SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Preconditioned Douglas-Rachford algorithms for TV and TGV regularized variational imaging problems

K. Bredies H. Sun

SFB-Report No. 2014-009 November 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**

Das Land Steiermark

Stadt GRAZ Wissenschaft

SFB MOBIS Report 2014-009

Preconditioned Douglas-Rachford algorithms for TV and TGV regularized variational imaging problems

Kristian Bredies · Hongpeng Sun

November 28, 2014

Abstract The recently introduced preconditioned Douglas-Rachford iteration (PDR) for convex-concave saddle-point problems is studied with respect to convergence rates and applied to variational imaging problems with total variation (TV) and total generalized variation (TGV) penalty. A rate of $\mathcal{O}(1/k)$ for restricted primal-dual gaps evaluated for ergodic sequences generated by the PDR iteration is established. Based on PDR, new fast iterative algorithms for TV-denoising, TV-deblurring and TGV-denoising of second order with L^2 and L^1 discrepancy are proposed. While for denoising, symmetric (block) Red-Black Gauss-Seidel preconditioners are effective, fast Fourier-transform (FFT) based preconditioners are employed for the deblurring problems. Finally, for the L^2-TGV denoising problem, an effective modified primal-dual gap is developed which may serve as a stopping criterion. All algorithms are tested and compared in numerical experiments. In particular, for problems where strong convexity does not hold, it turns out that the proposed preconditioning techniques are beneficial and lead to competitive results.

Keywords Preconditioned Douglas-Rachford iteration · Primal-dual algorithms · Variational image denoising and deblurring · Total Generalized Variation · Block preconditioner

1 Introduction

In this paper, we address the problem of efficiently computing solutions of TV and TGV-regularized variational imaging problems for instances of the type

$$\min_{u\in X}\ F(u) + \alpha\,\mathrm{TV}(u), \qquad\qquad \min_{u\in X}\ F(u) + \mathrm{TGV}^2_\alpha(u) \tag{1}$$

where X is a discrete image space, $F : X \to \mathbf{R}_\infty$ a convex fidelity functional for recovering the image u from some given data and TV and TGV^2_α suitable discretizations of the total variation and total generalized variation semi-norm, respectively. We propose iterative algorithms which base on the recently introduced Preconditioned Douglas-Rachford (PDR) method for the solution of certain convex-concave saddle-point problems [5]. For this iteration, we derive a $\mathcal{O}(1/k)$ ergodic convergence rate in terms of restricted primal-dual gap functionals. Dedicated linear preconditioners are developed which account for the differential structure of TV and TGV as well as the possible underlying structure of the image reconstruction problem.

Kristian Bredies, Hongpeng Sun
Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstr. 36, A-8010 Graz, Austria
E-mail: kristian.bredies@uni-graz.at, hongpeng.sun@uni-graz.at

Even in finite dimensions, the numerical optimization of the functionals in (1) is a challenge due to the non-smooth nature of the problems. For convex optimization, first-order iterative algorithms based on the resolvent of the subgradient of a proper, convex and lower semi-continuous functional, or, equivalently, on the associated proximal mapping are an efficient and widely used choice. They are most commonly employed in conjunction with splitting methods for the underlying monotone operator inclusions. Today's state-of-the art first-order iterative solvers for (1) utilize a reformulation to such a monotone inclusion problem associated with a convex-concave saddle-point problem as well as accelerated step-size strategies, the latter where applicable. Chambolle and Pock's first order primal-dual algorithm [9] is in particular efficient for TV-regularized problems and can also be used for TGV penalty terms [3]. As it involves forward evaluations with respect to a linear operator, it is conditionally stable, i.e., step-size constraints have to be respected in order to ensure convergence. In contrast, the Douglas-Rachford splitting method is a powerful and fully implicit algorithm for the solution of monotone operator inclusions, i.e., unconditionally stable. However, it suffers from the need of solving, in each iteration step, implicit equations which can be very expensive even in an inexact manner. In [5], the authors introduced a preconditioned Douglas-Rachford splitting method which could help to deal with the linear subproblems appearing in certain convex-concave saddle-point problems. In particular, under mild assumptions one as well as finitely many inner iterations of an approximate linear solver is sufficient to establish convergence. The effectiveness of this approach for L^2-TV denoising problems (also known as the ROF model [23]) has already been shown in [5].

