eta})} \right\}.
\]
(23)

(ii) For LS2
\[
\liminf_{k \to +\infty} \alpha_k \geq \min \left\{ 1, \frac{\beta \gamma}{L} \right\}.
\]
(24)

(iii) For LS3 and LS4
\[
\liminf_{k \to +\infty} \alpha_k \geq \min \left\{ 1, \frac{\beta \gamma m}{L} \right\}.
\]
(25)

Now, from Theorems 1-3 we can conclude that once the function value and iterate sequences are convergent, they must (sub)linearly converge under the local gradient-Lipschitz-continuity condition.

5 Conclusion & Future Work

In this paper, we study the global complexity of an algorithmic framework of inexact SQA methods with four types of line search strategies under mild assumptions. On one hand, with the QG property and the uniform positive lower bound condition on stepsizes, we derive the Q-linear convergence of the function value sequence and the R-linear convergence of the iterate sequence. On the other hand, without the QG property, we obtain the \( o(k^{-1}) \) complexity, which improves existing results. Finally, we give a uniform positive lower bound of the stepsizes for LS1-4 with the local gradient-Lipschitz-continuity assumption.

We believe that the new analysis developed in this paper might be extended to other related algorithms, such as inexact Bregman-type methods. We leave it as future work.
References

1. Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, second edition, 2017. With a foreword by Hedy Attouch.

2. José Yunier Bello Cruz and W. de Oliveira. On weak and strong convergence of the projected gradient method for convex optimization in real Hilbert spaces. *Numer. Funct. Anal. Optim.*, 37(2):129–144, 2016.

3. José Yunier Bello Cruz, G. Li, and Tran T. A. Nghia. On the q-linear convergence of forward backward splitting method and uniqueness of optimal solution to lasso. *arXiv preprint arXiv:1806.06333*, 2018.

4. José Yunier Bello Cruz and Tran T. A. Nghia. On the convergence of the forward-backward splitting method with linesearches. *Optim. Methods Softw.*, 31(6):1209–1238, 2016.

5. Dimitri P. Bertsekas. *Convex optimization theory*. Athena Scientific, Nashua, NH, 2009.

6. Richard H. Byrd, Jorge Nocedal, and Figen Oztoprak. An inexact successive quadratic approximation method for \( \ell_1 \) regularized optimization. *Math. Program.*, 157(2, Ser. B):375–396, 2016.

7. Andrzej Cegielski. *Iterative methods for fixed point problems in Hilbert spaces*, volume 2057 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2012.

8. Patrick L. Combettes and Valérie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200, 2005.

9. Cho-Jui Hsieh, Inderjit S Dhillon, Pradeep K Ravikumar, and Mátyás A Sustik. Sparse inverse covariance matrix estimation using quadratic approximation. In *Advances in neural information processing systems*, pages 2330–2338, 2011.

10. Ching-pei Lee and Stephen J Wright. Inexact successive quadratic approximation for regularized optimization. *arXiv preprint arXiv:1803.01298*, 2018.

11. Ion Necoara, Yu Nesterov, and François Glineur. Linear convergence of first order methods for non-strongly convex optimization. *Math. Program.*, pages 1–39, 2018.

12. Lijun Qi and Xiaojun Chen. A preconditioning proximal Newton method for non-differentiable convex optimization. *Math. Program.*, 76(3, Ser. B):411–429, 1997.

13. Saverio Salzo. The variable metric forward-backward splitting algorithm under mild differentiability assumptions. *SIAM J. Optim.*, 27(4):2153–2181, 2017.

14. Katya Scheinberg and Xiaocheng Tang. Practical inexact proximal quasi-Newton method with global complexity analysis. *Math. Program.*, 160(1-2, Ser. A):495–529, 2016.

15. Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Exact Worst-Case Convergence Rates of the Proximal Gradient Method for Composite Convex Minimization. *J. Optim. Theory Appl.*, 178(2):455–476, 2018.

16. Z. Wei and L. Qi. Convergence analysis of a proximal Newton method. *Numer. Funct. Anal. Optim.*, 17(3-4):453–472, 1996.

