Inexect Accelerated Proximal Gradient

Author 1 Name, Author 2 Name *

September 22, 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 (epsilon subgradient) *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc. Let $\epsilon \ge 0$. Then the ϵ-subgradient of g at some $\bar{x} \in \operatorname{dom} g$ is given by:*

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

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 (epsilon 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].

We will now define inexact proximal point based on epsilon 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 $\tilde{x} \approx_\epsilon \operatorname{prox}_{\lambda g}(x)$.

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

{fact:resv-identity}

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 $\epsilon \ge 0, B \ge 0$. Then, $\tilde{x} \approx_\epsilon T_B(x)$ is an inexact proximal gradient if it satisfies variational inequality:*

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

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. Let $\epsilon \ge 0, B \ge 0$, then for all $x \approx_\epsilon T_B(x)$, it has the following equivalent representations:

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

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 . Let $\epsilon \ge 0$. Consider $\tilde{x} \approx_\epsilon T_{B+\beta}(x)$. If in addition, it satisfies $D_f(\tilde{x}, x) \le B/2\|x - \tilde{x}\|^2$, then it has $\forall z \in \mathbb{R}^n$:

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

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

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

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 $\epsilon \geq 0$, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex, proper and closed then*

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

{thm:dlty-gap-inxt-pp}

The theorem is from Villa et al. [1, Proposition 2.3]

Theorem 1.16 (duality gap of inexact proximal problem) *For all $\epsilon \geq 0$, $v \in \mathbb{R}^n$. Consider conditions*

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

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

Next, let's explore some options for minimizing the duality gap of the proximal problem.

STILL WRITING AND NOT FINISHED YET!

1.4 Literature reviews

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.