SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Preconditioned Douglas-Rachford Splitting Methods for Convex-Concave Saddle-Point Problems

K. Bredies H. Sun

SFB-Report No. 2014-002 June 2014

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 DOUGLAS-RACHFORD SPLITTING METHODS FOR CONVEX-CONCAVE SADDLE-POINT PROBLEMS

KRISTIAN BREDIES* AND HONGPENG SUN†

Abstract. We propose a preconditioned version of the Douglas-Rachford splitting method for solving convex-concave saddle-point problems associated with Fenchel-Rockafellar duality. It allows to use approximate solvers for the linear subproblem arising in this context. We prove weak convergence in Hilbert space under minimal assumptions. In particular, various efficient preconditioners are introduced in this framework for which only a few inner iterations are needed instead of computing an exact solution or controlling the error. The method is applied to a discrete total-variation denoising problem. Numerical experiments show that the proposed algorithms with appropriate preconditioners are very competitive to existing fast algorithms including the first-order primal-dual algorithm for saddle-point problems of Chambolle and Pock.

Key words. Saddle-point problems, Douglas-Rachford splitting, linear preconditioners, convergence analysis.

AMS subject classifications. 65K10, 65F08, 49K35, 90C25.

1. Introduction.

The Douglas-Rachford splitting method is a classical approach for finding a zero point of the sum of maximal set-valued monotone operators which is quite common for the minimization of the sum of convex functions [12, 20]. While originally, it was designed for the discretized solution of the heat conduction problem [11], it developed to be a powerful tool due to its general applicability and unconditional stability. The convergence analysis is given, for example, in [9, 10, 20]. The Douglas-Rachford splitting method can in particular be interpreted in the framework of proximal point algorithms [12] which provides a convenient way to the analysis of this algorithm and is also our starting point.

The paper focuses on the following widely-used generic saddle-point problem [8]:

$$\min_{x\in X}\max_{y\in Y}\ \langle Kx,y\rangle + F(x) - G(y). \tag{1.1}$$

Here, X and Y are real Hilbert spaces and $K : X \to Y$ is a continuous linear mapping. The functionals $F : X \to \mathbf{R}_\infty$ (where $\mathbf{R}_\infty$ are the extended real numbers) and $G : Y \to \mathbf{R}_\infty$ are proper, convex and lower semi-continuous. With the Fenchel conjugate functionals $F^* : X \to \mathbf{R}_\infty$, $G^* : Y \to \mathbf{R}_\infty$, (1.1) is the associated Fenchel-Rockafellar primal-dual formulation of the primal and dual problem

$$\min_{x\in X}\ F(x) + G^*(Kx), \qquad \max_{y\in Y}\ -F^*(-K^*y) - G(y), \tag{1.2}$$

respectively [13, 18]. These are popular for discrete image denoising and deblurring problems where K is the ∇ operator and $G^*(Kx)$ is associated with the total variation functional [5, 8].

In order to solve (1.2), different kinds of operator splittings can be employed. It is shown by Gabay [15] that choosing $A = \partial[F^*(-K^*)]$, $B = \partial G$ and applying the Douglas-Rachford splitting method for solving $0 \in Ay + By$ results in the well-known *alternating direction method of multipliers* (ADMM) [12]. The splitting approach in this paper is based on the primal-dual optimality system for (1.1):

$$\begin{cases} 0 \in -Kx + \partial G(y), \\ 0 \in K^*y + \partial F(x) \end{cases} \tag{1.3}$$

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

†Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstr. 36, A-8010 Graz, Austria. (`hongpeng.sun@uni-graz.at`)

which becomes $0 \in Az + Bz$ for $z = (x, y)$, $z \in \mathcal{H} = X \times Y$ provided that

$$A = \begin{pmatrix} \partial F & 0 \\ 0 & \partial G \end{pmatrix}, \quad B = \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix}. \tag{1.4}$$

This kind of monotone + skew symmetric operator splitting has already been discussed in [7]. Indeed, A is maximally monotone and B is a skew-symmetric linear and continuous operator which is also maximally monotone. With an auxiliary variable $v \in \mathcal{H}$, the iteration resulting from the Douglas-Rachford splitting method reads as

$$\begin{cases} z^{k+1} = J_{\sigma B}(v^k), \\ v^{k+1} = v^k + J_{\sigma A}(2z^{k+1} - v^k) - z^{k+1}. \end{cases} \tag{1.5}$$

where $\sigma > 0$, and $J_{\sigma A} = (I + \sigma A)^{-1}$, $J_{\sigma B} = (I + \sigma B)^{-1}$ are the resolvent operators associated with σA and σB, respectively. The sequence $\{v^k\}$ can then be shown to converge weakly to a v^* such that $J_{\sigma B}(v^*)$ is a root of $A + B$, provided that such a point exists [12].

However, no matter how the operators are split, one has to solve some implicit equations. In case of (1.4), this amounts to computing the proximal mappings of F, G and solving linear equations of the form $(I + \sigma B)z^{k+1} = v^k$. In this respect, most implicit unconditionally stable methods (including the Douglas-Rachford splitting method) are quite different from forward-backward-type methods which perform forward evaluation of B at the cost of conditional stability in the form of step-size constraints, see for instance, the primal-dual method proposed in [8]. Nevertheless, the advantage of unconstrained step-sizes comes with the price that the implicit equations can in practice only be solved approximately. One way of dealing with this is to control the error in each evaluation of the resolvents such that it remains absolutely summable [9, 10, 12]: If

$$\|w^{k+1} - J_{\sigma B}(v^k)\| \leq \beta_k, \quad \|h^{k+1} - J_{\sigma A}(2w^{k+1} - v^k)\| \leq \alpha_k, \quad \text{and} \quad \sum_{k=0}^{\infty} (\beta_k + \alpha_k) < \infty, \tag{1.6}$$

then the iteration (1.5) with the resolvent evaluation replaced by w^{k+1} and h^{k+1}, respectively, still converges weakly to a root of $A + B$. This inexact Douglas-Rachford splitting method is the basis for some algorithms including the inexact ADMM [14] and the split Bregman method [16]. There, one or several Gauss Seidel iterations are introduced as an inexact solver which is much cheaper compared to solving the subproblem exactly for the whole iteration. The problem is that in practice, (1.6) is hard to verify unless a rigorous control of the error is implemented which is usually not done (see, for instance, [14, 16]). However, such an error control is crucial in order to maintain convergence.

The contribution of the paper is a preconditioned Douglas-Rachford method for the solution of (1.1) which bases on different kinds of splitting, including the skew splitting (1.4), and allows for the inexact solution of the linear subproblem without the need of controlling the error or restricting the step-size. We prove, in Theorem 2.3, weak convergence in Hilbert spaces on mild assumptions on the symmetric preconditioners. Furthermore, we study variants of the preconditioned Douglas-Rachford method for special classes of (1.1) which might increase efficiency. These methods are applied to discrete total-variation-based denoising, a classical variational problem in image processing. It turns out that in this situation, where $K = \nabla$, the symmetric Red-Black Gauss-Seidel method is preferable as a good preconditioner. In particular, any number of inner iterations for the subproblem is sufficient for convergence of the whole iteration.

Actually, introducing preconditioners for iterative methods dealing with saddle-point problems is not new. In [4, 17], preconditioned Uzawa algorithms are investigated for a special class of saddle-point problems in finite dimensions. Likewise, in [23], preconditioning techniques are studied for the class of forward-backward primal-dual algorithms introduced in [8]. All of these preconditioned iterations are conditionally stable leading to a constraint on the step size. In contrast, the preconditioned Douglas-Rachford splitting methods introduced in this paper are fully implicit and unconditionally stable.

The organization of the paper is as follows. In Section 2, we give a general framework for the preconditioned Douglas-Rachford method and introduce the notion of feasible preconditioners which is then investigated. In particular, a variety of classical iteration methods can be covered by this framework. In Section 3, we discuss the case where one can split off a linear-quadratic functional from F and G in (1.1) without losing convexity, paying particular attention to the pure linear-quadratic setting. In Section 4, we study the problem of image denoising with total-variation penalty within the framework of Sections 2 and 3. The symmetric Red-Black Gauss-Seidel method [1, 26] turns out to be a good preconditioner and yields very efficient algorithms when applied to these cases. In Section 5, we present numerical tests which demonstrate the efficiency of our preconditioned Douglas-Rachford splitting algorithms. Finally, some discussions on this technique are presented in the last section.

2. The Douglas-Rachford splitting method as a proximal point method and its preconditioned version.

2.1. The classical Douglas-Rachford iteration as a proximal point method. It is well-known that the Douglas-Rachford splitting method can be interpreted as a special instance of the proximal point algorithm [12]. Here, we give a new formulation which is convenient for the usage of preconditioners.

Let us first fix some notions and definitions. In the following, $\mathcal{H}$ always denotes a real Hilbert space. A mapping $T : \mathcal{H} \to \mathcal{H}$ is *non-expansive* if $\|Tu^1 - Tu^2\| \leq \|u^1 - u^2\|$ for all $u^1, u^2 \in \mathcal{H}$. It is *firmly non-expansive* if

$$\|Tu^1 - Tu^2\|^2 + \|(I - T)u^1 - (I - T)u^2\|^2 \leq \|u^1 - u^2\|^2 \qquad \text{for all} \quad u^1, u^2 \in \mathcal{H}.$$

We consider in particular multivalued mappings or operators, i.e. $A : \mathcal{H} \to 2^{\mathcal{H}}$ where all operations have to be understood in the pointwise sense. Such an operator A is said to be *monotone*, if $\langle v^1 - v^2, u^1 - u^2\rangle \geq 0$ whenever $v^i \in Au^i$ for $i = 1, 2$. It is *maximal* monotone, if additionally $\mathrm{rg}(I+A) = \mathcal{H}$. In this case, $(I+A)^{-1}$ is single-valued and firmly non-expansive [21].

We also utilize bounded linear operators $M : \mathcal{H} \to \mathcal{H}$. Recall that M defined to be *self-adjoint* if $M^* = M$, i.e., $\langle Mu^1, u^2\rangle = \langle u^1, Mu^2\rangle$ for all $u^1, u^2 \in \mathcal{H}$. It is *positive semi-definite* if $\langle Mu, u\rangle \geq 0$ for all $u \in \mathcal{H}$ and *positive definite* if there exists a $c > 0$ such that $\langle Mu, u\rangle \geq c\|u\|^2$ for all $u \in \mathcal{H}$. We denote positive semi-definiteness and positive definiteness by $M \geq 0$ and $M > 0$, respectively.

