Inexact Accelerated Proximal Gradient

Author 1 Name, Author 2 Name *

October 1, 2025

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

Abstract

This is still a draft. [4].

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 Fenchel subgradient ∂g.

{fact:esp-fenchel-ineq}

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

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 [3, 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. [2, 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.

No, we develop the theory based on the use of epsilon subgradient as in Definition 1.1. Let $\rho > 0$, the exact proximal gradient operator defined for (f, g, L) satisfying Assumption 1.9 has

$$\begin{aligned} T_\rho(x) &= \operatorname*{argmin}_{z \in \mathbb{R}^n} \left\{ g(z) + \langle \nabla f(x), z \rangle + \frac{\rho}{2}\|z - x\|^2 \right\} \\ &= \operatorname{prox}_{\rho^{-1} g}\left(x - \rho^{-1}\nabla f(x)\right). \end{aligned}$$

{def:inxt-pg}

The following definition extends the proximal gradient operator to the inexact case using the concept of ϵ-subgradient as given by Definition 1.1.

Definition 1.10 (inexact proximal gradient) *Let (f, g, L) satisfies Assumption 1.9. Let $\epsilon \geq 0, \rho > 0$. 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:other-repr-inxt-pg}

Lemma 1.12 (other representations of inexact proximal gradient) *Let (f, g, L) satisfies Assumption 1.9, $\epsilon \geq 0, \rho > 0$, then for all $\tilde{x} \approx_\epsilon T_\rho(x)$, 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, $\epsilon \geq 0, B \geq 0, \rho > 0$. Consider $\tilde{x} \approx_\epsilon T_{B+\rho}(x)$. Denote $F = f + g$. If in addition, $\tilde{x}, B$ satisfies the line search condition $D_f(\tilde{x}, x) \leq B/2\|x - \tilde{x}\|^2$, then it has $\forall z \in \mathbb{R}^n$:*

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

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 $z \in \mathbb{R}^n$:

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

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 $-D_f(z, x) \leq 0$ always, and in the statement hypothesis we assumed that B has $D_f(\tilde{x}, x) \leq B/2\|\tilde{x} - x\|^2$. We also used $F = f + g$. ■

Remark 1.14 When $\epsilon = 0, \rho = 0$, this reduces to proximal gradient inequality in the exact case. In this inequality, observe that the parameter ϵ controls the inexactiness of the proximal gradient evaluation. Morespecifically, ϵ_k controls the absolute pertubatnions of the proximal gradient inequality compared to its exact counterpart. ρ on the otherhand, it is the over-relaxation of proximal gradient operator and it compensates the pertubations caused by ϵ relative to the term $\|\tilde{x} - x\|^2$.

1.3 Optimizing the inexact proximal point problem

In this section we will present the optimization problem that obtains a $\tilde{x}$ such that $\tilde{x} \approx_\epsilon \text{prox}_{\lambda g}(z)$. Eventually we want to evaluate $T_\rho(x)$ of some $F = f + g$ inexactly using Lemma

1.12. To do that one would need to evaluate $\operatorname{prox}_{\rho^{-1}g}$ inexactly which is defined in Definition 1.5.

Most of these results that will follow 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 1.15 (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:

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

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

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

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:

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

{thm:primal-dual-trans}
The following theorem quantifies a sufficient conditions for $\tilde{x} \approx_\epsilon \operatorname{prox}_{\lambda g}(x)$. The theorem below is from [2, Proposition 2.2].

Theorem 1.16 (primal translate to dual [2, Proposition 2.2]) *Let (g, ω, A) satisfies assumption 1.15, $\epsilon \geq 0$, then*

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

This theorem that follows is from Villa et al. [2, Proposition 2.3].

{thm:dlty-gap-inxt-pp}
Theorem 1.17 (duality gap of inexact proximal problem [2, Proposition 2.3])
Let (g, ω, A) satisfies Assumption 1.15, 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 is from Villa et al. [2, Theorem 5.1].

