Inexect Accelerated Proximal Gradient

Author 1 Name, Author 2 Name *

September 25, 2025

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

Abstract

This is still a draft. [3].

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

1 Introduction

Notations. Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, we denote $g^\star$ to be the Fenchel conjugate. $I : \mathbb{R}^n \to \mathbb{R}^n$ denotes the identity operator. For a multivalued mapping $T : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, $\operatorname{gra} T$ denotes the graph of the operator, defined as $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : y \in Tx\}$.

1.1 Epsilon subgradient and inexact proximal point

{def:esp-subgrad}

Definition 1.1 (ϵ-subgradient) *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc. Let $\epsilon \geq 0$. Then the ϵ-subgradient of g at some $\bar x \in \operatorname{dom} g$ is given by:*

$$\partial g_\epsilon(\bar x) := \{v \in \mathbb{R}^n \mid \langle v, x - \bar x\rangle \leq g(x) - g(\bar x) + \epsilon \;\forall x \in \mathbb{R}^n\}.$$

When $\bar x \notin \operatorname{dom} g$, it has $\partial g_\epsilon(\bar x) = \emptyset$.

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

Remark 1.2 $\partial_\epsilon g$ is a multivalued operator and, it's not monotone, unless $\epsilon = 0$, which makes it equivalent to French subgradient ∂g.

If we assume lsc, proper and convex g, we will now introduce results in the literatures that we will use.

{fact:esp-fenchel-ineq}

Fact 1.3 (ϵ-Fenchel inequality) *Let $\epsilon \geq 0$, then:*

$$x^* \in \partial_\epsilon f(\bar{x}) \iff f^\star(x^*) + f(\bar{x}) \leq \langle x^*, \bar{x}\rangle + \epsilon \implies \bar{x} \in \partial_\epsilon f^\star(x^*).$$

They are all equivalent if $f^{\star\star}(\bar{x}) = f(\bar{x})$.

Remark 1.4 The above fact is taken from Zalinascu [2, Theorem 2.4.2].

{def:inxt-pp}

We will now define inexact proximal point based on ϵ-subgradient

Definition 1.5 (inexact proximal point) *For all $x \in \mathbb{R}^n, \epsilon \geq 0, \lambda > 0$, $\tilde{x}$ is an inexact evaluation of proximal point at x, if and only if it satisfies:*

$$\lambda^{-1}(x - \tilde{x}) \in \partial_\epsilon g(\tilde{x}).$$

We denote it by $\tilde{x} \approx_\epsilon \operatorname{prox}_{\lambda g}(x)$.

{fact:resv-identity}

Remark 1.6 This definition is nothing new, for example see Villa et al. [1, Definition 2.1]

Fact 1.7 (the resolvant identity) *Let $T : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, then it has:*

$$(I + T)^{-1} = (I - (I + T^{-1})^{-1}).$$

Theorem 1.8 (inexact Moreau decomposition) *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a closed, convex and proper function. It has the equivalence*

$$\tilde{y} \approx_\epsilon \operatorname{prox}_{\lambda^{-1} g^\star}(\lambda^{-1} y) \iff y - \lambda\tilde{y} \approx_\epsilon \operatorname{prox}_{\lambda g}(y).$$

Proof. Consider $\tilde{y} \approx_\epsilon \operatorname{prox}_{\lambda^{-1} g^\star}(\lambda^{-1} y)$, then it has:

$$\begin{aligned}
& \tilde{y} \in (I + \lambda^{-1}\partial_\epsilon g^\star)^{-1}(\lambda^{-1} y) \\
\iff & (\lambda^{-1} y, \tilde{y}) \in \operatorname{gra}(I + \lambda^{-1}\partial_\epsilon g^\star)^{-1} \\
\underset{(1)}{\iff} & (\lambda^{-1} y, \tilde{y}) \in \operatorname{gra}(I - (I + \partial_\epsilon g \circ (\lambda I))^{-1}) \\
\iff & (\lambda^{-1} y, \lambda^{-1} y - \tilde{y}) \in \operatorname{gra}(I + \partial_\epsilon g \circ (\lambda I))^{-1} \\
\iff & (\lambda^{-1} y - \tilde{y}, \lambda^{-1} y) \in \operatorname{gra}(I + \partial_\epsilon g \circ (\lambda I)) \\
\iff & (y - \lambda\tilde{y}, \lambda^{-1} y) \in \operatorname{gra}(\lambda^{-1} I + \partial_\epsilon g) \\
\iff & (y - \lambda\tilde{y}, y) \in \operatorname{gra}(I + \lambda\partial_\epsilon g) \\
\iff & y - \lambda\tilde{y} \in (I + \lambda\partial_\epsilon g)^{-1} y \\
\iff & y - \lambda\tilde{y} \approx_\epsilon \operatorname{prox}_{\lambda g}(y).
\end{aligned}$$

At (1) we can use Fact 1.7, and it has $(\lambda^{-1}\partial_\epsilon g^\star)^{-1} = \partial_\epsilon g \circ (\lambda I)$ by Fact 1.3 and the assumption that g is closed, convex and proper. ■

1.2 Inexact proximal gradient inequality

{ass:for-inxt-pg-ineq}

Assumption 1.9 (for inexact proximal gradient) The assumption is about (f, g, L). We assume that

(i) $f : \mathbb{R}^n \to \mathbb{R}$ is a convex, L Lipschitz function.
(ii) $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex, proper, and lsc function which we do not have its exact proximal operator.

{def:inxt-pg} We develop the theory based on the use of epsilon subgradient as in Definition 1.1.

Definition 1.10 (inexact proximal gradient) *Let (f, g, L) satisfies Assumption 1.9. Let $\epsilon \ge 0, \rho > 0$. Then, $\tilde x \approx_\epsilon T_\rho(x)$ is an inexact proximal gradient if it satisfies variational inequality:*

$$\mathbf{0} \in \nabla f(x) + \rho(x - \tilde x) + \partial_\epsilon g(\tilde x).$$

Remark 1.11 We assumed that we can get exact evaluation of ∇f at any points $x \in \mathbb{R}^n$.

Lemma 1.12 (other representations of inexact proximal gradient) *Let (f, g, L) satisfies Assumption 1.9, $\epsilon \ge 0, \rho > 0$, then for all $x \approx_\epsilon T_\rho(x)$, it has the following equivalent representations:*

$$\begin{aligned}
&(x - \rho^{-1}\nabla f(x)) - \tilde x \in \rho^{-1}\partial_\epsilon g(\tilde x) \\
\iff &\tilde x \in (I + \rho^{-1}\partial_\epsilon g(\tilde x))^{-1}(x - \rho^{-1}\nabla f(x)) \\
\iff &x \approx_\epsilon \operatorname{prox}_{\rho^{-1}g}\left(x - \rho^{-1}\nabla f(x)\right)
\end{aligned}$$

Proof. It's direct. ■

{thm:inxt-pg-ineq} **Theorem 1.13 (inexact over-regularized proximal gradient inequality)** *Let (f, g, L) satisfies Assumption 1.9, $\epsilon \ge 0, B \ge 0, \rho > 0$. Consider $\tilde x \approx_\epsilon T_{B+\rho}(x)$. Denote $F = f + g$. If in addition, $\tilde x, B$ satisfies the line search condition $D_f(\tilde x, x) \le B/2\|x - \tilde x\|^2$, then it has $\forall z \in \mathbb{R}^n$:*

$$-\epsilon \le F(z) - F(\tilde x) + \frac{B+\rho}{2}\|x - z\|^2 - \frac{B+\rho}{2}\|z - \tilde x\|^2 - \frac{\rho}{2}\|\tilde x - x\|^2.$$

Proof. By Definition 1.10 write the variational inequality that describes $\tilde{x} \approx_\epsilon T_B(x)$, and the definition of epsilon subgradient (Definition 1.1) it has for all $z \in \mathbb{R}^n$:

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

At (1), we used considered the following:

$$\begin{aligned}
\langle \nabla f(x), z - x\rangle &= \langle \nabla f(x), z - x + x - \tilde{x}\rangle \\
&= \langle \nabla f(x), z - x\rangle + \langle \nabla f(x), x - \tilde{x}\rangle \\
&= -D_f(z, x) + f(z) - f(x) + D_f(\tilde{x}, x) - f(\tilde{x}) + f(x) \\
&= -D_f(z, x) + f(z) + D_f(\tilde{x}, x) - f(\tilde{x}).
\end{aligned}$$

At (2), we used the fact that f is convex hence $-D_f(z, x) \le 0$ always, and in the statement hypothesis we assumed that B has $D_f(\tilde{x}, x) \le B/2\|\tilde{x} - x\|^2$. ■

1.3 Optimizing the inexact proximal point problem

In this section we will show an optimization problem that allows us to solve for some $\tilde{x} \approx_\epsilon \operatorname{prox}_{\lambda g}(z)$. Most of these results are from the literature. To start, we must assume the following about a function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, with g closed, convex and proper.

{ass:for-inxt-prox} **Assumption 1.14 (for inexact proximal operator)**
This assumption is about (g, ω, A). Let $m \in \mathbb{N}, n \in \mathbb{R}^n$, we assume that

(i) $A \in \mathbb{R}^{m\times n}$ is a matrix.
(ii) $\omega : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a closed and convex function such that it admits proximal operator $\operatorname{prox}_{\lambda\omega}$ and, its conjugate $\omega^\star$ is known.
(iii) $g := \omega(Ax)$ such that $\operatorname{rng} A \cap \operatorname{ri}\operatorname{dom} g \neq \emptyset$.

Now, we are ready to discuss how to choose $\tilde{x} \approx_\epsilon \operatorname{prox}_{\lambda g}(x)$. Fix $y \in \mathbb{R}^n, \lambda > 0$, we are ultimately interested in minimizing:

$$\Phi_\lambda(u) := \omega(Au) + \frac{1}{2\lambda}\|u - y\|^2 \tag{1.1}$$

This problem admits dual objective in $\mathbb{R}^m$:

$$\Psi_\lambda(v) := \frac{1}{2\lambda}\|\lambda A^\top v - y\|^2 + \omega^\star(v) - \frac{1}{2\lambda}\|y\|^2. \tag{1.2}$$

We define the duality gap

$$\mathbf{G}_\lambda(u, v) := \Phi_\lambda(u) + \Psi_\lambda(v). \tag{1.3}$$

If strong duality holds, it exists $(\hat u, \hat v)$ such that we have the following:

$$\mathbf{G}_\lambda(\hat u, \hat v) = 0 = \min_u \Phi_\lambda(u) + \min_v \Psi_\lambda(v)$$

The following theorem quantifies a sufficient conditions for $\tilde x \approx_\epsilon \operatorname{prox}_{\lambda g}(x)$. The theorem below is from [1, Proposition 2.2].

{thm:primal-dual-trans}

Theorem 1.15 (primal translate to dual) *Let (g, ω, A) satisfies assumption 1.14, $\epsilon \geq 0$, then*

$$\left(\forall z \approx_\epsilon \operatorname{prox}_{\lambda g}(y)\right)(\exists v \in \operatorname{dom}\omega^\star) : z = y - \lambda A^\top v.$$

This theorem that follows is from Villa et al. [1, Proposition 2.3], but put into our symbols and, Definition

{thm:dlty-gap-inxt-pp}

Theorem 1.16 (duality gap of inexact proximal problem) *Let (g, ω, A) satisfies Assumption 1.14, for all $\epsilon \geq 0$, $v \in \mathbb{R}^n$ consider the following conditions:*

(i) $\mathbf{G}_\lambda(y - \lambda A^\top v, v) \leq \epsilon$.
(ii) $A^\top v \approx_\epsilon \operatorname{prox}_{\lambda^{-1} g^\star}(\lambda^{-1} y)$.
(iii) $y - \lambda A^\top v \approx_\epsilon \operatorname{prox}_{\lambda g}(y)$.

