Linear Convergence of Accelerated Gradient without Restart requires some very strict conditions

Author 1 Name, Author 2 Name [*]

March 2, 2020

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

Abstract

This is still a note for a draft so no abstract [1]

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

1 Introduction

Necoara et al. [2] introduced the definition of quasi strongly convex function (Q-SCNVX), Quadratic Under approximations (QUA), Quadratic Gradient Growth (QGG), Proximal Error Bound (PEB) and, Quadratic Function Growth (QFG). These conditions are relaxation of strong convexity which enables linear convergence rate of first order method, including Nesterov's accelerated variants. In this file, we showed a new perspective of their works. Our goal is to relax their definitions and, to extend the linear convergence results, using completely new ideas and perspective.

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 $\mathbb{R}^n$.

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

The following definitions and assumptions are their.

{ass:necoara-linear} **Assumption 1.1 (Necoara's linear convergence assumptions)** The following assumptions are about (f, X, X^*, L_f).

(i) $f : \mathbb{R}^n \to \mathbb{R}$ is an L_f Lipschitz smooth function.
(ii) $X \subseteq \mathbb{R}^n$ is a closed convex non-empty set.
(iii) $X^* = \underset{x\in X}{\operatorname{argmin}}\, f(x) \neq \emptyset$.

Under this assumption, the following definitions are proposed.

Definition 1.2 (Necoara's weaker characterizations of strong convexity) *Suppose that (f, X, X^*, L_f) are given by Assumption 1.1. For all $x \in X$, denote $\bar{x} = \Pi_{X^*}x$. The following definitions are relaxations of strong convexity.*

(i) *f is Q-SCNVX if there exists $\kappa_f > 0$ such that $f(\bar{x}) - f(x) - \langle \nabla f(x), \bar{x} - x\rangle \geq \frac{\kappa_f}{2}\|x - \bar{x}\|^2$. Which we denote it by $f \in q\mathcal{S}(f, L_f, \kappa_f)$.*
(ii) *f is QUA if there exists $\kappa_f > 0$ such that $f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x}\rangle \geq \frac{\kappa_f}{2}\|x - \bar{x}\|^2$. We denote it by $f \in \mathcal{U}(f, L_f, \kappa_f)$.*
(iii) *f is QGG if there exists $\kappa_f > 0$ such that $\langle \nabla f(x) - \nabla f(\bar{x}), x - \bar{x}\rangle \geq \frac{\kappa}{2}\|x - \bar{x}\|^2$. We denote it by $f \in \mathcal{G}(f, L_f, \kappa_f)$.*
(iv) *f is PEB if there exists $\kappa_f > 0$ such that $\|x - L^{-1}\Pi_X(x - L^{-1}\nabla f(x))\| \geq \kappa_f\|x - \bar{x}\|$. We denote it by $f \in \mathcal{E}(f, L_f, \kappa_f)$.*
(v) *f is QFG if there exists $\kappa_f > 0$ such that $f(x) - f(\bar{x}) \geq \frac{\kappa_f}{2}\|x - \bar{x}\|^2$. We denote it by $f \in \mathcal{F}(f, L_f, \kappa_f)$.*

These definitions are the keys which Necoara's used to prove the linear convergence of projected gradient, and Nesterov's accelerated gradient method. In this paper, we will show a simple perspective that simplifies their arguments on linear convergence Accelerated Gradient method, without restart.

Unfortunately, and this is said at the start, so far the new perspective doesn't produce new results that are not in the literature.

