Abstract

In this work, we develop a level-set subdifferential error bound condition aiming towards convergence rate analysis of a variable Bregman proximal gradient (VBPG) method for a broad class of nonsmooth and nonconvex optimization problems. It is proved that the aforementioned condition guarantees linear convergence of VBPG, and is weaker than Kurdyka-Łojasiewicz property, weak metric subregularity and Bregman proximal error bound. Along the way, we are able to derive a number of verifiable conditions for level-set subdifferential error bounds to hold, and necessary conditions and sufficient conditions for linear convergence relative to a level set for nonsmooth and nonconvex optimization problems. The newly established results not only enable us to show that any accumulation point of the sequence generated by VBPG is at least a critical point of the limiting subdifferential or even a critical point of the proximal subdifferential with a fixed Bregman function in each iteration, but also provide a fresh perspective that allows us to explore inner-connections among many known sufficient conditions for linear convergence of various first-order methods.

Keywords: Level-set subdifferential error bound, Variable Bregman proximal gradient method, Linear convergence, Bregman proximal error bound, Metric subregularity, Weak metric-subregularity, Linear convergence relative to a level set
1 Introduction

This paper studies the following nonconvex and nonsmooth optimization problem:

\[
(P) \quad \min_{x \in \mathbb{R}^n} \quad F(x) = f(x) + g(x) \tag{1}
\]

where \( f : \mathbb{R}^n \to (-\infty, \infty) \) is a proper lower semi-continuous (l.s.c) function that is smooth in \( \text{dom} f \), and \( g : \mathbb{R}^n \to (-\infty, \infty) \) is a proper l.s.c function. We say that \( (P) \) is a convex problem (a fully nonconvex problem) if both \( f \) and \( g \) are convex (both \( f \) and \( g \) are nonconvex).

Problem \( (P) \) arises naturally in diverse areas such as compressed sensing \([10, 16]\), machine learning and statistics \([36]\), principal component analysis \([27]\) and principal component pursuit \([2]\). Typically these problems are of large scale. As the number of decision variables is huge, first-order methods and their enhanced versions are viewed to be a practical way to solve \( (P) \) \([13, 24, 30]\).

By incorporating a Newton-like approach, we propose to solve \( (P) \) by a variable Bregman proximal gradient (VBPG) method first introduced in \([13]\) with the name of Auxiliary Problem Principal (APP) method. An iteration of the method takes the form:

\[
(A_P_k) \quad x^{k+1} \in \arg\min_{x \in \mathbb{R}^n} \left\{ \langle \nabla f(x^k), x - x^k \rangle + g(x) + \frac{1}{\epsilon_k} D^k(x^k, x) \right\}, \tag{2}
\]

where \( D^k \) is a variable Bregman distance (see Section 2.1 for the definition of a Bregman distance). A nonsmooth version is investigated in \([14]\). The classical proximal gradient (PG) method is \( D^k(x, y) = \frac{1}{2} \| x - y \|^2 \). The second-order information through \( D^k \) can be used to enhance the rate of convergence of the method \([7, 12]\). Some other choices of \( D^k \) can be found in \([4]\). Moreover, the VBPG method can be combined with extrapolation, proximal alternating linearization and line search process \([41]\). VBPG can also be viewed as a forward-backward splitting method to find a critical point of \( (P) \): \( x^{k+1} = (\Gamma^k + A)^{-1}(\Gamma^k - B)x^k \) with \( A = \nabla f(x), B = \partial P g(x), \Gamma^k = \nabla K(x)/\epsilon^k \) (for notation of \( K(x), \epsilon^k \) and \( \partial P g(x) \), see Section 2 for details).

Theory of error bounds (EB) has long been known playing an important role in optimization theory \([28, 35]\), and a central role in convergence analysis and convergence rate analysis of various iterative methods \([32]\). As we are interested in finding an optimal solution, or a critical point, or an optimal value for \( (P) \), it is natural to look at the following types of error bounds: the first type EB is an inequality that bounds the distance from a set of test points to a target set (e.g., critical-point set of \( (P) \), optimal solution set of \( (P) \), or a level set of \( F \)) by a residual function; while the second type EB is an inequality that bounds certain absolute values of the difference between function \( F \) values at a set of test points and a target value (e.g., a critical value of \( F \), or the optimal value of \( (P) \)) by a residual function. Prominent examples of first type error bounds include \([8, 9, 15, 20, 34]\). Pioneering contributions to second type error bounds include \([33]\) and Łojasiewicz inequality \([25]\).

When \( (P) \) is a convex problem, PG methods exhibit sublinear convergence rates \([7, 30]\) and achieve a linear
convergence rate if is strongly convex. Without strong convexity, examines sufficient conditions for linear convergence of PG and acceleration techniques.

Recently there is a surge of interest in developing some first type error bound (EB) conditions that guarantee linear convergence for PG methods, and in applying a generalized second type EB is (Kurdyka-Łojasiewicz (KL) property) to obtain linear convergence of PG methods as well as a variety of other optimization methods. proposes a proximal-PL inequality that leads to an elegant linear convergence rate analysis for sequences of function-values generated by the PG method. We remark that the proximal-PL inequality condition combines and extends an idea originated from metric functions for variational inequalities (VI) by reformulating a VI as a constrained continuous differentiable optimization problem through certain gap functions, see. In addition, there are two major lines of research on error bound conditions to achieve linear convergence guarantee for gradient descent methods. The first line of research is to find connections among existing error bound conditions. Examples of such work include. Another line of research is the study of (P) when (P) is fully nonconvex. A sample of such work can be found in.

Motivated by the aforementioned works for a quest for linear convergence of PG methods, we are led to ask the following basic question: What are fundamental properties associated with itself so that linear convergence of VBPG is guaranteed? This question leads us to look into error bounds involving level sets, subdifferentials and various level-set error bounds. A significant departure of our work to the above cited works is the use of level sets as target sets to establish error bound conditions whereas the above cited works typically use optimal solution sets or sets of critical points (in the nonconvex case) as target sets to establish error bound conditions. In this work, we have discovered a number of interesting results on level sets of , revealed the roles of level-set based error bounds in achieving linear convergence of VBPG, and uncovered interconnections among level-set based error bounds and other known error bounds in the literature.

This is a simplified version of the manuscript entitled “An variational approach on level sets and linear convergence of variable Bregman proximal gradient method for nonconvex optimization problems”. By introducing and examining a level-set subdifferential error bound condition carefully, we are able to derive linear convergence of VBPG under this condition. Some interesting features of this condition are as follows:

(i) In the fully nonconvex setting (i.e., both and are nonconvex), this condition is sufficient for and linear convergence of generated by VBPG. Moreover, all known sufficient conditions for linear convergence of PG methods imply this condition;

(ii) The level-set subdifferential EB condition along with associated theorems provides a unique perspective that allows us to make connections with many known conditions in the literature which are shown to guarantee the linear convergence of PG.

In addition to the above contributions, we also provide necessary conditions and sufficient conditions for linear convergence with respect to level sets for VBPG. By examples of Subsection we have shown
that the notion of the level-set subdifferential EB condition is weaker than that of the KL property, that of weak metric subregularity, and that of Bregman proximal error bound. To our knowledge, this is the first comprehensive work on convergence rate analysis of VBPG. Moreover, a number of new results obtained in this work for VBPG are also new results even for PG methods.

The rest of this paper is structured as follows. Section 2 provides notation and preliminaries. Section 3 presents the results on convergence and linear convergence analysis for VBPG. Section 4 introduces level-set analysis, studies level-set type error bounds, and provides the necessary and sufficient condition of linear convergence of VBPG under level-set based error bounds. Section 5 investigates connections of various level-set error bounds established in this work with existing error bounds. Section 6 lists known sufficient conditions to guarantee the existence of level-set subdifferential error bounds. Finally, we supply Figure 1 in Section 5 and Figure 2 in Section 6 to aid the reader to see easily inner relationships of these conditions and results.

2 Notations and preliminaries

Throughout this paper, \( \langle \cdot, \cdot \rangle \) and \( \| \cdot \| \) denote the Euclidean scalar product of \( \mathbb{R}^n \) and its corresponding norm respectively. Let \( C \) be a subset of \( \mathbb{R}^n \) and \( x \) be any point in \( \mathbb{R}^n \). Define
\[
\text{dist}(x, C) = \inf \{ \| x - z \| : z \in C \}.
\]
When \( C = \emptyset \), we set \( \text{dist}(x, C) = \infty \).

The definitions we will use throughout the paper are standard in variational analysis (\cite{35} and \cite{28}).

**Definition 2.1** (\cite{35}). Let \( \psi : \mathbb{R}^n \to \mathbb{R} \cup \{ +\infty \} \) be a proper lsc function.

(i) For each \( \bar{x} \in \text{dom} \psi \), the Fréchet subdifferential of \( \psi \) at \( \bar{x} \), written \( \partial_F \psi(\bar{x}) \), is the set of vectors \( \xi \in \mathbb{R}^n \), which satisfy
\[
\liminf_{x \to \bar{x}} \frac{1}{\| x - \bar{x} \|} [\psi(x) - \psi(\bar{x}) - \langle \xi, x - \bar{x} \rangle] \geq 0.
\]
If \( x \notin \text{dom} \psi \), then \( \partial_F \psi(x) = \emptyset \).

(ii) The limiting-subdifferential (\cite{28}), or simply the subdifferential for short, of \( \psi \) at \( \bar{x} \in \text{dom} \psi \), written \( \partial_L \psi(\bar{x}) \), is defined as follows:
\[
\partial_L \psi(\bar{x}) := \{ \xi \in \mathbb{R}^n : \exists x_n \to \bar{x}, \psi(x_n) \to \psi(\bar{x}), \xi_n \in \partial_F \psi(x_n) \to \xi \}.
\]

(iii) The proximal subdifferential of \( \psi \) at \( \bar{x} \in \text{dom} \psi \) written \( \partial_P \psi(\bar{x}) \), is defined as follows:
\[
\partial_P \psi(\bar{x}) := \{ \xi \in \mathbb{R}^n : \exists \rho > 0, \eta > 0 \text{ s.t. } \psi(x) \geq \psi(\bar{x}) + \langle \xi, x - \bar{x} \rangle - \rho \| x - \bar{x} \|^2 \forall x \in B(\bar{x}; \eta) \},
\]
where \( B(\bar{x}; \eta) \) is the open ball of radius \( \eta > 0 \), centered at \( \bar{x} \).
Definition 2.2 (Definition 13.27 of [35]). Let $\psi : \mathbb{R}^n \to (-\infty, \infty]$ be a proper lsc function.

(i) (Definition 13.27 of [35]) A lsc function $\psi$ is said to be prox-regular at $\bar{x} \in \text{dom} \, \psi$ for subgradient $\bar{v} \in \partial_L \psi(\bar{x})$, if there exist parameters $\eta > 0$ and $\rho \geq 0$ such that for every point $(x, \nu) \in \text{gph} \partial_L \psi$ obeying $\|x - \bar{x}\| < \eta$, $|\psi(x) - \psi(\bar{x})| < \eta$ and $\|\nu - \bar{v}\| < \eta$ and $\nu \in \partial_L \psi(x)$, one has

$$\psi(x') \geq \psi(x) + \langle \nu, x' - x \rangle - \frac{\rho}{2} \|x' - x\|^2 \quad \forall \ x' \in B(\bar{x}; \eta).$$

(ii) (Proposition 3.3 of [35]) A lsc function $\psi$ is said to be uniformly prox-regular around $\bar{x} \in \text{dom} \, \psi$, if there exist parameters $\eta > 0$ and $\rho \geq 0$ such that for every point $x, x' \in B(\bar{x}; \eta)$ and $\nu \in \partial_L \psi(x)$, one has

$$\psi(x') \geq \psi(x) + \langle \nu, x' - x \rangle - \frac{\rho}{2} \|x' - x\|^2.$$

(iii) (Definition 10 of [6]) A lsc function $\psi$ is semi-convex on $\text{dom} \, \psi$ with modulus $\rho > 0$ if there exists a convex function $h : \mathbb{R}^n \to \mathbb{R}$ such that $\psi = h(x) - \frac{\rho}{2} \|x\|^2$.

The following inclusions always hold: $\partial_P \psi(x) \subset \partial_L \psi(x) \subset \partial_L \psi(x)$. If $\psi$ is uniformly prox-regular around $\bar{x}$ on $B(\bar{x}; \eta)$ with $\eta > 0$, we have $\partial_P \psi(x) = \partial_L \psi(x)$ for all $x \in B(\bar{x}; \eta)$. In particular, $\partial_P \psi(x) = \partial_L \psi(x)$ if $\psi$ is a semi-convex (convex) function.

Throughout the rest of this paper, we make the following assumption on $f$ and $g$.

Assumption 1. (i) $f : \text{dom} \to (-\infty, \infty]$ is a differentiable function with $\text{dom} \, f$ convex and with gradient $L$-Lipschitz continuous.

(ii) $g$ is proper lower semicontinuous on $\text{dom} \, g$, and $\text{dom} \, g$ is a convex set.

(iii) $F$ is level-bounded i.e., the set $\{x \in \mathbb{R}^n : F(x) \leq r\}$ is bounded (possibly empty) for every $r \in \mathbb{R}$.

A few remarks about Assumption 1 are in order. By Theorem 3.2.12 of [31], the following descent property of $f$ holds

$$\frac{L}{2} \|y - x\|^2 + \langle \nabla f(x), y - x \rangle \geq f(y) - f(x) \quad \forall \ x, y \in \text{dom} \, f.$$ 

From (i) and (ii), $\text{dom} \, F$ is a convex set. As a consequence of (iii), the optimal value $F^*$ of $(P)$ is finite and the optimal solution set $X^*$ of $(P)$ is non-empty.

