Convergence of PALM on non-convex problems

Part 1 : generated sequence is non-increasing with finite length trajectory

Andersen Ang

Mathématique et recherche opérationnelle
UMONS
Belgium

Email: manshun.ang@umons.ac.be
Homepage: angms.science

August 17, 2018

Reference

Proximal alternating linearized minimization or nonconvex and nonsmooth problems
J Bolte, S Sabach, M Teboulle
Mathematical Programming 146 (1-2), 459-494, 2014

The problem

Class of function

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \Phi(x,y) = f(x) + g(y) + H(x,y),$$

- Optimization variables : x, y
- Constraint sets $\mathcal{X} \subset \mathbb{R}^n, \mathcal{Y} \subset \mathbb{R}^m$ moved into f, g by using indicator functions $\mathcal{I}_{\mathcal{X}}$, $\mathcal{I}_{\mathcal{Y}}$
- f, g are extended value functions : e.g. $f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$
- H is smooth (it is partially Lipschitz)
- No convexity will be assumed on f, g, H

PALM

Problem : minimize $\Phi(x,y) = f(x) + g(y) + H(x,y)$.

Starts wtih $(x^0, y^0) \in \mathrm{dom}\Phi$, PALM generate (x^k, y^k) as

$$\begin{aligned} x_k &\in \mathrm{prox}_{f,c_k}\left(x^k - \frac{1}{c_k}\nabla_x H(x^k, y^k)\right), \\ y_k &\in \mathrm{prox}_{g,d_k}\left(y^k - \frac{1}{d_k}\nabla_x H(x^{k+1}, y^k)\right), \end{aligned}$$

for some $\gamma_{1,2} > 1$, the parameters c_k, d_k are selected as

$$c_k = \gamma_1 L_1(y^k), \quad d_k = \gamma_2 L_2(x^{k+1}).$$

Convergence condition of PALM - everything in one slide

Theorem (Bolte14) For $\Phi(x, y) = f(x) + g(y) + H(x, y)$, sequence produced by PALM converges to a stationary point of Φ if :

Assumption 1

- $f : \mathbb{R}^n \to (-\infty + \infty]$ and $g : \mathbb{R}^m \to (-\infty + \infty]$ are proper and lower semicontinuous
- $H : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a C^1/smooth function

Assumption 2

- $\inf_{\mathbb{R}^n \times \mathbb{R}^m} \Phi > -\infty$, $\inf_{\mathbb{R}^n} f > -\infty$, $\inf_{\mathbb{R}^m} g > -\infty$
- Partial gradient $\nabla_x H(x, y)$ is globally Lipschitz with $L_1(y)$:

$$\|\nabla_x H(x_1, y) - \nabla_x H(x_2, y)\| \leq L_1(y)\|x_1 - x_2\| \forall x_1, x_2 \in \mathbb{R}^n$$

- Partial gradient $\nabla_y H(x, y)$ is globally Lipschitz with $L_2(x)$:

$$\|\nabla_x H(x, y_1) - \nabla_x H(x, y_2)\| \leq L_2(x)\|y_1 - y_2\| \forall y_1, y_2 \in \mathbb{R}^n$$

- Lipschitz modulus $L_1(y^k), L_2(x^k)$ are bounded

$$L_1^{\text{min}} \leq L_1(y^k) \leq L_1^{\text{max}}, \quad L_2^{\text{min}} \leq L_2(x^k) \leq L_2^{\text{max}}, \forall k$$

- ∇H is Lipschitz on bounded subsets of $\mathbb{R}^n \times \mathbb{R}^m$

$$\left\| \Big(\nabla_x H(x_1, y_1) - \nabla_x H(x_2, y_2), \nabla_y H(x_1, y_1) - \nabla_y H(x_2, y_2) \Big) \right\| \leq M \|(x_1 - x_2, y_1 - y_2)\|$$

Assumption 3 Φ satisfies Kurdyka-Łojasiewicz property

Convergence condition of PALM - the focus of part 1

Theorem (Bolte14) For $\Phi(x,y) = f(x) + g(y) + H(x,y)$, sequence produced by PALM converges to a stationary point of Φ if :

