SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Preconditioned Alternating Direction Method of Multipliers for the Minimization of Quadratic plus Non-Smooth Convex Functionals

K. Bredies H. Sun

SFB-Report No. 2015-006 April 2015

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

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

Preconditioned Alternating Direction Method of Multipliers for the Minimization of Quadratic plus Non-Smooth Convex Functionals

Kristian Bredies   Hongpeng Sun*

April 8, 2015

Abstract

We propose a preconditioned ADMM (alternating direction method of multipliers) for non-smooth regularized problems. ADMM is a popular first-order method for general constrained optimization problems. However, it suffers from solving implicit subproblems, especially linear subproblems in various applications. In this paper we derive a preconditioned ADMM method by writing ADMM and its preconditioned version as a new kind of proximal point method directly on the primal problem. The weak (or strong) convergence in infinite (or finite) dimensional space is proved. Various efficient preconditioners with any number of inner iterations may be used in the preconditioned ADMM framework. Connections between the preconditioned ADMM and the PDRQ in [3] for general non-smooth problems are established. The methods are applied to total variation denoising problems and their benefits are shown in numerical experiments.

1 Introduction

The *alternating direction method of multipliers* (ADMM) is a widely used technique for the solution of large scale optimization problems [2,10]. Although developed in the 1970s, it regained popularity in the past several years in various fields of applied mathematics, in particular including variational imaging [11, 21, 25]. It is appropriate for solving the following kind of convex optimization problem:

$$\min_{u \in X} \; F(u) + G(Ku). \tag{1.1}$$

Here, X and Y are real Hilbert spaces, $K \in \mathcal{L}(X, Y)$, where $\mathcal{L}(X_1, X_2)$ denotes all the linear and continuous mapping between Hilbert spaces X_1 and X_2, and $\mathcal{L}(X_1)$ is used if

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

$X_2 = X_1$. The Hilbert space adjoint operator of K is $K^* \in \mathcal{L}(Y, X)$. Assume $F \in \Gamma_0(X)$, $G \in \Gamma_0(Y)$, where $\Gamma_0(\cdot)$ is defined as in [1],

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

The ADMM method was introduced and developed in the middle 1970s by R. Glowinski, A. Marrocco, and M. Fortin [13,16], and D. Gabay, B. Mercier [14,15,19]. An extensive study including a discussion of its history can be found in [13]. It was originally designed for partial differential equation (PDE) based variational problems [16]. Its application to numerical solutions of PDEs and the corresponding convergence properties are studied in [15,19].

Classical approaches for the convergence analysis of the ADMM method often involve convex duality. With the Fenchel conjugate functionals $F^* : X \to (-\infty, +\infty]$, $G^* : Y \to (-\infty, +\infty]$, the associated Fenchel-Rockafellar dual problem of (1.4) reads as [1,8,18,28],

$$\max_{\lambda \in Y} \; -F^*(-K^*\lambda) - G^*(\lambda). \tag{1.2}$$

Under appropriate conditions, the maximum in (1.2) indeed exists and coincides with the minimum in (1.1), for instance, it is sufficient that [1,28]

$$\bigcup_{\lambda \geq 0} \lambda(K \operatorname{dom} F - \operatorname{dom} G) \text{ is a closed linear subspace of } \; Y, \tag{1.3}$$

where the domain of F is defined by $\operatorname{dom} F := \{x \in X \mid F(x) < +\infty\}$. It is proved in [14] that ADMM is a variant of Douglas-Rachford splitting method to the dual problem (1.2) with monotone operator splitting $\partial(F^* \circ (-K^*))$ and ∂G^*. Moreover, ADMM may also be interpreted as a kind of more general proximal point method [9]. Operator splitting methods as well as the proximal point method are very important since they provide a guided way to study various splitting algorithms which are also starting points for the proposed preconditioned ADMM method.

In this paper, we focus on analysing the ADMM method on the primal side, i.e., without considering the dual problem (1.2). Writing the minimization problem (1.1) as the following constrained problem

$$\min_{u \in X} \; F(u) + G(p), \quad \text{subject to} \quad Ku = p, \tag{1.4}$$

we can associate the equivalent saddle-point problem

$$\min_{(u,p) \in \operatorname{dom} F \times \operatorname{dom} G} \; \max_{\lambda \in Y} \; L(u, p, \lambda), \quad L(u, p, \lambda) = F(u) + G(p) + \langle \lambda, Ku - p \rangle \tag{1.5}$$

where the dual variable $\lambda \in Y$ is a Lagrange multiplier of the constraint $Ku = p$ [9]. ADMM can directly be motivated by considering the augmented Lagrangian:

$$L_r(u, p, \lambda) := F(u) + G(p) + \langle \lambda, Ku - p \rangle + \frac{r}{2}\|Ku - p\|^2, \tag{1.6}$$

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

$$\begin{cases} u^{k+1} := \arg\min_{u \in X} \ L_r(u, p^k, \lambda^k), \\ p^{k+1} := \arg\min_{p \in Y} \ L_r(u^{k+1}, p, \lambda^k), \\ \lambda^{k+1} := \lambda^k + r(Ku^{k+1} - p^{k+1}). \end{cases} \tag{ADMM}$$

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

$$K^*(rp^k - \lambda^k) \in (rK^*K + \partial F)u^{k+1} \tag{1.7a}$$

$$\lambda^k + rKu^{k+1} \in (rI + \partial G)p^{k+1}, \tag{1.7b}$$

where we assume that

$$rK^*K + \partial F \quad \text{is strongly monotone,} \tag{1.8}$$

such that $(rK^*K+\partial F)^{-1}$ exists and is Lipschitz continuous. Note that (1.7b) amounts to evaluating a proximal operator for G which might assumed to be computable. Indeed, in a lot applications, while the resolvent $(rI + \partial G)^{-1}$ has explicit representations although it is nonlinear or nonsmooth, e.g., $G(y) = \|y\|_1$. In contrast to that, the evaluation of $(rK^*K + \partial F)^{-1}$ can be challenging, even if ∂F is linear. However, if one is able to overcome this challenge, one benefits, as ADMM can be seen as a variant of the Douglas-Rachford iteration, from its unconditional stability. That is, if a solution exists, the iterations converges (or weakly converges) to a solution for any positive step-size r. Such a behaviour is essentially different from forward-backward type first-order algorithms which are usually conditionally stable, see [5], for instance.