17. Man-Chung Yue, Zirui Zhou, and Anthony Man-Chak So. A family of inexact sqa methods for non-smooth convex minimization with provable convergence guarantees based on the luo-tseng error bound property. *Math. Program.*, pages 1–32, 2016.

18. Hui Zhang. New analysis of linear convergence of gradient-type methods via unifying error bound conditions. *arXiv preprint arXiv:1606.00269v7*, 2016.

19. T. Zolezzi. On equiwellset minimum problems. *Appl. Math. Optim.*, 4(3):209–223, 1977/78.

Appendix

A. Proof of Lemma

It is easy to see that if \( x_k = \bar{x}_{k+1} \), then the statement trivially holds. So we consider \( x_k \neq \bar{x}_{k+1} \). If \( x_k \in \mathcal{X} \), we have \( Q^*_k = Q_k(x_k) \), which implies that
$Q_k(\bar{x}_{k+1}) = Q_k(x_k)$. Due to the strong convexity of $Q_k(\cdot)$, it follows that
$\bar{x}_{k+1} = x_k$. Therefore, we only need to consider $x_k \notin \mathcal{X}$, which implies $Q_k^* < 0$
and hence $\Delta_k(\bar{x}_{k+1}) < 0$.

**LS1:** By contradiction suppose that for all $\alpha \in \mathcal{Q} := \{\alpha, \bar{\alpha}, \bar{\alpha}^2, \cdots \}$,

$$
\frac{F(x_k + \alpha(\bar{x}_{k+1} - x_k)) - F(x_k)}{\alpha} > \gamma \Delta_k(\bar{x}_{k+1}).
$$

With

$$
\alpha(g(\bar{x}_{k+1}) - g(x_k)) \geq g(x_k + \alpha(\bar{x}_{k+1} - x_k)) - g(x_k),
$$

it follows that

$$
\frac{f(x_k + \alpha(\bar{x}_{k+1} - x_k)) - f(x_k)}{\alpha} + g(\bar{x}_{k+1}) - g(x_k) > \gamma \Delta_k(\bar{x}_{k+1}).
$$

Taking $\alpha \downarrow 0$, due to $f$ is Fréchet differentiable at $x_k$, we obtain

$$
\Delta_k(\bar{x}_{k+1}) \geq \gamma \Delta_k(\bar{x}_{k+1}),
$$

a contradiction with $\Delta_k(\bar{x}_{k+1}) < 0$.

**LS2:** By contradiction suppose that for all $\alpha \in \mathcal{Q} := \{\bar{\alpha}, \bar{\alpha} \bar{\beta}, \bar{\alpha}^2 \bar{\beta}, \cdots \}$,

$$
\|\nabla f(x_k + \alpha(\bar{x}_{k+1} - x_k)) - \nabla f(x_k)\| > \gamma \|\bar{x}_{k+1} - x_k\|.
$$

Taking $\alpha \downarrow 0$, we have $\nabla f(x_k + \alpha(\bar{x}_{k+1} - x_k)) \to \nabla f(x_k)$. Then the continuity
of $\nabla f$ at $x_k$ yields the contradiction $0 \geq \gamma \|\bar{x}_{k+1} - x_k\|$.

**LS3:** By contradiction suppose that for all $\alpha \in \mathcal{Q} := \{\bar{\alpha}, \bar{\alpha} \bar{\beta}, \bar{\alpha}^2 \bar{\beta}, \cdots \}$,

$$
F(x_k + \alpha(\bar{x}_{k+1} - x_k)) - F(x_k) > \alpha \left(\Delta_k(\bar{x}_{k+1}) + \frac{\gamma}{2} \|\bar{x}_{k+1} - x_k\|^2\right).
$$

Using (26), dividing both sides by $\alpha$ and then taking $\alpha \downarrow 0$, due to that $f$
is Fréchet differentiable at $x_k$, we obtain

$$
\Delta_k(\bar{x}_{k+1}) \geq \Delta_k(\bar{x}_{k+1}) + \frac{\gamma}{2} \|\bar{x}_{k+1} - x_k\|^2,
$$

a contradiction with $\bar{x}_{k+1} - x_k \neq 0$.

