

# Linear Convergence Rates for Variants of the Alternating Direction Method of Multipliers in Smooth Cases

Pauline Tan<sup>1</sup> 

Received: 23 January 2017 / Accepted: 15 December 2017 / Published online: 20 December 2017  
© Springer Science+Business Media, LLC, part of Springer Nature 2017

**Abstract** In the present paper, we propose a novel convergence analysis of the alternating direction method of multipliers, based on its equivalence with the overrelaxed primal–dual hybrid gradient algorithm. We consider the smooth case, where the objective function can be decomposed into one differentiable with Lipschitz continuous gradient part and one strongly convex part. Under these hypotheses, a convergence proof with an optimal parameter choice is given for the primal–dual method, which leads to convergence results for the alternating direction method of multipliers. An accelerated variant of the latter, based on a parameter relaxation, is also proposed, which is shown to converge linearly with same asymptotic rate as the primal–dual algorithm.

**Keywords** Alternating direction method of multipliers · Primal–dual algorithm · Strong convexity · Linear convergence rate

**Mathematics Subject Classification** 65K10 · 49J52 · 49K35

## 1 Introduction

The alternating direction method of multipliers (ADMM) was initially introduced by Gabay–Mercier [1] and Glowinski–Marrocco [2] for solving convex composite

---

Communicated by Jalal M. Fadili.

✉ Pauline Tan  
pauline.tan@cmla.ens-cachan.fr

<sup>1</sup> Centre de Mathématiques Appliquées, CNRS, École polytechnique, Université Paris Saclay, Palaiseau, France

problems. This method has been intensively studied in the past years. One may see for instance a comprehensive review in [3]. The point of main interest is the convergence of ADMM and its convergence rate. Under some assumptions, linear rates can be achieved [4].

The ADMM is linked to another famous algorithm, which is known as the primal–dual hybrid gradient (PDHG) method [5]. This method can be accelerated [6–9], thanks to an overrelaxation step by Nesterov [10], which leads to the overrelaxed PDHG (OPDHG) method. This algorithm is also shown to converge linearly under some regularity assumptions [7].

Other linear convergence results for variants of the ADMM can be found in the literature. Generally speaking, they differ from our result on the hypotheses made on the problem (both on the regularity of the objective function and on the operators). In [11], the authors propose to add an overrelaxation step in the spirit of Nesterov’s acceleration. They show linear convergence rate for special case requiring strong convexity, Lipschitz continuity of the gradient, invertibility and full column-rank hypotheses. In [12], the authors study the ADMM in a wider framework, by allowing in each partial minimization to add an extra proximal term, which leads to a generalized ADMM. Linear convergence rates are proved for four scenarios, which require both strong convexity and smoothness, but an explicit convergence rate is given for only one scenario [12, Corollary 3.6]. It can be shown that the ADMM iterations are also equivalent to applying the Douglas–Rachford splitting (DRS) to the dual problem [1]. A relaxed version of the DRS, called relaxed Peaceman–Rachford splitting, thus leads to a relaxed version of the ADMM [13]. In [13, Theorem 6.3], the authors prove the linear convergence rate of this variant of the ADMM in various cases (including the one we studied here), which depends on the assumptions made on the problem regularity. However, the study is theoretical and does not provide explicit optimal convergence rates.

The OPDHG has also been subject of numerous studies. A convergence analysis of the method is investigated in [7] under many different assumptions on the problem. A key point is the choice of the method parameters, which can affect the convergence rate [14].

In this paper, we provide a new analysis of the ADMM based on the known equivalence between the ADMM and the OPDHG method. More specifically, we use this analysis to derive convergence rate for the ADMM in a case we refer to be *smooth*. We indeed made restrictive assumptions on the linear coupling constraint and on the objective function, which is supposed to be strongly convex, with a smooth part. We first establish new linear convergence rates of the OPDHG method, by generalizing the proofs of [7,9]. This leads to a linear convergence rate for the ADMM and its relaxed variant. The latter is then shown to achieve faster convergence rates.

The present paper is organized as follows. In Sect. 2, we define what we call the *smooth case*, which will be considered throughout this paper. In Sect. 3, we establish the linear convergence of the OPDHG method, and we provide the best parameter choice. In Sect. 4, we take advantage of the equivalence between the ADMM and the OPDHG to derive new linear convergence rates for the ADMM. We also propose a relaxed variant of the ADMM, which achieves the same asymptotic convergence rate as the OPDHG method, if the parameters are properly chosen. Eventually, in Sect. 5,

we applied our relaxed ADMM on two problems, and compared the results with the classical ADMM, the OPDHG and FISTA [15].

## 2 The Problem and Some Preliminaries

Let  $X$  and  $Y$  be two finite-dimensional real Hilbert spaces. The inner product is denoted by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$  is the induced norm.

### 2.1 Strong Convexity and Duality

Here are some facts about strong convexity that will be used in this paper.

**Definition 2.1** A function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  is said to be *strongly convex* of parameter  $\alpha > 0$  or  $\alpha$ -convex iff for any  $(x_1, x_2) \in \text{dom}(f)^2$  and  $p \in \partial f(x_1)$ ,

$$f(x_2) \geq f(x_1) + \langle p, x_2 - x_1 \rangle + \frac{\alpha}{2} \|x_2 - x_1\|^2. \tag{1}$$

where  $\text{dom}(f)$  is the domain of  $f$ , i.e., the set of  $x \in X$  for which  $f(x) < +\infty$ .

**Proposition 2.1** [16] *Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be an  $\alpha$ -convex function. Then its convex conjugate  $f^*$ , defined for any  $y \in X$  by*

$$f^*(y) = \sup_{x \in X} \{ \langle x, y \rangle - f(x) \} \tag{2}$$

*is differentiable, and its gradient  $\nabla f^*$  is Lipschitz continuous of constant  $1/\alpha$ .*

### 2.2 Smooth Composite Problem

We consider the composite problem

$$\min_{x \in X} \{ f(x) := g(x) + h(Ax) \}, \tag{3}$$

under the following assumptions:

**Assumption (S)** The convex objective function  $f$  in (3) is supposed to satisfy the following *smooth case* condition:

- (a)  $g : X \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper,<sup>1</sup> **strongly convex** of parameter  $\gamma$ , and lower semi-continuous (l.s.c.) function.
- (b)  $h : Y \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, convex, and a continuously differentiable function, and its gradient  $\nabla h$  is **Lipschitz continuous** of constant  $1/\delta$ .
- (c)  $A : X \rightarrow Y$  is a **bounded linear operator**, of norm  $L_A$  and of adjoint  $A^*$ .

---

<sup>1</sup> that is,  $g$  is not identically equal to  $+\infty$

Let us briefly discuss the assumptions made above. Though they may seem to be very restrictive, they can be satisfied in many applications. Assumption **(S)**-(a) is fulfilled for instance with the least squares error which is widely used as a fidelity term. Assumption **(S)**-(b) is a typical assumption for many convex optimization algorithms, including the forward–backward algorithm and its variants. When  $h(Ax)$  stands for the regularity term, **(S)**-(b) can be satisfied by considering smooth approximations of classical regularizations (e.g., Huber regularization). Assumption **(S)**-(c) is also fulfilled in many applications, since it is typically a gradient operator, or a degradation kernel.

*Remark 2.1* Problem (3) is a particular instance of the general problem

$$\min_{(x,z) \in X \times Z: Ax+Bz=c} g(x) + h(z) \quad (4)$$

with  $Z = Y$ ,  $B = -\text{Id}$  and  $c = 0$ . For the sake of simplicity, and since problems of standard form (3) arise in many contexts, we focus on problems of this form. However, the interested reader will easily extend our result to the case, when  $B$  has full row rank (see Remark 4.3).

*Remark 2.2* According to Assumption **(S)**,  $h \circ A$  is Lipschitz continuous, of constant  $L_A^2/\delta$ . Hence, one can define  $\kappa_f := L_A^2/(\delta\gamma)$  the *condition number* of  $f$  as the ratio between the Lipschitz constant of the smooth part and the strong convexity parameter of the nonsmooth part  $g$ . When  $f$  is both strongly convex and smooth with  $\nabla f$  Lipschitz continuous, this definition recovers the usual definition, and  $\kappa_f$  is always larger than 1 [17]. In the general case, it can be less than 1. When  $\kappa_f$  is large, the function  $f$  is said ill-conditioned.

### 3 New Convergence Analysis for OPDHG in the Smooth Case

As shown in Sect. 4, our convergence analysis of the ADMM and its variants rely on an equivalence with the OPDHG algorithm. Thus, let us introduce this method and state some convergence results.

#### 3.1 Primal–Dual Problem and the OPDHG Algorithm

Let us consider the following saddle-point problem:

$$\min_{\xi \in Z} \sup_{y \in Y} \left\{ \mathcal{L}(\xi; y) := G(\xi) + \langle K\xi, y \rangle - H^*(y) \right\} \quad (5)$$

where the convex–concave function  $\mathcal{L}$  satisfies the following assumption:

**Assumption (S2)** Let  $Z$  and  $Y$  be two finite-dimensional real Hilbert spaces.

- (a)  $G : Z \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper,  $\tilde{\gamma}$ -convex, and l.s.c. function.
- (b)  $H^* : Y \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper,  $\tilde{\delta}$ -convex, and l.s.c. function.