In this paper, we are interested in studying variants in which the latter inversion step is replaced by computationally more tractable approaches. In this context, we put our attention mainly (but not entirely) on the special case where F is quadratic-linear, i.e.,

$$F(u) = \langle Su, u\rangle/2 - \langle f_0, u\rangle, \quad \forall u \in X, \tag{1.9}$$

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, in various regularized linear inverse problems with L^2 data fidelity terms in imaging and signal processing. The representation (1.9) 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 $rK^*K + \partial F$ is strongly monotone if

$$S + rK^*K > 0 \tag{1.10}$$

and (1.7a) amounts to solving the linear equation

$$(S + rK^*K)u^{k+1} = K^*(rp^k - \lambda^k) + f_0. \tag{1.11}$$

As the exact solution of the linear equation (1.11) might already be challenging and numerically expensive, computationally less demanding alternatives which still allow for unconditional stability are desirable, even in this simplified case. A well-known alternative is the inexact version of ADMM proposed by Eckstein and Bertsekas in [9]. However, the condition that the errors for solving (1.7) are summable is very strict and not flexible to use in practical, especially if the linear subproblem (1.11) is of large scale.

Adopting the framework of [26], we consider the preconditioned update step for u according to

$$u^{k+1} := \arg\min_u L_r(u, p^k, \lambda^k) + \tfrac{1}{2}\|u - u^k\|^2_{N - rK^*K} \tag{1.12}$$

where $N \in \mathcal{L}(X)$ is a self-adjoint "preconditioner" such that $N - rK^*K$ is positive semi-definite and the squared norm reads as $\|u\|^2_{N-rK^*K} = \langle (N - rK^*K)u, u\rangle$. This step amounts to solving

$$K^*(rp^k - \lambda) + (N - rK^*K)u^k \in (N + \partial F)u^{k+1}, \tag{1.13}$$

i.e., the rK^*K-term has been replaced by N which we are free to choose. In particular, inspired by the recently studied preconditioning techniques for the Douglas-Rachford iteration in [3], we are allowed, in the case of quadratic-linear F, to perform the following preconditioning step

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

with M being a suitable preconditioner for the linear operator $T = S + rK^*K$. Such a step then allows a flexible and efficient inexact solution of the linear equation without the need of controlling the errors.

In conclusion, the contribution of this paper is the following preconditioned ADMM which reads as follows:

$$\begin{cases} u^{k+1} = (N + \partial F)^{-1}[K^*(rp^k - \lambda^k) + (N - rK^*K)u^k], \\ p^{k+1} = (rI + \partial G)^{-1}[\lambda^k + rKu^{k+1}], \\ \lambda^{k+1} = \lambda^k + r(Ku^{k+1} - p^{k+1}). \end{cases} \tag{PADMM}$$

It allows for the inexact solution of the linear subproblem without controlling the error as, for instance, in [9]. We prove, in Theorem 3.5, weak convergence in Hilbert space on mild assumptions on the self-adjoint preconditioners. In particular, any 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 the preconditioned ADMM and PDRQ which is a kind of preconditioned Douglas-Rachford splitting method [3, 4] for problems of the type (1.4) with quadratic-linear F.

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 (PADMM), prove its well-definition and discuss its connection to the classical ADMM. In Section 3, we prove the convergence of the preconditioned ADMM by proving the convergence of the proximal point iteration, in both infinite and finite dimensional Hilbert spaces. In Section 4, we discuss some feasible preconditioners for preconditioned ADMM that were analyzed in [3], and give a comparison of preconditioned ADMM and PDRQ developed in [3] which is also aimed at quadratic-linear problems with non-smooth regularizer. In the last two sections, we present its application to demonstrate its efficiency. Finally, we give a brief conclusion.

2 Preconditioned ADMM as a proximal point method in primal form

Throughout this section as well as Section 3, denote by $A = \partial F$ and $B = \partial G$ which both are maximal monotone operators as $F \in \Gamma_0(X)$ and $G \in \Gamma_0(Y)$. Under the same condition as for duality (1.3), we have (e.g., see [1, Theorem 16.37])

$$\partial(F + G \circ K) = \partial F + K^* \circ \partial G \circ K$$

where $\partial(G \circ K) = K^* \circ \partial G \circ K$ and the symbol $\circ$ denotes the composition of (multivalued) mappings. Writing short-hand K^*BK for the latter composition, the primal problem (1.1) is equivalent to the following monotone inclusion problem

$$\text{find } \hat{u} \in X, \text{ such that } \quad 0 \in (A + K^*BK)(\hat{u}). \tag{2.1}$$

Introduce three auxiliary variables $\hat{p} \in Y$, $\hat{v} \in X$ and $\hat{x} \in X$ and suppose

$$\hat{p} = K\hat{u}, \qquad \hat{v} - \hat{x} \in B\hat{p}. \tag{2.2}$$

Then, the monotone inclusion problem (2.1) is equivalent to

$$0 \in A\hat{u} + K^*(\hat{v} - \hat{x}). \tag{2.3}$$

Combining (2.2) and (2.3), we get the following equivalent form of (2.1),

$$\text{find } \begin{pmatrix} \hat{x} \\ \hat{u} \\ \hat{v} \\ \hat{p} \end{pmatrix}, \text{ such that } \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} 0 & K & 0 & -I \\ -K^* & A & K^* & 0 \\ 0 & -K & 0 & I \\ I & 0 & -I & B \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{u} \\ \hat{v} \\ \hat{p} \end{pmatrix}. \tag{2.4}$$

With $\mathcal{U} = Y \times X \times Y \times Y$, $q = (x, u, v, p) \in \mathcal{U}$, and denoting $\mathcal{A} : \mathcal{U} \to 2^{\mathcal{U}}$ as the operator matrix in (2.4), 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

$$0 \in \mathcal{M}(q^{k+1} - q^k) + \mathcal{A}q^{k+1}. \tag{2.5}$$

We will see that the (preconditioned) ADMM iteration (ADMM) and, (PADMM) corresponds to the specific choice

$$\mathcal{M} := \begin{pmatrix} \frac{2}{r}I & K & 0 & -I \\ K^* & N & K^* & 0 \\ 0 & K & \frac{2}{r}I & I \\ -I & 0 & I & rI \end{pmatrix} \tag{2.6}$$

where $N \in \mathcal{L}(X)$ is self-adjoint and satisfies

$$N - rK^*K \geq 0. \tag{2.7}$$

We will see in Section 3 that $\mathcal{M}$ according to (2.6) is indeed linear, continuous, self-adjoint and positive semi-definite as soon as (2.7) is fulfilled.