They have $(a) \implies (b) \iff (c)$. If in addition $\omega^\star(v) = g^\star(A^\top v)$, then all three conditions are equivalent.

The following fact from the literature indicates that it's sufficient to minimize the dual problem Ψ_λ to obtain an element of the inexact proximal point operator. The following fact is Proposition [1, Theorem 5.1].

Fact 1.17 (minimizing dual of the proximal problem) *Let $\bar v$ be a solution of Ψ. Suppose that $(v_n)_{n\geq 0}$ is a minimizing sequence for Ψ. Let $z_n = y - \lambda A^\top v_n$, and $\bar z = y - \lambda A^\top \bar v$. If in addition, Φ_λ is L_1 Lipschitz continuous, then it has for all $k \geq 0$ the inequality:*

$$\Phi_\lambda(z_n) - \Phi_\lambda(\bar x) \leq L_1\|z_n - \bar z\| \leq L_1\sqrt{2\lambda}(\Psi_\lambda(v_n) - \Psi_\lambda(\bar v))^{1/2}.$$

We remark that the above fact translates any algorithm that optimizes the function value of the dual problem into optimizing duality gap $\mathbf{G}(z_n, v_n)$. For this reasons, the inner loop interation required to achieve $\mathbf{G}(z_n, v_n) < \epsilon$ is now related to the convergence rate of the algorithms used to optimize $\Psi_\lambda(v_n)$.

STILL WRITING AND NOT FINISHED YET!

1.4 Literature reviews

1.5 Our contributions

2 The accelerated proximal gradient with controlled errors

In this section, we present an accelerated algorithm with controlled error using Definition 1.10, and show that it can have a convergence rate under certain error conditions.

{def:inxt-apg} **Definition 2.1 (our inexact accelerated proximal gradient)**
Suppose that (F, f, g, L) and, sequences $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\ge 0}$ satisfies the following

(i) $(\alpha_k)_{k\ge 0}$ *is a sequence such that* $\alpha \in (0, 1]$ *for all* $k \ge 0$.
(ii) $(B_k)_{k\ge 0}$ *is a non-negative sequence, characterizing the potential line search routine.*
(iii) $(\rho_k)_{k\ge 0}$ *be a sequence such that* $\rho_k > 0$, *characterizing the over-relaxation of the proximal gradient operator.*
(iv) $(\epsilon_k)_{k\ge 0}$ *is a non-negative sequence characterizing the errors of inexact proximal evaluation.*
(v) (f, g, L) *satisfies Assumption 1.9, and let* $F = f + g$.

Denote $L_k = B_k + \rho_k$ *for short. Given any initial condition* $v_{-1}, x_{-1} \in \mathbb{R}^n$, *the algorithm generates the sequences* $(y_k, x_k, v_k)_{k\ge 0}$ *such that they satisfy for all* $k \ge 0$:

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

{lemma:inxt-apg-cnvg-prep1} **Lemma 2.2 (inexact accelerated proximal gradient preparation stage I)**
Let (f, g, L), *and* $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\ge 0}$, *be given by Definition 2.1. Denote* $L_k = B_k + \rho_k$. *Then,*

for any $\bar x \in \mathbb{R}^n$, the sequences $(y_k, x_k, v_k)_{k\ge 0}$ generated satisfy for all $k \ge 1$ the inequality:

$$
\begin{aligned}
&\frac{\rho_k}{2}\|x_k - y_k\|^2 - \epsilon_k \\
&\le (1-\alpha_k)(F(x_{k-1}) - F(\bar x)) + F(\bar x) - F(x_k) \\
&+ \max\left(1-\alpha_k, \frac{\alpha_k^2 L_k}{\alpha_{k-1}^2 L_{k-1}}\right)\frac{\alpha_{k-1}^2 L_{k-1}}{2}\|\bar x - v_{k-1}\|^2 - \frac{\alpha_k^2 L_k}{2}\|\bar x - v_k\|^2.
\end{aligned}
$$

