# The Proximal Alternating Minimization Algorithm

Gert Wanka [1]
joint work with Sandy Bitterlich [1],
Radu Ioan Boț [2] and Ernö Robert Csetnek [2]

[1]Chemnitz University of Technology

[2]University of Vienna

Games, Dynamics and Optimization (GDO 2019)
Babeș Bolyai University Cluj-Napoca, April 9-11, 2019

# Table of contents

# Motivation

Figure: Blurred and noisy image

- For deblurring and denoising of an image we consider the nonsmooth optimization problem:

$$\inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \lambda \mathrm{TV}(x) \right\},$$

where $A \in \mathbb{R}^{n \times n}$ is a blur operator, $b \in \mathbb{R}^n$ is the given blurred and noisy image, $\lambda > 0$ is a regularization parameter and $\mathrm{TV} : \mathbb{R}^n \to \mathbb{R}$ is a discrete total variation functional.

# Motivation

Figure: Blurred and noisy image

Figure: Solution of the problem

- For deblurring and denoising of an image we consider the nonsmooth optimization problem:

$$\inf_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \lambda \mathrm{TV}(x) \right\},$$

where $A \in \mathbb{R}^{n \times n}$ is a blur operator, $b \in \mathbb{R}^n$ is the given blurred and noisy image, $\lambda > 0$ is a regularization parameter and $\mathrm{TV} : \mathbb{R}^n \to \mathbb{R}$ is a discrete total variation functional.

# Definitions (1)

- Let $\mathcal{H}$ be a Hilbert space. Then we define

  $\Gamma(\mathcal{H}) = \{f : \mathcal{H} \to \overline{\mathbb{R}} : f$ is proper, convex and lower semicontinuous$\}$.

- Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ and $\gamma > 0$. We call $f$ strongly convex with modulus $\gamma$ if for all $x, y \in \mathcal{H}$ and $t \in [0, 1]$ holds

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{1}{2}\gamma t(1-t)\|x - y\|^2$$

- Let $f \in \Gamma((H))$ and $\sigma > 0$. Then the Proximal Point Operator of $f$ is defined as:

$$\text{Prox}_{\sigma f}(x) = \underset{y \in \mathcal{H}}{\text{argmin}} \left\{ \sigma f(y) + \frac{1}{2}\|y - x\|^2 \right\}.$$

# Definitions (2)

- We set

$$S_+(\mathcal{H}) = \{M : \mathcal{H} \to \mathcal{H} : M \text{ is linear, continuous, self-adjoint and positive semidefinite}\}.$$

- For $M \in S_+(\mathcal{H})$ we define the semi-norm $\|x\|_M^2 = \langle x, Mx\rangle \; \forall x \in \mathcal{H}$.
- We denote for $M_1, M_2 \in \mathcal{S}_+(\mathcal{H})$ the Loewner partial ordering by

$$M_1 \succcurlyeq M_2 \Leftrightarrow \|x\|_{M_1}^2 \geq \|x\|_{M_2}^2 \; \forall x \in \mathcal{H}.$$

- Furthermore, we define for $\alpha > 0$

$$\mathcal{P}_\alpha(\mathcal{H}) = \{M \in \mathcal{S}_+(\mathcal{H}) : M \succcurlyeq \alpha \mathrm{Id}\}.$$

- Let $A : \mathcal{H} \to \mathcal{G}$ be a linear continuous operator. The operator $A^* : \mathcal{G} \to \mathcal{H}$, fulfilling

$$\langle A^* y, x\rangle = \langle y, Ax\rangle$$

for all $x \in \mathcal{H}$ and $y \in \mathcal{G}$, denotes the adjoint operator of $A$, while $\|A\| := \sup\{\|Ax\| : \|x\| \leq 1\}$ denotes the norm of $A$.

