Novel steplength possibilities for proximal-type algorithms

Federica Porta[1] and Ignace Loris[2]

[1]Università di Modena e Reggio Emilia, [2]Université Libre de Bruxelles

Abstract

In this work we analyze a number of novel steplength selection rules for iterative proximal-based convex optimization algorithms. Proximal methods are a very useful tool for addressing the minimization of a function that includes both a differentiable term with Lipschitz continuous gradient and a non-smooth term. Classical proximal approaches heavily exploit the knowledge of the gradient Lipschitz constant in order to select the steplength in the updating iteration. Since the value of this constant is not always available, we discuss two possibilities of realizing convergent proximal methods completely independent on the Lipschitz parameter. Numerical experiments carried out on signal recovering test problems show the performance of the methods in comparison to existing schemes.

Proposed proximal-type algorithms [1]

We consider the problem of minimizing the sum of two given functions:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \mathcal{F}(\boldsymbol{x}) \equiv f(\boldsymbol{x}) + g(\boldsymbol{x}) \qquad (1)$$

- $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ is a convex, continuously differentiable function;
- $g : \mathbb{R}^N \longrightarrow \overline{\mathbb{R}}$ is an extended-value convex function.

Proximal Khobotov-like Method (PKM)

Choose the starting point $\boldsymbol{x}^0$ and the parameters $\alpha_0 > 0$, $\rho \in (0,1)$.

for $k = 0, 1, 2, \dots$ **do**

STEP 1. $\bar{\boldsymbol{x}}_n = \mathbf{prox}_{\alpha_n g}(\boldsymbol{x}_n - \alpha_n \nabla f(\boldsymbol{x}_n))$

STEP 2.

if $\alpha_n^2 \dfrac{\|\nabla f(\boldsymbol{x}_n) - \nabla f(\bar{\boldsymbol{x}}_n)\|}{\|\boldsymbol{x}_n - \bar{\boldsymbol{x}}_n\|} > \rho^2$

then $\alpha_n = \min\left\{\dfrac{\alpha_n}{2}, \dfrac{\|\nabla f(\boldsymbol{x}_n) - \nabla f(\bar{\boldsymbol{x}}_n)\|}{\sqrt{2}\|\boldsymbol{x}_n - \bar{\boldsymbol{x}}_n\|}\right\}$, **goto** STEP 1

else **goto** STEP 3

endif

STEP 3. $\boldsymbol{x}_{n+1} = \mathbf{prox}_{\alpha_n g}(\boldsymbol{x}_n - \alpha_n \nabla f(\bar{\boldsymbol{x}}_n))$, $\alpha_{n+1} = \alpha_n$

end for

- **Convergence.** Let $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ be the sequence generated by PKM, then $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ converges to a minimizer of problem (1).
- **Remark.** If g is the indicator function of a set, PKM reduces to the Khobotov extra-gradient method.
- **Advantage.** It is possible to prove convergence for PKM even when the proximal operator of g cannot be computed exactly.

Proximal Armijo-like Method (PAM)

Choose the starting point $\boldsymbol{x}^0$ and the parameters $\alpha_0 > 0$, $0 < \beta, \sigma < 1$.

for $k = 0, 1, 2, \dots$ **do**

STEP 1. $\bar{\boldsymbol{x}}_n = \mathbf{prox}_{\alpha_n g}(\boldsymbol{x}_n - \alpha_n \nabla f(\boldsymbol{x}_n))$;
$\boldsymbol{d}_n = \bar{\boldsymbol{x}}_n - \boldsymbol{x}_n$, set $\lambda_n = 1$

STEP 2.

if $\mathcal{F}(\boldsymbol{x}_n + \lambda_n \boldsymbol{x}_n) < \mathcal{F}(\boldsymbol{x}_n) + \sigma\lambda_n [\langle \nabla f(\boldsymbol{x}_n), \boldsymbol{d}_n\rangle - g(\boldsymbol{x}_n) + g(\bar{\boldsymbol{x}}_n)]$

then **goto** STEP 3

else

$\lambda_n = \beta\lambda_n$, **goto** STEP 2

endif

STEP 3. $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n \boldsymbol{d}_n$; $\alpha_{n+1} \in [\alpha_{min}, \alpha_{max}]$.

end for

- **Convergence.** Let $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ be the sequence generated by PAM, then every limit point of $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ is a stationary point.
- **Remark.** If g is the indicator function of a set, PAM reduces to the Armijo gradient projection method.
- **Advantage.** The convergence of PAM is assured by the backtracking procedure on λ_n (STEP 2), therefore it is possible to exploit the choice of the steplength α_n to obtain a better convergence rate.

Poissonian sparse signal recovering

Signal formation process: an inverse problem

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{\eta}$$

- $\boldsymbol{y} \in \mathbb{R}^M$: observed data;
- $\boldsymbol{x} \in \mathbb{R}^N$: signal to be recovered;
- $H \in \mathbb{R}^{M\times N}$: measurement matrix;
- $\boldsymbol{b} \in \mathbb{R}^M$: background;
- $\boldsymbol{\eta} \in \mathbb{R}^M$: Poisson noise corrupting the data.