The proximal point iteration (2.5), in case it is well-defined, also be written as $\mathcal{M}q^k \in (\mathcal{M} + \mathcal{A})q^{k+1}$, giving rise to the fixed-point mapping

$$\mathcal{T} : \mathcal{U} \to \mathcal{U}, \quad q \mapsto (\mathcal{M} + \mathcal{A})^{-1}\mathcal{M}q. \tag{2.8}$$

With these notations, we can establish a relation between (2.5) and (ADMM).

Theorem 2.1. *The mapping $\mathcal{T}$ is well-defined under conditions* (1.8) *and* (2.7)*, 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* (2.5) *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* (2.5) *the following identity holds*

$$(\bar{u}^k, \bar{p}^k, \bar{\lambda}^k) = (u^{k-1}, p^k, -x^k + v^k + rp^k - rKu^k) \qquad \text{for each } k \geq 1.$$

Proof. For a given $q^k = (x^k, u^k, v^k, p^k)$ denote by $q^{k+1} := \mathcal{T}q^k$, i.e., $\mathcal{M}q^k \in (\mathcal{M}+\mathcal{A})q^{k+1}$. By (2.6) and as $\mathcal{A}$ denotes the operator matrix in (2.4), the iteration can be written componentwise as

$$p^{k+1} = (rI + B)^{-1}(-x^k + v^k + rp^k), \tag{2.9a}$$

$$v^{k+1} = v^k + \frac{r}{2}[Ku^k + p^k - 2p^{k+1}], \tag{2.9b}$$

$$u^{k+1} = (N + A)^{-1}[K^*(x^k + v^k - 2v^{k+1}) + Nu^k], \tag{2.9c}$$

$$x^{k+1} = x^k + \frac{r}{2}[Ku^k - 2Ku^{k+1} - p^k + 2p^{k+1}]. \tag{2.9d}$$

As B is maximal monotone, we see by (2.9a) that p^{k+1} is uniquely determined by q^k. With the newly updated p^{k+1} and (2.9b), v^{k+1} is uniquely determined. For the update of u^k (2.9c), note that by conditions (1.8) and (2.7), $N + A$ is strongly monotone, hence $(N + A)^{-1}$ is a well-defined Lipschitz-continuous mapping. Consequently, u^{k+1} is uniquely determined as well as x^{k+1}, the latter by (2.9d). Thus, q^{k+1} is uniquely determined by q^k which leads to the well-definition of $\mathcal{T}$.

In order to prove the stated equivalence of (PADMM) and (2.5), introduce the variable λ^k according to

$$\lambda^k := -x^k + v^k + rp^k - rKu^k. \tag{2.10}$$

Then, the update of p^k as in (2.9a) becomes

$$p^{k+1} = (rI + B)^{-1}(\lambda^k + rKu^k). \tag{2.11}$$

For the update of λ^k, consider (2.10) for $k+1$, i.e.,

$$\lambda^{k+1} = -x^{k+1} + v^{k+1} + rp^{k+1} - rKu^{k+1}.$$

Plugging in v^{k+1} and x^{k+1} from (2.9b) and (2.9d) to the right hand side of above equation, we get

$$\lambda^{k+1} = -x^k + v^k + rp^k - rp^{k+1}. \tag{2.12}$$

Substituting λ^k according to (2.10) yields

$$\lambda^{k+1} = \lambda^k + r(Ku^k - p^{k+1}). \tag{2.13}$$

Finally, for the update of u^k, Actually, we can write (2.9c) as follows

$$u^{k+1} = (N + A)^{-1}[K^*(x^k + v^k - 2v^{k+1} + rKu^k) + (N - rK^*K)u^k]. \tag{2.14}$$

With (2.9b) and (2.12), the argument for K^* becomes

$$\begin{aligned} x^k + v^k - 2v^{k+1} + rKu^k &= x^k + v^k - 2v^k - r[Ku^k + p^k - 2p^{k+1}] + rKu^k \\ &= x^k - v^k + 2rp^{k+1} - rp^k = rp^{k+1} - \lambda^{k+1}, \end{aligned}$$

leading to the identity

$$u^{k+1} = (N + A)^{-1}[K^*(rp^{k+1} - \lambda^{k+1}) + (N - rK^*K)u^k]. \tag{2.15}$$

Thus, (2.11), (2.13) and (2.15) 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

$$\begin{aligned} &x^0 = 0, \quad u^0 = (N + A)^{-1}[K^*(r\bar{p}^0 - \bar{\lambda}^0) + (N - rK^*K)\bar{u}^0], \\ &v^0 = 0, \quad p^0 = \tfrac{1}{r}\bar{\lambda}^0 + Ku^0. \end{aligned}$$

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 (2.10), the desired statement. $\square$

3 Convergence of preconditioned ADMM

In Theorem 2.1, we have shown the equivalence between the preconditioned ADMM (PADMM) with the proximal point method (2.5). 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 (1.8) and (2.7) be satisfied such that Theorem 2.1 is applicable. We start with the relation between the fixed point of $\mathcal{T}$ and solutions of the saddle point problem (1.5).

Lemma 3.1. *If a solution of saddle point problem* (1.5) *exists, then fixed points of* $\mathcal{T}$ *according to* (2.8) *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* (1.5).

Proof. Suppose $(\hat{u}, \hat{p}, \hat{\lambda})$ is a saddle point of (1.5), then by optimality, $-K^*\hat{\lambda} \in A\hat{u}$ and $\hat{\lambda} \in B\hat{p}$ as well as $K\hat{u} - \hat{p} = 0$. Choosing $\hat{v} - \hat{x} = \hat{\lambda}$, equations (2.2) and (2.3) are satisfied, i.e., $\hat{q} = (\hat{x}, \hat{u}, \hat{v}, \hat{p})$ fulfills $0 \in \mathcal{A}\hat{q}$ according to (2.4). But this means 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 (2.2) and (2.3) are satisfied, meaning that $(\hat{u}, \hat{p}, \hat{v} - \hat{x})$ fulfills the optimality conditions for (1.5). □

For the weak convergence of the proximal iteration (2.5), let us first study the "preconditioner" $\mathcal{M}$. It can easily be checked that $\mathcal{M}$ in (2.6) is a self-adjoint operator in $\mathcal{L}(\mathcal{U})$, i.e.,