{fact:minimizing-dual-pp}

Fact 1.18 (minimizing dual of the proximal problem [2, Theorem 5.1]) *Let* $\bar{v}$ *be a solution of* Ψ_λ. *Suppose that* $(v_n)_{n\geq 0}$ *is a minimizing sequence for* Ψ_λ. *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:*

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

We remark that the above fact translates any algorithm that optimizes the function value of the dual problem Ψ_λ into optimizing duality gap $\mathbf{G}(z_n, v_n)$. For this reason, the number of iterations of the inner loop required to achieve $\mathbf{G}(z_n, v_n) < \epsilon$ for a given ϵ is related to the convergence rate of the algorithms used to optimize $\Psi_\lambda(v_n)$. With the theorem derived above, and using Theorem 1.17 it implies that any algorithm which can optimize function value Ψ_λ will produce iterates sufficient to achieve $\approx_\epsilon \operatorname{prox}_{\lambda g}(y)$.

1.4 Literature reviews

1.5 Our contributions

2 The inexact 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* $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\geq 0}$ *satisfies the following*

(i) $(\alpha_k)_{k\geq 0}$ *is a sequence such that* $\alpha \in (0, 1]$ *for all* $k \geq 0$.

(ii) $(B_k)_{k\geq 0}$ *is a non-negative sequence, characterizing any potential line search routine.*
(iii) $(\rho_k)_{k\geq 0}$ *be a sequence such that* $\rho_k > 0$*, characterizing the over-relaxation of the proximal gradient operator.*
(iv) $(\epsilon_k)_{k\geq 0}$ *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* $(y_k, x_k, v_k)_{k\geq 0}$ *such that they satisfy for all* $k \geq 0$*:*

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

{lemma:inxt-apg-cnvg-prep1}
Lemma 2.2 (inexact accelerated proximal gradient preparation stage I) *Let* (F, f, g, L)*, and* $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\geq 0}$*, be given by Definition 2.1. Denote* $L_k = B_k + \rho_k$*. Then, for any* $\bar{x} \in \mathbb{R}^n$*, the sequences* $(y_k, x_k, v_k)_{k\geq 0}$ *generated satisfy for all* $k \geq 1$ *the inequality:*

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

When, $k = 1$ *it instead has:*

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

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 $k \geq 1$ 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 $k \geq 1$ 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:

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

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:

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

■

{ass:valid-err-schedule}

The following assumption encapsulate assumptions on the errors such that a near optimal convergence rate is still attainable by an algorithm that satisfies Definition 2.1.

Assumption 2.3 (valid error schedule) The following assumption is about an algorithm satisfying Definition 2.1, its parameters $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\geq 0}$ in relation to its iterates $(y_k, x_k, v_k)_{k\geq 0}$ and, some additional parameters $(\beta_k)_{k\geq 0}, \mathcal{E}_0$ and p. Let

(i) $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\geq 0}, (F, f, g, L)$ and $(y_k, x_k, v_k)_{k\geq 0}$ be given by Definition 2.1.
(ii) $\mathcal{E}_0 \geq 0$ be arbitrary;
(iii) the sequence $(\beta_k)_{k\geq 0}$ be defined as $\beta_k := \prod_{i=1}^k \max\left(1-\alpha_i, \frac{\alpha_i^2 L_i}{\alpha_{i-1}^2 L_{i-1}}\right)$ for all $k \geq 1$, with the base case being $\beta_0 = 1$;
(iv) $p \geq 1$ is some constant which will bound the error ϵ_k relative to ρ_k.

In addition, we assume that the error parameter ϵ_k and over-relaxation parameter ρ_k, iterates x_k, y_k and β_k together satisfies for all $k \geq 0$ the relations:

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

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 2.4 (generic convergence rate under valid error schedule) *Let (F, f, g, L), $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\geq 0}$, $(\beta_k)_{k\geq 0}, \mathcal{E}_0, p$ as assumed in Assumption 2.3. Fix any $\bar{x} \in \mathbb{R}^n$ for all $k \geq 0$ and assume that $\alpha_0 = 1$. Then for the iterates generated $(y_k, x_k, v_k)_{k\geq 0}$ by the algorithm, for all $k \geq 0$ they will satisfy:*

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

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

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

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 $\frac{-\mathcal{E}_0\beta_k}{k^p} \leq \frac{\rho_k}{2}\|x_k - y_k\|^2 - \epsilon_k$ (which was stated in Assumption 2.3). 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.

