First Order Nonconvex and Nonsmooth Optimization Algorithms: Regularity Conditions, Convergence Rates and Applications

Ziyuan Wang

University of British Columbia, Okanagan Campus

Proposal Defense

July 19, 2022

Overview

This talk is based on selected results from published papers

- X. Wang and Z. Wang, *The exact modulus of the generalized concave Kurdyka-Łojasiewicz property*, Mathematics of Operations Research, https://doi.org/10.1287/moor.2021.1227 (2022).
- X. Wang and Z. Wang, *Malitsky-Tam forward-reflected-backward splitting method for nonconvex minimization problems*, Computational Optimization and Applications, 82 (2022), pp. 441–463.

and preprints under review

- X. Wang and Z. Wang, *Calculus rules of the generalized concave Kurdyka-Łojasiewicz property*, preprint, arXiv:2110.03795, (2021).
- X. Wang and Z. Wang, *A Bregman inertial forward-reflected-backward method for nonconvex minimization*, preprint, arXiv:2207.01170, (2022).

Overview

Part I: Introduction

- Notation and preliminaries
- Background
- Our goal

Part II: Published and submitted work

- Malitsky-Tam forward-reflected-backward splitting method for nonconvex minimization
- A Bregman inertial forward-reflected-backward method for nonconvex minimization
- A comparison to the BDLM desingularizing function
- A sum rule of the generalized concave KL property

Part III: Future work

- An open question of Malitsky and Tam
- An open question of Yu, Li, and Pong

Part I: Introduction

Notation and preliminaries

Throughout this talk,

$\mathbb{R}^n$ is the standard Euclidean space

equipped with dot product $\langle x, y\rangle = x^T y$ and the Euclidean norm $\|x\| = \sqrt{\langle x, x\rangle}$.

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}} = (-\infty, \infty]$ be a proper function. We say that

(i) $v \in \mathbb{R}^n$ is a Fréchet subgradient of f at $\bar{x} \in \operatorname{dom} f$, denoted by $v \in \hat{\partial} f(\bar{x})$, if

$$\liminf_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x}\rangle}{\|x - \bar{x}\|} \geq 0.$$

(ii) $v \in \mathbb{R}^n$ is a limiting subgradient of f at $\bar{x} \in \operatorname{dom} f$, denoted by $v \in \partial f(\bar{x})$, if there exist $x_k \xrightarrow{f} \bar{x}$, $v_k \in \hat{\partial} f(x_k)$ such that $v_k \to v$, where $x_k \xrightarrow{f} \bar{x} \Leftrightarrow x_k \to \bar{x}$ with $f(x_k) \to f(\bar{x})$.

Notation and preliminaries

For $\eta \in (0, \infty]$, let Φ_η be the class of functions $\varphi : [0, \eta) \to \mathbb{R}_+$ satisfying: (i) φ is continuous on $[0, \eta)$ with $\varphi(0) = 0$; (ii) φ is strictly increasing.

Definition 1

f has the generalized concave Kurdyka-Łojasiewicz (KL) property at $\bar{x} \in \operatorname{dom} \partial f$, if there exist neighborhood $U \ni \bar{x}$, $\eta \in (0, \infty]$ and concave $\varphi \in \Phi_\eta$, such that

$$\varphi'_-\big(f(x) - f(\bar{x})\big) \cdot \operatorname{dist}\big(0, \partial f(x)\big) \geq 1, \tag{1}$$

for $x \in U \cap \{x : 0 < f(x) - f(\bar{x}) < \eta\}$.

Examples:

- The zero "norm" $x \mapsto \|x\|_0 = \sum_{i=1}^n |\operatorname{sgn}(x_i)|$, where $\operatorname{sgn}(0) = 0$.
- Polynomials $x \mapsto \|Ax - b\|^2$, $(X, Y) \mapsto \|XY - M\|_F^2$.
- Squared distance $x \mapsto \operatorname{dist}^2(x, C)$ to convex set C.

Notation and preliminaries

Definition 2

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \operatorname{dom} \partial f$ and let $U \subseteq \operatorname{dom} \partial f$ be a neighborhood of $\bar{x}$. Let $\eta \in (0, \infty]$. Furthermore, define $h : (0, \eta) \to \mathbb{R}$ by

$$h(s) = \sup \left\{ \operatorname{dist}^{-1} \big(0, \partial f(x)\big) : x \in U \cap [s \le f - f(\bar{x}) < \eta] \right\}$$

Suppose that $h(s) < \infty$ for $s \in (0, \eta)$. The exact modulus of the generalized concave KL property of f at $\bar{x}$ with respect to U and η is the function

$$(\forall t \in (0, \eta)) \ \tilde{\varphi} : [0, \eta) \to \mathbb{R}_+ : t \mapsto \int_0^t h(s) ds,$$

with $\tilde{\varphi}(0) = 0$.