2 Precursors materials for our proofs of linear convergence

For a function $f : \mathbb{R}^n \to \mathbb{R}$ we define its Bregman Divergence to be a mapping of $\mathbb{R}^n \times \operatorname{dom} \nabla f \mapsto \overline{\mathbb{R}}$ for all $x, y \in \mathbb{R}^n$:

$$D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y\rangle$$

Let $F = f + g$ where f is L Lipschitz smooth and g is convex. Let $B > 0$. We define the proximal gradient operator

$$T_B(x) = \operatorname*{argmin}_z \left\{ g(z) + \langle \nabla f(x), z - x\rangle + \frac{B}{2}\|z - x\|^2 \right\}.$$

Take note that it's also $T_B(x) = \operatorname{prox}_{B^{-1}g}(x - B^{-1}\nabla f(x))$, which is another equivalent representation.

{ass:obj-fxn} **Assumption 2.1 (objective function)**
The following assumption is about $(F, f, g, h, A, b, \mu, L)$. Let $m \in \mathbb{N}, n \in \mathbb{N}$ be arbitrary.

(i) $h : \mathbb{R}^m \mapsto \mathbb{R}$ is a L Lipschitz smooth and, $\mu \ge 0$ strongly convex function.
(ii) $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^n$ are a matrix and a vector.
(iii) $f = h(Ax - b)$.
(iv) $g = \delta_X$, so g is an indicator function of the set $X \subseteq \mathbb{R}^n$ which we assume that it's closed onvex and non-empty.

{lemma:aff-scnvx-smooth-lin-comp} **Lemma 2.2 (strongly and Lipscthiz smooth convex affine composite)**
Let $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^n$. Let $h : \mathbb{R}^m \mapsto \mathbb{R}$ be a L Lipschitz and $\mu \ge 0$ strongly convex function. Then it has for all $x, y \in \mathbb{R}^n$:

$$\frac{\sigma(A)^2\mu}{2}\|\Pi_{\ker A}(x - y)\|^2 \le D_f(x, y) \le \frac{L\|A\|^2}{2}\|x - y\|^2.$$

Proof. The lower bound is given by:

$$\begin{aligned}
&f(x) - f(y) - \langle \nabla f(y), x - y\rangle \\
&= h(Ax) - h(Ay) - \langle A^T\nabla h(y), x - y\rangle \\
&= h(Ax) - h(Ay) - \langle \nabla h(y), Ax - Ay\rangle \\
&\ge \frac{\mu}{2}\|Ax - Ay\|^2 \\
&\ge \frac{\mu}{2}\left(\|A\Pi_{\ker A}(x - y) + A(I - \Pi_{\ker A})(x - y)\|^2\right) \\
&\ge \frac{\mu\sigma(A)^2}{2}\|(I - \Pi_{\ker A})(x - y)\|^2.
\end{aligned}$$

The upper bound is direct from smoothness, which has:

$$\begin{aligned}
&f(x) - f(y) - \langle \nabla f(y), x - y\rangle \\
&= h(Ax) - h(Ay) - \langle \nabla h(y), Ax - Ay\rangle \\
&\le \frac{L}{2}\|Ax - Ay\|^2 \\
&\le \frac{L\|A\|^2}{2}\|x - y\|^2.
\end{aligned}$$

{thm:pg-ineq} ■

Theorem 2.3 (proximal gradient inequality) *Suppose that $(F, f, g, h, A, b, \mu, L)$ satisfies Assumption 2.1. Consider any $x, z \in \mathbb{R}^n$, then there exists a $B \ge 0$ such that for $x^+ = T_B(x)$, it has $D_f(x, x^+) \le \frac{B}{2}\|x - x^+\|^2$, and it satisfies:*

$$0 \le F(z) - F(x^+) + \frac{B - \mu\sigma(A)^2}{2}\|z - x\|^2 - \frac{B}{2}\|z - x^+\|^2 + \frac{\mu\sigma(A)^2}{2}\|\Pi_{\ker A}(z - x)\|^2.$$