**LS4:** By contradiction suppose that for all $\alpha \in \mathcal{Q} := \{\bar{\alpha}, \bar{\alpha} \bar{\beta}, \bar{\alpha}^2 \bar{\beta}, \cdots \}$,

$$
f(x_k, x \in \mathcal{X}) \geq \alpha \left(\nabla f(x_k) \cdot \bar{x}_{k+1} - x_k\right) + \frac{\gamma}{2} \|\bar{x}_{k+1} - x_k\|^2.
$$

Dividing both sides by $\alpha$ and then taking $\alpha \downarrow 0$, we obtain

$$
\langle \nabla f(x_k), \bar{x}_{k+1} - x_k\rangle \geq \langle \nabla f(x_k), \bar{x}_{k+1} - x_k\rangle + \frac{\gamma}{2} \|\bar{x}_{k+1} - x_k\|^2,
$$

a contradiction with $\bar{x}_{k+1} - x_k \neq 0$. 

B. Proof of Lemma 2

**LS1:** Combining (10) with (14), we obtain
\[
\frac{1}{\alpha_k} (F(x_{k+1}) - F(x_k)) + \frac{1}{2} \|\bar{x}_{k+1} - x_k\|^2_k \leq Q_k(\bar{x}_{k+1}),
\]
which implies the statement(i).

**LS2:** Due to the convexity of \( f \) and (11), we have
\[
f(x_{k+1}) - f(x_k) \leq \langle \nabla f(x_{k+1}), x_{k+1} - x_k \rangle
= \langle \nabla f(x_{k+1}) - \nabla f(x_k), x_{k+1} - x_k \rangle + \langle \nabla f(x_k), x_{k+1} - x_k \rangle
\leq \gamma \|x_{k+1} - x_k\|^2 + \langle \nabla f(x_k), x_{k+1} - x_k \rangle
\leq \gamma \alpha_k \|\bar{x}_{k+1} - x_k\|_k^2 + \langle \nabla f(x_k), x_{k+1} - x_k \rangle
= \alpha_k \langle \nabla f(x_k), \bar{x}_{k+1} - x_k \rangle + \gamma \alpha_k \|\bar{x}_{k+1} - x_k\|^2_k.
\]
Using the convexity of \( g \) with \( \alpha_k \in (0, 1] \), we have
\[
g(x_{k+1}) - g(x_k) \leq \alpha_k (g(\bar{x}_{k+1}) - g(x_k)),
\]
Adding (28) to (27) and then dividing \( \alpha_k \) on both sides of the resulted inequality, we obtain
\[
\frac{1}{\alpha_k} (F(x_{k+1}) - F(x_k)) + \left(\frac{1}{2} - \gamma \frac{1}{m}\right) \|\bar{x}_{k+1} - x_k\|^2_k \leq Q_k(\bar{x}_{k+1}).
\]
Since we select \( \gamma < m/2 \), the statement(ii) for LS2 is proved.

**LS3:** Combining (12) with (14), we obtain
\[
\frac{1}{\alpha_k} (F(x_{k+1}) - F(x_k)) + \frac{1}{2} \|\bar{x}_{k+1} - x_k\|^2_k \leq Q_k(\bar{x}_{k+1}).
\]
Since \( \gamma \in (0, 1) \), discarding the second term, then we proved the statement(ii) for LS3.

**LS4:** Combining (13) and (14), we have
\[
f(x_{k+1}) - f(x_k) + \alpha_k (g(\bar{x}_{k+1}) - g(x_k)) \leq \alpha_k Q_k(\bar{x}_{k+1}) + \alpha_k \left(\frac{\gamma - 1}{2}\right) \|x_k - \bar{x}_{k+1}\|^2_k.
\]
Adding (28) to the inequality above and then dividing \( \alpha_k \) on both sides of the resulted inequality, we obtain
\[
\frac{1}{\alpha_k} (F(x_{k+1}) - F(x_k)) + \left(\frac{1}{2} - \gamma \frac{1}{m}\right) \|\bar{x}_{k+1} - x_k\|^2_k \leq Q_k(\bar{x}_{k+1}).
\]
Since \( \gamma \in (0, 1) \), the statement(ii) is proved.
C. Proof of Lemma 3

