Alternating structure-adapted proximal gradient descent for nonconvex regularized problems (ASAP)

Pauline Tan
Laboratoire Jacques-Louis Lions, Sorbonne Université

Scientific Day in Memory of Prof. Mila Nikolova

ENS Paris-Saclay, Cachan, October 15th, 2018

Scientific collaboration from 2016 to 2018:

- at Onera - The French Aerospace Lab (2016 – 2017)
- at CMLA (DIM Math Innov postdoctoral position) (2017 – 2018)

Submitted papers:

- *Alternating structure-adapted proximal gradient descent for nonconvex block-regularised problems*
  Mila Nikolova, P. T., *submitted*, 2017 (HAL-01677456)
- *Inertial Alternating Generalized Forward-Backward Splitting for Image Colorization*
  P.T., Fabien Pierre, Mila Nikolova, *submitted*, 2018 (HAL-01792432)

Motivation

Question: How to jointly estimate two objects x^* and y^*, where

$$x^* = \arg\min_x F(x) + H(x, y^*) \qquad \text{convex problem}$$

$$y^* = \arg\min_y G(y) + H(x^*, y) \qquad \text{convex problem}$$

$\longrightarrow$ Biconvex (nonconvex) optimization problem

$$\min_{x,y} J(x, y) = F(x) + G(y) + H(x, y)$$

Applications: joint optimization, blind source separation, blind deconvolution, nonnegative matrix factorization, structured total least squares, multimodal learning for image classification, etc.

Alternating optimization

If J convex in (x, y), then block coordinate descent (BCD)-like strategy:

- alternating partial minimization
- alternating explicit/implicit gradient descent
- alternating forward-backward splitting (FBS)

$\longrightarrow$ Such algorithms are applicable as soon as the x-problem and the y-problem are convex, but convergence?

Related work: PALM

Assume $$J(x,y) = \underbrace{F(x) + G(y)}_{\text{nonsmooth and prox-friendly}} + \underbrace{H(x,y)}_{\text{smooth}}$$

Optimization in the x-direction

$$\underbrace{\text{prox}_{\tau^k F}}_{\text{implicit /}}(x^k \underbrace{- \tau^k \nabla_x H(x^k, y^k)}_{\text{explicit gradient descent}}) \qquad \text{(same for } y^k\text{)}$$

Introduced by Xu and Yin ('13) and Bolte, Sabach and Teboulle ('14)

What if

- F or G (regularizers) are not prox-friendly?
- step-sizes (τ^k, σ^k) (which depend on the Lipschitz constants of $\nabla_x H(\cdot, y^k)$ and $\nabla_y H(x^{k+1}, \cdot)$) are hard to estimated at each update?

Proposed ASAP algorithm

If

- H has a nice structure in the sense that H is partially prox-friendly (in x/y-direction)
- regularizers F and G are smooth (or can be replaced by smooth approximations)

Idea: Invert the roles of the components of J in the FBS, *i.e.* replace

$$\mathrm{prox}_{\tau^k F}(x^k - \tau^k \nabla_x H(x^k, y^k)) \qquad \text{(PALM)}$$

by

$$\mathrm{prox}_{\tau H(\cdot, y^k)}(x^k - \tau \nabla F(x^k)) \qquad \text{(ASAP)}$$

$\longrightarrow$ ASAP = 'mirror' of PALM

Theoretical issues

$$J(x,y) = F(x) + G(y) + H(x,y)$$

For ASAP, the coupling term H does not need to be smooth

- subdifferential vs. partial subdifferentials:

$$\partial J(x,y) \neq \partial_x J(x,y) \times \partial_y J(x,y)$$

- parametric closedness of the partial subdifferentials:

$$\begin{cases} (x^k, y^k) \underset{k\to+\infty}{\longrightarrow} (x^*, y^*) \\ (\tilde{x}^k, \tilde{y}^k) \underset{k\to+\infty}{\longrightarrow} (x^*, y^*) \\ p_x^k \in \partial_x J(x^k, y^k) \\ q_y^k \in \partial_y J(\tilde{x}^k, \tilde{y}^k) \\ p_x^k \underset{k\to+\infty}{\longrightarrow} p_x^*, \, q_y^k \underset{k\to+\infty}{\longrightarrow} q_y^* \end{cases} \quad \not\Longrightarrow (p_x^*, q_y^*) \in \partial_x J(x^*, y^*) \times \partial_y J(x^*, y^*)$$

Partial optimization

$x^{k+1} \longleftarrow$ optimization of $J(\cdot, y^k)$
$y^{k+1} \longleftarrow$ optimization of $J(x^{k+1}, \cdot)$

First-order optimality conditions give
$p_x^k \in \partial_x J(x^{k+1}, y^k)$ and $q_y^k \in \partial_y J(x^{k+1}, y^{k+1})$

Usually, one proves that $p_x^k \to 0$ and $q_y^k \to 0$ with $(x^k, y^k) \to (x^*, y^*)$

Under some conditions (convexity, smoothness...) this implies that

- $(0, 0) \in \partial_x J(x^*, y^*) \times \partial_y J(x^*, y^*)$ (*e.g.* H differentiable)
- $\partial_x J(x^*, y^*) \times \partial_y J(x^*, y^*) = \partial J(x^*, y^*)$ (*e.g.* J convex or H smooth)