Proof. Denote $\sigma = \sigma(A)$ for short, and $\Pi = \Pi_{\ker A}, \Pi_\perp = I - \Pi_{\ker A}$ for short. The function in the proximal gradient operator is L strongly convex, so it has quadratic growth over the minimizer x^+:

$$\begin{aligned}
&\frac{B}{2}\|z - x^+\|^2 \\
&\le \left(g(z) + \langle \nabla f(x), z - x\rangle + \frac{B}{2}\|z - x\|^2\right) - \left(g(x^+) + \langle \nabla f(x), x^+ - x\rangle + \frac{B}{2}\|x - x^+\|^2\right) \\
&= \left(F(z) - D_f(z, x) + \frac{B}{2}\|z - x\|^2\right) - \left(F(x^+) - D_f(x^+, x) + \frac{B}{2}\|x - x^+\|^2\right) \\
&\underset{(1)}{\le} \left(F(z) + \frac{B}{2}\|z - x\|^2 - \frac{\mu\sigma^2}{2}\|\Pi_\perp(z - x)\|^2\right) - \left(F(x^+) - D_f(x^+, x) + \frac{B}{2}\|x - x^+\|^2\right) \\
&\underset{(2)}{\le} F(z) + \frac{B}{2}\|z - x\|^2 - \frac{\mu\sigma^2}{2}\|\Pi_\perp(z - x)\|^2 - F(x^+).
\end{aligned}$$

At (1), we used the fact that $f = h(Ax + b)$, so it has $D_f(z, x) \ge \frac{\mu\sigma(A)^2}{2}\|\Pi_\perp(z - x)\|^2$. At (2), we used the fact that B makes $-D_f(x, x^+) + \frac{B}{2}\|x - x^+\|^2 \ge 0$. Continuing it has

$$\begin{aligned}
0 &\le F(z) + \frac{B}{2}\|z - x\|^2 - \frac{\mu\sigma^2}{2}\|\Pi_\perp(z - x)\|^2 - F(x^+) - \frac{B}{2}\|z - x^+\|^2 \\
&= F(z) + \frac{B}{2}\|z - x\|^2 - \frac{\mu\sigma^2}{2}\left(\|z - x\|^2 - \|\Pi(z - x)\|^2\right) - F(x^+) - \frac{B}{2}\|z - x^+\|^2 \\
&\le F(z) - F(x^+) + \frac{B - \mu\sigma^2}{2}\|z - x\|^2 - \frac{B}{2}\|z - x^+\|^2 + \frac{\mu\sigma^2}{2}\|\Pi(z - x)\|^2.
\end{aligned}$$

■

{thm:jen-ineq} **Theorem 2.4 (Jesen's inequality)** *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a L Lipschitz smooth and $\mu \geq 0$ strongly convex function. Let $A \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n$. Denote $\Pi_\perp = I - \Pi_{ker\, A}$ for short. Suppose that $f(x) = h(Ax + b)$, then it satisfies for all $x, y \in \mathbb{R}^n, \lambda \in [0, 1]$ the inequality:*

$$
\begin{aligned}
&f(\lambda x + (1-\lambda)y) \\
&\leq f(\lambda x + (1-\lambda)y) + \lambda f(x) - (1-\lambda)f(y) - \frac{\mu\sigma(A)^2\lambda(1-\lambda)}{2}\|\Pi_\perp(x-y)\|^2.
\end{aligned}
$$

Proof. To simplify notations, we denote $\sigma = \sigma(A)$ for short, and use $\Pi = \Pi_{\ker A}$, $\Pi_\perp = I - \Pi_{\ker A}$ for short. From Assumption 4 it has $f = h(Ax + b)$ therefore it has the following