Assumption 1

- $f : \mathbb{R}^n \to (-\infty + \infty]$ and $g : \mathbb{R}^m \to (-\infty + \infty]$ are proper and lower semicontinuous
- $H : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a C^1/smooth function

Assumption 2

- $\inf_{\mathbb{R}^n \times \mathbb{R}^m} \Phi > -\infty$, $\inf_{\mathbb{R}^n} f > -\infty$, $\inf_{\mathbb{R}^m} g > -\infty$
- Partial gradient $\nabla_x H(x,y)$ is globally Lipschitz with $L_1(y)$:

$$\|\nabla_x H(x_1, y) - \nabla_x H(x_2, y)\| \leq L_1(y)\|x_1 - x_2\| \forall x_1, x_2 \in \mathbb{R}^n$$

- Partial gradient $\nabla_y H(x,y)$ is globally Lipschitz with $L_2(x)$:

$$\|\nabla_x H(x, y_1) - \nabla_x H(x, y_2)\| \leq L_2(x)\|y_1 - y_2\| \forall y_1, y_2 \in \mathbb{R}^n$$

- Lipschitz modulus $L_1(y^k), L_2(x^k)$ are bounded

$$L_1^{\text{min}} \leq L_1(y^k) \leq L_1^{\text{max}}, \quad L_2^{\text{min}} \leq L_2(x^k) \leq L_2^{\text{max}}, \forall k$$

- ∇H is Lipschitz on bounded subsets of $\mathbb{R}^n \times \mathbb{R}^m$

$$\left\| \Big(\nabla_x H(x_1, y_1) - \nabla_x H(x_2, y_2), \nabla_y H(x_1, y_1) - \nabla_y H(x_2, y_2) \Big) \right\| \leq M \|(x_1 - x_2, y_1 - y_2)\|$$

Assumption 3 Φ satisfies Kurdyka-Łojasiewicz property

Convergence condition of PALM – in other words

For $\Phi(x,y) = f(x) + g(y) + H(x,y)$, sequence generated by PALM converges if

- f, g are proper, lower semicontinuous, lower bounded, extended value
- H is smooth such that
  - All partial gradients are globally Lipschitz with $L_{1,2}$
  - All Lipschitz constants $L_1(y^k), L_2(x^k)$ are bounded

Theorem (Bolte14) Let $\{z^k\}_{k\in\mathbb{N}} = \{x^k, y^k\}_{k\in\mathbb{N}}$ be the sequence generated by PALM which is assumed to be bounded, if the above are true, then
(1) The sequence of function value $\{\Phi(z^k)\}_{k\in\mathbb{N}}$ is non-increasing. More precisely

$$\Phi(z^k) - \Phi(z^{k+1}) \geq \frac{\rho_1}{2}\|z^{k+1} - z^k\|^2, \quad \forall k \in \mathbb{N},$$

where $\rho_1 = \min\left\{(\gamma_1 - 1)L_1^{\min}, (\gamma_2 - 1)L_2^{\min}\right\}$
(2) The trajectory of the sequence $\{z^k\}_{k\in\mathbb{N}}$ has finite length

$$\sum_{k=1}^{\infty} \|z^{k+1} - z^k\| < \infty,$$

which implies the gap between consecutive iterates decrease with k and

$$\lim_{k\to\infty} \|z^{k+1} - z^k\| = 0.$$

Material for the proof

1. From assumption 2 we have
   - Partial gradient $\nabla_x H(x,y)$ is globally Lipschitz with $L_1(y)$:

   $$\|\nabla_x H(x_1,y) - \nabla_x H(x_2,y)\| \leq L_1(y)\|x_1 - x_2\| \forall x_1, x_2 \in \mathbb{R}^n$$

   - Partial gradient $\nabla_y H(x,y)$ is globally Lipschitz with $L_2(x)$:

   $$\|\nabla_x H(x,y_1) - \nabla_x H(x,y_2)\| \leq L_2(x)\|y_1 - y_2\| \forall y_1, y_2 \in \mathbb{R}^n$$

2. Sufficient descrease lemma of proximal-gradient operator : for all u, on $f(x) + g(x)$, where f is L_f smooth but not convex and g is convex but non-smooth function, we have

$$f(u) + g(u) \geq f(u^+) + g(u^+) + \left(t - \frac{L_f}{2}\right)\|u^+ - u\|^2,$$

where $u^+ = \text{prox}_t^g\left(u - t\nabla f(u)\right)$ with $t > L_f$.