Background

Consider the inclusion problem of maximally monotone operators:

$$\text{find } x \in \mathbb{R}^n \text{ such that } 0 \in (A+B)(x), \tag{2}$$

where $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ and $B : \mathbb{R}^n \to \mathbb{R}^n$ are (maximally) monotone operators with B being Lipschitz continuous or cocoercive.

Convergence of splitting methods for solving (2) is well-understood.

- Lions-Mercier'79: Douglas-Rachford method
- Lions-Mercier'79, Passty'79: Forward-backward method
- Tseng'00: Tseng's method
- Malitsky-Tam'20: Forward-reflected-backward (FRB) method

Our goal

Consider

$$\min_{x\in\mathbb{R}^n} f(x) + g(x),$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper lsc and $g : \mathbb{R}^n \to \mathbb{R}$ has Lipschitz gradient, which corresponds to (2) with $A = \partial f$ and $B = \nabla g$ if f and g are convex. Examples:

- $\min_{x\in\mathbb{R}^n} (1/2) \|Ax - b\|^2 + \lambda \|x\|_0$.
- $\min_{x\in\mathbb{R}^n} (1/2) \operatorname{dist}^2(x, C) + \delta_D(x)$, where $C, D \subseteq$ are nonempty closed sets with C being convex.

However, in the absence of convexity, the usual convergence analysis of splitting methods for (2) collapses. This is where various regularity conditions come into play.

Our goal

Considering problem

$$\min_{x\in\mathbb{R}^n} f(x)+g(x), \tag{3}$$

where $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ is proper lsc and $g:\mathbb{R}^n\to\mathbb{R}$ has Lipschitz gradient, our goals are

to understand the generalized concave KL property,

which is the regularity condition of interest, and

to study the convergence behavior of FRB

in the full nonconvex setting of problem (3) by using the generalized KL.

Part II: Published and submitted work

FRB

In 2020, Malitsky and Tam proposed a *forward-reflected-backward* (FRB) splitting method for solving the inclusion problem of monotone operators (2).

Theorem 3 (Malitsky-Tam'20)

Let $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *be maximally monotone, let* $B : \mathbb{R}^n \to \mathbb{R}^n$ *be monotone and Lipschitz continuous with modulus* $L > 0$*, and suppose that* $(A + B)^{-1}(0) \neq \varnothing$*. Choose* $\lambda \in (0, 1/(2L))$*. Given* $x_{-1}, x_0 \in \mathbb{R}^n$*, define the sequence* $(x_k)_{k\in\mathbb{N}^*}$ *by*

$$x_{k+1} = J_{\lambda A}(x_k - 2\lambda B(x_k) + \lambda B(x_{k-1})), \ \forall k \in \mathbb{N}. \tag{4}$$

Then $(x_k)_{k\in\mathbb{N}^*}$ *converges to a point in* $(A + B)^{-1}(0)$*.*

Published work: nonconvex FRB

Now we present main results from

- X. Wang and Z. Wang, *Malitsky-Tam forward-reflected-backward splitting method for nonconvex minimization problems*, Computational Optimization and Applications, 82 (2022), pp. 441–463.

Our goal is to

establish convergence of FRB in the full nonconvex setting

of minimization problem (3).

Nonconvex FRB

Recall that we are interested in the following minimization problem (3):

$$\min_{x \in \mathbb{R}^n} F(x) = f(x) + g(x),$$

where

- $f : \mathbb{R}^n \to \overline{\mathbb{R}} = (-\infty, \infty]$ is proper lsc and prox-bounded with threshold $\lambda_f > 0$.
- $g : \mathbb{R}^n \to \mathbb{R}$ has a Lipschitz continuous gradient with modulus $L > 0$.
- $\inf(f + g) > -\infty$.