$$
\begin{aligned}
0 &\leq D_f(x, y) - \frac{\sigma^2\mu}{2}\|\Pi_\perp(x-y)\|^2 \\
&= f(x) - f(y) - \langle \nabla f(y), x - y\rangle - \frac{\sigma^2\mu}{2}\|\Pi_\perp(x-y)\|^2 \\
&= f(x) - f(y) - \langle \nabla f(y), x - y\rangle - \frac{\sigma^2\mu}{2}\|\Pi_\perp x\|^2 - \frac{\sigma^2\mu}{2}\|\Pi_\perp y\|^2 + \sigma^2\mu\langle \Pi_\perp x, \Pi_\perp y\rangle \\
&= f(x) - \frac{\sigma^2\mu}{2}\|\Pi_\perp x\|^2 - \left(f(y) - \frac{\mu\sigma^2}{2}\|\Pi_\perp y\|^2\right) \\
&\quad - \sigma^2\mu\|\Pi_\perp y\|^2 - \langle \nabla f(y), x - y\rangle + \sigma^2\mu\langle \Pi_\perp x, \Pi_\perp y\rangle \\
&= f(x) - \frac{\sigma^2\mu}{2}\|\Pi_\perp x\|^2 - \left(f(y) - \frac{\mu\sigma^2}{2}\|\Pi_\perp y\|^2\right) - \langle \nabla f(y), x - y\rangle + \sigma^2\mu\langle \Pi_\perp(x-y), \Pi_\perp y\rangle \\
&= f(x) - \frac{\sigma^2\mu}{2}\|\Pi_\perp x\|^2 - \left(f(y) - \frac{\mu\sigma^2}{2}\|\Pi_\perp y\|^2\right) - \langle \nabla f(y), x - y\rangle + \sigma^2\mu\left\langle x - y, \Pi_\perp^\top \Pi_\perp y\right\rangle \\
&= f(x) - \frac{\sigma^2\mu}{2}\|\Pi_\perp x\|^2 - \left(f(y) - \frac{\mu\sigma^2}{2}\|\Pi_\perp y\|^2\right) + \left\langle x - y, \sigma^2\mu\Pi_\perp^\top \Pi_\perp y - \nabla f(y)\right\rangle.
\end{aligned}
$$

Next, if $\phi(x) = f(x) + \frac{\sigma^2\mu}{2}\|\Pi_\perp x\|^2$, then $\nabla\phi(x) = \nabla f(x) + \mu\sigma^2\Pi_\perp^\top\Pi_\perp x$. From here, it's not hard to see that the above inequality is: $0 \leq \phi(x) - \phi(y) - \langle \nabla\phi(y), x - y\rangle$. Hence, the function $f(x) - \frac{\sigma\mu}{2}\|\Pi_\perp x\|^2$ is a convex function. From here, we will be ready to show the Jensen's formula. For all $x, y \in \mathbb{R}^n$, consider the convexity of ϕ, it has for all $\lambda \in [0, 1]$ the inequality:

$$
\begin{aligned}
&0 \leq \lambda\phi(x) + (1-\lambda)\phi(x) - \phi(\lambda x + (1-\lambda)y) \\
&\underset{(1)}{=} f(\lambda x + (1-\lambda)y) - \lambda f(x) - (1-\lambda)f(y) - \frac{\mu\sigma^2\lambda(1-\lambda)}{2}\|\Pi_\perp(x-y)\|^2.
\end{aligned}
$$

At (1), we skipped a lot of algebra that the reader should be familiar if they already seem the proof in for f that is strongly convex. ∎

3 the accelerated proximal gradient algorithm again

{def:st-form} In this section we introduce the algorithms usually used in Nesterov's accelerated gradient.

Definition 3.1 (similar triangle form) *Suppose that*

(i) *$F = f + g$ with f Lipschitz smooth and, g closed convex proper,*

(ii) *Denote T_B to be the proximal gradient operator on $F = f + g$,*

(iii) *Let $q_k \in [0, 1)$ for all $k \geq 0$ be a parameter of the algorithm,*

(iv) *$(B_k)_{k \geq 0}$ is a sequence such that $B_k \geq 0$ for all $k \geq 0$.*

Given the initial condition $v_{-1}, x_{-1} \in \mathbb{R}^n$. An algorithm is a similar triangle form of the accelerated proximal gradient if the iterates generated $(y_k, x_k, v_k)_{k \geq 1}$ satisfies for all $k \geq 1$:

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

{def:r-nes-seq} **Definition 3.2 (relaxed Nesterov's momentum sequence)**
We let

(i) *$(q_k)_{k \geq 0}$ be a sequence such that $q_k \in (0, 1]$ for all $k \geq 1$,*

(ii) *$(\alpha_k)_{k \geq 0}$ be a sequence such that, $\alpha_0 \in (0, 1]$, and $\alpha_k \in (q_k, 1)$ for all $k \geq 1$.*

For all $k \geq 1$ we define the sequence $(\rho_k)_{k \geq 0}$:

$$
\rho_{k-1} = \frac{\alpha_k(\alpha_k - q_k)}{(1 - \alpha_k)\alpha_{k-1}^2}.
$$

4 Convergence of accelerated gradient

{lemma:cnvg-prep-part1}

4.1 preparing for the convergence proof

Lemma 4.1 (Convergence preparation part I) *Let the sequence $(y_k, x_k, v_k)_{k \geq 1}$ be a sequence generated by Definition 3.1. Let $\bar{x} \in \mathbb{R}^n$ be arbitrary, define $z_k = \alpha_k \bar{x} + (1 - \alpha_k)x_{k-1} -$*

y_k. *Then the sequence of iterates satisfies for all $k \geq 1$ all of the following equalities:*

$$\begin{aligned}
z_k - y_k &= (1-q_k)^{-1}\left((\alpha_k - q_k)(\bar{x} - v_{k-1}) + q_k(1-\alpha_k)(\bar{x} - x_{k-1})\right), \\
z_k - x_k &= \alpha_k(\bar{x} - v_k), \\
y_k &= x_{k-1} + (1-q_k)^{-1}(\alpha_{k-1} - 1)(\alpha_k - q_k)(x_{k-1} - x_{k-2}).
\end{aligned}$$

Proof. From the update of y_k in Definition 3.1, we have:

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

Recall that $z_k = \alpha_k \bar{x} + (1-\alpha_k)x_{k-1} - y_k$. Therefore, using previous results we have:

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

Using some basic algebra we also have:

$$\begin{aligned}
z_k - x_k &= \alpha_k \bar{x} + (1-\alpha_k)x_{k-1} - x_k \\
&= \alpha_k \bar{x} + x_{k-1} - \alpha_k x_{k-1} - x_k \\
&= \alpha_k(\bar{x} + \alpha_k^{-1} x_{k-1} - x_{k-1} - \alpha_k^{-1} x_k) \\
&= \alpha_k(\bar{x} + \alpha_k^{-1}(x_{k-1} - x_k) - x_{k-1}) \\
&= \alpha_k(\bar{x} - v_k).
\end{aligned}$$

Finally, y_k can be expressed using x_k only. Starting with the first recurrence relation it has:

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

After some algebra it has:

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

■

{lemma:cnvg-prep-part2} **Lemma 4.2 (Convergence preparations part II)**
Suppose that sequences $(y_k, x_k, v_k)_{k\ge 0}$ satisfies Definition 3.1. In addition assume that the sequence $(\alpha_k)_{k\ge0}$ satisfies Definition 3.2. For any arbitrary $\bar{x} \in \mathbb{R}^n$, define for all $k \ge 1$ the sequence $z_k = \alpha_k\bar{x} + (1-\alpha_k)x_{k-1}$. Then, for all $k \ge 1$, it has:

$$
\begin{aligned}
&\frac{1-q_k}{2}\|z_k - y_k\|^2 - \frac{q_k\alpha_k(1-\alpha_k)}{2}\|\bar{x} - x_{k-1}\|^2 \\
&= \frac{\alpha_{k-1}\rho_{k-1}(1-\alpha_k)}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{q_k(1-\alpha_k)(q_k - \alpha_k)}{1-q_k}\|v_{k-1} - x_{k-1}\|^2 \\
&\le \frac{\alpha_{k-1}\rho_{k-1}(1-\alpha_k)}{2}\|\bar{x} - v_{k-1}\|^2.
\end{aligned}
$$

