Extending the Convergence Results of Nesterov's Acceleration using the Proximal Gradient Gap

Author 1 Name, Author 2 Name *

March 2, 2020

This paper is currently in draft mode. Check source to change options. [1]

Abstract

This paper gives definitive answers to open problems in Necoara et al. [2]. Linear convergence of Nesterov's accelerated gradient method is possible in a boarder context of functions.

2010 Mathematics Subject Classification: Primary 47H05, 52A41, 90C25; Secondary 15A09, 26A51, 26B25, 26E60, 47H09, 47A63. **Keywords:**

1 Introduction

Notations. Unless specified, our ambient space is $\mathbb{R}^n$ with Euclidean norm $\|\cdot\|$. Let $C \subseteq \mathbb{R}^n$, $\Pi_C(\cdot)$ denotes the projection onto the set C, i.e: the closest point in C to another point in $\mathbb{R}^n$. We denote δ_C to be the indicator function for the set C. For a function of $F = f + g$, and a $B \geq 0$ where f is $\mathcal{C}^1$ differentiable, and g is l.s.c, we consider the proximal gradient operator:

$$
\begin{aligned}
T_B(x) &= \underset{z}{\operatorname{argmin}} \left\{ g(z) + f(x) + \langle \nabla f(x), z - x\rangle + \frac{B}{2}\|x - z\|^2 \right\} \\
&= \operatorname{prox}_{B^{-1}g}(x - B^{-1}\nabla f(x)).
\end{aligned}
$$

*Subject type, Some Department of Some University, Location of the University, Country. E-mail: `author.name@university.edu`.

We also define the gradient mapping operator $\mathcal{G}_B(x) = B^{-1}(x - T_B(x))$.

Major results.

2 Precursors materials for our proofs of linear convergence

The following two definitions defines the accelerated proximal gradient algorithm.

{def:st-apg} **Definition 2.1 (similar triangle form of accelerated proximal gradient)** *The definition is about* $((\alpha_k)_{k\ge 0}, (q_k)_{k\ge 0}, (B_k)_{k\ge 0}, (y_k)_{k\ge 0}, (x_k)_{k\ge -1}, (v_k)_{k\ge -1})$. *These sequences satisfy:*

(i) $x_{-1}, y_{-1} \in \mathbb{R}^n$ *are arbitrary initial condition of the algorithm;*
(ii) $(q_k)_{k\ge 1}$ *be a sequence such that* $q_k \in [0, 1)$ *for all* $k \ge 1$*;*
(iii) $(\alpha_k)_{k\ge 1}$ *be a sequence such that* $\alpha_0 \in (0, 1]$*, and for all* $k \ge 1$ *it has* $\alpha_k \in (q_k, 1)$*;*
(iv) $(B_k)_{k\ge 0}$ *has* $B_k \ge 0$*, it's a nonnegative sequence.*

Then an algorithm satisfies the similar triangle form of Nesterov's accelerated gradient if it generates iterates $(y_k, x_k, v_k)_{k\ge 1}$ *such that for all* $k \ge 0$*:*

$$
\begin{aligned}
y_k &= \left(\frac{\alpha_k - q_k}{1 - q_k}\right) v_{k-1} + \left(\frac{1 - \alpha_k}{1 - q_k}\right) x_{k-1}, \\
x_k &= T_{B_k}(y_k), D_f(x_k, y_k) \le \frac{B_k}{2}\|x_k - y_k\|^2, \\
v_k &= x_{k-1} + \alpha_k^{-1}(x_k - x_{k-1}).
\end{aligned}
$$

Remark 2.2 If we choose $a_k = 1$ then the definition is equivalent to proximal gradient algorithm, with $v_k = x_k$ for all $k \ge 0$.

{def:rlx-momentum-seq} **Definition 2.3 (relaxed momentum sequence)** *The following definition is about sequences* $((\alpha_k)_{k\ge 0}, (q_k)_{k\ge 0}$. *They satisfy:*

(i) $(q_k)_{k\ge 0}$ *is a sequence such that* $q_k \in [0, 1)$ *for all* $k \ge 0$*;*
(ii) $(\alpha_k)_{k\ge 0}$ *has* $\alpha_k \in (0, 1]$*, and for all* $k \ge 0$.

{def:pg-gap} **Definition 2.4 (proximal gradient gap)** *Let* $F = f + g$ *where* f *is* L *Lipschitz smooth and* g *is convex. Then the proximal gradient mapping* $T_B(x) = \operatorname{prox}_{B^{-1}g}(x - B^{-1}\nabla f(x))$ *is a singleton, and* $\operatorname{dom} T_B = \mathbb{R}^n$. *Let* μ, B *be parameters such that* $B > \mu \ge 0$. *We define the proximal gradient gap* $\mathcal{E}(z, y, \mu, B)$ *is the mapping:*

$$
\mathcal{E}(z, y, \mu, B) := F(z) - F(T_B(y)) - \langle B(y - T_B(y)), z - y\rangle - \frac{\mu}{2}\|z - y\|^2 - \frac{B}{2}\|y - T_B(y)\|^2.
$$

Remark 2.5 This expression is the same as the proximal gradient inequality up to a negative sign, after moving everything to one side.

For the sum of smooth and nonsmooth objective, a lot can be said about the proximal gradient gap of the function.

{lemma:pg-under-cnvx}

Lemma 2.6 (proximal gradient gap under convexity)

3 Deriving the proximal gap inequality

To derive the convergence rate of algorithm satisfying Definition 2.1, 2.3, we leverage Definition 2.4. The first two subsections prepare for the results and the convergence results are derived by the end.

The following assumption is about the Proximal Gradient gap.

{ass:pggap-dscnt}

Assumption 3.1 (generic assumptions for proximal gradient gap inequality) The following assumption is about $(F, f, g, \mathcal{E}, \mu, L)$, it is the configuration needed to derive the convergence rate of algorithms that satisfy Definition 2.1. We assume that there exists $\mu \geq 0$ such that the followings are true:

{ass:pggap-dscnt-itm1}
{ass:pggap-dscnt-itm2}

(i) Let $F = f + g$ where f is L Lipschitz smooth and, g is closed convex and proper.
(ii) $\forall y \in \mathbb{R}^n, \exists B \geq 0 \; \exists \bar{y}$ such that $\mathcal{E}(\bar{y}, y, \mu, B) \geq 0$.