In this paper, we study more TV- and TGV-regularized imaging problems, and introduce new preconditioned Douglas-Rachford splitting methods for TV-regularized problems in imaging, such as L^1-TV denoising, L^2-TV and L^1-TV deblurring, which have not been discussed in [5]. The denoising problems are preconditioned by a symmetric Red-Black Gauss-Seidel method while for deblurring, fast Fourier-transform (FFT) based preconditioners are introduced. Moreover, we derive dedicated methods for TGV-regularized variational imaging problems where the underlying linear system is more complicated and difficult to solve in a traditional manner. This leads to symmetric block Red-Black Gauss-Seidel preconditioners and provides new ways to deal numerically with TGV-regularized imaging problems, in particular, L^2-TGV and L^1-TGV denoising. Additionally, a new primal-dual gap for L^2-TGV denoising is derived which constitutes a provable upper bound for the functional distance to the minimum as well as an estimate for the L^2-norm of the distance to the minimizer.

Let us mention some recent results from the vastly available literature on methods for the solution of monotone inclusion problems related to the considered non-smooth TV- and TGV-regularized minimization problems. Common approaches typically base on monotone splittings and lead to methods involving forward and backward (resolvent) evaluation. Indeed, classical methods like proximal point [22], forward-backward splitting and the Douglas-Rachford iteration [19], the Uzawa algorithm [26] as well as alternating direction method of multipliers (ADMM) [15] fall into this class and interconnections between them exist [10, 14]. In this context, recent developments focus on algorithms exploiting a specific structure of the monotone inclusions applicable to specific classes of minimization problems. While [5, 9] as well as this work build on saddle-point formulations associated with primal-dual problems according to Fenchel-Rockafellar duality, the papers [1, 2, 20, 27] aim at minimizing sums of convex functionals involving the composition with a linear operator and infimal convolution, allowing essentially the same flexibility. Notably, the Douglas-Rachford iteration for the same splitting as for the PDR algorithm (but without preconditioning) has been mentioned as a side remark in [8]. Moreover, Douglas-Rachford iterations for a certain class of primal-dual problems have, for instance, also been introduced in [1]. Furthermore, preconditioners have been discussed, in the context of nonsmooth optimization, e.g., in [21, 29]. However, to the best knowledge of the authors, the combination of preconditioning techniques and Douglas-Rachford iteration based on saddle-point formulations as in [5] and their application to the imaging problems presented in this paper can not be found in the existing literature.

The organization of the paper is as follows. First, we briefly recall in Section 2 the abstract preconditioned Douglas-Rachford splitting methods introduced in [5]. The ergodic convergence of the PDR method including the classical Douglas-Rachford splitting method with respect to restricted primal-dual gap functions is derived in Section 3. Section 4 then deals with the application of these algorithms to certain TV-regularized variational imaging problems, mainly L^2- and L^1-penalized denoising and deblurring. With the exception of L^2-TV-denoising which has already been treated in [5], this results in new iterative solution techniques. Section 5 then deals with preconditioning TGV-regularized denoising problems. After shortly reviewing TGV of second order and presenting the basic outline of the algorithms, we investigate the linear system which needs to be preconditioned during the iteration. The actual preconditioner then turns out to be a symmetric block Gauss-Seidel method on a modified equation that allows for Red-Black ordering. Discussing the associated finite-difference stencils, we give details for the implementation of the preconditioner as well as the resulting concrete methods. Stopping criteria based on the primal-dual gap are discussing in Section 6. While for TV-regularized problems, the algorithms allow for an efficient evaluation of the primal-dual gap, the situation is more difficult for TGV-regularization. We introduce a modified primal-dual gap which establishes a provable bound for the functional distance to the minimizer. In Section 7, thorough numerical tests are performed which demonstrate the efficiency of all preconditioned algorithms discussed in this paper. Finally, some discussions and conclusions are presented in the last section.