(c)  $K : Z \rightarrow Y$  is a bounded linear operator of norm  $L_K$ .

We recall that, for any convex function  $F : X \mapsto \mathbb{R} \cup \{+\infty\}$ , one can define the proximal operator of  $F$  [18] as

$$\forall x_0 \in X, \quad \text{prox}_F(x_0) := \arg \min_{x \in X} \left\{ F(x) + \frac{1}{2} \|x - x_0\|^2 \right\} \tag{6}$$

which is uniquely defined for any  $x_0 \in X$  as the minimizer of a strongly convex problem. In this section, we establish the convergence proof of the following primal–dual algorithm

**Algorithm 1 (OPDHG with overrelaxation on the primal variable)**

Initialization:  $(y_0, \xi_0) \in Y \times Z$ ,  $\bar{\xi}_0 = \xi_0$  and  $0 < \tau, \sigma, 0 < \theta \leq 1$   
 General Step: for  $n = 0, 1, \dots$ , compute

$$y_{n+1} = \text{prox}_{\sigma H^*}(y_n + \sigma K \bar{\xi}_n) \tag{7}$$

$$\xi_{n+1} = \text{prox}_{\tau G}(\xi_n - \tau K^* y_{n+1}) \tag{8}$$

$$\bar{\xi}_{n+1} = \xi_{n+1} + \theta (\xi_{n+1} - \xi_n) \tag{9}$$

until stopping criterion is satisfied.

which aims at solving problem (5). The parameters  $\tau, \sigma > 0$  and  $0 < \theta \leq 1$  are to be specified.

*Remark 3.1* When  $\theta = 0$ , this algorithm is known as the PDHG method [5] or as the Arrow-Hurwicz algorithm [19]. It consists in a proximal gradient ascent step for the dual variable (7), followed by a proximal gradient descent step for the primal variable (8). The overrelaxation step (9) has been added in [6] and studied in [8,9]. When  $\theta = 1$  and  $\tau\sigma \neq 1/L_K^2$ , the iterations are equivalent to a preconditioned version of the ADMM [7,8].

**3.2 Convergence Analysis for OPDHG**

We set  $\mathcal{G}(\xi', \xi; y, y') := \mathcal{L}(\xi'; y) - \mathcal{L}(\xi; y')$  for any  $(\xi, \xi') \in Z^2$  and  $(y, y') \in Y^2$ . Moreover, if  $0 < \omega < 1$ , then we denote

$$T_N := \sum_{n=1}^N \frac{1}{\omega^{n-1}} = \frac{1 - \omega^N}{\omega^{N-1}(1 - \omega)}, \quad \Xi_N := \frac{1}{T_N} \sum_{n=1}^N \frac{\xi_n}{\omega^{n-1}}, \quad Y_N := \frac{1}{T_N} \sum_{n=1}^N \frac{y_n}{\omega^{n-1}}. \tag{10}$$

Let us formulate our main result for the OPDHG algorithm in the smooth case.

**Theorem 3.1** *Let  $\mathcal{L}$  satisfy Assumption (S2) and  $(y_n, \xi_n, \bar{\xi}_n)_n$  a sequence generated by Algorithm 1. Assume problem (5) has a solution, which is a saddle point of  $\mathcal{L}$ , denoted by  $(\xi^*, y^*)$ . Choose  $\tau > 0, \sigma > 0$  and  $0 < \theta \leq 1$  such that*

$$\max \left\{ \frac{1}{\tau \tilde{\gamma} + 1}, \frac{1}{\sigma \tilde{\delta} + 1} \right\} \leq \theta \leq \frac{1}{L_K^2 \tau \sigma}. \tag{11}$$

*Then, for any  $\omega < 1$  such that*

$$\max \left\{ \frac{1}{\tau \tilde{\gamma} + 1}, \frac{\theta + 1}{\sigma \tilde{\delta} + 2} \right\} \leq \omega \leq \theta \tag{12}$$

*we have the following bound for any  $(\xi, y) \in Z \times Y$  and for any  $N \geq 1$ :*

$$\begin{aligned} & \frac{1 - \omega}{\omega(1 - \omega^N)} \left( \frac{\|\xi - \xi_N\|^2}{2\tau} + (1 - \omega L_K^2 \tau \sigma) \frac{\|y - y_N\|^2}{2\sigma} \right) + \mathcal{G}(\Xi_N, \xi; y, Y_N) \\ & \leq \frac{1}{T_N} \left( \frac{\|\xi - \xi_0\|^2}{2\tau} + \frac{\|y - y_0\|^2}{2\sigma} \right). \end{aligned} \tag{13}$$

This theorem is an improvement on a result proved in [7, Theorem 3]. The rate we provide here is indeed better, since no restrictive assumptions are made on the parameters values, unless necessary. The proof of Theorem 3.1 is proposed in Appendix A. A consequence of this proof is the linear convergence of the iterates:

**Corollary 3.1** *Let  $\mathcal{L}$  satisfy Assumption (S2) and  $(y_n, \xi_n, \bar{\xi}_n)_n$  a sequence generated by Algorithm 1. Assume problem (5) has a solution, which is a saddle point of  $\mathcal{L}$ , denoted by  $(\xi^*, y^*)$ . Suppose there exist  $\tau, \sigma, \theta$  and  $\omega$  satisfying both conditions (11) and (12). Then, for any  $N \geq 1$ ,*

$$\|\xi^* - \xi_N\|^2 \leq \omega^N \left( \|\xi^* - \xi_0\|^2 + \frac{\tau}{\sigma} \|y^* - y_0\|^2 \right). \tag{14}$$

*Moreover, if  $\omega L_K^2 \tau \sigma < 1$ , then we also have*

$$\|y^* - y_N\|^2 \leq \frac{\omega^N}{1 - \omega L_K^2 \tau \sigma} \left( \frac{\sigma}{\tau} \|\xi^* - \xi_0\|^2 + \|y^* - y_0\|^2 \right). \tag{15}$$

### 3.3 Choice of Parameters and Convergence Rates

Theorem 3.1 holds, provided one can properly choose the steps  $\tau$  and  $\sigma$  and the relaxation parameter  $\theta$ . In particular, one can tune  $(\tau, \sigma, \theta)$ , so that the resulting convergence rate  $\omega$  is minimal.

**Theorem 3.2** Let  $\kappa_F := L_K^2/(\tilde{\delta}\tilde{\gamma})$ . The best convergence rate in Theorem 3.1 is achieved when the parameters are  $(\tau, \sigma, \theta) = (\tau^*, \sigma^*, \theta^*)$ , given by

$$\tau^* = \frac{1 + \sqrt{1 + 4\kappa_F}}{2\tilde{\gamma}\kappa_F}, \quad \sigma^* = \frac{1 + \sqrt{1 + 4\kappa_F}}{2\tilde{\delta}\kappa_F} \quad \text{and} \quad \theta^* = \frac{\sqrt{1 + 4\kappa_F} - 1}{\sqrt{1 + 4\kappa_F} + 1} \quad (16)$$

which satisfy  $\tau\tilde{\gamma} = \sigma\tilde{\delta}$ . The resulting convergence rate is  $\omega = \theta^*$ .

Let us first prove the following lemma, which gives conditions on  $\tau$  and  $\sigma$ , so that (11) is well-defined:

**Lemma 3.1** For any  $\tau > 0$ , we define a quantity  $\sigma_{\max}(\tau) > 0$  given by

$$\sigma_{\max}(\tau) = \begin{cases} (1/\tau + \tilde{\gamma})/L_K^2, & \text{if } 0 < \tau \leq \tau^*, \\ 1/(L_K^2\tau - \tilde{\delta}), & \text{if } \tau > \tau^*, \end{cases} \quad (17)$$

where  $\tau^*$  is given by (16). Then, for any  $(\tau, \sigma) \in ]0, +\infty[ \times ]0, \sigma_{\max}(\tau)[$ , there exists  $\theta$  satisfying (11).

*Proof* Let  $\tau > 0$ . The inequality (11) is feasible, if  $1/(\tau\tilde{\gamma} + 1) \leq 1/(L_K^2\tau\sigma)$  and  $1/(\sigma\tilde{\delta} + 1) \leq 1/(L_K^2\tau\sigma)$ , which, respectively, read

$$\sigma \leq \frac{1/\tau + \tilde{\gamma}}{L_K^2} \quad \text{and} \quad (L_K^2\tau - \tilde{\delta})\sigma \leq 1. \quad (18)$$

Let us show that these inequalities hold iff  $\sigma$  is less than  $\sigma_{\max}(\tau)$ . Because of the second inequality in (18), one should consider two cases.