Remark 3.2 Note that:

(i) If the function f is convex and smooth, g is convex then, all conditions are satisfied for $\mu = 0$, and for all $\bar{y} \in \mathbb{R}^n$.

(ii) If $\mu \geq 0$ satisfies item (ii), (iii), then it also satisfies for all $\tilde{\mu}$ such that $0 \leq \tilde{\mu} \leq \mu$. For the best convergence rate, the largest such μ is of our interest.

Using the above assumption, in this section, we can build up to Lemma 3.7. The lemma is crucially important and, it gives us the inequality on what happen between iterates x_k, v_k and v_{k-1}, x_{k-1} generated by algorithm satisfying Definition 2.1.

{lemma:st-iterates-alt-form-part1}

Lemma 3.3 (equivalent representations of the iterates part I) *Suppose that the sequences of $\alpha_k, q_k, y_k, v_k, x_k, B_k$ satisfy the similar triangle form, then for all $k \geq 0$ the iterates v_k admits the following equivalent representations:*

$$v_k = v_{k-1} + \alpha_k^{-1} q_k (y_k - v_{k-1}) - \alpha_k^{-1} B_k^{-1} \mathcal{G}_{B_k}(y_k).$$

Proof. Consider all $k \geq 1$. The relations are direct, immediately from the update rule in Definition 2.1 of y_k we have

(a) $(\alpha_k - 1)x_{k-1} = (\alpha_k - q_k)v_{k-1} - (1 - q_k)y_k$.
(b) $x_k = y_k - B_k^{-1}\mathcal{G}_{B_k}(y_k)$.

Using the above and the update rule for v_k in Definition 2.1.

$$
\begin{aligned}
v_k &= x_{k-1} + \alpha_k^{-1}(x_k - x_{k-1}) \\
&= (1 - \alpha_k^{-1})x_{k-1} + \alpha_k^{-1}x_k \\
&= \alpha_k^{-1}(\alpha_k - 1)x_{k-1} + \alpha_k^{-1}x_k \\
&\underset{\text{(a)}}{=} \alpha_k^{-1}(\alpha_k - q_k)v_{k-1} - \alpha_k^{-1}(1 - q_k)y_k + \alpha_k^{-1}x_k \\
&\underset{\text{(b)}}{=} (1 - \alpha_k^{-1}q_k)v_{k-1} - (\alpha_k^{-1} - \alpha_k^{-1}q_k)y_k + \alpha_k^{-1}(y_k - B_k^{-1}\mathcal{G}_{B_k}(y_k)). \\
&= (1 - \alpha_k^{-1}q_k)v_{k-1} + \alpha_k^{-1}q_k y_k - \alpha_k^{-1}B_k^{-1}\mathcal{G}_{B_k}(y_k) \\
&= v_{k-1} + \alpha_k^{-1}q_k(y_k - v_{k-1}) - \alpha_k^{-1}B_k^{-1}\mathcal{G}_{B_k}(y_k).
\end{aligned}
$$

■

{lemma:st-iterates-alt-form-part2} **Lemma 3.4 (equivalent representations of the iterates part II)** *Suppose the sequences of $\alpha_k, q_k, y_k, v_k, x_k, B_k$ satisfy the similar triangle form, then for all $k \geq 0$ the iterates v_k admits the following equivalent representations:*

$$
y_k = x_{k-1} + (1 - q_k)^{-1}(\alpha_{k-1}^{-1} - 1)(\alpha_k - q_k)(x_{k-1} - x_{k-2}).
$$

Proof. For all $k \geq 1$, from the update rules in Definition 2.1:

$$
\begin{aligned}
(1 - q_k)y_k &= (\alpha_k - q_k)v_{k-1} + (1 - \alpha_k)x_{k-1} \\
&= (\alpha_k - q_k)\left(x_{k-2} + \alpha_{k-1}^{-1}(x_{k-1} - x_{k-2})\right) + (1 - \alpha_k)x_{k-1} \\
&= (\alpha_k - q_k)((1 - \alpha_{k-1}^{-1})x_{k-2} + \alpha_{k-1}^{-1}x_{k-1}) + (1 - \alpha_k)x_{k-1} \\
&= (\alpha_k - q_k)(1 - \alpha_{k-1}^{-1})x_{k-2} + \left(\frac{\alpha_k - q_k}{\alpha_{k-1}} + 1 - \alpha_k\right)x_{k-1}.
\end{aligned}
$$

Divide by $(1 - q_k)$ on both sides yield:

$$
\begin{aligned}
y_k &= \frac{(\alpha_k - q_k)(1 - \alpha_{k-1}^{-1})}{1 - q_k}x_{k-2} + \left(\frac{\alpha_k - q_k}{\alpha_{k-1}(1 - q_k)} + \frac{1 - \alpha_k}{1 - q_k}\right)x_{k-1} \\
&= \frac{(\alpha_k - q_k)(1 - \alpha_{k-1}^{-1})}{1 - q_k}x_{k-2} + \left(\frac{(\alpha_{k-1}^{-1} - 1)(\alpha_k - q_k) + \alpha_k - q_k + 1 - \alpha_k}{1 - q_k}\right)x_{k-1} \\
&= \frac{(\alpha_k - q_k)(1 - \alpha_{k-1}^{-1})}{1 - q_k}x_{k-2} + \left(\frac{(\alpha_{k-1}^{-1} - 1)(\alpha_k - q_k)}{1 - q_k} + 1\right)x_{k-1} \\
&= x_{k-1} + (1 - q_k)^{-1}\left(\alpha_{k-1}^{-1} - 1\right)(\alpha_k - q_k)(x_{k-1} - x_{k-2}).
\end{aligned}
$$

■

3.1 Preparations for the convergence rate proof

The following lemma summarize important results that give a swift exposition for the proofs show up at the end for the convergence rate.