Now, let $A, B : \mathcal{H} \to 2^{\mathcal{H}}$ be two maximally monotone operators for which we like to solve the problem of finding $z \in \mathcal{H}$ such that $0 \in Az + Bz$. Setting $\sigma > 0$ and introducing an auxiliary variable $w \in \mathcal{H}$ with $w \in \sigma Az$, this problem can equivalently be written as

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \sigma Bz + w \\ -z + (\sigma A)^{-1}w \end{pmatrix}. \tag{2.1}$$

With $\mathcal{U} = \mathcal{H} \times \mathcal{H}$, $u = (z, w)$, and denoting by $\mathcal{A} : \mathcal{U} \to 2^{\mathcal{U}}$ the operator in (2.1), the problem becomes $0 \in \mathcal{A}u$ for which the proximal point method can be applied. Introducing a linear and continuous "preconditioner" $\mathcal{M} : \mathcal{U} \to \mathcal{U}$ the latter results in the iteration

$$0 \in \mathcal{M}(u^{k+1} - u^k) + \mathcal{A}u^{k+1} \tag{2.2}$$

with

$$\mathcal{M} = \begin{pmatrix} I & -I \\ -I & I \end{pmatrix}, \qquad \mathcal{A} = \begin{pmatrix} \sigma B & I \\ -I & (\sigma A)^{-1} \end{pmatrix}. \tag{2.3}$$

We will also write $\mathcal{M}u^k \in (\mathcal{M}+\mathcal{A})u^{k+1}$ for (2.2) and introduce the iteration mapping

$$\mathcal{T} : \mathcal{U} \to \mathcal{U}, \qquad u \mapsto (\mathcal{M}+\mathcal{A})^{-1}\mathcal{M}u. \tag{2.4}$$

Writing both (2.2) and (2.3) in terms of the components $u^k = (z^k, w^k)$ and resolvents, we see that the iteration is well-defined and amounts to

$$\begin{cases} z^{k+1} = J_{\sigma B}(z^k - w^k) \\ w^{k+1} = J_{(\sigma A)^{-1}}(2z^{k+1} - z^k + w^k). \end{cases} \tag{2.5}$$

Using Moreau's identity $J_{(\sigma A)^{-1}} + J_{\sigma A} = I$ and substituting $v^k = z^k - w^k$ it follows that

$$v^{k+1} = z^{k+1} - w^{k+1} = v^k + J_{\sigma A}(2z^{k+1} - v^k) - z^{k+1},$$

hence, (2.5) may be expressed in terms of z and v resulting in the Douglas-Rachford iteration (1.5). We have shown:

Theorem 2.1. *The classical Douglas-Rachford iteration* (1.5) *is equivalent to the proximal point method* (2.2) *with data* (2.3).

2.2. Preconditioned Douglas-Rachford iteration for saddle-point problems. We now turn to saddle-point problems of the form (1.1) and split the optimality system (1.3) according to (1.4) such that the problem becomes $0 \in Az + Bz$ for primal-dual pairs $z = (x, y)$ in $\mathcal{H} = X \times Y$. An auxiliary variable $w = (\tilde{x}, \tilde{y})$ with $w \in \sigma A z$ corresponds to

$$\begin{aligned} \tilde{x} &\in \sigma \partial F(x), \\ \tilde{y} &\in \sigma \partial G(y), \end{aligned} \tag{2.6}$$

so the optimality condition according to (2.1) can be written as

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} 0 & \sigma K^* & I & 0 \\ -\sigma K & 0 & 0 & I \\ -I & 0 & (\sigma\partial F)^{-1} & 0 \\ 0 & -I & 0 & (\sigma \partial G)^{-1} \end{pmatrix} \begin{pmatrix} x \\ y \\ \tilde{x} \\ \tilde{y} \end{pmatrix}. \tag{2.7}$$

The matrix can again be written as an operator $\mathcal{A} : \mathcal{U} \to 2^{\mathcal{U}}$. With $u = (x, y, \tilde{x}, \tilde{y})$ the equation becomes $0 \in \mathcal{A}u$.

Now, we like to introduce more flexibility with respect to the preconditioner $\mathcal{M}$. In view of primal-dual variables, it makes sense to precondition each variable separately. Introduce the following two operators such that

$$\begin{cases} N_1 : X \to X \text{ linear, continuous, self-adjoint and } N_1 - I \geq 0, \\ N_2 : Y \to Y \text{ linear, continuous, self-adjoint and } N_2 - I \geq 0. \end{cases} \tag{2.8}$$

Our goal is to use these preconditioners for the solution of linear equations involving the linear operator $I + \sigma B$ with B defined in (1.4). Therefore, we define $\mathcal{M} : \mathcal{U} \to \mathcal{U}$ according to

$$\mathcal{M} := \begin{pmatrix} N_1 & 0 & -I & 0 \\ 0 & N_2 & 0 & -I \\ -I & 0 & I & 0 \\ 0 & -I & 0 & I \end{pmatrix}. \tag{2.9}$$

It could be seen that by (2.8), $\mathcal{M} : \mathcal{U} \to \mathcal{U}$ is a linear, continuous and self-adjoint positive semi-definite operator. We will make use of the following symmetric, positive semi-definite bilinear form on $\mathcal{U}$:

$$\langle u, v\rangle_{\mathcal{M}} = \langle \mathcal{M}u, v\rangle, \qquad \forall\ u, v \in \mathcal{U} \tag{2.10}$$

as well as $\|u\|_{\mathcal{M}} = \sqrt{\langle u, u\rangle_{\mathcal{M}}}$ which can easily be verified to constitute a continuous semi-norm.

With $\mathcal{M}$ and $\mathcal{A}$ defined, one can now consider the iteration (2.2) which can again be written componentwise in terms of $x, y, \tilde{x}, \tilde{y}$. Likewise, the involved resolvents of the form $(I+(\sigma\partial F)^{-1})^{-1}$ and $(I + (\sigma\partial G)^{-1})^{-1}$ are expressible in terms of $(I + \sigma\partial F)^{-1}$ and $(I + \sigma\partial G)^{-1}$. Introducing $\bar{x} = x - \tilde{x}$ and $\bar{y} = y - \tilde{y}$, the iteration then reads as follows:

$$\begin{cases} x^{k+1} = N_1^{-1}[(N_1 - I)x^k + \bar{x}^k - \sigma K^* y^{k+1}], \\ y^{k+1} = N_2^{-1}[(N_2 - I)y^k + \bar{y}^k + \sigma K x^{k+1}], \\ \bar{x}^{k+1} = \bar{x}^k + (I + \sigma\partial F)^{-1}[2x^{k+1} - \bar{x}^k] - x^{k+1}, \\ \bar{y}^{k+1} = \bar{y}^k + (I + \sigma\partial G)^{-1}[2y^{k+1} - \bar{y}^k] - y^{k+1}. \end{cases} \tag{2.11}$$

Note that the equations for x^{k+1} and y^{k+1} are implicit. They can, however, be solved by plugging one variable into the other. One possibility is to compute x^{k+1} first, i.e.,

$$x^{k+1} = (N_1 + \sigma^2 K^* N_2^{-1} K)^{-1}\big((N_1 x^k + \bar{x}^k - x^k) + \sigma K^*(N_2^{-1}(y^k - \bar{y}^k) - y^k)\big) \tag{2.12}$$

and updating y^{k+1} afterwards, but of course, the roles of x^{k+1} and y^{k+1} may be interchanged resulting in a slightly different update rule.

Remark 2.2. *Note that for $N_1 = I$, $N_2 = I$, we recover the classical Douglas-Rachford iteration as discussed in [7]. The identity* (2.12) *however allows us to write the updates for x^{k+1} and y^{k+1} according to*

$$\begin{cases} x^{k+1} = (I + \sigma^2 K^* K)^{-1}[\bar{x}^k - \sigma K^* \bar{y}^k], \\ y^{k+1} = \bar{y}^k + \sigma K x^{k+1}. \end{cases} \tag{2.13}$$

In equation (2.23) in [7], the solution of a slightly different linear system for the update of y^{k+1} is suggested which is, in practice, usually computationally more expensive.

As the solution step for $I+\sigma^2 K^* K$ in (2.13) might be expensive, we like replace it by suitable approximative solvers, i.e., preconditioners. Let us next discuss how N_1 and N_2 can be used to achieve this goal. Set $N_2 = I$, then (2.12) amounts to solving

$$(N_1 + \sigma^2 K^* K)x^{k+1} = (N_1 - I)x^k + \bar{x}^k - \sigma K^* \bar{y}^k. \tag{2.14}$$

Introduce an operator M such that

$$N_1 = M - \sigma^2 K^* K \tag{2.15}$$

where M is assumed to be positive definite. Then we have

$$x^{k+1} = x^k + M^{-1}[\bar{x}^k - \sigma K^* \bar{y}^k - (I + \sigma^2 K^* K)x^k]. \tag{2.16}$$

Introducing $T = I + \sigma^2 K^* K$ and $b^k = \bar{x}^k - \sigma K^* \bar{y}^k$, this can be rewritten to

$$x^{k+1} = x^k + M^{-1}(b^k - Tx^k) \tag{2.17}$$

PDR **Objective:**	Solve $\min_{x\in X}\max_{y\in Y}\ \langle Kx,y\rangle + F(x) - G(y)$
Initialization:	$(x^0,\bar x^0,y^0,\bar y^0)\in X^2\times Y^2$ initial guess, $\sigma>0$ step-size, $T = I+\sigma^2K^*K$, $M = N_1+\sigma^2K^*K$
Iteration:	$$\begin{cases} b^k = \bar x^k - \sigma K^*\bar y^k, \\ x^{k+1} = x^k + M^{-1}(b^k - Tx^k), \\ y^{k+1} = \bar y^k + \sigma K x^{k+1}, \\ \bar x^{k+1} = \bar x^k + (I+\sigma\partial F)^{-1}[2x^{k+1}-\bar x^k] - x^{k+1}, \\ \bar y^{k+1} = \bar y^k + (I+\sigma\partial G)^{-1}[2y^{k+1}-\bar y^k] - y^{k+1}, \end{cases} \quad \text{(PDR)}$$

Table 2.1: The abstract preconditioned Douglas-Rachford iteration for saddle point problems of the type (1.1).

which is one iteration of the splitting method associated with the representation $T = M-(N_1-I)$. Therefore, M may be interpreted as a preconditioner for the operator equation $Tx^{k+1} = b^k$. As this variant is mainly used throughout the paper, we refer to it as the **PDR** method. It is summarized in Table 2.1.

We will see in Subsection 2.3 that we are flexible enough in the choice of M to implement various well-known preconditioners. But first, we like to establish convergence of the iteration (2.11). Recall that (2.11) amounts, via (2.4), to a fixed-point operator $\mathcal{T}:\mathcal{U}\to\mathcal{U}$ which is defined by $u^{k+1} = \mathcal{T}u^k$ according to (2.11) for each $u^k\in\mathcal{U}$.

THEOREM 2.3. *If a solution of the saddle point problem* (1.1) *exists and condition* (2.8) *is satisfied, then the iteration* (2.11) *converges weakly to a fixed-point $u^* = (x^*,y^*,\bar x^*,\bar y^*)$. The pair (x^*,y^*) is a solution of the saddle point problem.*

If $N_2 = I$, then the iteration equivalent to the method (PDR).