# AMA-Alternating Minimization Algorithm (1)

- Consider the following convex minimization problem

$$\min \; f(x) + g(z),$$
$$\text{s.t. } Ax + Bz = b, x \in \mathbb{R}^n, z \in \mathbb{R}^m$$

- where $f \in \Gamma(\mathbb{R}^n)$ is $\gamma$-strongly convex and $g \in \Gamma(\mathbb{R}^m)$ , $A \in \mathbb{R}^{r \times n}, B \in \mathbb{R}^{r \times m}$ are linear operator and $b \in \mathbb{R}^r$.
- We have the following Lagrangian for this optimization problem

$$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \overline{\mathbb{R}}$$
$$L(x, z, p) = f(x) + g(z) + \langle p, b - Ax - Bz \rangle.$$

# AMA-Alternating Minimization Algorithm (2)

## Algorithm

Choose $p^0 \in \mathbb{R}^r$ and a sequence of stepsizes $(c_k)_{k\geq 0} \subseteq (0,+\infty)$.

$$\forall k \geq 0 \begin{cases} x^k & := \operatorname{argmin}_{x\in\mathbb{R}^n}\{f(x) - \langle p^k, Ax\rangle\}, \\ z^k & \in \operatorname{argmin}_{z\in\mathbb{R}^m}\{g(z) - \langle p^k, Bz\rangle \\ & \quad + \frac{c_k}{2}\|Ax^k + Bz - b\|^2\}, \\ p^{k+1} & := p^k + c_k(b - Ax^k - Bz^k). \end{cases}$$

# Convergence

## Theorem (Tseng,1991)

*Let* $A \neq 0$ *and* $(x, z) \in \text{ri}(\text{dom}\, f) \times \text{ri}(\text{dom}\, g)$ *be such that* $Ax + Bz = b$. *Assume that the sequence of stepsizes* $(c_k)_{k \geq 0}$ *satisfies*

$$\epsilon \leq c_k \leq \frac{2\gamma}{\|A\|^2} - \epsilon \ \forall k \geq 0,$$

*where* $\epsilon \in \left(0, \frac{\gamma}{\|A\|^2}\right)$. *Let* $(x^k, z^k, p^k)_{k \geq 0}$ *be the sequence generated by the algorithm above. Then there exist* $x^* \in \mathbb{R}^n$ *and an optimal Lagrange multiplier* $p^* \in \mathbb{R}^r$ *associated with the constraint* $Ax + Bz = b$ *such that*

$$x^k \to x^*, Bz^k \to b - Ax^*, p^k \to p^*(k \to +\infty).$$

*If the function* $z \mapsto g(z) + \|Bz\|^2$ *has bounded level sets, then* $(z^k)_{k \geq 0}$ *is bounded and any of its cluster points* $z^*$ *provides with* $(x^*, z^*)$ *an optimal solution of the problem above.*

# Problem Formulation (1)

- Let $\mathcal{H}$, $\mathcal{G}$ and $\mathcal{K}$ be real Hilbert spaces. Consider the following convex minimization problem

$$\min\{f(x) + g(z) + h_1(x) + h_2(z)\}$$
$$\text{s.t. } Ax + Bz = b$$

- where $f \in \Gamma(\mathcal{H})$ $\gamma$-strongly convex and $g \in \Gamma(\mathcal{G})$ , $h_1 : \mathcal{H} \to \mathbb{R}$ and $h_2 : \mathcal{G} \to \mathbb{R}$ convex and Fréchet differentiable functions with $L_1$- and $L_2$-Lipschitz continuous gradients $(L_1, L_2 \geq 0)$,
  $A : \mathcal{H} \to \mathcal{K}$ and $B : \mathcal{G} \to \mathcal{K}$ linear continuous operators such that $A \neq 0$ and $b \in \mathcal{K}$.
- There exists $x \in \text{ri}(\text{dom}(f))$ and $z \in \text{ri}(\text{dom}(g))$ satisfying $Ax + Bz = b$.

# Problem Formulation (2)

- We have the following Lagrangian for this optimization problem

$$L : \mathcal{H} \times \mathcal{G} \times \mathcal{K} \to \overline{\mathbb{R}}$$
$$L(x, z, p) = f(x) + g(z) + h_1(x) + h_2(z) + \langle p, b - Ax - Bz \rangle.$$

- We say that $(x^*, z^*, p^*) \in \mathcal{H} \times \mathcal{G} \times \mathcal{K}$ is a saddle point of the Lagrangian $L$, if

$$L(x^*, z^*, p) \leq L(x^*, z^*, p^*) \leq L(x, z, p^*)$$

holds for all $(x, z, p) \in \mathcal{H} \times \mathcal{G} \times \mathcal{K}$.