Nonconvex FRB

We formulate the FRB scheme for solving (3) as below:

Algorithm 4 (Forward-reflected-backward splitting method)

1. Initialization: Pick $x_{-1}, x_0 \in \mathbb{R}^n$ *and real number* $\lambda > 0$.
2. For $k \in \mathbb{N}$*, compute*

$$y_k = x_k + \lambda\big(\nabla g(x_{k-1}) - \nabla g(x_k)\big), \tag{5}$$

$$x_{k+1} \in \operatorname{prox}_{\lambda f}\big(y_k - \lambda \nabla g(x_k)\big). \tag{6}$$

For $0 < \lambda < \lambda_f$, the set $\operatorname{prox}_{\lambda f}(x)$ is nonempty for $x \in \mathbb{R}^n$, in which case the nonconvex FRB scheme is well-defined.

FRB merit function

Definition 5 (FRB merit function)

Let $\lambda > 0$. Define the FRB merit function $H : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ by

$$H(x, y) = f(x) + g(x) + \left(\frac{1}{4\lambda} - \frac{L}{4}\right) \|x - y\|^2 . \tag{7}$$

It turns out that the FRB merit function has the decreasing property under a suitable step-size assumption.

In the remainder, let $(x_k)_{k \in \mathbb{N}^*}$ be a sequence generated by FRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$.

FRB: decreasing property

Lemma 6 (decreasing property)

Assume that $0 < \lambda < \min\left\{1/(3L), \lambda_f\right\}$ *and* $\inf(f+g) > -\infty$. *Le* $M_1 = 1/(4\lambda) - (3L)/4$. *Then the following hold:*
(i) For $k \in \mathbb{N}$, *we have*

$$M_1 \left\|z_k - z_{k-1}\right\|^2 \leq H(z_{k-1}) - H(z_k), \tag{8}$$

which means that $H(z_k) \leq H(z_{k-1})$. *Hence, the sequence* $\big(H(z_k)\big)_{k\in\mathbb{N}^*}$ *is convergent.*
(ii) $\sum_{k=0}^{\infty} \left\|z_k - z_{k-1}\right\|^2 < \infty$, *consequently* $\lim_{k\to\infty} \left\|z_k - z_{k-1}\right\| = 0$.

FRB: function value convergence

Theorem 7 (function value convergence)

Assume that conditions in Lemma 6 are satisfied and $(z_k)_{k\in\mathbb{N}^*}$ *is bounded. Suppose that a subsequence* $(z_{k_l})_{l\in\mathbb{N}}$ *of* $(z_k)_{k\in\mathbb{N}^*}$ *converges to some* $z^* = (x^*, y^*)$ *as* $l \to \infty$. *Then the following hold:*

(i) $\lim_{l\to\infty} H(z_{k_l}) = f(x^*) + g(x^*) = F(x^*)$. *In fact, one has* $\lim_{k\to\infty} H(z_k) = F(x^*)$.

(ii) We have $x^* = y^*$ *and* $0 \in \partial H(x^*, y^*)$, *which implies* $0 \in \partial F(x^*) = \partial f(x^*) + \nabla g(x^*)$.

(iii) The set $\omega(z_{-1})$ *is nonempty, compact and connected, on which the FRB merit function* H *is finite and constant. Moreover, we have* $\lim_{k\to\infty} \operatorname{dist}(z_k, \omega(z_{-1})) = 0$.

FRB: sequential convergence

Theorem 8 (Global convergence of FRB)

Assume that $(z_k)_{k\in\mathbb{N}^*}$ *is bounded,* $\inf(f+g) > -\infty$*, and* $0 < \lambda < \min\left\{1/(3L), \lambda_f\right\}$*. Suppose that the FRB merit function* H *has the generalized concave KL property on* $\omega(z_{-1})$*. Then the sequence* $(x_k)_{k\in\mathbb{N}^*}$ *has finite length and converges to some* x^* *with* $0 \in \partial F(x^*)$*. To be specific, there exist* $M > 0$*,* $k_0 \in \mathbb{N}$*,* $\varepsilon > 0$ *and* $\eta \in (0, \infty]$ *such that for* $i \geq k_0 + 1$

$$\sum_{k=i}^{\infty} \|x_{k+1} - x_k\| \leq \|z_i - z_{i-1}\| + M\tilde{\varphi}\left(H(z_i) - H(z^*)\right), \tag{9}$$

where $\tilde{\varphi} \in \Phi_\eta$ *is the exact modulus associated with the setwise generalized concave KL property of* H *on* $\omega(z_{-1})$ *with respect to* ε *and* η*.*

Submitted work: BiFRB

Now we turn to selected results from

- X. Wang and Z. Wang, *A Bregman inertial forward-reflected-backward method for nonconvex minimization*, preprint, arXiv:2207.01170, (2022).