A vector $x$ satisfying $0 \in \partial_P F(x)$ is called a proximal critical point. The set of all proximal critical points of $F$ is denoted by $\tilde{X}_P$. By Assumption 1, $\tilde{X}_P \neq \emptyset$. The limiting critical point is defined as:

$$\bar{X}_L := \{x : 0 \in \nabla f(x) + \partial_L g(x)\}.$$

In general $\tilde{X}_P \subseteq \bar{X}_L$ and the equality holds if $\partial_P g(x) = \partial_L g(x)$. By Proposition 2.3 of [38], $\partial_P F(x) = \nabla f(x) + \partial_P g(x)$.
2.1 Variable Bregman distance, Bregman type mappings and functions

Let a sequence of functions \( \{K^k, k \in \mathbb{N}\} \) and positive numbers \( \{\epsilon^k, k \in \mathbb{N}\} \) be given, where the function \( K^k \) is strongly convex and differentiable with Lipschitz gradient. For each \( k \), define a variable Bregman distance

\[
D^k(x, y) = K^k(y) - [K^k(x) + \langle \nabla K^k(x), y - x \rangle].
\]  

(3)

The variable Bregman distance \( D^k \) measures the proximity between two points \((x, y)\); that is, \( D^k(x, y) \geq 0 \) and \( D^k(x, y) = 0 \) if and only if \( x = y \). We make the following standing assumption on the functions \( K^k(x) \).

Assumption 2. (i) For each \( k \), \( K^k \) is strongly convex with uniformly modulus \( m \) and with its gradient \( \nabla K^k \) being uniformly \( M \)-Lipschitz.

(ii) The parameter \( \epsilon^k \) satisfies: \( 0 < \underline{\epsilon} \leq \epsilon^k \leq \overline{\epsilon} \).

Under Assumption 2, \( \{D^k | k \in N\} \) uniformly satisfies:

\[
\begin{align*}
    m ||x - y||^2 &\leq \left\langle \nabla_x D^k(x, y), x - y \right\rangle \leq M ||x - y||^2, \\
    m ||x - y||^2 &\leq \left\langle \nabla_y D^k(x, y), y - x \right\rangle \leq M ||x - y||^2, \\
    \frac{m}{2} ||x - y||^2 &\leq D^k(x, y) \leq \frac{M}{2} ||x - y||^2.
\end{align*}
\]

To simplify our analysis, in what follows, we will drop the sub-index \( k \). Thanks to Assumption 2, the results we will establish hold for all \( k \). To this end, let a strongly twice differentiable convex function \( K \) along with a positive \( \epsilon \in (\underline{\epsilon}, \overline{\epsilon}) \) be given. Suppose a Bregman distance \( D \) is constructed based on \( K \). We introduce the following Bregman type mappings and functions which will play a key role for the convergence analysis of the VBPG method.

Bregman Proximal Envelope Function \( E_{D, \epsilon} \) is defined by

\[
E_{D, \epsilon}(x) = \min_{y \in \mathbb{R}^n} \{ f(x) + \langle \nabla f(x), y - x \rangle + g(y) + \frac{1}{\epsilon} D(x, y) \} \ \forall x \in \mathbb{R}^n,
\]

which is expressed as the value function of optimization problem (AP\(^k\)) (see (2)), where \( x^k \) is replaced by \( x \).

Bregman Proximal Mapping \( T_{D, \epsilon} \) is defined by

\[
T_{D, \epsilon}(x) = \arg\min_{y \in \mathbb{R}^n} (\nabla f(x), y - x) + g(y) + \frac{1}{\epsilon} D(x, y) \ \forall x \in \mathbb{R}^n,
\]

(5)

which can be viewed as the set of optimizers of optimization problem (AP\(^k\)). By Assumptions 1 and 2, \( T_{D, \epsilon}(x) \) is non-empty, the mapping \( T_{D, \epsilon}(x) \) could be multi-valued.

Bregman proximal gap function \( G_{D, \epsilon} \) is defined by

\[
G_{D, \epsilon}(x) = -\frac{1}{\epsilon} \min_{y \in \mathbb{R}^n} (\nabla f(x), y - x) + g(y) - g(x) + \frac{1}{\epsilon} D(x, y) \ \forall x \in \mathbb{R}^n.
\]

(6)
Obviously, we have \( G_{D,e}(x) \geq 0 \) for all \( x \). If \( g \) is semi-convex, the following optimization problem is equivalent to the differential inclusion problem \( 0 \in \partial_P F(x) \) associated with problem (P) (see Proposition 2.4):

\[
\min_{x \in \mathbb{R}^n} G_{D,e}(x).
\]

The above mappings and functions enjoy some favorable properties summarized in the following:

**Proposition 2.1. (Global properties of Bregman type mappings and functions)** Let a Bregman function \( D \) be given. Suppose that Assumptions [1] and [2] hold, and that \( \epsilon \in (0, m/L) \). Then for any \( x \in \mathbb{R}^n \), \( t_{D,e}(x) \in T_{D,e}(x) \), we have

1. \( E_{D,e}(x) = F(x) - \epsilon G_{D,e}(x) \);
2. \( F(t_{D,e}(x)) \leq E_{D,e}(x) - a \| x - t_{D,e}(x) \|^2 \) with \( a = \frac{1}{2}(\frac{m}{\epsilon} - L) \);
3. \( F(t_{D,e}(x)) \leq F(x) - a \| x - t_{D,e}(x) \|^2 \).

**Proof.** (i): This follows immediately from the definitions \( G_{D,e}(x) \) and \( E_{D,e}(x) \).

(ii) & (iii): Since \( \nabla f \) is \( L \)-Lipschitz, one has

\[
E_{D,e}(x) = f(x) + \langle \nabla f(x), t_{D,e}(x) - x \rangle + g(t_{D,e}(x)) + \frac{1}{\epsilon} D(x, t_{D,e}(x)) \geq f(t_{D,e}(x)) - \frac{L}{2} \| x - t_{D,e}(x) \|^2 + g(t_{D,e}(x)) + \frac{1}{\epsilon} D(x, t_{D,e}(x)).
\]

Thus, by (i) and the fact \( D(x, t_{D,e}(x)) \geq \frac{m}{\epsilon} \| x - t_{D,e}(x) \|^2 \), we get

\[
F(t_{D,e}(x)) \leq E_{D,e}(x) - \frac{1}{\epsilon} D(x, t_{D,e}(x)) + \frac{L}{2} \| x - t_{D,e}(x) \|^2 \leq F(x) - \frac{1}{2} \left( \frac{m}{\epsilon} - L \right) \| x - t_{D,e}(x) \|^2.
\]

This completes the proof. \( \square \)

**Proposition 2.2. (Properties of \( \partial_P F \))** Suppose that Assumptions [1] and [2] hold. Then for all \( t_{D,e}(x) \in T_{D,e}(x) \) we have

1. \( \text{dist}(0, \partial_P F(t_{D,e}(x))) \leq (L + \frac{M}{\epsilon}) \| x - t_{D,e}(x) \| \);
2. If \( x \in T_{D,e}(x) \), then \( 0 \in \partial_P F(x) \).

**Proof.** (i): The optimality condition of optimizer \( t_{D,e}(x) \in T_{D,e}(x) \) yields

\[
0 \in \nabla f(x) + \partial_P g(t_{D,e}(x)) + \frac{1}{\epsilon} \nabla D(x, t_{D,e}(x)).
\]
Let $\xi = \nabla f(t_{D,e}(x)) - \nabla f(x) - \frac{1}{\epsilon} \nabla_D D(x, t_{D,e}(x))$. Then we have

$$\xi \in \partial P F(t_{D,e}(x)) = \nabla f(t_{D,e}(x)) + \partial pg(t_{D,e}(x)).$$

By Assumptions [1] and [2] we have

$$\|\xi\| \leq \|\nabla f(t_{D,e}(x)) - \nabla f(x)\| + \frac{1}{\epsilon} \|\nabla_D D(x, t_{D,e}(x))\| \leq (L + M) \|x - t_{D,e}(x)\|,$$

which follows the desired statement.

(ii): The claim follows directly from statement (i). \qed

Under assumptions of Proposition 2.1 if $\tau < \frac{M}{L}$, then it’s easy to show that functions $E_{D,e}(x)$ and $G_{D,e}(x)$ are continuous, mapping $T_{D,e}(x)$ is closed and is continuous whenever $T_{D,e}(x)$ is single valued (see Proposition 6.1 of [42]).

Before the end of this section, we introduce the following lemma about the generalized descent inequality.

**Lemma 2.1** (Generalized descent inequality in the nonconvex case). *Suppose that Assumptions [1] and [2] hold. For any $t_{D,e}(x) \in T_{D,e}(x)$, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^n$, we have that*

$$a \left[ F(t_{D,e}(x)) - F(u) \right] \leq b \|u - x\|^2 - \|u - t_{D,e}(x)\|^2 - c \|x - t_{D,e}(x)\|^2,$$

*where $a = 2$, $b = \frac{4M}{L} + 2 + 3L$ and $c = \frac{L}{M} - (L + 2).$*

**Proof.** Denote $\Delta = \langle \nabla f(x), t_{D,e}(x) - u \rangle + g(t_{D,e}(x)) - g(u)$. First, we estimate the lower bound of $\Delta$:

$$\Delta = \langle \nabla f(x), t_{D,e}(x) - u \rangle + g(t_{D,e}(x)) - g(u)
= \langle \nabla f(x), t_{D,e}(x) - x \rangle + \langle \nabla f(x), x - u \rangle + g(t_{D,e}(x)) - g(u)
\geq f(t_{D,e}(x)) - f(x) - \frac{L}{2} \|x - t_{D,e}(x)\|^2 + \langle \nabla f(x), x - u \rangle + g(t_{D,e}(x)) - g(u)
\quad \text{(since $f$ is gradient Lipschitz with modulus $L$)}
= F(t_{D,e}(x)) - F(u) - \frac{L}{2} \|x - t_{D,e}(x)\|^2 + \langle f(u) - f(x) - \nabla f(x), x - u \rangle \quad \text{.}$$

By the gradient Lipschitz continuity of $f$, we estimate the term $\delta_1$ in (8):

$$\delta_1 = f(u) - f(x) - \langle \nabla f(u), u - x \rangle + \langle \nabla f(x) - \nabla f(u), x - u \rangle
\geq - \frac{L}{2} \|u - x\|^2 - \|\nabla f(x) - \nabla f(u)\| \cdot \|x - u\|
\geq - \frac{3L}{2} \|u - x\|^2.$$
Therefore, we have that
\[ \Delta \geq F(t_{D,e}(x)) - F(u) - \frac{L}{2} \| x - t_{D,e}(x) \|^2 - \frac{3L}{2} \| u - x \|^2. \] (9)

Since \( t_{D,e}(x) \) solves the minimization problem (5), we have

\[ \Delta = \langle \nabla f(x), t_{D,e}(x) - u \rangle + g(t_{D,e}(x)) - g(u) \leq \frac{1}{\epsilon} [D(x, u) - D(x, t_{D,e}(x))] \leq \frac{M}{2\epsilon} \| x - t_{D,e}(x) \|^2. \] (by Assumption 2) (10)

Since \(-\frac{1}{2} \| x - t_{D,e}(x) \|^2 + \| u - x \|^2 + \| x - t_{D,e}(x) \|^2 \geq 0\), (10) follows that

\[ \Delta \leq \frac{M}{2\epsilon} \| u - x \|^2 - \frac{1}{2} \| u - t_{D,e}(x) \|^2 + \| u - x \|^2 + \| x - t_{D,e}(x) \|^2 - \frac{m}{2\epsilon} \| x - t_{D,e}(x) \|^2 \]
\[ \leq \left( \frac{M}{2\epsilon} + 1 \right) \| u - x \|^2 - \frac{1}{2} \| u - t_{D,e}(x) \|^2 - \frac{m}{2\epsilon} \| x - t_{D,e}(x) \|^2. \] (11)

The desired result follows by combining (9) and (11). \( \Box \)

Remark 2.1 (Cost-to-go inequality [46]). From this lemma with \( \kappa = \max \{ \frac{2b - 1}{\alpha}, \frac{2b - c}{\alpha} \} > 0 \), we also get for \( x, u \in \mathbb{R}^n \),

\[ F(t_{D,e}(x)) - F(u) \leq \frac{1}{\alpha} \left\{ 2b \| u - t_{D,e}(x) \|^2 + 2b \| t_{D,e}(x) - x \|^2 - \| u - t_{D,e}(x) \|^2 - c \| x - t_{D,e}(x) \|^2 \right\} \]
\[ \leq \kappa \left( \| u - t_{D,e}(x) \|^2 + \| x - t_{D,e}(x) \|^2 \right) \] (12)

which is one cost-to-go estimate [46].

2.2 The properties of Bregman type mapping and function under semiconvexity of \( g \)

Proposition 2.3. (Single-valueness of Bregman proximal mappings) Suppose that Assumptions 7 and 2 hold, and that \( g \) is semiconvex on \( \mathbb{R}^n \) with constant \( \rho \) and \( \overline{\rho} < \min\{\frac{m}{\overline{T}}, \frac{m}{\overline{P}}\} \). Then for all \( x \in \mathbb{R}^n \), \( T_{D,e}(x) \) is single-valued.

Proof. The claim is derived directly by the definition of semi-convexity. \( \Box \)

Proposition 2.4. (Further properties of Bregman type mappings and functions) Suppose that the assumptions of Proposition 2.3 hold. Then for \( x \in \mathbb{R}^n \) and \( \overline{\rho} < \min\{\frac{m}{\overline{T}}, \frac{m}{\overline{P}}\} \) the following statements hold:

(i) \( E_{D,e}(x) \leq F(x) - \frac{1}{2}(\frac{m}{\overline{T}} - \rho) \| x - T_{D,e}(x) \|^2 \).
(ii) \( \frac{1}{2\epsilon}(m - \epsilon \rho)||x - T_{D,\epsilon}(x)||^2 \leq G_{D,\epsilon}(x) \);

(iii) \( G_{D,\epsilon}(x) \leq \frac{m}{2\epsilon^2}dist^2(0, \partial P F(x)) \);

(iv) \( ||x - T_{D,\epsilon}(x)|| \leq \left( \frac{m}{m - \epsilon \rho} \right) \sqrt{2}dist(0, \partial P F(x)) \);

(v) \( G_{D,\epsilon}(x) = 0 \) if only if \( x = T_{D,\epsilon}(x) \) or \( 0 \in \partial P F(x) \).