1. Case  $L_K^2\tau - \tilde{\delta} \leq 0$ . In this case, the second inequality is always true and (18) reduces to its first inequality, which is  $\sigma \leq \sigma_{\max}(\tau)$ , since  $\tau \leq \tilde{\delta}/L_K^2 < \tau^*$ .
2. Case  $L_K^2\tau - \tilde{\delta} > 0$ . Then one can divide the second inequality by  $L_K^2\tau - \tilde{\delta}$  and (18) reads  $\sigma \leq \min\{(1/\tau + \tilde{\gamma})/L_K^2, 1/(L_K^2\tau - \tilde{\delta})\}$ . One can check that

$$\frac{1/\tau + \tilde{\gamma}}{L_K^2} \leq \frac{1}{L_K^2\tau - \tilde{\delta}} \iff L_K^2\tau^2 - \tilde{\delta}\tau - \frac{\tilde{\delta}}{\tilde{\gamma}} \leq 0 \iff \tau \leq \tau^*. \quad (19)$$

Thus,  $\sigma \leq \sigma_{\max}(\tau)$ . □

*Proof of Theorem 3.2* Let  $(\tau, \sigma) \in ]0, +\infty[ \times ]0, \sigma_{\max}(\tau)[$ . We denote

$$\theta_{\min}(\tau, \sigma) := \max \left\{ \frac{1}{\tau\tilde{\gamma} + 1}, \frac{1}{\sigma\tilde{\delta} + 1} \right\} \quad \text{and} \quad \theta_{\max}(\tau, \sigma) := \frac{1}{L_K^2\tau\sigma}. \quad (20)$$

One has  $\theta_{\min}(\tau, \sigma) \leq \theta_{\max}(\tau, \sigma)$  (Lemma 3.1). Let  $\theta \in [\theta_{\min}(\tau, \sigma), \theta_{\max}(\tau, \sigma)]$ . The best convergence rate that appears in Theorem 3.1 for  $(\tau, \sigma, \theta)$  is given by the lower bound of (12), namely

$$\omega(\tau, \sigma, \theta) = \max \left\{ \frac{1}{\tau\tilde{\gamma} + 1}, \frac{\theta + 1}{\sigma\tilde{\delta} + 2} \right\}. \quad (21)$$

The aim is thus to find stepsizes  $\tau^* \in ]0, +\infty[$ ,  $\sigma^* \in ]0, \sigma_{\max}(\tau^*)]$ , and parameter  $\theta^* \in [\theta_{\min}(\tau^*, \sigma^*), \theta_{\max}(\tau^*, \sigma^*)]$ , which minimize  $\omega(\tau, \sigma, \theta)$ . We first note that, for any feasible  $(\tau, \sigma, \theta)$ , one has

$$\omega(\tau, \sigma, \theta_{\min}(\tau, \sigma)) \leq \omega(\tau, \sigma, \theta). \quad (22)$$

Thus, Problem (21) is equivalent to minimizing  $\omega(\tau, \sigma, \theta_{\min}(\tau, \sigma))$  subject to  $(\tau, \sigma) \in ]0, +\infty[ \times ]0, \sigma_{\max}(\tau)]$ . Let  $\sigma_{\min}(\tau) := \tau\tilde{\gamma}/\tilde{\delta}$ . One can verify that

$$\theta_{\min}(\tau, \sigma) = \begin{cases} 1/(\sigma\tilde{\delta} + 1), & \text{if } \tau > \tau^*, \\ 1/(\sigma\tilde{\delta} + 1), & \text{if } (\tau \leq \tau^* \text{ and } 0 < \sigma \leq \sigma_{\min}(\tau)), \\ 1/(\tau\tilde{\gamma} + 1), & \text{if } (\tau \leq \tau^* \text{ and } \sigma_{\min}(\tau) < \sigma \leq \sigma_{\max}(\tau)). \end{cases} \quad (23)$$

This yields

$$\frac{\theta_{\min}(\tau, \sigma) + 1}{\sigma\tilde{\delta} + 2} = \begin{cases} 1/(\sigma\tilde{\delta} + 1), & \text{if } \tau > \tau^*, \\ 1/(\sigma\tilde{\delta} + 1), & \text{if } (\tau \leq \tau^* \text{ and } 0 < \sigma \leq \sigma_{\min}(\tau)), \\ \frac{\tau\tilde{\gamma} + 2}{(\tau\tilde{\gamma} + 1)(\sigma\tilde{\delta} + 2)}, & \text{if } (\tau \leq \tau^* \text{ and } \sigma_{\min}(\tau) < \sigma \leq \sigma_{\max}(\tau)). \end{cases} \quad (24)$$

Obviously, for any feasible  $(\tau, \sigma)$ , one has

$$\frac{\theta_{\min}(\tau, \sigma_{\max}(\tau)) + 1}{\sigma_{\max}(\tau)\tilde{\delta} + 2} \leq \frac{\theta_{\min}(\tau, \sigma) + 1}{\sigma\tilde{\delta} + 2}. \quad (25)$$

As a consequence,  $\omega(\tau, \sigma_{\max}(\tau), \theta_{\min}(\tau, \sigma_{\max}(\tau))) \leq \omega(\tau, \sigma, \theta_{\min}(\tau, \sigma))$ . Thus, Problem (21) reduces to minimize  $\omega^*(\tau) := \omega(\tau, \sigma_{\max}(\tau), \theta_{\min}(\tau, \sigma_{\max}(\tau)))$ . Let  $\tau > 0$ . According to (24) and the definition of  $\sigma_{\max}(\tau)$ ,

$$\frac{\theta_{\min}(\tau, \sigma_{\max}(\tau)) + 1}{\sigma_{\max}(\tau)\tilde{\delta} + 2} = \begin{cases} 1/(\sigma_{\max}(\tau)\tilde{\delta} + 1), & \text{if } \tau > \tau^*, \\ \frac{\tau\tilde{\gamma} + 2}{(\tau\tilde{\gamma} + 1)(\sigma_{\max}(\tau)\tilde{\delta} + 2)}, & \text{if } \tau \leq \tau^*. \end{cases} \quad (26)$$

It can be deduced from (23) that  $\tau\tilde{\gamma} > \sigma_{\max}(\tau)\tilde{\delta}$  iff  $\tau > \tau^*$ . This proves that

$$\omega^*(\tau) = \begin{cases} 1 - \tilde{\delta}/(L_K^2 \tau), & \text{if } \tau > \tau^*, \\ 1/(\tau\tilde{\gamma} + 1), & \text{if } \tau \leq \tau^*, \end{cases} \quad (27)$$



which is minimal for  $\tau = \tau^*$ . This leads to the rate  $\omega = \omega^*(\tau^*)$ , which is attained for  $\tau = \tau^*, \sigma = \sigma_{\max}(\tau^*) = \sigma^*$  and  $\theta = \theta_{\min}(\tau^*, \sigma^*) = \theta^*$ . □

### 3.4 Overrelaxation on the Dual Variable

Thanks to the symmetry of (5), similar results still hold, if the relaxation is done on the dual variable  $y$ , which leads to Algorithm 2.

**Algorithm 2 (OPDHG with overrelaxation on the dual variable)**

Initialization:  $(y_0, \xi_0) \in Y \times Z, \bar{y}_0 = y_0$  and  $0 < \tau, \sigma, 0 < \theta \leq 1$   
 General Step: for  $n = 0, 1, \dots$ , compute

$$\xi_{n+1} = \text{prox}_{\tau G}(\xi_n - \tau K^* \bar{y}_n) \tag{28}$$

$$y_{n+1} = \text{prox}_{\sigma H^*}(y_n + \sigma K \xi_{n+1}) \tag{29}$$

$$\bar{y}_{n+1} = y_{n+1} + \theta (y_{n+1} - y_n) \tag{30}$$

until stopping criterion is satisfied.

As it will be seen later on, such an overrelaxation will be useful for the analysis of the ADMM. Algorithm 2 is obtained by inverting the role of the dual and the primal variables in the saddle-point problem (5). Hence, applying Theorem 3.1 yields the following result:

**Theorem 3.3** *Let  $\mathcal{L}$  satisfy Assumption (S2) and  $(\xi_n, y_n, \bar{y}_n)_n$  a sequence generated by Algorithm 2. Assume problem (5) has a solution, which is a saddle point of  $\mathcal{L}$ , denoted by  $(\xi^*, y^*)$ . Choose  $\tau > 0, \sigma > 0$  and  $0 < \theta \leq 1$  such that (11) holds. Then, for any  $\omega < 1$  such that*

$$\max \left\{ \frac{\theta + 1}{\tau \bar{\gamma} + 2}, \frac{1}{\sigma \bar{\delta} + 1} \right\} \leq \omega \leq \theta, \tag{31}$$

we have the following bound for any  $(\xi, y) \in Z \times Y$  and for any  $N \geq 1$ :

$$\begin{aligned} & \frac{1 - \omega}{\omega(1 - \omega^N)} \left( (1 - \omega L_K^2 \tau \sigma) \frac{\|\xi - \xi_N\|^2}{2\tau} + \frac{\|y - y_N\|^2}{2\sigma} \right) + \mathcal{G}(\Xi_N, \xi; y, Y_N) \\ & \leq \frac{1}{T_N} \left( \frac{\|\xi - \xi_0\|^2}{2\tau} + \frac{\|y - y_0\|^2}{2\sigma} \right). \end{aligned} \tag{32}$$

Note that the conditions on the parameters (31) now slightly differ from the ones in the previous case (see (12)). A variant can be found in [20, Appendix C2]. Similar computations as in the previous section show the linear convergence of the iterates. The parameters, that minimize the convergence rate  $\omega$  appearing in (32), can also be

computed as in Sect. 3.3. In particular, the best rate can be shown to be the same as in Theorem 3.2.