Proof.

$$
\begin{aligned}
\frac{(\alpha_k - q_k)^2}{2(1-q_k)} - \frac{\alpha_{k-1}\rho_{k-1}(1-\alpha_k)}{2} &\underset{(1)}{=} \frac{(\alpha_k - q_k)^2}{2(1-q_k)} - \frac{\alpha_k(\alpha_k - q_k)}{2} \\
&= \frac{1}{2(1-q_k)}\left((\alpha_k - q_k)^2 - (1-q_k)\alpha_k(\alpha_k - q_k)\right) \\
&= \frac{\alpha_k - q_k}{2(1-q_k)}\left(\alpha_k - q_k - (1-q_k)\alpha_k\right) \\
&= \frac{\alpha_k - q_k}{2(1-q_k)}\left(-q_k + q_k\alpha_k\right) \\
&= \frac{(\alpha_k - q_k)q_k(\alpha_k - 1)}{1-q_k}.
\end{aligned}
\tag{4.1}
$$

{eqn:cnvg-prep-part2-eqn1}

At (1), we used the relation $\alpha_{k-1}\rho_{k-1}(1-\alpha_k) = \alpha_k(\alpha_k - q_k)$ because the sequence $(\alpha_k)_{k\ge 1}$ satisfies Definition 3.2. Next, we have:

$$
\begin{aligned}
\frac{q_k^2(1-\alpha_k)^2}{1-q_k} - q_k\alpha_k(1-\alpha_k) &= \frac{q_k(1-\alpha_k)}{1-q_k}\left(q_k(1-\alpha_k) - \alpha_k(1-q_k)\right) \\
&= \frac{q_k(1-\alpha_k)(q_k - \alpha_k)}{1-q_k}.
\end{aligned}
\tag{4.2}
$$

{eqn:cnvg-prep-part2-eqn2}

Next, using two of the above results we have:

$$
\begin{aligned}
&\frac{1-q_k}{2}\|z_k - y_k\|^2 - \frac{q_k\alpha_k(1-\alpha_k)}{2}\|\bar{x} - x_{k-1}\|^2 \\
&\underset{(1)}{=} \frac{1-q_k}{2}\left\|\left(\frac{\alpha_k - q_k}{1-q_k}\right)(\bar{x} - v_{k-1}) + \left(\frac{q_k(1-\alpha_k)}{1-q_k}\right)(\bar{x} - x_{k-1})\right\|^2 - \frac{q_k\alpha_k(1-\alpha_k)}{2}\|\bar{x} - x_{k-1}\|^2 \\
&= \frac{(\alpha_k - q_k)^2}{2(1-q_k)}\|\bar{x} - v_{k-1}\|^2 + \frac{q_k^2(1-\alpha_k)^2}{2(1-q_k)}\|\bar{x} - x_{k-1}\|^2 + \frac{(\alpha_k - q_k)q_k(1-\alpha_k)}{1-q_k}\langle \bar{x} - v_{k-1}, \bar{x} - x_{k-1}\rangle \\
&\quad - \frac{1}{2}q_k\alpha_k(1-\alpha_k)\|\bar{x} - x_{k-1}\|^2 \\
&= \frac{(\alpha_k - q_k)^2}{2(1-q_k)}\|\bar{x} - v_{k-1}\|^2 + \frac{1}{2}\left(\frac{q_k^2(1-\alpha_k)^2}{1-q_k} - q_k\alpha_k(1-\alpha_k)\right)\|\bar{x} - x_{k-1}\|^2 \\
&\quad + \frac{(\alpha_k - q_k)q_k(1-\alpha_k)}{1-q_k}\langle \bar{x} - v_{k-1}, \bar{x} - x_{k-1}\rangle \\
&= \frac{(\alpha_k - q_k)^2}{2(1-q_k)}\|\bar{x} - v_{k-1}\|^2 + \frac{1}{2}\left(\frac{q_k^2(1-\alpha_k)^2}{1-q_k} - q_k\alpha_k(1-\alpha_k)\right)\|\bar{x} - x_{k-1}\|^2 \\
&\quad + \frac{(\alpha_k - q_k)q_k(1-\alpha_k)}{1-q_k}\langle \bar{x} - v_{k-1}, \bar{x} - x_{k-1}\rangle \\
&= \frac{1}{2}\left(\frac{(\alpha_k - q_k)^2}{1-q_k} - \alpha_{k-1}^2\rho_{k-1}(1-\alpha_k)\right)\|\bar{x} - v_{k-1}\|^2 \\
&\quad + \frac{1}{2}\left(\frac{q_k^2(1-\alpha_k)^2}{1-q_k} - q_k\alpha_k(1-\alpha_k)\right)\|\bar{x} - x_{k-1}\|^2
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{(\alpha_k - q_k)q_k(1-\alpha_k)}{1-q_k}\langle \bar{x} - v_{k-1}, \bar{x} - x_{k-1}\rangle - \frac{\alpha_{k-1}^2\rho_{k-1}(1-\alpha_k)}{2}\|\bar{x} - v_{k-1}\|^2 \\
\underset{(2)}{=}\ & \frac{1}{2}\frac{(\alpha_k - q_k)q_k(\alpha_k - 1)}{1-a_k}\|\bar{x} - v_{k-1}\|^2 + \frac{1}{2}\left(\frac{q_k^2(1-\alpha_k)^2}{1-q_k} - q_k\alpha_k(1-\alpha_k)\right)\|\bar{x} - x_{k-1}\|^2 \\
&+ \frac{(\alpha_k - q_k)q_k(1-\alpha_k)}{1-q_k}\langle \bar{x} - v_{k-1}, \bar{x} - x_{k-1}\rangle - \frac{\alpha_{k-1}^2\rho_{k-1}(1-\alpha_k)}{2}\|\bar{x} - v_{k-1}\|^2 \\
\underset{(3)}{=}\ & \frac{1}{2}\frac{(\alpha_k - q_k)q_k(\alpha_k - 1)}{1-a_k}\|\bar{x} - v_{k-1}\|^2 + \frac{1}{2}\frac{q_k(1-\alpha_k)(q_k - \alpha_k)}{1-q_k}\|\bar{x} - x_{k-1}\|^2 \\
&+ \frac{(\alpha_k - q_k)q_k(1-\alpha_k)}{1-q_k}\langle \bar{x} - v_{k-1}, \bar{x} - x_{k-1}\rangle - \frac{\alpha_{k-1}^2\rho_{k-1}(1-\alpha_k)}{2}\|\bar{x} - v_{k-1}\|^2 \\
=\ & \frac{\alpha_{k-1}^2\rho_{k-1}(1-\alpha_k)}{2}\|\bar{x} - v_{k-1}\|^2 + \frac{q_k(1-\alpha_k)(q_k - \alpha_k)}{1-q_k}\|v_{k-1} - x_{k-1}\|^2. \\
\underset{(4)}{\le}\ & \frac{\alpha_{k-1}^2\rho_{k-1}(1-\alpha_k)}{2}\|\bar{x} - v_{k-1}\|^2.
\end{aligned}
$$

At (1) we used one of the results from Lemma 4.1, at (2) we used (4.1), and at (3) we used (4.2). At (4), we used the fact that $\alpha_k \in (q_k, 1)$, which means that the coefficient $q_k(1-\alpha_k)(q_k - \alpha_k) \le 0$, therefore we can simplify the term away. ■

Theorem 4.3 (the convergence results)

Proof.

■

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.