2 Abstract Preconditioned Douglas-Rachford Iterations

Let us shortly review the techniques and results from [5] for the solution of saddle-point problems of the type

$$\min_{x \in \operatorname{dom} \mathcal{F}} \max_{y \in \operatorname{dom} \mathcal{G}} \langle \mathcal{K}x, y \rangle + \mathcal{F}(x) - \mathcal{G}(y). \tag{2}$$

We fix X and Y as real Hilbert spaces and $\mathcal{K} : X \to Y$ as a continuous linear mapping. The functionals $\mathcal{F} : X \to \mathbf{R}_\infty$ (where $\mathbf{R}_\infty = (-\infty, +\infty]$ are the extended real numbers) and $\mathcal{G} : Y \to \mathbf{R}_\infty$ are proper, convex and lower semi-continuous. Recall that the Fenchel conjugate functionals $\mathcal{F}^* : X \to \mathbf{R}_\infty$, $\mathcal{G}^* : Y \to \mathbf{R}_\infty$, given by

$$\mathcal{F}^*(x) = \sup_{x' \in X} \langle x, x' \rangle - \mathcal{F}(x'), \qquad \mathcal{G}^*(y) = \sup_{y' \in Y} \langle y, y' \rangle - \mathcal{G}(y')$$

are again proper, convex and lower semi-continuous. With these notions, (2) can be regarded as the Fenchel-Rockafellar primal-dual formulation associated with the primal and dual problem

$$\min_{x \in X} \mathcal{F}(x) + \mathcal{G}^*(\mathcal{K}x), \qquad \max_{y \in Y} -\mathcal{F}^*(-\mathcal{K}^*y) - \mathcal{G}(y), \tag{3}$$

respectively [11, 17]. Under suitable conditions, namely those which guarantee that the minimum and maximum in (3) exist and coincide, primal-dual solution pairs $(x^*, y^*) \in X \times Y$ of (3) are exactly the solutions of the saddle-point problem (2).

The preconditioned Douglas-Rachford iteration comes in two variants. The first, called PDR, is suitable for general $\mathcal{F}$ and $\mathcal{G}$ for which the resolvents can be computed while the second, called PDRQ, assumed that either the primal or dual functional is purely quadratic-linear. They both employ linear preconditioners M for the solution of the equation $Tx = b$ with T given by

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

Table 1: The abstract preconditioned Douglas-Rachford iteration for saddle-point problems of the type (2).

PDRQ	Solve $\min_{x\in X} \max_{y\in\operatorname{dom}\mathcal{G}} \langle \mathcal{K}x, y\rangle + \langle \frac{1}{2}Qx - f_0, x\rangle - \mathcal{G}(y)$
Initialization:	$(x^0, \bar{y}^0) \in X \times Y$ initial guess, $\sigma > 0$ step-size, $T = \sigma Q + \sigma^2\mathcal{K}^*\mathcal{K}$, $M - T \geq 0$
Iteration:	$\begin{cases} b^k = -\sigma\mathcal{K}^*\bar{y}^k + \sigma f_0 \\ x^{k+1} = x^k + M^{-1}(b^k - Tx^k) \\ y^{k+1} = \bar{y}^k + \sigma\mathcal{K}x^{k+1} \\ \bar{y}^{k+1} = \bar{y}^k + (I + \sigma\partial\mathcal{G})^{-1}[2y^{k+1} - \bar{y}^k] - y^{k+1}, \end{cases}$ (PDRQ)

Table 2: The abstract preconditioned Douglas-Rachford iteration for purely quadratic-linear primal functionals.

the problem. The preconditioner is applied by performing one step with respect to the splitting $T = M - (M - T)$, i.e.,

$$x^{\text{new}} = x^{\text{old}} + M^{-1}(b - Tx^{\text{old}}).$$

As we will see immediately, T essentially depends on $\mathcal{K}$ and $\mathcal{K}^*$ for which the application to vectors can be assumed to be easily computable. Hence, the preconditioner M should be chosen such that the application of M^{-1} is also easily computable and, in addition, a good approximation of T^{-1}. We will see in the course of the article that such a choice is indeed possible. To this end, let us note the central condition on M which is needed to ensure convergence.

Definition 1 Let $T, M : X \to X$ be linear, continuous, symmetric and positive definite. Then, M is called a *feasible preconditioner* for T if $M - T$ is positive semi-definite.

In the PDR method, we set $T = T_{\text{PDR}} = I + \sigma^2\mathcal{K}^*\mathcal{K}$ for some step-size $\sigma > 0$ and M a feasible preconditioner for T, resulting in the iteration outlined in Table 1. For the PDRQ method, either $\mathcal{F}$ or $\mathcal{G}$ is assumed to be purely quadratic-linear, here we restrict ourselves to the primal case, i.e.,

$$\mathcal{F}(x) = \frac{1}{2}\langle Qx, x\rangle - \langle f_0, x\rangle$$

for $Q : X \to X$ linear, continuous and symmetric positive semi-definite and $f_0 \in X$. In this case, M should be a feasible preconditioner for $T = T_{\text{PDRQ}} = \sigma Q + \sigma^2\mathcal{K}^*\mathcal{K}$ for some $\sigma > 0$ (which means in particular that we assume T to be positive definite). The iteration is then performed according to Table 2 and needs less variables compared to PDR.