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

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

$$\|q\|_{\mathcal{M}}^2 = \langle \mathcal{M}q, q \rangle = \frac{1}{r}\|rKu + x + v\|_Y^2 + \frac{1}{r}\|rp - (x - v)\|_Y^2 + \|(N - rK^*K)^{\frac{1}{2}}u\|_X^2. \tag{3.2}$$

Here, $(N - rK^*K)^{\frac{1}{2}}$ exists due to (2.7), see e.g., [22], and is a linear, continuous and positive semi-definite operator. In particular, if $\|q\|_{\mathcal{M}} = 0$ for $q = (x, u, v, p)$, we have

$$rKu + x + v = 0, \tag{3.3a}$$
$$rp - (x - v) = 0, \tag{3.3b}$$
$$(N - rK^*K)u = 0, \tag{3.3c}$$

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, $u_i' - K^*(v_i - x_i) \in Au_i$ and $p_i' - (x_i - v_i) \in Bp_i$, hence

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

and, as $x_i' = Ku_i - p_i$ and $v_i' = p_i - Ku_i$,

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

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 (2.8), we can choose $q_1' \in \mathcal{A}\mathcal{T}q_1 \subset 2^{\mathcal{U}}$, $q_2' \in \mathcal{A}\mathcal{T}q_2 \subset 2^{\mathcal{U}}$, 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:

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

To establish (3.5), introduce the notations

$$\eta = rKu + x + v, \qquad \xi = rp - x + v, \qquad \mu = (N - rK^*K)^{\frac{1}{2}}u$$

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. By (2.9a), we have

$$\|p_1' - p_2'\| \leq \|\eta_1 - \eta_2\| \leq \sqrt{r}\|q_1 - q_2\|_{\mathcal{M}}.$$

Note that (2.9b) implies $v_i' = \frac{1}{2}(\eta_i + \xi_i) - r p_i'$, hence

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

for some $C > 0$. (We will use, in the following, the same C for different constants as we are only interested in the qualitative estimate.) Likewise, with (2.9b) it follows that $x_i + v_i - 2v_i' + rKu_i = 2rp_i' - \eta_i$, so (2.9c) as well as the Lipschitz continuity of $(N+A)^{-1}$ (which is due to the strong monotonicity of $(N + A)$, see (1.8) and (2.7)), yields

$$\begin{aligned}\|u_1' - u_2'\| &\leq C\big(2r\|K^*\|\|p_1' - p_2'\| + \|K^*\|\|\eta_1 - \eta_2\| + \|(N - rK^*K)^{\frac{1}{2}}\|\|\mu_1 - \mu_2\|\big) \\ &\leq C\|q_1 - q_2\|_{\mathcal{M}}.\end{aligned}$$

Finally, (2.9d) yields $x_i' = \frac{1}{2}(\eta_i - \xi_i) + r(p_i' - Ku_i')$, hence

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

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 [3]. Thus, we just give the result without proof.

Lemma 3.3. *If there is a solution to problem* (1.5)*, 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* $k \to \infty$*.*

Now, the main technical difficulty in establishing weak convergence following the lines of [20] and [3] is deriving the demi-closedness of the operator $I - \mathcal{T}$ in $\mathcal{U}$. We will discuss three sufficient conditions for this property to hold, the first being the compactness of K, the second the positive definiteness of $N - rK^*K$ while the third states that A is affine-linear and an injectivity property holds.

More precisely, in the latter case we assume that F according to (1.9) is quadratic-linear which imposes the representation

$$A = S - \mathbf{1}_{f_0} \tag{3.6}$$

for $S \in \mathcal{L}(X)$ linear, continuous and positive semi-definite and $f_0 \in X$. Denoting by $\mathcal{N}$ the kernel of a linear operator, the following injectivity is furthermore assumed:

$$\mathcal{N}\begin{pmatrix} N - rK^*K \\ S - rK^*K \end{pmatrix} = \{0\}, \tag{3.7}$$

i.e., from $(N - rK^*K)u = 0$ and $(S - rK^*K)u = 0$ follows that $u = 0$. This is clearly weaker than injectivity of one of the operators $N - rK^*K$ and $S - rK^*K$ and does not require full range, i.e., continuous invertibility.

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

Lemma 3.4. *Let one of the following conditions be true:*

(i) K *is compact.*

(ii) $N - rK^*K$ *is positive definite.*

(iii) The operator A *satisfies* $A = S - \mathbf{1}_{f_0}$ *according to* (3.6) *and the injectivity condition* (3.7) *holds.*

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 $\hat{q} = \mathcal{T}q + \bar{q}$. We first prove

$$\|q - \hat{q}\|_{\mathcal{M}} = 0. \tag{3.8}$$

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 $\mathcal{T}$, we have

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

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

$$\begin{aligned}
rK(u - u' - \bar{u}) + (x - x' - \bar{x}) + (v - v' - \bar{v}) &= 0, \\
(v - v' - \bar{v}) - (x - x' - \bar{x}) + r(p - p' - \bar{p}) &= 0, \\
(N - rK^*K)(u - u' - \bar{u}) &= 0.
\end{aligned} \tag{3.9}$$

Our aim is to show that $q - q' - \bar{q} = 0$ which is the claimed statement. By definition of $\mathcal{T}$ (also see (2.9b) and (2.9d)), we have

$$v - v' = -\frac{r}{2}[Ku + p - 2p'], \qquad x - x' = -\frac{r}{2}[K(u - 2u') - (p - 2p')].$$

Denoting $\mathcal{T}q^k = \big((x^k)', (u^k)', (v^k)', (p^k)'\big)$, the updates for v^k and x^k according to (2.9b) and (2.9d) leads to

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

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

$$\begin{aligned}
0 &= (v - v' - \bar v) - (x - x' - \bar x) + r(p - p' - \bar p)\\
&= \frac{r}{2}\big[2p' - p - Ku + Ku + 2\bar p - p\big] + \frac{r}{2}\big[K(u - 2u') - (p - 2p')\\
&\qquad - K(2\bar u - u) + (2\bar p - p)\big] + r(p - p' - \bar p)\\
&= r\big[K(u - u' - \bar u) - (p - p' - \bar p)\big],
\end{aligned} \tag{3.10}$$

or, expressed differently, $K(u - u' - \bar u) = (p - p' - \bar p)$. We can now use this to obtain