Our goal is to

generalize and accelerate the nonconvex FRB method for (3).

In turn, we answered a question of Malitsky and Tam regarding whether FRB can be adapted to incorporate a Nesterov-type acceleration.

BiFRB: notation and background

Inertial effect is a powerful acceleration scheme. Examples:

- Nesterov'83 Accelerated gradient descent method for smooth convex problems with convergence rate $O(1/k^2)$.
- Beck-Teboulle'09 Fast iterative shrinkage-thresholding (FISTA) algorithm for linear inverse problems with convergence rate $O(1/k^2)$.

Let $h : \mathbb{R}^n \to \mathbb{R}$ be differentiable. The Bregman distance induced by kernel h is $D_h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$(x, y) \mapsto h(x) - h(y) - \langle x - y, \nabla h(y)\rangle.$$

For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and parameter $\omega \in \mathbb{R}^n$, define

$$T_\lambda(\cdot;\omega) = \mathrm{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) + \langle x - \cdot, \omega\rangle + \frac{1}{\lambda} D_h(x, \cdot) \right\}.$$

BiFRB and its variants

Algorithm 9 (BiFRB)

1. Initialization: Pick $x_{-1}, x_0 \in \mathbb{R}^n$*. Let* $0 < \underline{\lambda} \leq \overline{\lambda} \leq \lambda_f$*.*
Choose $\underline{\lambda} \leq \lambda_{-1} \leq \overline{\lambda}$*.*
2. For $k \in \mathbb{N}$*, choose* $\underline{\lambda} \leq \lambda_k \leq \overline{\lambda}$ *and* $0 \leq \alpha_k < 1$*. Compute*

$$y_k = x_k + \lambda_{k-1}\big(\nabla g(x_{k-1}) - \nabla g(x_k)\big), \tag{10}$$

$$x_{k+1} \in T_{\lambda_k}\left(y_k; \frac{\alpha_k}{\lambda_k}(x_{k-1} - x_k) + \nabla g(x_k)\right). \tag{11}$$

BiFRB and its variants

Setting the kernel $h = \|\cdot\|^2/2$ gives Algorithm 10, which is an inertial forward-reflected-backward method (iFRB).

Algorithm 10 (iFRB)

1. Initialization: Pick $x_{-1}, x_0 \in \mathbb{R}^n$*. Let* $0 < \underline{\lambda} \leq \overline{\lambda}$*. Choose* $\underline{\lambda} \leq \lambda_{-1} \leq \overline{\lambda}$*.*
2. For $k \in \mathbb{N}$*, choose* $\underline{\lambda} \leq \lambda_k \leq \overline{\lambda}$ *and* $0 \leq \alpha_k < 1$*. Compute*

$$y_k = x_k + \lambda_{k-1}\big(\nabla g(x_{k-1}) - \nabla g(x_k)\big),$$

$$x_{k+1} \in \operatorname{prox}_{\lambda_k f}\left(y_k - \lambda_k \nabla g(x_k) + \alpha_k(x_k - x_{k-1})\right).$$

Choosing $\alpha_k = 0$ and constant stepsize λ_k in Algorithm 10, one recovers the nonconvex FRB (Algorithm 4).

BiFRB merit function

In the remainder, let (α_k) be a sequence of inertial parameters, (λ_k) a sequence of stepsizes, (x_k) a sequence generated by BiFRB, and $z_k = (x_{k+1}, x_k)$ for all k.

Let $p_{-1} \in \mathbb{R}$. Define recursively

$$p_k = \left(\frac{(L_{\nabla h} - \sigma) L_{\nabla g}^2}{2} \right) \frac{\lambda_{k-1}^2}{\lambda_k} + \frac{1}{2} \left(\frac{\sigma}{\lambda_k} - L_{\nabla g} \right) - p_{k-1}, \quad (12)$$

$$M_{1,k} = p_{k-1} - \frac{\alpha_k + \sigma L_{\nabla g} \lambda_{k-1}}{2\lambda_k} - \frac{(L_{\nabla h} - \sigma) L_{\nabla g}^2 \lambda_{k-1}^2}{2\lambda_k}. \quad (13)$$

We say that $H_{p_k}(x, y) = F(x) + p_k \|x - y\|^2$ is the kth BiFRB merit function.