Due to that $\bar{x}_{k+1}$ is an $\eta$-approximate minimizer, we have

$$Q_k(\bar{x}_{k+1}) \leq \eta Q_k(J_k) \leq \eta Q_k(x_k + \lambda(\bar{x}_{k+1} - x_k)) \quad (\forall \lambda \in [0, 1])$$

where the last inequality follows from the strong convexity of $Q_k$:

$$Q_k(x_k + \lambda(\bar{x}_{k+1} - x_k)) \leq \lambda Q_k(x_k) + (1-\lambda)Q_k(\bar{x}_{k+1}) - \frac{\lambda(1-\lambda)}{2}\|\bar{x}_{k+1} - x_k\|^2_k$$

and the fact $Q_k(x_k) = 0$. The inequality (30) leads to

$$Q_k(\bar{x}_{k+1}) \leq -\eta \lambda (1-\lambda) \frac{1}{2} \|\bar{x}_{k+1} - x_k\|^2_k.$$  

Setting $\lambda := \frac{1}{1+\sqrt{1-\eta}}$, which lies on $[1/2, 1]$, we have:

$$Q_k(\bar{x}_{k+1}) \leq -\eta \frac{2}{1+\sqrt{1-\eta}} \|\bar{x}_{k+1} - x_k\|^2_k. \quad (31)$$

Revoking Lemma 2 we obtain the following sufficient descent properties:

$$\textbf{LS1} : F(x_{k+1}) - F(x_k) \leq -\frac{\alpha_k \gamma \eta}{2(1+\sqrt{1-\eta})} \|\bar{x}_{k+1} - x_k\|^2_k. \quad (32)$$

$$\textbf{LS2-4} : F(x_{k+1}) - F(x_k) \leq -\frac{\alpha_k \eta}{2(1+\sqrt{1-\eta})} \|\bar{x}_{k+1} - x_k\|^2_k. \quad (33)$$

Therefore, $\{F(x_k)\}$ is monotone decreasing and

$$\forall k \geq 0, F(x_{k+1}) - F(x_k) \leq -\alpha_k c_1 \|\bar{x}_{k+1} - x_k\|^2_k \quad (34)$$

for some positive constant $c_1$. Summing up (34) for all $k \geq 0$, we have

$$\sum_{k=0}^{+\infty} \alpha_k \|\bar{x}_{k+1} - x_k\|^2_k \leq F(x_0) - \lim_{k \to \infty} F(x_k) \leq F(x_0) - F^* < \infty.$$  

D. Proof of Lemma 4

If $\delta_T = 0$ for some $T$, then $\delta_k = 0$ holds for $k \geq T$. Hence we assume $\delta_k \neq 0$ without loss of generality. The statement (i) is immediately obtained from [10, Lemma 6]. We consider the statement (ii). Since $A_k$ has a limit, then $A_k$ is upper bounded and thus (i) holds so that $\delta_k \to 0$.

Since

$$\delta_{k+1} \leq \begin{cases} (1-c)\delta_k + \frac{\delta_k}{2} & \text{if } \delta_k > A_k \\ \delta_k - \frac{\delta_k^2}{2} & \text{otherwise,} \end{cases}$$

we have

...
and note that $\delta_k$ is monotone decreasing. With a slight abuse of notation that let $c/0 = +\infty$ when $A_k = 0$, dividing $\delta_k \delta_{k+1}$ on both sides, then we obtain

$$\frac{1}{\delta_{k+1}} \geq \min \left\{ \frac{1}{\delta_k} + \frac{c}{\delta_k} \frac{1}{2 - c}, \frac{1}{\delta_k} + \frac{c}{2A_k} \right\},$$

which implies

$$\frac{1}{\delta_{k+1}} \geq \frac{1}{\delta_k} + z_k$$

with $z_k = \min \left\{ \frac{c}{\delta_k (2 - c)}, \frac{c}{2A_k} \right\}$. Since $z_k \to +\infty$, we immediately have

$$\delta_k \leq \frac{1}{\delta_0 + \sum_{i=0}^{k-1} z_i} \sim o \left( \frac{1}{k} \right).$$