**Proof.** By Proposition [2,3], \( T_{D,\epsilon}(x) \) is single-valued.

(i): The optimality condition for the minimization problem in (2) follows that

\[
0 \in \nabla f(x) + \partial Lg(T_{D,\epsilon}(x)) + \frac{1}{\epsilon} \nabla D(x, T_{D,\epsilon}(x)).
\]  

Since \( g \) is l.s.c and semiconvex with \( \rho \), we have

\[
g(x) \geq g(T_{D,\epsilon}(x)) - \frac{\rho}{2}||x - T_{D,\epsilon}(x)||^2 - \langle \nabla f(x) + \frac{1}{\epsilon} \nabla D(x, T_{D,\epsilon}(x)), x - T_{D,\epsilon}(x) \rangle
\]

\[
\geq g(T_{D,\epsilon}(x)) - \frac{\rho}{2}||x - T_{D,\epsilon}(x)||^2 - \langle \nabla f(x), x - T_{D,\epsilon}(x) \rangle
\]

\[
+ \frac{1}{\epsilon}D(x, T_{D,\epsilon}(x)) - \frac{1}{\epsilon}D(x, x) + \frac{m}{2\epsilon}||x - T_{D,\epsilon}(x)||^2
\]

\[
\geq g(T_{D,\epsilon}(x)) - \langle \nabla f(x), x - T_{D,\epsilon}(x) \rangle + \frac{1}{\epsilon}D(x, T_{D,\epsilon}(x)) + \frac{1}{\epsilon}(\frac{m}{\epsilon} - \rho) ||x - T_{D,\epsilon}(x)||^2.
\]

Adding \( f(x) \) to both sides, by the definition of \( E_{D,\epsilon}(x) \) the claim is provided.

(ii): From statement (i) of Proposition 2.1 we have

\[
G_{D,\epsilon}(x) = \frac{1}{\epsilon}(F(x) - E_{D,\epsilon}(x)) \geq \frac{1}{2\epsilon^2}(m - \epsilon \rho)||x - T_{D,\epsilon}(x)||^2.
\]  

(by i) of this proposition)

(iii): For \( \epsilon < \min\{\frac{\rho}{\sqrt{T}}, \frac{\rho}{\rho} \} \), we have

\[
e G_{D,\epsilon}(x) = -\langle \nabla f(x), T_{D,\epsilon}(x) - x \rangle - g(T_{D,\epsilon}(x)) + g(x) - \frac{1}{\epsilon}D(x, T_{D,\epsilon}(x)).
\]

Let \( \nu \in \partial D g(x) \), thanks the semiconvex of \( g \), we get

\[
e G_{D,\epsilon}(x) \leq -\langle \nabla f(x), T_{D,\epsilon}(x) - x \rangle - \langle \nu, T_{D,\epsilon}(x) - x \rangle + \frac{\rho}{2}||x - T_{D,\epsilon}(x)||^2 - \frac{m}{2\epsilon}||x - T_{D,\epsilon}(x)||^2
\]

\[
= -\langle \nabla f(x) + \nu, T_{D,\epsilon}(x) - x \rangle - \frac{1}{2}(\frac{m}{\epsilon} - \rho)||x - T_{D,\epsilon}(x)||^2
\]

\[
\leq ||\nabla f(x) + \nu|| \cdot ||x - T_{D,\epsilon}(x)|| - \frac{1}{2}(\frac{m}{\epsilon} - \rho)||x - T_{D,\epsilon}(x)||^2
\]

\[
\leq \frac{\epsilon}{2(m - \epsilon \rho)}||\nabla f(x) + \nu||^2.
\]  

\[
(14)
\]

Therefore \( G_{D,\epsilon}(x) \leq \frac{\epsilon}{2(m - \epsilon \rho)}||\nabla f(x) + \nu||^2, \forall \nu \in \partial D g(x) \), and the claim is verified.

(iv) and (v): The statement (iv) is a simple consequence of (ii) and (iii). (v) follows directly from statements (ii), (iii) and ([13]). \( \square \)
3 Convergence analysis of VBPG

Section 3 studies convergence behaviors of sequences generated by the update formula of VBPG \( (2) \). We will use \( D_k \) explicitly and assume that \( D_k \) and parameters \( \epsilon_k \) satisfy Assumption \( 2 \) uniformly throughout Sections 3. A number of basic properties of sequences \( \{x^k\} \) and \( \{F(x^k)\} \) are summarized in the following proposition.

**Proposition 3.1.** Suppose that the Assumptions \( 1 \) and \( 2 \) hold and \( \epsilon < \frac{m}{L} \). Let \( \{x^k\} \) be a sequence generated by the VBPG method. Then the following assertions hold:

(i) The sequence \( \{F(x^k)\} \) is strictly decreasing (unless \( x^k \in \bar{X}_P \) for some \( k \)).

(ii) \( \{x^k\} \) is bounded, and any cluster point \( \bar{x} \) of \( \{x^k\} \) is a limiting critical point of \( F \): \( 0 \in \partial_L F(\bar{x}) \). Furthermore, \( \bar{x} \) is also actually a proximal critical point of \( F \): \( 0 \in \partial_P F(\bar{x}) \).

(iii) Let \( \Omega \) be the set of accumulation points of \( \{x^k\} \). Then \( F_\zeta := \lim_{k \to +\infty} F(x^k) \) exists and \( F(\bar{x}) \leq F_\zeta \) for every \( \bar{x} \) in \( \Omega \).

**Proof.** (i): By (iii) of Proposition \( 2.1 \) we have

\[
F(x^{k+1}) \leq F(x^k) - a\|x^k - x^{k+1}\|^2 \quad \text{with} \quad a = \frac{1}{2}(\frac{m}{\epsilon} - L). \tag{15}
\]

If \( x^k = x^{k+1} \), then by (ii) of Proposition \( 2.2 \) we have \( x^k \in \bar{X}_P \). Otherwise \( \{F(x^k)\} \) is strictly decreasing.

(ii): By summation for (15), we have

\[
a \sum_{k=0}^{N} \|x^k - x^{k+1}\|^2 \leq F(x^0) - F(x^{N+1}) \leq F(x^0) - F^* \forall N.
\]

and it follows that

\[
\sum_{k=0}^{\infty} \|x^k - x^{k+1}\|^2 < +\infty, \quad \|x^k - x^{k+1}\| \to 0, \text{ when } k \to \infty.
\]

The boundedness of \( \{x^k\} \) comes from Assumption \( 1 \) \( F = (f + g) \) is level bounded along with the fact that \( \{F(x^k)\} \) is strictly decreasing. Thus \( \{x^k\} \) has at least one cluster point. Let \( \bar{x} \) denote such a point and \( x^{k'} \to \bar{x}, \) \( k' \to \infty \). From statement (i) of Proposition \( 2.2 \) we have

\[
\text{dist} \left( 0, \partial_L F(x^{k'}) \right) \leq \text{dist} \left( 0, \partial_P F(x^{k'}) \right) \leq \left( L + \frac{M}{\epsilon} \right)\|x^{k'} - x^{k'-1}\|.
\]

Thus \( \text{dist} \left( 0, \partial_L F(x^{k'} \right) \to 0 \) as \( k' \to \infty \). By the closedness of \( \partial_L F(\cdot) \), we have \( 0 \in \partial_L F(\bar{x}) \).
Furthermore, by the update formula of VBPG in (AP), we get
\[
\langle \nabla f(x^k), x^{k+1} - x^k \rangle + g(x^{k+1}) + \frac{1}{\epsilon_k}D^k(x^k, x^{k+1})
\]
\[
\leq \langle \nabla f(x^k), x - x^k \rangle + g(x) + \frac{1}{\epsilon_k}D^k(x^k, x)
\]
\[
\leq f(x) - f(x^k) + \frac{L}{2}\|x - x^k\|^2 + g(x) + \frac{1}{\epsilon_k}D^k(x^k, x) \forall x \in \text{dom}(F).
\]

By Assumption 2 for \(D^k\), the above inequality yields
\[
\langle \nabla f(x^k), x^{k+1} - x^k \rangle + g(x^{k+1}) + \frac{m}{2\epsilon_k}\|x^k - x^{k+1}\|^2
\]
\[
\leq f(x) - f(x^k) + \frac{L}{2}\|x - x^k\|^2 + g(x) + \frac{M}{2\epsilon_k}\|x^k - x\|^2 \forall x \in \text{dom}(F).
\]

Taking \(k = k'\) such that \(\lim_{k' \to \infty} x^{k'} = \bar{x}\), from the continuity of \(f\) and lower semicontinuous of \(g\), we obtain that
\[
\lim_{k' \to \infty} f(x^{k'}) = f(\bar{x}) \quad \text{and} \quad \lim_{k' \to \infty} g(x^{k'}) \geq g(\bar{x}).
\]

Therefore, by taking \(k' \to \infty\) on both sides of (16) and one has
\[
F(\bar{x}) \leq F(x) + \left(\frac{L}{2} + \frac{M}{2\epsilon_k}\right)\|x - \bar{x}\|^2 \forall x \in \text{dom}(F),
\]
which implies 0 \(\in \partial_p F(\bar{x})\).

(iii): By (15), \(\lim_{k \to \infty} F(x^k) \to F_\xi \geq F^*\). Let \(\bar{x} \in \Omega\), then there exists a subsequence \(\{x^{k'}\}\) of \(\{x^k\}\) such that \(x^{k'} \to \bar{x}\). By the lower semicontinuity of \(F\) on \(\text{dom} F\) and the convergence of \(\{F(x^k)\}\), we have \(F(\bar{x}) \leq \lim_{k' \to \infty} F(x^{k'}) = F_\xi. \quad \square\)

In order to study the linear convergence of the sequence generated by VBPG, we need the following concept of value proximity error bound. Let \(\bar{x} \in \mathbb{R}^n\) and \(\bar{F} = F(\bar{x})\). For given positive numbers \(\eta\) and \(\mu\), let
\[
\mathcal{B}(\bar{x}; \eta, \nu) = \mathcal{B}(\bar{x}; \eta) \cap \{x \in \mathbb{R}^n : \bar{F} < F(x) < \bar{F} + \nu\}.
\]

**Definition 3.1** (VBPG iteration based Value Proximity Error Bound (VP-EB)). Let \(\{x^k\}\) be the sequence generated by VBPG method and \(\bar{x}\) be an accumulation point of \(\{x^k\}\). We say the VP-EB holds at \(\bar{x}\) if there exist \(\kappa', \eta\) and \(\nu > 0\) such that
\[
F(x^{k+1}) - F(\bar{x}) \leq \kappa'\|x^k - x^{k+1}\|^2 \forall x^{k+1} \in \mathcal{B}(\bar{x}; \eta, \nu).
\]

(17)

We next show that a sequence generated by (3) is convergent and has a finite length property.

**Proposition 3.2.** (Finite length property of whole sequence \(\{x^k\}\)) Let the sequence \(\{x^k\}\) be generated by VBPG method and \(\bar{x}\) be an accumulation point of \(\{x^k\}\). Suppose that the Assumptions 1 and 2 hold and \(\bar{F} < \frac{n}{4}\) and that the VP-EB holds at the point \(\bar{x}\) with \(\kappa', \eta\) and \(\nu > 0\). Let \(\lambda\) be the constant given in Proposition 2.1 and \(\bar{F} = F(\bar{x})\). Then the following statements hold:
(i) \( x^k \in \mathcal{B}(\bar{x}; \eta, \nu) \) \( \forall k \geq k_0; \)

(ii) \( \sum_{i=0}^{+\infty} \|x^i - x^{i+1}\| < +\infty \) (finite length property);

(iii) The sequence \( \{x^i\} \) actually converges to the point \( \bar{x} \) which is a proximal critical point of \( F. \)

Proof. (i): By (i) and (iii) of Proposition 3.1 there is \( k_0 \) such that we have \( \bar{F} < F(x^k) \) \( \forall k \geq k_0. \) From assumptions, without loss of generality, we assume that

\[
\bar{F} < F(x^{k_0}) < \bar{F} + \nu \forall k \geq k_0.
\]

We will use the Principle of Mathematical Introduction to prove that the sequence \( \{x^i\} \subset \mathcal{B}(\bar{x}; \eta, \nu) \) \( \forall k \geq k_0. \) It is clear that \( x^{k_0} \in \mathcal{B}(\bar{x}; \eta, \nu) \) by (18) and (19). The inequalities \( \bar{F} < F(x^{k_0+1}) \leq F(x^{k_0}) < \bar{F} + \nu \) hold trivially. On the other hand, by (15), we have

\[
\|x^{k_0+1} - x^{k_0}\| \leq \sqrt{\frac{F(x^{k_0}) - F(x^{k_0+1})}{a}} \leq \sqrt{\frac{F(x^{k_0}) - \bar{F}}{a}} \tag{18}
\]

and

\[
\|x^{k_0+1} - \bar{x}\| \leq \|x^{k_0} - \bar{x}\| + \|x^{k_0} - x^{k_0+1}\| \leq \|x^{k_0} - \bar{x}\| + \sqrt{\frac{F(x^{k_0}) - \bar{F}}{a}} < \eta. \tag{19}
\]

Thus \( x^{k_0+1} \in \mathcal{B}(\bar{x}; \eta, \nu). \) Now suppose that \( x^i \in \mathcal{B}(\bar{x}; \eta, \nu) \) for \( i = k_0 + 1, \ldots, k_0 + k \) and \( x^{k_0+k} \neq x^{k_0+k+1}. \) Note that \( F(x^{k_0+1}) > F(x^{k_0+2}) > \cdots > F(x^{k_0+k}) > F(x^{k_0+k+1}) > \bar{F}. \) We need to show that \( x^{k_0+k+1} \in \mathcal{B}(\bar{x}; \eta, \nu). \) By the concavity of function \( h(y) = y^\frac{1}{2}, \) we have, for \( i = k_0 + 1, k_0 + 2, \ldots, k_0 + k, \) that

\[
\left( F(x^i) - \bar{F} \right)^{\frac{1}{2}} - \left( F(x^{i+1}) - \bar{F} \right)^{\frac{1}{2}} \geq \frac{1}{2} \left( F(x^i) - F(x^{i+1}) \right) \tag{20}
\]

