# An Alternating Augmented Lagrangian method for constrained nonconvex optimization

6th July 2018 - ISMP Bordeaux, France

Giulio Galvan (speaker),
Matteo Lapucci,
Tommaso Levato,
Marco Sciandrone,

*DINFO, Università degli Studi di Firenze.*

# Introduction

## Problem Statement

$$\begin{aligned} \min_{x} \quad & f(x) \\ s.t. \quad & x \in X \cap Y, \end{aligned}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function,
- $X$, $Y$ are nonempty, convex, closed sets, "easy to treat separately".

**Example**

$$\begin{aligned} \min_{x} \quad & f(x) \\ s.t. \quad & a^T x = b, \\ & l \leq x \leq u. \end{aligned}$$

# Problem Statement

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & x \in X \cap Y, \end{aligned}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function,
- $X$, $Y$ are nonempty, convex, closed sets, "easy to treat separately".

**Example**

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & a^T x = b, \\ & l \leq x \leq u. \end{aligned}$$

# Problem Statement

$$\begin{aligned} \min_{x} \quad & f(x) \\ s.t. \quad & x \in X \cap Y, \end{aligned}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function,
- $X$, $Y$ are nonempty, convex, closed sets, "easy to treat separately".

**Example**

$$\begin{aligned} \min_{x} \quad & f(x) \\ s.t. \quad & a^T x = b, \\ & l \leq x \leq u. \end{aligned}$$

## Problem Statement

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & x \in X \cap Y, \end{aligned}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function,
- $X$, $Y$ are nonempty, convex, closed sets, "easy to treat separately".

**Example**

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & a^T x = b, \\ & l \le x \le u. \end{aligned}$$

# ADMM (DRS)

**Reformulation**

$$\begin{aligned} \min_{x,y} \quad & f(x) \\ s.t. \quad & x \in X,\ y \in Y, \\ & x - y = 0. \end{aligned}$$

**ADMM**

$$\begin{aligned} x^{k+1} &\in \operatorname*{argmin}_{x \in X} f(x) + {\lambda^k}^T (x - y^k) + \frac{\tau}{2}\|x - y^k\|^2, \\ y^{k+1} &= \operatorname*{argmin}_{y \in Y} {\lambda^k}^T (x^{k+1} - y) + \frac{\tau}{2}\|x^{k+1} - y\|^2, \\ \lambda^{k+1} &= \lambda^k + \tau(x^{k+1} - y^{k+1}). \end{aligned}$$

# ADMM (DRS)

**Reformulation**

$$\begin{aligned} \min_{x,y} \quad & f(x) \\ s.t. \quad & x \in X, \; y \in Y, \\ & x - y = 0. \end{aligned}$$

**ADMM**

$$\begin{aligned} x^{k+1} &\in \operatorname*{argmin}_{x \in X} f(x) + {\lambda^k}^T (x - y^k) + \frac{\tau}{2} \|x - y^k\|^2, \\ y^{k+1} &= \operatorname*{argmin}_{y \in Y} {\lambda^k}^T (x^{k+1} - y) + \frac{\tau}{2} \|x^{k+1} - y\|^2, \\ \lambda^{k+1} &= \lambda^k + \tau (x^{k+1} - y^{k+1}). \end{aligned}$$

# Convergence of ADMM

- convergence is well understood in the convex case,
- the non convex case is an open reasearch topic,
- $\tau$ is a "fixed" parameter,
- $\tau$ is a critical parameter for the performances of ADMM.

**$\tau$ in practice**

$$\tau_{k+1} = \begin{cases} \alpha_1 \tau_k & \text{if } \|x^{k+1} - y^{k+1}\| > \mu \|y^{k+1} - y^k\|, \\ \tau_k / \alpha_2 & \text{if } \|y^{k+1} - y^k\| \quad > \mu \|x^{k+1} - y^{k+1}\|, \\ \tau_k & \text{otherwise.} \end{cases}$$

# Convergence of ADMM

- convergence is well understood in the convex case,
- the non convex case is an open reasearch topic,
- $\tau$ is a "fixed" parameter,
- $\tau$ is a critical parameter for the performances of ADMM.