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

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 $(\alpha_k)_{k\geq 0}$.

2.1 convergence results of the outer loop

This section will give specific instances of the error control sequence $(\epsilon_k)_{k\geq 0}, (\rho_k)_{k\geq 0}$ and, momentum sequence $(\alpha_k)_{k\geq 0}$ such that an optimal convergence rate of $\mathcal{O}(1/k^2)$ can be achieved.

{prop:opt-cnvg-outr-loop}

Proposition 2.5 ($\mathcal{O}(1/k^2)$ optimal convergence rate of the outer loop) *Let (f, g, L), $(\alpha_k, B_k, \rho_k, \epsilon_k)_{k\geq 0}$, $(\beta_k)_{k\geq 0}, \mathcal{E}_0, p$ as assumed in Assumption 2.3. Assume in addition that*

(i) *there exists $\bar{x} \in \mathbb{R}^n$ that is an minimizer of $F = f + g$;*
(ii) *$\alpha_0 = 1, p > 0$;*
(iii) *the sequence $(\alpha_k)_{k\geq 0}$ satisfies for all $k \geq 0$ the equality $(1 - \alpha_k) = \alpha_k^2 L_k \alpha_{k-1}^{-2} L_{k-1}^{-1}$;*
(iv) *the sequence $L_k := B_k + \rho_k$ is bounded, and there exists an $L_{\max}$ such that for all $k \geq 0$ it has $L_{\max} \geq \max_{k\geq i\geq 0} L_i$.*

Then, $\alpha_k \in (0, 1]$ hence valid and, it satisfies for all $k \geq 0$:

$$F(x_k) - F(\bar{x}) + \frac{\alpha_k^2 L_k}{2}\|\bar{x} - v_k\|^2 \leq \left(1 + \frac{k\alpha_0\sqrt{L_0}}{2\sqrt{L_{\max}}}\right)^{-2} \left(\frac{L_0}{2}\|\bar{x} - v_{-1}\|^2 + \mathcal{E}_0 \sum_{n=1}^{k} \frac{1}{n^p}\right).$$

Finally, since $p > 0$ the sum is convergent and hence it has an over all convergence rate $\mathcal{O}(1/k^2)$.

Proof. From the assumption that $(\alpha_k)_{k\geq 0}$ has $(1 - \alpha_k) = \alpha_k^2 L_k \alpha_0^{-2} L_0^{-1}$ for all $k \geq 0$ and, the definition of β_k, it yields the following equalities:

$$\beta_k = \prod_{i=1}^{k} \max\left(1 - \alpha_i, \frac{\alpha_i^2 L}{\alpha_{i-1}^2 L_{i-1}}\right) = \prod_{i=1}^{k}(1 - \alpha_i) = \prod_{i=1}^{k} \frac{\alpha_i^2 L}{\alpha_0^2 L_0} = \frac{\alpha_k^2 L_k}{\alpha_0^2 L_0}.$$

With the above relation, and the definitions of the sequences $(\alpha_k)_{k\geq 0}, (\beta_k)_{k\geq 0}$ it satisfies for all $k \geq 1$ the properties:

(i) β_k is nonincreasing and $\beta_k > 0$ for all $k \geq 0$ because $\beta_k = \prod_{i=1}^{k}(1-\alpha_i)$ and, $\alpha_k \in (0, 1]$.
(ii) It has the equalities $\beta_k/\beta_{k-1} = (1 - \alpha_k) = \frac{\alpha_k^2 L_k}{\alpha_{k-1}^2 L_{k-1}}$ for all $k \geq 1$.