#### 4 A Relaxed Variant of the ADMM in the Smooth Case

Let us now consider the initial primal problem (3), which can be interpreted as a constrained problem (see (4) in Remark 2.1). The ADMM aims at solving it by finding a saddle point of the augmented Lagrangian

$$\min_{\substack{x \in X \\ z \in Z}} \sup_{y \in Y} \left\{ L_\tau(x, z; y) := g(x) + h(z) + \langle Ax - z, y \rangle + \frac{1}{2\tau} \|Ax - z\|^2 \right\} \quad (33)$$

for  $\tau > 0$ , by alternating in partial minimizations and gradient ascent:

##### Algorithm 3 (ADMM)

Initialization:  $(y_0, z_0) \in Y \times Z$  and  $0 < \tau$

General Step: for  $n = 0, 1, \dots$ , compute

$$x_{n+1} = \arg \min_{x \in X} \left\{ g(x) + \langle Ax, y_n \rangle + \|Ax - z_n\|^2 / (2\tau) \right\} \quad (34)$$

$$z_{n+1} = \arg \min_{z \in Y} \left\{ h(z) - \langle z, y_n \rangle + \|Ax_{n+1} - z\|^2 / (2\tau) \right\} \quad (35)$$

$$y_{n+1} = y_n + (Ax_{n+1} - z_{n+1}) / \tau \quad (36)$$

until stopping criterion is satisfied.

*Remark 4.1* If  $A$  is the identity operator, the minimization steps in Algorithm 3 can be interpreted as proximal gradient descent steps for the ordinary Lagrangian  $L_\infty(x, z; y) := g(x) + h(z) + \langle x - z, y \rangle$  associated with Problem (3). We recall that both the Lagrangian  $L_\infty$  and the augmented Lagrangian  $L_\tau$ ,  $\tau > 0$  have same saddle points [1, Theorem 2.2].

#### 4.1 Relaxed Variant of the ADMM

We propose to relax the choice of step  $\tau$  in the updates of  $z$  and of  $y$  in Algorithm 3. Replacing  $\tau$  by  $\tau' \leq \tau$  in these two updates leads to:

**Algorithm 4 (Relaxed ADMM)**

Initialization:  $(y_0, z_0) \in Y \times Z$  and  $0 < \tau' \leq \tau$   
 General Step: for  $n = 0, 1, \dots$ , compute

$$x_{n+1} = \arg \min_{x \in X} \left\{ g(x) + \langle Ax, y_n \rangle + \|Ax - z_n\|^2 / (2\tau) \right\} \tag{37}$$

$$z_{n+1} = \arg \min_{z \in Y} \left\{ h(z) - \langle z, y_n \rangle + \|Ax_{n+1} - z\|^2 / (2\tau') \right\} \tag{38}$$

$$y_{n+1} = y_n + (Kx_{n+1} - z_{n+1}) / \tau' \tag{39}$$

until stopping criterion is satisfied.

One recovers the ADMM (Algorithm 3), when  $\tau = \tau'$ .

**4.2 Equivalence Between the Relaxed ADMM and OPDHG**

We show in Appendix B that, for  $n \geq 2$ , the iterations in Algorithm 4 are equivalent to those of OPDHG with overrelaxation on the dual variable (Algorithm 2), with overrelaxation parameter  $\theta = \tau' / \tau$  and stepsize  $(\tau, \sigma = 1 / \tau')$  applied to

$$\min_{\xi \in Y} \sup_{y \in Y} \left\{ \mathcal{L}(\xi; y) := g_A(\xi) + \langle \xi, y \rangle - h^*(y) \right\}, \tag{40}$$

namely

$$\begin{aligned} \xi_{n+1} &= \text{prox}_{\tau g_A}(\xi_n - \tau \bar{y}^n) \\ y_{n+1} &= \text{prox}_{h^* / \tau'}(y_n + \xi_{n+1} / \tau') \\ \bar{y}_{n+1} &= y_{n+1} + (\tau' / \tau)(y_{n+1} - y_n) \end{aligned} \tag{41}$$

with  $g_A(\xi) := \inf\{g(x) : Ax = \xi\}$  and  $\xi_{n+1} := Ax_{n+1}$ .

*Remark 4.2* The equivalence between the ADMM ( $\tau = \tau'$ ) and OPDHG has already been investigated in [8].

*Remark 4.3* The equivalence between the relaxed ADMM and the OPDHG iterations as demonstrated in Appendix B still holds, when  $B \neq -\text{Id}$  has full row rank. Indeed, one can check that Appendix B remains true, if  $h^* \circ B^*$  is strongly convex, so that  $h_{-B} : \nu \mapsto \inf\{h(z) : -Bz = \nu\}$  has a Lipschitz continuous gradient, which is the case if  $B$  has full row rank, see [21].

Because of the equivalence between the relaxed ADMM and OPDHG recalled above, the convergence analysis of the relaxed ADMM and the classical ADMM may be investigated by applying Theorem 3.3. Let us first check that the equivalent primal–dual problem (40) satisfies Assumption (S2), when the problem (3) satisfies

**Assumption (S).** According to Assumption (S),  $h^*$  is  $\delta$ -convex. Let us prove that, if  $g$  is  $\gamma$ -convex, then  $g_A$  is  $(\gamma/L_A^2)$ -convex. Indeed,

$$g_A^*(y) = \sup_{\substack{\xi \in Y \\ Ax=\xi}} \{ \langle \xi, y \rangle - g(x) \} = \sup_{x \in X} \{ \langle Ax, y \rangle - g(x) \} = g^*(A^*y) \quad (42)$$

where  $\nabla g^*$  is  $(1/\gamma)$ -Lipschitz continuous according to Proposition 2.1. By composition,  $g_A^*$  is differentiable and  $\nabla g_A^*$  is  $(L_A^2/\gamma)$ -Lipschitz continuous. Thus,  $g_A$  is  $(\gamma/L_A^2)$ -convex (Proposition 2.1). Hence,  $\mathcal{L}$  in (40) satisfies Assumption (S2). Then, if we denote for any  $\omega > 0$  and for any  $N > 1$

$$X_N := \frac{1}{T_N} \sum_{n=1}^N \frac{x_n}{\omega^{n-1}} \quad (43)$$

applying Theorem 3.3 yields

**Proposition 4.1** *Let  $f$  satisfy Assumption (S) and  $(x_n, y_n, z_n)_n$  a sequence generated by Algorithm 4. Assume problem (3) has a solution denoted by  $x^*$ . Choose  $\tau \geq \tau' > 0$  such that*

$$\max \left\{ \frac{1}{\tau\gamma/L_A^2 + 1}, \frac{1}{\delta/\tau' + 1} \right\} \leq \frac{\tau'}{\tau} \leq 1. \quad (44)$$

*Then, for any  $\omega < 1$  such that*

$$\max \left\{ \frac{\tau'/\tau + 1}{\tau\gamma/L_A^2 + 2}, \frac{1}{\delta/\tau' + 1} \right\} \leq \omega \leq \frac{\tau'}{\tau}, \quad (45)$$

*we have the following bound for any  $(x, y) \in X \times Y$  and for any  $N \geq 2$ :*

$$\begin{aligned} & \frac{1 - \omega}{\omega(1 - \omega^N)} (1 - \omega\tau/\tau') \frac{\|Ax - Ax_N\|^2}{2\tau} + \frac{1 - \omega}{\omega(1 - \omega^N)} \frac{\|y - y_N\|^2}{2/\tau'} \\ & + \mathcal{G}(AX_N, Ax; y, Y_N) \leq \frac{1}{T_N} \left( \frac{\|Ax - Ax_1\|^2}{2\tau} + \frac{\|y - y_1\|^2}{2/\tau'} \right). \end{aligned} \quad (46)$$

An interesting consequence of this proposition is the ergodic convergence of the relaxed ADMM in terms of objective error and solution error.