Recalling that \( x^{i+1} \in T_{D_i,x^i}(x^i) \) and applying (iii) of Proposition 2.1 and (17) of VP-EB to \( F(x^i) - F(x^{i+1}) \) and \( (F(x^i) - \bar{F})^{\frac{1}{2}} \) respectively one has

\[
\frac{2 \sqrt{\kappa}}{a} \|x^i - x^{i-1}\| \left( F(x^i) - \bar{F} \right)^{\frac{1}{2}} - \left( F(x^{i+1}) - \bar{F} \right)^{\frac{1}{2}} \geq ||x^i - x^{i+1}||^2.
\]

It follows from \( 2 \sqrt{d_1 d_2} \leq d_1 + d_2 \) with nonnegative \( d_1 \) and \( d_2 \) that

\[
2 \|x^{i+1} - x^i\| \leq ||x^i - x^{i-1}|| + \frac{2 \sqrt{\kappa}}{a} \left( F(x^i) - \bar{F} \right)^{\frac{1}{2}} - \left( F(x^{i+1}) - \bar{F} \right)^{\frac{1}{2}} \tag{21}
\]

Summing (20) for \( i = k_0 + 1, \ldots, k_0 + k, \) we obtain

\[
\sum_{i=k_0+1}^{k_0+k} \|x^{i+1} - x^i\| + \|x^{k_0+k+1} - x^{k_0+k}\|
\]

\[
\leq \|x^{k_0+1} - x^{k_0}\| + \frac{2 \sqrt{\kappa}}{a} \left( F(x^{k_0+1}) - \bar{F} \right)^{\frac{1}{2}} - \left( F(x^{k_0+k+1}) - \bar{F} \right)^{\frac{1}{2}} \tag{21}
\]
Using (21) along with the triangle inequality, we have

\[
\|\bar{x} - x^{k_0+k+1}\| \leq \|\bar{x} - x^{k_0}\| + \|x^{k_0} - x^{k_0+1}\| + \sum_{i=k_0+1}^{k_0+k} \|x^{i+1} - x^i\|
\]

\[
\leq \|\bar{x} - x^{k_0}\| + 2\|x^{k_0} - x^{k_0+1}\| + \frac{2\sqrt{\kappa'}}{a} \left[ \left(F(x^{k_0+1}) - \bar{F}\right)^\frac{1}{2} \right]
\]

\[
\leq \|\bar{x} - x^{k_0}\| + \sqrt{2\frac{\|F(x^{k_0}) - \bar{F}\|}{a}} + \frac{2\sqrt{\kappa'}}{a} \left[ \left(F(x^{k_0}) - \bar{F}\right)^\frac{1}{2} \right]
\]

\[
< \eta. \quad \text{(by (19))}
\]

This shows that \(x^{k_0+k+1} \in B(\bar{x}; \eta, \nu)\), and (i) is proved by the Principle of Mathematical Induction.

(ii) and (iii): A direct consequence of (21) is, for all \(k\),

\[
\sum_{i=k_0+1}^{k_0+k} \|x^{i+1} - x^i\| \leq \|x^{k_0+1} - x^{k_0}\| + \frac{2\sqrt{\kappa'}}{a} \left[ \left(F(x^{k_0+1}) - \bar{F}\right)^\frac{1}{2} \right] < +\infty.
\]

Therefore, we have

\[
\sum_{i=0}^{+\infty} \|x^{i+1} - x^i\| < +\infty.
\]

In particular, this implies that the sequence \(\{x^k\}\) is convergent and thus it actually converges to the point \(\bar{x}\). And \(\bar{x}\) is a desired critical point of \(F\) by Proposition 3.1.

The main result of this section follows.

**Theorem 3.1** (Sufficient conditions for local linear convergence). Suppose that all the conditions of Proposition 3.2 hold. Then \(\{F(x^k)\}\) converges to value \(\bar{F}\) at the Q-linear rate of convergence; that is, there are some \(\beta \in (0, 1)\) and \(k_0\) such that

\[
F(x^{k+1}) - \bar{F} \leq \beta(F(x^k) - \bar{F}) \quad \forall k \geq k_0.
\]

Moreover, the sequence \(\{x^k\}\) converges at the R-linear rate to the critical point \(\bar{x}\), which is either a limiting critical point or a proximal critical point of \(F\).

**Proof.** In view of Proposition 3.2 there is \(k_0\) such that for \(k \geq k_0\) such that \(\{x^k\} \subset B(\bar{x}; \eta, \nu)\). It follows that

\[
F(x^{k+1}) - \bar{F} = F(x^k) - \bar{F} + (F(x^{k+1}) - F(x^k))
\]

\[
\leq F(x^k) - \bar{F} - a\|x^{k+1} - x^k\|^2 \quad \text{(by (iii) in Proposition 2.1)}
\]

\[
\leq F(x^k) - \bar{F} - \frac{1}{k^2} \left(F(x^{k+1}) - \bar{F}\right). \quad \text{(by VP-EB condition)}
\]

\[
\leq \frac{1}{k^2} (F(x^{k+1}) - \bar{F}).
\]

\[
< \eta. \quad \text{(by (19))}
\]

This shows that \(x^{k_0+k+1} \in B(\bar{x}; \eta, \nu)\), and (i) is proved by the Principle of Mathematical Induction.

(ii) and (iii): A direct consequence of (21) is, for all \(k\),

\[
\sum_{i=k_0+1}^{k_0+k} \|x^{i+1} - x^i\| \leq \|x^{k_0+1} - x^{k_0}\| + \frac{2\sqrt{\kappa'}}{a} \left[ \left(F(x^{k_0+1}) - \bar{F}\right)^\frac{1}{2} \right] < +\infty.
\]

Therefore, we have

\[
\sum_{i=0}^{+\infty} \|x^{i+1} - x^i\| < +\infty.
\]

In particular, this implies that the sequence \(\{x^k\}\) is convergent and thus it actually converges to the point \(\bar{x}\). And \(\bar{x}\) is a desired critical point of \(F\) by Proposition 3.1.

The main result of this section follows.

**Theorem 3.1** (Sufficient conditions for local linear convergence). Suppose that all the conditions of Proposition 3.2 hold. Then \(\{F(x^k)\}\) converges to value \(\bar{F}\) at the Q-linear rate of convergence; that is, there are some \(\beta \in (0, 1)\) and \(k_0\) such that

\[
F(x^{k+1}) - \bar{F} \leq \beta(F(x^k) - \bar{F}) \quad \forall k \geq k_0.
\]

Moreover, the sequence \(\{x^k\}\) converges at the R-linear rate to the critical point \(\bar{x}\), which is either a limiting critical point or a proximal critical point of \(F\).

**Proof.** In view of Proposition 3.2 there is \(k_0\) such that for \(k \geq k_0\) such that \(\{x^k\} \subset B(\bar{x}; \eta, \nu)\). It follows that

\[
F(x^{k+1}) - \bar{F} = F(x^k) - \bar{F} + (F(x^{k+1}) - F(x^k))
\]

\[
\leq F(x^k) - \bar{F} - a\|x^{k+1} - x^k\|^2 \quad \text{(by (iii) in Proposition 2.1)}
\]

\[
\leq F(x^k) - \bar{F} - \frac{1}{k^2} \left(F(x^{k+1}) - \bar{F}\right). \quad \text{(by VP-EB condition)}
\]

\[
\leq \frac{1}{k^2} (F(x^{k+1}) - \bar{F}).
\]

\[
< \eta. \quad \text{(by (19))}
\]
Therefore, one has
\[
F(x^{k+1}) - F \leq \frac{1}{1 + a\left(\frac{1}{2}\right)} \left(F(x^k) - \bar{F}\right) \quad \forall k \geq k_0. \tag{24}
\]

The above estimation shows that \(\{F(x^k)\}\) converges to \(\bar{F}\) at the Q-linear rate; that is,
\[
F(x^{k+1}) - \bar{F} \leq \beta \left(F(x^k) - \bar{F}\right) \quad \forall k \geq k_0, \tag{25}
\]
where \(\beta := \frac{1}{1 + a\left(\frac{1}{2}\right)} \in (0, 1)\). We now derive the R-linear rate of convergence of \(\{x^k\}\). We have
\[
\|x^k - x^{k+1}\|^2 \leq \frac{1}{a} \left[ (F(x^k) - \bar{F}) - (F(x^{k+1}) - \bar{F}) \right] \quad \text{(by (15))}
\]
\[
\leq \frac{1}{a} (F(x^k) - \bar{F})
\]
\[
\leq \frac{\beta^{k-k_0}}{a} (F(x^{k_0}) - \bar{F}). \quad \text{(by (25))}
\]

From the above inequality, we see that
\[
\|x^k - x^{k+1}\| \leq \hat{M}\left(\sqrt{\beta}\right)^{(k-k_0)} \quad \forall k > k_0, \quad \hat{M} := \sqrt{\frac{F(x^{k_0}) - \bar{F}}{a}}.
\]

By Proposition 3.2, we have \(\{x^k\}\) converges to the critical point \(\bar{x}\) and \(\|x^k - \bar{x}\| \leq \sum_{i=k}^{\infty} \|x^i - x^{i+1}\| \leq \frac{\hat{M}}{1 - \sqrt{\beta}} \left(\sqrt{\beta}\right)^{(k-k_0)}\). This shows that \(\{x^k\}\) converges to the desired critical point \(\bar{x}\) at the R-linear rate; that is,
\[
\limsup_{k \to \infty} \left(\frac{\alpha - k}{\sqrt{\|x^k - \bar{x}\|}}\right) = \sqrt{\beta} < 1.
\]

This completes the proof. \(\square\)

4 Level-set based error bounds and necessary and sufficient conditions for linear convergence of VBPG

4.1 Level-set subdifferential EB implies VP-EB

In the rest of this paper, unless otherwise stated, we will always choose \(\bar{F} = F(\bar{x})\) for some given \(\bar{x} \in \text{dom}\, F\). Set \([F \leq \bar{F}] = \{x \in \mathbb{R}^n : F(x) \leq F(\bar{x})\}\) and \([F > \bar{F}] = \{x \in \mathbb{R}^n : F(x) > F(\bar{x})\}\). In this subsection, we examine level-set subdifferential and level-set Bregman proximal error bounds.

Definition 4.1 (Level-set error bounds). (i) (Level-set subdifferential error bound) The function \(F\) is said to satisfy the level-set subdifferential error bound (EB) condition at \(\bar{x}\) with exponent \(\gamma > 0\) if there exist \(\eta > 0\), \(\nu > 0\), and \(c_1 > 0\) such that the following inequality holds:
\[
dist^\gamma(x, [F \leq \bar{F}]) \leq c_1 \dist\left(0, \partial F(x)\right) \quad \forall x \in \mathcal{B}(\bar{x}; \eta, \nu).
\tag{26}
\]
(ii) **(Level-set Bregman proximal error bound)** Given a Bregman function $D$ along with $\epsilon > 0$, $F$ is said to satisfy the level-set Bregman proximal EB condition at $\bar{x}$ with exponent $p > 0$, if there exist $\eta, \nu > 0$, and $\theta > 0$ such that the following inequality holds:

$$
\text{dist}^p(x, [F \leq \bar{F}]) \leq \theta \text{dist}(x, T_{D,\epsilon}(x)) \forall x \in \mathcal{B}(\bar{x}; \eta, \nu).
$$

**(Proposition 4.1)** (Level-set subdifferential EB implies level-set Bregman proximal EB). Suppose Assumptions[7] and[2] hold with $\bar{\epsilon} < \frac{m}{L}$. Assume the level-set subdifferential EB holds at $\bar{x}$ with exponent $\gamma \in (0, \infty)$ over $\mathcal{B}(\bar{x}; \eta, \nu)$. Then there are $N > \max\{\frac{\sqrt{p}}{m-\epsilon L}/(\frac{2}{\gamma})^2, 1\}$ and $\theta > 0$ such that

$$
\text{dist}^p(x, [F \leq \bar{F}]) \leq \theta \text{dist}(x, T_{D,\epsilon}(x)) \forall x \in \mathcal{B}(\bar{x}, \frac{\eta}{2}, \frac{\nu}{N}),
$$

where $p = \frac{1}{\min\{\eta, 1\}}$, $\theta_1 = 1 + c^1(L + \frac{M}{\epsilon})^{\frac{1}{p}}$, $\theta_2 = (\frac{p}{2})^{1-\frac{1}{p}} + c^1(L + \frac{M}{\epsilon})^{\frac{1}{p}}$ and $\theta = \max\{\theta_1, \theta_2\}$.