# Proximal AMA

## Algorithm

Let $(M_1^k)_{k\geq 0} \subseteq \mathcal{S}_+(\mathcal{H})$ and $(M_2^k)_{k\geq 0} \subseteq \mathcal{S}_+(\mathcal{G})$. Choose $(x^0, z^0, p^0) \in \mathcal{H} \times \mathcal{G} \times \mathcal{K}$ and a sequence of stepsizes $(c_k)_{k\geq 0} \subseteq (0, +\infty)$.

$$\forall k \geq 1 \begin{cases} x^{k+1} & := \operatorname{argmin}_{x\in\mathcal{H}} \Big\{ f(x) - \langle p^k, Ax\rangle + \langle x - x^k, \nabla h_1(x^k)\rangle \\ & \qquad + \frac{1}{2}\|x - x^k\|^2_{M_1^k} \Big\}, \\ z^{k+1} & \in \operatorname{argmin}_{z\in\mathcal{G}} \Big\{ g(z) - \langle p^k, Bz\rangle + \frac{c_k}{2}\|Ax^{k+1} + Bz - b\|^2 \\ & \qquad + \langle z - z^k, \nabla h_2(z^k)\rangle + \frac{1}{2}\|z - z^k\|^2_{M_2^k} \Big\}, \\ p^{k+1} & := p^k + c_k(b - Ax^{k+1} - Bz^{k+1}). \end{cases}$$

- The sequence $(z^k)_{k\geq 0}$ is uniquely determined if there exists $\alpha_k > 0$ such that $c_k B^* B + M_2^k \in \mathcal{P}_{\alpha_k}(\mathcal{G})$ for all $k \geq 0$.
- For $M_2^k := \frac{1}{\sigma_k}\mathrm{Id} - c_k B^* B$ with $\sigma_k > 0$ and $\sigma_k c_k \|B\|^2 \leq 1$ the update of $z^{k+1}$ is a proximal step.

# Convergence

## Theorem

*Let the set of saddle points of the Lagrangian L be nonempty and* $M_1^k - \frac{L_1}{2} Id \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}, M_2^k - \frac{L_2}{2} Id \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$ *for all* $k \geq 0$*. Assume that the sequence* $(x^k, z^k, p^k)_{k\geq 0}$ *is generated by the Algorithm above and* $(c_k)_{k\geq 0}$ *is monotonically decreasing satisfying:*

$$\epsilon \leq c_k \leq \frac{2\gamma}{\|A\|^2} - \epsilon, \quad \forall k \geq 0,$$

*where* $\epsilon \in (0, \frac{\gamma}{\|A\|^2})$*. If one of the following assumptions hold true:*

- *there exists* $\alpha > 0$ *such that* $M_2^k - \frac{L_2}{2} Id \in \mathcal{P}_\alpha(\mathcal{G})$ *for all* $k \geq 0$*;*
- *there exists* $\beta > 0$ *such that* $B^* B \in \mathcal{P}_\beta(\mathcal{G})$*;*

*then* $(x^k, z^k, p^k)_{k\geq 0}$ *converges weakly to a saddle point of the Lagrangian L.*

# Image deblurring and denoising

- For deblurring and denoising of an image we consider the nonsmooth optimization problem:

$$\inf_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|^2+\lambda\text{TV}(x)\right\},$$

  where $A \in \mathbb{R}^{n\times n}$ is a blur operator, $b \in \mathbb{R}^n$ is the given blurred and noisy image, $\lambda > 0$ is a regularization parameter and $\text{TV} : \mathbb{R}^n \to \mathbb{R}$ is a discrete total variation functional.
- The vector $x \in \mathbb{R}^n$ is the vectorized image $X \in \mathbb{R}^{M\times N}$, where $n = MN$ and $x_{i,j} := X_{i,j}$ stands for the normalized value of the pixel in the $i$-th row and the $j$-th column, $1 \le i \le M, 1 \le j \le N$. For color images we have $X \in \mathbb{R}^{M\times N\times 3}$ and $n = 3MN$.