In general $(0, 0) \notin \partial J(x^*, y^*)$

i.e. (x^*, y^*) is NOT a critical point

Convergence of ASAP

Assume $\inf_{x,y} J(x,y) = F(x) + G(y) + H(x,y) > -\infty$

ASAP: Choose $\tau < 1/\mathrm{Lip}(\nabla F)$ and $\sigma < 1/\mathrm{Lip}(\nabla G)$ and compute

$$\begin{cases} x^{k+1} \in \mathrm{prox}_{\tau H(\cdot, y^k)}(x^k - \tau \nabla F(x^k)) \\ y^{k+1} \in \mathrm{prox}_{\sigma H(x^{k+1}, \cdot)}(y^k - \sigma \nabla G(y^k)) \end{cases}$$

Convergence in value

If the iterations are computable

(*i.e.* F and G smooth, and prox of $H(\cdot, y^k)$ and $H(x^{k+1}, \cdot)$ computable)

then the sequence $J(x^k, y^k)$ decreases to a finite value J^*

(ASAP is a descent scheme)

Convergence of ASAP (2)

Convergence to the set of critical points

If the iterations are computable, and

- H is continuous on its closed domain
- $\partial_x H(x,y) \times \partial_y H(x,y) \subset \partial H(x,y)$
  (*e.g.* H is differentiable)
- the parametric closedness of the partial subdifferentials holds for H
  (*e.g.* H is smooth or H is biconvex)
- $\{(x^k, y^k)\}$ is bounded (*e.g.* $\mathrm{dom} H$ is bounded)

then any limit point of $\{(x^k, y^k)\}$ is a critical point and

$$\mathrm{dist}\big((x^k, y^k), \mathrm{crit}(J)\big) \underset{k \to +\infty}{\longrightarrow} 0$$

Convergence of ASAP (3)

In addition,

- if $\nabla_x H(x, \cdot)$ is locally Lipschitz
- and J has the Kurdyka-Łojasiewicz property at a critical point of J (*e.g.* most of the sums/compositions of real-analytic and semi-algebraic functions)

then ASAP generates a Cauchy (convergent) sequence

Extensions of ASAP

Partial convexity

If $H(\cdot, y)$ is convex, then the stepsize τ can be chosen twice larger
If $H(x, \cdot)$ is convex, then the stepsize σ can be chosen twice larger

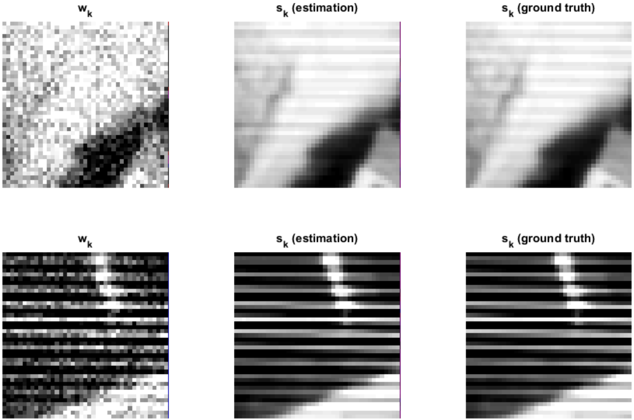
Bregman generalization

In the backward step, the proximity operator can be replaced by a generalized one using a Bregman distance
(*e.g.* for the optimization on a simplex)

Acceleration using inertia

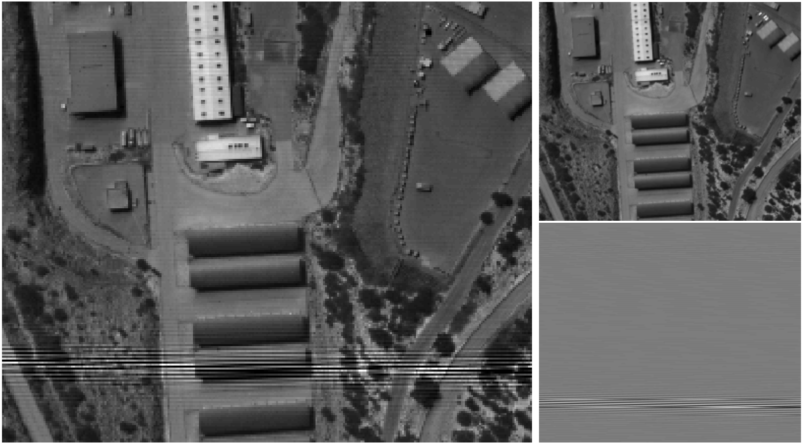
Overrelaxation steps can be added and may lead to empirical acceleration

Denoising in interferometric imagery (Onera)



Correction par méthode variationnelle des non uniformités des détecteurs d'un interféromètre imageur, P. T., Yann Ferrec, Laurent Rousset-Rouvière, colloque GRETSI, 2017

Fringe separation (Onera)



Fast and Accurate Multiplicative Decomposition for Fringe Removal in Interferometric Images, Daniel Chen Soncco, Clara Barbanson, Mila Nikolova, Andrés Almansa, Yann Ferrec, IEEE Transactions on Computational Imaging, 2017

Image colorization

Inertial Alternating Generalized Forward-Backward Splitting for Image Colorization
P.T., Fabien Pierre, Mila Nikolova, *submitted*, 2018 (HAL-01792432)

Some concluding notes

- The proposed ASAP is an alternative scheme to PALM for solving nonsmooth and nonconvex optimization problem
- Choice between ASAP and PALM depends on the structure and the regularity of the objective
- Biconvexity of the coupling term gives nice properties (large stepsizes)
- Promising applications on image processing
- Open questions: critical points vs. (local) minimum, initialization, theoretical convergence rate

Thank you for your attention!