E. Proof of Proposition 1

(i) Define a level set sequence $\{C_k\}$ associated with $\{l_k\}$, i.e., $\forall k \geq 0$

$$C_k := \{ x \in \mathcal{H} | F(x) \leq l_k \},$$

where we set $l_k := F(x_k) (\forall k \geq 0)$ and thus $l_k \downarrow F^*$. We will illustrate that $F(x_k) \to F^*$ implies $\text{dist}(x, \mathcal{X}) \to 0$ by contradiction. Assume that there exist a subsequence $\{x_{k_i}\}$ and $D > 0$ such that $\text{dist}(x_{k_i}, \mathcal{X}) \geq D$ for every $i \geq 0$. The sequence $\{x_{k_i}\}$ is in the compact set $C_0$. Thus without any loss of generality, we assume $x_{k_i} \to x^* \subset C_0$. Then we have $\text{dist}(x^*, \mathcal{X}) \geq D > 0$.

$C_k$ are closed due to $F$ is lsc. Because of the closedness of $C_k$ and $C_k \subset C_0$, each $C_k$ is compact, so is $\mathcal{X}$.

$$\forall \varepsilon > 0, \exists I \geq 0, \forall i \geq I, \text{we have} \|x_{k_i} - x^*\| < \varepsilon/2.$$ Then we have $B_{\varepsilon/2}(x^*) \cap C_{k_i} \neq \emptyset$ due to $x_{k_i} \in C_{k_i}$. Using $\forall k \geq 0$ $C_{k+1} \subset C_k$, then we have $\forall k \geq 0$, $B_{\varepsilon/2}(x^*) \cap C_k \neq \emptyset$. Denote $E_k := B_{\varepsilon/2}(x^*) \cap C_k$, then $E_k$ is compact due to compactness of $C_k$ and closedness of $B_{\varepsilon/2}(x^*)$. Via $\forall k \geq 0$ $E_{k+1} \subset E_k$, we have $\cap_{k=0}^{\infty} E_k \neq \emptyset$, which leads to

$$B_{\varepsilon/2}(x^*) \cap \mathcal{X} = B_{\varepsilon/2}(x^*) \cap (\cap_{k=0}^{\infty} C_k) = \cap_{k=0}^{\infty} E_k \neq \emptyset.$$ Therefore, $x^*$ is in the closure of $\mathcal{X}$. Note that $\mathcal{X}$ is compact and hence closed, we have $x^* \in \mathcal{X}$, which contradicts $\text{dist}(x^*, \mathcal{X}) \geq D$.

(ii) Let $C_k$ be defined as above. Then $C_k = L_F + \tilde{C}_k$, where $\tilde{C}_k \subset L_F^\perp$ and $\tilde{C}_k$ is compact [1, Proposition 1.4.11]. Each $x_k$ could be uniquely decomposed as $x_k = y_k + P_{L_F^\perp}(x_k)$ where $y_k \in L_F, P_{L_F^\perp}(x_k) \in \tilde{C}_k$. Define $\tilde{F} : \mathbb{R}^n \to (-\infty, \infty]$ as

$$\tilde{F}(x) = \begin{cases} F(x) & \text{if } x \in L_F^\perp \\ \infty & \text{otherwise.} \end{cases}$$

Note that $\forall x \in \mathbb{R}^n$, we have $F(x) = F(P_{L_F^\perp}(x)) = \tilde{F}(P_{L_F^\perp}(x))$. It is easy to show that $\tilde{F}$ is convex and lsc. Its minimum is $\inf_{x \in \mathbb{R}^n} \tilde{F}(x) = F^*$ and the
optimal set is $\tilde{x} = P_{L_F}(X)$. Consider the sequence $\{z_k\}$ where $z_k := P_{L_F}(x_k)$. Then

$$\tilde{F}(z_k) = F(x_k) \downarrow \min_{x \in \mathbb{R}^n} \tilde{F}(x).$$

(35)

Note that the set $\{x \in \mathbb{R}^n | \tilde{F}(x) \leq \tilde{F}(z_0)\} = \{x \in L_F^+|F(x) \leq F(x_0)\} = \tilde{C}_0$ is compact. Using (i) and (35), we have $\text{dist}(z_k, \tilde{x}) \to 0$. Thus,

$$\text{dist}(x_k, \mathcal{X}) = \text{dist}(x_k, \tilde{x} + L_F) = \text{dist}(z_k, \tilde{x}) \to 0$$

which shows the statement (ii).