# Discrete total variation (1)

We consider the discrete *isotropic total variation* $\mathrm{TV}_{\mathrm{iso}} : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{TV}_{\mathrm{iso}}(x) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}$$

$$+ \sum_{i=1}^{M-1} |x_{i+1,N} - x_{i,N}| + \sum_{j=1}^{N-1} |x_{M,j+1} - x_{M,j}|,$$

and the discrete *anisotropic total variation* $\mathrm{TV}_{\mathrm{aniso}} : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{TV}_{\mathrm{aniso}}(x) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$$

$$+ \sum_{i=1}^{M-1} |x_{i+1,N} - x_{i,N}| + \sum_{j=1}^{N-1} |x_{M,j+1} - x_{M,j}|.$$

# Discrete total variation (2)

- We define the linear operator
  $L : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n, x_{i,j} \mapsto (L_1 x_{i,j}, L_2 x_{i,j})$, where

$$L_1 x_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j}, & \text{if } i < M \\ 0, & \text{if } i = M \end{cases} \quad \text{and}$$

$$L_2 x_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j}, & \text{if } j < N \\ 0, & \text{if } j = N \end{cases}$$

- The problem above can be written as

$$\inf_{x \in \mathbb{R}^n} \left\{ f(Ax) + g(Lx) \right\},$$

where $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \frac{1}{2}\|x - b\|^2$ and $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, where in the case of the anisotropic total variation $g(y, z) = \lambda\|(y, z)\|_1$ and in the case of the isotropic total variation $g(y, z) = \lambda\|(y, z)\|_\times := \lambda \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{y_{i,j}^2 + z_{i,j}^2}$.

# Image deblurring and denoising

- The Fenchel dual problem is given by (strong duality holds):

$$\inf_{p\in\mathbb{R}^n, q\in\mathbb{R}^n\times\mathbb{R}^n} \{f^*(p)+g^*(q)\}$$
$$\text{s.t. } A^*p+L^*q=0.$$

- As $f^*(p)=\frac{1}{2}\|p\|^2+\langle p,b\rangle$ for all $p\in\mathbb{R}^n$, $f^*$ is 1-strongly convex.
- The conjugate of $g$ is the indicator function of the set

$$[-\lambda,\lambda]^n\times[-\lambda,\lambda]^n$$

(in the anisotrpic case) or the indicator function of the set

$$S:=\left\{(v,w)\in\mathbb{R}^n\times\mathbb{R}^n : \max_{1\le i\le n}\sqrt{v_i^2+w_i^2}\le\lambda\right\}$$

(in the isotropic case).

# Proximal-AMA-Algorithm

We choose $M_1^k = 0$ and $M_2^k = \frac{1}{\sigma_k}\mathrm{I} - c_k LL^*$ for every $k \geq 0$ and obtain for Proximal AMA:

## Algorithm

Choose $x^0 \in \mathbb{R}^n$ and $(c_k)_{k\geq 0} > 0$. For all $k \geq 0$ generate the sequence $(p^k, q^k, x^k)_{k\geq 0}$ as follows:

$$\begin{aligned} p^{k+1} &= \operatorname*{argmin}_{p\in\mathbb{R}^n} \left\{ f^*(p) - \langle x^k, A^* p\rangle \right\} = Ax^k - b \\ q^{k+1} &= \mathrm{Prox}_{\sigma_k g^*}\left( q^k + \sigma_k c_k L(-A^* p^{k+1} - L^* q^k) + \sigma_k L(x^k) \right) \\ x^{k+1} &= x^k + c_k(-A^* p^{k+1} - L^* q^{k+1}). \end{aligned}$$

The proximal operator of the $q^{k+1}$-Update is a projection operator.

# AMA-Algorithm

## Algorithm

Choose $x^0 \in \mathbb{R}^n$ and $(c_k)_{k\geq 0} > 0$. For all $k \geq 0$ generate the sequence $(p^k, q^k, x^k)_{k\geq 0}$ as follows:

$$p^k = \underset{p\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ f^*(p) - \langle x^k, A^*p \rangle \right\} = Ax^k - b$$

$$q^k = \underset{q\in\mathbb{R}^m}{\operatorname{argmin}} \{ g^*(q) - \langle x^k, L^*q \rangle + \frac{c_k}{2} \|A^*p^k + L^*q\|^2 \}$$

$$x^{k+1} = x^k + c_k(-A^*p^k - L^*q^k).$$

In Proximal AMA a closed formula is available for the computation of $q^k$, in AMA we solved the resulting optimization subproblem in every iteration $k \geq 0$ by making some steps of the FISTA method.