Details of sufficient descrease lemma : see this slide. (The slide gives the sufficient descrease lemma for convex g, however the lemma also holds even if g is not convex, which is the case here)

Material for the proof

Sufficient descrease lemma :

$$f(u) + g(u) \geq f(u^+) + g(u^+) + \left(t - \frac{L_f}{2}\right) \|u^+ - u\|^2,$$

Implies

$$f(u) + g(u) \geq f(u^+) + g(u^+) + \frac{t - L_f}{2} \|u^+ - u\|^2,$$

Re-arrange gives

$$f(u^+) + g(u^+) \leq f(u) + g(u) - \frac{t - L_f}{2} \|u^+ - u\|^2.$$

Material for the proof

We have sufficient descrease lemma for proximal gradient operator :

$$f(u^+) + g(u^+) \leq f(u) + g(u) - \frac{t - L_f}{2}\|u^+ - u\|^2.$$

Recall problem : minimize $\Phi(x, y) = f(x) + g(y) + H(x, y)$, here H is smooth and f, g are not smooth and PALM on Φ steps are

$$\begin{aligned} x_k &\in \text{prox}_{f,c_k}\left(x^k - \frac{1}{c_k}\nabla_x H(x^k, y^k)\right), \\ y_k &\in \text{prox}_{g,d_k}\left(y^k - \frac{1}{d_k}\nabla_x H(x^{k+1}, y^k)\right), \end{aligned}$$

Therefore, we can see that, we should apply the sufficient descrease lemma on Φ by putting the smooth function as $H(x)$, $H(y)$ and the non-smooth parts as f, g

The proof - 1. apply sufficient descrease lemma

Apply the sufficient descrease lemma for proximal gradient operator :

$$f(u^+) + g(u^+) \leq f(u) + g(u) - \frac{t - L_f}{2}\|u^+ - u\|^2.$$

On PALM twice :
First time on x : smooth part $H(.,y)$, non-smooth part $f(x)$, step size $t = c_k = \gamma_1 L_1(y)$ with $\gamma_1 > 1$ so that $t \geq L_{H(.,y)} = L_1(y)$

$$H(x^{k+1}, y^k) + f(x^{k+1}) \leq H(x^k, y^k) + f(x^k) - \frac{c_k - L_1(y^k)}{2}\|x^{k+1} - x^k\|_2^2$$

Second time on y : smooth part $H(x,.)$, non-smooth part $g(x)$, step size $t = d_k = \gamma_2 L_2(x)$ with $\gamma_2 > 1$ so that $t \geq L_{H(x,.)} = L_2(x)$

$$H(x^{k+1}, y^{k+1}) + g(y^{k+1}) \leq H(x^{k+1}, y^k) + g(x^k) - \frac{d_k - L_2(x^{k+1})}{2}\|y^{k+1} - y^k\|_2^2$$

The proof - 1. apply sufficient descrease lemma

We now have

$$H(x^{k+1}, y^k) + f(x^{k+1}) \leq H(x^k, y^k) + f(x^k) - \frac{c_k - L_1(y^k)}{2} \|x^{k+1} - x^k\|_2^2$$

$$H(x^{k+1}, y^{k+1}) + g(y^{k+1}) \leq H(x^{k+1}, y^k) + g(x^k) - \frac{d_k - L_2(x^{k+1})}{2} \|y^{k+1} - y^k\|_2^2$$

Put $c_k = \gamma_1 L_1(y^k)$ and $d_k = \gamma_2 L_2(x^{k+1})$ gives

$$H(x^{k+1}, y^k) + f(x^{k+1}) \leq H(x^k, y^k) + f(x^k) - \frac{\gamma_1 - 1}{2} L_1(y^k) \|x^{k+1} - x^k\|_2^2$$

$$H(x^{k+1}, y^{k+1}) + g(y^{k+1}) \leq H(x^{k+1}, y^k) + g(x^k) - \frac{\gamma_2 - 1}{2} L_2(x^{k+1}) \|y^{k+1} - y^k\|_2^2$$