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

For the following steps we need to distinguish cases.
Case (i). If K is compact, it follows by complete continuity that $Ku^k \to Ku$ in Y as $k \to \infty$. Thus, $\{p^k\}$ also converges strongly to p which follows, for instance, from

$$\frac{r}{2}p^k = \lim_{k\to\infty} v^k - (v^k)' + \frac{r}{2}\big[Ku^k + 2\big(p^k - (p^k)'\big)\big] \to \bar v + \frac{r}{2}\big[Ku + 2\bar p\big] = \frac{r}{2}p,$$

as $k \to \infty$, the latter since weak and strong limit coincide. Consequently, $(p^k)' = p^k + (p^k)' - p^k \to p - \bar p$ as well as $-x^k + v^k + r\big(p^k - (p^k)'\big) \rightharpoonup -x + v + r\bar p$ as $k \to \infty$. The update rule (2.9a) for p^k implies

$$r\big(p^k - (p^k)'\big) + v^k - x^k \in B(p^k)' \qquad \text{for each } k,$$

so by strong-weak closedness of maximally monotone operators, $r\bar p + v - x \in B(p - \bar p)$ or, equivalently, $p - \bar p = (rI + B)^{-1}(-x + v + rp)$. By definition of $\mathcal{T}$, the right-hand side coincides with p', hence $p - p' - \bar p = 0$. Together with (3.10) and (3.11), it also follows that $v - v' - \bar v = 0$.

We will use almost the same argument to obtain $u - u' - \bar u = 0$. The update rule (2.9c) for u^k gives

$$N\big(u^k - (u^k)'\big) + K^*\big(x^k + v^k - 2(v^k)'\big) \in A(u^k)' \qquad \text{for each } k,$$

where $(u^k)' \rightharpoonup u - \bar u$ as $k \to \infty$ and, as N is continuous and K^* is completely continuous,

$$N\big(u^k - (u^k)'\big) + K^*\big(x^k + v^k - 2(v^k)'\big) \to N\bar u + K^*(x - v + 2\bar v) \qquad \text{as } k \to \infty.$$

Maximally monotone operators are also weak-strong closed (as they have maximally monotone inverses), hence, plugging in $\bar{v} = v - v'$, we get $N\bar{u} + K^*(x+v-2v') \in A(u-\bar{u})$ which is equivalent to $u - \bar{u} = (N+A)^{-1}\big(K^*(x+v-2v')+Nu\big)$. 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 (ii). If $N - rK^*K$ is positive definite, then it is in particular injective, so (3.9) implies $u - u' - \bar{u} = 0$. Then, combining (3.10), (3.11) and (3.9) leads to $p - p' - \bar{p} = 0$, $v - v' - \bar{v} = 0$, consequently, $q - q' - \bar{q} = 0$.

Case (iii). Now, suppose that $A = S - \mathbf{1}_{f_0}$, so it follows that $(N + A)^{-1} = (N + S)^{-1}(I + \mathbf{1}_{f_0})$, and thus, (2.9c) and the weak sequential continuity of linear operators gives us

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

implying, together with $v - v' - \bar{v} = rK(\bar{u} + u' - u)$ which follows from (3.9), (3.10) and (3.11), that

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

Applying $(N + S)$ on both sides, using $(N - rK^*K)(u - u' - \bar{u}) = 0$ (see (3.9)) and rearranging, this turns out to be equivalent to

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

As condition (3.7) is assumed to hold, we can conclude $u - u' - \bar{u} = 0$. Finally, (3.10) together with (3.11) and (3.9) gives $p - p' - \bar{p} = 0$ and $v - v' - \bar{v} = 0$, hence $q - q' - \bar{q} = 0$. □

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

Theorem 3.5. *Let* (1.8) *be true and let N satisfy* (2.7). *If a solution of the saddle point problem* (1.5) *exists, then the proximal iteration sequence $\{q^k\}$ according to* (2.5) *converges weakly to a fixed-point $q^* = (x^*, u^*, v^*, p^*)$, under the condition one of the conditions (i), (ii) or (iii) in Lemma 3.4 holds. The triple $(u^*, p^*, v^* - x^*)$ is a solution of the saddle point problem* (1.5).

Proof. Denote the set of the fixed points of $\mathcal{T}$ by $\mathcal{F}$ which is a closed and convex set. As a solution for (1.5) 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

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

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):

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

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 3.6. *If, in the situation of Theorem 3.5, K is compact, then the primal variables $\{u^k\}$ and $\{p^k\}$ converge strongly.*

Proof. For compact K we already have weak convergence of the sequence $\{q^k\}$ generated by (2.5) to a limit q^* as well as $q^k - q^{k+1} \to 0$ as $k \to \infty$. Thus, $Ku^k \to Ku^*$ as $k \to \infty$ by compactness of K and passing to the limit, for instance, in (2.9b) leads to strong convergence of $\{p^k\}$. Likewise, as K^* is also compact and (2.9c) can be rewritten to

$$u^{k+1} = (rK^*K + A)^{-1}\big[K^*(x^k + v^k - 2v^{k+1} + rKu^{k+1}) + N(u^k - u^{k+1})\big]$$

with $(rK^*K + A)^{-1}$ being continuous and its argument strongly convergent. Consequently, also the sequence $\{u^k\}$ converges strongly. □

Corollary 3.7. *If X is finite-dimensional, then the primal variables $\{u^k\}$ and $\{p^k\}$ converge strongly. If Y is finite-dimensional, primal and dual variables converge strongly.*

Proof. In both cases, K has to be compact, so Corollary 3.6 gives the strong convergence for the primal variables. In case Y is finite-dimensional, weak and strong convergence for $\{x^k\}$ and $\{v^k\}$ coincide. □

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 $A = S - \mathbf{1}_{f_0}$ according to (3.6). Furthermore, let $S + rK^*K$ be positive definite, see (1.10). Finally, let us introduce the notation $T = S + rK^*K$ and $M = N + S$.

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 3.5, i.e.,

$$M \geq T \qquad \text{and} \qquad \begin{cases} M > T \text{ or} \\ \mathcal{N}\begin{pmatrix} M - T \\ M - 2S \end{pmatrix} = \{0\} \end{cases} \qquad \text{if } K \text{ is not compact.} \tag{4.1}$$

It is easily checked that $M \geq T$ and the injectivity condition are equivalent to (2.7) and (3.7), respectively. Hence, Theorem 3.5 is applicable, implying (weak) convergence of the iteration.