**Theorem 4.1** *Let  $f$  satisfy Assumption (S) and  $(x_n, y_n, z_n)_n$  a sequence generated by Algorithm 4. Assume problem (3) has a solution denoted by  $x^*$ . Choose  $\tau \geq \tau' > 0$  satisfying (44). Then, for any  $\omega < 1$  satisfying (45) and for any  $N \geq 1$ , one has*

$$\begin{aligned} f(X_N) - f(x^*) & \leq \frac{1}{T_N} \frac{L_A^2}{2\tau} \left( 1 + \frac{N}{T_N} \frac{2\tau}{\delta^2/\tau'} \frac{\omega}{1 - \omega\tau/\tau'} \right) \|x^* - x_1\|^2 \\ & + \frac{1}{T_N} \frac{1}{1/\tau'} \left( 1 + \frac{N}{T_N} \frac{\tau}{\delta^2/\tau'} \frac{\omega}{1 - \omega\tau/\tau'} \right) \|y^* - y_1\|^2 \end{aligned} \quad (47)$$

with  $N/T_N \rightarrow 0$  as  $N$  goes to infinity. Moreover, one has for any  $N \geq 1$

$$\begin{aligned} \|X_N - x^*\|^2 &\leq \frac{2}{T_N} \frac{L_A^2}{\gamma\tau} \left(1 + \frac{N}{T_N} \frac{2\tau}{\delta^2/\tau'} \frac{\omega}{1 - \omega\tau/\tau'}\right) \|x^* - x_1\|^2 \\ &\quad + \frac{4}{T_N} \frac{1}{\gamma/\tau'} \left(1 + \frac{N}{T_N} \frac{\tau}{\delta^2/\tau'} \frac{\omega}{1 - \omega\tau/\tau'}\right) \|y^* - y_1\|^2. \end{aligned} \tag{48}$$

*Proof* For any  $(x, y) \in X \times Y$ , one has  $\mathcal{L}(Ax; y) = g(x) + \langle Ax, y \rangle - h^*(y)$  and  $f(x) = \sup_{y \in Y} \mathcal{L}(Ax; y)$ . Let us apply inequality (46) to  $x = x^*$ . Since

$$f(x^*) = \mathcal{L}(Ax^*; y^*) = \sup_{y \in Y} \mathcal{L}(Ax^*; y) \geq \mathcal{L}(Ax^*; Y_N) \tag{49}$$

we get  $\mathcal{L}(AX_N; y) - f(x^*) \leq \mathcal{L}(AX_N; y) - \mathcal{L}(Ax^*; Y_N) = \mathcal{G}(AX_N, Ax^*; y, Y_N)$  for any  $y \in Y$ , which implies

$$\begin{aligned} &\frac{1 - \omega}{\omega(1 - \omega^N)} (1 - \omega\tau/\tau') \frac{\|Ax^* - AX_N\|^2}{2\tau} + \frac{1 - \omega}{\omega(1 - \omega^N)} \frac{\|y - y_N\|^2}{2/\tau'} \\ &\quad + \mathcal{L}(AX_N; y) - f(x^*) \leq \frac{1}{T_N} \left( \frac{\|Ax^* - Ax_1\|^2}{2\tau} + \frac{\|y - y_1\|^2}{2/\tau'} \right). \end{aligned} \tag{50}$$

Let us define  $y_N^* \in Y$  as the maximizer of  $\mathcal{L}(AX_N; \cdot)$  so that  $\mathcal{L}(AX_N; y_N^*) = f(X_N)$ . The left-hand side in (50) is then nonnegative for  $y = y_N^*$  and yields

$$0 \leq f(X_N) - f(x^*) \leq \frac{1}{T_N} \frac{1}{2\tau} \|Ax^* - Ax_1\|^2 + \frac{1}{T_N} \frac{1}{2/\tau'} \|y_N^* - y_1\|^2. \tag{51}$$

Let us bound that the quantity  $\|y_N^* - y_1\|$ . One can check by first-order optimality condition that  $y_N^* = \nabla h(AX_N)$  and  $y^* = \nabla h(Ax^*)$ . Thanks to the Lipschitz continuity of  $\nabla h$  (according to Assumption (S)), this implies that  $\|y_N^* - y_1\| \leq \|Ax^* - AX_N\|/\delta + \|y^* - y_1\|$ . Using that  $(a + b)^2 \leq 2a^2 + 2b^2$  for any  $(a, b) \in \mathbb{R}^2$ , one gets

$$\|y_N^* - y_1\|^2 \leq 2 \frac{\|Ax^* - AX_N\|^2}{\delta^2} + 2 \|y^* - y_1\|^2 \tag{52}$$

However, applying (50) to  $y = y^*$  and  $N = n$  yields after simplification

$$\|Ax^* - Ax_n\|^2 \leq \frac{\omega(1 - \omega^n)}{(1 - \omega)(1 - \omega\tau/\tau')} \frac{\tau}{T_n} \left( \frac{1}{\tau} \|Ax^* - Ax_1\|^2 + \frac{1}{1/\tau'} \|y^* - y_1\|^2 \right). \tag{53}$$

Using the definition of  $X_N$  and the convexity of the quadratic norm, we get

$$\|Ax^* - AX_N\|^2 \leq \frac{\tau N}{T_N} \frac{\omega}{1 - \omega\tau/\tau'} \left( \frac{1}{\tau} \|Ax^* - Ax_1\|^2 + \frac{1}{1/\tau'} \|y^* - y_1\|^2 \right) \tag{54}$$

since  $T_N = (1 - \omega^N)/(\omega^{N-1}(1 - \omega))$  according to (10). Injecting (54) in (52), and using the definition of  $L_A$ , we get (47). The ergodic convergence of  $(X_N)_N$  comes from the strong convexity of  $f$  [22, Theorem 9.2.6].  $\square$

*Remark 4.4* The convergence rate for the (relaxed) ADMM slightly differs from the linear rate for OPDHG stated in Corollary 3.1 (and, similarly, in Theorem 3.3). Indeed, for the (relaxed) ADMM, the bounds for the objective error (47) and the solution error (48) are of form

$$C \omega^N \frac{1}{1 - \omega^N} \left( 1 + C' \frac{N \omega^N}{1 - \omega^N} \right) \quad C, C' > 0 \text{ constant} \tag{55}$$

while the bounds for the OPDHG are only of form  $C \omega^N$ . The extra multiplicative factor, which appears in (55), decreases to 1. Hence, asymptotically, OPDHG and the (relaxed) ADMM are expected to have same linear convergence rate, provided the same value  $\omega$  can be chosen in both bounds. However, in the first iterations, the decrease in the multiplicative factor can visibly affect the convergence (see Sect. 5.1).

### 4.3 New Convergence Rate for the Classical ADMM and Acceleration

Let us estimate the best convergence rate, which can be achieved by the ADMM and its relaxed variant in the smooth case.

**Theorem 4.2** *Let  $f$  satisfy Assumption (S). Assume problem (3) has a solution denoted by  $x^*$ . Let  $\kappa_f := L_A^2/(\delta\gamma)$ .*

- (i) *The best convergence rate in Proposition 4.1 and Theorem 4.1 for the ADMM (Algorithm 3) is reached, when the parameter is  $\tau = \sqrt{2\delta^2\kappa_f}$ . We call this parameter the optimal parameter for the ADMM. The resulting convergence rate is  $\omega = 1/(\sqrt{1/(2\kappa_f)} + 1)$ .*
- (ii) *The best convergence rate in Proposition 4.1 and Theorem 4.1 for the relaxed ADMM (Algorithm 4) is reached, when the parameters are  $\tau = \tau^*$  and  $\tau' = 1/\sigma^*$  given by*

$$\tau^* = \frac{\delta}{2} \left( 1 + \sqrt{1 + 4\kappa_f} \right) \quad \text{and} \quad \sigma^* = \frac{\gamma}{2L_A^2} \left( 1 + \sqrt{1 + 4\kappa_f} \right) \tag{56}$$

*We call these parameter the optimal parameters for the relaxed ADMM. The resulting rate is  $\omega = (\sqrt{1 + 4\kappa_f} - 1)/(\sqrt{1 + 4\kappa_f} + 1)$ , which is the same that the one achieved in Theorem 3.2.*

*Proof* (i) When  $\tau' = \tau$ , the condition (44) is fulfilled for any  $\tau > 0$ . Then, the convergence rate for any  $\tau > 0$  is given by the left-hand side of (45), that is:

$$\omega(\tau) := \max \left\{ \frac{2}{\tau\gamma/L_A^2 + 2}, \frac{1}{\delta/\tau + 1} \right\} \tag{57}$$

Let us minimize this quantity with respect to  $\tau$ . One can first check that

$$\omega(\tau) = \begin{cases} 1/(\tau\gamma/(2L_A^2) + 1), & \text{if } \tau < \sqrt{2\delta L_A^2/\gamma}, \\ 1/(\delta/\tau + 1), & \text{if } \tau \geq \sqrt{2\delta L_A^2/\gamma}. \end{cases} \tag{58}$$

Hence, the minimizer is obviously  $\tau = \sqrt{2\delta L_A^2/\gamma}$ .