{lemma:cnvg-prep-part1}

Lemma 3.5 (convergence preparations part I) *Let $(F, f, g, \mathcal{E}, \mu, L)$ satisfies Assumption 3.1. Suppose that*

(i) *The iterates $(y_k, v_k, x_k)_{k\geq 0}$ satisfies Definition 2.1 where T_B defined using $F = f + g$.*
(ii) *The sequences $(\alpha_k)_{k\geq 0}, (q_k)_{k\geq 0}$ satisfy Definition 2.3.*
(iii) *We choose the parameters q_k has $q_k = \mu/B_k$, with $B_k > \mu$, for all $k \geq 0$.*

Then, for all $\bar{x} \in \mathbb{R}^n$, $k \geq 0$:

$$
\begin{aligned}
&\frac{B_k\alpha_k^2}{2}\|\bar{x} - v_k\|^2 - \frac{B_k\alpha_k(\alpha_k - q_k)}{2}\|\bar{x} - v_{k-1}\|^2 \\
&= \frac{\alpha_k\mu}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{q_k\mu}{2}\|y_k - v_{k-1}\| + \frac{q_k}{2\mu}\|\mathcal{G}_{B_k}(y_k)\|^2 \\
&\quad - \langle q_k(y_k - v_{k-1}) + \alpha_k(v_{k-1} - \bar{x}), \mathcal{G}_{B_k}(y_k)\rangle + \alpha_k\mu\,\langle v_{k-1} - \bar{x}, y_k - v_{k-1}\rangle\,.
\end{aligned}
$$

Proof. Consider any $\bar{x} \in \mathbb{R}^n$.

$$
\begin{aligned}
&\frac{B_k\alpha_k^2}{2}\|\bar{x} - v_k\|^2 \\
&\underset{(1)}{=} \frac{B_k\alpha_k^2}{2}\left\|\bar{x} - \left(v_{k-1} + \alpha_k^{-1}q_k(y_k - v_{k-1}) - \alpha_k^{-1}B_k^{-1}\mathcal{G}_{B_k}(y_k)\right)\right\|^2 \\
&= \frac{B_k\alpha_k^2}{2}\left\|(\bar{x} - v_{k-1}) - \alpha_k^{-1}\left(q_k(y_k - v_{k-1}) - B_k^{-1}\mathcal{G}_{B_k}(y_k)\right)\right\|^2 \\
&= \frac{B_k\alpha_k^2}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{B_k}{2}\left\|q_k(y_k - v_{k-1}) - B_k^{-1}\mathcal{G}_{B_k}(y_k)\right\|^2 \\
&\quad - \alpha_k B_k\left\langle \bar{x} - v_{k-1}, q_k(y_k - v_{k-1}) - B_k^{-1}\mathcal{G}_{B_k}(y_k)\right\rangle \\
&= \frac{B_k\alpha_k^2}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{B_k q_k^2}{2}\|y_k - v_{k-1}\| + \frac{1}{2B_k}\|\mathcal{G}_{B_k}(y_k)\|^2 - q_k\langle y_k - v_{k-1}, \mathcal{G}_{B_k}(y_k)\rangle \\
&\quad + B_k\alpha_k\left\langle v_{k-1} - \bar{x}, q_k(y_k - v_{k-1}) - B_k^{-1}\mathcal{G}_{B_k}(y_k)\right\rangle \\
&= \frac{B_k\alpha_k^2}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{B_k q_k^2}{2}\|y_k - v_{k-1}\| + \frac{1}{2B_k}\|\mathcal{G}_{B_k}(y_k)\|^2 - q_k\langle y_k - v_{k-1}, \mathcal{G}_{B_k}(y_k)\rangle \\
&\quad + B_k\alpha_k\left\langle v_{k-1} - \bar{x}, q_k(y_k - v_{k-1})\right\rangle - \alpha_k\left\langle v_{k-1} - \bar{x}, \mathcal{G}_{B_k}(y_k)\right\rangle \\
&= \frac{\alpha_k^2 B_k}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{q_k^2 B_k}{2}\|y_k - v_{k-1}\| + \frac{1}{2B_k}\|\mathcal{G}_{B_k}(y_k)\|^2 \\
&\quad - \langle q_k(y_k - v_{k-1}) + \alpha_k(v_{k-1} - \bar{x}), \mathcal{G}_{B_k}(y_k)\rangle \\
&\quad + \alpha_k q_k B_k\langle v_{k-1} - \bar{x}, y_k - v_{k-1}\rangle .
\end{aligned}
$$

At (1) we used Lemma 3.3. Subtracting $\frac{B_k\alpha_k(\alpha_k - q_k)}{2}\|\bar{x} - v_{k-1}\|^2$ from both sides, the coefficient of $\|\bar{x} - v_{k-1}\|^2$ comes out to be:

$$
\frac{\alpha_k^2 B_k}{2} - \frac{B_k\alpha_k(\alpha_k - q_k)}{2} = \frac{B_k}{2}(\alpha_k^2 - \alpha_k(\alpha_k - q_k)) \underset{(1)}{=} \frac{B_k\alpha_k q_k}{2}.
$$

At (1), we used the relation $(1-\alpha_k)\rho_{k-1}\alpha_{k-1}^2 = \alpha_k(\alpha_k - q_k)$ as in Definition 2.3, for all $k \geq 1$. Therefore, we have the equality:

$$
\begin{aligned}
&\frac{B_k\alpha_k^2}{2}\|\bar{x} - v_k\|^2 - \frac{B_k\alpha_k(\alpha_k - q_k)}{2}\|\bar{x} - v_{k-1}\|^2 \\
&= \frac{\alpha_k q_k B_k}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{q_k^2 B_k}{2}\|y_k - v_{k-1}\| + \frac{1}{2B_k}\|\mathcal{G}_{B_k}(y_k)\|^2 \\
&\quad - \langle q_k(y_k - v_{k-1}) + \alpha_k(v_{k-1} - \bar{x}), \mathcal{G}_{B_k}(y_k)\rangle \\
&\quad + \alpha_k q_k B_k\langle v_{k-1} - \bar{x}, y_k - v_{k-1}\rangle . \\
&\underset{(1)}{=} \frac{\alpha_k\mu}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{q_k\mu}{2}\|y_k - v_{k-1}\| + \frac{q_k}{2\mu}\|\mathcal{G}_{B_k}(y_k)\|^2 \\
&\quad - \langle q_k(y_k - v_{k-1}) + \alpha_k(v_{k-1} - \bar{x}), \mathcal{G}_{B_k}(y_k)\rangle + \alpha_k\mu\langle v_{k-1} - \bar{x}, y_k - v_{k-1}\rangle .
\end{aligned}
$$

At (1), we used the relation that $B_k = \mu/q_k$, for all $k \geq 0$. ■

{lemma:cnvg-prep-part2}

Lemma 3.6 (convergence preparations part II) *The iterates $(y_k, x_k, v_k)_{k\geq 0}$ satisfies Definition 2.1 then, for all $k \geq 0, \bar{x} \in \mathbb{R}^n$ the following identities:*

$$\alpha_k(v_{k-1} - \bar{x}) + q_k(y_k - v_{k-1}) + x_{k-1} - y_k = \alpha_k(x_{k-1} - \bar{x}).$$