The reason for introducing M and T is that the update step for u^k, see (2.9c), becomes, as announced in the introduction,

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

which corresponds to one step of a splitting method for the solution of $Tu^{k+1} = b^k$ with $b^k = K^*(rp^k - \lambda^k) + f_0$ 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

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

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

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

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).

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

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

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

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

Table 1: Choices for different preconditioners for $T = S + rK^*K$.

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

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

If K is compact, the iteration converges for $\mu \geq r\|K\|^2$, otherwise, $\mu > r\|K\|^2$ leads to weak convergence, see Theorem 3.5 and Corollary 3.6. The whole iteration (PADMM) *then corresponds to a* split inexact Uzawa method*, see [12, 26], and is equivalent to the primal-dual method in [5] (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 *[17, 24]. 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 [3]. 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 [3]. 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 [3] 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.,*

$$\begin{cases} x^{k+(i+1)/n} = x^{k+i/n} + M^{-1}(b^k - Tx^{k+i/n}) \\ \qquad\qquad i = 0, \ldots, n-1 \end{cases}$$

corresponds to $x^{k+1} = x^k + M_n^{-1}(b^k - Tx^k)$ where M_n is a feasible preconditioner.

For more sufficient conditions for feasibility, we once more refer to [3].

4.2 The relation to PDRQ

We next 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 [3, 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 [3], the PDRQ iteration then reads as

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

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 $\lambda^{k+1} = \lambda^k + r(Ku^{k+1} - p^k)$ as well as $2y^{k+1} - \bar{y}^k = \lambda^{k+1} + rKu^{k+1}$ 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

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

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 [27] for a general discussion of this phenomenon). Hence, PDRQ can alternatively be expressed as

$$\begin{cases} u^{k+1} = \operatorname*{arg\,min}_{u \in X} \; L_r(u, p^k, \lambda^k) + \frac{1}{2}\|u - u^k\|^2_{N - rK^*K} \\ \lambda^{k+1} = \lambda^k + r(Ku^{k+1} - p^k) \\ p^{k+1} = \operatorname*{arg\,min}_{p \in Y} \; L_r(u^{k+1}, p, \lambda^{k+1}) \end{cases} \tag{PDRQ-ALM}$$

where L_r is given by (1.6) with F according to (1.9) and $\|\cdot\|^2_{N-rK^*K}$ is defined as for (1.12). Despite the similarity, note that u and p play different roles in the original formulation (1.4): The variable u is associated with an operator K 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 [3].

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)*:*

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

where in case of linear quadratic F according to (1.9)*, the update for x^k can be written as*

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

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 [6], 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.

5.1 Discretization

We start with describing the discrete setting. Following essentially the presentation in [3], consider the image domain $\Omega \subset \mathbf{Z}^2$ as the discretized rectangular grid

$$\Omega = \{(i, j) \mid i, j \in \mathbb{N}, \; 0 \le i \le N_x - 1, 0 \le j \le N_y - 1\}$$

where N_x, N_y are the image dimensions. Finite differences are used to discretize the operator ∇ and its adjoint operator $\nabla^* = -\operatorname{div}$ with homogeneous Neumann and Dirichlet boundary conditions, respectively. We define ∇ as the following operator

$$(\nabla u) = \begin{pmatrix} \partial_x^+ u \\ \partial_y^+ u \end{pmatrix},$$

where forward differences are taken according to

$$(\partial_x^+ u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j}, & \text{if } 0 \le i < N_x - 1, \\ 0, & \text{if } i = N_x - 1, \end{cases} \qquad (\partial_y^+ u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j}, & \text{if } 0 \le j < N_y - 1, \\ 0, & \text{if } j = N_y - 1. \end{cases}$$

With the following two vector spaces of functions and vector fields,

$$X = \{u : \Omega \to \mathbb{R}\}, \qquad Y = \{v : \Omega \to \mathbb{R}^2\},$$

the operator maps $\nabla : X \to Y$. The discrete divergence is then the negative adjoint of ∇, i.e., the unique linear mapping $\operatorname{div} : Y \to X$ which satisfies

$$\langle \nabla u, v \rangle_Y = \langle u, \nabla^* v \rangle_X = -\langle u, \operatorname{div} v \rangle_X, \quad \forall u \in X,\ v \in Y.$$

It can be computed to read as

$$\operatorname{div} v = \partial_x^- v^1 + \partial_y^- v^2$$

involving the backward difference operators

$$(\partial_x^- u)_{i,j} = \begin{cases} u_{0,j}, & \text{if } i = 0, \\ u_{i,j} - u_{i-1,j}, & \text{for } 0 < i < N_x - 1, \\ -u_{N_x-1,j}, & \text{for } i = N_x - 1, \end{cases} \quad (\partial_y^- u)_{i,j} = \begin{cases} u_{i,0}, & \text{if } j = 0, \\ u_{i,j} - u_{i,j-1}, & \text{for } 0 < j < N_y - 1, \\ -u_{i,N_y-1}, & \text{for } j = N_y - 1. \end{cases}$$

The discrete Laplacian is then given by $\Delta = \operatorname{div} \nabla$. Finally, in order to define the discrete version of TV-regularized variational problems, we still need the discrete versions of the L^1, L^2 and L^∞ norms: For $u \in X$, $v = (v^1, v^2) \in Y$ and $1 \le p < \infty$ let

$$\|u\|_p = \Big(\sum_{(i,j)\in\Omega} |u_{i,j}|^p \Big)^{1/p}, \quad \|u\|_\infty = \max_{(i,j)\in\Omega} |u_{i,j}|,$$

$$\|v\|_p = \Big(\sum_{(i,j)\in\Omega} \big((v^1_{i,j})^2 + (v^2_{i,j})^2\big)^{p/2} \Big)^{1/p}, \quad \|v\|_\infty = \max_{(i,j)\in\Omega} \sqrt{(v^1_{i,j})^2 + (v^2_{i,j})^2}.$$

5.2 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 [23] for the L^2-case which is usually called the *ROF model*)

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

and $f : \Omega \to \mathbf{R}$ a given noisy image and $\alpha > 0$ a regularization parameter.
The case $p = 2$. We see that for $p = 2$, (5.1) is equivalent to (1.4) with F according to (1.9) and data

