SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

A Proximal-Point Analysis of the Preconditioned Alternating Direction Method of Multipliers

K. Bredies H. Sun

SFB-Report No. 2015-006 October 2016

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

Der Wissenschaftsfonds.

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Das Land Steiermark

Stadt GRAZ Wissenschaft

A Proximal-Point Analysis of the Preconditioned Alternating Direction Method of Multipliers

Kristian Bredies* Hongpeng Sun†

November 9, 2016

Abstract

We study preconditioned ADMM (alternating direction method of multipliers) algorithms for non-smooth optimization problems. ADMM is a popular first-order method for general constrained optimization problems. However, one of its drawbacks is the need to solve implicit subproblems. In various applications, these subproblems are either easily solvable or linear but nevertheless challenging. We derive a preconditioned ADMM method that allows for flexible and efficient preconditioning for these linear subproblems. ADMM and its preconditioned version is written as a new kind of proximal point method directly on the primal problem and the weak (strong) convergence in infinite (finite) dimensional Hilbert spaces is proved. Various efficient preconditioners with any number of inner iterations may be used in this preconditioned ADMM framework. Furthermore, connections between the preconditioned ADMM and the preconditioned Douglas–Rachford method for linear-quadratic functionals (PDRQ) in [1] for general non-smooth problems are established. The methods are applied to total variation denoising problems and their benefits are shown in numerical experiments.

Keywords Alternating direction method of multipliers proximal point algorithm Weak convergence analysis Linear preconditioning techniques

Mathematics Subject Classification (2000) 65K10 · 90C25 · 65F08

1 Introduction

The *alternating direction method of multipliers* (ADMM) is a widely used technique for the solution of large-scale optimization problems [2,3]. Although developed in the 1970s, it regained popularity in the past several years in various fields of applied mathematics, in particular including variational imaging [4–6]. It is appropriate for solving optimization

*Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria, Email: kristian.bredies@uni-graz.at

†Institute for Mathematical Sciences, Renmin University of China, No. 59, Zhongguancun Street, Beijing, China, Email: hpsun@amss.ac.cn

problems involving two variables, an objective being the sum of two convex functionals depending on one of these variables, respectively, and a coupling in terms of general affine linear constraints. The ADMM method was introduced and developed in the middle 1970s by R. Glowinski, A. Marrocco, and M. Fortin [7, 8] as well as D. Gabay, B. Mercier [9–11]. An extensive study including a discussion of its history can be found in [7]. It was originally designed for partial differential equation (PDE) based variational problems [8]. Its application to numerical solutions of PDEs and the corresponding convergence properties are studied, for instance, in [10, 11].

Classical approaches for the convergence analysis of the ADMM method usually involve convex duality. In the special case where Fenchel–Rockafellar duality can be employed, it has been proven in [9] that ADMM is a variant of the Douglas–Rachford splitting method applied to the dual problem and a particular monotone operator splitting of the subgradients. Moreover, ADMM may also be interpreted as a proximal point method [12]. Operator splitting methods as well as the proximal point method are very important since they provide a guided way to study various splitting algorithms. The latter is also a starting point for the preconditioned ADMM method proposed in this paper.

The ADMM method can also be analyzed on the primal side, i.e., without considering the dual problem. This corresponds to introducing a Lagrange multiplier for the affine linear constraint and considering the augmented Lagrangian (see, e.g., [12]) associated with a positive "step-size". While the classical augmented Lagrangian method updates each of the two variables simultaneously, ADMM performs consecutive updates. Doing so, one has to solve implicit inclusion relations involving the sum of a linear operator and the subgradient of a convex functional. While often, one linear operator is the identity and solving the inclusion relations amounts to applying a proximal operator, dealing with operators other than the identity is challenging, even if the overall problem is still linear. However, if one is able to overcome this challenge, one benefits from the unconditional stability of the method. That is, if a solution exists, the iteration converges (or weakly converges) to a solution regardless of the choice of the step-size in the augmented Lagrangian functional. Such a behaviour is essentially different from forward-backward type first-order algorithms which are usually conditionally stable, see [13], for instance.

In this paper, we are interested in studying variants of the ADMM method for which the inversion steps are replaced by computationally tractable approaches. In this context, we consider in particular the special case where one functional in the objective is quadratic-linear as it is the case for various regularized linear inverse problems. This yields an implicit linear inversion step which might already be challenging and numerically expensive to solve, so computationally less demanding alternatives which still allow for unconditional stability are desirable. A well-known alternative is the inexact version of ADMM proposed by Eckstein and Bertsekas in [12]. However, the condition that the errors are summable is very strict and not flexible to use in practice, especially if the linear subproblem is of large scale. In order to overcome this drawback, we adopt the framework of [14, 15] and consider preconditioned update steps which correspond to classical symmetric splitting methods. This allows a flexible and efficient inexact solution of the linear equation without the need of controlling the errors.

The contribution of this paper is the study of a class of preconditioned ADMM algorithms. These algorithms allow for the inexact solution of the linear subproblem without controlling the error as, for instance, in [12]. We carry out a proximal-point analysis of the method in infinite-dimensional Hilbert spaces. This allows to prove weak convergence of the involved quantities under novel mild assumptions on the problem or the preconditioners. In particular, preconditioning with an arbitrary number of inner iterations for well-known and simple linear solution strategies is sufficient for the convergence of the whole iteration. Furthermore, we also compare, for special problem setups, the preconditioned ADMM and PDRQ which is a variant of a preconditioned Douglas–Rachford splitting method [1, 16] for the same problem class.

We would like to point out related research efforts in this direction in the literature as well as the novelties of this paper. In existing works, similar generalized ADMM or majorized ADMM are discussed, where general positive semidefinite or indefinite [14, 17], positive definite [18] proximal terms are added to the augmented Lagrangian functions, also with the aim of making the computation for the respective subproblems easy. Convergence properties are established, however, only finite dimensional spaces are considered in these works. Further, by writing the preconditioned and classical ADMM as a novel general proximal-point iteration, we give a rigorous analysis with some new conditions on the (weak) convergence in infinite-dimensional Hilbert spaces which are different from the finite-dimensional case as in [14, 17–20] and the infinite-dimensional case as in [9, 21]. We also introduce the concrete framework of preconditioners that was developed in [1, 16] into ADMM, including the classical iteration methods such as damped Jacobi, (successive over-relaxation) symmetric Gauss–Seidel and Richardson. And finally, differently from some of the linearization or prox-linear strategy as in [17, 18] that could possibly bring out new constraints on step size, many of the proposed preconditioners maintain the unconditional convergence properties of classical ADMM.

The organization of this paper is as follows. In Section 2, we give a new type proximal point iteration that is equivalent to preconditioned ADMM, prove its well-definition and discuss its connection to the classical ADMM. In Section 3, the convergence of the preconditioned ADMM is proven by reformulating it in terms of a proximal-point iteration and showing its weak and strong convergence in infinite and finite dimensional Hilbert spaces, respectively, under mild assumptions. In Section 4, we discuss some feasible preconditioners for preconditioned ADMM with quadratic-linear primal functional and give a comparison of preconditioned ADMM and PDRQ developed in [1], a method that aims at solving the same class of problems. In the last two sections, we present applications in order to demonstrate the efficiency of the preconditioning approaches for ADMM. Finally, we give a brief conclusion.

2 Preconditioned ADMM as a Proximal Point Method in Primal Form

In this paper, we consider ADMM algorithms for the following problem class:

$$\min_{u\in X, p\in Y} \ F(u) + G(p) \qquad \text{subject to } Au + Bp = c. \tag{2.1}$$

Here, X, Y, Z are real Hilbert spaces, along with $A \in \mathcal{L}(X, Z)$, $B \in \mathcal{L}(Y, Z)$, $c \in Z$, where $\mathcal{L}(X_1, X_2)$ denotes the space of linear and continuous mappings between Hilbert spaces X_1 and X_2, and $\mathcal{L}(X_1)$ is used if $X_2 = X_1$. The Hilbert space adjoint operator of A, B is denoted by $A^* \in \mathcal{L}(Z, X)$ and $B^* \in \mathcal{L}(Z, Y)$, respectively. Also, $F \in \Gamma_0(X)$ and $G \in \Gamma_0(Y)$ is assumed, where $\Gamma_0(\cdot)$ is defined as in [22], i.e.,

$$\Gamma_0(X) := \{f : X \to (-\infty, +\infty] \mid f \text{ is proper, convex and lower semi-continuous}\}.$$

Under the condition

$$\bigcup_{\lambda\geq 0} \lambda(A \operatorname{dom} F + B \operatorname{dom} G - c) = Z,$$

where the domain of F is given by $\operatorname{dom} F := \{x \in X \mid F(x) < +\infty\}$ (with an analogous definition for $\operatorname{dom} G$), Fenchel–Rockafellar duality is applicable [22] and hence, problem (2.1) is equivalent to the saddle-point problem

$$\min_{(u,p)\in \operatorname{dom} F\times \operatorname{dom} G} \ \max_{\lambda\in Z} \ \mathcal{L}(u, p, \lambda), \qquad \mathcal{L}(u, p, \lambda) = F(u) + G(p) + \langle \lambda, Au + Bp - c\rangle \tag{2.2}$$

where the dual variable $\lambda \in Z$ is a Lagrange multiplier of the constraint $Au + Bp = c$, see, for instance [12]. Recall that ADMM can directly be motivated by considering the augmented Lagrangian:

$$\mathcal{L}_r(u, p, \lambda) := F(u) + G(p) + \langle \lambda, Au + Bp - c\rangle + \frac{r}{2}\|Au + Bp - c\|^2, \tag{2.3}$$

where $r > 0$ is a step-size parameter. Differently from the classical augmented Lagrangian method which updates u and p simultaneously, ADMM now updates u and p separately:

$$\begin{cases} u^{k+1} := \operatorname*{arg\,min}_{u\in X} \ \mathcal{L}_r(u, p^k, \lambda^k), \\ p^{k+1} := \operatorname*{arg\,min}_{p\in Y} \ \mathcal{L}_r(u^{k+1}, p, \lambda^k), \\ \lambda^{k+1} := \lambda^k + r(Au^{k+1} + Bp^{k+1} - c). \end{cases} \tag{ADMM}$$

The updates of u^{k+1} and p^{k+1} involve the solution of two implicit equations involving the subgradients of F and G:

$$A^*(-rBp^k - \lambda^k + rc) \in (rA^*A + \partial F)u^{k+1}, \tag{2.4a}$$

$$B^*(-\lambda^k - rAu^{k+1} + rc) \in (rB^*B + \partial G)p^{k+1}, \tag{2.4b}$$

where in the following, we assume that

$$rA^*A + \partial F, \quad rB^*B + \partial G \text{ are strongly monotone,} \tag{2.5}$$

such that $(rA^*A + \partial F)^{-1}$ and $(rB^*B + \partial G)^{-1}$ exist and are Lipschitz continuous. Note that in the case $A^*A = I$ or $B^*B = I$, the inclusion relation amounts to evaluating a proximal operator for F and G, respectively, which is commonly assumed to be computable. In contrast to that, the evaluation of $(rA^*A + \partial F)^{-1}$ and $(rB^*B + \partial G)^{-1}$ can be challenging in general, even if ∂F or ∂G are (affine) linear.