# Image

Figure: Original image "office _ 4"

Figure: Blurred and noisy image

Figure: Image after 50s cpu-time

# Comparison Proximal-AMA and AMA

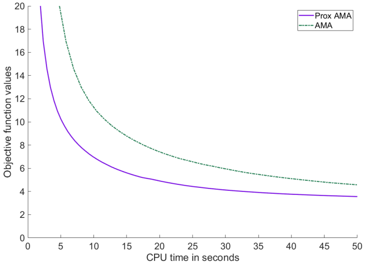


Figure: Objective function values for anisotropic TV with $\lambda = 5 \cdot 10^{-5}$

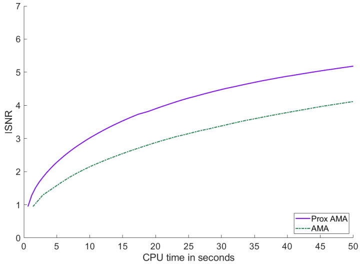


Figure: ISNR value for anisotropic TV with $\lambda = 5 \cdot 10^{-5}$

# Kernel based machine learning (1)

- For Kernel based machine learning we have a given training data set

$$\mathcal{Z} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \subseteq \mathbb{R}^d \times \{+1, -1\}.$$

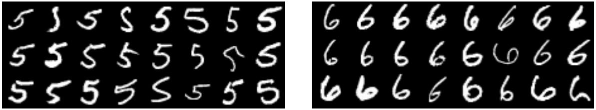

Figure: A sample of images belonging to the classes $+1$ and $-1$.

- The symmetric and finitely positive definite Gaussian kernel function is given by

$$\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \ \kappa(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right).$$

- By $K \in \mathbb{R}^{n\times n}$ we denoted the symmetric and positive definite Gram matrix with entries $K_{ij} = \kappa(X_i, X_j)$ for $i, j = 1, \ldots, n$.

# Kernel based machine learning (2)

- We consider the nonsmooth optimization problem:

$$\inf_{x \in \mathbb{R}^n} \{f(x) + g(Kx)\}$$

which is equivalent to

$$\inf_{x \in \mathbb{R}^n} \{f(x) + g(z)\}, \quad \text{s.t. } Kx - z = 0$$

where $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \frac{1}{2}x^T Kx$, $g : \mathbb{R}^n \to \mathbb{R}$ and $g(z) = C \sum_{i=1}^n \max\{1 - z_i Y_i, 0\}$ for a $C > 0$.

- So $f$ is $\lambda_{\min}(K)$-strongly convex and differentiable and $\nabla f(x) = Kx \ \ \forall x \in \mathbb{R}^n$.
- For $p \in \mathbb{R}^n$, we have

$$g^*(p) = \begin{cases} \sum_{i=1}^n p_i Y_i, & \text{if } p_i Y_i \in [-C, 0], i = 1, \dots, n \\ +\infty & \text{otherwise.} \end{cases}$$

# Proximal-AMA-Algorithm

## Algorithm

Choose $x^0 \in \mathbb{R}^n$, $z^0 \in \mathbb{R}^n$, $p^0 \in \mathbb{R}^n$, for an $\epsilon > 0$ the sequence $(c_k)_{k\geq 0} \in (\epsilon, \frac{2\lambda_{\min}(K)}{\|K\|^2} - \epsilon)$, $(M_1^k)_{k\geq 1}$ positive semidefinite and $(\sigma_k)_{k\geq 0} > 0$ such that $\sigma_k \leq \frac{1}{c_k}$. For all $k \geq 1$ generate the sequence $(p^k, q^k, x^k)_{k\geq 0}$ as follows:

$$x^{k+1} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x) - \langle p^k, Kx \rangle + \frac{1}{2}\|x - x^k\|^2_{M_1^k} \right\}$$

$$= (K + M_1^k)^{-1}(Kp^k + M_1^k x^k) \quad (1)$$

$$z^{k+1} = \operatorname{Prox}_{\sigma_k g}\left((1 - c_k\sigma_k)z^k + \sigma_k(c_k Kx^{k+1} - p^k)\right) \quad (2)$$