REMARK 2.4. *Theorem 2.3 immediately implies the weak convergence of the sequence*

$$\begin{cases} x_{\text{test}}^{k+1} := (I+\sigma\partial F)^{-1}[2x^{k+1}-\bar x^k] \\ y_{\text{test}}^{k+1} := (I+\sigma\partial G)^{-1}[2y^{k+1}-\bar y^k] \end{cases} \tag{2.18}$$

to (x^*,y^*), the weak limit of $\{(x^k,y^k)\}$. As the resolvents are proximal mappings, their components are always contained in $\operatorname{dom}\partial F$ *and* $\operatorname{dom}\partial G$*. In particular, $F(x_{\text{test}}^k)$ and $G(y_{\text{test}}^k)$ are finite for each k. This might be helpful to evaluate primal-dual gap functions, for instance for the L^2-TV denoising problem [8] and will be discussed in Section 5.*

The proof of Theorem 2.3 is done in five steps, each formulated as a lemma.

LEMMA 2.5. *The mapping $\mathcal{T}:\mathcal{U}\to\mathcal{U}$*

$$u\mapsto(\mathcal{M}+\mathcal{A})^{-1}\mathcal{M}u \tag{2.19}$$

is well-defined. It is $\mathcal{M}$-firmly non-expansive in the sense that for $u^1,u^2\in\mathcal{U}$,

$$\|\mathcal{T}u^1-\mathcal{T}u^2\|_{\mathcal{M}}^2 + \|(I-\mathcal{T})u^1-(I-\mathcal{T})u^2\|_{\mathcal{M}}^2 \le \|u^1-u^2\|_{\mathcal{M}}^2. \tag{2.20}$$

Finally, there exists a constant $C>0$ such that for all $u^1,u^2\in\mathcal{U}$,

$$\|\mathcal{T}u^1-\mathcal{T}u^2\| \le C\|u^1-u^2\|_{\mathcal{M}} \tag{2.21}$$

Proof. Note that $\mathcal{M}u \in (\mathcal{M}+\mathcal{A})u'$ is equivalent to

$$\begin{cases} N_1 x' + \sigma K^* y' = N_1 x - \tilde{x} \\ N_2 y' - \sigma K x' = N_2 y - \tilde{y} \end{cases} \quad \text{and} \quad \begin{cases} \tilde{x}' = (I - (I + \sigma\partial F)^{-1})[\tilde{x} - x + 2x'] \\ \tilde{y}' = (I - (I + \sigma\partial G)^{-1})[\tilde{y} - y + 2y'], \end{cases} \tag{2.22}$$

where $u' = (x', y', \tilde{x}', \tilde{y}')$. The first two equations correspond to solving a linear equation involving the operator

$$N = \begin{pmatrix} N_1 & \sigma K^* \\ -\sigma K & N_2 \end{pmatrix} \tag{2.23}$$

which is linear and continuous mapping $N : \mathcal{H} \to \mathcal{H}$ on the Hilbert space $\mathcal{H} = X \times Y$. One verifies with the help of the Woodbury formula that its inverse has to be the product

$$N^{-1} = \begin{pmatrix} (N_1 + \sigma^2 K^* N_2^{-1} K)^{-1} & 0 \\ 0 & (N_2 + \sigma^2 K N_1^{-1} K^*)^{-1} \end{pmatrix} \begin{pmatrix} I & -\sigma K^* N_2^{-1} \\ \sigma K N_1^{-1} & I \end{pmatrix}$$

which is linear and continuous. Note that the operators whose inverse is taken are all positive definite as $N_1 - I \geq 0$, $N_2 - I \geq 0$ and $N_1 + \sigma^2 K^* N_2^{-1} K \geq N_1$, $N_2 + \sigma^2 K N_1^{-1} K^* \geq N_2$. Therefore, one can uniquely solve the first two equations of (2.22) for any given u. The solution (x', y') is given by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = N^{-1} \begin{pmatrix} (N_1 - I)x + (x - \tilde{x}) \\ (N_2 - I)y + (y - \tilde{y}) \end{pmatrix}, \tag{2.24}$$

and can be plugged in the third and fourth equation of (2.22) which yields a unique $(\tilde{x}', \tilde{y}')$ as the resolvent mappings are unique. Hence $\mathcal{T} : u \mapsto u'$ is well-defined.

Now, for any $u^1 = (x^1, y^1, \tilde{x}^1, \tilde{y}^1)$, $u^2 = (x^2, y^2, \tilde{x}^2, \tilde{y}^2) \in \mathcal{U}$, it holds that

$$(u^i)' = \begin{pmatrix} \sigma K^* y^i + \tilde{x}^i \\ -\sigma K x^i + \tilde{y}^i \\ -x^i + v^i \\ -y^i + w^i \end{pmatrix} \in \mathcal{A}u^i \qquad \text{if} \qquad \begin{cases} v^i \in (\sigma\partial F)^{-1}(\tilde{x}^i), \\ w^i \in (\sigma\partial G)^{-1}(\tilde{y}^i) \end{cases}$$

for $i = 1, 2$. It it then easily checked that

$$\langle (u^1)' - (u^2)', u^1 - u^2 \rangle = \langle v^1 - v^2, \tilde{x}^1 - \tilde{x}^2 \rangle + \langle w^1 - w^2, \tilde{y}^1 - \tilde{y}^2 \rangle \geq 0,$$

since $(\sigma\partial F)^{-1}$ and $(\sigma\partial G)^{-1}$ are both monotone operators. Hence, $\mathcal{A}$ is monotone in $\mathcal{U}$. By definition of $\mathcal{T}$, we can choose $(u^1)' \in \mathcal{A}\mathcal{T}u^1$, $(u^2)' \in \mathcal{A}\mathcal{T}u^2$, such that

$$\mathcal{M}\mathcal{T}u^1 + (u^1)' = \mathcal{M}u^1, \qquad \mathcal{M}\mathcal{T}u^2 + (u^2)' = \mathcal{M}u^2.$$

Then, we have $\langle (u^1)' - (u^2)', \mathcal{T}u^1 - \mathcal{T}u^2 \rangle \geq 0$ as $\mathcal{A}$ is monotone and, consequently,

$$\begin{aligned} \|\mathcal{T}u^1 - \mathcal{T}u^2\|_{\mathcal{M}}^2 &\leq \langle \mathcal{M}(\mathcal{T}u^1 - \mathcal{T}u^2), \mathcal{T}u^1 - \mathcal{T}u^2 \rangle + \langle (u^1)' - (u^2)', \mathcal{T}u^1 - \mathcal{T}u^2 \rangle \\ &= \langle \mathcal{M}(u^1 - u^2), \mathcal{T}u^1 - \mathcal{T}u^2 \rangle = \langle \mathcal{T}u^1 - \mathcal{T}u^2, u^1 - u^2 \rangle_{\mathcal{M}}. \end{aligned}$$

Employing this estimate, the desired inequality (2.20) is obtained as follows:

$$\begin{aligned}
&\|\mathcal{T}u^1 - \mathcal{T}u^2\|_{\mathcal{M}}^2 + \|(I-\mathcal{T})u^1 - (I-\mathcal{T})u^2\|_{\mathcal{M}}^2 \\
&\quad = \|\mathcal{T}u^1 - \mathcal{T}u^2\|_{\mathcal{M}}^2 + \|(u_1 - u_2) - (\mathcal{T}u^1 - \mathcal{T}u^2)\|_{\mathcal{M}}^2 \\
&\quad = \|\mathcal{T}u^1 - \mathcal{T}u^2\|_{\mathcal{M}}^2 + \|u^1 - u^2\|_{\mathcal{M}}^2 - 2\langle \mathcal{T}u^1 - \mathcal{T}u^2, u^1 - u^2\rangle_{\mathcal{M}} \\
&\qquad + \|\mathcal{T}u^1 - \mathcal{T}u^2\|_{\mathcal{M}}^2 \\
&\quad = \|u^1 - u^2\|_{\mathcal{M}}^2 - 2\big(\langle \mathcal{T}u^1 - \mathcal{T}u^2, u^1 - u^2\rangle_{\mathcal{M}} - \|\mathcal{T}u^1 - \mathcal{T}u^2\|_{\mathcal{M}}^2\big) \\
&\quad \le \|u^1 - u^2\|_{\mathcal{M}}^2.
\end{aligned}$$

To verify (2.21), observe that for $u = (x, y, \tilde{x}, \tilde{y})$,

$$\begin{aligned}
\|u\|_{\mathcal{M}}^2 &= \langle N_1 x, x\rangle - \|x\|^2 + \|x\|^2 - 2\langle x, \tilde{x}\rangle + \|\tilde{x}\|^2 \\
&\quad + \langle N_2 y, y\rangle - \|y\|^2 + \|y\|^2 - 2\langle y, \tilde{y}\rangle + \|\tilde{y}\|^2 \\
&= \langle (N_1 - I)x, x\rangle + \langle (N_2 - I)y, y\rangle + \|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2.
\end{aligned}$$

As $N_1 - I$ and $N_2 - I$ are positive semi-definite by (2.8), $(N_1 - I)^{1/2}$ and $(N_1 - I)^{1/2}$ exist (by Theorem VI.9 of [24]) and are bounded linear operators. We can hence write

$$\|u\|_{\mathcal{M}}^2 = \|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2 + \|(N_1 - I)^{1/2}x\|^2 + \|(N_2 - I)^{1/2}y\|^2. \tag{2.25}$$

and get, with $\mathcal{T}u = (x', y', \tilde{x}', \tilde{y}')$ and by virtue of (2.24), that

$$\begin{aligned}
\|x'\|^2 + \|y'\|^2 &\le 2\|N^{-1}\|^2\big(\|(N_1 - I)^{1/2}\|^2\|(N_1 - I)^{1/2}x\|^2 + \|x - \tilde{x}\|^2\big) \\
&\quad + 2\|N^{-1}\|^2\big(\|(N_2 - I)^{1/2}\|^2\|(N_2 - I)^{1/2}y\|^2 + \|y - \tilde{y}\|^2\big) \\
&\le C_0^2\|u\|_{\mathcal{M}}^2
\end{aligned}$$

for some $C_0 > 0$. If $u^1, u^2 \in \mathcal{U}$ this implies

$$\|(x^1)' - (x^2)'\|^2 + \|(y^1)' - (y^2)'\|^2 \le C_0^2\|u^1 - u^2\|_{\mathcal{M}}^2$$

by linearity of (2.24). Note that both $I - (I + \sigma\partial F)^{-1}$ and $I - (I + \sigma\partial G)^{-1}$ are non-expansive (by Moreau's identity), hence (2.22) allows to estimate

$$\begin{cases}
\|(\tilde{x}^1)' - (\tilde{x}^2)'\| \le \big(\|(\tilde{x}^1 - \tilde{x}^2) - (x^1 - x^2)\| + 2\|(x^1)' - (x^2)'\|\big) \le C_1\|u^1 - u^2\|_{\mathcal{M}}, \\
\|(\tilde{y}^1)' - (\tilde{y}^2)'\| \le \big(\|(\tilde{y}^1 - \tilde{y}^2) - (y^1 - y^2)\| + 2\|(y^1)' - (y^2)'\|\big) \le C_2\|u^1 - u^2\|_{\mathcal{M}}
\end{cases}$$

for suitable $C_1, C_2 > 0$. This immediately gives the existence of a $C > 0$ such that (2.21) holds. ◻

Fixed-points of $\mathcal{T}$ are indeed solutions of the considered saddle-point problem in a certain sense.