In order to motivate the preconditioned ADMM algorithm, assume that F is quadratic-linear, i.e.,

$$F(u) = \langle Su, u \rangle / 2 - \langle f_0, u \rangle, \quad \forall u \in X, \tag{2.6}$$

where $S \in \mathcal{L}(X)$ is a self-adjoint, positive semi-definite operator, and $f_0 \in X$. This kind of F is employed, for instance, for solving regularized linear inverse problems with L^2 data fidelity terms in imaging and signal processing. The representation (2.6) yields $\partial F = S - \mathbf{1}_{f_0}$ with $\mathbf{1}_{f_0}$ being the constant mapping to f_0, thus, it is immediate that $rA^*A + \partial F$ is strongly monotone if

$$S + rA^*A > 0 \tag{2.7}$$

and (2.4a) amounts to solving the linear equation

$$(S + rA^*A)u^{k+1} = A^*(-rBp^k - \lambda^k - rc) + f_0. \tag{2.8}$$

As the exact solution of the linear equation (2.8) might already be challenging and numerically expensive, we are looking for computationally less demanding alternatives which still allow for unconditional stability. In order to do so, we adopt the framework of [14, 15] and consider preconditioned update steps for both u and p according to

$$u^{k+1} := \operatorname*{arg\,min}_{u \in X} \ \mathcal{L}_r(u, p^k, \lambda^k) + \tfrac{1}{2}\|u - u^k\|^2_{N - rA^*A}, \tag{2.9}$$

$$p^{k+1} := \operatorname*{arg\,min}_{p \in Y} \ \mathcal{L}_r(u^{k+1}, p, \lambda^k) + \tfrac{1}{2}\|p - p^k\|^2_{L - rB^*B}, \tag{2.10}$$

where $N \in \mathcal{L}(X)$, $L \in \mathcal{L}(Y)$ are self-adjoint "preconditioners" such that both $N - rA^*A$, $L - rB^*B$ are positive semi-definite and the squared norms read as $\|u\|^2_{N-rA^*A} := \langle (N - rA^*A)u, u\rangle$, $\|p\|^2_{L-rB^*B} := \langle (L - rB^*B)p, p\rangle$, respectively. These steps amount to solving

$$A^*(-rBp^k - \lambda^k + rc) + (N - rA^*A)u^k \in (N + \partial F)u^{k+1}, \tag{2.11}$$

$$B^*(-\lambda^k - rAu^{k+1} + rc) + (L - rB^*B)p^k \in (L + \partial G)p^{k+1}, \tag{2.12}$$

i.e., the rA^*A-term, rB^*B-term have been replaced by N and L respectively, which we are free to choose. In particular, inspired by the recently studied preconditioning techniques for the Douglas–Rachford iteration in [1], we are able, as we will see later in

Section 4, in the case of quadratic-linear F, to perform the following linear preconditioning step

$$u^{k+1} = u^k + M^{-1}[A^*(-rBp^k - \lambda^k + rc) + f_0 - Tu^k], \tag{2.13}$$

with M being a suitable preconditioner for the linear operator $T = S + rA^*A$. Such a step then allows a flexible and efficient inexact solution of the linear equation without the need of controlling the errors. The preconditioned ADMM method we study within this paper therefore reads as follows:

$$\begin{cases} u^{k+1} = (N + \partial F)^{-1}[A^*(-rBp^k - \lambda^k + rc) + (N - rA^*A)u^k], \\ p^{k+1} = (L + \partial G)^{-1}[B^*(-\lambda^k - rAu^{k+1} + rc) + (L - rB^*B)p^k], \\ \lambda^{k+1} = \lambda^k + r(Au^{k+1} + Bp^{k+1} - c). \end{cases} \tag{PADMM}$$

For the purpose of analyzing the convergence of this method in Hilbert spaces, we rewrite the iteration (PADMM) as a proximal point method. Denote by $(\hat{u}, \hat{p}, \hat{\lambda})$ in $X \times Y \times Z$ a saddle-point for problem (2.2), which is assumed to exist. Then, by the Karush–Kuhn–Tucker (KKT) conditions [23], we have

$$0 \in A^*\hat{\lambda} + \partial F(\hat{u}), \quad 0 \in B^*\hat{\lambda} + \partial G(\hat{p}), \quad A\hat{u} + B\hat{p} - c = 0. \tag{2.14}$$

Introduce two auxiliary variables $\hat{x} \in Z$ and $\hat{v} \in Z$, for which it holds that

$$\hat{\lambda} = \hat{v} - \hat{x}. \tag{2.15}$$

Combining (2.14) and (2.15), we get the following equivalent form of (2.14),

$$\text{find } \begin{pmatrix} \hat{x} \\ \hat{u} \\ \hat{v} \\ \hat{p} \end{pmatrix}, \text{ such that } \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} 0 & A & 0 & B \\ -A^* & \partial F & A^* & 0 \\ 0 & -A & 0 & -B \\ -B^* & 0 & B^* & \partial G \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{u} \\ \hat{v} \\ \hat{p} \end{pmatrix} + \begin{pmatrix} -c \\ 0 \\ c \\ 0 \end{pmatrix}. \tag{2.16}$$

With $\mathcal{U} = Z \times X \times Z \times Y$, $q = (x, u, v, p) \in \mathcal{U}$, and denoting by $\mathcal{A} : \mathcal{U} \to 2^{\mathcal{U}}$ the operator on the right-hand side of (2.16), the problem becomes $0 \in \mathcal{A}q$. We would like to employ a proximal point method in order to solve this inclusion relation. Introducing a possibly degenerate metric on $\mathcal{U}$ in form of a linear, continuous, self-adjoint and positive semi-definite "preconditioner" $\mathcal{M} : \mathcal{U} \to \mathcal{U}$, the proximal point method amounts to the iteration

$$0 \in \mathcal{M}(q^{k+1} - q^k) + \mathcal{A}q^{k+1}. \tag{2.17}$$

We will see that the both the preconditioned and unpreconditioned iteration schemes (ADMM) and (PADMM) follow from the specific choice

$$\mathcal{M} := \begin{pmatrix} \frac{2}{r}I & A & 0 & B \\ A^* & N & A^* & 0 \\ 0 & A & \frac{2}{r}I & -B \\ B^* & 0 & -B^* & L \end{pmatrix} \tag{2.18}$$

where $N \in \mathcal{L}(X)$, $L \in \mathcal{L}(Y)$ are self-adjoint and satisfy

$$N - rA^*A \geq 0, \qquad L - rB^*B \geq 0. \tag{2.19}$$

Indeed, $\mathcal{M}$ according to (2.18) is linear, continuous, self-adjoint and positive semi-definite as soon as (2.19) is fulfilled; we postpone the proof to Section 3.

The proximal point iteration (2.17), in case it is well-defined, can also be written as $\mathcal{M}q^k \in (\mathcal{M} + \mathcal{A})q^{k+1}$, giving rise to the fixed-point mapping

$$\mathcal{T} : \mathcal{U} \to \mathcal{U}, \quad q \mapsto (\mathcal{M} + \mathcal{A})^{-1}\mathcal{M}q. \tag{2.20}$$

With these notations, we can establish a relation between (2.17) and (ADMM).

Theorem 2.1. *The mapping $\mathcal{T}$ is well-defined under conditions* (2.5) *and* (2.19)*, i.e., for each $q \in \mathcal{U}$, there is a unique $q' \in \mathcal{U}$ with $\mathcal{M}q \in (\mathcal{M} + \mathcal{A})q'$.*

Moreover, (PADMM) *can be written in terms of* (2.17) *in the following sense: For $(\bar{u}^k, \bar{p}^k, \bar{\lambda}^k)$ generated by* (PADMM) *with initial value $(\bar{u}^0, \bar{p}^0, \bar{\lambda}^0)$, there is an initial value $q^0 = (x^0, u^0, v^0, p^0)$ such that for $q^k = (x^k, u^k, v^k, p^k)$ generated by* (2.17) *the following identity holds*

$$(\bar{u}^k, \bar{p}^k, \bar{\lambda}^k) = \big(u^{k-1}, p^k, -x^k + v^k - r(Au^k + Bp^k - c)\big) \qquad \textit{for each} \qquad k \geq 1.$$

Proof. For a given $q^k = (x^k, u^k, v^k, p^k)$ denote by $q^{k+1} := \mathcal{T}q^k$ when whenever the mapping is well-defined, i.e., when $\mathcal{M}q^k \in (\mathcal{M} + \mathcal{A})q^{k+1}$ is uniquely satisfied for q^{k+1}. By (2.18) and as $\mathcal{A}$ denotes the operator on the right-hand side of (2.16), the iteration can be written componentwise as

$$p^{k+1} = (L + \partial G)^{-1}[B^*(x^k - v^k) + Lp^k], \tag{2.21a}$$

$$v^{k+1} = v^k + \frac{r}{2}[Au^k - B(p^k - 2p^{k+1}) - c], \tag{2.21b}$$

$$u^{k+1} = (N + \partial F)^{-1}[A^*(x^k + v^k - 2v^{k+1}) + Nu^k], \tag{2.21c}$$

$$x^{k+1} = x^k + \frac{r}{2}[A(u^k - 2u^{k+1}) + B(p^k - 2p^{k+1}) + c]. \tag{2.21d}$$

Note that as ∂G is maximally monotone, conditions (2.5) and (2.19) ensure that $L + \partial G$ is strongly monotone with full range (see [22, Proposition 22.8]), hence $(L + \partial G)^{-1}$ is a well-defined Lipschitz-continuous mapping. This implies that p^{k+1} according to (2.21a) exists and is uniquely determined by q^k. With the newly updated p^{k+1} and (2.21b), v^{k+1} is uniquely determined. For the update of u^k according to (2.21c) one can argue analogously to the update of p^k, taking the new value v^{k+1} into account. Finally, again by analogy, x^{k+1} according to (2.21d) is uniquely determined. This leads to the well-definition of $\mathcal{T}$.

In order to prove the stated equivalence of (PADMM) and (2.17), introduce the variable λ^k according to

$$\lambda^k := -x^k + v^k - r(Au^k + Bp^k - c). \tag{2.22}$$

Then, the update of p^k as in (2.21a) becomes

$$\begin{aligned} p^{k+1} &= (L + \partial G)^{-1}[B^*(-\lambda^k - rAu^k + rc) + (L - rB^*B)p^k] \\ &= p^k + (L + \partial G)^{-1}[B^*(-\lambda^k - rAu^k + rc) - (rB^*B + \partial G)p^k]. \end{aligned} \tag{2.23}$$

For the update of λ^k, consider (2.22) for $k+1$, i.e.,

$$\lambda^{k+1} = -x^{k+1} + v^{k+1} - r(Au^{k+1} + Bp^{k+1} - c).$$

Plugging in v^{k+1} and x^{k+1} from (2.21b) and (2.21d) to the right hand side of above equation, we get

$$\lambda^{k+1} = -x^k + v^k + rB(p^{k+1} - p^k). \tag{2.24}$$

Substituting λ^k according to (2.22) yields

$$\lambda^{k+1} = \lambda^k + r(Au^k + Bp^{k+1} - c). \tag{2.25}$$

Finally, for the update of u^k, one writes (2.21c) as

$$u^{k+1} = (N + \partial F)^{-1}[A^*(x^k + v^k - 2v^{k+1} + rAu^k) + (N - rA^*A)u^k]. \tag{2.26}$$

With (2.21b) and (2.24), the argument for A^* becomes

$$\begin{aligned} x^k + v^k - 2v^{k+1} + rAu^k &= x^k + v^k - 2v^k - r[Au^k - B(p^k - 2p^{k+1}) - c] + rAu^k \\ &= x^k - v^k - rB(2p^{k+1} - p^k) + rc = -rBp^{k+1} - \lambda^{k+1} + rc, \end{aligned}$$

leading to the identity

$$u^{k+1} = (N + \partial F)^{-1}[A^*(-rBp^{k+1} - \lambda^{k+1} + rc) + (N - rA^*A)u^k]. \tag{2.27}$$

Thus, (2.23), (2.25) and (2.27) corresponds to (PADMM), however, with the update order $p^{k+1} \to \lambda^{k+1} \to u^{k+1}$ instead of $u^{k+1} \to p^{k+1} \to \lambda^{k+1}$. To account for that, let $(\bar{u}^0, \bar{p}^0, \bar{\lambda}^0)$ be the initial values of (PADMM) and choose

$$\begin{aligned} &x^0 = 0, \quad u^0 = (N + \partial F)^{-1}[A^*(-rB\bar{p}^0 - \bar{\lambda}^0 + rc) + (N - rA^*A)\bar{u}^0], \\ &p^0 = \bar{p}^0, \quad v^0 = \bar{\lambda}^0 + r(Au^0 + Bp^0 - c). \end{aligned}$$

Then, denoting by $(\bar{u}^k, \bar{p}^k, \bar{\lambda}^k)$ the sequence generated by (PADMM), it follows that $(\bar{u}^k, \bar{p}^k, \bar{\lambda}^k) = (u^{k-1}, p^k, \lambda^k)$ for $k \geq 1$ which is, plugging in (2.22), the desired statement. □

3 Convergence of Preconditioned ADMM

In Theorem 2.1, we have shown the equivalence of (PADMM) and the proximal point method (2.17). With this result, we can carry out a convergence analysis which is based on non-expansive mappings. We will obtain weak (and strong) convergence under assumptions on the problem and preconditioner which will in particular give strong convergence in finite-dimensional spaces. The result will be immediate using the following intermediate statements, formulated in a series of lemmas.

Throughout this section, let (2.5) and (2.19) be satisfied such that Theorem 2.1 is applicable. We start with the relation between fixed points of $\mathcal{T}$ and solutions of the saddle point problem (2.2).

Lemma 3.1. *If a solution of the saddle point problem* (2.2) *exists, then fixed points of* $\mathcal{L}$ *according to* (2.20) *exist. For each fixed point* $(\hat{x}, \hat{u}, \hat{v}, \hat{p})$ *of* $\mathcal{T}$*, the triple* $(\hat{u}, \hat{p}, \hat{\lambda})$ *with* $\hat{\lambda} = \hat{v} - \hat{x}$ *is a saddle point of* (2.2).