Proof. We first establish two intermediate results. From Definition 2.1, it has for all $k \geq 0$:

$$\begin{aligned} y_k &= \left(\frac{\alpha_k - q_k}{1 - q_k}\right) v_{k-1} + \left(\frac{1 - \alpha_k}{1 - q_k}\right) x_{k-1} \\ &= \left(1 - \frac{1 - \alpha_k}{1 - q_k}\right) v_{k-1} + \left(\frac{1 - \alpha_k}{1 - q_k}\right) x_{k-1} \\ \iff y_k - v_{k-1} &= \left(\frac{1 - \alpha_k}{1 - q_k}\right)(x_{k-1} - v_{k-1}). \end{aligned}$$

Similarly:

$$\begin{aligned} y_k &= \left(\frac{\alpha_k - q_k}{1 - q_k}\right) v_{k-1} + \left(\frac{1 - \alpha_k}{1 - q_k}\right) x_{k-1} \\ &= \left(\frac{\alpha_k - q_k}{1 - q_k}\right) v_{k-1} + \left(1 - \frac{\alpha_k - q_k}{1 - q_k}\right) x_{k-1} \\ \iff y_k - x_{k-1} &= \left(\frac{\alpha_k - q_k}{1 - q_k}\right)(v_{k-1} - x_{k-1}). \end{aligned}$$

Now, we use the above two results and, it derives

$$\begin{aligned} &\alpha_k(v_{k-1} - \bar{x}) + q_k(y_k - v_{k-1}) + x_{k-1} - y_k \\ &= \alpha_k(v_{k-1} - \bar{x}) + q_k\left(\frac{1 - \alpha_k}{1 - q_k}\right)(x_{k-1} - v_{k-1}) - \left(\frac{\alpha_k - q_k}{1 - q_k}\right)(v_{k-1} - x_{k-1}). \\ &= \alpha_k(v_{k-1} - \bar{x}) + (1 - q_k)^{-1}(q_k - q_k\alpha_k + (\alpha_k - q_k))(x_{k-1} - v_{k-1}) \\ &= \alpha_k(v_{k-1} - \bar{x}) + \alpha_k(x_{k-1} - v_{k-1}) \\ &= \alpha_k(x_{k-1} - \bar{x}). \end{aligned}$$

■

{lemma:cnvg-prep-part3}

Lemma 3.7 (convergence preparations part III)
Suppose that all the following are satisfied

(i) $(F, f, g, \mathcal{E}, \mu, L)$ *satisfies Assumption 3.1.*
(ii) *The sequences $(\alpha_k)_{k\geq 0}, (q_k)_{k\geq 0}$ satisfies Definition 2.3.*

(iii) *We choose* $(q_k)_{k\geq 0}$ *is given by* $q_k = \mu/B_k$ *for all* $k \geq 0$.
(iv) *The sequence* $(y_k, v_k, x_k)_{k\geq 0}$ *satisfies Definition 2.1.*

Then, $\forall k \geq 1$*, there exists* $\bar{x}_k \in \mathbb{R}^n$*, such that:*

$$
\begin{aligned}
&F(x_k) - F(\bar{x}_k) + \frac{B_k\alpha_k^2}{2}\|\bar{x} - v_k\|^2 \\
&\leq (1-\alpha_k)(F(x_{k-1}) - F(\bar{x}_k)) + \frac{B_k\alpha_k(\alpha_k - q_k)}{2}\|\bar{x}_k - v_{k-1}\|^2 \\
&\quad - (1-\alpha_k)\left(\mathcal{E}(x_{k-1}, y_k, \mu, B_k) + \frac{\mu}{2}\|x_{k-1} - y_k\|^2\right).
\end{aligned}
$$

Proof. Recall Definition 2.4, consider:

$$
\begin{aligned}
&\mathcal{E}(x_{k-1}, y_k, \mu, B_k) \\
&= F(x_{k-1}) - F(T_{B_k}(y_k)) - \langle B_k(y_k - T_{B_k}(y_k)), x_{k-1} - y_k\rangle - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 - \frac{B_k}{2}\|y_k - x_k\|^2 \\
&= F(x_{k-1}) - F(x_k) - \langle \mathcal{G}_{B_k}(y_k), x_{k-1} - y_k\rangle - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 - \frac{q_k}{2\mu}\|\mathcal{G}_{B_k}(y_k)\|^2.
\end{aligned}
$$

On the second equality above, we used $B_k(y_k - x_k) = \mathcal{G}_{B_k}(y_k)$, and $B_k = \mu/q_k$. For all $k \geq 0$, we define Ξ_k and simplify using the above result:

$$
\begin{aligned}
\Xi_k &:= \mathcal{E}(x_{k-1}, y_k, \mu, B_k) + F(x_k) - F(\bar{x}_k) - (1-\alpha_k)(F(x_{k-1}) - F(\bar{x}_k)) \\
&= \mathcal{E}(x_{k-1}, y_k, \mu, B_k) + F(x_k) - \alpha_k F(\bar{x}_k) - (1-\alpha_k)F(x_{k-1}) \\
&= \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \langle \mathcal{G}_{B_k}(y_k), x_{k-1} - y_k\rangle - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 - \frac{q_k}{2\mu}\|\mathcal{G}_{B_k}(y_k)\|^2.
\end{aligned}
$$

Now consider the new term Ξ'_k which we defined and simplify below:

$$
\begin{aligned}
\Xi'_k &:= \Xi_k + \frac{B_k\alpha_k^2}{2}\|\bar{x}_k - v_k\|^2 - \frac{B_k\alpha_k(\alpha_k - q_k)}{2}\|\bar{x}_k - v_{k-1}\|^2 \\
&\underset{(1)}{=} \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \langle \mathcal{G}_{B_k}(y_k), x_{k-1} - y_k\rangle - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 - \frac{q_k}{2\mu}\|\mathcal{G}_{B_k}(y_k)\|^2 \\
&\quad + \frac{\alpha_k\mu}{2}\|\bar{x}_k - v_{k-1}\|^2 + \frac{q_k\mu}{2}\|y_k - v_{k-1}\|^2 + \frac{q_k}{2\mu}\|\mathcal{G}_{B_k}(y_k)\|^2 \\
&\quad - \langle q_k(y_k - v_{k-1}) + \alpha_k(v_{k-1} - \bar{x}_k), \mathcal{G}_{B_k}(y_k)\rangle + \alpha_k\mu\langle v_{k-1} - \bar{x}_k, y_k - v_{k-1}\rangle \\
&= \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \langle \mathcal{G}_{B_k}(y_k), x_{k-1} - y_k\rangle - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 \\
&\quad + \frac{\alpha_k\mu}{2}\|\bar{x}_k - v_{k-1}\|^2 + \frac{q_k\mu}{2}\|y_k - v_{k-1}\|^2 \\
&\quad - \langle q_k(y_k - v_{k-1}) + \alpha_k(v_{k-1} - \bar{x}_k), \mathcal{G}_{B_k}(y_k)\rangle + \alpha_k\mu\langle v_{k-1} - \bar{x}_k, y_k - v_{k-1}\rangle \\
&= \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 + \frac{\alpha_k\mu}{2}\|\bar{x}_k - v_{k-1}\|^2 + \frac{q_k\mu}{2}\|y_k - v_{k-1}\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \langle x_{k-1} - y_k + q_k(y_k - v_{k-1}) + \alpha_k(v_{k-1} - \bar{x}_k), \mathcal{G}_{B_k}(y_k)\rangle \\
&\quad + \alpha_k\mu \left\langle v_{k-1} - \bar{x}_k, q_k(y_k - v_{k-1})\right\rangle \\
&\underset{(2)}{=} \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 \\
&\quad + \frac{\alpha_k\mu}{2}\|\bar{x}_k - v_{k-1}\|^2 + \frac{q_k\mu}{2}\|y_k - v_{k-1}\|^2 \\
&\quad - \alpha_k\langle x_{k-1} - \bar{x}_k, \mathcal{G}_{B_k}(y_k)\rangle + \alpha_k\mu \left\langle v_{k-1} - \bar{x}_k, q_k(y_k - v_{k-1})\right\rangle \\
&= \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 \\
&\quad + \frac{\alpha_k\mu}{2}\|\bar{x}_k - v_{k-1}\|^2 + \frac{\alpha_k\mu}{2}\|y_k - v_{k-1}\|^2 + \alpha_k\mu \left\langle v_{k-1} - \bar{x}_k, q_k(y_k - v_{k-1})\right\rangle \\
&\quad - \alpha_k\langle x_{k-1} - \bar{x}_k, \mathcal{G}_{B_k}(y_k)\rangle + \frac{q_k\mu - \mu\alpha_k}{2}\|y_k - v_{k-1}\|^2 \\
&= \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 \\
&\quad + \frac{\alpha_k\mu}{2}\|y_k - \bar{x}_k\|^2 - \alpha_k\langle x_{k-1} - \bar{x}_k, \mathcal{G}_{B_k}(y_k)\rangle + \frac{q_k\mu - \mu\alpha_k}{2}\|y_k - v_{k-1}\|^2 \\
&\underset{(3)}{\le} \alpha_k F(x_{k-1}) - \alpha_k F(\bar{x}_k) - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 \\
&\quad + \frac{\alpha_k\mu}{2}\|y_k - \bar{x}_k\|^2 - \alpha_k\langle x_{k-1} - \bar{x}_k, \mathcal{G}_{B_k}(y_k)\rangle. \\
&= \alpha_k\left(F(x_{k-1}) - F(\bar{x}_k) - \langle x_{k-1} - \bar{x}_k, \mathcal{G}_{B_k}(y_k)\rangle + \frac{\mu}{2}\|y_k - \bar{x}_k\|^2\right) - \frac{\mu}{2}\|x_{k-1} - y_k\|^2.
\end{aligned}
$$

At (1), we used Lemma 3.5, and substituted Ξ_k. At (2), we used Lemma 3.6 to simplify the inner product. At (3), we used the $\alpha_k > q_k$ as in Definition 2.3, hence it makes the coefficient $q_k\mu - \mu\alpha_k \le 0$, which gives us the inequality. Now, subtracting $\mathcal{E}(x_{k-1}, y_k, \mu, B_k)$ from both sides of the inequality will yield:

$$
\begin{aligned}
&\Xi_k' - \mathcal{E}(x_{k-1}, y_k, \mu, B_k) \\
&\le \alpha_k\left(F(x_{k-1}) - F(\bar{x}_k) - \langle x_{k-1} - \bar{x}_k, \mathcal{G}_{B_k}(y_k)\rangle + \frac{\mu}{2}\|y_k - \bar{x}_k\|^2 - \mathcal{E}(x_{k-1}, y_k, \mu, B_k)\right) \\
&\quad - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 - (1-\alpha_k)\mathcal{E}(x_{k-1}, y_k, \mu, B_k) \\
&\underset{(1)}{=} \alpha_k\left(F(x_k) - F(\bar{x}_k) - \langle \mathcal{G}_{B_k}(y_k), x_{k-1} - \bar{x}_k - (x_{k-1} - y_k)\rangle\right) \\
&\quad + \alpha_k\left(\frac{\mu}{2}\|y_k - \bar{x}_k\|^2 + \frac{\mu}{2}\|x_{k-1} - y_k\|^2 + \frac{B_k}{2}\|y_k - x_k\|^2\right) \\
&\quad - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 - (1-\alpha_k)\mathcal{E}(x_{k-1}, y_k, \mu, B_k) \\
&= \alpha_k\left(F(x_k) - F(\bar{x}_k) - \langle \mathcal{G}_{B_k}(y_k), y_k - \bar{x}_k\rangle + \frac{\mu}{2}\|y_k - \bar{x}_k\|^2 + \frac{B_k}{2}\|y_k - x_k\|^2 + \frac{\mu}{2}\|x_{k-1} - y_k\|^2\right) \\
&\quad - \frac{\mu}{2}\|x_{k-1} - y_k\|^2 - (1-\alpha_k)\mathcal{E}(x_{k-1}, y_k, \mu, B_k) \\
&= \alpha_k\left(-\mathcal{E}(\bar{x}_k, y_k, \mu, B_k) + \frac{\mu}{2}\|x_{k-1} - y_k\|^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\mu}{2}\|x_{k-1}-y_k\|^2-(1-\alpha_k)\mathcal{E}(x_{k-1},y_k,\mu,B_k)\\
&\underset{(2)}{\le}\frac{\alpha_k\mu}{2}\|x_{k-1}-y_k\|^2-\frac{\mu}{2}\|x_{k-1}-y_k\|^2-(1-\alpha_k)\mathcal{E}(x_{k-1},y_k,\mu,B_k)\\
&=-(1-\alpha_k)\left(\mathcal{E}(x_{k-1},y_k,\mu,B_k)+\frac{\mu}{2}\|x_{k-1}-y_k\|^2\right).
\end{aligned}
$$