(ii) To prove this statement, it is sufficient to find  $\tau \geq \tau'$  satisfying both (44) and (45), which minimize

$$\omega(\tau, \tau') := \max \left\{ \frac{\tau'/\tau + 1}{\tau\gamma/L_A^2 + 2}, \frac{1}{\delta/\tau' + 1} \right\} \tag{59}$$

Set  $\tilde{\gamma} = \gamma/L_A^2$  and  $\tilde{\delta} = \delta$ . Then, (59) is exactly the lower bound of (31), applied to  $(\tau, \sigma = 1/\tau', \theta = \tau'/\tau)$ . Because of the link between  $\tau$  and  $\theta$ , it follows that

$$\min_{\substack{(\tau, \sigma, \theta) \\ \text{satisfying (11)}}} \max \left\{ \frac{\theta + 1}{\tau\tilde{\gamma} + 2}, \frac{1}{\sigma\tilde{\delta} + 1} \right\} \leq \min_{\substack{(\tau, \sigma, \theta) \\ \text{satisfying (11) and } \theta=1/(\tau\sigma)}} \max \left\{ \frac{\theta + 1}{\tau\tilde{\gamma} + 2}, \frac{1}{\sigma\tilde{\delta} + 1} \right\}. \tag{60}$$

According to Sect. 3.4, the left-hand side equals the rate given by Theorem 3.2. Otherwise said,

$$\frac{\sqrt{1 + 4\kappa_f} - 1}{\sqrt{1 + 4\kappa_f} + 1} \leq \min_{\substack{\tau \geq \tau' \\ \text{satisfying (44) and (45)}}} \omega(\tau, \tau'). \tag{61}$$

Hence, it suffices to check that  $(\tau, \tau') = (\tau^*, 1/\sigma^*)$  satisfy both (44) and (45), and realize the equality. □

## 5 Numerical Applications

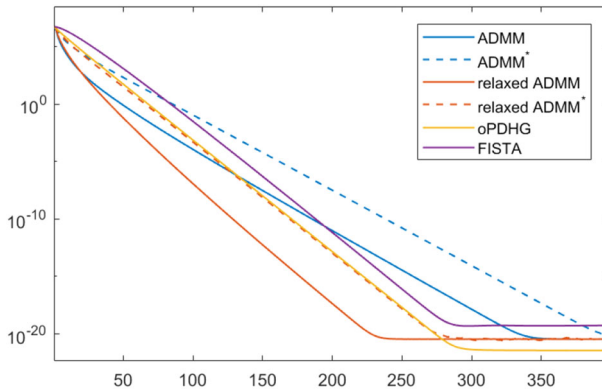
### 5.1 Denoising with TV-Huber

We apply the relaxed ADMM to a denoising problem. Let  $u \in \mathbb{R}^{3N_x N_y}$  be a color (noisy) image. We want to solve the following problem:

$$\min_{x \in \mathbb{R}^{3N_x N_y}} \left\{ f(x) := \frac{\mu}{2} \|x - u\|_2^2 + h(\nabla x) \right\}, \tag{62}$$

where the gradient linear operator  $\nabla = (\delta_x, \delta_y)^T$  is defined by the forward differences. The TV-Huber regularization term is defined by

$$h(\nabla x) = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} h_0(\|(\nabla x)_{i,j}\|) \quad \text{with } h_0(z) = \begin{cases} |z|^2/2, & \text{if } |z| \leq 1, \\ |z| - 1/2, & \text{if } |z| > 1. \end{cases} \tag{63}$$



**Fig. 1** Empirical convergence of the solution error for TV-Huber denoising. The plain lines stand for the sequence  $(x_n)$  while the dashed line stand for  $(X_n)$ . The y-scale is logarithmic. The x-axis corresponds to the iteration number.

Hence, this term acts like a quadratic regularization, when the image variations are small, and like a TV regularization, when they are larger. The quantity  $\mu > 0$  is a weight parameter. Let  $g := \mu \| \cdot - u \|_2^2 / 2$ . Functions  $h^*$  and  $g$  are, respectively, 1-convex and  $\mu$ -convex. The gradient operator is bounded of norm  $L \leq 2\sqrt{2}$  (this bound being tight when  $N_x$  or  $N_y$  go to  $+\infty$ ). Thus, we set  $L = 2\sqrt{2}$ .

The noisy image  $u$  is generated by adding Gaussian white noise of standard variation of 10 to a RGB color image of size  $201 \times 201$ . The image values are between 0 and 255. The code used is written in MATLAB. We chose  $\mu = 0.1$ . Iterations for the ADMM and the relaxed ADMM can be explicitly computed, thanks to the Euler equation. We tested two sets of parameters:

1. the optimal parameter for the ADMM (Algorithm 3)  $\tau = 4/\sqrt{\mu}$ ; the theoretical rate is  $\omega = 1/(\sqrt{\mu}/4 + 1) \approx 0.9267$ ;
2. the optimal parameters for the relaxed ADMM (Algorithm 4)  $\tau = (\sqrt{1 + 32/\mu} + 1)/2$  and  $\tau' = (\sqrt{1 + 32/\mu} - 1)/2$ ; the theoretical rate is  $\omega = (\sqrt{1 + 32/\mu} - 1)/(\sqrt{1 + 32/\mu} + 1) \approx 0.8943$ .

For comparison, we also applied OPDHG with the optimal parameters  $\tau = (1 + \sqrt{1 + 32/\mu})/16$ ,  $\sigma = \mu\tau$ , and  $\theta = (\sqrt{1 + 32/\mu} - 1)/(\sqrt{1 + 32/\mu} + 1)$  (as given in Theorem 3.2) and FISTA [15], which solves Problem (62) thanks to an accelerated forward-backward splitting algorithm (with an explicit gradient step for  $h(\nabla \cdot)$  and a proximal step for the quadratic fidelity term  $g$ ). The optimal parameter for FISTA are chosen according to [20, 23], namely the stepsize is  $1/8$  and the overrelaxation parameter is  $(1 - \sqrt{\mu/(8 + \mu)})^2(1 + \mu/8)$ . Note that for OPDHG and FISTA, an initialization of  $x$  has to be given, from which  $x_1$  is computed thanks to a gradient descent-like step. This is not the case for the (relaxed) ADMM, where  $x_1$  is defined as a minimizer which depends on  $(y_0, z_0)$ . Thus, for a fair comparison of the methods,  $x_0$  in OPDHG and FISTA is set to be  $x_1$  generated by the relaxed ADMM, while all the other variables are initialized by zeros.

To compare the methods, we compute the solution error (see Fig. 1). This quantity depends on the knowledge of the solution  $x^*$  of Problem (62), which is obviously not



available. However, since the OPDHG method has been shown to linearly converge to the solution, we use it to compute an approximation of  $x^*$  by computing a large number of iterations. Note that, even not proved theoretically, the same convergence rates are observed for both the ergodic sequence  $(X_n)$  and the iterate sequence  $(x_n)$  in the case of the (relaxed) ADMM. The empirical rates, computed by linear regression between iteration 60 and iteration 200, are, respectively, 0.8478 for the ADMM and 0.7826 for the relaxed ADMM. Note that, empirically, we observe that the theoretical rate becomes tight, compared to the observed rate, as the condition number grows to infinity.

Moreover, though the parameters were chosen so that the theoretical rates  $\omega$  are the same for the relaxed ADMM and OPDHG (see Theorems 3.2(ii) and 4.2), the solution error bound for the (relaxed) ADMM has a multiplicative factor (see Remark 4.4) which has an influence on the observed decay. Indeed, if one compares for instance the slope for the ADMM (for which the rate  $\omega$  given in 3.2(i) is greater than that of the OPDHG) and for OPDHG at the 500 first iterations, they look similar, while the slope for the ADMM becomes smaller, as the number of iterations grows. This explains why the ADMM and the relaxed ADMM present a better error decay, at least for the first iterations in the case of the ADMM.

Besides, one should keep in mind that both the ADMM and its relaxed variant require an operator inversion, unlike the OPDHG method and FISTA. Hence, even if comparable numbers of iterations are needed to achieve convergence, the (relaxed) ADMM iterations are more time consuming than the other methods and should be used only when the operator inversion can be implemented efficiently.

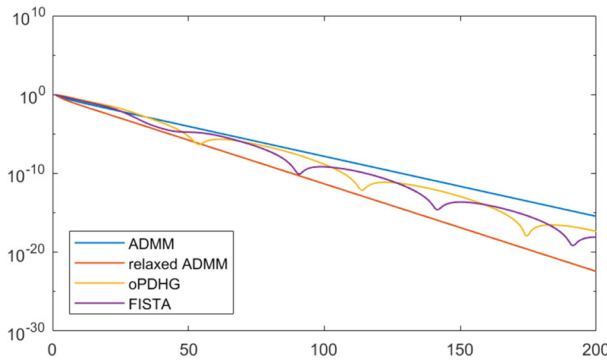
### 5.2 Local Oscillation Behavior

In this section, we consider a simple quadratic problem which reveals oscillating behaviors for OPDHG and FISTA. The ADMM and the relaxed ADMM, as we will see it, are not affected by these artifacts. Let  $N$  be a integer. We consider the following constrained problem:

$$\min_{x=(x_i) \in \mathbb{R}^N : x_0=1} \left\{ f(x) := \frac{M-m}{2} \|K_N x\|_2^2 + \frac{m}{2} \|x\|_2^2 \right\}, \tag{64}$$

where  $K_N : \mathbb{R}^N \rightarrow \mathbb{R}^{N-1}$  is defined by  $(K_N x)_i = (x_{i+1} - x_i)/2$ . The condition number of this problem is  $\kappa_f = (M - m)/m$ . Hence, the problem is ill-conditioned when  $m$  is negligible compared to  $M$ . Note that the minimizer of problem (64) can be explicitly computed. We chose  $N = 15$ ,  $m = 0.1$  and  $M = 10$ , so that  $\kappa_f = 99$ .