Proof. Suppose $(\hat{u}, \hat{p}, \hat{\lambda})$ is a saddle point of $\mathcal{L}$ according to (2.2), then by optimality, $-A^*\hat{\lambda} \in \partial F(\hat{u})$ and $-B^*\hat{\lambda} \in \partial G(\hat{p})$ as well as $A\hat{u} + B\hat{p} - c = 0$. Choosing $\hat{v} - \hat{x} = \hat{\lambda}$, equations (2.14) and (2.15) are satisfied, i.e., $\hat{q} = (\hat{x}, \hat{u}, \hat{v}, \hat{p})$ fulfills $0 \in \mathcal{A}\hat{q}$ according to (2.16). But this implies that $\hat{q}$ is a fixed-point of $\mathcal{T}$.

Conversely, for $\hat{q} = (\hat{x}, \hat{u}, \hat{v}, \hat{p})$ a fixed-point of $\mathcal{T}$, we see that $0 \in \mathcal{A}\hat{q}$ follows, equations (2.14) and (2.15) are satisfied, consequently, $(\hat{u}, \hat{p}, \hat{v} - \hat{x})$ fulfills the optimality conditions for (2.2). □

For the weak convergence of the proximal iteration (2.17), let us first study the "preconditioner" $\mathcal{M}$. It can easily be checked that $\mathcal{M}$ in (2.18) is a self-adjoint operator in $\mathcal{L}(\mathcal{U})$, i.e., $\mathcal{M} : \mathcal{U} \to \mathcal{U}$ is linear, continuous and satisfies

$$\langle \mathcal{M}q_1, q_2 \rangle = \langle q_1, \mathcal{M}q_2 \rangle, \quad \forall q_1, q_2 \in \mathcal{U}. \tag{3.1}$$

By direct calculation, we see that $\mathcal{M}$ is positive semi-definite with associated $\|\cdot\|_{\mathcal{M}}$-seminorm which reads as

$$\begin{aligned} \|q\|_{\mathcal{M}}^2 := \langle \mathcal{M}q, q \rangle = &\frac{1}{r}\|rAu + x + v\|_Y^2 + \frac{1}{r}\|rBp + (x - v)\|_Y^2 \\ &+ \|u\|_{N-rA^*A}^2 + \|p\|_{L-rB^*B}^2 \end{aligned} \tag{3.2}$$

where the norms $\|\cdot\|_{N-rA^*A}^2$ and $\|\cdot\|_{L-rB*B}^2$ are defined as for (2.9) and (2.10), respectively. As (2.19) is satisfied, $(N - rA^*A)^{1/2}$ and $(L - rB^*B)^{1/2}$ exist, see e.g., [24], and are linear, continuous and positive semi-definite operators. In particular, $\|u\|_{N-rA^*A}^2 = \|(N - rA^*A)^{1/2}u\|_X^2$ and $\|p\|_{L-rB^*B}^2 = \|(L - rB^*B)^{1/2}p\|_Y^2$, so if $\|q\|_{\mathcal{M}} = 0$ for $q = (x, u, v, p)$, then

$$\begin{aligned} rAu + x + v = 0, \quad (N - rA^*A)u = 0, \\ rBp + (x - v) = 0, \quad (L - rB^*B)p = 0, \end{aligned} \tag{3.3}$$

which might not necessarily imply that $(x, u, v, p) = 0$, i.e., the metric associated with $\mathcal{M}$ is in general degenerate. Nevertheless, the iteration is non-expansive with respect to the $\mathcal{M}$-seminorm.

Lemma 3.2. *The mapping $\mathcal{T}$ is $\mathcal{M}$-firmly non-expansive in the sense that for $q_1, q_2 \in \mathcal{U}$,*

$$\|\mathcal{T}q_1 - \mathcal{T}q_2\|_{\mathcal{M}}^2 + \|(I-\mathcal{T})q_1 - (I-\mathcal{T})q_2\|_{\mathcal{M}}^2 \le \|q_1 - q_2\|_{\mathcal{M}}^2. \tag{3.4}$$

Further, there exists a constant $C > 0$ such that for all $q_1, q_2 \in \mathcal{U}$,

$$\|\mathcal{T}q_1 - \mathcal{T}q_2\| \le C\|q_1 - q_2\|_{\mathcal{M}}. \tag{3.5}$$

Proof. The statement (3.4) is immediate from the properties of the proximal point method, we provide a proof for the sake of completeness. Let us first verify that $\mathcal{A}$ is monotone in $\mathcal{U}$. Choose $q_i = (x_i, u_i, v_i, p_i) \in \mathcal{U}$ and $q_i' = (x_i', u_i', v_i', p_i')$, $i = 1, 2$ such that $q_i' \in \mathcal{A}q_i$. In particular, $u_i' - A^*(v_i - x_i) \in \partial F(u_i)$ and $p_i' - B^*(v_i - x_i) \in \partial G(p_i)$, hence

$$\begin{aligned}\langle u_1' - u_2', u_1 - u_2\rangle &\ge \langle A^*(v_1 - x_1) - A^*(v_2 - x_2), u_1 - u_2\rangle,\\ \langle p_1' - p_2', p_1 - p_2\rangle &\ge \langle B^*(v_1 - x_1) - B^*(v_2 - x_2), p_1 - p_2\rangle,\end{aligned}$$

and, as $x_i' = Au_i + Bp_i - c$ and $v_i' = -Au_i - Bp_i + c$,

$$\begin{aligned}\langle x_1' - x_2', x_1 - x_2\rangle &= \langle A(u_1 - u_2) + B(p_1 - p_2), x_1 - x_2\rangle,\\ \langle v_1' - v_2', v_1 - v_2\rangle &= \langle -A(u_1 - u_2) - B(p_1 - p_2), v_1 - v_2\rangle.\end{aligned}$$

Summing up all scalar products yields

$$\langle q_1' - q_2', q_1 - q_2\rangle \ge 0,$$

hence $\mathcal{A}$ is monotone in $\mathcal{U}$.

Next, fix $q_1, q_2 \in \mathcal{U}$ and consider $\mathcal{T}q_1$, $\mathcal{T}q_2$. By the definition (2.20), we can choose $q_1' \in \mathcal{A}\mathcal{T}q_1 \subset \mathcal{U}$, $q_2' \in \mathcal{A}\mathcal{T}q_2 \subset \mathcal{U}$, such that

$$\mathcal{M}\mathcal{T}q_1 + q_1' = \mathcal{M}q_1, \qquad \mathcal{M}\mathcal{T}q_2 + q_2' = \mathcal{M}q_2.$$

Then, we have $\langle q_1' - q_2', \mathcal{T}q_1 - \mathcal{T}q_2\rangle \ge 0$ as $\mathcal{A}$ is monotone, consequently,

$$\begin{aligned}\|\mathcal{T}q_1 - \mathcal{T}q_2\|_{\mathcal{M}}^2 &\le \langle \mathcal{M}(\mathcal{T}q_1 - \mathcal{T}q_2), \mathcal{T}q_1 - \mathcal{T}q_2\rangle + \langle q_1' - q_2', \mathcal{T}q_1 - \mathcal{T}q_2\rangle\\ &= \langle \mathcal{M}(q_1 - q_2), \mathcal{T}q_1 - \mathcal{T}q_2\rangle = \langle \mathcal{T}q_1 - \mathcal{T}q_2, q_1 - q_2\rangle_{\mathcal{M}}.\end{aligned}$$

Employing this estimate, the desired inequality (3.4) is obtained as follows:

$$\begin{aligned}&\|\mathcal{T}q_1 - \mathcal{T}q_2\|_{\mathcal{M}}^2 + \|(I-\mathcal{T})q_1 - (I-\mathcal{T})q_2\|_{\mathcal{M}}^2\\ &\quad = \|q_1 - q_2\|_{\mathcal{M}}^2 - 2\underbrace{\big(\langle \mathcal{T}q_1 - \mathcal{T}q_2, q_1 - q_2\rangle_{\mathcal{M}} - \|\mathcal{T}q_1 - \mathcal{T}q_2\|_{\mathcal{M}}^2\big)}_{\ge 0} \le \|q_1 - q_2\|_{\mathcal{M}}^2.\end{aligned}$$

To establish (3.5), introduce the notations

$$\eta = rAu + x + v, \quad \xi = rBp + x - v, \quad \mu = (N - rA^*A)^{\frac{1}{2}}u, \quad \zeta = (L - rB^*B)^{\frac{1}{2}}p,$$

for $q = (x, u, v, p)$. Suppose that $q_i = (x_i, u_i, v_i, p_i) \in \mathcal{U}$ and $q_i' = (x_i', u_i', v_i', p_i') = \mathcal{T}q_i$ for $i = 1, 2$. We will estimate, in terms of the $\mathcal{M}$-seminorm, the difference of each component in q_i' separately. (In the following, the same C stands for different constants as we are only interested in the qualitative estimate.) By (2.21a) and the Lipschitz continuity of $(L+\partial G)^{-1}$ (which is due to the strong monotonicity of $(L+\partial G)$, see (2.5) and (2.19)), we have

$$\begin{aligned}\|p_1'-p_2'\| &\le C\big(\|B^*\|\|\xi_1-\xi_2\|+\|(L-rB^*B)^{1/2}\|\|\zeta_1-\zeta_2\|\big)\\ &\le C(r\|B^*\|^2+\|L-rB^*B\|)^{\frac{1}{2}}(\tfrac{1}{r}\|\xi_1-\xi_2\|^2+\|\zeta_1-\zeta_2\|^2)^{\frac{1}{2}} \le C\|q_1-q_2\|_{\mathcal{M}}\end{aligned}$$

for some $C>0$. Note that (2.21b) implies $v_i' = \frac{1}{2}(\eta_i-\xi_i)+rBp_i'-\frac{r}{2}c$, hence

$$\|v_1'-v_2'\| \le \tfrac{1}{2}\|\eta_1-\eta_2\|+\tfrac{1}{2}\|\xi_1-\xi_2\|+r\|B\|\|p_1'-p_2'\| \le C\|q_1-q_2\|_{\mathcal{M}}$$

Moreover, (2.21b) also implies that $x_i+v_i-2v_i'+rAu_i = -2rBp_i'+\xi_i+rc$, so (2.21c) as well as the Lipschitz continuity of $(N+\partial F)^{-1}$ (which follows analogously to the Lipschitz continuity of $(L+\partial G)^{-1}$), yields

$$\begin{aligned}\|u_1'-u_2'\| &\le C\big(2r\|A^*\|\|B\|\|p_1'-p_2'\|+\|A^*\|\|\xi_1-\xi_2\|+\|(N-rA^*A)^{\frac{1}{2}}\|\|\mu_1-\mu_2\|\big)\\ &\le C\|q_1-q_2\|_{\mathcal{M}}.\end{aligned}$$

Finally, (2.21d) yields $x_i' = \frac{1}{2}(\eta_i+\xi_i)-r(Bp_i'+Au_i')+\frac{r}{2}c$, hence

$$\|x_1'-x_2'\| \le \tfrac{1}{2}\|\eta_1-\eta_2\|+\tfrac{1}{2}\|\xi_1-\xi_2\|+r\|B\|\|p_1'-p_2'\|+r\|A\|\|u_1'-u_2'\| \le C\|q_1-q_2\|_{\mathcal{M}}.$$

Putting the estimates together then gives the desired inequality (3.5). □

The next lemma is concerned with the asymptotic regularity of the iteration sequence. With the help of Lemma 3.2, the proof is nearly immediate and the same as in [1]. Thus, we just give the result without proof.

Lemma 3.3. *If there is a solution to problem* (2.2)*, then the sequence* $\{q^k\}$ *given by* $q^k := \mathcal{T}^k q^0$*,* $q^0 \in \mathcal{U}$*, satisfies*
(i) $\{q^k\}$ *is bounded,*
(ii) $\|q^{k+1}-q^k\| \to 0$ *as* $k\to\infty$*.*

The main technical difficulty in establishing weak convergence of the iterates following the lines of [25] and [1] is deriving the demi-closedness of the operator $I-\mathcal{T}$ in $\mathcal{U}$. We will discuss sufficient conditions for this property to hold: the compactness of A or B, the weak closedness of the operators ∂F and ∂G and various injectivity properties. On the one hand, denoting by $\mathcal{N}$ the kernel of an operator, we consider the injectivity condition

$$\mathcal{N}\begin{pmatrix} A & B \\ N-rA^*A & 0 \\ 0 & L-rB^*B\end{pmatrix} = \{0\}, \tag{3.6}$$

meaning that $Au+Bp=0$, $(N-rA^*A)u=0$ and $(L-rB^*B)p=0$ implies $u=0$ and $p=0$. On the other hand, we are interested in the special case that F according to (2.6) is quadratic-linear which imposes the representation

$$\partial F = S - \mathbf{1}_{f_0} \tag{3.7}$$

for $S \in \mathcal{L}(X)$ linear, continuous and positive semi-definite and $f_0 \in X$. In that case, we assume the injectivity condition

$$\mathcal{N}\begin{pmatrix} N - rA^*A \\ S - rA^*A \end{pmatrix} = \{0\}. \tag{3.8}$$

With these prerequisites, we are able to formulate the demi-closedness result.