When, $k = 1$ it instead has:

$$
\begin{aligned}
&\frac{\rho_0}{2}\|x_0 - y_0\|^2 - \epsilon_0 \\
&\le (1-\alpha_0)(F(x_{-1}) - F(\bar x)) + F(\bar x) - F(x_0) + \frac{\alpha_0^2 L_0}{2}\|\bar x - v_{-1}\|^2 - \frac{\alpha_0^2 L_0}{2}\|\bar x - v_0\|^2.
\end{aligned}
$$

Proof. Two intermediate results are in order before we can prove the inequality. Define $z_k := \alpha_k \bar x + (1-\alpha_k)x_{k-1}$ for short. It has for all $k \ge 1$ the equality:

{eqn:inxt-apg-cnvg-prep1-a}

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

It also has for all $k \ge 1$ the equality:

{eqn:inxt-apg-cnvg-prep1-b}

$$
\begin{aligned}
z_k - y_k &= \alpha_k \bar x + (1-\alpha_k)x_{k-1} - y_k \\
&= \alpha_k \bar x - \alpha_k v_{k-1}.
\end{aligned} \tag{b}
$$

Let's denote $L_k = B_k + \rho_k$ for short. Recall that (f, g, L) satisfies Assumption 1.9, if we choose $x = y_k$ so $\tilde x = x_k \approx_\epsilon T_{L_k}(y_k)$, and set $z = z_k, \epsilon = \epsilon_k$ then Theorem 1.13 has:

$$
\begin{aligned}
&\frac{\rho_k}{2}\|x_k - y_k\|^2 - \epsilon_k \\
&\le F(z_k) - F(x_k) + \frac{L_k}{2}\|y_k - z_k\|^2 - \frac{L_k}{2}\|z_k - x_k\|^2 \\
&\underset{(1)}{\le} \alpha_k F(\bar x) + (1-\alpha_k)F(x_{k-1}) - F(x_k) + \frac{L_k}{2}\|y_k - z_k\|^2 - \frac{L_k}{2}\|z_k - x_k\|^2 \\
&\underset{(2)}{=} (1-\alpha_k)(F(x_{k-1}) - F(\bar x)) + F(\bar x) - F(x_k) + \frac{\alpha_k^2 L_k}{2}\|\bar x - v_{k-1}\|^2 - \frac{\alpha_k^2 L_k}{2}\|\bar x - v_k\|^2 \\
&\le (1-\alpha_k)(F(x_{k-1}) - F(\bar x)) + F(\bar x) - F(x_k) \\
&+ \max\left(1-\alpha_k, \frac{\alpha_k^2 L_k}{\alpha_{k-1}^2 L_{k-1}}\right)\frac{\alpha_{k-1}^2 L_{k-1}}{2}\|\bar x - v_{k-1}\|^2 - \frac{\alpha_k^2 L_k}{2}\|\bar x - v_k\|^2.
\end{aligned}
$$

At (1) we used the fact that $F = f + g$ hence F is convex. At (2) we used (a), (b). Finally, if $k = 0$, then take the RHS of $\underset{(1)}{=}$ then:

$$
\begin{aligned}
&\frac{\rho_0}{2}\|x_0 - y_0\|^2 - \epsilon_0 \\
&\le (1-\alpha_0)(F(x_{-1}) - F(\bar{x})) + F(\bar{x}) - F(x_0) + \frac{\alpha_0^2 L_0}{2}\|\bar{x} - v_{-1}\|^2 - \frac{\alpha_0^2 L_0}{2}\|\bar{x} - v_0\|^2.
\end{aligned}
$$

■

The following proposition is a prototype of the convergence rate together with the error schedule that delivers convergence of algorithms satisfying Definition 2.1.