Parameter rules

Assumption 11 (Parameter rules for BiFRB merit function)

(i) $\liminf_{k\to\infty} p_k \geq 0$ *and* $\liminf_{k\to\infty} M_{1,k} > 0$.
(ii) $p_k \leq \bar{p}$ *for all* k.

Assumption 11 generalizes merit function properties in the literature[1]. Examples:

- Inertial forward-backward (Ochs-Chen-Brox-Pock'14) (p_k) decreasing and positive. $M_{1,k} \geq c > 0$ for all k.
- Bregman inertial Tseng's (Boţ-Csetnek'16) (p_k) and $(M_{1,k})$ are constant and positive.
- Forward-reflected-backward with fixed stepsize (Wang-Wang'21) (p_k) and $(M_{1,k})$ are constant and positive.

[1] Parameters for these methods may have different forms.

BiFRB: abstract function value convergence

Theorem 12 (function value convergence)

Suppose that Assumption 11 holds and $(z_k)_{k\in\mathbb{N}^*}$ *is bounded. Let* $(z_{k_l})_{l\in\mathbb{N}}$ *be a subsequence such that* $z_{k_l} \to z^*$ *for some* $z^* = (x^*, y^*)$ *as* $l \to \infty$*. Then the following hold:*

(i) We have $\lim_{l\to\infty} H_{p_{k_l}}(z_{k_l}) = F(x^*)$*. In fact,* $\lim_{k\to\infty} H_{p_k}(z_k) = F(x^*)$*. Consequently* $\lim_{k\to\infty} H_{\bar{p}}(z_k) = F(x^*)$*, where* $\bar{p}$ *is given in Assumption 11(ii).*

(ii) The limit point $z^* = (x^*, y^*)$ *satisfies* $x^* = y^*$ *and* $0 \in \partial H_{\bar{p}}(x^*, y^*)$*, which implies that* $0 \in \partial F(x^*)$*.*

(iii) The set $\omega(z_{-1})$ *is nonempty, connected, and compact, on which the function* $H_{\bar{p}}$ *is constant* $F(x^*)$*. Moreover,* $\lim_{k\to\infty} \operatorname{dist}(z_k, \omega(z_{-1})) = 0$*.*

BiFRB: abstract sequential convergence

Theorem 13 (global sequential convergence)

Suppose that Assumption 11 holds and $(x_k)_{k\in\mathbb{N}^*}$ *is bounded. Suppose further that the dominating BiFRB merit function* $H_{\bar{p}}$ *has the generalized concave KL property on* $\omega(z_{-1})$*. Then the sequence* $x_k \to x^*$ *for some* x^* *with* $0 \in \partial F(x^*)$*, and has finite length property. To be specific, there exist index* $k_0 \in \mathbb{N}$*,* $\varepsilon > 0$ *and* $\eta \in (0, \infty]$ *such that for* $i \geq k_0 + 1$*,*

$$\sum_{k=i}^{\infty} \|x_{k+1} - x_k\| \leq \|z_i - z_{i-1}\| + C\tilde{\varphi}\left(H_{p_i}(z_i) - F(x^*)\right), \quad (14)$$

where

$$C = \frac{2\sqrt{2}\max\left\{\frac{L_{\nabla h}}{\underline{\lambda}} + L_{\nabla g} + 6\bar{p}, \frac{L_{\nabla h}L_{\nabla g}\overline{\lambda}+1}{\underline{\lambda}}\right\}}{\liminf_{k\to\infty} M_{1,k}}.$$

BiFRB: stepsize rules

In the remainder, let

$$a = (L_{\nabla h} - \sigma)L^2_{\nabla g}, b = \sigma, c = L_{\nabla g}. \tag{15}$$

Proposition 14 (fixed stepsize)

Let $(\lambda_k)_{k\in\mathbb{N}^*}$ be a constant sequence with $(\forall k \in \mathbb{N}^*)$ $\lambda_k = \lambda > 0$. Suppose that $\sigma > 2$, $(L_{\nabla h} - \sigma)\sigma > 1/4$ and $(\forall k \in \mathbb{N}^*)$ $0 \le \alpha_k < \min(1, \sigma/2) = 1$. Define

$$\lambda^* = \frac{\sqrt{(2bc + c)^2 + 4a(b - 2)} - 2bc - c}{2a} > 0.$$

If $\lambda < \min\left\{\lambda^*, (\sigma - 1)/[(\sigma + 1)L_{\nabla g}]\right\}$, then there exists p_{-1} such that $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}}$ generated by (12) and (13) satisfy Assumption 11.