Lemma 2.6. *A $u^* = (x^*, y^*, \tilde{x}^*, \tilde{y}^*)$ is a fixed-point of $\mathcal{T}$ if and only if (x^*, y^*) is a solution of the saddle-point problem* (1.1) *and* $(\tilde{x}^*, \tilde{y}^*) = (-\sigma K^* y^*, \sigma K x^*)$.

Proof. Let u^* be a fixed-point of $\mathcal{T}$ which is equivalent to $0 \in \mathcal{A}u^*$ where $\mathcal{A}$ is the operator according to (2.7). The first two lines are, in turn, equivalent to $\tilde{x}^* = -\sigma K^* y^*$ and $\tilde{y}^* = \sigma K x^*$ while the last two lines can be characterized by $\tilde{x}^* \in \sigma\partial F(x^*)$ and $\tilde{y}^* \in \sigma\partial G(y^*)$. Plugging in the former into the latter yields the equivalence to (1.3) which is finally a characterization of the saddle points. ◻

The operator $\mathcal{T}$ has the important *asymptotic regularity* property [22].

Lemma 2.7. *If there is a solution to the saddle point problem* (1.1), *then the sequence* $\{u^k\}$ *given by* $u^k = \mathcal{T}^k u^0$, $u^0 \in \mathcal{U}$, *satisfies*

(i) $\{u^k\}$ *is bounded,*
(ii) $\|u^{k+1}-u^k\| \to 0$ *as* $k \to \infty$.

Proof. (i) According to Lemma 2.6, a solution (x^*, y^*) of the saddle-point problem gives a fixed point of $\mathcal{T}$ via $u^* = (x^*, y^*, \tilde{x}^*, \tilde{y}^*)$. Hence, by Lemma 2.5 we have

$$\|u^k - u^*\|_{\mathcal{M}} \le \|u^{k-1} - u^*\|_{\mathcal{M}} \le \cdots \le \|u^0 - u^*\|_{\mathcal{M}}, \tag{2.26}$$

for each k which implies that $\{\|u^k - u^*\|_{\mathcal{M}}\}$ is bounded. From (2.21) then follows that $\{\|\mathcal{T}u^k - \mathcal{T}u^*\|\}$ and hence, $\{u^k\}$ must be bounded.

(ii) As u^* is a fixed point, the $\mathcal{M}$-firm nonexpansiveness of $\mathcal{T}$ in (2.20) gives

$$\|\mathcal{T}u^k - \mathcal{T}u^*\|_{\mathcal{M}}^2 + \|(I-\mathcal{T})u^k - (I-\mathcal{T})u^*\|_{\mathcal{M}}^2 \le \|u^k - u^*\|_{\mathcal{M}}^2$$

which is equivalent to

$$\|u^{k+1} - u^*\|_{\mathcal{M}}^2 + \|u^k - u^{k+1}\|_{\mathcal{M}}^2 \le \|u^k - u^*\|_{\mathcal{M}}^2.$$

Induction with respect to k yields

$$\sum_{k=0}^{\infty} \|u^k - u^{k+1}\|_{\mathcal{M}}^2 \le \|u^0 - u^*\|_{\mathcal{M}}^2. \tag{2.27}$$

In particular, $\|u^k - u^{k+1}\|_{\mathcal{M}} \to 0$ as $k \to +\infty$ for which again (2.21) implies that $\|u^k - u^{k+1}\| = \|\mathcal{T}u^{k-1} - \mathcal{T}u^k\| \to 0$ as $k \to \infty$. □

Next, we show that $I - \mathcal{T}$ is demiclosed which is needed for weak limits of the iteration being fixed points. As $\mathcal{T}$ is only $\mathcal{M}$-non-expansive, this does not immediately follow from the results in [22].

LEMMA 2.8. *The operator* $I - \mathcal{T}$ *is demiclosed, i.e.,* $u^k \rightharpoonup u$ *and* $(I-\mathcal{T})u^k \to v$ *implies* $(I-\mathcal{T})u = v$.

Proof. Denote the components of u by $(x, y, \tilde{x}, \tilde{y})$, analogously for u^k and let $v = (z_1, z_2, w_1, w_2)$. As linear and continuous operators are weakly sequentially continuous, we have, according to (2.24) for $\mathcal{T}u = (x', y', \tilde{x}', \tilde{y}')$ that

$$\begin{pmatrix} x^k \\ y^k \end{pmatrix} - N^{-1}\begin{pmatrix} (N_1 - I)x^k + (x^k - \tilde{x}^k) \\ (N_2 - I)y^k + (y^k - \tilde{y}^k) \end{pmatrix} \rightharpoonup \begin{pmatrix} x \\ y \end{pmatrix} - N^{-1}\begin{pmatrix} (N_1 - I)x + (x - \tilde{x}) \\ (N_2 - I)y + (y - \tilde{y}) \end{pmatrix} = \begin{pmatrix} x - x' \\ y - y' \end{pmatrix} \tag{2.28}$$

Since $(I - \mathcal{T})u^k \to v$, this shows that the first two components v and $u - \mathcal{T}u$ are equal, i.e., $z_1 = x - x'$ and $z_2 = y - y'$.

For the remaining two components, observe that the $\mathcal{M}$-seminorm is continuous on $\mathcal{U}$ (see (2.9)), so we have $\|u^k - \mathcal{T}u^k - v\|_{\mathcal{M}} \to 0$. Hence, as $\mathcal{T}$ is $\mathcal{M}$-non-expansive, the respective weak and strong convergence yields

$$\begin{aligned}
\liminf_{k\to\infty} \|u^k - u\|_{\mathcal{M}}^2 &\ge \liminf_{k\to\infty} \|\mathcal{T}u^k - \mathcal{T}u\|_{\mathcal{M}}^2 \\
&= \liminf_{k\to\infty} \|u^k - \mathcal{T}u - v\|_{\mathcal{M}}^2 - 2\lim_{k\to\infty} \langle u^k - \mathcal{T}u - v, u^k - \mathcal{T}u^k - v\rangle_{\mathcal{M}} \\
&\quad + \lim_{k\to\infty} \|u^k - \mathcal{T}u^k - v\|_{\mathcal{M}}^2 = \liminf_{k\to\infty} \|u^k - \mathcal{T}u - v\|_{\mathcal{M}}^2.
\end{aligned}$$

On the other hand, if $u' = \mathcal{T}u + v$ satisfies $\|u' - u\|_{\mathcal{M}} > 0$, then by weak convergence,

$$\begin{aligned}
\liminf_{k\to\infty} \|u^k - u\|_{\mathcal{M}}^2 &= \liminf_{k\to\infty} \|u^k - u' + u' - u\|_{\mathcal{M}}^2 \\
&= \liminf_{k\to\infty} \|u^k - u'\|_{\mathcal{M}}^2 + \|u' - u\|_{\mathcal{M}}^2 + \lim_{k\to\infty} 2\langle u^k - u', u' - u\rangle_{\mathcal{M}} \\
&> \liminf_{k\to\infty} \|u^k - u'\|_{\mathcal{M}}^2
\end{aligned}$$

which contradicts the above inequality. Hence $\|u - \mathcal{T}u - v\|_{\mathcal{M}} = 0$. By (2.25) this implies $(x - x') - (\tilde{x} - \tilde{x}') = (z_1 - w_1)$ and $(y - y') - (\tilde{y} - \tilde{y}') = (z_2 - w_2)$. As $z_1 = x - x'$ and $z_2 = y - y'$, all components of $u - \mathcal{T}u$ and v coincide. □

Finally, with these preparations, we can now prove our main Theorem 2.3 which also benefits from the celebrated idea of [22].

Proof of Theorem 2.3. Denote the set of the fixed points of $\mathcal{T}$ by $\mathcal{F}$ which is a closed and convex set. As a saddle-point for (1.1) exists, $\mathcal{F}$ is not empty according to Lemma 2.6. For any $u^* \in \mathcal{F}$, the sequence $\{\|u^k - u_0^*\|_{\mathcal{M}}\}$ is non-increasing and hence possesses a limit $d(u^*)$. This defines a non-negative function on $\mathcal{F}$.

According to Lemma 2.7 (i), $\{u^k\}$ is bounded. Now, each subsequence possesses a weakly convergent subsequence $\{u^{k_i}\}$ with limit u^* which must be a fixed point of $\mathcal{T}$: From Lemma 2.7 (ii) follows that $(I - \mathcal{T})u^{k_i} \to 0$, hence demiclosedness in Lemma 2.8 implies $u^* = \mathcal{T}u^*$.

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

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

and observe that the right-hand side converges as $k \to \infty$ to some $c \in \mathbf{R}$. Plugging in $\{u^{k_i}\}$ and $\{u^{l_i}\}$ on the left-hand side implies, by weak convergence that $\langle u^*, u^{**} - u^* \rangle_{\mathcal{M}} = \langle u^{**}, u^{**} - u^* \rangle_{\mathcal{M}}$. Consequently, $\|u^* - u^{**}\|_{\mathcal{M}} = 0$. Using that both u^* and u^{**} are fixed-points yields, by virtue of (2.21):

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

Thus, $u^{**} = u^*$ so u^* is the only weak accumulation point. As each subsequence of $\{u^k\}$ possesses a weakly convergent subsequence, we get $u^k \rightharpoonup u^*$ for the whole sequence. Recalling that $\bar{x}^k = x^k - \tilde{x}^k$, $\bar{y}^k = y^k - \tilde{y}^k$ and noting that $\bar{u}^* = (x^*, y^*, x^* - \tilde{x}^*, y^* - \tilde{y}^*)$ is a fixed point of (2.11) yields the convergence result.

Finally, we have already seen that for $N_2 = I$, (PDR) is equivalent to (2.11). □

2.3. Feasible preconditioners. In view of Theorem 2.3, choosing $N_2 = I$ leads to $M = N_1 + \sigma^2 K^* K$ being a preconditioner for $T = I + \sigma^2 K^* K$. The other way around: If we choose M as a preconditioner for T, then N_1 satisfies (2.8) if and only if $M : X \to X$ is a linear, continuous and self-adjoint operator for which $N_1 = M - T$ is positive semi-definite. In this case, Theorem 2.3 yields the convergence of the iteration (PDR).

Definition 2.9. *Let $M, T : X \to X$ be linear, continuous, self-adjoint and positive semi-definite. Then, M is called a* feasible preconditioner *for T if M is positive definite and $M - T$ is positive semi-definite.*

Our aim is to state general results about when a given preconditioner is feasible which may easily be checked in practice. We start with some obvious examples for feasible preconditioners.

example 2.10.