**$\tau$ in practice**

$$\tau_{k+1} = \begin{cases} \alpha_1 \tau_k & \text{if } \|x^{k+1} - y^{k+1}\| > \mu \|y^{k+1} - y^k\|, \\ \tau_k / \alpha_2 & \text{if } \|y^{k+1} - y^k\| \quad > \mu \|x^{k+1} - y^{k+1}\|, \\ \tau_k & \text{otherwise.} \end{cases}$$

# Convergence of ADMM

- convergence is well understood in the convex case,
- the non convex case is an open reasearch topic,
- $\tau$ is a "fixed" parameter,
- $\tau$ is a critical parameter for the performances of ADMM.

**$\tau$ in practice**

$$\tau_{k+1} = \begin{cases} \alpha_1 \tau_k & \text{if } \|x^{k+1} - y^{k+1}\| > \mu \|y^{k+1} - y^k\|, \\ \tau_k / \alpha_2 & \text{if } \|y^{k+1} - y^k\| \quad > \mu \|x^{k+1} - y^{k+1}\|, \\ \tau_k & \text{otherwise.} \end{cases}$$

## Convergence of ADMM

- convergence is well understood in the convex case,
- the non convex case is an open reasearch topic,
- $\tau$ is a "fixed" parameter,
- $\tau$ is a critical parameter for the performances of ADMM.

**$\tau$ in practice**

$$\tau_{k+1} = \begin{cases} \alpha_1 \tau_k & \text{if } \|x^{k+1} - y^{k+1}\| > \mu \|y^{k+1} - y^k\|, \\ \tau_k / \alpha_2 & \text{if } \|y^{k+1} - y^k\| \quad > \mu \|x^{k+1} - y^{k+1}\|, \\ \tau_k & \text{otherwise.} \end{cases}$$

# Convergence of ADMM

- convergence is well understood in the convex case,
- the non convex case is an open reasearch topic,
- $\tau$ is a "fixed" parameter,
- $\tau$ is a critical parameter for the performances of ADMM.

**$\tau$ in practice**

$$\tau_{k+1} = \begin{cases} \alpha_1 \tau_k & \text{if } \|x^{k+1} - y^{k+1}\| > \mu \|y^{k+1} - y^k\|, \\ \tau_k / \alpha_2 & \text{if } \|y^{k+1} - y^k\| \quad > \mu \|x^{k+1} - y^{k+1}\|, \\ \tau_k & \text{otherwise.} \end{cases}$$

# An Augmented Lagrangian approach

# Augmented Lagrangian framework

Solve (approximately) a sequence of subproblems of the form

$$\min_{x \in X, y \in Y} q_{\tau_k, \lambda^k}(x, y) \triangleq L_{\tau_k}(x, y, \lambda^k) = f(x) + {\lambda^k}^T (x - y) + \frac{\tau_k}{2} \|x - y\|^2,$$

with

- increasing $\tau^k$,
- bounded $\lambda^k$.

# The framework (ALF)

$\alpha > 1, \sigma \in (0,1), \tau_0 > 0,$

$\{\epsilon_k\}$ s.t. $\lim_{k\to\infty} \epsilon_k = 0$, $\Lambda \leftarrow$ bounded set;

**for** *k=0, 1, 2 . . .* **do**

$(x^{k+1}, y^{k+1}) \leftarrow \epsilon_k$-stationary point for $\min_{x\in X, y\in Y} q_{\tau_k,\lambda^k}(x,y)$;

$\lambda^{k+1} \leftarrow \lambda^k + \tau_k(x^{k+1} - y^{k+1})$;

$\lambda^{k+1} \leftarrow [\lambda^{k+1}]_\Lambda^+$;

**if** $\|x^{k+1} - y^{k+1}\| \le \sigma \|x^k - y^k\|$ **then**

$\tau_{k+1} = \tau_k$;

**else**

$\tau_{k+1} = \alpha\tau_k$;

**end**

**end**

## The framework (ALF)

$\alpha > 1, \sigma \in (0,1), \tau_0 > 0,$
$\{\epsilon_k\}$ s.t. $\lim_{k\to\infty} \epsilon_k = 0,\ \Lambda \leftarrow$ bounded set;