Lemma 3.4. *For general operators A and B satisfying the condition* (2.5) *and with preconditioners N and L satisfying* (2.19)*, let one of the following conditions be true:*

(i) The injectivity property (3.6) *holds.*

(ii) ∂F and ∂G are weakly closed.

(iii) A or B is compact.

(iv) ∂G is weakly closed and N is injective on $\mathcal{N}(A)$.

(v) The subgradient of F satisfies $\partial F = S - \mathbf{1}_{f_0}$ according to (3.7)*, the injectivity property* (3.8) *holds and L is injective on $\mathcal{N}(B)$.*

Then, $I - \mathcal{T}$ is weak-strong closed (or demiclosed) in $\mathcal{U}$, i.e., from $q^k \rightharpoonup q$ and $(I - \mathcal{T})q^k \to \bar{q}$ it follows that $(I - \mathcal{T})q = \bar{q}$.

Proof. Observe that in any case, the previous lemmas and the results of Section 2 are applicable. Let $\hat{q} = \mathcal{T}q + \bar{q}$ for $q^k \rightharpoonup q$ and $(I - \mathcal{T})q^k \to \bar{q}$ as $k \to \infty$. We first prove that

$$\|q - \hat{q}\|_{\mathcal{M}} = 0. \tag{3.9}$$

Since the $\mathcal{M}$-seminorm is continuous on $\mathcal{U}$, we have $\|q^k - \mathcal{T}q^k - \bar{q}\|_{\mathcal{M}} \to 0$. By the $\mathcal{M}$-nonexpansivity of $\mathcal{T}$, it furthermore follows that

$$\begin{aligned} \liminf_{k\to\infty} \|q^k - q\|_{\mathcal{M}}^2 &\geq \liminf_{k\to\infty} \|\mathcal{T}q^k - \mathcal{T}q\|_{\mathcal{M}}^2 \\ &= \liminf_{k\to\infty} \|q^k - \mathcal{T}q - \bar{q}\|_{\mathcal{M}}^2 - 2\lim_{k\to\infty} \langle q^k - \mathcal{T}q - \bar{q}, q^k - \mathcal{T}q^k - \bar{q}\rangle \\ &\quad + \lim_{k\to\infty} \|q^k - \mathcal{T}q^k - \bar{q}\|_{\mathcal{M}}^2 = \liminf_{k\to\infty} \|q^k - \hat{q}\|_{\mathcal{M}}^2. \end{aligned}$$

However, we also have

$$\begin{aligned} \liminf_{k\to\infty} \|q^k - \hat{q}\|_{\mathcal{M}}^2 &= \liminf_{k\to\infty} \|q^k - q + q - \hat{q}\|_{\mathcal{M}}^2 \\ &= \liminf_{k\to\infty} \|q^k - q\|_{\mathcal{M}}^2 + 2\lim_{k\to\infty} \langle q^k - q, q - \hat{q}\rangle + \|q - \hat{q}\|_{\mathcal{M}}^2 \\ &= \liminf_{k\to\infty} \|q^k - q\|_{\mathcal{M}}^2 + \|q - \hat{q}\|_{\mathcal{M}}^2. \end{aligned}$$

This implies (3.9). Denoting the components by $q =: (x, u, v, p)$, $\bar{q} =: (\bar{x}, \bar{u}, \bar{v}, \bar{p})$ and $q' = \mathcal{T}q =: (x', u', v', p')$, the representation (3.2) allows us to conclude that

$$\begin{aligned} rA(u - u' - \bar{u}) + (x - x' - \bar{x}) + (v - v' - \bar{v}) &= 0, \\ rB(p - p' - \bar{p}) + (x - x' - \bar{x}) - (v - v' - \bar{v}) &= 0, \\ (N - rA^*A)(u - u' - \bar{u}) &= 0, \\ (L - rB^*B)(p - p' - \bar{p}) &= 0. \end{aligned} \tag{3.10}$$

Our aim is to show that $q - q' - \bar{q} = 0$ which is the claimed statement. By definition of $\mathcal{T}$ (also see (2.21b) and (2.21d)), we have

$$v - v' = -\frac{r}{2}[Au + B(2p' - p) - c], \qquad x - x' = -\frac{r}{2}[A(u - 2u') + B(p - 2p') + c].$$

Denoting by $\mathcal{T}q^k = \big((x^k)', (u^k)', (v^k)', (p^k)'\big)$, the updates for v^k and x^k according to (2.21b) and (2.21d) lead to

$$\begin{aligned}\bar{v} \leftarrow v^k - (v^k)' &= -\frac{r}{2}\big[Au^k - 2B\big(p^k - (p^k)'\big) + Bp^k - c\big] \rightharpoonup -\frac{r}{2}\big[Au - B(2\bar{p} - p) - c\big],\\ \bar{x} \leftarrow x^k - (x^k)' &= -\frac{r}{2}\big[2A\big(u^k - (u^k)'\big) + 2B\big(p^k - (p^k)'\big) - Au^k - Bp^k + c\big]\\ &\rightharpoonup -\frac{r}{2}\big[A(2\bar{u} - u) + B(2\bar{p} - p) + c\big],\end{aligned}$$

hence the left- and right-hand sides coincide. Together, we get

$$\begin{aligned}0 &= (x - x' - \bar{x}) - (v - v' - \bar{v}) + rB(p - p' - \bar{p})\\ &= \frac{r}{2}\big[A(2u' - u) + B(2p' - p) + A(2\bar{u} - u) + B(2\bar{p} - p)\big]\\ &+ \frac{r}{2}\big[Au + B(2p' - p) - Au + B(2\bar{p} - p)\big] + rB(p - p' - \bar{p})\\ &= r\big[-A(u - u' - \bar{u}) - B(p - p' - \bar{p})\big].\end{aligned} \tag{3.11}$$

We can now use this to obtain from (3.10) that

$$x - x' - \bar{x} = -\frac{r}{2}\big[A(u - u' - \bar{u}) + B(p - p' - \bar{p})\big] = 0. \tag{3.12}$$

For the following steps we need to distinguish cases.
Case (i). If the injectivity property (3.6) holds, then (3.10) and (3.11) imply that $u - u' - \bar{u} = 0$ and $p - p' - \bar{p} = 0$ which in turn yields $v - v' - \bar{v} = 0$ by virtue of (3.10), so $q - q' - \bar{q} = 0$.
Case (ii). If ∂F and ∂G are weakly closed, then rewriting the update rule (2.21a) as

$$L\big(p^k - (p^k)'\big) + B^*(x^k - v^k) \in \partial G[(p^k)'] \qquad \text{for each } k, \tag{3.13}$$

and observing $L\big(p^k - (p^k)'\big) + B^*(x^k - v^k) \rightharpoonup L\bar{p} + B^*(x - v)$, $(p^k)' = p^k + (p^k)' - p^k \rightharpoonup p - \bar{p}$ as $k \to \infty$ allows to conclude $L\bar{p} + B^*(x - v) \in \partial G(p - \bar{p})$. This is equivalent to $p - \bar{p} = (L + \partial G)^{-1}[B^*(x - v) + Lp]$ where the right-hand side coincides, by definition, with p', hence $p - p' - \bar{p} = 0$. Using (3.10) and (3.12) then implies $v - v' - \bar{v} = 0$.

We will use an analogous argument to obtain $u - u' - \bar{u} = 0$. The update rule (2.21c) for u^k gives

$$N\big(u^k - (u^k)'\big) + A^*\big(x^k + v^k - 2(v^k)'\big) \in \partial F[(u^k)'] \qquad \text{for each } k, \tag{3.14}$$

where $N\big(u^k - (u^k)'\big) + A^*\big(x^k + v^k - 2(v^k)'\big) \rightharpoonup N\bar{u} + A^*(x - v + 2\bar{v})$ as well as $(u^k)' \rightharpoonup u - \bar{u}$ as $k \to \infty$. Weak closedness as well as $\bar{v} = v - v'$ imply that $N\bar{u} + A^*(x + v - 2v') \in \partial F(u - \bar{u})$

which is equivalent to $u-\bar{u} = (N+\partial F)^{-1}\big(A^*(x+v-2v')+Nu\big)$. Again, by definition, the right-hand side coincides with u', hence $u-u'-\bar{u} = 0$. Together, this means $q-q'-\bar{q} = 0$ and we are done in this case.

Case (iii). If A is compact, it follows by complete continuity that $Au^k \to Au$ in Z as $k \to \infty$. Thus, $\{Bp^k\}$ also converges strongly to Bp which follows, for instance, from

$$-\frac{r}{2}Bp^k = v^k - (v^k)' + \frac{r}{2}\big[Au^k - 2B\big(p^k - (p^k)'\big) - c\big] \to \bar{v} + \frac{r}{2}\big[Au - 2B\bar{p} - c\big] = -\frac{r}{2}Bp,$$

as $k \to \infty$, the latter since weak and strong limits coincide. If B is compact, it follows by analogy that $Bp^k \to Bp$ and $Au^k \to Au$ as $k \to \infty$, so we have the latter convergence in both situations.

We would like to proceed analogously to Case (ii) under these weakened conditions. For this purpose, observe that $B(p^k)' \to B(p - \bar{p})$, $L\big(p^k - (p^k)'\big) \to L\bar{p}$ as $k \to \infty$ and consider the scalar product sequence between the left-hand side and argument of the subgradient in (3.13) which satisfies

$$\begin{aligned} \langle L\big(p^k - (p^k)'\big) + B^*(x^k - v^k), (p^k)'\rangle &= \langle L(p^k - (p^k)'), (p^k)'\rangle + \langle x^k - v^k, B(p^k)'\rangle \\ &\to \langle L\bar{p}, p - \bar{p}\rangle + \langle x - v, B(p - p')\rangle = \langle L\bar{p} + B^*(x - v), p - \bar{p}\rangle \end{aligned}$$

as $k \to \infty$, the latter since scalar products of weakly and strongly convergent sequences are taken. As ∂G is maximally monotone, it follows that $L\bar{p}+B^*(x-v) \in \partial G(p-\bar{p})$ and we can proceed as in Case (ii) to conclude that $p - p' - \bar{p} = 0$ as well as $v - v' - \bar{v} = 0$. As $A(u^k)' \to A(u - \bar{u})$ and $N\big(u^k - (u^k)'\big) \to N\bar{u}$ as $k \to \infty$, an analogous argument can be applied to the scalar product sequence between the left-hand side and subgradient argument in (3.14) such that also $N\bar{u} + A^*(x + v - 2v') \in \partial F(u - \bar{u})$. Also here, the arguments in Case (ii) lead to $u - u' - \bar{u} = 0$, meaning that indeed $q - q' - \bar{q} = 0$.

Case (iv). If ∂G is weakly closed, then the arguments in Case (ii) still yield $p-p'-\bar{p} = 0$ as well as $v - v' - \bar{v} = 0$. The identities (3.10) then imply $A(u - u' - \bar{u}) = 0$ and in particular, $N(u - u' - \bar{u}) = (N - rA^*A)(u - u' - \bar{u}) = 0$, so if N is injective on $\mathcal{N}(A)$, then $u - u' - \bar{u} = 0$ follows, and, consequently, $q - q' - \bar{q} = 0$.