- *Obviously, $M = T$ is a feasible preconditioner for T. This choice reproduces the original Douglas-Rachford iteration for* (1.1) *according to the splitting* (1.4).
- *The choice $M = \lambda I$ with $\lambda \geq 1 + \sigma^2\|K\|^2$ also yields a feasible preconditioner. This is corresponding to the Richardson method. The update for x^{k+1} can be seen as an explicit step while the condition on λ realizes some kind of step-size restriction.*

example 2.11. *If X is finite-dimensional, then $M = D$, the diagonal part of T, i.e., the Jacobi method suggests itself as a preconditioner. However, unless T is diagonal, the positive semi-definiteness of $M - T$ can not be established.*

Nevertheless, choosing $M = (\lambda+1)D$ *where* $\lambda \geq \lambda_{\max}(T-D)$ *the greatest eigenvalue of* $T-D$*, we can verify that*

$$\langle (M-T)x, x\rangle \geq \langle \lambda D x, x\rangle - \langle (T-D)x, x\rangle \geq \langle \lambda(D-I)x, x\rangle \geq 0$$

the latter since each diagonal entry of T *is at least* 1*. The preconditioner* M *is then feasible and corresponds to a* damped Jacobi *method. The condition on* λ *again realizes what can be interpreted as a step-size restriction.*

It may occur that a preconditioner for T is not symmetric. Symmetry can, however, easily be achieved by composing it with the adjoint preconditioner which leads to a feasible preconditioner in many cases.

Proposition 2.12. *For* $M_0 : X \to X$ *linear, continuous such that* $M_0 - \frac{1}{2}T$ *is positive definite,* M *according to*

$$M = M_0(M_0 + M_0^* - T)^{-1} M_0^*$$

is a well-defined feasible preconditioner. In (PDR)*, it corresponds to*

$$\begin{cases} x^{k+1/2} = x^k + M_0^{-1}(b^k - Tx^k) \\ \quad x^{k+1} = x^{k+1/2} + M_0^{-*}(b^k - Tx^{k+1/2}) \end{cases} \tag{2.29}$$

Proof. Observe that the update step (2.29) corresponds to

$$x^{k+1} = (I - M_0^{-*}T)(I - M_0^{-1}T)x^k + (M_0^{-*} + M_0^{-1} - M_0^{-*}TM_0^{-1})b^k$$

which is of the form $x^{k+1} = x^k + M^{-1}(b^k - Tx^k)$ in (PDR) if M is chosen as

$$M = (M_0^{-*} + M_0^{-1} - M_0^{-*}TM_0^{-1})^{-1} = M_0(M_0 + M_0^* - T)^{-1}M_0^*.$$

Now, $M_0 + M_0^* - T$ is symmetric and since $M_0 - \frac{1}{2}T$ is positive definite,

$$\langle (M_0 + M_0^* - T)x, x\rangle = 2\langle (M_0 - \tfrac{1}{2}T)x, x\rangle \geq c\|x\|^2$$

for each $x \in X$ and some $c > 0$. Hence $(M_0 + M_0^* - T)^{-1}$ exists and consequently, M is well-defined and symmetric. Finally, writing $D = (M_0 + M_0^* - T)$ and $E = M_0 - T$:

$$M - T = (D - E^*)D^{-1}(D - E) - T = D - E - E^* + E^*D^{-1}E - T = E^*D^{-1}E$$

where the right-hand side is positive semi-definite since D is positive definite. □

example 2.13. *If* X *is finite-dimensional,* T *can be split into*

$$T = D - E - E^*$$

where D *is the diagonal part and* $-E$ *represents the strict lower triangular part. Choosing* $M_0 = \frac{1}{\omega}D - E$ *for* $\omega \in \,]0,2[$ *yields the well-known successive over-relaxation (SOR) method which becomes the Gauss-Seidel method for* $\omega = 1$*. We verify*

$$\langle (M_0 - \tfrac{1}{2}T)x, x\rangle = (\tfrac{1}{\omega} - \tfrac{1}{2})\langle Dx, x\rangle + \tfrac{1}{2}\langle (E^* - E)x, x\rangle = (\tfrac{1}{\omega} - \tfrac{1}{2})\langle Dx, x\rangle \geq (\tfrac{1}{\omega} - \tfrac{1}{2})\|x\|^2,$$

since $T = I + \sigma^2 K^*K$ *implies that each diagonal entry is at least* 1*, yielding the positive semi-definiteness. The update* (2.29) *then corresponds to the* symmetric successive over-relaxation

(SSOR) *method (the* symmetric Gauss-Seidel *method for* $\omega = 1$*) and establishes a feasible preconditioner. We denote the preconditioners by* M_{SSOR} *and* M_{SGS} *for SSOR and the* symmetric Gauss-Seidel *method, respectively.*

Feasible preconditioners can be iterated giving a new feasible preconditioner.

PROPOSITION 2.14. *Let* M *be a feasible preconditioner for* T *and* $n \geq 1$*. Then, applying the preconditioner* n *times, i.e.,*

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

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

Proof. We first prove that for any linear, self-adjoint and positive definite operator B in the Hilbert space X, the following three statements are equivalent: $B - T \geq 0$, $I - B^{-1/2}TB^{-1/2} \geq 0$ and $\sigma(I - B^{-1}T) \subset [0, \infty[$ where σ denotes the spectrum. As B is positive definite, the square root $B^{1/2}$ exists and is a linear, self-adjoint and positive operator. The same holds true for $B^{-1/2}$. We have that $B - T \geq 0$ is equivalent to

$$\langle Bx, x\rangle - \langle Tx, x\rangle \geq 0, \ \forall x \in X \iff \langle y, y\rangle - \langle B^{-\frac{1}{2}}TB^{-\frac{1}{2}}y, y\rangle \geq 0, \ \forall y \in X,$$

i.e., $I - B^{-1/2}TB^{-1/2} \geq 0$, hence the first equivalence. In turn, the latter holds if and only if $\sigma(I - B^{-1/2}TB^{-1/2}) \subset [0, \infty[$ (by Proposition 3.D.(c) of [19]). However, as $I - B^{-1}T = B^{-1/2}(I - B^{-1/2}TB^{-1/2})B^{1/2}$, we have $\sigma(I - B^{-1}T) = \sigma(I - B^{-1/2}TB^{-1/2})$ and hence, the second equivalence.

Now, one can see by induction that n-fold application of the preconditioner M corresponds to $x^{k+1} = x^k + M_n^{-1}(b^k - Tx^k)$ where M_n^{-1} satisfies the formal recursion

$$M_{n+1}^{-1} = M^{-1} + M_n^{-1} - M^{-1}TM_n^{-1} = M_n^{-1} + M^{-1} - M_n^{-1}TM^{-1}$$

with $M_1^{-1} = M^{-1}$. In particular, each M_n^{-1} is symmetric and we have the identity $I - M_n^{-1}T = (I - M^{-1}T)^n$ for each n. Hence,

$$M_{n+1}^{-1} = M_n^{-1} + (I - M^{-1}T)^n M^{-1} = M_n^{-1} + M^{-1/2}(I - M^{-1/2}TM^{-1/2})^n M^{-1/2}$$

As $M - T \geq 0$, the operator $I - M^{-1/2}TM^{-1/2}$ is positive semi-definite, so we see again by induction that each M_n^{-1} is positive definite and M_n is well-defined as well as positive definite.

Since $\sigma(I - M_n^{-1}T) = \sigma((I - M^{-1}T)^n)$, we have $\sigma(I - M_n^{-1}T) \subset [0, \infty[$ as $\sigma(I - M^{-1}T) \subset [0, \infty[$, the latter by the equivalence statement at the beginning of the proof. It follows that $M_n - T \geq 0$, hence M_n is feasible. ☐

REMARK 2.15. *It could analogously be seen that if* $M - T$ *is positive definite, then* $qI - M^{-1/2}TM^{-1/2} \geq 0$ *for some* $q \in [0, 1[$*, hence* $\sigma(I - M^{-1}T) = \sigma(I - M^{-1/2}TM^{-1/2}) \subset [0, q]$*, in particular,* $I - M^{-1}T$ *is a strong contraction. Hence,* $I - M_n^{-1}T = (I - M^{-1}T)^n \to 0$ *as* $n \to \infty$ *in the operator norm and with linear convergence rate. Consequently,* $M_n \to T^{-1}$ *as* $n \to \infty$ *in the same sense, a property we would indeed expect from a preconditioner.*

With the notations T, D, M_{SGS}, M_{SSOR} as before in this section and by using SGS and DR to represent the *symmetric Gauss-Seidel* method and the original Douglas-Rachford splitting method, we can give the Table 2.2 which summarizes the properties for the different preconditioners and may help making a specific choice.

REMARK 2.16. *Throughout this section, we derive symmetric preconditioners for the symmetric operator* $I + \sigma^2 K^*K$ *as a consequence of the convergence result. This seems to be a natural approach. Indeed, it has been observed in [3, p. 450]: "In general, preconditioned iterative techniques for symmetric problems are much more effective when applied with symmetric preconditioners. The use of a nonsymmetric preconditioner is inappropriate in this case".*

	Example 2.10		Example 2.11		Example 2.13	
Preconditioner	T	λI	D	$(\lambda+1)D$	M_{SGS}	M_{SSOR}
Feasible?	Yes	Yes	No	Yes	Yes	Yes
Conditions	—	$\lambda \geq 1+\sigma^2\\|K\\|^2$	—	$\lambda \geq \lambda_{\max}(T-D)$	—	$\omega \in (0,2)$
Iteration type	DR	Richardson	Jacobi	Damped Jacobi	SGS	SSOR

Table 2.2: Summary of the properties of the discussed preconditioners.

3. Linear-quadratic functionals. In this section, we like to discuss the case where one can split off a linear-quadratic functional from F and G in (1.1) without losing convexity. We assume the following:

$$\begin{cases} F(x) = \frac{1}{2}\langle Qx, x\rangle + \langle f, x\rangle + \tilde{F}(x), \\ G(y) = \frac{1}{2}\langle Ry, y\rangle + \langle g, y\rangle + \tilde{G}(y) \end{cases} \tag{3.1}$$

for $Q : X \to X$, $R : Y \to Y$ linear, continuous, self-adjoint positive semi-definite, $f \in X$, $g \in Y$ and $\tilde{F} : X \to \mathbf{R}_\infty$, $\tilde{G} : Y \to \mathbf{R}_\infty$ proper, convex and lower semi-continuous on X and Y, respectively. It will turn out that this allows for variants of the preconditioned Douglas-Rachford splitting algorithm which also incorporate Q and R into the preconditioner.