At (1), we substituted $\mathcal{E}(x_{k-1},y_k,\mu,B_k)$. At (2), we used the $\mathcal{E}(\bar{x}_k,y_k,\mu,B_k)\le 0$ by choosing $\bar{x}_k=\bar{y}$ in Assumption 3.1(ii) to make the inequality. Now, recall the definitions of Ξ_k,Ξ'_k which means we proved the following:

$$
\begin{aligned}
&\Xi'_k-\mathcal{E}(x_{k-1},y_k,\mu,B_k)\\
&=\Xi_k+\frac{B_k\alpha_k^2}{2}\|\bar{x}_k-v_k\|^2-\frac{B_k\alpha_k(\alpha_k-q_k)}{2}\|\bar{x}_k-v_{k-1}\|^2-\mathcal{E}(x_{k-1},y_k,\mu,B_k)\\
&=\mathcal{E}(x_{k-1},y_k,\mu,B_k)+F(x_k)-F(\bar{x}_k)-(1-\alpha_k)(F(x_{k-1})-F(\bar{x}_k))\\
&\quad+\frac{B_k\alpha_k^2}{2}\|\bar{x}_k-v_k\|^2-\frac{B_k\alpha_k(\alpha_k-q_k)}{2}\|\bar{x}_k-v_{k-1}\|^2-\mathcal{E}(x_{k-1},y_k,\mu,B_k)\\
&=F(x_k)-F(\bar{x}_k)-(1-\alpha_k)(F(x_{k-1})-F(\bar{x}_k))\\
&\quad+\frac{B_k\alpha_k^2}{2}\|\bar{x}_k-v_k\|^2-\frac{B_k\alpha_k(\alpha_k-q_k)}{2}\|\bar{x}_k-v_{k-1}\|^2\\
&\le-(1-\alpha_k)\left(\mathcal{E}(x_{k-1},y_k,\mu,B_k)+\frac{\mu}{2}\|x_{k-1}-y_k\|^2\right).
\end{aligned}
$$

■

3.2 Proving the convergence rate

To finally find the convergence rate, we will strengthen Assumption 3.1. Our convergence rate is expressed for sequences $(\alpha_k)_{k\ge 0},(q_k)_{k\ ge0}$ if they satisfy Definition 2.3. This means that any sequence with $\alpha_k\in(q_k,1)$ would work.

The following assumptions describe the behaviors of an algorithm satisfying Definition 2.1, its parameters, and the properties of the objective function.

{ass:lin-cnvg} **Assumption 3.8 (Assumptions for linear convergence rate)**
Let $(F,f,g,\mathcal{E},\mu,L)$ satisfies Assumption 3.1. In addition, we strengthened the prior assumptions:

(i) Define the set of minimizers $X^+=\operatorname*{argmin}_{z\in\mathbb{R}^n}\{f(z)+g(z)\}$, it has $X^+\neq\emptyset$.
(ii) Strenghening Assumption 3.1(ii), assume $\exists\mu>0$ such that $\forall y\in\mathbb{R}^n$, it has $\mathcal{E}\left(\Pi_{X^+}y,y,\mu,B\right)\ge 0$.
(iii) For all $y,z\in\mathbb{R}^n$, it satisfies that $\mathcal{E}(z,y,\mu,B)+\mu/2\|z-y\|^2\ge 0$.

Now, suppose an algorithm which optimizes a $F = f + g$ satisfying all previous assumptions, and it generates iterates $(y_k, x_k, v_k)_{k\ge 0}$ that satisfies Definition 2.4. In addition, we assume that the parameter sequences $(\alpha_k)_{k\ge0}, (q_k)_{k\ge0}, (B_k)_{k\ge0}$ satisfy the following:

(iv) The sequences α_k, q_k are given by Definition 2.3.
(v) The sequence $q_k = \mu/B_k$, with $B_k > \mu$.
(vi) For all $k \ge 0$, $\Pi_{X^+} y_k$ is unique.

{ass:forever-cnvx-cnvg} **Assumption 3.9 (forever sublinear convergence rate in the convex case)**
Let $(F, f, g, \mathcal{E}, \mu, L)$ satisfies Assumption 3.1. In addition, we strengthened the prior assumptions:

(i) Define the set of minimizers $X^+ = \operatorname*{argmin}_{z\in\mathbb{R}^n}\{f(z) + g(z)\}$, it has $X^+ \neq \emptyset$.
(ii) The function f is convex. Hence assumption Assumtion 3.1 (ii) satisfied for all $\bar y$, $\mu = 0$.

Now, suppose an algorithm which optimizes a $F = f + g$ satisfying all previous assumptions, and it generates iterates $(y_k, x_k, v_k)_{k\ge 0}$ that satisfies Definition 2.4. In addition, we assume that the parameter sequences $(\alpha_k)_{k\ge0}, (q_k)_{k\ge0}, (B_k)_{k\ge0}$ satisfy the following:

(i) The sequence $B_k \ge B_{k-1}$, it's a non-decreasing sequence.
(ii) $\alpha_k \ge \alpha_{k-1}$, it's a non-increasing sequence.

Assumption 3.10 (optimal convergence rate in the convex case)

3.3 convergence results based on these assumptions

{thm:cnvg-generic-seq} **Theorem 3.11 (convergence with generic sequence)**
Let $(F, f, g, \mathcal{E}, \mu, L)$, and sequences $(\alpha_k)_{k\ge0}, (q_k)_{k\ge0}, (B_k)_{k\ge0}$ satisfy Assumption 3.8. Denote

$$\beta_k = \prod_{i=1}^{k} \max\left(1 - \alpha_i, \frac{B_i\alpha_i(\alpha_i - q_i)}{\alpha_{i-1}^2 B_{i-1}}\right).$$

Then, there exists a unique $\bar x \in \mathbb{R}^n$ such that for all $k \ge 1$, $\bar x = \Pi_{X^+} y_k$, and it satisfies

$$F(x_k) - F(\bar x) + \frac{B_k}{\alpha_k^2}\|\bar x - v_k\|^2 \le \beta_k\left(F(x_0) - F(\bar x) + \frac{\alpha_0 B_0}{2}\|\bar x - v_0\|^2\right).$$