**for** *k=0, 1, 2 . . .* **do**
  $(x^{k+1}, y^{k+1}) \leftarrow \epsilon_k$-stationary point for $\min_{x\in X, y\in Y} q_{\tau_k,\lambda^k}(x,y)$;
  $\lambda^{k+1} \leftarrow \lambda^k + \tau_k(x^{k+1} - y^{k+1})$;
  $\lambda^{k+1} \leftarrow [\lambda^{k+1}]^+_\Lambda$;
  **if** $\|x^{k+1} - y^{k+1}\| \le \sigma\|x^k - y^k\|$ **then**
    $\tau_{k+1} = \tau_k$;
  **else**
    $\tau_{k+1} = \alpha\tau_k$;
  **end**
**end**

## The framework (ALF)

$\alpha > 1, \sigma \in (0,1), \tau_0 > 0,$

$\{\epsilon_k\}$ s.t. $\lim_{k\to\infty} \epsilon_k = 0,\ \Lambda \leftarrow$ bounded set;

**for** *k=0, 1, 2 . . .* **do**

- $(x^{k+1}, y^{k+1}) \leftarrow \epsilon_k$-stationary point for $\min_{x\in X, y\in Y} q_{\tau_k,\lambda^k}(x,y)$;
- $\lambda^{k+1} \leftarrow \lambda^k + \tau_k(x^{k+1} - y^{k+1})$;
- $\lambda^{k+1} \leftarrow [\lambda^{k+1}]^+_\Lambda$;
- **if** $\|x^{k+1} - y^{k+1}\| \le \sigma\|x^k - y^k\|$ **then**
  - $\tau_{k+1} = \tau_k$;
- **else**
  - $\tau_{k+1} = \alpha\tau_k$;
- **end**

**end**

## Convergence

**Theorem**

*Let* $\{(x^k, y^k)\}$ *be the sequence generated by Algorithm* $\mathrm{ALF}$. *Then every cluster point* $(\bar{x}, \bar{y})$ *of* $\{(x^k, y^k)\}$ *is such that* $\bar{x} = \bar{y}$ *and* $\bar{x}$ *is a critical point of the original problem i.e.*

$$\|\bar{x} - [\bar{x} - \nabla f(\bar{x})]^+_{X \cap Y}\| = 0.$$

# Towards ADMM

# Roadmap

- design a solver for the subproblems that employ the ADMM steps,
- change the **argmin** operation to something suitable for the non convex setting,
- embed it in the ALF.

# Roadmap

- design a solver for the subproblems that employ the ADMM steps,
- change the **argmin** operation to something suitable for the non convex setting,
- embed it in the ALF.

## Roadmap

- design a solver for the subproblems that employ the ADMM steps,
- change the **argmin** operation to something suitable for the non convex setting,
- embed it in the ALF.

# Alternate Minimization (AM)

Let $x^0 \in X$, $y^0 \in Y$,

$t = 0$;

**while** $(x^t, y^t)$ *is not an $\epsilon$-stationary point* **do**

find $x^{t+1}$ such that $q_{\tau,\lambda}(x^{t+1}, y^t) \leq q_{\tau,\lambda}(x^t, y^t)$ and

$$\nabla_x q_{\tau,\lambda}(x^{t+1}, y^t)^T (x - x^{t+1}) \geq 0 \quad \forall x \in X;$$

find $y^{t+1}$ such that

$$y^{t+1} \in \operatorname*{argmin}_{y \in Y} q_{\tau,\lambda}(x^{t+1}, y);$$

$t = t + 1$;

**end**

- checking $\nabla_x q_{\tau,\lambda}(x^{t+1}, y^t)^T (x - x^{t+1}) \geq 0 \;\; \forall x \in X$ may be difficult in practice.

# Alternate Minimization (AM)

Let $x^0 \in X$, $y^0 \in Y$,

$t = 0$;

**while** $(x^t, y^t)$ *is not an $\epsilon$-stationary point* **do**

find $x^{t+1}$ such that $q_{\tau,\lambda}(x^{t+1}, y^t) \le q_{\tau,\lambda}(x^t, y^t)$ and

$$\nabla_x q_{\tau,\lambda}(x^{t+1}, y^t)^T (x - x^{t+1}) \ge 0 \;\; \forall x \in X;$$

find $y^{t+1}$ such that

$$y^{t+1} \in \operatorname*{argmin}_{y \in Y} q_{\tau,\lambda}(x^{t+1}, y);$$

$t = t + 1$;

**end**

- checking $\nabla_x q_{\tau,\lambda}(x^{t+1}, y^t)^T (x - x^{t+1}) \ge 0 \;\; \forall x \in X$ may be difficult in practice.