**Proof.** By assumptions, $\bar{\epsilon} < m/L$, $T_{D,\epsilon}(x) \neq \emptyset$. For $x \in \mathcal{B}(\bar{x}, \frac{\eta}{2}, \frac{\nu}{N})$, let $t_p(x) \in \text{Proj}_{T_{D,\epsilon}(x)}(x)$. If $F(t_p(x)) \leq \bar{F}$, then $t_p(x) \in [F \leq \bar{F}]$, and we have

$$
\text{dist}(x, [F \leq \bar{F}]) \leq \|x - t_p(x)\|.
$$

Now we consider the non-trivial case $F(t_p(x)) > \bar{F}$. If $x \in \mathcal{B}(\bar{x}; \frac{\eta}{2}, \frac{\nu}{N}) \subset \mathcal{B}(\bar{x}; \eta, \nu)$ with $N$ satisfying the assumptions, then by (iii) of Proposition[7,1] we have that

$$
\frac{1}{2} \left(\frac{m}{\epsilon} - L\right) \|x - t_p(x)\| \leq F(x) - F(t_p(x)) \leq F(x) - \bar{F} \leq \frac{\nu}{N}.
$$

Since $N > \frac{2\nu}{m-\epsilon L}/(\frac{2}{\gamma})^2$, we get $\sqrt{\frac{2\nu}{N(m-\epsilon L)}} < \frac{\nu}{2}$ and $\|x - t_p(x)\| < \frac{\nu}{2}$. As $\|x - \bar{x}\| \leq \frac{\nu}{2}$, it follows that $\|t_p(x) - \bar{x}\| < \eta$, which yields $t_p(x) \in \mathcal{B}(\bar{x}; \eta)$. Hence for any $x \in \mathcal{B}(\bar{x}; \frac{\eta}{2}, \frac{\nu}{N})$, by the level-set subdifferential EB we have

$$
\text{dist}(x, [F \leq \bar{F}]) \leq \|x - t_p(x)\| + \text{dist}(t_p(x), [F \leq \bar{F}]) = \|x - t_p(x)\| + c^1 \text{dist}^\frac{1}{p}(0, \partial_p F(t_p(x))).
$$

Therefore, for any $x \in \mathcal{B}(\bar{x}; \frac{\eta}{2}, \frac{\nu}{N})$, by (i) of Proposition[7,2] we have that

$$
\text{dist}(x, [F \leq \bar{F}]) \leq \|x - t_p(x)\| + c^1(L + \frac{M}{\epsilon})^{\frac{1}{p}}\|x - t_p(x)\|^{\frac{1}{p}}.
$$

Since $\|x - t_p(x)\| < \frac{\nu}{2}$, by the above inequality, we have the following estimate

$$
\text{dist}(x, [F \leq \bar{F}]) \leq \left\{ \begin{array}{ll}
\theta_1 \|x - t_p(x)\| & \text{if } 0 < \gamma \leq 1, \\
\theta_2 \|x - t_p(x)\|^{\frac{1}{p}} & \text{if } \gamma > 1
\end{array} \right.
$$

$$
\leq \theta \|x - t_p(x)\|^{\frac{1}{p}} = \theta \text{dist}^\frac{1}{p}(x, T_{D,\epsilon}(x)).
$$

This completes the proof.

\[\square\]

Next proposition shows that level-set subdifferential EB plays a key role for linear convergence of VBPG.
**Proposition 4.2 (Level-set subdifferential EB implies VP-EB).** Suppose that the Assumptions 1 and 2 hold. Let \( \{x^k\} \) be the sequence generated by VBPG method and \( \bar{x} \) be an accumulation point of \( \{x^k\} \). Assume that the level-set subdifferential EB holds at the point \( \bar{x} \) with exponent \( \gamma \in (0, 1] \), \( \eta > 0 \) and \( \nu > 0 \). Then there exists \( k'_0 > 0 \) such that VP-EB holds at \( \bar{x} \) for all \( k \geq k'_0 \).

**Proof.** Let \( x^k_p \in [F \leq \bar{F}] \) such that \( \|x^{k+1} - x^{k+1}_p\| = \text{dist}(x^{k+1}, [F \leq \bar{F}]) \). From the cost-to-go inequality in Lemma 2.1 with \( u = x^{k+1}_p \), we have

\[
F(x^{k+1}) - F(x^{k+1}_p) \leq \kappa(\|x^{k+1}_p - x^{k+1}\|^2 + \|x^k - x^{k+1}\|^2).
\] (29)

The VBPG update process implies that there is \( k'_0 > 0 \) such that for \( k \geq k'_0 \), we have \( \|x^k - x^{k+1}\| \leq 1 \). Since level-set subdifferential EB holds at \( \bar{x} \) with \( \gamma \in (0, 1] \), \( \eta, \nu > 0 \), for \( x^{k+1} \in \mathcal{B}(\bar{x}; \eta, \nu) \) and \( k \geq k'_0 \), we have

\[
\|x^{k+1} - x^k\|^2 = \text{dist}^2(x^{k+1}, [F \leq \bar{F}]) \leq (c_1)^2 \text{dist}^2(0, \partial_p F(x^{k+1})) \\
\leq (c_1)^2 \left( L + \frac{M}{\epsilon} \right)^2 \|x^k - x^{k+1}\|^2 \\
\leq (c_1)^2 \left( L + \frac{M}{\epsilon} \right)^2 \|x^k - x^{k+1}\|^2.
\] (30)

Using (29) and (30), it yields

\[
F(x^{k+1}) - \bar{F} \leq F(x^{k+1}) - F(x^{k+1}_p) \leq \kappa' \|x^k - x^{k+1}\|^2 \\
(\kappa' = \max((c_1)^2 \left( L + \frac{M}{\epsilon} \right)^2 \kappa, \kappa))
\]

which implies the claim. \( \square \)

### 4.2 The strong level-set error bounds and necessary and sufficient conditions for linear convergence

We now turn to study the notion of the strong level-set error bounds holding on a set \( \bar{F} \leq F \leq \bar{F} + \nu \). This notion plays an important role in deriving a sufficient condition and a necessary condition for linear convergence relative to level sets.

**Definition 4.2 (Strong level-set error bounds).** (i) **(Strong level-set subdifferential error bound)** We say that \( F \) satisfies the strong level-set subdifferential EB condition on \( \bar{F} < F < \bar{F} + \nu \) with the value \( \bar{F} \) and \( \nu > 0 \) if there exists \( c'_1 > 0 \) such that

\[
\text{dist}(x, [F \leq \bar{F}]) \leq c'_1 \text{dist}(0, \partial_p F(x)) \quad \forall x \in [\bar{F} < F < \bar{F} + \nu].
\]

(ii) **(Strong level-set Bregman proximal error bound)** Given a Bregman distance \( D \) along with \( \epsilon > 0 \), \( F \) satisfies the strong level-set Bregman proximal EB condition on \( \bar{F} < F < \bar{F} + \nu \) with \( \bar{F} \) and \( \nu > 0 \) if there exists \( \theta' > 0 \) such that

\[
\text{dist}(x, [F \leq \bar{F}]) \leq \theta' \text{dist}(x, T_{D, \epsilon}(x)) \quad \forall x \in [\bar{F} < F < \bar{F} + \nu].
\]
Corollary 4.1 (Strong level-set subdifferential EB ⇒ Strong level-set Bregman proximal EB). Suppose Assumptions 2 and 3 hold with \( \bar{\tau} < \frac{m}{L} \). Assume the strong level-set subdifferential EB holds over \([\bar{F} < F < \bar{F} + \nu] \). Then there is \( \theta' = 1 + c'_i (L + \frac{M}{\lambda}) > 0 \) such that

\[
\text{dist}(x, [F \leq \bar{F}]) \leq \theta' \text{dist} (x, T_{D,\epsilon}(x)) \forall x \in [\bar{F} < F < \bar{F} + \nu].
\]

Proof. The claim is proved by the same argument for the proof of Proposition 4.1 with \( \gamma = 1 \). \( \square \)

The following theorem gives a necessary condition and a sufficient condition for linear convergence relative to a level set.

Theorem 4.1. (Necessary and sufficient conditions for linear convergence relative to \([F \leq \bar{F}]\)) Suppose that Assumptions 2 and 3 hold. Let a sequence \( \{x^k\} \) be generated by the VBPG method, let \( \bar{x} \) be an accumulation point of \( \{x^k\} \), and let \( \nu > 0 \) be given.

(i) For any initial point \( x^0 \in [\bar{F} < F < \bar{F} + \nu] \), if the strong Bregman proximal EB condition holds on \([\bar{F} < F < \bar{F} + \nu] \) with \( \theta' \in \left( \sqrt{\frac{b}{b + 1}}, \sqrt{\frac{1}{b + 1}} \right) \), then the VBPG method converges linearly respect to level-set \([F \leq \bar{F}]\), i.e.,

\[
\text{dist} \left( x^{k+1}, [F \leq \bar{F}] \right) \leq \beta \text{dist} \left( x^k, [F \leq \bar{F}] \right) \quad \forall k \geq 0,
\]

with \( \beta := \sqrt{\frac{b - \frac{c}{\theta'}}{\theta'}} \in (0, 1) \), where the values of \( b \) and \( c \) are appeared in Lemma 2.1.

(ii) If \( g \) is semi-convex on \( \mathbb{R}^n \), \( \tau < \min\{\frac{m}{L}, \frac{M}{\lambda}\} \) and the VBPG method converges linearly in the sense of (31) with \( \beta \in (0, 1) \), then \( F \) satisfies the strong level-set subdifferential EB condition on \([\bar{F} < F < \bar{F} + \nu]\) with \( c'_i = \frac{\bar{\tau}}{1 - \beta(m - \rho p)} \frac{1}{\sqrt{2}} \) and strong level-set Bregman proximal EB with \( \theta' = 1 + c'_i (L + \frac{M}{\lambda}) \).

Proof. (i) Since \( \{F(x^k)\} \) is strictly decreasing and converges to \( F_{\bar{F}} \geq \bar{F} \), then we must have \( F(x^k) \geq \bar{F} \). Note that the equality implies \( x^k \in \bar{x}_p \). Therefore, for given \( x^0 \in [\bar{F} < F < \bar{F} + \nu] \), we have \( x^k \in [\bar{F} < F < \bar{F} + \nu] \).

Let \( x^k_p = \text{Proj}_{F \leq \bar{F}}(x^k) \). Then \( F(x^k_p) \leq \bar{F} \). By Lemma 2.1 with \( u = x^k_p \) in (7), we have \( F(x^{k+1}) \geq \bar{F} \) and

\[
0 \leq \alpha [F(x^{k+1}) - F(x^k_p)] \leq b||x^k_p - x^k||^2 - ||x^k_p - x^{k+1}||^2 - c||x^k - x^{k+1}||^2.
\]

This together with strong Bregman proximal EB condition yields

\[
||x^k_p - x^{k+1}||^2 \leq b||x^k_p - x^k||^2 - c||x^k - x^{k+1}||^2 \leq b||x^k_p - x^k||^2 - \frac{c}{(\theta')^2}||x^k_p - x^k||^2.
\]

Thus, one has

\[
\text{dist} \left( x^{k+1}, [F \leq \bar{F}] \right) \leq ||x^k_p - x^{k+1}|| \leq \left( b - \frac{c}{(\theta')^2} \right) \frac{1}{2} \text{dist} \left( x^k, [F \leq \bar{F}] \right).
\]
(ii) By semi-convexity of \( g \) and \( \bar{\varepsilon} < \min\{\frac{\varepsilon}{T}, \frac{m}{\rho}\} \), it follows from Proposition 2.3 that \( T_{D,\varepsilon}(x) \) is single-valued. Let \( T_{D,\varepsilon}(x)_p = \text{Proj}_{\{F \leq \bar{F}\}}(T_{D,\varepsilon}(x)) \). Then we see that

\[
\text{dist}\left(x, [F \leq \bar{F}]\right) \leq \|x - T_{D,\varepsilon}(x)_p\| \\
\leq \|T_{D,\varepsilon}(x) - T_{D,\varepsilon}(x)_p\| + \|x - T_{D,\varepsilon}(x)\| \\
= \text{dist}\left(T_{D,\varepsilon}(x), [F \leq \bar{F}]\right) + \text{dist}\left(x, T_{D,\varepsilon}(x)\right) \\
\leq \beta \text{dist}\left(x, [F \leq \bar{F}]\right) + \text{dist}\left(x, T_{D,\varepsilon}(x)\right).
\]

By the statement (iv) of Proposition 2.4, we have

\[
\text{dist}\left(x, [F \leq \bar{F}]\right) \leq \frac{1}{(1 - \beta)} \text{dist}\left(x, T_{D,\varepsilon}(x)\right) \leq \frac{\bar{\varepsilon}}{(1 - \beta)(m - \bar{\varepsilon})} \sqrt{\text{dist}(0, \partial_p F(x))},
\]

which shows that \( F \) satisfies the strong level-set subdifferential EB on \([\bar{F} < F < \bar{F} + \nu]\). Then by Corollary 4.1 we get that \( F \) satisfies the strong level-set Bregman proximal EB. □

**Remark 4.1.** For problem (P), if \( F \) attains the global minimum value \( F^* \) at every critical point, then solution set \( X^* = [F \leq F^*] \), \( x^k \in [F^* < F < F^* + \nu] \), and the inequality (31) with respect to \([F \leq F^*]\) becomes

\[
\text{dist}\left(x^{k+1}, X^*\right) \leq \beta \text{dist}\left(x^k, X^*\right).
\]

Observe that a convex or an invex function \( F \) satisfies (32). Furthermore, conditions such as proximal-PL, a global version of KL and proximal EB in [21] also guarantee (32).

### 5 Connections with known error bounds in literature and applications

This section examines the novelty of level-set error bounds and their relationships with existing error bounds. The established linear convergence results of VBPG allow us to exploit the novel convergence results for various existing algorithms. Although we only study the “local” version error bounds on \( \mathcal{B}(\bar{x}; \eta, \nu) \) in this section, but the same analysis used in this section can be readily extended to “global” version error bounds on \([\bar{F} < F < \bar{F} + \nu]\).

#### 5.1 Error bounds with target set \( \bar{X}_p \) or target value \( F(\bar{x}) \)

Let \( \bar{x} \in \bar{X}_p \), first we study conditions under which the distance from any vector \( x \in \mathcal{B}(\bar{x}; \eta, \nu) \) to the set \( \bar{X}_p \) is bounded by a residual function \( R_1(x) \), raised to a certain power, evaluated at \( x \). Specifically, we study the
existence of some $\gamma_1, \delta_1$, such that

$$\text{dist}^{\gamma_1}(x, \hat{x}_P) \leq \delta_1 R_1(x) \ \forall x \in \mathcal{B}(\bar{x}; \eta, \nu).$$

An expression of this kind is called a first type error bound with target set $\bar{x}_P$ for (P).