{prop:inxt-apg-cnvg-generic} **Proposition 2.3 (valid error schedule and convergence rate)** *Let (f, g, L), $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\ge 0}$ be given by Definition 2.1. Fix any $\bar{x} \in \mathbb{R}^n$ for all $k \ge 0$ and assume that $\alpha_0 = 1$. Denote for brevity $\beta_0 = 1$, $\beta_k = \prod_{i=1}^k \max\left(1 - \alpha_i, \frac{\alpha_i^2 L_i}{\alpha_{i-1}^2 L_{i-1}}\right)$ and $L_k = B_k + \rho_k$. If for some fixed $\mathcal{E}_0 \ge 0, p \ge 1$ the parameter ρ_k, ϵ_k can satisfy for all $k \ge 0$ the condition*

$$
\frac{-\mathcal{E}_0 \beta_k}{k^p} \le \frac{\rho_k}{2}\|x_k - y_k\|^2 - \epsilon_k.
$$

Then for the sequence generated $(y_k, x_k, v_k)_{k\ge 0}$ by the algorithm, for all $k \ge 0$ they satisfy:

$$
F(x_k) - F(\bar{x}) + \frac{\alpha_k^2 L_k}{2}\|\bar{x} - v_k\|^2 \le \beta_k\left(\frac{L_0}{2}\|\bar{x} - v_{-1}\|^2 + \mathcal{E}_0 \sum_{n=1}^{k} \frac{1}{n^p}\right).
$$

Proof. Consider results from Lemma 2.2 has $\forall k \ge 1$:

$$
\begin{aligned}
&\frac{\rho_k}{2}\|x_k - y_k\|^2 - \epsilon_k \\
&\le (1-\alpha_k)(F(x_{k-1}) - F(\bar{x})) + F(\bar{x}) - F(x_k) \\
&+ \max\left(1-\alpha_k, \frac{\alpha_k^2 L_k}{\alpha_{k-1}^2 L_{k-1}}\right)\frac{\alpha_{k-1}^2 L_{k-1}}{2}\|\bar{x} - v_{k-1}\|^2 - \frac{\alpha_k^2 L_k}{2}\|\bar{x} - v_k\|^2. \\
&\le \max\left(1-\alpha_k, \frac{\alpha_k^2 L_k}{\alpha_{k-1}^2 L_{k-1}}\right)\left(F(x_{k-1}) - F(\bar{x}) + \frac{\alpha_{k-1}^2 L_{k-1}}{2}\|\bar{x} - v_{k-1}\|^2\right) \\
&\quad + F(\bar{x}) - F(x_k) - \frac{\alpha_k^2 L_k}{2}\|\bar{x} - v_k\|^2
\end{aligned}
$$

For notation brevity, we introduce β_k, Λ_k:

$$
\begin{aligned}
&\beta_0 = 1, \\
&\beta_k := \prod_{i=1}^{k} \max\left(1-\alpha_i, \frac{\alpha_i^2 L_i}{\alpha_{i-1}^2 L_{i-1}}\right), \\
&\Lambda_k := -F(\bar{x}) + F(x_k) + \frac{\alpha_k^2 L_k}{2}\|\bar{x} - v_k\|^2.
\end{aligned}
$$

Now, suppose that in addition there is a non-negative sequence $(\mathcal{E}_k)_{k\geq 0}$ such that

(i) For all $k \geq 0$, it has $\frac{-\mathcal{E}_k}{k^p} \leq (\rho_k/2)\|x_k - y_k\|^2 - \epsilon_k$ where $p \geq 1$,
(ii) For all $k \geq 1$, it has $\mathcal{E}_k = \frac{\beta_k}{\beta_{k-1}}\mathcal{E}_{k-1}$, with $\mathcal{E}_0 \geq 0$.