Using the above observations, we can show the chain of equalities $\alpha_k^2 = (1 - \beta_k/\beta_{k-1})^2 = \beta_k L_0 \alpha_0^2 L_k^{-1}$ for all $k \geq 0$. This is true by first considering the relations $\prod_{i=1}^{k}(1 - \alpha_i) = \beta_k$:

{eqn:opt-cnvg-outr-loop-pitem1}

$$\begin{aligned} & (1 - \alpha_k) = \beta_k/\beta_{k-1} \\ \iff & \alpha_k = 1 - \beta_k/\beta_{k-1} \\ \implies & \alpha_k^2 = (1 - \beta_k/\beta_{k-1})^2. \end{aligned} \tag{2.2}$$

Next, the recursive relation of $(\alpha_k)_{k\geq 0}$ gives

{eqn:opt-cnvg-outr-loop-pitem2}

$$\begin{aligned}\alpha_k^2 &= (1-\alpha_k)\alpha_{k-1}^2 L_{k-1} L_k^{-1} \\ &= (1-\alpha_k)\frac{\alpha_{k-1}^2 L_{k-1}}{\alpha_0^2 L_0}\frac{L_{k-1}\alpha_0^2 L_0}{L_k} \\ &= (\beta_k \beta_{k-1}^{-1})\beta_{k-1} L_0 \alpha_0^2 L_k^{-1} \\ &= \beta_k L_0 \alpha_0^2 L_k^{-1}.\end{aligned} \tag{2.3}$$

Combining (2.2), (2.3) it would mean for all $i \geq 1$ it has:

$$\begin{aligned}L_0\alpha_0^2 L_i^{-1} &= \beta_i^{-1}\left(1-\frac{\beta_k}{\beta_{k-1}}\right)^2 \\ &= \beta_i\left(\beta_i^{-1}-\beta_{i-1}^{-1}\right)^2 \\ &= \beta_i\left(\beta_i^{-1/2}-\beta_{i-1}^{-1/2}\right)^2\left(\beta_i^{-1/2}+\beta_{i-1}^{-1/2}\right)^2 \\ &\underset{(1)}{\leq} \beta_i\left(\beta_i^{-1/2}-\beta_{i-1}^{-1/2}\right)^2\left(2\beta_i^{-1/2}\right)^2 \\ &= 4\left(\beta_i^{-1/2}-\beta_{i-1}^{-1/2}\right)^2 \\ \implies \sqrt{\frac{L_0\alpha_0^2}{L_i}} &\leq 2\left(\beta_i^{-1/2}-\beta_{i-1}^{-1/2}\right).\end{aligned}$$

Telescope for all $i \geq 1$:

$$\frac{\alpha_0\sqrt{L_0}}{2}\sum_{i=1}^{k}\sqrt{L_k^{-1}} \leq \beta_k^{-1/2}-\beta_0^{-1/2} = \beta_k^{-1/2}-1.$$

Re-arranging, yields:

$$\beta_k \leq \left(1+\frac{\alpha_0\sqrt{L_0}}{2}\sum_{i=1}^{k}\sqrt{L_i^{-1}}\right)^{-2} \leq \left(1+\frac{k\alpha_0\sqrt{L_0}}{2\sqrt{L_{\max}}}\right)^{-2}.$$

Recall that $L_k = B_k + \rho_k$, and by our assumptions, the sequence $(B_k)_{k\geq 0}, (\rho_k)_{k\geq 0}$ are both bounded above, then it's not hard to see that $\beta_k \leq \mathcal{O}(1/k^2)$. Combine it with Lemma 2.2, using $\alpha_0 = 1$ will yield the desired result. Finally, one can solves for α_k by the recursive equality, and it should yield:

■

3 Linear convergence for the proximal problem

In this section, we continue the discussion from [REF PREVIOUS SECTION]. The inner loop of the algorithm evaluates $x_k \approx_\epsilon T_{(B+\rho)}(y_k)$ for a given value of ϵ. To attain x_k, one approach is to utilize Theorem 1.17 numerically, usually using iterative algorithms.

The following assumption places additional assumption to the proximal problem for the inner loop.

{ass:pg-eb} **Assumption 3.1 (gradient mapping error bound)**
The following assumption is about (F, f, g, L, S, γ). Assume that

(i) (f, g, L) satisfies Assumption 1.9,
(ii) let $\tau > 0$ be the step size inverse, let T_τ be the proximal operator of $f + g$ as given by $T_\tau(x) := \operatorname{prox}_{\tau^{-1}g}(x - \tau^{-1}\nabla f(x))$,
(iii) $S = \operatorname*{argmin}_x f(x) + g(x) \neq \emptyset$,
(iv) the objective function is given by $F = f + g$.