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

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 (1.5) [1,8,18]. Furthermore, observe that $S + rK^*K = I - r\Delta > 0$ 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 [7]. 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 (1.9). In total, the problem has the pair (v, p) as dual variable and the data

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

Again, $S + rK^*K = r(I - \Delta) > 0$ 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 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 [4, 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.*

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.,*

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

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.

for $L : X \to Z$ a linear operator and $p \in \{1, 2\}$. While ADMM is generally applicable here, see for instance [11], 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 [4] 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 [5] 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 [5] 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. 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 [3]. 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

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

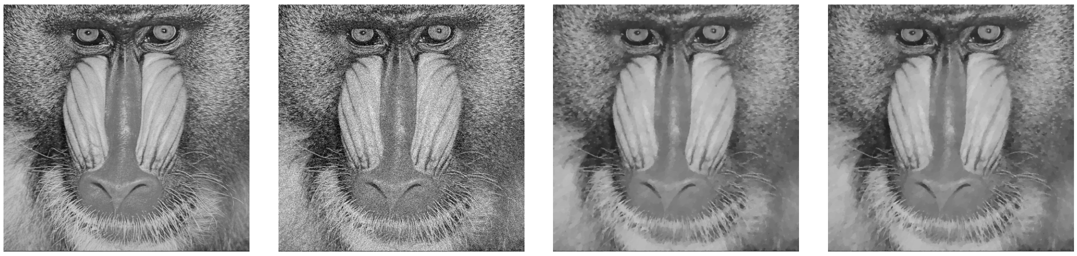
as the basis for the comparison as well as a stopping criterion by plugging, in each iteration step, a primal-dual test pair $(u_{\text{test}}^k, p_{\text{test}}^k)$ and normalizing by the number of pixels. The iteration is stopped once this normalized primal-dual gap is below a specified tolerance $\varepsilon > 0$, 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 [5] and [3], 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 (MATLAB and Image Processing Toolbox Release R2012b, The MathWorks, Inc., Natick, Massachusetts, United States) 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. 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

	$\alpha = 0.1$		$\alpha = 0.3$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
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})

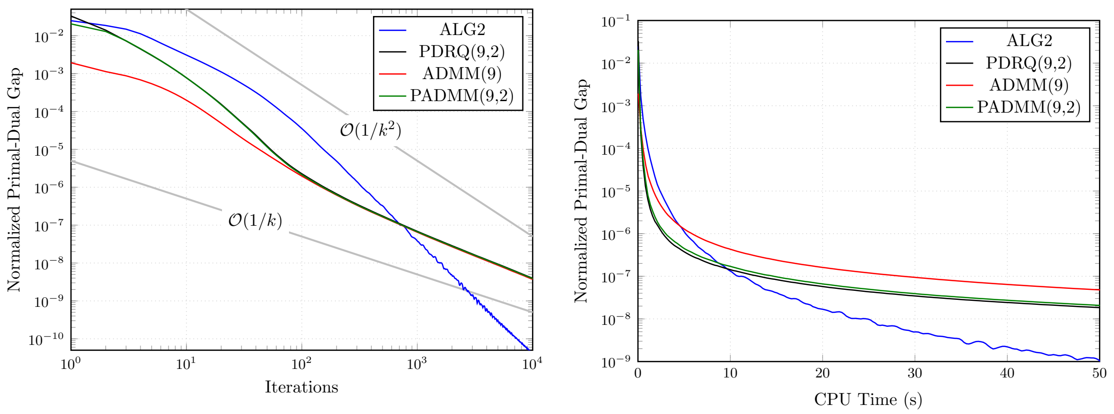
Figure 1: Image denoising with random Gaussian noise. (a) shows the 512×512 input Baboon image and (b) is a noisy image which has been corrupted by 10% random Gaussian noise. (c) and (d) are the denoised images by the PADMM algorithm with normalized primal energy less than 10^{-4} and 10^{-6} for $\alpha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

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}.

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 [5] 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 [5].
- ADMM: Alternating direction method of multipliers (ADMM). Again, DCT is employed for linear subproblem (5.5) instead of preconditioners.



(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 error of energy is compared in terms of iteration number and computation time for Figure 2 with $\alpha = 0.3$. 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.

- PADMM: Preconditioned alternating direction method of multipliers. Symmetric Red-Black Gauss-Seidel preconditioner is employed for linear subproblem (5.5).
- PDRQ: Preconditioned Douglas-Rachford method for pure quadratic-linear primal (or dual) functionals in [3]. 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 than 4×10^5 iterations and represents the minimal value among all tested algorithms.

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

	$\alpha = 1.0$		
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$
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.

the best performance both with respect to iteration numbers and computation time, achieving a speedup of more than 2 compared to ADMM and ALG1.

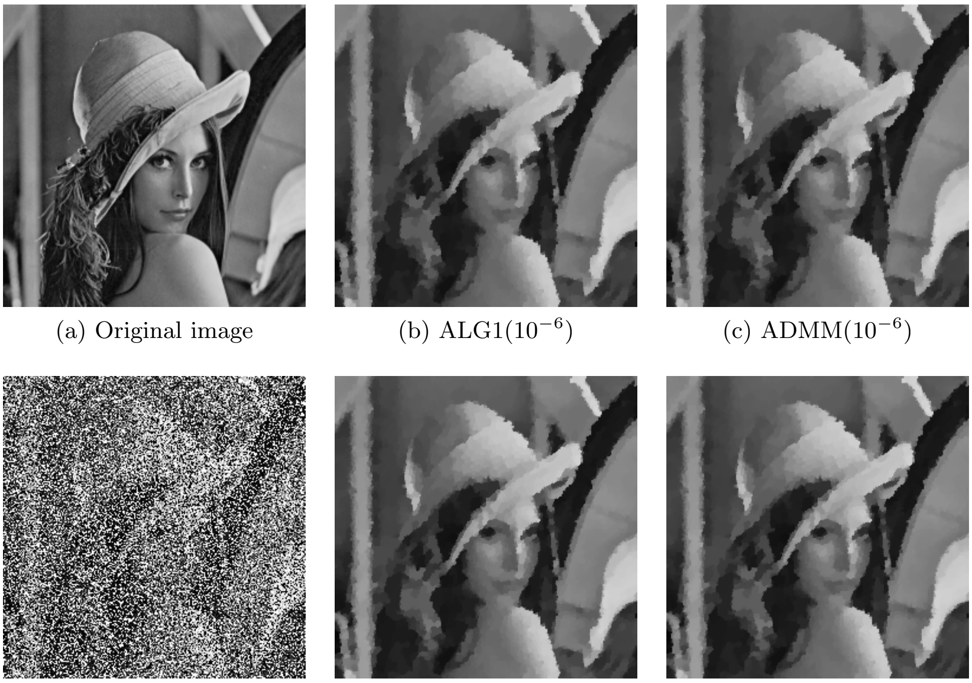
7 Conclusions

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 complex problems for which the subproblems also have to be solved iteratively. We also present applications for image denoising. The numerical results tell us that the preconditioned ADMM has the potential to bring out appealing benefits and fast algorithms.