These conditions are equivalent to the assumption that $\frac{-\mathcal{E}_0\beta_k}{k^p} \leq \frac{\rho_k}{2}\|x_k - y_k\|^2 - \epsilon_k$. One can show that by unrolling recurrence on $\mathcal{E}_k$. Then (2.1) implies $\forall k \geq 1$:

{ineq:inxt-apg-cnvg-generic-pitem-1}

$$
\frac{-\mathcal{E}_k}{k^p} \leq \frac{\beta_k}{\beta_{k-1}}\Lambda_{k-1} - \Lambda_k \iff \Lambda_k \leq \frac{\beta_k}{\beta_{k-1}}\Lambda_{k-1} + \frac{\mathcal{E}_k}{k^p}. \tag{2.1}
$$

Now, we show the convergence of Λ_k, using the relations of $\mathcal{E}_k, \Lambda_k, \beta_k$ above.

$$
\begin{aligned}
\Lambda_k &\leq \frac{\beta_k}{\beta_{k-1}}\Lambda_{k-1} + \frac{\mathcal{E}_k}{k^p} \\
&\leq \frac{\beta_k}{\beta_{k-1}}\Lambda_{k-1} + \frac{\beta_k}{\beta_{k-1}}\frac{\mathcal{E}_{k-1}}{k^p} \\
&= \frac{\beta_k}{\beta_{k-1}}\left(\Lambda_{k-1} + \frac{\mathcal{E}_{k-1}}{k^p}\right) \\
&\leq \frac{\beta_k}{\beta_{k-1}}\left(\frac{\beta_{k-1}}{\beta_{k-2}}\Lambda_{k-2} + \frac{\mathcal{E}_{k-1}}{(k-1)^p} + \frac{\mathcal{E}_{k-1}}{k^p}\right) \\
&= \frac{\beta_k}{\beta_{k-2}}\left(\Lambda_{k-2} + \frac{\mathcal{E}_{k-2}}{(k-1)^p} + \frac{\mathcal{E}_{k-2}}{k^p}\right) \\
&\ldots \\
&\leq \frac{\beta_k}{\beta_1}\left(\Lambda_1 + \mathcal{E}_1\sum_{n=2}^{k}\frac{1}{n^p}\right) \\
&\leq \frac{\beta_k}{\beta_1}\left(\frac{\beta_1}{\beta_0}\Lambda_0 + \mathcal{E}_1\sum_{n=1}^{k}\frac{1}{n^p}\right) \\
&= \frac{\beta_k}{\beta_0}\left(\Lambda_0 + \mathcal{E}_0\sum_{n=1}^{k}\frac{1}{n^p}\right).
\end{aligned}
$$

Therefore, it points to the following inequality:

$$
\begin{aligned}
&F(x_k) - F(\bar{x}) + \frac{\alpha_k^2 L_k}{2}\|\bar{x} - v_k\|^2 \\
&\leq \beta_k \left(F(x_0) - F(\bar{x}) + \frac{\alpha_0^2 L_0}{2}\|\bar{x} - v_0\|^2 + \mathcal{E}_0 \sum_{n=1}^{k} \frac{1}{n^p} \right).
\end{aligned}
$$

Finally, when $\alpha_0 = 1$, then the results from 2.2 with $k = 0$ simplifies the above inequality and give:

$$
F(x_k) - F(\bar{x}) + \frac{\alpha_k^2 L_k}{2}\|\bar{x} - v_k\|^2 \leq \beta_k \left(\frac{L_0}{2}\|\bar{x} - v_{-1}\|^2 + \mathcal{E}_0 \sum_{n=1}^{k} \frac{1}{n^p} \right).
$$

■

Now, it only remains to determine the sequence α_k to derive a type of convergence rate for the algorithm because from the above theorem, we have the convergence rate β_k and, the error parameters ϵ_k, ρ_k both controlled by the sequence $(\alpha_k)_{k\geq 0}$.

3 Linear convergence for the inner loop proximal problem

References

[1] S. Villa, S. Salzo, L. Baldassarre, and A. Verri, *Accelerated and inexact forward-backward algorithms*, SIAM Journal on Optimization, 23 (2013), pp. 1607–1633.

[2] C. Zalinescu, *Convex analysis in general vector spaces*, World Scientific, River Edge, N.J. ; London, 2002.

[3] M. Zhang, M. Zhang, F. Zhang, A. Chaddad, and A. Evans, *Robust brain MR image compressive sensing via re-weighted total variation and sparse regression*, Magnetic Resonance Imaging, 85 (2022), pp. 271–286.