$$p^{k+1} = p^k + c_k(-Kx^{k+1} + z^{k+1}). \quad (3)$$

# Comparison Proximal-AMA and AMA (1)

- For $M_1^k = 0$ and $\sigma_k = \frac{1}{c_k}$ the algorithm above is the AMA-Algorithm which performs for the update of $z^k$ the proximal-step: $z^{k+1} = \text{Prox}_{\frac{1}{c_k}g}(Kx^{k+1} - \frac{1}{c^k}p^k) = \left(Kx^{k+1} - \frac{1}{c^k}p^k\right) - \frac{1}{c_k}\text{Prox}_{c_k g^*}\left(c^k Kx^{k+1} - p^k\right)$ by means of the Moreau decomposition formula for $\gamma > 0$ $\text{Prox}_{\gamma f}(x) + \gamma\text{Prox}_{(1/\gamma)f^*}(\gamma^{-1}x) = x, \quad \forall x \in \mathcal{H}$ $(= \mathbb{R}^n$ here).
- In the numerical experiments $\sigma_k = \frac{1}{c_k}$ was the best choice for the Proximal AMA algorithm, so the update of $z^{k+1}$ is the same as in the AMA algorithm.
- But the choice of $M_1^k = \tau_k K$ was for some $\tau_k > 0$ better than $M_1^k = 0$. So the update of $x^k$ for Proximal-AMA becomes $x^{k+1} = \frac{1}{1+\tau_k}(p^k + \tau_k x^k)$ instead of $x^{k+1} = p^k$ like in AMA.
- We used for both algorithms a constant sequence of stepsizes $c_k = 2 \cdot \frac{\lambda_{\min}(K)}{\|K\|^2} - 10^{-8}$ for all $k \geq 0$.

# Comparison Proximal-AMA and AMA (2)

| Algorithm | misclassification rate at 0.7027 % | RMSE $\leq 10^{-3}$ |
|---|---|---|
| Proximal AMA | 8.18s (145) | 23.44s (416) |
| AMA | 8.65s (153) | 26.64s (474) |

Table: Performance evaluation for the SVM problem using $C = 1$, $\sigma = 0.2$ (standard deviation of the gaussian kernel function) and for Proximal AMA $\tau_k = 10$. The entries refer to the CPU times in secondes and the number of iterations.

| Algorithm | misclassification rate at 0.7027 % | RMSE $\leq 10^{-3}$ |
|---|---|---|
| Proximal AMA | 141.78 s (2448) | 629.52 s (10,940) |
| AMA | 147.99 s (2574) | 652.61 s (11,368) |

Table: Performance evaluation for the SVM problem using $C = 1$, $\sigma = 0.25$ and for Proximal AMA $\tau_k = 102$. The entries refer to the CPU times in secondes and the number of iterations.

# Literature (1)


- Banert, Sebastian; Boţ, Radu Ioan; Csetnek, Ernö Robert: *Fixing and extending some recent results on the ADMM algorithm*. Preprint. arXiv:1612.05057, 2017.
- Bitterlich, Sandy; Boţ, Radu Ioan; Csetnek, Ernö Robert; Wanka, Gert: *The Proximal Alternating Minimization Algorithm for Two-Block Separable Convex Optimization Problems with Linear Constraints*. Journal of Optimization Theory and Applications. https://doi.org/10.1007/s10957-018-01454-y (to appear).
- Boţ, Radu Ioan; Csetnek, Ernö Robert; Heinrich, Andre; Hendrich, Christopher: *On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems*. Mathematical Programming 150(2), 251-279, 2015.
- Hendrich, Christopher: *Proximal Splitting Methods in Nonsmooth Convex Optimization*. Dissertation, TU-Chemnitz, Fakultät für Mathematik, 2014.

# Literature (2)

Rudin,Leonid I.; Osher, Stanley and Fatemi, Emad: *Nonlinear total-variation-based noise removal algorithms*. Physica D: Nonlinear Phenom. 60 (1–4), 259–268, 1992.

Tseng, Paul: *Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities*. SIAM J. Control Optimization 29(1), 119–138, 1991.