3.1. The general case. First note the optimality conditions for the saddle-point problem (1.1) with (3.1):

$$\begin{cases} 0 \in Qx + f + K^*y + \partial\tilde{F}(x), \\ 0 \in Ry + g - Kx + \partial\tilde{G}(y). \end{cases} \tag{3.2}$$

This allows for the splitting

$$A = \begin{pmatrix} \sigma\partial\tilde{F} & 0 \\ 0 & \sigma\partial\tilde{G} \end{pmatrix}, \qquad B = \begin{pmatrix} \sigma Q + \sigma\mathbf{1}_f & \sigma K^* \\ -\sigma K & \sigma R + \sigma\mathbf{1}_g \end{pmatrix}.$$

Here and in the following, $\mathbf{1}_f$ denotes the constant mapping which yields f for each argument (analogous for g). Proceeding along the lines of Subsection 2.2, we introduce

$$\mathcal{A} = \begin{pmatrix} \sigma Q + \sigma\mathbf{1}_f & \sigma K^* & I & 0 \\ -\sigma K & \sigma R + \sigma\mathbf{1}_g & 0 & I \\ -I & 0 & (\sigma\partial\tilde{F})^{-1} & 0 \\ 0 & -I & 0 & (\sigma\partial\tilde{G})^{-1} \end{pmatrix} \tag{3.3}$$

and again perform, with $\mathcal{M}$ from (2.9), the proximal point iteration $0 \in \mathcal{M}(u^{k+1} - u^k) + \mathcal{A}u^{k+1}$ for $u^k = (x^k, y^k, \tilde{x}^k, \tilde{y}^k)$. Setting again $\bar{x}^k = x^k - \tilde{x}^k$, $\bar{y}^k = y^k - \bar{y}^k$ yields the iteration

$$\begin{cases} x^{k+1} = (N_1 + \sigma Q)^{-1}\big[(N_1 - I)x^k + \bar{x}^k - \sigma(K^*y^{k+1} + f)\big], \\ y^{k+1} = (N_2 + \sigma R)^{-1}\big[(N_2 - I)y^k + \bar{y}^k + \sigma(Kx^{k+1} - g)\big], \\ \bar{x}^{k+1} = \bar{x}^k + (I + \sigma\partial\tilde{F})^{-1}[2x^{k+1} - \bar{x}^k] - x^{k+1}, \\ \bar{y}^{k+1} = \bar{y}^k + (I + \sigma\partial\tilde{G})^{-1}[2y^{k+1} - \bar{y}^k] - y^{k+1}, \end{cases} \tag{3.4}$$

where again the first two equations are coupled. One can see that Theorem 2.3 can easily be modified to yield weak convergence of this iteration under assumption (2.8).

Remark 3.1. *Let us discuss a special case where the iteration becomes similar to* (PDR). *One sees, for instance, if $N_2 = \mu I$ and $R = \lambda I$ for some scaling parameter $\mu \geq 1$ and $\lambda \geq 0$, then the method reads as*

$$\begin{cases} b^k = \bar{x}^k - \frac{\sigma}{\mu+\sigma\lambda} K^*((\mu-1)y^k + \bar{y}^k - \sigma g) - \sigma f \\ x^{k+1} = x^k + M^{-1}(b^k - Tx^k) \\ y^{k+1} = \frac{1}{\mu+\sigma\lambda}\big((\mu-1)y^k + \bar{y}^k + \sigma(Kx^{k+1} - g)\big) \\ \bar{x}^{k+1} = \bar{x}^k + (I + \sigma\partial\tilde{F})^{-1}[2x^{k+1} - \bar{x}^k] - x^{k+1}, \\ \bar{y}^{k+1} = \bar{y}^k + (I + \sigma\partial\tilde{G})^{-1}[2y^{k+1} - \bar{y}^k] - y^{k+1}, \end{cases} \tag{3.5}$$

where $M = N_1 + \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda}K^*K$ and $T = I + \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda}K^*K$. As $N_1 - I = M - T$, we have that also this iteration converges weakly if M is a feasible preconditioner for T.

Of course, in this situation we have to find feasible preconditioners for $T = I + \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda}K^*K$ or, more general, $Q_1 + Q_2$ for $Q_1 > 0, Q_2 \geq 0$. If one has, in this case, a feasible preconditioner for Q_1, it may not be feasible for $Q_1 + Q_2$. The following procedure can then be used:

Proposition 3.2. *Let $Q_1, Q_2 : X \to X$ be linear, continuous, self-adjoint, $Q_1 > 0$, $Q_2 \geq 0$, M be a feasible preconditioner for $T = Q_1$ and let $M - Q_2$ be boundedly invertible. Then, for each $K_2 : X \to \tilde{X}$ linear, continuous with values in the Hilbert space $\tilde{X}$ and such that $Q_2 = K_2^*K_2$, the operator*

$$\tilde{M} = M + K_2^*(I - K_2M^{-1}K_2^*)^{-1}K_2$$

exists and is a feasible preconditioner for $\tilde{T} = Q_1 + Q_2$.

The application of $\tilde{M}$ then can be written in terms of M with modified data:

$$\begin{cases} x^{k+1/2} = x^k + M^{-1}\big((b^k - Q_2x^k) - Tx^k\big) \\ x^{k+1} = x^k + M^{-1}\big((b^k - Q_2x^{k+1/2}) - Tx^k\big). \end{cases} \tag{3.6}$$

Proof. It is clear that $(M - K_2^*K_2)^{-1}$ exists and is a bounded linear operator. The Woodbury formula then yields

$$(M - K_2^*K_2)^{-1} = M^{-1} + M^{-1}K_2^*(I - K_2M^{-1}K_2^*)^{-1}K_2M^{-1} = M^{-1}\tilde{M}M^{-1} \tag{3.7}$$

showing that $\tilde{M}$ is indeed well-defined. One easily sees

$$\tilde{M} - \tilde{T} = M - T + K_2^*\big((I - K_2M^{-1}K_2^*)^{-1} - I\big)K_2$$

where the right-hand side is positive semi-definite if $I - K_2M^{-1}K_2^* \leq I$ which is obviously always the case. Hence $\tilde{M}$ is a feasible preconditioner for $\tilde{T}$.

For the representation of the application of $\tilde{M}$ in terms of M, we compute with $\tilde{T} = T + Q_2$ and the identity $\tilde{M}^{-1} = M^{-1} - M^{-1}Q_2M^{-1}$ (which follows from (3.7)),

$$I - \tilde{M}^{-1}\tilde{T} = I - M^{-1}(\tilde{T} - Q_2M^{-1}\tilde{T}) = I - M^{-1}\big(T + Q_2(I - M^{-1}\tilde{T})\big).$$

This gives, with the help of the definition of $x^{k+1/2}$,

$$\begin{aligned} x^{k+1} &= x^k + \tilde{M}^{-1}(b^k - \tilde{T}x^k) = x^k + M^{-1}(b^k - Q_2(x^k + M^{-1}(b^k - \tilde{T}x^k)) - Tx^k) \\ &= x^k + M^{-1}((b^k - Q_2x^{k+1/2}) - Tx^k). \end{aligned}$$

□

REMARK 3.3. *This result can also be used for the general iteration* (PDR) *in case Y has a product structure, i.e., $Y = Y_1 \times Y_2$ and $Kx = (K_1x, K_2x)$ for all $x \in X$. In this case, $\tilde{T} = I + \sigma^2 K^*K = I + \sigma^2 K_1^*K_1 + \sigma^2 K_2^*K_2$, hence if one has a feasible preconditioner M for $T = I + \sigma^2 K_1^* K_1$, Proposition 3.2 may give one for $\tilde{T}$ by letting $Q_2 = \sigma^2 K_2^* K_2$.*

3.2. The purely quadratic-linear case. Let us now assume that $\tilde{F} = 0$ in (3.1), i.e., the primal functional is purely quadratic-linear. Then, in view of (3.3), there is no need to introduce the additional variable $\tilde{x}$. The following data for the proximal point algorithm can be used:

$$\mathcal{M} = \begin{pmatrix} N_1 & 0 & 0 \\ 0 & N_2 & -I \\ 0 & -I & I \end{pmatrix}, \qquad \mathcal{A} = \begin{pmatrix} \sigma Q + \sigma \mathbf{1}_f & \sigma K^* & 0 \\ -\sigma K & \sigma R + \sigma \mathbf{1}_g & I \\ 0 & -I & (\sigma \partial \tilde{G})^{-1} \end{pmatrix}. \tag{3.8}$$

This obviously leads to the iteration (3.4) with the auxiliary variable $\bar{x}$ replaced by x, i.e., $\bar{x}^k = x^k$ for each $k \geq 0$. While this is nothing new in terms of algorithms, the prerequisites for convergence can be weakened.

THEOREM 3.4. *Consider the case* (3.1) *with $\tilde{F} = 0$ and assume that the associated saddle-point problem* (1.1) *has a solution. If*

$$\begin{cases} N_1 : X \to X \text{ linear, continuous, self-adjoint with } N_1 \geq 0, \\ N_2 : Y \to Y \text{ linear, continuous, self-adjoint with } N_2 - I \geq 0, \end{cases} \tag{3.9}$$

and $N_1 + \sigma Q + \sigma^2 K^*[N_2 + \sigma R]^{-1} K > 0$, then the iteration (3.4) *converges to a fixed-point $(x^*, y^*, \bar{x}^*, \bar{y}^*)$ for which (x^*, y^*) is a saddle-point of* (1.1).

Proof. The proof is completely analogous to the proof of Theorem 2.3, keeping in mind that $\mathcal{M}$ still induces a semi-norm and that the update operator for (x, y) is affine linear and continuous. ∎

REMARK 3.5. *The assumption $N_1 - I \geq 0$ replaced by $N_1 \geq 0$, $N_1 + \sigma Q > 0$ and $\bar{x}^k = x^k$ for all k allows to consider more general preconditioners. In the situation of Remark 3.1 we see that the update for x and y can further be simplified to yield the **PDRQ** method, see Table 3.1. The data for the linear parts now read as $T = \sigma Q + \frac{\sigma^2}{\mu + \sigma\lambda} K^*K$ and $M = N_1 + \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda} K^*K$. It is immediate that $N_1 \geq 0$ and $N_1 + \sigma Q + \sigma^2 K^*[N_2 + \sigma R]^{-1}K > 0$ if and only if $M - T \geq 0$ and $M > 0$.*

In other words: The iteration still converges if M is a feasible preconditioner for $T = \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda} K^*K$. Compared to Remark 3.1, where the preconditioner must be feasible for the operator $I + \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda} K^*K$, this allows more flexibility. Note that in this situation, Proposition 3.2 can in particular be employed to construct such a preconditioner.

A similar discussion can be carried out if $\tilde{G} = 0$ instead of $\tilde{F} = 0$, also leading to weaker conditions on the preconditioners. We leave the details to the interested reader.

4. Application to total-variation denoising. In the following, the proposed preconditioned Douglas-Rachford iteration framework will be applied to a discrete total-variation denoising problem, i.e.,

$$\min_{u \in X} \frac{\|u - f_0\|_2^2}{2} + \alpha \|\nabla u\|_1$$

for a given noisy image f_0. This problem is a simple instance of a variational problem with the non-smooth total-variation semi-norm $\|\nabla u\|_1$ as penalty. Such penalty functionals were a preferred choice for imaging problems in the time they were introduced [25].