## Alternate Minimization v2

Let $x^0 \in X$, $y^0 \in Y$,

$t = 0$;

**while** $(x^t, y^t)$ *is not an* $\epsilon$*-stationary point* **do**

find $x^{t+1}$ such that

$$q_{\tau,\lambda}(x^{t+1}, y^t) \leq q_{\tau,\lambda}(x^t + \alpha_t d^t, y^t),$$

where $d^t = [x^t - \nabla_x q_{\tau,\lambda}(x^t, y^t)]_X^+ - x^t$ and $\alpha_t$ is obtained with an Armijo-type line search;

$y^{t+1} = \operatorname{argmin}_{y \in Y} q_{\tau,\lambda}(x^{t+1}, y)$;

$t = t + 1$;

**end**

## Convergence of AM

**Proposition**

Algorithm $\mathrm{AM(v2)}$ determines in a finite number of iterations a point $(x(\epsilon), y(\epsilon))$ which is an $\epsilon$-stationary point for problem

$$\min_{x \in X, y \in Y} L_\tau(x, y, \lambda).$$

## The complete algorithm (ALTALM)

$\alpha > 1,\ \sigma \in (0,1),\ \tau_0 > 0,\ \{\epsilon_k\}$ s.t. $\lim_{k\to+\infty} \epsilon_k = 0,\ \Lambda \leftarrow$ bounded set;

**for** *k=0, 1, 2...* **do**

  $x^t, y^t \leftarrow x^k, y^k$;

  **while** $(x^t, y^t)$ *is not an $\epsilon$-stationary* **do**

    find $x^{t+1}$ such that

$$q_{\tau,\lambda}(x^{t+1}, y^t) \leq q_{\tau,\lambda}(x^t + \alpha_t d^t, y^t);$$

    $y^{t+1} = \text{argmin}_{y\in Y}\, q_{\tau,\lambda}(x^{t+1}, y)$;

    $t = t + 1$;

  **end**

  $x^{k+1}, y^{k+1} \leftarrow x^t, y^t$;

  $\lambda^{k+1} \leftarrow \lambda^k + \tau_k(x^{k+1} - y^{k+1})$

  $\lambda^{k+1} \leftarrow [\lambda^{k+1}]^+_\Lambda$

  **if** $\|x^{k+1} - y^{k+1}\| \leq \sigma\|x^k - y^k\|$ **then**

    $\tau_{k+1} = \tau_k$,

  **else**

    $\tau_{k+1} = \alpha\tau_k$.

  **end**

**end**

# The complete algorithm (ALTALM)

$\alpha > 1,\ \sigma \in (0,1),\ \tau_0 > 0,\ \{\epsilon_k\}$ s.t. $\lim_{k\to+\infty} \epsilon_k = 0,\ \Lambda \leftarrow$ bounded set;

**for** *k=0, 1, 2...* **do**

$x^t, y^t \leftarrow x^k, y^k$;

**while** $(x^t, y^t)$ *is not an* $\epsilon$*-stationary* **do**

find $x^{t+1}$ such that

$$q_{\tau,\lambda}(x^{t+1}, y^t) \leq q_{\tau,\lambda}(x^t + \alpha_t d^t, y^t);$$

$y^{t+1} = \text{argmin}_{y\in Y}\, q_{\tau,\lambda}(x^{t+1}, y)$;

$t = t + 1$;

**end**

$x^{k+1}, y^{k+1} \leftarrow x^t, y^t$;

$\lambda^{k+1} \leftarrow \lambda^k + \tau_k(x^{k+1} - y^{k+1})$

$\lambda^{k+1} \leftarrow [\lambda^{k+1}]_\Lambda^+$

**if** $\|x^{k+1} - y^{k+1}\| \leq \sigma\|x^k - y^k\|$ **then**

$\tau_{k+1} = \tau_k$,

**else**

$\tau_{k+1} = \alpha\tau_k$.