Next step : sum the two equations

The proof - 2. sum the two equations

Sum the two equations gives

$$\begin{aligned} & H(x^{k+1}, y^k) + f(x^{k+1}) + H(x^{k+1}, y^{k+1}) + g(y^{k+1}) \\ \leq \quad & H(x^k, y^k) + f(x^k) - \frac{\gamma_1 - 1}{2} L_1(y^k) \|x^{k+1} - x^k\|_2^2 \\ + \quad & H(x^{k+1}, y^k) + g(x^k) - \frac{\gamma_2 - 1}{2} L_2(x^{k+1}) \|y^{k+1} - y^k\|_2^2 \end{aligned}$$

Recall : $\Phi(x, y) = f(x) + g(y) + H(x, y)$, so

$$\begin{aligned} & H(x^{k+1}, y^k) + f(x^{k+1}) + H(x^{k+1}, y^{k+1}) + g(y^{k+1}) \\ \leq \quad & H(x^k, y^k) + f(x^k) - \frac{\gamma_1 - 1}{2} L_1(y^k) \|x^{k+1} - x^k\|_2^2 \\ + \quad & H(x^{k+1}, y^k) + g(x^k) - \frac{\gamma_2 - 1}{2} L_2(x^{k+1}) \|y^{k+1} - y^k\|_2^2 \end{aligned}$$

Becomes

$$\begin{aligned} & \Phi(x^{k+1}, y^{k+1}) \\ \leq \quad & \Phi(x^k, y^k) - \frac{\gamma_1 - 1}{2} L_1(y^k) \|x^{k+1} - x^k\|_2^2 - \frac{\gamma_2 - 1}{2} L_2(x^{k+1}) \|y^{k+1} - y^k\|_2^2 \end{aligned}$$

Notice that the red parts cancel each other.

The proof - 3. some simple algebra

Now we have

$$\begin{aligned} & \Phi(x^{k+1}, y^{k+1}) \\ \leq \quad & \Phi(x^k, y^k) - \frac{\gamma_1 - 1}{2} L_1(y^k) \|x^{k+1} - x^k\|_2^2 - \frac{\gamma_2 - 1}{2} L_2(x^{k+1}) \|y^{k+1} - y^k\|_2^2 \end{aligned}$$

Rearrange gives

$$\begin{aligned} & \Phi(x^k, y^k) - \Phi(x^{k+1}, y^{k+1}) \\ \geq \quad & \frac{\gamma_1 - 1}{2} L_1(y^k) \|x^{k+1} - x^k\|_2^2 + \frac{\gamma_2 - 1}{2} L_2(x^{k+1}) \|y^{k+1} - y^k\|_2^2 \end{aligned}$$

From assumption 2

$$L_1(y^k) \geq L_1^{\min}, \quad L_2(x^k) \geq L_2^{\min},$$

We have

$$\begin{aligned} & \Phi(x^k, y^k) - \Phi(x^{k+1}, y^{k+1}) \\ \geq \quad & \frac{\gamma_1 - 1}{2} L_1^{\min} \|x^{k+1} - x^k\|_2^2 + \frac{\gamma_2 - 1}{2} L_2^{\min} \|y^{k+1} - y^k\|_2^2 \end{aligned}$$

The proof - 3. some simple algebra

Now we have

$$\begin{aligned} & \Phi(x^k, y^k) - \Phi(x^{k+1}, y^{k+1}) \\ \geq \quad & \frac{\gamma_1 - 1}{2} L_1^{\min} \|x^{k+1} - x^k\|_2^2 + \frac{\gamma_2 - 1}{2} L_2^{\min} \|y^{k+1} - y^k\|_2^2 \end{aligned}$$