If in addition, we assume $x_{-1} = v_{-1}, \alpha_0 = 1$, then the above inequality simplifies:

$$F(x_k) - F(\bar x) + \frac{B_k}{\alpha_k^2}\|\bar x - v_k\|^2 \le \frac{\beta_k B_0}{2}\|\bar x - x_{-1}\|^2.$$

Proof. Set $\bar{x}_k = \bar{x}$ in Lemma 3.7 then for all $k \geq 1$ it has

$$
\begin{aligned}
&F(x_k) - F(\bar{x}) + \frac{B_k\alpha_k^2}{2}\|\bar{x} - v_k\|^2 \\
&\leq (1-\alpha_k)\left(F(x_{k-1}) - F(\bar{x})\right) + \frac{(1-\alpha_k)B_k\alpha_{k-1}^2\rho_{k-1}}{2}\|\bar{x} - v_{k-1}\|^2 \\
&\underset{(1)}{=} (1-\alpha_k)\left(F(x_{k-1}) - F(\bar{x})\right) + \frac{\alpha_k(\alpha_k - q_k)B_k\alpha_{k-1}^2}{\alpha_{k-1}^2}\|\bar{x} - v_{k-1}\|^2 \\
&= \max\left(1-\alpha_k, \frac{B_i\alpha_k(\alpha_k - q_k)}{\alpha_{k-1}^2 B_{k-1}}\right)\left(F(x_{k-1}) - F(\bar{x}) + \frac{B_{k-1}\alpha_{k-1}^2}{2}\|\bar{x} - v_{k-1}\|^2\right).
\end{aligned}
$$

From the above, a recurrence relation is formed for $k \geq 1$, unrolling the recurrence relation it has

$$
\begin{aligned}
&F(x_k) - F(\bar{x}) + \frac{B_k\alpha_k^2}{2}\|\bar{x} - v_k\|^2 \\
&\leq \prod_{i=1}^{k}\max\left(1-\alpha_i, \frac{B_i\alpha_i(\alpha_i - q_i)}{\alpha_{i-1}^2 B_{i-1}}\right)\left(F(x_0) - F(\bar{x}) + \frac{B_0\alpha_0^2}{2}\|\bar{x} - v_0\|^2\right).
\end{aligned}
$$

When $x_{-1} = v_{-1}$, from Definition 2.1, when $k = 0$ it has $y_0 = v_{-1} = x_{-1}$, so $x_0 = T_{B_0}(y_0)$. Because $\alpha_0 = 1$, it also has $v_0 = x_0$. Choose $z = \bar{x}, y = x_{-1}$ we have from Assumption 3.1 (iii) that

$$
\begin{aligned}
0 &\leq \mathcal{E}(\bar{x}, x_{-1}, \mu, B_0) - \frac{\mu}{2}\|\bar{x} - x_{-1}\|^2 \\
&= F(\bar{x}) - F(x_{-1}) - B_0\left\langle x_{-1} - T_{B_0}(x_{-1}), \bar{x} - x_{-1}\right\rangle - \frac{B_0}{2}\left\|x_{-1} - T_{B_0}x_{-1}\right\|^2 \\
&= F(\bar{x}) - F(x_{-1}) - \frac{B_0}{2}\|\bar{x} - T_{B_0}x_{-1}\|^2 + \frac{B_0}{2}\|\bar{x} - x_{-1}\|^2. \\
&= F(\bar{x}) - F(x_{-1}) - \frac{B_0}{2}\|\bar{x} - v_0\|^2 + \frac{B_0}{2}\|\bar{x} - x_{-1}\|^2.
\end{aligned}
$$

Substitute the above into the RHS of the inequality of previous results to complete the proof. ■

SOMETHING HAS LIMITATIONS, LET US EXPLAIN. The parameter $q_k = \mu/B_k$ cannot be eliminated from Definition 2.1. If one were to implement any algorithm without the exact knowledge of μ, then the convergence results we just derived are not applicable. Implementations of the algorithm, and the convergence results requires the presence of both parameters q_k, α_k. Doomed. Because in the applications of our interests, the exact knowledge of μ is impossible to know.

The theorem below will definitively close the case to show that there exists choices for the sequence such that a linear convergence exists, meaning that β_k will decrease at a linear rate.

{lemma:beta-seq} **Lemma 3.12 (beta sequence linear bound)**
Let sequences $(\alpha_k)_{k\ge 0}, (q_k)_{k\ge 0}, (B_k)_{k\ ge0}$ *satisfy Assumption 3.8. Define*

$$\beta_k := \prod_{i=1}^{k} \max\left(1-\alpha_i, \frac{B_i\alpha_i(\alpha_i - q_i)}{\alpha_{i-1}^2 B_{i-1}}\right).$$

Assume in addition $\forall k \ge 0$, $\exists B : B_k = B_{k-1} = B$ *hence there exists* q *such that* $(\forall k \ge 0)\ q_k = q$. *Then, linear convergence of* β_k *is possible under the following scenarios:*

(i) *If for all* $k \ge 1$, $\sqrt{q} \le \alpha_k \le \alpha_{k-1}$ *so it's non-increasing, then* $\beta_k = \Pi_{i=1}^k(1 - q/\alpha_{i-1})$. *Since* $\alpha_k \in (0,1)$ *and it's monotone, it has upper bound* $\beta_k \le (1-q)^k$.

(ii) *If for all* $k \ge 1$, $\alpha_k = \alpha_{k-1}$, *so there exists* $\alpha = \alpha_k$, *making the sequence a constant, then* $\beta_k = \max(1 - q/\alpha, 1-\alpha)^k$. *And, it's lowest when* $\alpha_k = \sqrt{q}$.

By the assumption that $q \in (0,1)$ *hence we had a linear convergence rate in each of the above scenarios.*

Proof. Since B_k is a constant, and by Definition 2.1 it has $B_k = \frac{\mu}{q_k}$ it would make q_k to be a constant for all $k \ge 0$, which we denote $q := q_k$.