**end**

**end**

## The complete algorithm (ALTALM)

$\alpha > 1,\ \sigma \in (0,1),\ \tau_0 > 0,\ \{\epsilon_k\}$ s.t. $\lim_{k\to+\infty} \epsilon_k = 0,\ \Lambda \leftarrow$ bounded set;

**for** *k=0, 1, 2...* **do**

$x^t, y^t \leftarrow x^k, y^k$;

**while** $(x^t, y^t)$ *is not an $\epsilon$-stationary* **do**

find $x^{t+1}$ such that

$$q_{\tau,\lambda}(x^{t+1}, y^t) \leq q_{\tau,\lambda}(x^t + \alpha_t d^t, y^t);$$

$y^{t+1} = \text{argmin}_{y\in Y}\, q_{\tau,\lambda}(x^{t+1}, y)$;

$t = t+1$;

**end**

$x^{k+1}, y^{k+1} \leftarrow x^t, y^t$;

$\lambda^{k+1} \leftarrow \lambda^k + \tau_k(x^{k+1} - y^{k+1})$

$\lambda^{k+1} \leftarrow [\lambda^{k+1}]^+_\Lambda$

**if** $\|x^{k+1} - y^{k+1}\| \leq \sigma\|x^k - y^k\|$ **then**

$\tau_{k+1} = \tau_k$,

**else**

$\tau_{k+1} = \alpha\tau_k$.

**end**

**end**

## The complete algorithm (ALTALM)

$\alpha > 1,\ \sigma \in (0,1),\ \tau_0 > 0,\ \{\epsilon_k\}$ s.t. $\lim_{k\to+\infty} \epsilon_k = 0,\ \Lambda \leftarrow$ bounded set;

**for** *k=0, 1, 2...* **do**

$x^t, y^t \leftarrow x^k, y^k$;

**while** $(x^t, y^t)$ *is not an $\epsilon$-stationary* **do**

find $x^{t+1}$ such that

$$q_{\tau,\lambda}(x^{t+1}, y^t) \le q_{\tau,\lambda}(x^t + \alpha_t d^t, y^t);$$

$y^{t+1} = \text{argmin}_{y\in Y}\, q_{\tau,\lambda}(x^{t+1}, y)$;

$t = t + 1$;

**end**

$x^{k+1}, y^{k+1} \leftarrow x^t, y^t$;

$\lambda^{k+1} \leftarrow \lambda^k + \tau_k(x^{k+1} - y^{k+1})$

$\lambda^{k+1} \leftarrow [\lambda^{k+1}]^+_\Lambda$

**if** $\|x^{k+1} - y^{k+1}\| \le \sigma\|x^k - y^k\|$ **then**

$\tau_{k+1} = \tau_k$,

**else**

$\tau_{k+1} = \alpha\tau_k$.

**end**

**end**

# Numerical Experiments

# SVM training - Indefinite Kernels

$$\min_x f(x) \triangleq \frac{1}{2} x^T Q x - e^T x$$
$$\text{s.t. } a^T x = 0,$$
$$0 \leq x_i \leq C \quad \forall\, i \in 1, \dots, n.$$

$Q_{ij} \triangleq a_i a_j K(u_i, u_j)$,
$u_i$ are the patterns of the dataset,
$a_i$ are the labels.

- The *sigmoid kernel*:

$$K(u, v) = \tanh(a u^T v + b),$$

- The *gaussian combination kernel*:

$$k(u, v) = \exp\left(-\frac{\|u - v\|^2}{\sigma_1}\right) + \exp\left(-\frac{\|u - v\|^2}{\sigma_2}\right) - \exp\left(-\frac{\|u - v\|^2}{\sigma_3}\right).$$

## SVM training - Indefinite Kernels

$$\min_x f(x) \triangleq \frac{1}{2}x^T Qx - e^T x$$
$$\text{s.t. } a^T x = 0,$$
$$0 \leq x_i \leq C \quad \forall\, i \in 1, \ldots, n.$$

$Q_{ij} \triangleq a_i a_j K(u_i, u_j)$,
$u_i$ are the patterns of the dataset,
$a_i$ are the labels.