BiFRB: an open question of Malitsky and Tam

In 2020, Malitsky and Tam posted a question regarding whether FRB can be adapted to incorporate a Nesterov-type acceleration.

With inertial parameter $\alpha_k \in [0, 1)$ independent of stepsize λ in Proposition 14, this question is resolved. Indeed, the Nesterov acceleration scheme corresponds to Proposition 14 with

$$(\forall k \in \mathbb{N}^*) \ \alpha_k = \frac{t_k - 1}{t_{k+1}},$$

where $(\forall k \in \mathbb{N}^*) \ t_{k+1} = \left(1 + \sqrt{1 + 4t_k^2}\right)/2$ and $t_{-1} = 1$.

BiFRB: subproblem formulae

Set $\alpha \geq 0, \beta > 0$. Fix $\omega \in \mathbb{R}^n$ and $\lambda > 0$. In the remainder of this section, let

$$h(x) = \alpha\sqrt{1 + \|x\|^2} + \frac{\beta}{2}\|x\|^2.$$

We now present formuale of the BiFRB subproblem. Define $p_\lambda(u) = \lambda\omega - \nabla h(u)$. Then the subproblem satisfies

$$T_\lambda(u;\omega) = \mathrm{argmin}_{x\in\mathbb{R}^n}\left\{\lambda f(x) + \langle x, p_\lambda(u)\rangle + h(x)\right\}.$$

BiFRB: subproblem formulae

Let $D = \{x \in \mathbb{R}^n : \|x\|_0 \leq r, \|x\| \leq R\}$ for integer an $r \leq n$ and real number $R > 0$. Consider l_0-constrained problems of the form $\min_{x \in D} g(x)$.

Proposition 15 (l_0-constrained problems)

Let $f = \delta_D$ be the indicator function of set D. Define

$$x^* = \begin{cases} 0, & \text{if } \nabla h(u) = \lambda\omega, \\ -t^* \left\| H_r\big(p_\lambda(u)\big) \right\|^{-1} H_r\big(p_\lambda(u)\big), & \text{if } \nabla h(u) \neq \lambda\omega, \end{cases} \tag{16}$$

where t^* satisfies $t^* = R$ if $\|H_r(p_\lambda(u))\| \geq \alpha(1+R^2)^{-1/2} + \beta R$; otherwise $t^* \in (0, R)$ is the unique solution of

$$\alpha(1+t^2)^{-1/2} t + \beta t - \|H_r(p_\lambda(u))\| = 0.$$

Then $x^* \in T_\lambda(u;\omega)$.

BiFRB: subproblem formulae

Consider $\min f + g$, where f is positively homogeneous and convex.

Proposition 16

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc and convex. Suppose that f is positively homogeneous. Then $T_\lambda(u;\omega) = t^* \operatorname{prox}_{\lambda f}(-p_\lambda(u))$, where t^* is the unique root of the equation

$$1 - \alpha t\Big(1 + t^2 \left\|\operatorname{prox}_{\lambda f}(-p_\lambda(u))\right\| \Big)^{-1/2} - \beta t = 0.$$

Example (l_1-penalized problems) Let $f = \|\cdot\|_1$. Then $T_\lambda(u;\omega) = -t^* S_\lambda\left(p_\lambda(u)\right)$, where S_λ denotes the soft-threshold and t^* is the unique solution of

$$1 - \alpha t\left(1 + t^2 \left\|S_\lambda\left(p_\lambda(u)\right)\right\|^2\right)^{-1/2} - \beta t = 0.$$

BiFRB: application to nonconvex feasibility problems

Let $A \in \mathbb{R}^{m\times n}$ $(m < n)$, $b \in \mathbb{R}^m$, $r \in \mathbb{N}\backslash\{0\}$, and $R > 0$. Define

$$C = \{x \in \mathbb{R}^n : Ax = b\} \text{ and } D = \{x \in \mathbb{R}^n : \|x\|_0 \le r, \|x\| \le R\}.$$

Consider

$$\min_{x\in D} \frac{1}{2}\operatorname{dist}^2(x, C). \tag{17}$$

Then a global minimizer x^* of (17) satisfies

$$x^* \in C \cap D.$$

We shall benchmark BiFRB and iFRB against the Douglas-Rachford (DR) method[2] with $R = 1$ and $r = \lceil m/5 \rceil$.