PDRQ	Solve $\min_{x\in X} \max_{y \in Y} \ \langle Kx, y\rangle + \langle \frac{1}{2}Qx + f, x\rangle - \langle \frac{\lambda}{2} y + g, y\rangle - G(y)$
Initialization:	$(x^0, y^0, \bar{y}^0) \in X \times Y^2$ initial guess, $\sigma > 0$ step-size, $\mu \geq 1$ scaling
	$T = \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda} K^*K$, $M = N_1 + \sigma Q + \frac{\sigma^2}{\mu+\sigma\lambda}K^*K$
Iteration:	$$\begin{cases} b^k = -\frac{\sigma}{\mu+\sigma\lambda}K^*((\mu-1)y^k + \bar{y}^k - \sigma g) - \sigma f \\ x^{k+1} = x^k + M^{-1}(b^k - Tx^k) \\ y^{k+1} = \frac{1}{\mu+\sigma\lambda}\big((\mu-1)y^k + \bar{y}^k + \sigma(Kx^{k+1} - g)\big) \\ \bar{y}^{k+1} = \bar{y}^k + (I + \sigma\partial G)^{-1}[2y^{k+1} - \bar{y}^k] - y^{k+1}, \end{cases} \quad \text{(PDRQ)}$$

Table 3.1: The abstract preconditioned Douglas-Rachford iteration for purely quadratic-linear primal functionals.

4.1. Discretization. We start with the discretization. Following essentially the presentation in [5, 6], consider the image domain $\Omega \subset \mathbf{Z}^2$ as the discretized rectangular grid

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

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

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

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

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

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

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

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

4.2. Symmetric Red-Black Gauss-Seidel preconditioners for TV-denoising problems. Let us apply the discrete framework for the total-variation regularized L^2-type denoising problems (see [25] for the L^2 case which is usually called the *ROF model*)

$$\min_{u\in X} \ F(u) + \alpha\|\nabla u\|_1, \qquad F(u) = \tfrac{1}{2}\|u - f_0\|_2^2 \tag{4.1}$$

for $f_0 : \Omega \to \mathbf{R}$ a noisy image and $\alpha > 0$ a regularization parameter. As we are in finite dimensions, both F and $\alpha\|\cdot\|_1$ are continuous, so one can employ Fenchel-Rockafellar duality to obtain the equivalent saddle-point problem [13, 18]

$$\min_{u\in X} \max_{v\in Y} \ \langle Ku, v\rangle + F(u) - G(v) \tag{4.2}$$

with $K = \nabla$ and $G = \mathcal{I}_{\{\|v\|_\infty \le \alpha\}}$ while $\mathcal{I}_C$ denotes the indicator function of the set C, i.e.,

$$\mathcal{I}_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{else.} \end{cases}$$

Now, preconditioned Douglas-Rachford methods can be derived.

The PDR method. For the general PDR iteration, the data corresponding to Table 2.1 are

$$K = \nabla, \quad F = \frac{\|\cdot - f_0\|_2^2}{2}, \quad G = \mathcal{I}_{\{\|v\|_\infty \le \alpha\}}.$$

The operator M is thus supposed to be a feasible preconditioner for $I - \sigma^2 \operatorname{div} \nabla = I - \sigma^2 \Delta$ (where $\Delta = \operatorname{div} \nabla$) and will be discussed later. The iteration then needs the resolvent

$$(I + \sigma\partial F)^{-1}(u) = \operatorname*{arg\,min}_{u'\in X} \ \frac{1}{2}\|u' - u\|_2^2 + \frac{\sigma}{2}\|u' - f_0\|_2^2 = \frac{u + \sigma f_0}{1+\sigma}$$

as well as $(I + \sigma\partial G)^{-1} = \mathcal{P}_\alpha$ which reads as [5]

$$\begin{aligned} (I + \sigma\partial G)^{-1}(v) &= \operatorname*{arg\,min}_{v'\in Y} \ \frac{1}{2}\|v' - v\|_2^2 + \mathcal{I}_{\{\|v\|_\infty \le \alpha\}}(v') \\ &= \mathcal{P}_\alpha(v) = \frac{v}{\max(1, |v|/\alpha)}. \end{aligned} \tag{4.3}$$

The PDRQ method. The functional F in (4.1) is purely quadratic-linear (up to a constant), so one can use the iteration (PDRQ) as outlined in Table 3.1. With the notations in Section 3, the data of the method reads as

$$K = \nabla, \quad Q = I, \quad f = -f_0, \quad \tilde{F} = 0, \quad R = 0, \quad g = 0, \quad \tilde{G} = \mathcal{I}_{\{\|v\|_\infty \le \alpha\}}, \quad N_2 = I.$$

Here, M has to be a feasible preconditioner for $\sigma I - \sigma^2 \operatorname{div} \nabla = \sigma I - \sigma^2 \Delta$ while the resolvent $(I + \sigma\partial\tilde{G})^{-1} = (I + \sigma\partial G)^{-1}$, i.e., is given by $\mathcal{P}_\alpha$ according to (4.3).

The preconditioners. Observe that in both cases, M is required to be a feasible preconditioner for operators of type $T = \alpha I - \beta\Delta$ for $\alpha, \beta > 0$ where $\Delta = \operatorname{div} \nabla$ can be interpreted as a discrete Laplace operator with homogeneous Neumann boundary conditions [27]. In other words: solving $Tu = b$ correspond to a discrete version of the boundary value problem

$$\begin{cases} \alpha u - \beta\Delta u = b, \\ \frac{\partial u}{\partial \nu}|_{\partial\Omega} = 0. \end{cases} \tag{4.4}$$

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 well-known and popular *Red-Black ordering* [1, 26, 28] then yields a symmetric block representation according to

$$T = \begin{pmatrix} D_1 & A \\ A^* & D_2 \end{pmatrix},$$

with $D_1, D_2 > 0$ diagonal matrices and A with at most four entries per row and column. Hence, the Gauss-Seidel method (or SOR method) can efficiently be employed in this situation, i.e.,

$$M_0 = \begin{pmatrix} D_1 & A \\ 0 & D_2 \end{pmatrix}.$$

In order to obtain a feasible preconditioner for T, we have to symmetrize it (see Proposition 2.12 and Example 2.13), leading to the *symmetric Red-Black Gauss-Seidel preconditioner*, denoted by M. Of course, several steps of this preconditioner can be performed (see Proposition 2.14). We denote the n-fold application of the symmetric Red-Black to the initial guess u and right-hand side b by

$$\mathrm{SRBGS}^n_{\alpha,\beta}(u, b) = (I + M^{-1}(\mathbf{1}_b - T))^n u \tag{4.5}$$

making it again explicit that M and T depend on α and β.

Remark 4.1. *Of course, an SSOR method based on Red-Black ordering can be used instead of the symmetric Red-Black Gauss-Seidel update. This gives a slightly modified preconditioner which is feasible for all $\omega \in]0, 2[$, see Example 2.13.*

The algorithms. The building blocks (4.3) and (4.5) eventually give all the needed ingredients to perform the iterations. For the solution of (4.1), we may employ both preconditioned Douglas-Rachford splitting method according to (PDR) and (PDRQ). Details for (PDRQ) can be found in Table 4.1.

Remark 4.2. *Note that the methods PDRQ, ADMM [14] and split Bergman [30] for the solution of* (4.1) *are all designed on the basis of the Douglas-Rachford splitting method. However, in the concrete realization, PDRQ differs in the following points. First, the update of u^{k+1} is performed in a flexible preconditioned manner. Second, the updates of v^{k+1} and $\bar{v}^{k+1}$ in PDRQ involve the projection $\mathcal{P}_\alpha$ at $2v^{k+1} - \bar{v}^k$ and a linear update which is different from the shrinkage operator and multiplier-type update that are involved in ADMM and split Bergman methods.*

5. Numerical Results.

We now illustrate the efficiency of the methods proposed in Section 4 for the solution of the L^2-TV denoising problem (4.1). Here, the primal-dual gap function can be used to measure optimality. For the considered problem, this function is given by [8],

$$\mathcal{G}(u, v) := F(u) + G(\nabla v) + F^*(-\nabla^* v) + G^*(v) \tag{5.1}$$

$$= \frac{\|u - f\|^2}{2} + \alpha\|\nabla u\|_1 + \langle f, \operatorname{div} v\rangle + \frac{\|\operatorname{div} v\|^2}{2} + I_{\{\|v\|_\infty \le \alpha\}}(v). \tag{5.2}$$

PDRQ Objective: L^2-TV denoising		$\min_{u\in X}\ \frac{1}{2}\|u-f\|_2^2+\alpha\|\nabla u\|_1$
Initialization:	$(u^0,v^0,\bar v^0)\in X\times Y\times Y$ initial guess, $\sigma>0$ step-size, $n\geq 1$ inner iterations for symmetric Gauss-Seidel	
Iteration:	$u^{k+1}=\mathrm{SRBGS}^n_{\sigma,\sigma^2}\big(u^k,\sigma(f+\operatorname{div}\bar v^k)\big)$	according to (4.5)
	$v^{k+1}=\bar v^k+\sigma\nabla u^{k+1}$	
	$v^{k+1}_{\text{test}}=\mathcal{P}_\alpha(2v^{k+1}-\bar v^k)$	according to (4.3)
	$\bar v^{k+1}=\bar v^k+v^{k+1}_{\text{test}}-v^{k+1}$	

Table 4.1: The preconditioned Douglas-Rachford iteration for L^2-TV denoising.

In order to be independent from image size, the normalized primal-dual gap $\mathcal{G}(u,v)/(N_xN_y)$ is used in our experiments where N_xN_y is the total number of pixels. As in Remark 2.4, either $\{(u^{k+1}_{\text{test}},v^{k+1}_{\text{test}})\}$ in algorithm PDR or $\{(u^{k+1},v^{k+1}_{\text{test}})\}$ in algorithm PDRQ converges to a solution of the saddle point problem (4.2). And it could be seen from the algorithm PDR and PDRQ that v^{k+1}_{test} in the two algorithms satisfy the constraint $I_{\{\|v\|_\infty\leq\alpha\}}(v^{k+1}_{\text{test}})=0$. Hence, the primal-dual gap function evaluated at

$$\mathcal{G}(u^{k+1},v^{k+1}_{\text{test}})/(N_xN_y) \tag{5.3}$$

yields finite values and converges to zero by convergence of the iterates and continuity. As these values estimate the functional distance to the minimizer, they will be used to assess convergence speed in the numerical experiments. Note that this holds also for the algorithms we use for comparison. For the first order primal-dual algorithm ALG1 in [8] the constraint $\|v\|_\infty\leq\alpha$ is naturally satisfied. And for the FISTA algorithm [2], which is based on the following dual problem of (4.1), i.e.,

$$\min_{v\in Y}\frac{\|\operatorname{div}v\|^2}{2}+\langle f,\operatorname{div}v\rangle+I_{\{\|v\|_\infty\leq\alpha\}}(v), \tag{5.4}$$

the constraint $\|v\|_\infty\leq\alpha$ is also naturally satisfied during the iteration. Hence, convergence speed can also be measured by evaluating the primal-dual gap function. For other type of fast algorithms like ADMM [14, 16], the primal-dual gap is not easily computable, hence we do not use it in this comparison.