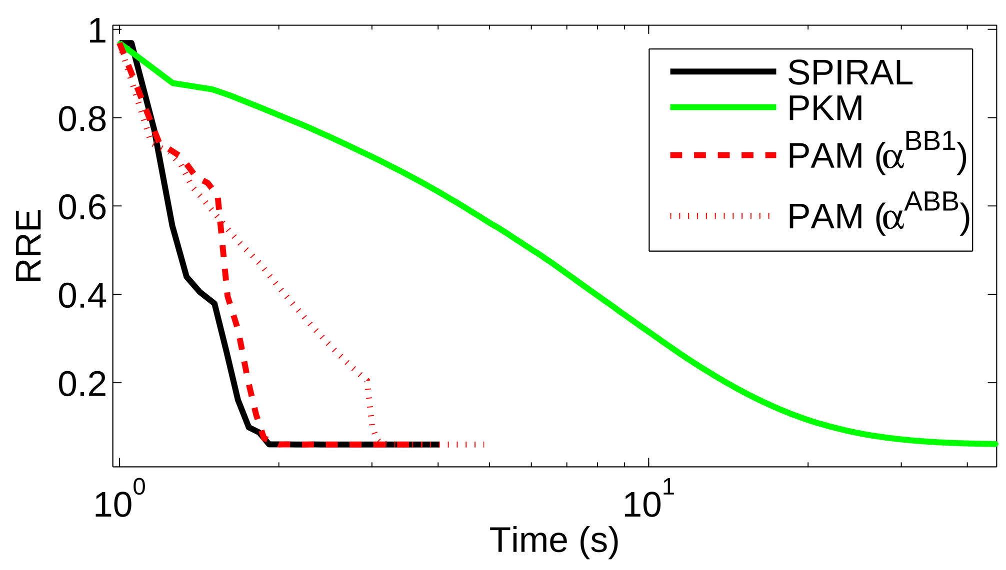
Signal restoration: optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \mathrm{KL}(\boldsymbol{x}) + \mu\|\boldsymbol{x}\|_1 + \mathcal{I}_{\{\boldsymbol{x}\geq 0\}}$$

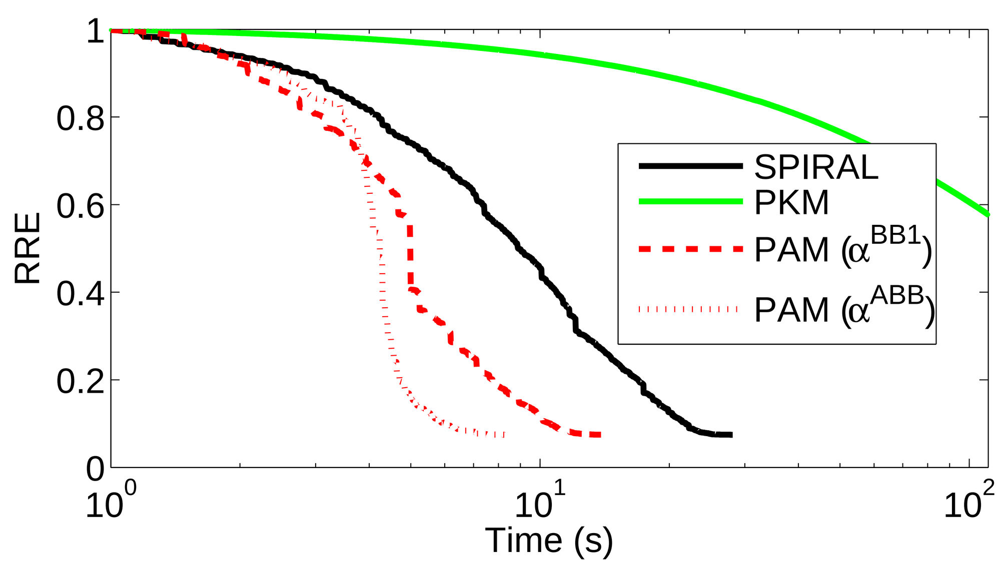
- $\mathrm{KL}(\boldsymbol{x}) = \sum_{i=1}^N \left\{\boldsymbol{y}_i \ln \frac{\boldsymbol{y}_i}{(H\boldsymbol{x}+\boldsymbol{b})_i} + (H\boldsymbol{x}+\boldsymbol{b})_i - \boldsymbol{y}_i\right\}$;
- $\mu > 0$: regularization parameter;
- $\mathcal{I}_{\{\boldsymbol{x}\geq 0\}}$ indicator function of the non-negative orthant

Numerical results: comparison with SPIRAL method [2]

First dataset ($N = 10^5, M = 4\times 10^3$)



Second dataset ($N = 5\times 10^3, M = 10^3$)



Conclusion. PAM shows very promising results in addressing this type of problems: it outperforms SPIRAL for the second dataset. PKM convergence rate is very slow, probably due to the line-search procedure on the steplength.

References

[1] Porta, F., Loris, I.: On some steplength approaches for proximal algorithms. (submitted to Applied Mathematics and Computation - 2014)

[2] Harmany, Z., Marcia, R., Willett, R.: This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms-Theory and Practice. IEEE Trans. Image Processing 21(3), 1084-1096 (2012)