Let the gradient mapping $\mathcal{G}_\tau$ be defined as: $\mathcal{G}_\tau(x) := \tau(x - T_\tau(x))$ In addition, assume that the optimization problem F satisfies the error bound condition if it has for all $\tau \geq L, x \in \mathbb{R}^n$ there exists $\gamma > 0$:

{def:ista}
$$\|\mathcal{G}_\tau(x)\| \geq \gamma \operatorname{dist}(x|S).$$

Definition 3.2 (proximal gradient method) *Suppose that (f, g, L) satisfies Assumption 1.9. Let $\tau \geq L$, and $x_0 \in \mathbb{R}^n$. Then an algorithm is a proximal gradient method if it generates iterates $(x_k)_{k\geq 0}$ such that they satisfies for all $k \geq 1$:*

$$x_{k+1} = \operatorname{prox}_{\tau^{-1}g}\left(x_k + \tau^{-1}\nabla f(x_k)\right).$$

{ass:eb-for-pp} **Assumption 3.3 (error bound for proximal problem)**
This assumption is about $(g, \omega, A, \Psi_\lambda, \gamma)$ Here are the assumptions

(i) Let function Ψ_λ as given by (1.2) which satisfies gradient mapping error bound (Assumption 3.1) where, $f(v) = \frac{1}{2\lambda}\|\lambda A^\top v - y\|^2, g(v) = \omega^\star(v) - \frac{1}{2\lambda}\|y\|^2$.
(ii) The primal objective Φ_λ is a L_1 Lipschitz continuous on its domain.
(iii) Let (f, g, L) satisfies Assumption 1.15, we can do this because f is quadratic and has a Lipschitz continuous gradient.

3.1 error bound and linear convergence

The following theorem characterize linear convergence of the proximal gradient method under gradient mapping error bound condition.

{thm:lin-cnvg-ista-eb}

Theorem 3.4 (linear convergence under gradient mapping error bound) *Assume that (F, f, g, L, S, γ) is given by Assumption 3.1. Under this assumption, the iterates $(x_k)_{k\geq 0}$ given by Definition 3.2 satisfies for all $k \geq 0, \bar{x} \in S$ the inequality:*

$$F(x_{k+1}) - F(\bar{x}) \leq \left(1 - \frac{\gamma}{2\tau}\right)(F(x_k) - F(\bar{x})).$$

Hence, the algorithm generates $F(x_k) - F(\bar{x}) \leq \mathcal{O}((1 - \gamma/(2\tau))^k)$.

Proof. Two important immediate results will be presented first. Consider the proximal gradient inequality from 1.13, but with $\rho = 0, \epsilon = 0, B = \tau$, then for all x such that $\|\mathcal{G}_\tau(x)\| > 0$ it has for $\tilde{x} = T_\tau(x), z \in \mathbb{R}^n$ the inequality

$$\begin{aligned} F(\tilde{x}) - F(z) &\leq \frac{\tau}{2}\|x - z\|^2 - \frac{\tau}{2}\|z - \tilde{x}\|^2 \\ &= -\frac{\tau}{2}\|x - \tilde{x}\|^2 + \tau\langle x - z, x - \tilde{x}\rangle \\ &= -\frac{1}{2\tau}\|\mathcal{G}_\tau(x)\|^2 + \langle x - z, \mathcal{G}_\tau(x)\rangle \\ &\leq -\frac{1}{2\tau}\|\mathcal{G}_\tau(x)\|^2 + \|x - z\|\|\mathcal{G}_\tau(x)\| \\ &= \|\mathcal{G}_\tau(x)\|^2 \left(\frac{\|x - z\|}{\|\mathcal{G}_\tau(x)\|} - \frac{1}{2\tau}\right). \end{aligned}$$

Now, for all $z = \bar{x} \in S$, from Assumption 3.3 $\exists\gamma > 0$ such that:

$$\frac{\|x - z\|}{\|\mathcal{G}_\tau(x)\|} \leq \frac{\|x - z\|}{\gamma \operatorname{dist}(x|S)} \leq \frac{1}{\gamma}.$$