Convergence of both methods is summarized in the following theorem whose proof can be found in [5, Theorem 2.3].

Preconditioner	T	λI	$(\lambda+1)D$	M_{SGS}	M_{SSOR}
Conditions	—	$\lambda \geq \|T\|$	$\lambda \geq \lambda_{\max}(T-D)$	—	$\omega \in (0,2)$
Iteration type	Douglas-Rachford	Richardson	Damped Jacobi	Symmetric Gauss-Seidel	Symmetric SOR

$D = \operatorname{diag}(T)$, $T = D - E - E^*$, E lower triangular, $M_{SGS} = (D-E)(T+D)^{-1}(D-E^*)$, $M_{SSOR} = (\frac{1}{\omega}D - E)(T + \frac{2-\omega}{\omega}D)^{-1}(\frac{1}{\omega} - E^*)$

Table 3: Summary of common preconditioners and possible conditions for feasibility.

Theorem 1 *Let M be a feasible preconditioner for T_{PDR} and T_{PDRQ}, respectively. If there exists a saddle point for* (2)*, then* (PDR) *and* (PDRQ) *converges weakly to a saddle point (x^*, y^*) of* (2)*, respectively.*

Regarding the choice of the preconditioner M from given T, the feasibility condition allows for great flexibility. The following proposition summarizes ways to construct feasible preconditioners, see [5, Propositions 2.12, 2.14 & 3.2].

Proposition 1 *Let $T : X \to X$ linear, continuous, symmetric and positive definite be given. Then, the following update schemes correspond to feasible preconditioners.*

(i) For $M_0 : X \to X$ linear, continuous such that $M_0 - \frac{1}{2}T$ is positive definite,

$$\begin{cases} x^{k+1/2} = x^k + M_0^{-1}(b^k - Tx^k), \\ x^{k+1} = x^{k+1/2} + M_0^{-*}(b^k - Tx^{k+1/2}). \end{cases} \tag{4}$$

(ii) For $M : X \to X$ feasible and $n \geq 1$,

$$\begin{cases} x^{k+(i+1)/n} = x^{k+i/n} + M^{-1}(b^k - Tx^{k+i/n}), \\ \qquad i = 0, \ldots, n-1. \end{cases} \tag{5}$$

(iii) For $T = T_1 + T_2$, with $T_1, T_2 : X \to X$ linear, continuous, symmetric, T_1 positive definite, T_2 positive semi-definite and $M : X \to X$ feasible for T_1, $M - T_2$ boundedly invertible,

$$\begin{cases} x^{k+1/2} = x^k + M^{-1}\big((b^k - T_2 x^k) - Tx^k\big), \\ x^{k+1} = x^k + M^{-1}\big((b^k - T_2 x^{k+1/2}) - Tx^k\big). \end{cases} \tag{6}$$

The update scheme (4) is useful if one has a non-symmetric "feasible" preconditioner M_0 for $\frac{1}{2}T$, then the concatenation with the adjoint preconditioner will be symmetric and feasible. Likewise, (5) corresponds to the n-fold application of a feasible preconditioner which is again feasible. Finally, (6) is useful if T can be split into $T_1 + T_2$ for which T_1 can be easily preconditioned by M. Then, one gets a feasible preconditioner for T only using forward evaluation of T_2.

Several standard preconditioners can be seen to be feasible, including symmetric Gauss-Seidel and SSOR. Table 3 provides an overview; for details, we refer to [5].

3 An Ergodic Convergence Rate

In this section, we derive an ergodic convergence rate for restricted primal-dual gaps associated with the Lagrangian of the problem (2) with respect to sequences generated by the general PDR method discussed in [5]. This includes, in particular, the cases in this paper (by letting $N_2 = I$)

as well as the original Douglas-Rachford method associated with the monotone+skew symmetric splitting for the solution of (2), see [8] (by letting $N_1 = N_2 = I$).