- The *sigmoid kernel*:

$$K(u, v) = \tanh(au^T v + b),$$

- The *gaussian combination kernel*:

$$k(u, v) = \exp\left(-\frac{\|u - v\|^2}{\sigma_1}\right) + \exp\left(-\frac{\|u - v\|^2}{\sigma_2}\right) - \exp\left(-\frac{\|u - v\|^2}{\sigma_3}\right).$$

# SVM training - Indefinite Kernels

$$\min_x f(x) \triangleq \frac{1}{2}x^T Qx - e^T x$$
$$\text{s.t. } a^T x = 0,$$
$$0 \le x_i \le C \quad \forall\, i \in 1, \dots, n.$$

$Q_{ij} \triangleq a_i a_j K(u_i, u_j)$,
$u_i$ are the patterns of the dataset,
$a_i$ are the labels.

- The *sigmoid kernel*:

$$K(u, v) = \tanh(au^T v + b),$$

- The *gaussian combination kernel*:

$$k(u, v) = \exp\left(-\frac{\|u - v\|^2}{\sigma_1}\right) + \exp\left(-\frac{\|u - v\|^2}{\sigma_2}\right) - \exp\left(-\frac{\|u - v\|^2}{\sigma_3}\right).$$

# The augmented Lagrangian subproblem

$$
\begin{aligned}
\min_{x,y}\ & \frac{1}{2}x^T Q x - e^T x + \lambda^T (x-y) + \frac{\tau}{2}\|x-y\|^2 \\
\text{s.t. } & 0 \le x_i \le C \quad \forall\, i \in 1,\dots,n \\
& a^T y = 0.
\end{aligned}
$$

- QP solver for the $x$ step,
- closed form solution for $y$

## The augmented Lagrangian subproblem

$$\min_{x,y} \frac{1}{2}x^T Qx - e^T x + \lambda^T x + \lambda^T y + \frac{\tau}{2}\|x - y\|^2$$
$$\text{s.t. } 0 \leq x_i \leq C \quad \forall\, i \in 1, \ldots, n$$
$$a^T y = 0.$$

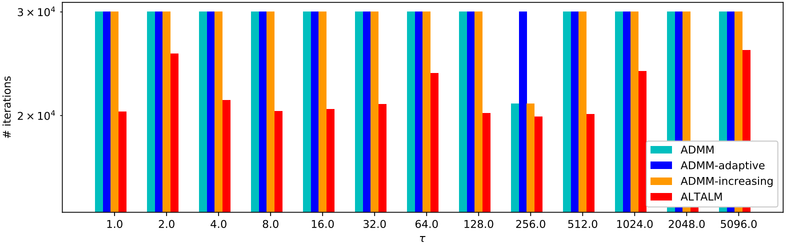
- QP solver for the $x$ step,
- closed form solution for $y$

## The augmented Lagrangian subproblem

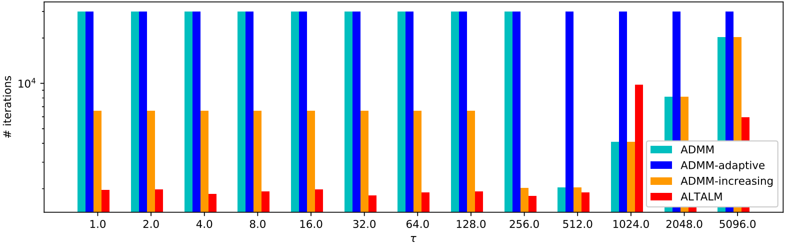
$$\min_{x,y} \frac{1}{2}x^T Qx - e^T x + \lambda^T x + \lambda^T y + \frac{\tau}{2}\|x - y\|^2$$
$$\text{s.t. } 0 \leq x_i \leq C \quad \forall\, i \in 1, \dots, n$$
$$a^T y = 0.$$

- QP solver for the $x$ step,
- closed form solution for $y$

UCI `liver-disorder` dataset (345 examples, 7 attributes)




**(a)** Sigmoidal kernel.




**(b)** Gaussian mixture kernel.

# Wrap up

- "same" 3 fundamental steps,
- different convergence proofs,
- ALTALM naturally deals with the non convex setting,
- different assumptions on and role of $\tau$,
- ALTALM has a double loop (weakness?),
- ALTALM compares favorably in the numerical experiments.

**Thank you!**