F. Proof of Proposition 2

Note that $Q_k(x_k) = 0$ and $Q_k(\bar{x}_{k+1}) \leq Q_k(x_k)$, then we have $Q_k(\bar{x}_{k+1}) \leq 0$, i.e.,

$$\langle \nabla f(x_k), d_k \rangle + g(x_k + d_k) - g(x_k) + \frac{1}{2}\|d_k\|^2 \leq 0 \quad (\forall k \geq 0).$$

Since $\|d_k\| = n\|\cdot\|$, the inequality above immediately leads to

$$\langle \nabla f(x_k), d_k \rangle + g(x_k + d_k) - g(x_k) + \frac{m}{2}\|d_k\|^2 \leq 0 \quad (\forall k \geq 0).$$

(36)

Suppose that there exists a subsequence $\{d_k\}$ and a positive number $D$ such that $\|d_k\| > D > 0$ for all $i \geq 0$. For an arbitrary positive number $\delta$ which satisfies $0 < \delta < \min\{\frac{\epsilon}{2}, D\}$, since $\text{dist}(x_k, \mathcal{X}) \to 0$, we have $\text{dist}(x_k, \mathcal{X}) < \delta$ for all large $i$. Denote $\omega_k := x_k + \delta d_k / \|d_k\|$. Then

$$\text{dist}(\omega_k, \mathcal{X}) = \|\omega_k - P_\mathcal{X}(\omega_k)\| = \left\|x_k + \delta \cdot \frac{d_k}{\|d_k\|} - P_\mathcal{X}(x_k)\right\|$$

$$\leq \delta + \|x_k - P_\mathcal{X}(x_k)\| = \delta + \text{dist}(x_k, \mathcal{X})$$

$$< \delta + \text{dist}(x_k, \mathcal{X}) \leq \epsilon,$$

which implies $\omega_k \in \mathbb{B}_\epsilon(\mathcal{X})$ for all large $i$. Also note that $\omega_k$ is on the line segment $[x_k, x_k + d_k] \subset \text{dom } f$, thus $\omega_k \in \mathbb{B}_\epsilon(\mathcal{X}) \cap \text{dom } F$. The L-Lipschitz continuity of $\nabla f$ on $\mathbb{B}_\epsilon(\mathcal{X}) \cap \text{dom } F$ implies

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathbb{B}_\epsilon(\mathcal{X}) \cap \text{dom } F.$$ 

(37)

Then using [11] Lemma 2.64(i)], we have

$$f(\omega_k) - f(x_k) \leq \frac{\delta}{\|d_k\|} \langle \nabla f(x_k), d_k \rangle + \frac{L\delta^2}{2}.$$

(38)

Due to the convexity of $g$ and $\delta / \|d_k\| < 1$, we have

$$g(\omega_k) \leq \frac{\delta}{\|d_k\|}g(x_k + d_k) + \left(1 - \frac{\delta}{\|d_k\|}\right)g(x_k),$$
a simple transformation of which yields
\[ g(\omega_{k_i}) - g(x_{k_i}) \leq \frac{\delta}{\|d_{k_i}\|} (g(x_{k_i} + d_{k_i}) - g(x_{k_i})). \] (39)

Combining (38) and (39), for all large \( i \), we have
\[ F(\omega_{k_i}) - F(x_{k_i}) \leq \frac{\delta}{\|d_{k_i}\|} \left( (\nabla f(x_{k_i}), d_{k_i}) + g(x_{k_i} + d_{k_i}) - g(x_{k_i}) \right) + \frac{L\delta^2}{2}. \]

Then employing (36) leads to
\[ F(\omega_{k_i}) - F(x_{k_i}) \leq -\frac{\delta}{\|d_{k_i}\|} \frac{m\|d_{k_i}\|^2}{2} + \frac{L\delta^2}{2} \leq \frac{L}{2} \delta^2 - \frac{\delta}{2} mD. \] (40)