In order to formulate the convergence statements, let us define the Lagrangian

$$\mathcal{L} : \operatorname{dom} \mathcal{F} \times \operatorname{dom} \mathcal{G} \to \mathbf{R}, \qquad \mathcal{L}(x, y) = \langle \mathcal{K}x, y\rangle + \mathcal{F}(x) - \mathcal{G}(y)$$

as well as the associated primal-dual gap:

$$\mathfrak{G}(x, y) = \mathcal{F}(x) + \mathcal{G}(\mathcal{K}^* x) + \mathcal{F}^*(-\mathcal{K}^* y) + \mathcal{G}(y) = \sup_{\substack{x' \in \operatorname{dom} \mathcal{F} \\ y' \in \operatorname{dom} \mathcal{G}}} \mathcal{L}(x, y') - \mathcal{L}(x', y). \tag{7}$$

Note that $\mathfrak{G}(x, y) = 0$ if and only if (x, y) is a saddle point of $\mathcal{L}$. In case that a saddle point exists, primal and dual solutions for (3) exist, the minimum and maximum coincides and

$$\begin{aligned} \mathfrak{G}(x, y) = \Big(\mathcal{F}(x) + \mathcal{G}^*(\mathcal{K}x) - \big(\min_{x' \in X} \mathcal{F}(x') + \mathcal{G}^*(\mathcal{K}x')\big)\Big) \\ + \Big(\mathcal{F}^*(-\mathcal{K}^* y) + \mathcal{G}(y) - \big(\min_{y' \in Y} \mathcal{F}^*(-\mathcal{K}^* y') + \mathcal{G}(y')\big)\Big), \end{aligned} \tag{8}$$

meaning that $\mathfrak{G}(x, y)$ estimates both the error in energy of the primal and dual problem for x and y, respectively.

In the following, we will not consider the "full" gap but rather restrictions of the supremum to bounded subsets $X_0 \subset \operatorname{dom} \mathcal{F}$, $Y_0 \subset \operatorname{dom} \mathcal{G}$ as, for instance, also done in [9]. This *restricted primal-dual gap* is then defined as

$$\mathfrak{G}_{X_0 \times Y_0}(x, y) = \sup_{(x', y') \in X_0 \times Y_0} \mathcal{L}(x, y') - \mathcal{L}(x', y),$$

which obviously satisfies $\mathfrak{G}_{X_0 \times Y_0}(x, y) \geq 0$ as soon as there exists a saddle point of $\mathcal{L}$ in $X_0 \times Y_0$. Our goal is to derive an asymptotic rate for an ergodic sequence generated by the general preconditioned Douglas-Rachford iteration introduced in [5]. Recall that the latter has the form

$$\begin{cases} x^{k+1} = N_1^{-1}[(N_1 - I)x^k + \bar{x}^k - \sigma \mathcal{K}^* y^{k+1}], \\ y^{k+1} = N_2^{-1}[(N_2 - I)y^k + \bar{y}^k + \sigma \mathcal{K} x^{k+1}], \\ \bar{x}^{k+1} = \bar{x}^k + (I + \sigma \partial \mathcal{F})^{-1}[2x^{k+1} - \bar{x}^k] - x^{k+1}, \\ \bar{y}^{k+1} = \bar{y}^k + (I + \sigma \partial \mathcal{G})^{-1}[2y^{k+1} - \bar{y}^k] - y^{k+1} \end{cases} \tag{9}$$

with $N_1 : X \to X$, $N_2 : Y \to Y$ and linear, continuous, self-adjoint operators satisfying $N_1 - I \geq 0$ and $N_2 - I \geq 0$. For the analysis of the method, it is useful to consider the vectors $u^k = (x^k, y^k, \bar{x}^k, \bar{y}^k)$ and an associated degenerate inner product/semi-norm according to

$$\langle u, u'\rangle_{\mathcal{M}} = \langle (N_1 - I)x, x'\rangle + \langle (N_2 - I)y, y'\rangle + \langle \bar{x}, \bar{x}'\rangle + \langle \bar{y}, \bar{y}'\rangle, \tag{10}$$

as well as $\|u\|_{\mathcal{M}}^2 = \langle u, u\rangle_{\mathcal{M}}$ for $u = (x, y, \bar{x}, \bar{y})$ and $u' = (x', y', \bar{x}', \bar{y}')$. In case of existence of a saddle point, the iteration (9) converges weakly to a fixed point $u^* = (x^*, y^*, \bar{x}^*, \bar{y}^*)$ for which the pair (x^*, y^*) is a solution of (2), see [5, Theorem 2.3]. Now, consider

$$\begin{cases} x_{\text{test}}^{k+1} = (I + \sigma \partial \mathcal{F})^{-1}[2x^{k+1} - \bar{x}^k] = \bar{x}^{k+1} - \bar{x}^k + x^{k+1}, \\ y_{\text{test}}^{k+1} = (I + \sigma \partial \mathcal{G})^{-1}[2y^{k+1} - \bar{y}^k] = \bar{y}^{k+1} - \bar{y}^k + y^{k+1}. \end{cases} \tag{11}$$