In Fig. 2, we observe oscillations for both the OPDHG method and strongly convex FISTA. Rippling for FISTA has been already observed for quadratic problems of this kind [24]. This phenomena occurs, when the overrelaxation parameter  $\theta$  is chosen too large compared to the eigenvalues of  $m I_N + (M - m) K_N^* K_N$ . For the OPDHG method, the reason remains unclear. An investigation of the oscillation behavior for polyhedra objectives can be found in [25]. Note that we do not observe such oscillations for ADMM-like schemes.



**Fig. 2** Empirical convergence of the solution error for the toy example. The y-scale is logarithmic. The x-axis corresponds to the iteration number

### 6 Conclusions

In this work, we studied the convergence of the OPDHG scheme, in the case where the composite problem is strongly convex with a differentiable with Lipschitz continuous gradient part. Using the equivalence between this algorithm and the ADMM, we provided a new convergence analysis of the latter. This analysis allowed us to design an accelerated variant of the ADMM, by changing the augmented Lagrangian parameter, which is proved to have same convergence rate as the OPDHG method. Hence, we showed that in the smooth case, the choice of the ADMM parameter(s) can be crucial in terms of convergence rate. Experimental results confirmed this theoretical analysis. It has also been observed that the accelerated ADMM does not introduce oscillations in some cases, unlike the OPDHG algorithm and FISTA.

### Appendix A: Proof of Theorem 3.1 and of Corollary 3.1

We proceed analogously to [7, Section 5.2], but we do not specify any parameter unless needed. This proof is also inspired by the one found in [9], which does not allow  $\theta \neq 1$ . For now, we only assume that  $0 < \theta \leq 1$ .

We suppose that  $\mathcal{L}$  satisfy Assumption (S2). Let  $(y_n, \xi_n, \bar{\xi}_n)_n$  a sequence generated by Algorithm 1. For any  $N \geq 1$ , points  $\xi_{n+1}$  and  $y_{n+1}$  can be seen as minimizers of convex problem, thus, the first-order optimality condition yields, respectively,  $-(\xi_{n+1} - \xi_n)/\tau - K^*y_{n+1} \in \partial G(\xi_{n+1})$  and  $-(y_{n+1} - y_n)/\sigma + K\xi_n \in \partial H^*(y_{n+1})$ . Using the definition of strong convexity, we get for any  $(\xi, y) \in Z \times Y$  (after expanding the scalar products and summing the two inequalities for  $G$  and  $H^*$ )

$$\begin{aligned} \mathcal{G}(\xi_{n+1}, \xi; y, y_{n+1}) &\leq \frac{\|\xi - \xi_n\|^2}{2\tau} - (1 + \tau\tilde{\gamma}) \frac{\|\xi - \xi_{n+1}\|^2}{2\tau} - \frac{\|\xi_n - \xi_{n+1}\|^2}{2\tau} \\ &\quad + \frac{\|y - y_n\|^2}{2\sigma} - (1 + \sigma\tilde{\delta}) \frac{\|y - y_{n+1}\|^2}{2\sigma} - \frac{\|y_n - y_{n+1}\|^2}{2\sigma} \\ &\quad + \langle K(\xi_{n+1} - \xi), y_{n+1} - y_n \rangle - \langle K(\xi_{n+1} - \bar{\xi}_n), y_{n+1} - y \rangle. \end{aligned} \tag{65}$$

For any  $N \geq 1$  and  $(\xi, y) \in Z \times Y$ , we set  $\Delta_n = \|\xi - \xi_n\|^2/(2\tau) + \|y - y_n\|^2/(2\sigma)$ . Let us first prove the following lemma:

**Lemma A.1** *Let  $\mathcal{L}$  satisfy Assumption (S2) and  $(y_n, \xi_n, \bar{\xi}_n)_n$  a sequence generated by Algorithm 1. Then, for any integer  $N \geq 1$ ,  $\tau, \sigma > 0$  and  $0 < \omega \leq \theta$ , we have for any  $(\xi, y) \in Z \times Y$*

$$\begin{aligned} \mathcal{G}(\xi_n, \xi; y, y_n) &\leq \Delta_n - \frac{\Delta_{n+1}}{\omega} + \omega \frac{\|\xi_{n-1} - \xi_n\|^2}{2\tau} + \frac{\|\xi_n - \xi_{n+1}\|^2}{2\tau} \\ &+ \omega \langle K(\xi_{n-1} - \xi_n), y - y_n \rangle - \langle K(\xi_n - \xi_{n+1}), y - y_{n+1} \rangle. \end{aligned} \tag{66}$$

*Proof* Replacing  $\bar{\xi}_n$  by  $\xi_n + \theta(\xi_n - \xi_{n-1})$  in (65), we get after simplification

$$\begin{aligned} \mathcal{G}(\xi_{n+1}, \xi; y, y_{n+1}) &\leq \frac{\|\xi - \xi_n\|^2}{2\tau} - (1 + \tau\tilde{\gamma}) \frac{\|\xi - \xi_{n+1}\|^2}{2\tau} - \frac{\|\xi_n - \xi_{n+1}\|^2}{2\tau} \\ &+ \frac{\|y - y_n\|^2}{2\sigma} - (1 + \sigma\tilde{\delta}) \frac{\|y - y_{n+1}\|^2}{2\sigma} - \frac{\|y_n - y_{n+1}\|^2}{2\sigma} \\ &+ \theta \langle K(\xi_{n-1} - \xi_n), y - y_{n+1} \rangle - \langle K(\xi_n - \xi_{n+1}), y - y_{n+1} \rangle. \end{aligned} \tag{67}$$

Now, we define  $\tau\tilde{\gamma} = \mu > 0$  and  $\sigma\tilde{\delta} = \mu' > 0$ . Let us bound the scalar products in (67). Introducing  $0 < \omega \leq \theta$ , we have

$$\begin{aligned} \theta \langle K(\xi_{n-1} - \xi_n), y - y_{n+1} \rangle &= \omega \langle K(\xi_{n-1} - \xi_n), y - y_n \rangle \\ &+ \omega \langle K(\xi_{n-1} - \xi_n), y_n - y_{n+1} \rangle \end{aligned} \tag{68}$$

Let us have a closer look at the last two terms. Let  $\alpha > 0$ . Since  $\langle K\Xi, Y \rangle \leq L_K(\alpha\|\Xi\|^2 + \|Y\|^2/\alpha)/2$ , the majoration (67) becomes, after simplification,

$$\begin{aligned} \mathcal{G}(\xi_{n+1}, \xi; y, y_{n+1}) &\leq \Delta_n - (1 + \mu) \Delta_{n+1} + \left( \frac{\omega L_K \sigma}{\alpha} - 1 \right) \frac{\|y_n - y_{n+1}\|^2}{2\sigma} \\ &+ \left( \frac{(\theta - \omega)L_K \sigma}{\alpha} + \mu - \mu' \right) \frac{\|y - y_{n+1}\|^2}{2\sigma} + \theta L_K \alpha \tau \frac{\|\xi_{n-1} - \xi_n\|^2}{2\tau} - \frac{\|\xi_n - \xi_{n+1}\|^2}{2\tau} \\ &+ \omega \langle K(\xi_{n-1} - \xi_n), y - y_n \rangle - \langle K(\xi_n - \xi_{n+1}), y - y_{n+1} \rangle. \end{aligned} \tag{69}$$

Choose  $\alpha = \omega L_K \sigma$ . Hence,  $\omega L_K / \alpha = 1/\sigma$ , so that the  $\|y_n - y_{n+1}\|^2$  term cancels. Since  $1 + \mu = 1/\omega + 1 + \mu - 1/\omega$ , one gets

$$\begin{aligned} \mathcal{G}(\xi_{n+1}, \xi; y, y_{n+1}) &\leq \Delta_n - \frac{\Delta_{n+1}}{\omega} + \omega \theta L_K^2 \tau \sigma \frac{\|\xi_n - \xi_{n-1}\|^2}{2\tau} - \frac{\|\xi_n - \xi_{n+1}\|^2}{2\tau} \\ &+ \left( \frac{1}{\omega} - \mu - 1 \right) \frac{\|\xi - \xi_{n+1}\|^2}{2\tau} + \left( \frac{\theta - \omega}{\omega} + \frac{1}{\omega} - \mu' - 1 \right) \frac{\|y - y_{n+1}\|^2}{2\sigma} \\ &+ \omega \langle K(\xi_{n-1} - \xi_n), y - y_n \rangle - \langle K(\xi_n - \xi_{n+1}), y - y_{n+1} \rangle. \end{aligned} \tag{70}$$

We can now set conditions on  $\omega, \theta, \tau$  and  $\sigma$ . First, choose  $\theta, \tau$  and  $\sigma$  so that  $\theta L_K^2 \tau \sigma \leq 1$ . Then, choose  $\theta$  so that both the parentheses are nonpositive, which implies that

$1/(\mu + 1) \leq \omega$  and  $(\theta + 1)/(\mu' + 2) \leq \omega$ . Since  $\omega \leq \theta$ , we get the wanted inequality.  $\square$