[2]Similarly to BiFRB, DR converges globally to a stationary point of F (Li-Pong'16).

BiFRB: application to nonconvex feasibility problems

Below we present selected experiment outcomes with $m = 100, 200$.

Size	BiFRB		iFRB		DR	
$(m = 100)$	iter[3]	$\text{fval}_{\min}$[4]	iter	$\text{fval}_{\min}$	iter	$\text{fval}_{\min}$
$n = 4000$	50	0.03251	93	0.03251	860	0.02819
$n = 5000$	52	0.02413	115	0.02413	684	0.02192
$n = 6000$	57	0.01369	145	0.01369	683	0.01253
Size	BiFRB		iFRB		DR	
$(m = 200\ )$	iter	$\text{fval}_{\min}$	iter	$\text{fval}_{\min}$	iter	$\text{fval}_{\min}$
$n = 4000$	870	0.2923	39	0.2745	5466	0.2711
$n = 5000$	1198	0.2367	41	0.2095	5864	0.2073
$n = 6000$	81	0.2306	43	0.2259	4199	0.2236

[3] Average number of iterations of 50 randomly generated instances.

[4] Minimum objective function value at termination out of 50 randomly generated instances

BiFRB: application to nonconvex feasibility problems

We can observe the following:

- BiFRB is the most advantageous method on "bad" problems (small m), but its performance is not robust with respect to problem size.
- DR tends to have the smallest function value at termination, however, it can converge very slowly.
- In most cases, iFRB has the smallest number of iterations with fair function value at termination.

Published work: a comparison to the BDLM desingularizing function

Next, we present results from published paper

- X. Wang and Z. Wang, *The exact modulus of the generalized concave Kurdyka-Łojasiewicz property*, Mathematics of Operations Research, https://doi.org/10.1287/moor.2021.1227 (2022).

which is partially completed in my master thesis.

Here we only present a new result concerning the difference between the exact modulus of the generalized concave KL property and a desingularizing function of Bolte, Daniilids, Ley and Mazet.

Published work: a comparison to the BDLM desingularizing function

Denote $\mathcal{K}_\eta \subseteq \Phi_\eta$ the class of functions $\varphi : [0, \eta) \to \mathbb{R}_+$ that satisfy the following three conditions: (i) φ is continuous with $\varphi(0) = 0$; (ii) φ is C^1 on $(0, \eta)$; (iii) $\varphi'(t) > 0$ for all $t \in (0, \eta)$.

Definition 17 (Bolte-Daniilids-Ley-Mazet'10)

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. We say f has the KL property at $\bar{x} \in \operatorname{dom} \partial f$ if there exist a neighborhood $U \ni \bar{x}$, $\eta \in (0, \infty]$ and a function $\varphi \in \mathcal{K}_\eta$ such that for all $x \in U \cap [0 < f - f(\bar{x}) < \eta]$,

$$\varphi'\big(f(x) - f(\bar{x})\big) \cdot \operatorname{dist}\big(0, \partial f(x)\big) \geq 1, \tag{18}$$

Note that here φ is C^1 and not necessarily concave.

Published work: a comparison to the BDLM desingularizing function

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc and semiconvex, in the sense that there exists $\alpha > 0$ such that $f + (\alpha/2)\left\|\cdot\right\|^2$ is convex.

Bolte-Daniilids-Ley-Mazet'10: Let $r_0 > 0$ and suppose that

$$u(r) = 1/\inf_{x \in \mathbb{B}(\bar{x};r_0)\cap[f=r]} \operatorname{dist}(0, \partial f(x))$$

is finite and $L^1(0, r_0)$. Then there exists a continuous majorant $\bar{u}$ of u such that $(\forall t \in (0, r_0))\ \varphi(t) = \int_0^t \bar{u}(s)ds \in \mathcal{K}_{r_0}$, and φ is a desingularizing function of f at $\bar{x}$.

Given the existence of concave desingularizing functions, the Bolte-Daniilids-Ley-Mazet (BDLM) desingularizing function fails to capture the optimal (minimal) one.