Since \( F(x_k) \downarrow F^* \), taking the limit inferior on both sides of (40), we have
\[ \liminf_{i \to +\infty} F(\omega_{k_i}) - F^* \leq -\frac{L}{2} \delta^2 - \frac{\delta}{2} mD. \]
By setting \( \delta := \min\{\varepsilon/4, mD/(2L), D/2\} \), we obtain
\[ \liminf_{i \to +\infty} F(\omega_{k_i}) < F^*, \]
a contradiction with \( F^* \) being the minimum. Therefore, we have \( \|d_k\| \to 0 \).

G. Proof of Theorem 1

By setting \( x := P_{\mathcal{X}}(x_k) \) in (15), we have
\[ Q_k(\bar{x}_{k+1}) \leq \eta \left( -\lambda(F(x_k) - F^*) + \frac{\lambda^2}{2} \|x_k - P_{\mathcal{X}}(x_k)\|^2 \right) \]
\[ \leq \eta \left( -\lambda(F(x_k) - F^*) + \frac{M\lambda^2}{\mu} (F(x_k) - F^*) \right). \] (41)

The second inequality is due to the \( \mu \)-quadratic growth condition of \( F \). By setting \( \lambda := \min\{\mu/2M, 1\} \), we have
\[ Q_k(\bar{x}_{k+1}) \leq -\zeta (F(x_k) - F^*), \] (42)
where \( \zeta \) is a constant in \( (0, 1) \) satisfying
\[ \zeta = \begin{cases} \eta\mu/(4M), & \text{if } \mu \leq 2M, \\ 1 - M/\mu, & \text{else}. \end{cases} \]

Note that \( \alpha_k \geq \alpha \), using Lemma 2(i), together with (42) then we obtain
\[ \forall k \geq 0, \quad F(x_{k+1}) - F^* \leq (1 - (\alpha_k^{-}) (F(x_k) - F^*) \]
for LS1. Similarly, for LS2-4 we have
\[
\forall k \geq 0, \quad F(x_{k+1}) - F^* \leq (1 - \zeta \Omega) (F(x_k) - F^*).
\]
Therefore, \(\{F(x_k)\}\) is Q-linearly convergent to \(F^*\):
\[
F(x_k) - F^* \leq c^k (F(x_0) - F^*),
\]
where \(c\) is a constant belonging to \((0, 1)\).

We now prove that \(\{x_k\}\) is R-linearly convergent. Using Lemma 3, we have
\[
\frac{c_1 \|x_{k+1} - x_k\|^2}{\bar{\alpha}} \leq \alpha_k c_1 \|\bar{x}_{k+1} - x_k\|^2 \leq F(x_k) - F(x_{k+1}) \leq F(x_k) - F^* \leq c^k (F(x_0) - F^*).\]
Thus, \(\{\|x_{k+1} - x_k\|\}\) is linearly convergent satisfying
\[
\forall k \geq 0, \quad \|x_k - x_{k+1}\| \leq c \sqrt{\frac{\bar{\alpha}(F_0 - F^*)}{c_1}},
\]
which implies that \(\{x_k\}\) is a Cauchy sequence. By supposing \(x_k \to x^*\), we have
\[
\|x_k - x^*\| \leq \sum_{i=k}^{+\infty} \|x_i - x_{i+1}\| \\
\leq \sqrt{\frac{\bar{\alpha}(F_0 - F^*)}{c_1}} \sum_{i=k}^{+\infty} c_i^2 \\
= \sqrt{\frac{\bar{\alpha}(F_0 - F^*)}{c_1}} \frac{c_k^2}{1 - c_k^2}.
\]
This is just the R-linear convergence of the iterate sequence \(\{x_k\}\) and hence the proof is completed.