Case (v). Suppose that $\partial F = S-\mathbf{1}_{f_0}$, so it follows that $(N+\partial F)^{-1} = (N+S)^{-1}(I+\mathbf{1}_{f_0})$, and thus, (2.21c) and the weak sequential continuity of linear operators gives us

$$\begin{aligned} u - u' &= u - (N + S)^{-1}\big[A^*(x + v - 2v') + Nu + f_0\big], \\ \bar{u} \leftarrow u^k - (u^k)' &= u^k - (N + S)^{-1}\big[A^*\big(x^k + 2\big(v^k - (v^k)'\big) - v^k\big) + Nu^k + f_0\big] \\ &\rightharpoonup u - (N + S)^{-1}\big[A^*(x + 2\bar{v} - v) + Nu + f_0\big], \end{aligned}$$

implying, together with $v - v' - \bar{v} = rA(\bar{u} + u' - u)$ which follows from (3.10), (3.11) and (3.12), that

$$u - u' - \bar{u} = 2(N + S)^{-1}A^*(\bar{v} + v' - v) = 2r(N + S)^{-1}A^*A(u - u' - \bar{u}). \tag{3.15}$$

Applying $(N + S)$ on both sides, using $(N - rA^*A)(u - u' - \bar{u}) = 0$ (see (3.10)) and rearranging, this turns out to be equivalent to

$$(S - rA^*A)(u - u' - \bar{u}) = 0.$$

As injectivity according to (3.8) is assumed to hold, we can conclude $u - u' - \bar{u} = 0$. Finally, (3.11) together with (3.12) and (3.10) gives $B(p - p' - \bar{p}) = 0$ and $v - v' - \bar{v} = 0$. Analogously to Case (iv), we also have $L(p - p' - \bar{p}) = (L - rB^*B)(p - p' - \bar{p}) = 0$, hence $p - p' - \bar{p} = 0$ by injectivity of L on $\mathcal{N}(B)$, meaning that $q - q' - \bar{q} = 0$. □

Remark 3.5. *There are several special cases of case* (*i*)*. For example, if $N - rA^*A$ is positive definite and L is injective, or $L - rB^*B$ is positive definite and N is injective, the injectivity property* (3.6) *is satisfied.*

Corollary 3.6. *If one of the spaces X, Y and Z is finite-dimensional, then $I - \mathcal{T}$ is demi-closed.*

Proof. In this situation, one of the operators A and B has to be compact, so Lemma 3.4 is applicable as condition (iii) is satisfied. □

Now we can give the following theorem guaranteeing the weak convergence of the preconditioned ADMM iteration (PADMM).

Theorem 3.7. *Let* (2.5) *be true and let N and L satisfy* (2.19)*. If a solution of the saddle point problem* (2.2) *exists, then the proximal iteration sequence $\{q^k\}$ according to* (2.17) *converges weakly to a fixed-point $q^* = (x^*, u^*, v^*, p^*)$, whenever one of the conditions (i)-(v) in Lemma 3.4 holds. The triple $(u^*, p^*, v^* - x^*)$ is a solution of the saddle point problem* (2.2)*.*

Proof. Denote the set of fixed points of $\mathcal{T}$ by $\mathcal{F}$ which is a closed and convex set. As a solution for (2.2) exists, $\mathcal{F}$ is not empty by Lemma 3.1. For any $q^* \in \mathcal{F}$, the sequence $\{\|q^k - q^*\|_{\mathcal{M}}\}$ is non-increasing by (3.4) in Lemma 3.2 and hence possesses a limit $d(q^*)$. This defines a non-negative function on $\mathcal{F}$.

According to Lemma 3.3 (i), $\{q^k\}$ is bounded. Now, each subsequence possesses a weakly convergent subsequence $\{q^{k_i}\}$ with limit q^* which must be a fixed point of $\mathcal{T}$: From Lemma 3.3 (ii) follows that $(I - \mathcal{T})q^{k_i} \to 0$, hence demi-closedness in Lemma 3.4 implies $q^* = \mathcal{T}q^*$.

Next, let q^{**} be another weak accumulation point of $\{q^k\}$, i.e., $q^{l_i} \rightharpoonup q^{**}$ as $i \to \infty$ for some index sequence $\{l_i\}$. The limit is also a fixed point, hence $\|q^k - q^{**}\|_{\mathcal{M}} \to d(q^{**})$ monotonically non-increasing as $k \to \infty$. We compute

$$2\langle q^k, q^{**} - q^* \rangle_{\mathcal{M}} = \|q^k - q^*\|_{\mathcal{M}}^2 - \|q^k - q^{**}\|_{\mathcal{M}}^2 - \|q^*\|_{\mathcal{M}}^2 + \|q^{**}\|_{\mathcal{M}}^2$$

and observe that the right-hand side converges as $k \to \infty$ to some $c \in \mathbf{R}$. Plugging in $\{q^{k_i}\}$ and $\{q^{l_i}\}$ on the left-hand side implies, by weak convergence that $\langle q^*, q^{**} - q^* \rangle_{\mathcal{M}} = \langle q^{**}, q^{**} - q^* \rangle_{\mathcal{M}}$. Consequently, $\|q^* - q^{**}\|_{\mathcal{M}} = 0$. Using that both q^* and q^{**} are fixed-points yields, by virtue of (3.5):

$$\|q^* - q^{**}\| = \|\mathcal{T}q^* - \mathcal{T}q^{**}\| \le C\|q^* - q^{**}\|_{\mathcal{M}} = 0.$$

Thus, $q^{**} = q^*$ so q^* is the only weak accumulation point. As each subsequence of $\{q^k\}$ possesses a weakly convergent subsequence, we get $q^k \rightharpoonup q^*$ for the whole sequence. □

Corollary 3.8. *In the situation of Theorem 3.7, $Au^k + Bp^k \to c$ as $k \to \infty$. If A is compact, then the primal iterates $\{u^k\}$ converge strongly. If B is compact, the dual iterates $\{p^k\}$ converge strongly.*

Proof. By Theorem 3.7, we already have weak convergence of the sequence $\{q^k\}$ generated by (2.17) to a limit q^* as well as $q^k - q^{k+1} \to 0$ as $k \to \infty$. In particular, $v^k - v^{k+1} \to 0$ and $p^k - p^{k+1} \to 0$ as $k \to \infty$, implying by virtue of (2.21b) that

$$\lim_{k\to\infty} \frac{r}{2}[Au^k + Bp^k - c] = \lim_{k\to\infty} v^{k+1} - v^k + rB(p^k - p^{k+1}) = 0,$$

or, equivalently, $Au^k + Bp^k \to c$ as $k \to \infty$.

If A is compact, A^* is also compact and (2.21c) can be rewritten to

$$u^{k+1} = (rA^*A + \partial F)^{-1}\big[A^*(x^k + v^k - 2v^{k+1} + rAu^{k+1}) + N(u^k - u^{k+1})\big]$$

with $(rA^*A + \partial F)^{-1}$ being continuous and its argument strongly convergent due to complete continuity of A^*. Consequently, also the sequence $\{u^k\}$ converges strongly. If B is compact, an analog argument applied to the update (2.21a) leads to strong convergence of $\{p^k\}$. □

Corollary 3.9. *If X and Y are finite-dimensional, then the primal and dual variables $\{u^k\}$ and $\{p^k\}$ converge strongly, respectively. If Z is finite-dimensional, all variables converge strongly.*

Proof. Again, if X and Y are finite-dimensional, A and B are compact, respectively, so the first statement follows by Corollary 3.8. If Z is finite-dimensional, both A and B are compact, so $\{u^k\}$ and $\{p^k\}$ converge strongly, once again by Corollary 3.8. Additionally, $\{v^k\}$, $\{x^k\}$ converge strongly due to equivalence of weak and strong convergence in finite-dimensional spaces, yielding the second statement. □

Remark 3.10. *As a consequence of the corollaries, it can be seen that Theorem 3.7 recovers in particular the convergence results of majorized ADMM or (unrelaxed) generalized ADMM that were discussed in finite dimensional spaces only in [14, 17, 18, 20]. Moreover, choosing the preconditioners according to $N = rA^*A + \frac{1}{r}I$ and $L = rB^*B + \frac{1}{r}I$ allows to recover the proximal ADMM method in infinite-dimensional Hilbert spaces as studied in [21]. Here, Theorem 3.7 leads to weak convergence as condition (i) in Lemma 3.4 is satisfied, see Remark 3.5.*

4 Feasible Preconditioners and the Relation between PADMM and PDRQ

4.1 Feasible Preconditioners

We are now interested in finding suitable preconditioners N in the case F is linear-quadratic. For this purpose, it is assumed, throughout this section that $\partial F = S - \mathbf{1}_{f_0}$

according to (3.7). Furthermore, let $S + rA^*A$ be positive definite, see (2.7). Finally, let us introduce the notation $T = S + rA^*A$ and $M = N + S$. Also here and in the following, we always assume $L = rB^*B$ with $B = -I$ and $c = 0$ for convenience, i.e., the spaces satisfy $Y = Z$ and the constraint reads as $Au = p$.

A feasible preconditioner for T then is defined as a linear, bounded, self-adjoint and positive definite operator M that satisfies the conditions of Theorem 3.7, i.e.,

$$M \geq T \qquad \text{and} \qquad \begin{cases} A \text{ compact or} \\ \mathcal{N}\begin{pmatrix} M - T \\ M - 2S \end{pmatrix} = \{0\}. \end{cases} \tag{4.1}$$

It is easily checked that $M \geq T$ and the injectivity condition are equivalent to (2.19) and (3.8), respectively. Hence, Theorem 3.7 is applicable, implying (weak) convergence of the iteration.

The reason for introducing M and T is that the update step for u^k, see (2.21c), becomes, as announced in the introduction (see (2.13)),

$$u^{k+1} = u^k + M^{-1}[A^*(rp^k - \lambda^k) + f_0 - Tu^k] \tag{4.2}$$

which corresponds to one step of a splitting method for the solution of $Tu^{k+1} = b^k$ with $b^k = A^*(rp^k - \lambda^k) + f_0$ with respect to the splitting $T = M - (M - T)$, i.e., writing T as the difference of a positive definite operator and a positive semi-definite operator. In case K is not compact (which necessarily refers to the infinite-dimensional setting), the injectivity condition in (4.1) has to be checked separately. An alternative way is ensuring convergence by dampening the iteration in an appropriate manner.

Proposition 4.1. *Let M be linear, bounded, self-adjoint, positive definite.*

(i) For $M \geq T$ and $\alpha \in (0, 1)$, the relaxed update step

$$u^{k+1} = (1 - \alpha)u^k + \alpha\big(u^k + M^{-1}[A^*(rp^k - \lambda^k) + f_0 - Tu^k]\big) \tag{4.3}$$

corresponds to the application of the feasible preconditioner $M_\alpha = \frac{1}{\alpha}M$.

(ii) For $M \geq T_\alpha$, $T_\alpha = \alpha I + T$ and $\alpha > 0$, the update step

$$u^{k+1} = u^k + M^{-1}[\alpha u^k + A^*(rp^k - \lambda^k) + f_0 - T_\alpha u^k] \tag{4.4}$$

corresponds to a feasible preconditioner.

Proof. In the case (i), by positive definiteness, $M \geq cI$ for some $c > 0$ and since $\alpha^{-1} > 1$, $M_\alpha - T \geq c(\frac{1}{\alpha} - 1)I$. Consequently, $M_\alpha > T$ and (4.1) is satisfied. The update procedure (4.2) with M_α then corresponds to (4.3). In the case (ii), we have $M - T \geq \alpha I$ as $M \geq T_\alpha$, leading to (4.1). The representation (4.4) follows immediately. □

Therefore, we restrict ourselves in the following to preconditioners M which are positive definite and satisfy $M \geq T$ as these always lead to convergent algorithms (where, possibly, a dampened update according to (4.3) or (4.4) has to be performed for u).

Preconditioner	T	μI	$(1+\mu)D$	M_{SGS}	M_{SSOR}
Conditions	—	$\mu \geq \|S\| + r\|A\|^2$	$\mu \geq \max \sigma_{T-D}$	—	$\omega \in (0,2)$
Iteration type	ADMM	Richardson	Damped Jacobi	SGS	SSOR

Table 1: Choices for different preconditioners for $T = S + rA^*A$.

Let us discuss preconditioners which fulfill the latter assumptions. This is obviously the case for $M = T$ which corresponds to the ADMM iteration without preconditioning. On the other extreme, the Richardson preconditioner $M = \mu I$ satisfies $M \geq T$ if

$$\mu \geq \|S\| + r\|A\|^2$$

leading to a convergent iteration with an almost trivial update step for u.

Remark 4.2. *In case of nonlinear ∂F, we may choose $N = \mu I$ resulting in resolvents of ∂F appearing in the update for u:*

$$u^{k+1} = (\mu I + \partial F)^{-1}[\mu u^k + A^*(rp^k - \lambda^k - rAu^k)].$$

If A is compact, the iteration converges for $\mu \geq r\|A\|^2$, otherwise, $\mu > r\|A\|^2$ leads to weak convergence, see Theorem 3.7. The whole iteration (PADMM) *then corresponds to a* split inexact Uzawa method*, see [15, 26], and is equivalent to the primal-dual method in [13] (where this special case is called preconditioned ADMM). In this light,* (PADMM) *may be interpreted as a generalization of all of these methods.*

Also note that $N = 0$ corresponds to the original (ADMM) *which is equivalent to the* split Bregman method *[27, 28]. Consequently, our convergence analysis is trivially also applicable for the latter.*

In finite dimensions, T can be represented by a matrix and classical matrix splitting methods are applicable. Denote by D the diagonal of T and consider the *damped Jacobi method* $M = (1 + \mu)D$ which can be shown to satisfy $M \geq T$ if μ is at least the greatest eigenvalue of $T - D$, i.e., $\mu \geq \max \sigma_{T-D}$ with σ_{T-D} denoting the spectrum of $T - D$, see [1]. Using SGS and SSOR to abbreviate the *symmetric Gauss-Seidel* and the *symmetric successive over-relaxation* method for T, respectively, the associated preconditioners M_{SGS} and M_{SSOR} (for an over-relaxation parameter $\omega \in (0, 2)$) may be considered. These are always feasible, i.e., $M_{SGS} \geq T$ as well as $M_{SSOR} \geq T$, again see [1]. In conclusion, we can give the Table 1 which summarizes the properties for the different preconditioners and may help making a specific choice.