Nevertheless, there are some interesting considerations, such as non-stationary iterative methods such as the CG (conjugate gradient) method that needs to be investigated. The CG method has already been used in numerical experiments for the linear subproblem of ADMM, e.g., for reconstruction in computed tomography (CT) [21]. Establishing convergence rates as well as accelerating ADMM and preconditioned ADMM to give optimal rates is also a very important topic. We believe that this framework will also be beneficial for other applications, 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.



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

References

[1] H. H. Bauschke, P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer Science+Business Media, LLC 2011, New York.

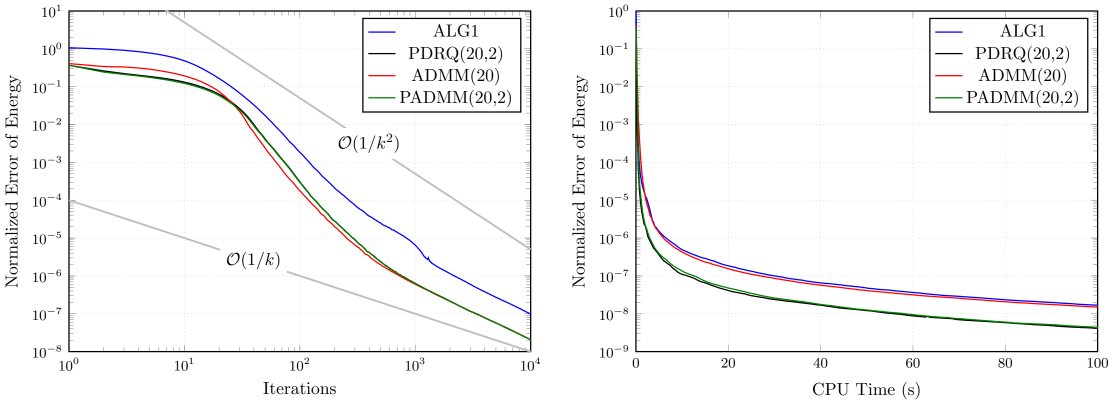
[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Found. Trends Mach. Learn., Vol.3, No.1, Jan, 2011, pp. 1-122.

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

[4] K. Bredies, H. P. Sun. *Preconditioned Douglas–Rachford Algorithms for TV- and TGV-Regularized Variational Imaging Problems*, J. Math. Imaging Vision (2015), DOI: 10.1007/s10851-015-0564-1.

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

[6] A. Chambolle, T. Pock. *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, IEEE International Conference on Computer Vision (ICCV), pp. 1762-1769, 2011.



(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.

[7] P. L. Combettes, V. R. Wajs. *Signal Recovery by Proximal Forward-Backward Splitting*, SIAM Multiscale Modeling & Simulation, Vol.4, No.4, 1168-1200 (2005).

[8] I. Eckland, R. Témam. *Convex Analysis and Variational Problems*, english ed, Classics in Applied Mathematics 28, SIAM, Philadelphia, PA, 1999, Translated from French.

[9] J. Eckstein, D. P. Bertsekas. *On the Douglas–Rachford splitting method and the proximal algorithm for maximal monotone operators*, Mathematical Programming 55, (1992), 293-318.

[10] J. Eckstein. *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, Rutcor Research Report, RRR 32-2012, Dec, 2012.

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

[12] E. Esser, X. Q. Zhang, T. F. Chan. *A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science*, SIAM Journal on Imaging Sciences, Vol.3, No.4, 2010, pp. 1015-1046.

[13] M. Fortin, R. Glowinski. *On decomposition-coordination methods using an augmented Lagrangian*, in: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*, North-Holland, Amsterdam, 1983.

[14] D. Gabay. *Applications of the method of multipliers to variational inequalities*, in: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*, North-Holland, Amsterdam, 1983.

[15] D. Gabay, B. Mercier. *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & Mathematics with Applications, Vol.2, No.1, 1976, pp.17-40.

[16] R. Glowinski, A. Marrocco. *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, tome 9, no. 2 (1975), p. 41-76.

[17] T. Goldstein, S. Osher. *The split Bregman method for L1 regularized problems*, SIAM J. Imaging Sci., 2(2), (2009), pp. 323–343.

[18] K. Ito, K. Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*, Advances in design and control 15, Philadelphia, SIAM, 2008.

[19] B. Mercier. *Approximation par éléments finis et résolution par un algorithme de pénalisation-dualité d'un problème d'élasto-plasticité*, C.R. Acad. Sc. Paris, 280A, (1975), pp. 287-290.

[20] Z. Opial. *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc. Vol.73, No.4 (1967), pp. 591-597.

[21] S. Ramani, J. A. Fessler. *A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction*, IEEE Trans Med Imaging. 2012 Mar;31(3):677-88.

[22] M. Reed, B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, 1980.

[23] L. Rudin, S. Osher, E. Fatemi. *Nonlinear total variation based noise removal algorithms*, Physica D, 60, (1992), pp. 259-268.

[24] S. Setzer. *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, Scale Space and Variational Methods in Computer Vision, Springer Berlin Heidelberg, 2009, pp. 464-476.

[25] C. L. Wu, X. C. Tai. *Augmented Lagrangian Method, Dual Methods, and Split Bregman Iteration for ROF, Vectorial TV, and High Order Models*, SIAM J. Imaging Sci., 3(3), (2010), pp. 300–339.

[26] X. Q. Zhang, M. Burger, S. Osher. *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput, Jan. 2011, Vol.46, Issue 1, pp.20-46.

[27] M. Yan, W. Yin. *Self Equivalence of the Alternating Direction Method of Multipliers*, UCLA CAM Report 14-59, 2014.

[28] C. Zălinescu. *Convex Analysis in General Vector Spaces*, World Scientific, Singapore, 2002.