The following five algorithms are tested for the ROF model. First, the proposed preconditioned Douglas-Rachford iterations based methods PDR and PDRQ that are stated as before. Both of them employ symmetric Red-Black Gauss-Seidel preconditioners. We denote them by PDR(σ,n) and PDRQ(σ,n) where σ is the step size and n the number of inner iterations for the preconditioner. Additionally, three algorithms are chosen for comparison with parameter settings as follows:

- ALG1: Fixed step-size primal-dual algorithm as described in Algorithm 1 in [8]. Here, $L=\sqrt{8}$, $\sigma\tau L^2=1$ and ALG1(τ) means ALG1 with dual step size τ.
- FISTA: Fast iterative shrinkage thresholding algorithm on the dual of ROF model (5.4) [2]. The Lipschitz constant L as in [2] here is chosen as 8.
- PDHG: Primal-dual hybrid gradient algorithm for the TV denoising problem with adaptive steps proposed in [31].

We note that all algorithms in this paper were implemented in Matlab (MATLAB and Image Processing Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States)

	$\alpha = 0.2$		$\alpha = 0.5$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
FISTA	68 (2.61s)	435 (14.64s)	222 (7.78s)	1271 (42.81s)
PDHG	156 (5.40s)	420 (14.17s)	279 (7.52s)	821 (22.90s)
ALG1	(0.05) 54 (1.65s)	(0.05) 366 (8.63s)	(0.05) 127 (3.20s)	(0.01) 672 (18.66s)
PDRQ	(4,2) 16 (1.08s)	(14,3) 79 (4.99s)	(12,3) 33 (2.49s)	(45,3) 189 (11.97s)
PDR	(4,2) 23 (1.93s)	(14,3) 141 (10.57s)	(6,3) 45 (3.46s)	(18,3) 278 (19.94s)

Table 5.1: Numerical results for the L^2-TV image denoising (ROF) problem (4.1) with noise level 0.1 and regularization parameters $\alpha = 0.2$, $\alpha = 0.5$. For FISTA and PDHG, we use the pair $\{k(t)\}$ to represent the iteration number k and CPU time cost t. For ALG1, we use the combination $\{(\tau)\ k(t)\}$ to represent the dual step size τ, the iteration number k and the iteration time t. For algorithms PDRQ and PDR, we use the two pairs $\{(\sigma, n)\ k(t)\}$ to represent the iteration number k, CPU time t with constant step size σ and inner iteration number n. The iteration is performed until the normalized primal-dual gap (5.3) is below ε.

(a) Original image (b) Noise level: $\sigma = 0.1$ (c) $\alpha = 0.2$ (d) $\alpha = 0.5$

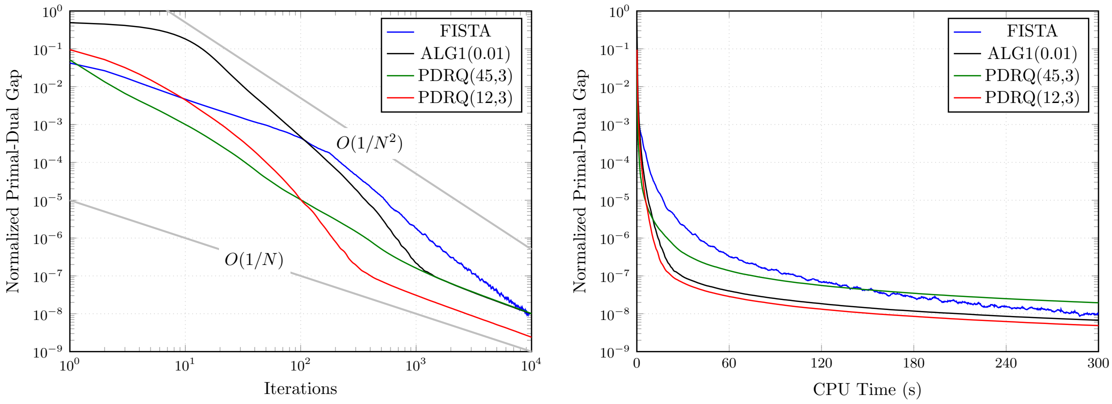
Fig. 5.1: Results for variational L^2-TV denoising. All denoised images are obtained with algorithm PDRQ with appropriate step sizes and inner iterations such that the normalized primal-dual gap is less than 10^{-6}. (a) is the original image: Fronalpstock (513×513 pixels, gray) which is from the Wikipedia article on the mountain 'Fronalpstock' [29]. (b) shows a noise-perturbed version of (a) (additive Gaussian noise, standard deviation 0.1), (c) and (d) are the denoised images with $\alpha = 0.2$ and $\alpha = 0.5$, respectively.

and executed on a Dell Precision T1650 workstation running a 64 Bit Linux system with 8 cores each at 3.40GHz.

Table 5.1 summarizes the numerical results for the ROF image denoising problem with the above five algorithms. Computations were performed for the image [29] (size 513×513), additive Gaussian noise (level 0.1) and different regularization parameters α in (4.1). Figure 5.1 includes the test images, noisy images and denoised images with different regularization parameters.

It could be seen from Table 5.1 that our algorithms PDRQ and PDR are competitive, especially PDRQ. For the ROF model with high regularization parameter α which is important when the image is corrupted by high level of noise, PDRQ is very fast for appropriate step size σ and inner iteration number n. It could also be seen that both algorithms PDRQ and PDR benefit from the preconditioner.

Figure 5.2 represents the comparison according to the iteration number and CPU time cost.



(a) Numerical convergence rate compared with iteration number.

(b) Numerical convergence rate compared with iteration time. Each algorithm runs for 300 seconds.

Fig. 5.2: Numerical convergence rate compared with iteration number and iteration time for the ROF model with regularization parameter $\alpha = 0.5$. The left plots utilizes a double-logarithmic scale while the right plot is semi-logarithmic.

It could be seen from Figure 5.2(a) that our PDRQ is superior to other algorithms in terms of the iteration number while Figure 5.2(b) tells that PDRQ is also fast compared to FISTA and ALG1 with respect to run time. The asymptotic convergence rate is, however, not as fast as for FISTA which is adaptive with respect to the step-size. Nevertheless, PDRQ is competitive in the relevant range of accuracy (roughly up to 10^{-7}).

6. Conclusions. We propose several preconditioned Douglas-Rachford splitting methods for the solution of general saddle-point problems which efficiently deal with the implicit linear equations arising in this context. We moreover establish a practical convergence theory within this framework. These could make the Douglas-Rachford method and its variants more safe and more flexible to use in applications. We also present an application for image denoising. The numerical results tell us that the preconditioned Douglas-Rachford method has the potential to bring out appealing benefits and fast algorithms.

Nevertheless, there are some open questions that need to be figured out. Other preconditioners such as multigrid and ADI (alternating direction implicit method) could also be feasible for the linear system as it is the case for the symmetric Red-Black Gauss-Seidel iteration. Another direction is to consider nonstationary iterative methods such as the conjugate gradient method. Finding an appropriate adaptive step-size strategy is furthermore an important issue. We believe that this framework will also be beneficial for other applications, for instance, for compressed sensing problems, provided that appropriate preconditioners could be found. Also, the total generalized variation [6] is worth to try where more complicated linear systems are involved.

Acknowledgements. The work of Kristian Bredies and Hongpeng Sun is supported by the Austrian Science Fund (FWF) under grant SFB32 (SFB "Mathematical Optimization and Applications in the Biomedical Sciences").

References.

[1] G. Alefeld, *On the convergence of the symmetric SOR method for matrices with red-black ordering*, Numerische Mathematik, 39 (1982), pp. 113–117.

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[3] J. Bramble and J. Pasciak, *New estimates for multilevel algorithms including the V-cycle*, Math. Comp., 60 (1993), pp. 447–471.

[4] J. Bramble, J. Pasciak, and A. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle point problems*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 1072–1092.

[5] Kristian Bredies, *Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty*, in Efficient Algorithms for Global Optimization Methods in Computer Vision, Andrs Bruhn, Thomas Pock, and Xue-Cheng Tai, eds., Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2014, pp. 44–77.

[6] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 492–526.

[7] L. Briceño Arias and P. L. Combettes, *A monotone+skew splitting model for composite monotone inclusions in duality*, SIAM Journal on Optimization, 21 (2011), pp. 1230–1250.

[8] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.

[9] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, 53 (2004), pp. 475–504.

[10] P. L. Combettes and J. Pesquet, *A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery*, IEEE Journal of Selected Topics in Signal Processing, 1 (2007), pp. 564–574.

[11] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Transactions of the American Mathematical Society, 82 (1956), pp. 421–439.

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

[13] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, no. 1 in Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 1999.

[14] E. Esser, *Applications of Lagrangian-based alternating direction methods and connections to split Bregman*, tech. report, CAM report 09-2009, UCLA, 2009.

[15] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, Michel Fortin and Roland Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, 1983, pp. 299 – 331.

[16] T. Goldstein and S. Osher, *The split Bregman method for L^1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.

[17] Q. Hu and J. Zou, *Nonlinear inexact Uzawa algorithms for linear and nonlinear saddle-point problems*, SIAM Journal on Optimization, 16 (2006), pp. 798–825.

[18] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Advances in Design and Control, Society for Industrial and Applied Mathematics, 2008.

[19] C. S. Kubrusly, *Spectral Theory of Operators on Hilbert Spaces*, Springer, 2012.

[20] P. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, 16 (1979), pp. 964–979.

[21] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Mathematical Journal, 29 (1962), pp. 341–346.

[22] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bulletin of the American Mathematical Society, 73 (1967), pp. 591–597.

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

[24] M. Reed and B. Simon, *Methods of Modern Mathematical Physics: Functional analysis*, no. 1 in Methods of Modern Mathematical Physics, Academic Press, 1980.

[25] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.

[26] Y. Saad, *Iterative Methods for Sparse Linear Systems: Second Edition*, Society for Industrial and Applied Mathematics, 2003.

[27] J. W. Thomas, *Numerical Partial Differential Equations: Conservation Laws and Elliptic Equations*, vol. 33 of Texts in Applied Mathematics, Springer New York, 1999.

[28] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, 2001.

[29] Wikipedia, *Fronalpstock*. `http://en.wikipedia.org/wiki/File:Fronalpstock_big.jpg`, 2014. [Online; accessed 07-February-2014].

[30] X. Q. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, Journal of Scientific Computing, 46 (2011), pp. 20–46.

[31] M. Q. Zhu and T. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, tech. report, UCLA CAM Report 08-2008, UCLA, 2008.