For multiple iterations of the preconditioner, we have the same proposition as in [1] which means one can use any number of iterations of a feasible preconditioner without losing feasibility.

Proposition 4.3. *Let M be a feasible preconditioner for T and $n \geq 1$. Then, applying the preconditioner n times, i.e.,*

$$\begin{cases} u^{k+(i+1)/n} := u^{k+i/n} + M^{-1}[A^*(rp^k - \lambda^k) + f_0 - Tu^{k+i/n}], \\ \qquad\qquad i = 0, \dots, n-1, \end{cases}$$

corresponds to $u^{k+1} = u^k + M_n^{-1}[A^*(rp^k - \lambda^k) + f_0 - Tu^k]$ *where* M_n *is a feasible preconditioner.*

For more sufficient conditions for feasibility, we once more refer to [1].

4.2 The Relation to PDRQ

Next, we would like to establish connections between the preconditioned ADMM iteration and the preconditioned Douglas–Rachford iteration for linear-quadratic functionals PDRQ which was introduced in [1, Table 3.1]. We employ this method with the following data:

$$Q = S, \quad f = -f_0, \quad R = 0, \quad g = 0, \quad \sigma = r, \quad \mu = 1, \quad \lambda = 0$$

and denote by $T' = \sigma T$ as well as $M' = \sigma M$. Taking into account that G corresponds to the Fenchel dual G^* in [1], the PDRQ iteration then reads as

$$\begin{cases} b^k = r(f_0 - A^*\bar{y}^k), \\ x^{k+1} = x^k + (M')^{-1}(b^k - T'x^k), \\ y^{k+1} = \bar{y}^k + rAx^{k+1}, \\ \bar{y}^{k+1} = \bar{y}^k + (I + r\partial G^*)^{-1}[2y^{k+1} - \bar{y}^k] - y^{k+1}, \end{cases} \tag{4.5}$$

which can be reformulated as follows: Using Moreau's identity yields $(I + r\partial G^*)^{-1} = I - (I + (r\partial G^*)^{-1})^{-1}$ such that with the substitutions

$$p^k = \tfrac{1}{r}(y^k - \bar{y}^k), \quad \lambda^k = y^k, \quad u^k = x^k$$

the update of p^k can be written as

$$p^{k+1} = \frac{1}{r}(I + (r\partial G^*)^{-1})^{-1}(2y^{k+1} - \bar{y}^k) = (rI + \partial G)^{-1}(2y^{k+1} - \bar{y}^k).$$

Plugging in the substitutions, we get $\lambda^{k+1} = \lambda^k + r(Au^{k+1} - p^k)$ as well as $2y^{k+1} - \bar{y}^k = \lambda^{k+1} + rAu^{k+1}$ and $-\bar{y}^k = \lambda^k - rp^k$. Furthermore, we observe that the factor σ cancels in the update for u^k. In total, the iteration (4.5) becomes

$$\begin{cases} u^{k+1} = u^k + M^{-1}\big(A^*(\lambda^k - rp^k) + f_0 - Tu^k\big), \\ \lambda^{k+1} = \lambda^k + r(Au^{k+1} - p^k), \\ p^{k+1} = (rI + \partial G)^{-1}(\lambda^{k+1} + rAu^{k+1}), \end{cases} \tag{4.6}$$

which converges weakly as soon as $M \geq T$ by weak convergence of the PDRQ method.

With the formulation (4.6), the connection to (PADMM) with preconditioner (4.2) for linear-quadratic F becomes immediate: While the actual update steps are the same, (PADMM) performs updates in the order

$$u^{k+1} \to p^{k+1} \to \lambda^{k+1},$$

while PDRQ corresponds to the order

$$u^{k+1} \to \lambda^{k+1} \to p^{k+1},$$

which can be identified as a preconditioned ADMM with u and p interchanged (also see [29] for a general discussion of this phenomenon). Hence, PDRQ can alternatively be expressed as

$$\begin{cases} u^{k+1} = \arg\min_{u \in X} \ \mathcal{L}_r(u, p^k, \lambda^k) + \frac{1}{2}\|u - u^k\|^2_{N - rA^*A}, \\ \lambda^{k+1} = \lambda^k + r(Au^{k+1} - p^k), \\ p^{k+1} = \arg\min_{p \in Y} \ \mathcal{L}_r(u^{k+1}, p, \lambda^{k+1}), \end{cases} \tag{PDRQ-ALM}$$

where $\mathcal{L}_r$ is given by (2.3) with F according to (2.6) and $\|\cdot\|^2_{N-rA^*A}$ is defined as for (2.9). Despite the similarity, note that u and p play different roles in the original formulation (2.1): The variable u is associated with an operator A while for p, only the identity is involved. Consequently, interchanging u and p has a qualitative effect on the algorithm. This seems to be particularly relevant in the case of infinite-dimensional Hilbert spaces X and Y, where weak convergence of (PDRQ-ALM) can be established under weaker conditions compared to (PADMM) as none of the conditions in Lemma 3.4 is required to establish the demi-closedness for the corresponding $I - \mathcal{T}$, see [1].

Remark 4.4. *Similar substitutions used to derive* (4.6) *from the PDRQ algorithm can also be used to obtain an alternate form of* (PADMM)*:*

$$\begin{cases} x^{k+1} = (N + \partial F)^{-1}[Nx^k - A^*\bar{y}^k], \\ y^{k+1} = (I + r\partial G^*)^{-1}[y^k + rAx^{k+1}], \\ \bar{y}^{k+1} = 2y^{k+1} - y^k, \end{cases} \tag{PADMM-PD}$$

where in case of linear quadratic F according to (2.6)*, the update for x^k can be written as*

$$x^{k+1} = x^k + M^{-1}[f_0 - A^*(\bar{y}^k + rAx^k) - Tx^k].$$

Note that as in (4.5)*, the resolvent of the subgradient of the dual functional G^* is involved. As this corresponds to the general method presented in [30], we refer to this equivalent iteration as* preconditioned ADMM in primal-dual form.

5 Application to Total-Variation regularized Problems

We would like to apply (PADMM) for linear-quadratic F for the solution of discrete total-variation denoising problems with L^2- and L^1-type discrepancy. For the discretization, please refer to Section 4.1 of [1] or Section 4 of [16] for details.

5.1 The Minimization Problems, Preconditioners and Algorithms

Let us apply the discrete framework for the total-variation regularized L^2- and L^1-type denoising problems (see [31] for the L^2-case which is usually called the *ROF model*)

$$\min_{u \in X} \frac{1}{p}\|u - f\|_p^p + \alpha\|\nabla u\|_1, \qquad \text{for } p = 1, 2 \tag{5.1}$$

where $f : \Omega \to \mathbf{R}$ is a given noisy image and $\alpha > 0$ is a regularization parameter.

The case $p = 2$. We see that for $p = 2$, (5.1) is equivalent to (2.1) with F according to (2.6) and data

$$A = \nabla, \quad S = I, \quad f_0 = f, \quad G = \alpha\|\cdot\|_1.$$

Note that a unique minimizer u of (5.1) exists. As we are in finite dimensions, both F and $\alpha\|\cdot\|_1$ are continuous, so one can employ Fenchel–Rockafellar duality to obtain existence of solutions for the associated saddle-point problem (2.2) [22, 23, 32]. Furthermore, observe that $S + rA^*A = I - r\Delta > 0$ for any $r > 0$. Thus, preconditioned ADMM algorithms can be derived. In order to implement the algorithm, we note that for $(rI + \partial G)^{-1}$, we have

$$(rI + \partial G)^{-1} = \Big(I + \frac{1}{r}\partial G\Big)^{-1}\Big(\frac{1}{r}\,\cdot\Big), \tag{5.2}$$

and since $G = \alpha\|\cdot\|_1$, the latter resolvent is given by the soft-shrinkage operator $\mathcal{S}_{\alpha/r}(\,\cdot\,/r)$, see [33]. This can in turn be expressed, for $p \in Y$, by

$$\mathcal{S}_{\alpha/r}(p/r) = \frac{\max(0, |p| - \alpha)}{r|p|}p \tag{5.3}$$

where $|p| = \sqrt{(p_1)^2 + (p_2)^2}$. We still have to choose a preconditioner for $T = I - r\Delta$. This will discussed after the next paragraph.

The case $p = 1$. In this case, the discrepancy $\|\cdot - f\|_1$ is not linear-quadratic and we have to reformulate (5.1) to

$$\min_{u,v \in X,\ p \in Y} \|v - f\|_1 + \alpha\|p\|_1 \qquad \text{subject to} \qquad \begin{bmatrix} I \\ \nabla \end{bmatrix} u = \begin{bmatrix} v \\ p \end{bmatrix}.$$

This leads to $F = 0$ which obviously satisfies (2.6). In total, the problem has the pair (v, p) as dual variable and the data

$$A = \begin{bmatrix} I \\ \nabla \end{bmatrix}, \quad S = 0, \quad f_0 = 0, \quad G(v, p) = \|v - f\|_1 + \alpha\|p\|_1.$$

Again, $S + rA^*A = r(I - \Delta) > 0$ for each $r > 0$, so preconditioned ADMM is applicable. The resolvent $(rI + G)^{-1}$ then decouples into

$$(rI + \partial G)^{-1}(v, p) = \big(\mathcal{S}_{1/r}(v - rf) + f, \mathcal{S}_{\alpha/r}(p/r)\big) \tag{5.4}$$

where the soft-shrinkage operator on X is involved which also reads as (5.3) with $|\cdot|$ denoting the usual (pointwise) absolute value. Here, a preconditioner for $T = r(I - \Delta)$ has to be chosen.

The preconditioners. Observe that in both cases, M is required to be a feasible preconditioner for operators of type $T = sI - r\Delta$ for $s > 0$ where $\Delta = \operatorname{div} \nabla$ can be interpreted as a discrete Laplace operator with homogeneous Neumann boundary conditions. In other words: solving $Tu = b$ corresponds to a discrete version of the boundary value problem

$$\begin{cases} su - r\Delta u = b, \\ \dfrac{\partial u}{\partial \nu}|_{\partial\Omega} = 0. \end{cases} \tag{5.5}$$

In particular, T can be represented as the application of a five-point finite-difference stencil involving, besides the center point, only neighboring points from left, right, above and below. The equation $Tu = b$ can therefore be solved with a Red-Black Gauss–Seidel method whose application is easily parallelizable with respect to each grid point. In order to give a feasible preconditioner, the update step has to be symmetric, for this reason, the symmetric Red-Black Gauss–Seidel method (SRBGS) has to be used, see Table 1. Furthermore, n steps of SRBGS are still feasible, see Proposition 4.3, i.e., corresponding to a symmetric M which satisfies $M \geq T$. We denote the n-fold application of the symmetric Red-Black Gauss–Seidel method to the initial guess u and right-hand side b by

$$\mathrm{SRBGS}^n_{r,s}(u, b) = (I + M^{-1}(\mathbf{1}_b - T))^n u \tag{5.6}$$

making it again explicit that M and T depend on s and r. For details concerning the efficient realization of this preconditioner, we refer to [16, Section 4.1.3].

The algorithms. The building blocks (5.3), (5.4) and (5.6) eventually give all the needed ingredients for performing (PADMM) to solve (5.1) for both $p = 2$ and $p = 1$. The algorithm for the L^2-case is summarized in Table 2 while Table 3 presents the L^1-case.