Let us now prove Corollary 3.1 and then Theorem 3.1. Set  $\xi^{-1} = \xi^0$ . Multiplying (15) by  $1/\omega^n$  and summing between  $n = 0$  and  $n = N - 1$  yield

$$\sum_{n=1}^N \frac{1}{\omega^{n-1}} \mathcal{G}(\xi_n, \xi; y, y_n) \leq \Delta_0 - \frac{\Delta_N}{\omega^N} - \frac{1}{\omega^{N-1}} \frac{\|\xi_{N-1} - \xi_N\|^2}{2\tau} - \frac{1}{\omega^{N-1}} \langle K(\xi_{N-1} - \xi_N), y - y_N \rangle. \tag{71}$$

One has, for any  $\beta > 0$ ,

$$- \langle K(\xi_{N-1} - \xi_N), y - y_N \rangle \leq L_K \left( \frac{\beta}{2} \|\xi_{N-1} - \xi_N\|^2 + \frac{1}{2\beta} \|y - y_N\|^2 \right) \tag{72}$$

and thus the right-hand side of (71) is bounded from above by

$$\Delta_0 - \frac{1}{\omega^N} \Delta_N + \frac{1}{\omega^{N-1}} (L_K \beta \tau - 1) \frac{\|\xi_{N-1} - \xi_N\|^2}{2\tau} + \frac{L_K}{\omega^{N-1}} \frac{1}{2\beta} \|y_N - y\|^2. \tag{73}$$

Choosing  $\beta = 1/(L_K \tau)$ , which cancels the  $\|\xi_{N-1} - \xi_N\|^2$  term, we have, since  $\omega L_K^2 \tau \sigma \leq \theta L_K^2 \tau \sigma \leq 1$  and  $\mathcal{G}(\xi_n, \xi^*; y^*, y_n) \geq 0$  for any  $N \geq 1$ ,

$$\begin{aligned} 0 &\leq \frac{\|\xi^* - \xi_N\|^2}{2\tau} + (1 - \omega L_K^2 \tau \sigma) \frac{\|y^* - y_N\|^2}{2\sigma} + \sum_{n=1}^N \frac{\omega^N}{\omega^{n-1}} \mathcal{G}(\xi_n, \xi^*; y^*, y_n) \\ &\leq \omega^N \left( \frac{\|\xi^* - \xi_0\|^2}{2\tau} + \frac{\|y^* - y_0\|^2}{2\sigma} \right). \end{aligned} \tag{74}$$

This inequality proves the linear convergence of the iterates (Corollary 3.1). We can now complete the proof of Theorem 3.1. Now, dividing (74) by  $\omega^N T_N$  and using convexity gives Theorem 3.1.  $\square$

### Appendix B: Equivalence Between Algorithm 4 and Algorithm 2

Let us prove that the iterations of Algorithm 4 are equivalent to (41). Let  $n \geq 0$ . By optimality in (37), one has for any  $x \in X$

$$g(x_{n+1}) + \langle Ax_{n+1}, y_n \rangle + \frac{\|Ax_{n+1} - z_n\|^2}{2\tau} \leq g(x) + \langle Ax, y_n \rangle + \frac{\|Ax - z_n\|^2}{2\tau} \tag{75}$$

For any  $\xi \in Y$ , we define

$$g_A(\xi) := \begin{cases} \inf \{g(x) : Ax = \xi\}, & \text{if } \{x \in X : Ax = \xi\} \neq \emptyset, \\ +\infty, & \text{otherwise.} \end{cases} \tag{76}$$

Let  $\xi \in Y$ . Suppose that  $\{x \in X : Ax = \xi\} \neq \emptyset$ . Then, taking the infimum over  $\{x \in X : Ax = \xi\}$ , we have

$$g(x_{n+1}) + \langle Ax_{n+1}, y_n \rangle + \frac{\|Ax_{n+1} - z_n\|^2}{2\tau} \leq g_A(\xi) + \langle \xi, y_n \rangle + \frac{\|\xi - z_n\|^2}{2\tau}. \tag{77}$$

Let us define  $\xi_{n+1} := Ax_{n+1}$ . Since  $g_A(\xi_{n+1}) \leq g(x_{n+1})$ , one has

$$g_A(\xi_{n+1}) + \langle \xi_{n+1}, y_n \rangle + \frac{\|\xi_{n+1} - z_n\|^2}{2\tau} \leq g_A(\xi) + \langle \xi, y_n \rangle + \frac{\|\xi - z_n\|^2}{2\tau} \tag{78}$$

for any  $\xi \in Y$ , including those such that  $\{x \in X : Ax = \xi\}$  is empty. Thus

$$g_A(\xi_{n+1}) + \langle \xi_{n+1}, y_n \rangle + \frac{\|\xi_{n+1} - z_n\|^2}{2\tau} = \min_{\xi \in Y} \left\{ g_A(\xi) + \langle \xi, y_n \rangle + \frac{\|\xi - z_n\|^2}{2\tau} \right\}. \tag{79}$$

In particular, one can check that  $\xi_{n+1} = \text{prox}_{\tau g_A}(z_n - \tau y_n)$ . Let us write the  $y$ -update (39). We have

$$y_{n+1} = y_n + \frac{\xi_{n+1} - z_{n+1}}{\tau'} \iff z_{n+1} = \xi_{n+1} - \tau'(y_{n+1} - y_n) \tag{80}$$

$$\iff z_{n+1} + \tau' y_{n+1} = \xi_{n+1} + \tau' y_n. \tag{81}$$

Let us define  $\bar{y}_{n+1}$  as in (30). Equation (80) implies that, for any  $n \geq 1$ , one has  $z_n - \tau y_n = \xi_n - \tau'(y_n - y_{n-1}) - \tau y_n = \xi_n - \tau(y_n + (\tau'/\tau)(y_n - y_{n-1}))$ . Injecting the latter equality in the expression of  $\xi_{n+1}$  yields (28). The  $z$ -update (38) can be rewritten as a proximal step

$$z_{n+1} = \text{prox}_{\tau'h}(\xi_{n+1} + \tau' y_n) = \text{prox}_{\tau'h}(z_{n+1} + \tau' y_{n+1}). \tag{82}$$

Using Moreau’s decomposition [18], we eventually obtain (29).

## References

1. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* **2**(1), 17–40 (1976)
2. Glowinski, R., Marocco, A.: Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. *Revue française d’automatique, informatique, recherche opérationnelle. Analyse numérique* **9**(2), 41–76 (1975)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
4. Hong, M., Luo, Z.-Q.: On the linear convergence of the alternating direction method of multipliers. *Math. Program.* **162**(1–2), 165–199 (2012)
5. Zhu, M., Chan, T.: An efficient primal–dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report.* 8-34 (2008)
6. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the Mumford–Shah functional. In: *IEEE International Conference on Computer Vision.* pp. 1133–1140 (2009)
7. Chambolle, A., Pock, T.: A first-order primal–dual algorithm for convex problems with applications to imaging. *JMIV* **40**(1), 120–145 (2011)

8. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal–dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.* **3**(4), 1015–1046 (2010)
9. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal–dual algorithm. *Math. Program.* **159**(1–2), 253–287 (2016)
10. Nesterov, Y.: A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ . *Sov. Math. Dokl.* **27**(2), 372–376 (1983)
11. Nishihara R., Lessard L., Recht B., Packard A., Jordan, M.: A general analysis of the convergence of ADMM. In: *International Conference on Machine Learning*, pp. 343–352 (2015)
12. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.* **66**(3), 889–916 (2016)
13. Davis, D., Yin, W.: Faster convergence rates of relaxed Peaceman–Rachford and ADMM under regularity assumptions. *Math. Oper. Res.* **42**, 783–805 (2017)
14. He, B., You, Y., Yuan, X.: On the convergence of primal–dual hybrid gradient algorithm. *SIAM J. Imaging Sci.* **7**(4), 2526–2537 (2014)
15. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
16. Rockafellar, R., Wets, R.: *Variational Analysis*. Springer, Berlin (2009)
17. Zhang, L., Mahdavi, M., Jin, R.: Linear convergence with condition number independent access of full gradients. In: *Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ (eds.) Advances in Neural Information Processing Systems*, pp. 980–988. Curran Associates, Inc., New York (2013)
18. Moreau, J.-J.: Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* **93**(2), 273–299 (1965)
19. Arrow, K., Hurwicz, L., Uzawa, H., Chenery, H.: *Studies in Linear and Non-linear Programming*. Stanford University Press, Redwood City (1958)
20. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. *Acta Numer.* **25**, 161–319 (2016)
21. Necoara, I., Nesterov, Y, Glineur, F.: Linear convergence of first order methods for non-strongly convex optimization. preprint [arXiv:1504.06298](https://arxiv.org/abs/1504.06298) (2015)
22. Allaire, G.: *Analyse numérique et optimisation: une introduction à la modélisation mathématique et à la simulation numérique*. Editions Ecole Polytechnique (2005)
23. Nesterov, Y.: *Introductory Lectures on Convex Optimization*. Kluwer Academic Publishers, Boston (2004)
24. O’Donoghue, B., Candes, E.: Adaptive restart for accelerated gradient schemes. *Found. Comput. Math.* **15**(3), 715–732 (2015)
25. Liang, J., Fadili, J., Peyré, G.: Local linear convergence analysis of primal–dual splitting methods. preprint [arXiv:1705.01926](https://arxiv.org/abs/1705.01926) (2017)