H. Proof of Theorem 2

Using (11), we have
\[
Q_k(\bar{x}_{k+1}) \leq \eta \left( -\lambda (F(x_k) - F^*) + \frac{\lambda^2 M}{2} \text{dist}(x_k, \mathcal{X})^2 \right). \quad (43)
\]
For LS1, combining (43) and Lemma 21, we have
\[
F(x_{k+1}) - F(x_k) \leq \alpha_k \gamma \eta \left( \lambda (F^* - F(x_k)) + \frac{\lambda^2 M \text{dist}(x_k, \mathcal{X})^2}{2} \right), \forall k \geq 0.
\]
Let \( \delta := F(x_k) - F^* \), \( c_k := \alpha_k \gamma (1 - \eta) \) and \( \hat{A} := MR_0^2 \) in Lemma 4; then we obtain
\[
F(x_k) - F^* \leq \frac{MR_0^2 + F(x_0) - F^*}{\sum_{i=0}^{k-1} a_i \gamma \eta} \sim O(k^{-1}).
\]
Via the identical routine, similar results can also be obtained for LS2-4:
\[
F(x_k) - F^* \leq \frac{MR_0^2 + F(x_0) - F^*}{\sum_{i=0}^{k-1} a_i \eta} \sim O(k^{-1}).
\]
Next, we will show \( o(1/k) \) convergence in the function value sequence. Since \( \text{dist}(x_k, X) \to 0 \), using (44) and Lemma 4(ii), then for LS1-4, we have
\[
F(x_k) - F^* \sim o\left(\frac{1}{k}\right).
\]
The proof is completed.

I. Proof of Theorem 3
Since \( \|d_k\| \to 0 \) and \( \text{dist}(x_k, X) \to 0 \), for all sufficiently large \( k \), \( \{x_k\} \) and \( \{x_k + d_k\} \) will eventually fall into \( B_{\varepsilon}(X) \cap \text{dom} F \). According to the \( L \)-Lipschitz continuity of \( \nabla f \), we have
\[
\|\nabla f(x_k + \alpha d_k) - \nabla f(x_k)\| \leq L \|\alpha d_k\|, \quad \forall \alpha \in [0, 1],
\]
which also implies
\[
f(x_k + \alpha d_k) - f(x_k) \leq \alpha \langle \nabla f(x_k), d_k \rangle + \frac{L \alpha^2}{2} \|d_k\|^2, \quad \forall \alpha \in [0, 1]. \tag{45}
\]
(i) Adding (28) (replacing \( \alpha_k \) by \( \alpha \)) to (45), for sufficiently large \( k \) we have
\[
F(x_k + \alpha d_k) - F(x_k) \leq \alpha \Delta_k(\bar{x}_{k+1}) + \frac{\alpha^2 L}{2} \|d_k\|^2. \tag{46}
\]
From (43), according to the relationship
\[
Q_k(\cdot) = \Delta_k(\cdot) + \frac{1}{2} \cdot \|\cdot\|_k^2 \geq \Delta_k(\cdot) + \frac{m}{2} \cdot \|\cdot\|^2, \tag{47}
\]
we have
\[
\Delta_k(\bar{x}_{k+1}) \leq -\left( \frac{mn}{2(1 + \sqrt{1 - \eta})} + \frac{m}{2} \right) \|d_k\|^2. \tag{48}
\]
Combining (46) and (48), canceling the term \( \|d_k\|^2 \), we have
\[
F(x_k + \alpha d_k) - F(x_k) \leq \left( \alpha - \frac{\alpha^2 L (1 + \sqrt{1 - \eta})}{m(\eta + 1 + \sqrt{1 - \eta})} \right) \Delta_k(\bar{x}_{k+1}). \tag{49}
\]
When $\alpha \leq \min\{1, (1-\gamma)\frac{m(\eta+1+\sqrt{1-\eta})}{2(1+\sqrt{1-\eta})}\}$, the stopping criterion in (10) must hold. Hence (i) is proved.

(ii) If $\alpha \leq \min\{1, \gamma/L\}$, then for all sufficiently large $k$,
\[\alpha\|\nabla f(x_k + \alpha d_k) - \nabla f(x_k)\| \leq \alpha L\|\alpha d_k\| \leq \gamma\|x_{k+1} - x_k\|.
\]
which satisfies the stopping criterion in (11) and thus (24) holds.

(iii) Using (46), if $\alpha \leq \min\{1, \frac{\gamma m}{L}\}$, the stopping criterion in (12) holds. The case of LS4 can be proved similarly.