Remark 5.1. *Note that the equivalent iteration* (PADMM-PD) *could also be used. This involves, in case $p = 2$, the resolvent $(I + r\partial G^*)^{-1}$ associated with the Fenchel dual*

$$G^*(p) = \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha\}}(p) = \begin{cases} 0 & \text{if } \|p\|_\infty \leq \alpha \\ \infty & \text{else,} \end{cases}$$

and can be seen to correspond to the projection operator

$$(I + r\partial G^*)^{-1}(p) = \mathcal{P}_{\{\|\cdot\|_\infty \leq \alpha\}}(p) = \frac{p}{\max(1, |p|/\alpha)}.$$

The case $p = 1$ follows completely analogous. While being equivalent iterations, we found in our numerical experiments that the (PADMM-PD) *versions admit a slight speed advantage compared to their* (PADMM) *counterparts. However, as they can easily be derived from Tables 2 and 3, we do not explicitly state them.*

PADMM	**Objective:** L^2-TV denoising $\quad \min_{u\in X}\ \frac{1}{2}\|u-f\|_2^2 + \alpha\|\nabla u\|_1$	
Initialization:	$(u^0, p^0, \lambda^0) \in X \times Y \times Y$ initial guess, $r > 0$ step-size, $n \geq 1$ inner iterations for symmetric Gauss–Seidel	
Iteration:	$u^{k+1} = \mathrm{SRBGS}^n_{1,r}\big(u^k, \mathrm{div}(\lambda^k - rp^k) + f\big)$	according to (5.6)
	$p^{k+1} = \mathcal{S}_{\alpha/r}\big((\lambda^k + r\nabla u^{k+1})/r\big)$	according to (5.3)
	$\lambda^{k+1} = \lambda^k + r(\nabla u^{k+1} - p^{k+1})$	

Table 2: The preconditioned ADMM iteration for L^2-TV denoising.

PADMM	**Objective:** L^1-TV regularization $\quad \min_{u\in X}\ \|u-f\|_1 + \alpha\|\nabla u\|_1$	
Initialization:	$(u^0, v^0, p^0, \lambda_v^0, \lambda_p^0) \in X \times (X \times Y)^2$ initial guess, $r > 0$ step-size, $n \geq 1$ inner iterations for symmetric Gauss–Seidel	
Iteration:	$b^k = rv^k - \lambda_v^k + \mathrm{div}(\lambda_p^k - rp^k)$	
	$u^{k+1} = \mathrm{SRBGS}^n_{r,r}(u^k, b^k)$	according to (5.6)
	$v^{k+1} = \mathcal{S}_{1/r}\big((\lambda_v^k + r(u^{k+1} - f))/r\big) + f$	according to (5.3)
	$p^{k+1} = \mathcal{S}_{\alpha/r}\big((\lambda_p^k + r\nabla u^{k+1})/r\big)$	according to (5.3)
	$\lambda_v^{k+1} = \lambda_v^k + r(u^{k+1} - v^{k+1})$	
	$\lambda_p^{k+1} = \lambda_p^k + r(\nabla u^{k+1} - p^{k+1})$	

Table 3: The preconditioned ADMM iteration for L^1-TV denoising.

Remark 5.2. *The application of the preconditioned ADMM methods are, of course, not limited to denoising problems. For instance, TV-regularized inverse problems may be considered, i.e.,*

$$\min_{u\in X} \frac{1}{p}\|Lu - f\|_p^p + \alpha\|\nabla u\|_1$$

for $L : X \to Z$ a linear operator and $p \in \{1, 2\}$. While ADMM is generally applicable here, see for instance [4], deblurring problems turn out to be particularly suitable for efficient preconditioning as they are diagonalizable via the Fourier transform. Such approaches has been pursued for the PDRQ method in [16] which lead to competitive algorithms for TV-regularized L^2- and L^1-deblurring. Due to the close structural similarity between PDRQ and (PADMM), *see Subsection 4.2, one may expect a similar efficiency for the respective preconditioned ADMM methods.*

6 Numerical Experiments

6.1 L^2-TV Denoising Problems

We now illustrate the efficiency of the methods proposed in Section 5 for the solution of the L^2-TV denoising problem (5.1), i.e., for the choice $p = 2$. The following four algorithms are tested for the ROF model,

- ALG2: $\mathcal{O}(1/k^2)$ accelerated primal-dual algorithm introduced in [13] with adaptive step sizes, $\tau_0 = 1/L$, $\tau_k \sigma_k L^2 = 1$, $\gamma = 0.7$ and $L = 8$. Here we use the same notations for the parameters τ, σ together with θ as in [13] throughout this section.

- ADMM: Alternating direction method of multipliers (ADMM). In each iteration step, the linear subproblem (5.5) is solved by diagonalization techniques via the *discrete cosine transform* (DCT).

- PADMM: Preconditioned alternating direction method of multipliers according to Table 2. A symmetric Red-Black Gauss–Seidel preconditioner is employed for the linear subproblem (5.5).

- PDRQ: Preconditioned Douglas-Rachford method for pure quadratic-linear primal (or dual) functionals in [1]. Again, a Symmetric Red-Black Gauss–Seidel preconditioner is employed for the corresponding subproblems (5.5).

Note that the ADMM and PADMM algorithms have been realized in their (PADMM-PD) versions. We use the primal-dual gap

$$\mathfrak{G}_{L^2\text{-}\mathrm{TV}}(u,p) = \frac{\|u-f\|_2^2}{2} + \alpha\|\nabla u\|_1 + \frac{\|\operatorname{div} p + f\|_2^2}{2} - \frac{\|f\|_2^2}{2} + \mathcal{I}_{\{\|p\|_\infty \le \alpha\}}(p) \tag{6.1}$$

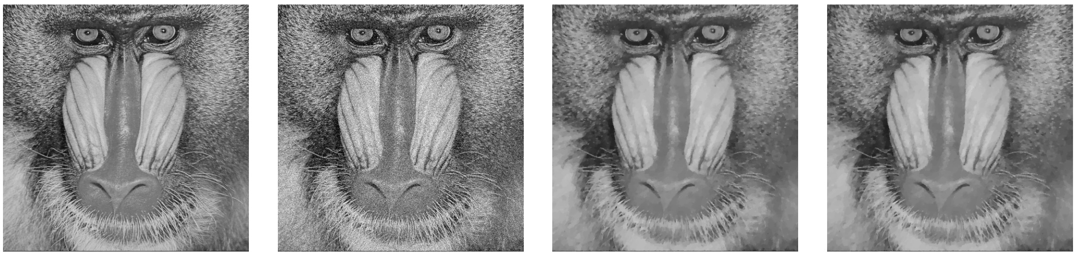
as the basis for comparison as well as a stopping criterion by plugging, in each iteration step, a primal-dual test pair $(u_{\text{test}}^k, p_{\text{test}}^k)$ and normalizing by the number of pixels. The iteration is stopped once this normalized primal-dual gap is below a specified tolerance $\varepsilon > 0$, i.e., when $\mathfrak{G}_k = \mathfrak{G}_{L^2\text{-}\mathrm{TV}}(u_{\text{test}}^k, p_{\text{test}}^k)/(N_x N_y) \le \varepsilon$. This guarantees that u_{test}^k is ε-optimal for (5.1). While the choice of the primal-dual test pair is clear for ALG2 and PDRQ from [13] and [1], respectively, the corresponding pair for ADMM and PADMM is given by $(u_{\text{test}}^k, p_{\text{test}}^k) = (u^k, \lambda^k)$. Note that $p^k \in \partial\mathcal{I}_{\{\|\cdot\|_\infty \le \alpha\}}(\lambda^k)$ for each $k \ge 1$, so, in particular, $\|p_{\text{test}}^k\|_\infty \le \alpha$ for each $k \ge 1$. By convergence of the iterations and since $\mathfrak{G}_{L^2\text{-}\mathrm{TV}}$ is continuous on its domain, we get convergence $\mathfrak{G}_k \to 0$ as $k \to \infty$ for the ADMM and PADMM algorithms.

We note that all algorithms in this paper were implemented in Matlab 2012b and executed on a workstation running a 64 Bit system with 8 cores each at 3.40GHz.

Table 4 summarizes the numerical results for the ROF image denoising problem with the above four algorithms. Computations were performed for the image “baboon” (size 512×512 pixels), additive Gaussian noise (level 0.1) and different regularization parameters α in (5.1). Figure 1 includes the test images, noisy images and denoised images with different stopping tolerances. Additionally, Figure 2 gives a more detailed overview of

	$\alpha = 0.1$		$\alpha = 0.3$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
ALG2	18 (0.27s)	81 (1.28s)	67 (1.02s)	334 (4.99s)
ADMM	(3) 9 (0.37s)	(3) 43 (1.84s)	(9) 15 (0.61s)	(9) 149 (6.31s)
PADMM	(3,2) 9 (0.17s)	(3,2) 42 (0.83s)	(9,2) 24 (0.47s)	(9,2) 153 (2.99s)
PDRQ	(3,2) 8 (0.15s)	(3,2) 43 (0.78s)	(9,2) 24 (0.44s)	(9,2) 157 (2.77s)

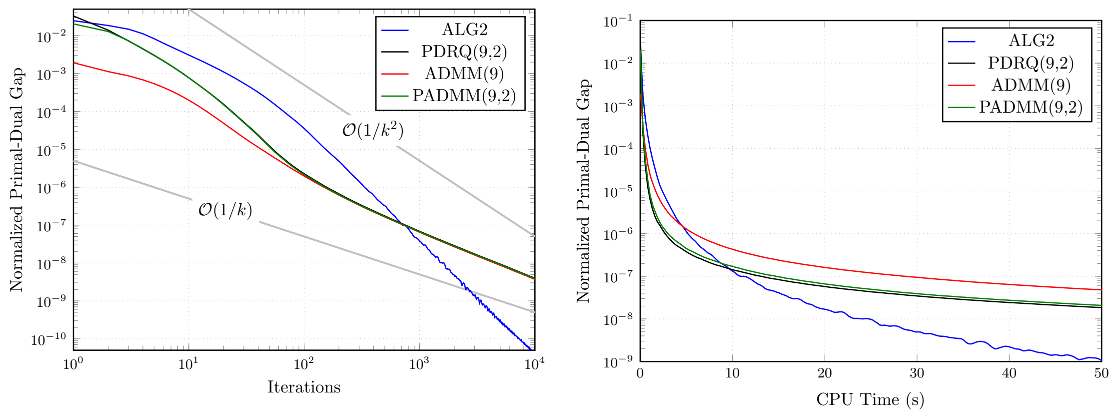
Table 4: Numerical results for the L^2-TV image denoising (ROF) problem (5.1) with noise level 0.1 and regularization parameters $\alpha = 0.1$, $\alpha = 0.3$. For algorithms PADMM and PDRQ, we use the two pairs $\{(r, n)\ k(t)\}$ to represent the iteration number k, CPU time t with constant step size r and inner iteration number n. The iteration is performed until the normalized primal-dual gap (6.1) is below ε.



(a) Original image (b) Noisy (level: 10%) (c) PADMM(10^{-4}) (d) PADMM(10^{-6})

Figure 1: Image denoising with random Gaussian noise. (a) shows the 512×512 input Baboon image and (b) is a noisy image which has been corrupted by 10% random Gaussian noise. (c) and (d) are the denoised images by the PADMM algorithm with normalized primal energy less than 10^{-4} and 10^{-6} for $\alpha = 0.1$.

the iteration numbers and CPU time cost for each algorithm. One can see from the results that in terms of iteration numbers, ADMM, PADMM and PDRQ admit nearly the same performance (which can be estimated to be $\mathcal{O}(1/k)$). However, as the symmetric Gauss–Seidel updates in PADMM and PDRQ are less complex compared to the DCT in ADMM, a speed advantage of factor ≈ 2 in terms of CPU time can be observed. Here, PDRQ can be observed to be slightly faster than PADMM, presumably due to the internal realization in Matlab. Compared to ALG2 which possesses an convergence rate of $\mathcal{O}(1/k^2)$ (and, in fact, actually appears to be even faster), these methods all suffer from slower asymptotic performance. Nevertheless, if one targets a moderate accuracy up to $\approx 10^{-6}$, the preconditioned iterations PADMM and PDRQ are competitive. For the L^2-TV denoising problem, the obtained solutions can hardly visually be distinguished from more accurate solutions, see, for example, Figure 1 for a comparison between accuracy of 10^{-4} and 10^{-6}.



(a) Numerical convergence rate of normalized error of energy compared with iteration number.

(b) Numerical convergence rate of normalized error of energy compared with iteration time.

Figure 2: L^2-TV denoising: Numerical convergence rates. The normalized primal-dual gap (6.1) is compared in terms of iteration number and computation time for $\alpha = 0.3$. The notations ADMM(r), PADMM(r, n) and PDRQ(r, n) are used to indicate the step-size r and n inner iterations. Note the double-logarithmic and semi-logarithmic scale, respectively, which is used in the plots.

6.2 L^1-TV Denoising

We performed experiments on the numerical solution of the L^1-TV denoising problem (5.1), i.e., $p = 1$.

- ALG1: $\mathcal{O}(1/k)$ primal-dual algorithm introduced in [13] with constant step sizes, the dual step size $\tau = 0.02$, $\tau\sigma L^2 = 1$, $\theta = 1$ with $L = \sqrt{8}$ as proposed in [13].
- ADMM: Alternating direction method of multipliers (ADMM). Again, DCT is employed for linear subproblem (5.5) instead of preconditioners.
- PADMM: Preconditioned alternating direction method of multipliers. A symmetric Red-Black Gauss–Seidel preconditioner is employed for the linear subproblem (5.5).
- PDRQ: Preconditioned Douglas-Rachford method for pure quadratic-linear primal (or dual) functionals in [1]. A symmetric Red-Black Gauss–Seidel preconditioner is employed for the corresponding subproblems (5.5).