**Definition 5.1 (First type error bounds).**

(i) *(Weak metric-subregularity)* We say that $\partial_p F$ is weakly metrically subregular at $\bar{x} \in \bar{X}_P$ for the zero vector 0 if there exist $\eta$, $\nu$ and $c_2$ such that

$$\text{dist}(x, \hat{x}_P) \leq c_2 \text{dist}(0, \partial_p F(x)) \ \forall x \in \mathcal{B}(\bar{x}; \eta, \nu).$$  \hspace{1cm} (33)

(ii) *(Bregman proximal error bound)* Given a Bregman function $D$ along with $\varepsilon > 0$, we say that the Bregman proximal error bound (EB) holds at $\bar{x} \in \bar{X}_P$ if there exist $\eta$, $\nu$ and $c_3$ such that

$$\text{dist}(x, \hat{x}_P) \leq c_3 \text{dist}(x, T_{D, \varepsilon}(x)) \ \forall x \in \mathcal{B}(\bar{x}; \eta, \nu).$$  \hspace{1cm} (34)

(iii) *(Luo-Tseng error bound)* We say the Luo-Tseng error bound (EB) holds if any $\xi \geq \inf_{x \in \mathbb{R}^n} F(x)$, there exists constant $c_4 > 0$ and $\sigma > 0$ such that

$$\text{dist}(x, \hat{x}_P) \leq c_4 \|x - T_{D, \varepsilon}(x)\| \ \text{with} \ D(x, y) = \frac{\|x - y\|^2}{2}$$

whenever $F(x) \leq \xi, \|x - T_{D, \varepsilon}(x)\| \leq \sigma$.

A few remarks about (33) are in order. Metric subregularity of a set-valued mapping is a well-known notion in variational analysis. See the monograph [17] by Dontchev and Rockafellar for motivations, theory, and applications. In (33) if $\mathcal{B}(\bar{x}; \eta, \nu)$ is replaced by $\mathcal{B}(\bar{x}; \eta)$, then (33) is equivalent to metric subregularity of the set-value mapping $\partial_p F$ at $\bar{x}$ for the vector 0 (see Exercise 3H.4 of [17]) for a proof.

The following proposition provides a sufficient condition for Bregman proximal error bound.

**Proposition 5.1.** The Luo-Tseng EB condition implies the Bregman proximal EB when $g$ is semiconvex.

**Proof.** Taking $\xi > \bar{F}, \nu \geq \xi - \bar{F}$, then $[\bar{F}, \bar{F} + \nu] \subseteq [F \leq \xi]$. By Luo-Tseng EB condition, for $\xi > \bar{F}$, there are $c_4$ and $\sigma_\xi$ such that

$$\text{dist}(x, \hat{x}_P) \leq c_4 \|x - T(x)\|, \quad F(x) \leq \xi, \quad \|x - T(x)\| \leq \sigma_\xi$$

Since $T(x)$ is continuous, then for $\sigma_\xi$, there is $\hat{\eta}_\xi$ such that

$$\|T(x) - T(\bar{x})\| \leq \frac{\sigma_\xi}{2} \quad \text{when} \quad \|x - \bar{x}\| \leq \hat{\eta}_\xi.$$ 

Now let $\eta = \min\{\sigma_\xi, \hat{\eta}_\xi\}$, if $\|x - \bar{x}\| \leq \eta$, we have

$$\|x - T(x)\| \leq \|x - \bar{x}\| + \|T(x) - T(\bar{x})\| = \|x - \bar{x}\| + \|T(x) - T(\bar{x})\|.$$
Therefore \( \forall x \in \mathbb{B}(\bar{x}; \eta) \cap [\bar{F} < F \leq \bar{F} + \nu] \), we have \( \| x - T(x) \| \leq \sigma \xi \), \( F(x) \leq \xi \). By Luo-Tseng EB, we have

\[
\text{dist}(x, \bar{X}_P) \leq c_4 \| x - T(x) \|,
\]

which shows that the Bregman proximal EB holds at \( \bar{x} \). \( \square \)

We next examine the second type error bounds with target value \( F(\bar{x}) \). These error bounds are used to bound the absolute difference of any function value \( F \) at \( \bar{x} \in \bar{X}_P \) from a test set to the value \( \bar{F} = F(\bar{x}) \) by a residual function \( R_2 \). Specifically we study if there exist some \( \gamma_2, \delta_2 \) such that

\[
R_2(x) \geq \delta_2 \left( F(\bar{x}) - \bar{F}\right)^{\gamma_2} \quad \forall x \in \mathbb{B}(\bar{x}; \eta, \nu).
\]

**Definition 5.2** (Second type error bounds). (i) **Kurdyka-Łojasiewicz property** The proper lower semi-continuous function \( F \) is said to satisfy the Kurdyka-Łojasiewicz (KL) property at \( \bar{x} \) with exponent \( \alpha \in (0, 1) \), if there exist \( \nu > 0 \), \( \eta > 0 \), and \( c_S > 0 \) such that

\[
\text{dist}(0, \partial L F(x)) \geq c_S [F(x) - \bar{F}]^\alpha \quad \forall x \in \mathbb{B}(\bar{x}; \eta, \nu).
\]

(ii) **Bregman proximal gap condition** Given a Bregman function \( D \) along with \( \epsilon > 0 \), we say that the function \( F \) satisfies the Bregman proximal (BP) gap condition relative to \( D \) and \( \epsilon \) at \( \bar{x} \in \text{dom} F \) with exponent \( q \in [0, 2) \) if there exist \( \nu > 0 \), \( \eta > 0 \), and \( \mu > 0 \) such that

\[
G_{D,\epsilon}(x) \geq \mu(F(x) - \bar{F})^q \quad \forall x \in \mathbb{B}(\bar{x}; \eta, \nu),
\]

where \( G_{D,\epsilon}(x) = -\frac{1}{\epsilon} \min_{y \in \mathbb{R}^n} [\langle \nabla f(x), y - x \rangle + g(y) - g(x) + \frac{1}{\epsilon} D(x, y)] \).

### 5.2 Relationships between level-set error bounds and other error bounds

**Assumption (H):** For \( \bar{x} \in \bar{X}_P \), there is a \( \delta > 0 \) such that \( F(y) \leq F(\bar{x}) \) whenever \( y \in \bar{X}_P \) and \( \| y - \bar{x} \| \leq \delta \).

The next proposition will establish the relationships between level-set based EB and the first type EB.

**Proposition 5.2.** Suppose that Assumption (H) holds at \( \bar{x} \in \bar{X}_P \), for \( x \in \mathbb{B}(\bar{x}; \eta, \nu) \) with \( \eta \leq \delta \), then we have

\[
\text{dist}(x, [F \leq \bar{F}]) \leq \text{dist}(x, \bar{X}_P).
\]

**Proof.** By Assumption (H), for \( \eta \leq \delta \), we have that \( \bar{X}_P \cap \mathbb{B}(\bar{x}; \eta, \nu) = \emptyset \). For given \( x \in \mathbb{B}(\bar{x}; \eta, \nu) \), let \( x_p = \text{proj}_{\bar{X}_P}(x) \), then we must have \( F(x_p) \leq \bar{F} \), \( x_p \in [F \leq \bar{F}] \). Therefore, for \( x \in \mathbb{B}(\bar{x}; \eta, \nu) \), we conclude the result. \( \square \)

We are ready to present a key result on how the level-set subdifferential EB condition relates to some important notions in variational analysis and optimization.
Theorem 5.1. (KL property, Bregman proximal EB, and weak metric-subregularity imply level-set subdifferential EB) For the proper l.s.c function $F$, the following assertions hold.

(a) Suppose that $F$ satisfies the KL property at $\bar{x}$ with exponent $\alpha \in [0, 1)$ over $\mathcal{B}(\bar{x}; \eta, \nu)$. Then $F$ satisfies the level-set subdifferential EB condition at $\bar{x}$ with $\gamma = \frac{\alpha}{1-\alpha}$ over $x \in \mathcal{B}(\bar{x}; \frac{\eta}{2}, \nu)$. So $\alpha = \frac{\gamma}{1+\gamma}$. As a consequence, if $\alpha \in [0, 1/2]$, then one has $\gamma \in (0, 1]$, and if $\alpha \in (1/2, 1)$, then we have $\gamma > 1$.

(b) Suppose Assumption 1 holds, and Assumption (H) holds at $\bar{x} \in \bar{X}_P$. If one of the following condition holds

(i) the Bregman proximal EB holds at $\bar{x}$, and $g$ is semiconvex;
(ii) $\partial_p F$ is weakly metric-subregular at $\bar{x}$ for the zero vector $0$;

then the level-set subdifferential EB condition holds at $\bar{x}$ with $\gamma = 1$.

Proof. (a): By Proposition 3.16 and Theorem 3.22 of [22], we can easily get that there exists some $c_1 > 0$ such that

$$\text{dist}'(x, \{F \leq \bar{F}\}) \leq c_1 \text{dist}(0, \partial_L F(x)) \forall x \in \mathcal{B}(\bar{x}; \frac{\eta}{2}, \nu),$$

where $\gamma = \frac{\alpha}{1-\alpha}$. Since $\partial_p F(x) \subset \partial_L F(x) \forall x \in \mathbb{R}^n$, the claim is proved.

(b): (i): By Proposition 5.2 and the Bregman proximal EB, we have

$$\text{dist}(x, \{F \leq \bar{F}\}) \leq \text{dist}(x, \bar{X}_P) \leq c_3 \text{dist}(x, T_{D,\epsilon}(x)).$$

Then by (iv) of Proposition 2.4 one has

$$\text{dist}'(x, \{F \leq \bar{F}\}) \leq c_1 \text{dist}(0, \partial_p F(x)) \text{ with } c_1 = c_3 \left( \frac{\epsilon}{m - \epsilon L} \right) \sqrt{\epsilon / L},$$

(ii): By Assumption (H), for $\eta \leq \delta$, we have that $\bar{X}_P \subset \{F \leq \bar{F}\}$. For $x \in \mathcal{B}(\bar{x}; \eta, \nu)$, since $\partial_p F$ satisfies weak metric subregularity, we have

$$c_2 \text{dist}(x, \partial_p F(x)) \geq \text{dist}(x, \bar{X}_P) \geq \text{dist}'(x, \{F \leq \bar{F}\}),$$

which yields the desired result. \qed

The following proposition provides the value proximity in terms of the distance between $x$ and the set $\{F \leq \bar{F}\}$. Thanks of this proposition, we will establish the connection of level set EB with second type EB.

Proposition 5.3 (Function-value proximity in terms of level sets). Suppose that Assumptions 1 and 2 hold. If $\bar{\epsilon} < \frac{m}{L}$, then there is some $c_0 = \frac{3}{2}L + \frac{M}{m \delta} > 0$ such that the following estimation holds.

$$F(t_{D,\epsilon}(x)) - \bar{F} \leq E_{D,\epsilon}(x) - \bar{F} \leq c_0 \text{dist}^2(x, \{F \leq \bar{F}\}) \forall x \in \{F > \bar{F}\} \text{ and } t_{D,\epsilon}(x) \in T_{D,\epsilon}(x).$$
Proof. With the given choice of $\epsilon$, $T_{D,\epsilon}(x)$ is nonempty by Proposition 2.1. So $E_{D,\epsilon}(x)$ has a finite value for any given $x$. For $x \in [F > \bar{F}]$, let $x_p \in [F \leq \bar{F}]$ such that $\|x - x_p\| = \text{dist}(x, [F \leq \bar{F}])$. Since $F(x_p) \leq \bar{F}$, then

\[
F(t_{D,\epsilon}(x)) - \bar{F} \leq E_{D,\epsilon}(x) - \bar{F} \quad \text{(by (i) of Proposition 2.1)}
\]

\[
\leq \min_{y \in \mathbb{R}^n} \{f(x) + \langle \nabla f(x), y - x \rangle + g(y) + \frac{1}{\epsilon}D(x, y) - (f + g)(x_p)\}
\]

\[
\leq f(x) + \langle \nabla f(x), x_p - x \rangle + g(x_p) + \frac{1}{\epsilon}D(x, x_p) - (f + g)(x_p)
\]

\[
= f(x) - f(x_p) + \langle \nabla f(x), x_p - x \rangle + \frac{1}{\epsilon}D(x, x_p)
\]

\[
\leq \langle \nabla f(x_p) - \nabla f(x), x - x_p \rangle + \frac{L}{2}\|x - x_p\|^2 + \frac{1}{\epsilon}D(x, x_p) \quad \text{(by Assumption 1)}
\]

\[
\leq \frac{3}{2}L\|x - x_p\|^2 + \frac{M}{2\epsilon}\|x - x_p\|^2 \quad \text{(by Assumptions 1 and 2)}
\]

\[
\leq c_0\|x - x_p\|^2 = c_0\text{dist}^2(x, [F \leq \bar{F}]) \quad \text{(where } c_0 = \frac{3}{2}L + \frac{M}{2\epsilon}). \quad (35)
\]

This completes the proof. \qed

Under the assumption of semi-convexity of $g$ at $\bar{x}$, we have the following theorem, which gives an answer to the converse of statement (a) of Theorem 5.1.

**Proposition 5.4. (Level-set Bregman EB implies BP gap condition and KL property)** Suppose that Assumption 1 holds and $g$ is semi-convex. For a given Bregman function $D$ along with $\epsilon > 0$ satisfying Assumption 2 we have the following statements:

(i) If $F$ satisfies level-set Bregman EB holds at $\bar{x}$ with exponent $p$ over $\mathcal{B}(\bar{x}; \eta, \nu)$, then BP gap condition holds at $\bar{x}$ with exponent $q = \frac{1}{\min\{\frac{p}{2}, 1\}}$ over $\mathcal{B}(\bar{x}; \eta, \nu)$.

(ii) If $F$ satisfies BP gap condition at $\bar{x}$ with exponent $q$ over $\mathcal{B}(\bar{x}; \eta, \nu)$, then function $F$ has the KL property at $\bar{x}$ with exponent of $\frac{p}{2}$ over $\mathcal{B}(\bar{x}; \eta, \nu)$.

