Proximal alternating linearized minimization

Introduction to the algorithm

Andersen Ang

Mathématique et recherche opérationnelle
UMONS
Belgium

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

August 14, 2018

Reference

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

The problem

Class of function

$$\min_{x\in\mathcal{X},y\in\mathcal{Y}} \Phi(x,y) = f(x) + g(y) + H(x,y)$$

- Optimization variables : x, y
- Constraint sets : $\mathcal{X} \subset \mathbb{R}^n, \mathcal{Y} \subset \mathbb{R}^m$
- Using indicator functions $\mathcal{I}_{\mathcal{X}}$, $\mathcal{I}_{\mathcal{Y}}$, constraints can be moved into Φ and we get unconstrained expression of the same problem

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

and f, g become *extended value functions*.

The problem

Class of function

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

- 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 !**

Proximal Alternating Linearized Minimization (PALM)

Problem :

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

PALM iterations

$$\begin{aligned} x^{k+1} &\in \arg\min_x \left\{ \hat{\Phi}(x, y^k) + \frac{c_k}{2}\|x - x^k\|^2 \right\} \\ y^{k+1} &\in \arg\min_x \left\{ \hat{\Phi}(x^{k+1}, y) + \frac{d_k}{2}\|y - y^k\|^2 \right\} \end{aligned}$$

i.e. PALM replaces the original function Φ in by approximation $\hat{\Phi}$

The linear approximation $\hat{\Phi}$ in PALM

Recall setting : H is smooth, f, g not necessarily smooth[1] for

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

Recall first order Taylor approximation

$$H(x) \approx \underbrace{H(x^k)}_{\text{constant}} + \underbrace{\left\langle x - x^k, \nabla_x H(x^k, y^k) \right\rangle}_{\text{important part}}$$

PALM : i.e. approximate H by the linearized H

$$\begin{aligned} \hat{\Phi}(x, y^k) &= f(x) + g(y) + \left\langle x - x^k, \nabla_x H(x^k, y^k) \right\rangle \\ \hat{\Phi}(x^k, y) &= f(x) + g(y) + \left\langle y - y^k, \nabla_y H(x^k, y^k) \right\rangle, \end{aligned}$$

* In convex case Taylor approximation is under-estimator so $\approx$ becomes $\geq$

[1] If f, g include the indicator function then they are non-smooth

Altenrating minimization in PALM

Function $\Phi(x,y) = f(x) + g(y) + H(x,y)$ has 2 variables.
So alternating minimization scheme gives

$$\begin{aligned}
\arg\min_x \hat{\Phi}(x, y^k) &= \arg\min_x \left\{ f(x) + \underbrace{g(y^k)}_{\text{no } x} + \langle x - x^k, \nabla_x H(x^k, y^k) \rangle \right\} \\
&= \arg\min_x \left\{ f(x) + \langle x - x^k, \nabla_x H(x^k, y^k) \rangle \right\}
\end{aligned}$$

$$\begin{aligned}
\arg\min_y \hat{\Phi}(x^{k+1}, y) &= \arg\min_y \left\{ \underbrace{f(x^{k+1})}_{\text{no } y} + g(y) + \langle y - y^k, \nabla_y H(x^{k+1}, y^k) \rangle \right\} \\
&= \arg\min_y \left\{ g(y) + \langle y - y^k, \nabla_y H(x^{k+1}, y^k) \rangle \right\}.
\end{aligned}$$

i.e. we have

$$\begin{aligned}
\arg\min_x \hat{\Phi}(x, y^k) &= \arg\min_x \left\{ f(x) + \langle x - x^k, \nabla_x H(x^k, y^k) \rangle \right\} \\
\arg\min_y \hat{\Phi}(x^{k+1}, y) &= \arg\min_y \left\{ g(y) + \langle y - y^k, \nabla_y H(x^{k+1}, y^k) \rangle \right\}.
\end{aligned}$$

Proximal operator in PALM

Add proximal term :

$$x_k = \arg\min_x \left\{ f(x) + \underbrace{\left\langle x - x^k, \nabla_x H(x^k, y^k) \right\rangle + \frac{c_k}{2}\|x - x^k\|_2^2}_{\psi_x} \right\}$$

$$y_k = \arg\min_y \left\{ g(y) + \underbrace{\left\langle y - y^k, \nabla_y H(x^{k+1}, y^k) \right\rangle + \frac{d_k}{2}\|y - y^k\|_2^2}_{\psi_y} \right\}.$$

The minimizer of the smooth part ψ (quadratic function) is gradient step :

set $\nabla_x \psi_x = 0$ get $x = x^k - \frac{1}{c_k}\nabla_x H(x^k, y^k)$, $c_k > 0$

set $\nabla_y \psi_y = 0$ get $y = y^k - \frac{1}{d_k}\nabla_y H(x^{k+1}, y^k)$, $d_k > 0$

From theory of gradient descent, (c_k, d_k) can be set to be the partial Lipschitz constant of ∇H

Proximal operator in PALM

Apply proximal operator on $\Phi = \underbrace{f + g}_{\text{non-smooth}} + H$ we have

$$
\begin{aligned}
x_k &\in \operatorname{prox}_{f,c_k}\left(x^k - \frac{1}{c_k}\nabla_x H(x^k, y^k)\right), c_k > 0, \\
y_k &\in \operatorname{prox}_{g,d_k}\left(y^k - \frac{1}{d_k}\nabla_x H(x^{k+1}, y^k)\right), d_k > 0.
\end{aligned}
$$

Recall, at a point u, the poximal map associated to a function $\sigma(x)$ is

$$
\operatorname{prox}_{\sigma,t}(u) = \arg\min_x \left\{\sigma(x) + \frac{t}{2}\|x - u\|_2^2\right\}
$$

Note : the standard gradient descent step $x^k - \frac{1}{c_k}\nabla_x H(x^k, y^k)$ is the "forward step". The proximal step is the "backward step".
So PALM = alternating proximal forward backward method

PALM algorithm

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

Starts wtih $(x^0, y^0) \in$ domΦ, PALM generate (x^k, y^k) as

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

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

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

In words :

- on x, perform gradient update on the smooth part of Φ
- on x, perform proximal update on the non-smooth part of Φ
- on y, perform gradient update on the smooth part of Φ
- on y, perform proximal update on the non-smooth part of Φ
- c_k,d_k are partial Lipschitz constants of H magnified larger than 1

Last page - summary

- Problem (non-convex)

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

- PALM iterations

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

where $c_k = \gamma_1 L_1(y^k)$, $d_k = \gamma_2 L_2(x^{k+1})$ some $\gamma_{1,2} > 1$.

Not covered : condition on Φ that sequence produced PALM converges to a stationary point.

End of document