Again, ADMM and PADMM have been realized in their (PADMM-PD) versions. We use the normalized primal energy R_k to realize a stopping criterion,

$$R_k := (E_k - E_*)/E_*, \qquad E_k := \|u^k - f\|_1 + \alpha\|\nabla u^k\|_1, \tag{6.2}$$

where E_* represents the primal energy that was obtained by PADMM after roughly 4×10^5 iterations and represents the minimal value among all tested algorithms.

	$\alpha = 1.0$		
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$
ALG1	267 (0.87s)	840 (2.53s)	2180 (6.65s)
ADMM (20)	120 (0.97s)	255 (2.07s)	733 (6.02s)
PADMM (20,2)	135 (0.46s)	288 (0.96s)	765 (2.73s)
PDRQ (20,2)	136 (0.48s)	290 (1.01s)	765 (2.77s)

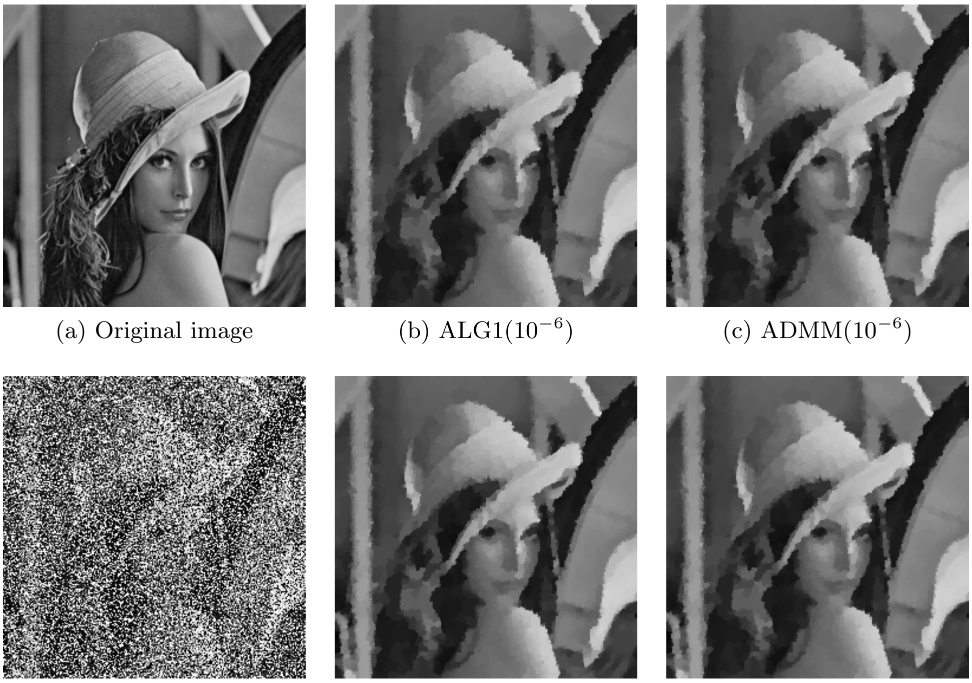
Table 5: L^1-TV denoising: Performance comparison. The results are shown in the format $k(t)$ where k and t denote iteration number and CPU time, respectively. The values (r) and (r, n) for ADMM and PADMM, PDRQ indicate the step-size r and inner iteration number n.

Computations were performed on the 256×256 "Lena" image (Figure 3(a)) corrupted by 60% salt-and-pepper noise (Figure 3(d)) with regularization parameter $\alpha = 1.0$, see Figure 3 for the outcome of the algorithms. Table 5 shows the numerical evaluation with comparison of the iteration numbers and CPU time cost for the four algorithms. Figure 4 depicts a detailed comparison in terms of iteration numbers and CPU time. It can be seen that both ADMM, PADMM, PDRQ and ALG1 effectively admit the asymptotic convergence rate of roughly $\mathcal{O}(1/k^2)$ which is good for L^1-type problems but not covered by the theory. In the example, PDRQ and PADMM show similar behaviour and admit the best performance both with respect to iteration numbers and computation time, achieving a speedup of more than 2 compared to ADMM and ALG1.

7 Conclusions

We proposed and analyzed the preconditioned ADMM method for general non-smooth regularized problems which efficiently deals with the implicit linear equations arising in this context. We moreover establish a practical convergence theory within this framework, allowing for weak convergence in infinite-dimensional Hilbert spaces and strong convergence in finite dimensions. These could make the ADMM more safe and more flexible to use in applications, in particular for large-scale problems in which the subproblems also have to be solved iteratively. We also presented applications for image denoising. The numerical results showed that preconditioned ADMM has the potential to bring out appealing benefits and fast algorithms.

Nevertheless, there are some open questions and possible directions for further investigation. For instance, non-stationary iterative methods such as the CG (conjugate gradient) method are currently not covered by the theory and might be worth investigating. The CG method has already been used in numerical experiments for linear subproblems in ADMM, e.g., for reconstruction in computed tomography (CT) [5]. Establishing convergence rates as well as accelerating preconditioned ADMM to give optimal rates is also a very important topic, see [34, 35] for recent work in this direction. We believe that



(a) Original image (b) ALG1(10^{-6}) (c) ADMM(10^{-6})

(d) Noisy image (60%) (e) PADMM(10^{-6}) (f) PDRQ(10^{-6})

Figure 3: Image denoising with salt-and-pepper noise. (a) shows the 256×256 input image and (d) is a version which has been corrupted by 60% salt-and-pepper noise. (b), (c), (e) and (f) are the denoised images obtained with ALG1, ADMM, PADMM and PDRQ, respectively, with normalized primal energy R_k less than 10^{-6} for $\alpha = 1.0$.

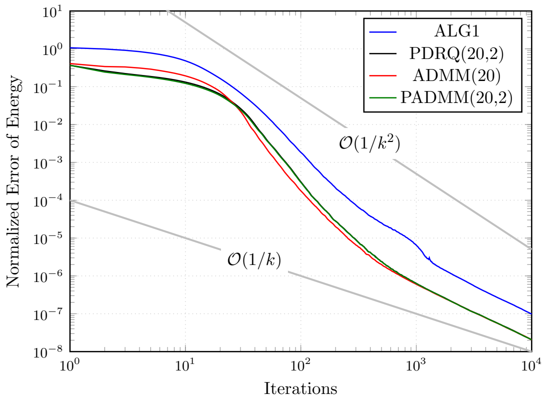
this framework will also be beneficial for applications beyond imaging, for instance, for compressed sensing problems and other large scale problems, provided that appropriate preconditioners could be found.

Acknowledgements

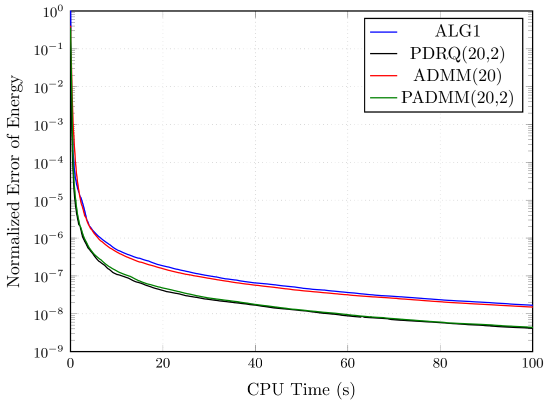
Kristian Bredies and Hongpeng Sun acknowledge support of the Austrian Science Fund (FWF) under grant SFB32 (SFB “Mathematical Optimization and Applications in the Biomedical Sciences”) during the preparation of this manuscript.

References

[1] Bredies, K., Sun, H.: Preconditioned Douglas–Rachford splitting methods for convex-concave saddle-point problems. SIAM J. Numer. Anal., 53(1), 421–444 (2015)



(a) Numerical convergence rate of normalized error of energy compared with iteration number.



(b) Numerical convergence rate of normalized error of energy compared with iteration time.

Figure 4: L^1-TV denoising: Numerical convergence rates. The normalized error of energy is compared in terms of iteration number and computation time for Figure 4 with $\alpha = 1.0$. The notations ADMM(r), PADMM(r, n) and PDRQ(r, n) are used to indicate the step-size r and n inner iterations. Note the double-logarithmic and semi-logarithmic scale, respectively, which is used in the plots.

[2] 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)

[3] Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. Pac. J. Optim. 11(4), 619–644 (2015)

[4] Esser, E.: Application of Lagrangian-based alternating direction methods and connections to split Bregman. CAM Report 09-31, UCLA, Center for Applied Math. (2009)

[5] Ramani, S., Fessler, J. A.: A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. IEEE Trans. Med. Imag. 31(3), 677–688 (2012)

[6] Wu, C., Tai, X. C.: Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. SIAM J. Imaging Sci. 3(3), 300–339 (2010)

[7] Fortin, M., Glowinski, R.: On decomposition-coordination methods using an augmented Lagrangian. In: Fortin, M., Glowinski, R. (eds.): Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems, pp. 97–146. North-Holland, Amsterdam (1983)

[8] Glowinski, R., Marrocco, 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. ESAIM Math. Model. Num. 9(2), 41–76 (1975)

[9] Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.): Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems, pp. 299–331. North-Holland, Amsterdam (1983)

[10] 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)

[11] Mercier, B.: Approximation par éléments finis et résolution par un algorithme de pénalisation-dualité d'un problème d'élasto-plasticité. C.R. Acad. Sci. Math. 280A, 287–290 (1975)

[12] Eckstein, J., Bertsekas, D. P.: On the Douglas–Rachford splitting method and the proximal algorithm for maximal monotone operators. Math. Program. 55, 293–318 (1992)

[13] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. 40, 120–145 (2011)

[14] Li, M., Sun, D., Toh, K. C.: A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. SIAM J. Optim. 26(2), 922–950 (2016)

[15] Zhang X., Burger M., Osher S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. 46(1), 20–46 (2011)

[16] Bredies, K., Sun, H.: Preconditioned Douglas–Rachford algorithms for TV- and TGV-regularized variational imaging problems. J. Math. Imaging Vis. 52(3), 317–344 (2015)

[17] 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)

[18] Fang, E. X., He, B., Liu, H., Yuan, X.: Generalized alternating direction method of multipliers: new theoretical insights and applications. Math. Prog. Comp. 7(2), 149–187 (2015)

[19] Monteiro, R. D. C., Svaiter, B. F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM J. Optim. 23(1), 475–507 (2013)

[20] Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. SIAM J. Optim. 24(1), 269–297 (2014)

[21] Attouch, H., Soueycatt, M.: Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control. Pac. J. Optim. 5(1), 17–37 (2008)

[22] Bauschke, H. H., Combettes, P. L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)

[23] Ito, K., Kunisch, K.: Lagrange Multiplier Approach to Variational Problems and Applications. SIAM, Philadelphia (2008)

[24] Reed, M., Simon, B.: Methods of Modern Mathematical Physics I: Functional Analysis. Academic Press, San Diego (1980)

[25] Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Amer. Math. Soc. 73(4), 591–597 (1967)

[26] 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)

[27] Goldstein, T., Osher, S.: The split Bregman method for L^1 regularized problems. SIAM J. Imaging Sci. 2(2), 323–343 (2009)

[28] Setzer, S.: Split Bregman algorithm, Douglas–Rachford splitting and frame shrinkage. In: Tai, X. C., Mørken, K., Lysaker, M., Lie, K. A. (eds.): Scale Space and Variational Methods in Computer Vision, pp. 464–476. Springer, Berlin Heidelberg (2009)

[29] Yan, M., Yin, W.: Self equivalence of the alternating direction method of multipliers. arXiv eprints, 1407:7400 (2014)

[30] Chambolle A., Pock T.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: Proceedings of the 2011 International Conference on Computer Vision (ICCV), pp. 1762–1769 (2011)

[31] Rudin L., Osher S., Fatemi E.: Nonlinear total variation based noise removal algorithms. Physica D 60, 259–268 (1992)
[32] Eckland, I., Témam, R.: Convex Analysis and Variational Problems. SIAM, Philadelphia, (1999)
[33] Combettes, P. L., Wajs, V. R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Sim. 4(4), 1168–1200 (2005)
[34] Bredies, K., Sun, H.: Accelerated Douglas–Rachford methods for the solution of convex-concave saddle-point problems. arXiv eprints, 1604:06282 (2016)
[35] Xu, Y.: Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. arXiv eprints, 1606:09155 (2016)