**Proof.** (i): For $x \in \mathcal{B}(\bar{x}; \eta, \nu)$, let $x_p \in [F \leq \bar{F}]$ s.t. $\|x - x_p\| = \text{dist}(x, [F \leq \bar{F}])$. We have $F(x_p) \leq F(\bar{x}) = \bar{F}$ and the estimate for term $E_{D,\epsilon}(x) - \bar{F}$ can obtained by Proposition 5.3 as following

\[
E_{D,\epsilon}(x) - \bar{F} \leq c_0\text{dist}^2(x, [F \leq \bar{F}]) \quad \text{with } c_0 = \frac{3}{2}L + \frac{M}{2\epsilon}. \quad (36)
\]

Furthermore, we obtain

\[
F(x) - \bar{F} = F(x) - E_{D,\epsilon}(x) + E_{D,\epsilon}(x) - \bar{F}
\]

\[
\leq F(x) - E_{D,\epsilon}(x) + c_0\text{dist}^2(x, [F \leq \bar{F}]) \quad \text{(by (36))}
\]

\[
\leq \epsilon G_{D,\epsilon}(x) + c_0\theta^2\text{dist}^\frac{2}{p}(x, T_{D,\epsilon}(x)) \quad \text{(by level-set Bregman EB condition)}
\]

\[
\leq \epsilon G_{D,\epsilon}(x) + c_0\theta^2 \left( \frac{2\epsilon^2}{m - \epsilon\rho} \right)^{\frac{1}{p}} (G_{D,\epsilon}(x))^{\frac{1}{p}}. \quad \text{(by (ii) of Proposition 2.4)}
\]

23
So, there is some $\mu > 0$ such that

$$G_{D,\epsilon}(x) \geq \mu (F(x) - \bar{F})^q, \quad \forall x \in \mathcal{B}(\bar{x}; \eta, \nu), \quad q = \frac{1}{\min\{\frac{1}{p}, 1\}}.$$  

The proof is completed.

(ii): By the hypothesis, the BP gap condition holds at $\bar{x}$ with exponent $q$ over $\mathcal{B}(\bar{x}; \eta, \nu)$, i.e.,

$$G_{D,\epsilon}(x) \geq \mu (F(x) - \bar{F})^q, \quad \forall x \in \mathcal{B}(\bar{x}; \eta, \nu).$$

By the assumptions for $g$, one has $\partial_p g(x) = \partial_L g(x)$ and $\partial_p F(x) = \partial_L F(x) \forall x \in \mathcal{B}(\bar{x}; \eta, \nu)$. Then by (iii) of Proposition 2.4 we have

$$G_{D,\epsilon}(x) \leq \frac{\bar{\epsilon}}{2\epsilon(m - \bar{\epsilon}\rho)} \text{dist}^2(0, \partial_L F(x)).$$

It follows that

$$\left(\frac{2\epsilon}{\bar{\epsilon}}\right)(m - \bar{\epsilon}\rho)\mu \left(F(x) - \bar{F}\right)^q \leq \left[\text{dist}(0, \partial_L F(x))\right]^2,$$

which implies that

$$\text{dist}(0, \partial_L F(x)) \geq \sqrt{\left(\frac{2\epsilon}{\bar{\epsilon}}\right)(m - \bar{\epsilon}\rho)\mu \left(F(x) - \bar{F}\right)^q}.$$  

The assertion is justified. \hfill $\square$

5.3 Examples illustrating the novelty of the level-set subdifferential EB condition

The following examples show that KŁ property, weak metric subregularity or Bregman proximal EB are not be necessary for level-set subdifferential EB.

**Example 5.1** (level-set subdifferential EB does not imply weak metric subregularity). Let $\bar{x} = (0, 0)^T$ and $F: \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$F(x) := \begin{cases} x_1^2 - x_2^3 & \text{if } x_2 > 0 \\ x_2^3 & \text{if } x_2 \leq 0. \end{cases}$$

By some direct calculations, we get that $F$ is a lower semicontinuous function, $\bar{x}$ is the unique proximal critical point,

$$[F \leq 0] = \mathbb{R} \times (-\infty, 0] \cup \{(x_1, x_2)^T | x_1^2 \leq x_2^3\}$$

and

$$\nabla F(x) = (2x_1, -3x_2^2)^T \text{ when } x \notin [F \leq 0].$$
Since for each $x$ closed enough to $\bar{x}$ with $v > F(x) > 0$ and some $v > 0$ one has

$$d(x, [F \leq 0]) \leq |x_1| \leq \frac{1}{2} ||\nabla F(x)||,$$

$F$ satisfies the level-set subdifferential EB condition at $\bar{x}$ with the exponent $\gamma = 1$. However, $F$ is not weakly metric-subregular at $\bar{x}$ for the zero vector $0$, since for $x_n = (\frac{1}{n}, \frac{1}{n})$ which is not in $[F \leq 0]$ one has

$$\frac{\sqrt{x_{n,1}^2 + x_{n,2}^2}}{\sqrt{4x_{n,1}^2 + 9x_{n,2}^4}} \to \infty \text{ as } n \to \infty.$$

**Example 5.2** (level-set subdifferential EB does not imply KL property). (See Example 3.19 of [22]) Let $F : \mathbb{R} \to \mathbb{R}$ be given by

$$F(x) := \begin{cases} 
0 & \text{if } x \leq 0 \\
 x^2 + \frac{1}{n} - \frac{1}{n^2} & \text{if } \frac{1}{n} < x \leq \frac{1}{n-1}, n = 3, 4, \cdots, \\
 x^2 + \frac{1}{4} & \text{if } x > \frac{1}{2}.
\end{cases}$$

In Example 3.19 of [22], the authors have obtained that $F$ is lower semicontinuous and for each $x > 0$

$$\partial F(x) = \begin{cases} 
[2x, \infty) & \text{if } x = \frac{1}{n-1}, n = 3, 4, \cdots, \\
 2x & \text{otherwise}.
\end{cases}$$

Let $\bar{x} = 0$ be the reference point. Then we easily get that $F$ satisfies the level-set subdifferential EB condition at $\bar{x}$ with the exponent $\gamma = 1$. However, $F$ does not have KL property at $\bar{x}$ with any exponent $\alpha \in [0, 1)$, since for $x_n = \frac{1}{n-1}$ one has

$$d(0, \partial F(x_n)) \leq \frac{2}{n-1} \leq \frac{2n^\alpha}{n-1} \to 0 \text{ as } n \to \infty.$$

Moreover, we can verify that $F$ is not semi-convex. Indeed, for any given $\frac{1}{n} < x < \frac{1}{n-1}$ with any positive integer $n \geq 3$ and any positive constant $\rho > 0$, one easily obtains that the following inequality does not hold when $x' < x$ and $x'$ sufficiently closes to $x$:

$$\frac{\rho + 2}{2} x'^2 \geq \frac{\rho + 2}{2} x^2 + (2x, x' - x),$$

which implies that $F$ is not semi-convex by Proposition 8.12 of [33].

**Example 5.3** (level-set subdifferential EB does not imply the Bregman proximal EB). Let $\bar{x} = (0, 0)^T$ and $F : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$F(x) := \begin{cases} 
x_1^2 - x_2^3 & \text{if } x_2 > 0 \\
x_2^3 & \text{if } x_2 \leq 0.
\end{cases}$$
Figure 1: The relationships among the notions of the level-set subdifferential EB, subregularity of subdifferential, Bregman proximal EB, KL property, level-set Bregman EB and Bregman gap condition

From Example 1.1, we get that \( F \) satisfies the level-set subdifferential EB condition at \( \bar{x} \) with the exponent \( \gamma = 1 \). Moreover, it is easy to verify that Assumption (H) holds. By the first-order optimality condition, we can easily calculate that \( T_{D,1}(x) = \{ (x_1, 0)^T \} \) with \( D(x, y) = \frac{1}{2} \| x - y \|^2 \) for \( x \notin [F \leq 0] \) sufficiently closing to \( \bar{x} \) and \( x_1 > 2x_2 > 0 \). However, \( F \) does not have the Bregman proximal EB at \( \bar{x} \), since for \( x_n = (\frac{1}{n}, \frac{1}{3n}) \) which is not in \( [F \leq 0] \) one has

\[
\sqrt{\frac{x_{n,1}^2 + x_{n,2}^2}{|x_{n,2}|}} \to \infty \text{ as } n \to \infty.
\]

Figure 1 summarizes the main results of this section.

5.4 Applications of level-set subdifferential error bounds

Application 1: Linear convergence of regularized Jacobi method
In big data applications, the regularizer $g$ in problem (P) may have block separable structures, i.e., $g(x) = \sum_{i=1}^{N} g_i(x_i), x_i \in \mathbb{R}_i^n$. In this setting, (P) can be specified as

$$\min_{x \in \mathbb{R}^n} f(x_1, ..., x_n) + \sum_{i=1}^{N} g_i(x_i)$$

(37)

If we take $K^k(x) = \sum_{i=1}^{N} f\left(R^k_i(x)\right) + \frac{\alpha}{2}||x_i - x^k_i||^2$ and $D^k(x, y) = K^k(y) - K^k(x) - \langle \nabla K^k(x), y - x \rangle$, where $R^k_i \triangleq (x^k_1, ..., x^k_{i-1}, x_i, x^k_{i+1}, ..., x^k_n)$. Thus VBPG become a regularized Jacobi algorithm. Recently, G. Bajac [4] provided the linear convergence of regularized Jacobi algorithm under quadratic growth condition for full convex problem (37).

By Theorem 3.1 for full nonconvex problem (37), the VBPG method provides the linear convergence under the level-set subdifferential EB condition at the point $\bar{x} \in \overline{X}_L$. For the convex problem (37), the quadratic growth condition is equivalent to strongly level set subdifferential EB condition see Theorem 3.3 and Corollary 3.6 [18] for more details. Together with Theorem 4.1, we can show that the quadratic growth condition is also necessary for linear convergence in the sense of (32).

**Application 2: Linear convergence under proximal-PL inequality and Bregman proximal gap**

[21] proposes the concept of proximal-PL inequality for solving problem (P) where $F$ is invex function, $g$ is convex, i.e., there is $\mu > 0$ such that the following inequality holds:

$$\frac{1}{2}D_g(x, L) \geq \mu (F(x) - F^*)$$

where $F^*$ is the global minimum value and

$$D_g(x, \alpha) = -2\alpha \min_{y \in \mathbb{R}^n} \left[ \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}||y - x||^2 + g(y) - g(x) \right],$$

which is a global version of Bregman proximal gap function with $D^k(x, y) = \frac{\|x - y\|}{2}$. [21] proves the sequence \{F(x^k)\} generated by PG method with a step size of $1/L$ linearly converges to $F^*$ under proximal-PL inequality. For the fully nonconvex case, Theorem 3.1 shows the Q-linear convergence of \{F(x^k)\} and the R-linear convergence of \{x^k\} under the Bregman proximal gap condition, which is weaker than the proximal-PL inequality. Observe that the proximal PL inequality implies that every critical point achieves an optimum $F^*$, and the strong level-set subdifferential EB condition holds. If $g$ is semi-convex, by Theorem 4.1 the proximal PL inequality is also a necessary condition for linear convergence in the sense of (32).

**Application 3: Linear convergence under KL property**

Various variable metric proximal gradient methods (VMPG) are provided in following algorithms for problem (P)

$$x^{k+1} \rightarrow \min_{x} \langle \nabla f(x^k), x - x^k \rangle + g(x) + \frac{1}{2}\|x - x^k\|^2_{B^k},$$

27
where $B_k$ is positive definite matrix. E. Chonzennx et. al. [12] proposed an inexact version of VMPG algorithm for problem (P) where $g$ is convex. And the authors also provided linear convergence of VMPG under KL property with exponent $\frac{1}{2}$. Noted that VMPG is the special case of VBPG with $D^k = \frac{\|x^k - y^k\|^2}{2}$, Theorem[31] states that VMPG has the linear convergence for $\{x^k\}$ and $\{F(x^k)\}$ under level-set subdifferential EB condition. Moreover, the strong level-set subdifferential error bound condition on $[\bar{F} < F < \bar{F} + \nu]$ is necessary and sufficient for linear convergence in the sense of (31). Mention that if $g$ is semi-convex, level-set subdifferential EB condition with exponent $\gamma = 1$ is equivalent to KL exponent $\frac{1}{2}$ condition.

6 Sufficient conditions for the level-set subdifferential EB condition to hold on $\mathcal{B}(\bar{x}; \eta, \nu)$ with $\bar{x} \in \bar{X}_P$

This section provides sufficient conditions to guarantee level-set subdifferential EB condition at $\bar{x} \in \bar{X}_P$ on $\mathcal{B}(\bar{x}; \eta, \nu)$, where $\bar{x}$ is a proximal critical point of $F = f + g$

First, we provide some new notions on relaxed strong convexity of function $f$ on $\mathcal{B}(\bar{x}; \eta)$. Given $z \in \mathcal{B}(\bar{x}; \eta)$, for brevity, we denote $Proj_{\mathcal{B}(\bar{x}; \eta) \cap \bar{X}_P}(z)$ by $z_p$. The following notations can be viewed as the local version for that in H. Karimi et al’s and I. Necoara et al’s paper [21], [29] respectively.

1. Local strong-convexity (LSC) on $\mathcal{B}(\bar{x}; \eta)$:
   
   $$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2, \quad \forall x, y \in \mathcal{B}(\bar{x}; \eta).$$

2. Local essentially-strong-convexity at $\bar{x}_p$ (LESC) on $\mathcal{B}(\bar{x}; \eta)$:
   
   $$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2, \quad \forall x, y \in \mathcal{B}(\bar{x}; \eta) \text{ with } \bar{x}_p = \bar{y}_p.$$

3. Local weak-strong-convexity at $\bar{x}_p$ (LWSC) on $\mathcal{B}(\bar{x}; \eta)$:
   
   $$f(\bar{x}_p) \geq f(x) + \langle \nabla f(x), \bar{x}_p - x \rangle + \frac{\mu}{2}\|ar{x}_p - x\|^2, \quad \forall x \in \mathcal{B}(\bar{x}; \eta).$$

4. Local quadratic-gradient-growth (LQGG) at $\bar{x}_p$ on $\mathcal{B}(\bar{x}; \eta)$:
   
   $$\langle \nabla f(x) - \nabla f(\bar{x}_p), x - \bar{x}_p \rangle \geq \mu\|ar{x}_p - x\|^2, \quad \forall x \in \mathcal{B}(\bar{x}; \eta).$$

For the case $g = 0$, the following two notions are introduced.