Let $\rho_1 = \min\left\{(\gamma_1 - 1)L_1^{\min}, (\gamma_2 - 1)L_2^{\min}\right\}$, then

$$\begin{aligned} \Phi(x^k, y^k) - \Phi(x^{k+1}, y^{k+1}) & \geq \frac{\rho_1}{2}\|x^{k+1} - x^k\|_2^2 + \frac{\rho_1}{2}\|y^{k+1} - y^k\|_2^2 \\ & = \frac{\rho_1}{2}\left(\|x^{k+1} - x^k\|_2^2 + \|y^{k+1} - y^k\|_2^2\right) \end{aligned}$$

Hence for $z^k = \{x_k, y_k\}$,

$$\Phi(z^k) - \Phi(z^{k+1}) \geq \frac{\rho_1}{2}\|z^{k+1} - z^k\|_2^2.$$

The proof of (1) is completed. Next we prove (2).

The proof of finite trajectory length (1/3)

Let N be a positive interger, consider the trajectory from iteration 0 to N :

$$\sum_{k=0}^{N-1} \|x^{k+1} - x^k\|_2^2 + \|y^{k+1} - y^k\|_2^2 = \sum_{k=0}^{N-1} \|z^{k+1} - z^k\|_2^2.$$

By (1) we have

$$\begin{aligned} \sum_{k=0}^{N-1} \|z^{k+1} - z^k\|_2^2 &\leq \sum_{k=0}^{N-1} \frac{2}{\rho_1}\Big(\Phi(z^k) - \Phi(z^{k+1})\Big) \\ &= \frac{2}{\rho_1} \sum_{k=0}^{N-1} \Phi(z^k) - \Phi(z^{k+1}) \end{aligned}$$

The term $\sum_{k=0}^{N-1} \Phi(z^k) - \Phi(z^{k+1})$ form s a telescoping series :

$$\sum_{k=0}^{N-1} \Phi(z^k) - \Phi(z^{k+1}) = \Phi(z^0) \underbrace{- \Phi(z^1) + \Phi(z^1) - \Phi(z^2) + ... + \Phi(z^{N-1})}_{\text{cancel out}} - \Phi(z^N)$$

The proof of finite trajectory length (2/3)

We have

$$\sum_{k=0}^{N-1} \|z^{k+1} - z^k\|_2^2 \quad \le \quad \frac{2}{\rho_1}\Big(\Phi(z^0) - \Phi(z^N)\Big)$$

Take limit $N \to \infty$

$$\sum_{k=0}^{\infty} \|z^{k+1} - z^k\|_2^2 \quad \le \quad \frac{2}{\rho_1}\Big(\Phi(z^0) - \Phi(z^\infty)\Big)$$

As we assume Φ is bounded below (assumption 2), i.e. $\inf \Phi > -\infty$, and by the fact $\Phi(z^k) \ge \inf \Phi$ for all k (even if $k \to \infty$), therefore we have

$$-\Phi(z^\infty) \le -\inf \Phi < -(-\infty) = +\infty$$

Hence

$$\sum_{k=0}^{\infty} \|z^{k+1} - z^k\|_2^2 \quad \le \quad \frac{2}{\rho_1}\Big(\Phi(z^0) - \inf \Phi\Big) < +\infty$$

The proof is completed.

Last page : summary

For $\Phi(x,y) = f(x) + g(y) + H(x,y)$, if

- f, g are proper, lower semicontinuous, lower bounded, extended value
- H is smooth such that
  - All partial gradients are globally Lipschitz with $L_{1,2}$
  - All Lipschitz constants $L_1(y^k), L_2(x^k)$ are bounded

Then the bounded sequence $\{x^k, y^k\}_{k\in\mathbb{N}}$ generated by PALM has the following properties

- The sequence $\{\Phi(z^k)\}_{k\in\mathbb{N}}$ is non-increasing.
- The trajectory of $\{x^k, y^k\}_{k\in\mathbb{N}}$ has finite length,
  and therefore $\{x^k, y^k\}_{k\in\mathbb{N}}$ converges (to some unknown point).

What is not proved : the above only proves $\{x^k, y^k\}_{k\in\mathbb{N}}$ converges to some unknown points, it does not say anything on $\{x^k, y^k\}_{k\in\mathbb{N}}$ converges to a stationary point of Φ. For convergence to a stationary point, see part 2.

End of document