Proof of (i). Using mononicity of sequence α_k, it has $\alpha_i/\alpha_{i-1} \le 1$ hence for all $i \ge 1$ it has

$$\begin{aligned}
&\max\left(1-\alpha_i, \frac{\alpha_i(\alpha_i - q)}{\alpha_{i-1}^2}\right) \\
&\le \max\left(1-\alpha_i, \frac{\alpha_i - q}{\alpha_{i-1}}\right) \\
&\le \max\left(1-\alpha_i, 1 - \frac{q}{\alpha_{i-1}}\right)
\end{aligned}$$

Then, observe that when $\alpha_i \ge \sqrt{q}$ it has:

$$\begin{aligned}
1 - \frac{q}{\alpha_{i-1}} &> 1 - \frac{q}{\sqrt{q}} \\
&= 1 - \sqrt{q} \ge 1 - \alpha_k.
\end{aligned}$$

Hence

$$\max\left(1-\alpha_i, \frac{\alpha_i(\alpha_i - q)}{\alpha_{i-1}^2}\right) \le \left(1 - \frac{q}{\alpha_{k-1}}\right).$$

The result now follows by the definition of β_k.

Proof of (ii). The proof is obvious by setting both q_i, α_i to be a constant. ■

4 Examples for our assumptions

We will check Assumption 3.8 and, propose examples for it.

4.1 quasi strongly convex function

In Necoara et al.'s setting, they have the following assumptions for their objective function.

{ass:necoara} **Assumption 4.1 (The settings for Necoara's et al)**
The assumption is about (f, X, X^+, f^+) where

(i) $X \subseteq \mathbb{R}^n$ is a closed convex set.
(ii) $f : X \to \mathbb{R}$ has L Lipschitz continuous gradient, and it's convex.
(iii) $X^+ = \operatorname*{argmin}_{x \in X} f(x) \neq \emptyset$, and we denote f^+ to be the minimum value.

The following definition on Quasi Strongly Convex function (Q-SCNVX) is taken from [2, Definition 1].

{def:Q-SCNVX}

Definition 4.2 (Q-SCNVX) *Let (f, X, X^+, f^+) be given by Assumption 4.1. We define f to be Quasi Strong Convex on X if there exists $\kappa_f > 0$ such that $\forall x \in \mathbb{R}^n$, with $\bar{x} = \Pi_{X^+} x$ it has:*

$$f^+ - f(x) - \langle \nabla f(x), \bar{x} - x \rangle - \frac{\kappa_f}{2} \|\bar{x} - x\|^2 \geq 0.$$

{prop:qscnvx-ass-ok}

Proposition 4.3 (Q-SCNVX is an example of our assumptions) *Let (f, X, X^+, f^+) satisfies Assumption 4.1. Let $\mathcal{E}$ be given by Definition 2.4 with $g = \delta_X$. Then for all $x \in \mathbb{R}^n$, let $\bar{x} = \Pi_{X^+} x, x^+ = T_B(x)$, there exists $B \geq 0$, such that*

$$0 \leq \mathcal{E}(\bar{x}, x, B, \kappa_f) = f(\bar{x}) - f(x^+) - \frac{B}{2} \|x - x^+\|^2 - B \langle \bar{x} - x, x - x^+ \rangle - \frac{\kappa_f}{2} \|\bar{x} - x\|^2.$$

Therefore, it satisfies Assumption 3.1.

Proof. Let $h = z \mapsto \delta_X(z) + \langle \nabla f(x), z \rangle + \frac{B}{2} \|z - x\|^2$. So it has $T_B(x) = \operatorname*{argmin}_{z \in \mathbb{R}^n} h(z)$. Furthermore, $h(z)$ is a B strongly convex function by convexity of X and, the fact that other parts are just the sum of a linear and quadratic function. Denote $x^+ = T_B(x)$, since x^+ is a minimizer therefore it has $(\forall x \in X)\ h(z) - h(x^+) \geq \frac{B}{2} \|z - x^+\|^2$. That was the quadratic growth condition of h. Next, denote $\bar{x} = \Pi_{X^+} x$, let $z = \bar{x}$ then the condition becomes:

$$
\begin{aligned}
&\frac{B}{2}\|\bar{x}-x^+\|^2 \\
&\le \langle \nabla f(x), \bar{x}\rangle + \frac{B}{2}\|\bar{x}-x\|^2 - \langle \nabla f(x), x^+\rangle + \frac{B}{2}\|x^+ - x\|^2 \\
&= \langle \nabla f(x), \bar{x}-x\rangle - \langle \nabla f(x), x^+ - x\rangle + \frac{B}{2}\|\bar{x}-x\|^2 - \frac{B}{2}\|x^+ - x\|^2 \\
&= -D_f(\bar{x}, x) - f(x) + f(\bar{x}) - f(x^+) + f(x) + D_f(x^+, x) + \frac{B}{2}\|\bar{x}-x\|^2 - \frac{B}{2}\|x^+ - x\|^2 \\
&= -D_f(\bar{x}, x) + f(\bar{x}) - f(x^+) + \frac{B}{2}\|\bar{x}-x\|^2 + D_f(x^+, x) - \frac{B}{2}\|x^+ - x\|^2 \\
&\underset{(1)}{\le} -\frac{\kappa_f}{2}\|\bar{x}-x\|^2 + f(\bar{x}) - f(x^+) + \frac{B}{2}\|\bar{x}-x\|^2 + 0
\end{aligned}
$$

At (1), we used the fact that f is assumed satisfy 4.2 because $\bar{x} = \Pi_{X^+}x$, in addition the inequality $D_f(x^+, x) \le \frac{B}{2}\|x^+ - x\|^2$ is true because of Lipschitz gradient Assumption in 4.1 hence for all $B \ge L$, it's true for all x. Rearranging it has

$$
\begin{aligned}
0 &\le f(\bar{x}) - f(x^+) + \frac{B}{2}\|\bar{x}-x\|^2 - \frac{B}{2}\|\bar{x}-x^+\|^2 - \frac{\kappa_f}{2}\|\bar{x}-x\|^2 \\
&= f(\bar{x}) - f(x^+) - \frac{B}{2}\|x-x^+\|^2 - B\langle \bar{x}-x, x-x^+\rangle - \frac{\kappa_f}{2}\|\bar{x}-x\|^2 \\
&= \mathcal{E}(\bar{x}, x, B, \kappa_f).
\end{aligned}
$$

The above results showed that Assumption 3.1 (ii) has been satisfied, with $\bar{y} = \Pi_{X^+}y$. To see (iii) in Assumption 3.1, observe that Assumption 4.1 has f is convex. ■

4.2 A nonconvex example

4.3 Examples of Q-SCNVX

References

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer International Publishing, Cham, 2017.

[2] I. Necoara, Y. Nesterov, and F. Glineur, *Linear convergence of first order methods for non-strongly convex optimization*, Mathematical Programming, 175 (2019), pp. 69–107.