5. Local restricted secant inequality (LRSI):
   
   $$\langle \nabla f(x), x - \bar{x}_p \rangle \geq \mu\|x - \bar{x}_p\|^2, \quad \forall x \in \mathcal{B}(\bar{x}; \eta).$$
6. Local Polyak-Łojasiewicz (LPL) inequality:

\[ \frac{1}{2} \| \nabla f(x) \|^2 \geq \mu \left( f(x) - f(\bar{x}) \right), \quad \forall x \in B(\bar{x}; \eta). \]

It’s easy to show that the following implications hold for the function \( f \) on \( B(\bar{x}; \eta) \).

\((LSC) \Rightarrow (LES C) \Rightarrow (LWSC)\).

For the case \( g = 0 \), the LQGG reduce to the local restricted secant inequality (LRSI). So we have:

\((LWSC) \Rightarrow (LRS I) \Rightarrow (LPL)\) (if \( g = 0 \)).

Along with Assumption 3, the following proposition allow us to establish the level-set subdifferential EB of \( F \), when \( g \) is uniformly prox-regular.

**Proposition 6.1** (Sufficient conditions for weak metric subregularity). Suppose \( \bar{x} \in \bar{x}_p \), \( g \) is uniformly prox-regular around \( \bar{x} \in \text{dom} \ g \) with modulus \( \rho \). If one of the following conditions holds

(i) \( f \) is local weak strongly convex (LWSC) at \( \bar{x}_p \) with modulus \( \mu \) and \( \mu > \rho \) on \( B(\bar{x}; \eta) \).

(ii) \( f \) satisfies local quadratic gradient growth condition (LQGG) at \( \bar{x}_p \) with modulus \( \mu \) and \( \mu > \rho \) on \( B(\bar{x}; \eta) \),

then \( F \) satisfies the weak metric subregularity condition at \( \bar{x} \).

**Proof.** (i): If \( f \) is LWSC at \( \bar{x}_p \) on \( B(\bar{x}; \eta) \), then we have

\[ f(\bar{x}_p) \geq f(x) + \langle \nabla f(x), \bar{x}_p - x \rangle + \frac{\mu}{2} \| \bar{x}_p - x \|^2. \] \hspace{1cm} (38)

Since \( g \) is uniformly prox-regular around \( \bar{x} \) with \( \rho \), then \( \partial_pg(x) = \partial_Lg(x) \) and

\[ g(\bar{x}_p) \geq g(x) + \langle \xi, \bar{x}_p - x \rangle - \frac{\rho}{2} \| \bar{x}_p - x \|^2, \quad \forall \xi \in \partial_pg(x). \] \hspace{1cm} (39)

Adding inequalities (38) and (39), we obtain

\[ F(\bar{x}_p) = F(\bar{x}) = F_\xi \geq F(x) + \langle \nabla f(x) + \xi, \bar{x}_p - x \rangle + \frac{(\mu - \rho)}{2} \| \bar{x}_p - x \|^2. \]

and

\[ \langle \nabla f(x) + \xi, x - \bar{x}_p \rangle \geq \frac{(\mu - \rho)}{2} \| \bar{x}_p - x \|^2, \quad \forall \xi \in \partial_pg(x), \quad \forall x \in B(\bar{x}; \eta, \nu). \]

Using Cauchy-Schwartz on above inequality, we conclude

\[ \text{dist}(0, \partial_pF(x)) \geq \frac{(\mu - \rho)}{2} \| \bar{x}_p - x \| \geq \frac{(\mu - \rho)}{2} \text{dist}(x, \bar{x}_p), \quad \forall x \in B(\bar{x}; \eta, \nu), \]
which yields the desired results.

(ii): If $f$ is LQGG at $\bar{x}_p$ on $\mathbb{B}(\bar{x}; \eta)$, then we have

$$\langle \nabla f(x) - \nabla f(\bar{x}_p), x - \bar{x}_p \rangle \geq \mu \|\bar{x}_p - x\|^2, \quad \forall x \in \mathbb{B}(\bar{x}; \eta, \nu).$$

Since $g$ is semi-convex, we have

$$\langle u - v, x - \bar{x}_p \rangle \geq -\rho \|x - \bar{x}_p\|^2, \quad \forall u \in \partial P g(x), \quad \forall v \in \partial P g(\bar{x}_p).$$

Adding the above two inequalities for $x \in \mathbb{B}(\bar{x}; \eta)$, we obtain

$$\langle (\nabla f(x) + u) - (\nabla f(\bar{x}_p) + v), x - \bar{x}_p \rangle \geq \frac{\mu}{2} \|\bar{x}_p - x\|^2.$$  

Since $\bar{x}_p$ is a proximal critical point, $0 = \nabla f(\bar{x}_p) + v$ for some $v \in \partial P g(\bar{x}_p)$. With this choice of $v$, the last above inequality yields

$$\langle \nabla f(x) + u, x - \bar{x}_p \rangle \geq \frac{(\mu - \rho)}{2} \|\bar{x}_p - x\|^2 \geq \frac{(\mu - \rho)}{2} \text{dist}(x, \bar{X}_p), \quad \forall u \in \partial P g(x), \quad \forall x \in \mathbb{B}(\bar{x}; \eta, \nu).$$

This is enough for the proof of proposition.  \qed

Now we are ready to present a main result on sufficient conditions to guarantee that the level-set subdifferential EB holds at $\bar{x}$ on $\mathbb{B}(\bar{x}, \eta, \nu)$, where $\bar{x}$ is an accumulation point of the sequence $\{x^k\}$ generated by VBPG.

**Theorem 6.1** (Sufficient conditions for the existence of a level-set subdifferential EB). Consider problem (P). Suppose that Assumption 1 and Assumption 2 hold, and $\bar{x} \in \bar{X}_p$. If one of following conditions hold, then $F$ satisfies the level-set subdifferential EB condition at $\bar{x}$ on $\mathbb{B}(\bar{x}; \eta, \nu)$.

(i) $F = f + g$ satisfies the KL exponent at $\bar{x}$ on $\mathbb{B}(\bar{x}; \eta, \nu)$ at $\bar{x}$.

(ii) $F = f + g$ satisfies Bregman proximal EB condition, Assumption (H) holds, $g$ is semi-convex or $g$ is uniformly prox-regular around $\bar{x}$, $x \in \mathbb{B}(\bar{x}, \frac{\eta}{2}, \nu)$ with $N \geq \frac{2\rho m}{\eta^2 \ell} / \left(\frac{\eta}{2}\right)^2$ satisfies Property (A).

(iii) $F = f + g$ satisfies weak metric subregularity at $\bar{x}$ and Assumption (H) holds.

(iv) With $g = 0$, $f = F$ satisfies the (LPL) inequality on $\mathbb{B}(\bar{x}; \eta)$.

**Proof.** (i)-(iii) see Theorem 5.1 (iv) For this case, the (LPL) inequality implies the KL property. Then the assertion follows from Proposition 5.1. \qed

**Remark 6.1.** For the optimization problem (P), if we consider the global solution $X^*$ instead of $\bar{X}_p$, then Assumption (H) is automatically satisfied. Weak metric subregularity and the Bregman proximal EB imply the level-set subdifferential EB.
Remark 6.2. From the definition of a level-set subdifferential EB, suppose that $\bar{x}$ is a critical point. If $x \in \mathcal{S}(\bar{x}; \eta, \nu)$ is also a critical point, then $0 \in \partial_P F(x)$ and $\text{dist}(x, \partial F(x)) = 0$. This fact follows $F(x) \leq F(\bar{x})$, which implies Assumption (H) is a necessary condition for a level-set subdifferential EB to hold. We mention that Assumption (H) is also necessary for KL property.

References

[1] Attouch, H., Bolte, J. and Svaiter, B. F. (2013). Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Mathematical Programming, 137(1-2), 91-129.

[2] Aybat, N. S. and Iyengar, G. (2014). A unified approach for minimizing composite norms. Mathematical Programming, 144(1-2), 181-226.

[3] Azé, D. and Corvellec, J. N. (2017). Nonlinear error bounds via a change of function. Journal of Optimization Theory and Applications, 172(1), 9-32.

[4] Banjac, G., Margellos K. and Goulart P. J. (2018). On the convergence of a regularized Jacobi algorithm for convex optimization. IEEE Transactions on Automatic control, 63(4), 1113-1119.

[5] Bernard, F. and Thibault, L. (2005). Uniform prox-regularity of functions and epigraphs in Hilbert spaces. Nonlinear analysis, 60, 187-207.
[6] Bolte, J., Daniilidis, A., Ley, O. and Mazet, L. (2010). Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Transactions of the American Mathematical Society, 362(6), 3319-3363.

[7] Bonettini, S., Loris, I., Porta, F. and Prato, M. (2016). Variable metric inexact line-search-based methods for nonsmooth optimization. SIAM Journal on Optimization, 26(2), 891-921.

[8] Burke J. V. and Deng S. (2002). Weak sharp minima revisited, part I: Basic theory. Control and Cybernetics 31, 439-469.

[9] Burke J. V. and Ferris M. C. (1993). Weak sharp minima in mathematical programming. SIAM Journal on Control and Optimization 31, 1340-1359.

[10] Candés, E. J. and Tao, T. (2005). Decoding by linear programming. IEEE transactions on information theory, 51(12), 4203-4215.

[11] Carpentier, P. and Cohen, G. (2017). Décomposition-coordination en optimisation déterministe et stochastique. Springer Berlin Heidelberg.

[12] Chouzenoux, E., Pesquet, J. C. and Repetti, A. (2014). Variable metric forwardbackward algorithm for minimizing the sum of a differentiable function and a convex function. Journal of Optimization Theory and Applications, 162(1), 107-132.

[13] Cohen, G. (1980). Auxiliary problem principle and decomposition of optimization problems. Journal of Optimization Theory and Applications, 32(3), 277-305.

[14] Cohen, G. and Zhu D. (1984). Decomposition and coordination methods in large scale optimization problems: The nondifferentiable case and the use of augmented Lagrangians. Advances in Large Scale Systems, 1, 203-266.

[15] Cromme. L. (1978). Strong Uniqueness. Numerische Mathematik, 29, 179-193.

[16] Donoho, D. L. (2006). Compressed sensing. IEEE Transactions on Information Theory, 52(4), 1289-1306.

[17] Dontchev, A. and Rockafellar, R.T. (2009). Implicit Functions and Solution Mappings. Springer Science & Business Media.

[18] Drusvyatskiy, D. and Lewis, A. S. (2018). Error bounds, quadratic growth, and linear convergence of proximal methods. Mathematics of Operations Research, 43(3), 919-948.

[19] Frankel, P., Garrigos, G. and Peypouquet, J. (2015). Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates. Journal of Optimization Theory and Applications, 165(3), 874-900.
[20] Hoffman, A. J. (1952). On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards, 49*, 263-265.

[21] Karimi, H., Nutini, J. and Schmidt, M. (2016). Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 795-811). Springer, Cham.

[22] Kruger Alexander Y, Lopez Marco A., Yang X.Q. and Zhu J. X. (2019). Holder error bounds and Holder calmness with applications to convex semi-infinite optimization. *Set-Valued and Variational Analysis, 27*, 995-1023.

[23] Li, G. and Pong, T. K. (2018). Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of Computational Mathematics, 18*(5), 1199-1232.

[24] Lions, P. L. and Mercier, B. (1979). Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis, 16*(6), 964-979.

[25] Lojasiewicz. S. (1963). A topological property of real analytic subsets (in French). *Coll du. CNRS. Les Equations Aux derivees Partielles*, 87-89.

[26] Luo, Z. Q. and Tseng, P. (1992). Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM Journal on Optimization, 2*(1), 43-54.

[27] Ma, S. and Aybat, N. S. (2018). Efficient optimization algorithms for robust principal component analysis and its variants. *Proceedings of the IEEE, 106*(8), 1411-1426.

[28] Mordukhovich, B. S. (2006). *Variational analysis and generalized differentiation I: Basic theory* (Vol. 330). Springer Science & Business Media.

[29] Necoara, I., Nesterov, Y. and Glineur, F. (2018). Linear convergence of first order methods for non-strongly convex optimization. *Mathematical Programming, 1*-39.

[30] Nesterov, Y. (2013). *Introductory lectures on convex optimization: A basic course* (Vol. 87). Springer Science & Business Media.

[31] Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables* (Vol. 30). SIAM.

[32] Pang, J. S. (1997). Error bounds in mathematical programming. *Mathematical Programming, 79*, 299-332.

[33] Polyak. B. T. (1963). Gradient methods for minimizing functionals (in Russian). *Zhurnal Vychislitel'noi Matematikii Matematicheskoi Fiziki*, 643-653.
[34] Robinson, S. M. (1980). Some continuity properties of polyhedral multifunctions. *Mathematics of Operations Research*, 5, 206-214.

[35] Rockafellar, R. T. and Wets, R. J. B. (2009). *Variational analysis* (Vol. 317). Springer Science & Business Media.

[36] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, 267-288.

[37] Tseng, P. and Yun, S. (2009). A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming*, 117(1-2), 387-423.

[38] Wang, X., Ye, J., Yuan, X., Zeng, S. and Zhang, J. (2018). Perturbation techniques for convergence analysis of proximal gradient method and other first-order algorithms via variational analysis. *arXiv preprint arXiv:1810.10051*.

[39] Zhang, H. (2020). New analysis of linear convergence of gradient-type methods via unifying error bound conditions. *Mathematical programming*, 180, 371-416.

[40] Zhu, D. and Marcotte, P. (1994). An extended descent framework for variational inequalities. *Journal of Optimization Theory and Applications*, 80(2), 349-366.

[41] Zhu, D. and Marcotte, P. (1995). Coupling the auxiliary problem principle with descent methods of pseudoconvex programming. *European Journal of Operational Research*, 83(3), 670-685.

[42] Zhu, D. and Deng, S. (2019). A Variational Approach on Level sets and Linear Convergence of Variable Bregman Proximal Gradient Method for Nonconvex Optimization Problems. *arXiv preprint arXiv:1905.08445*.