Hence for all $\bar{x} \in S$ it has

{ineq:lin-cnvg-ista-eb-pitem1}

$$0 \leq F(\tilde{x}) - F(\bar{x}) \leq \|\mathcal{G}_\tau(x)\|^2 \left(\frac{1}{\gamma} - \frac{1}{2\tau}\right). \tag{3.1}$$

Obviously it has $\gamma^{-1} - (1/2)\tau^{-1} > 0$. When $z = x$, we have the inequality:

{ineq:lin-cnvg-ista-eb-pitem2}

$$F(\tilde{x}) - F(x) \leq -\frac{1}{2\tau}\|\mathcal{G}_\tau(x)\|^2. \tag{3.2}$$

To derive the linear convergence, we use (3.1) with $x = x_k, \tilde{x} = x_{k+1}$:

$$\begin{aligned} 0 &\leq \|\mathcal{G}_\tau(x_k)\|^2 \left(\frac{1}{\gamma} - \frac{1}{2\tau}\right) - F(x_{k+1}) + F(\bar{x}) \\ &= \frac{1}{2\tau}\|\mathcal{G}_\tau(x_k)\|^2 \left(\frac{2\tau}{\gamma} - 1\right) - F(x_{k+1}) + F(\bar{x}) \end{aligned}$$

$$
\begin{aligned}
&\underset{(1)}{\leq} \left(\frac{2\tau}{\gamma} - 1\right)(F(x_k) - F(x_{k+1})) - F(x_{k+1}) + F(\bar{x}) \\
&= \left(\frac{2\tau}{\gamma} - 1\right)(F(x_k) - F(\bar{x}) + F(\bar{x}) - F(x_{k+1})) - F(x_{k+1}) + F(\bar{x}) \\
&= \frac{2\tau}{\gamma}(F(\bar{x}) - F(x_{k+1})) + \left(\frac{2\tau}{\gamma} - 1\right)(F(x_k) - F(\bar{x})).
\end{aligned}
$$

At (1) we used (3.2). Multiple bothside by $\frac{\gamma}{2\tau}$ then we are done. ■

3.2 characterizing linear convergence of the proximal problem

In this section, we will focus on the sufficient characterization of the proximal problem which allows proximal gradient method to achieve linear convergence rate. We can immediately claim quadratic growth condition when $\omega = \|\cdot\|_1$. The next proposition will characterize a precise case where Assumption 3.3 is true, and it's a case widely available in applications.

Proposition 3.5 (1-norm problem) *Let (g, ω, A) satisfies Assumption 1.15. In addition, if $g := \|\cdot\|_1$, then the function Ψ_λ satisfies Assumption 3.3.*

Proof. Take note that the dual has closed form $g^\star(z) = \delta(z|\{x : \|x\|_1 \leq 1\})$. The $g^\star$ is an indicator function which represents a box constraint. The objective function $\Psi_\lambda(v) = \frac{1}{2\lambda}\|\lambda A^\top v - y\|^2 + \omega^\star(v) - \frac{1}{2\lambda}\|y\|^2$ so it's Lipschitz smooth, and it must have a set of minimizers S because its domain is compact. In addition, Ψ_λ fits the assumption of Necoara et al. [1, Theorem 8]. Therefore, Ψ_λ is quasi-strongly convex then by [1, Theorem 4], and [1, Theorem 7], it satisfies error bound condition as given in Assumption 3.1, hence Assumption 3.3 also.

Let $\mu_f = \lambda\|A\|^2/\kappa_f$, let $\kappa_f = \theta^{-2}(A, C)$ be the Hoffman Constant as presented by Necoara et al. in [1, Section 4] where C is the constraint matrix of the inequality set $\delta(z|\{x : \|x\|_1 \leq 1\})$, then the error bound constant can be satisfied with

$$
\gamma = \frac{\kappa_f}{1 + \mu_f + \sqrt{1 + \mu_f}}.
$$

■ The following propositions precisely show that the linear convergence is achievable for the inner loop when $g = \|\cdot\|_1$ for our optimization objective.

{prop:inn-loop-lin-cnvg}

Proposition 3.6 (inner loop linear convergence scenario 1)

3.3 inner loop complexity

4 Total complexity of the algorithm

This section puts results regarding the total complexity of the proposed inexact proximal gradient algorithm.

References

[1] 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.

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

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

[4] 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.