Published work: a comparison to the BDLM desingularizing function

Example 18

Let $r_1 = \pi^2/6 - 1$ and $r_{k+1} = r_k - 1/[k^2(k+1)]$ for $k \in \mathbb{N}$. Define for $k \in \mathbb{N}$ and $x > 0$

$$f(x) = \frac{1}{k}\left(x - \frac{1}{k}\right) + r_k, \forall x \in \left(\frac{1}{k+1}, \frac{1}{k}\right].$$

Let $f(-x) = f(x)$ for $x < 0$ and $f(0) = 0$. Then the exact modulus of f at 0 is

$$(\forall k \in \mathbb{N})\ \tilde{\varphi}(t) = k(t - r_{k+1}) + \sum_{i=k+1}^{\infty} i(r_i - r_{i+1}), \forall t \in (r_{k+1}, r_k],$$

with $\tilde{\varphi}(0) = 0$, and every BDLM desingularizing function $\varphi : [0, r_1] \to \mathbb{R}_+$ satisfies $(\forall 0 < t \leq r_1)\ \varphi(t) > \tilde{\varphi}(t)$.

Submitted work: a sum rule of the generalized concave KL property

Finally, we turn to preprint under review

- X. Wang and Z. Wang, *Calculus rules of the generalized concave Kurdyka-Łojasiewicz property*, preprint, arXiv:2110.03795, (2021),

which is partially completed in my master thesis. Here we only present new result.

Submitted work: a sum rule of the generalized concave KL property

Theorem 19

Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper continuous function and let $f = \sum_{i=1}^m f_i$. Suppose that at most one of f_i is not locally Lipschitz. Pick $\bar{x} \in \cap_{i=1}^m \operatorname{int} \operatorname{dom} \partial f_i$. Assume that f_i has strictly concave desingularizing function φ_i at $\bar{x}$. Suppose that there exist $\varepsilon_0 > 0$ and $\alpha > 0$ such that for every $x_i \in \mathbb{B}(\bar{x}; \varepsilon_0)$

$$(\forall u_i \in \partial f_i(x_i)) \quad \left\| \sum_{i=1}^m u_i \right\| \geq \alpha \sum_{i=1}^m \|u_i\| . \tag{19}$$

Define $(0 < t < \eta)$ $\varphi(t) = \frac{1}{\alpha} \int_0^t \max_{1 \leq i \leq m} (\varphi_i)'_- \left(\frac{s}{m}\right) ds$ and $\varphi(0) = 0$. Then the sum f has desingularizing function φ at $\bar{x}$.

Part III: Future research

An open question of Malitsky and Tam

Consider finding $x \in \mathbb{R}^n$ such that $0 \in (A + B)(x)$, where $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is maximally monotone and $B = \sum_{i=1}^n B_i$ with $(\forall i \in \{1, \dots, n\})$ B_i being monotone and Lipschitz continuous.

In 2020, Malitsky and Tam proved almost sure convergence of the following stochastic variant of FRB:

$$\begin{aligned}
&\text{choose } i_k \text{ uniformly at random from } \{1,\dots, \text{n}\}, \\
&x_{k+1} = J_{\lambda A}\big(x_k - \lambda B(x_k) - \lambda(B_{i_k}(x_k) - B_{i_k}(x_{k-1}))\big). \qquad (20)
\end{aligned}$$

Algorithm (20) still requires full evaluation of operator B per iteration. Therefore, Malitsky and Tam posted an open question regarding *whether it is possible to obtain a block coordinate FRB with full stochastic approximation of* B*?*

An open question of Malitsky and Tam

Thus, one direction for future research is

to resolve the Malitsky-Tam question in nonconvex setting.

To be specific, we are interested in a stochastic extension of Algorithm 9 for solving minimization problems of the form

$$\min_{x\in\mathbb{R}^N} f(x) + \sum_{i=1}^{n} g_i(x_i)$$

that does not compute full gradient of $\sum_{i=1}^{n} g_i(x_i)$ per iteration, where $(\forall i \in \{1,\ldots,n\})$ $x_i \in \mathbb{R}^{n_i}, n_i \in \mathbb{N}$, $N = \sum_i n_i$, $f : \mathbb{R}^N \to \overline{\mathbb{R}}$ is proper lsc and $g_i : \mathbb{R}^{n_i} \to \mathbb{R}$ has Lipschitz gradient.

An open question of Yu, Li, and Pong

In 2021, Yu, Li, and Pong posted an open question regarding the behavior of KL exponents under supremum operation: *"It would be of interest to see, under what additional conditions, the supremum operation can preserve the KL exponents. "*.

The KL exponent corresponds to a special case of the generalized concave KL property. Therefore, one direction for our future research is to

study a supremum rule of the generalized concave KL property,

which could shed lights on the Yu-Li-Pong open question.

Thank you for the attention :)