It is clear that in case of weak convergence of the general algorithm, $\{(x_{\text{test}}^k, y_{\text{test}}^k)\}$ also converges weakly to a saddle point (x^*, y^*) of (2). Moreover, $\mathcal{F}(x_{\text{test}}^k)$ and $\mathcal{G}(y_{\text{test}}^k)$ are finite for each $k \geq 1$. Our aim is now to show a convergence rate for the associated ergodic sequences, i.e., for

$$x_{\text{erg}}^k = \frac{1}{k} \sum_{k'=1}^{k} x_{\text{test}}^{k'}, \qquad y_{\text{erg}}^k = \frac{1}{k} \sum_{k'=1}^{k} y_{\text{test}}^{k'}.$$

Theorem 2 *Let $X_0 \subset \operatorname{dom} F$, $Y_0 \subset \operatorname{dom} G$ be bounded sets such that $X_0 \times Y_0$ contains a saddle point of $\mathcal{L}$. Then,*

$$0 \leq \mathfrak{G}_{X_0 \times Y_0}(x_{\text{erg}}^k, y_{\text{erg}}^k) \leq \frac{1}{k}\Big[\sup_{(x',y') \in X_0 \times Y_0} \frac{\|u' - u^0\|_{\mathcal{M}}^2}{2\sigma}\Big] = \mathcal{O}(1/k) \tag{12}$$

where u' denotes the vector $u' = (x', y', x' + \sigma\mathcal{K}^* y', y' - \sigma\mathcal{K}x')$.

The proof of Theorem 2 relies on the following two lemmas.

Lemma 1 *For the iteration sequence $\{u^k\} = \{(x^k, y^k, \bar{x}^k, \bar{y}^k)\}$, we have for all k:*

$$\langle x^{k+1} - \bar{x}^{k+1}, x_{\text{test}}^{k+1} \rangle + \langle y^{k+1} - \bar{y}^{k+1}, y_{\text{test}}^{k+1} \rangle = \langle u^k - u^{k+1}, u^{k+1} \rangle_{\mathcal{M}}.$$

Proof Note that the first two equations in (9) may be written as

$$\begin{cases} x^{k+1} = \bar{x}^k - \sigma\mathcal{K}^* y^{k+1} + (N_1 - I)(x^k - x^{k+1}) \\ y^{k+1} = \bar{y}^k + \sigma\mathcal{K}x^{k+1} + (N_2 - I)(y^k - y^{k+1}) \end{cases} \tag{13}$$

Hence, with the definition of the test elements (11),

$$\begin{aligned} \langle x^{k+1} - \bar{x}^{k+1}, x_{\text{test}}^{k+1} \rangle &= \langle x^{k+1} - \bar{x}^{k+1}, \bar{x}^{k+1} - \bar{x}^k \rangle + \langle x^{k+1} - \bar{x}^{k+1}, x^{k+1} \rangle \\ &= \langle x^{k+1} - \bar{x}^{k+1}, \bar{x}^{k+1} - \bar{x}^k \rangle \\ &\quad + \langle \bar{x}^k - \bar{x}^{k+1} - \sigma\mathcal{K}^* y^{k+1} + (N_1 - I)(x^k - x^{k+1}), x^{k+1} \rangle \\ &= \langle \bar{x}^k - \bar{x}^{k+1}, \bar{x}^{k+1} \rangle + \langle (N_1 - I)(x^k - x^{k+1}), x^{k+1} \rangle - \sigma \langle \mathcal{K}^* y^{k+1}, x^{k+1} \rangle \end{aligned}$$

and, similarly,

$$\langle y^{k+1} - \bar{y}^{k+1}, y_{\text{test}}^{k+1} \rangle = \langle \bar{y}^k - \bar{y}^{k+1}, \bar{y}^{k+1} \rangle + \langle (N_2 - I)(y^k - y^{k+1}), y^{k+1} \rangle + \sigma \langle \mathcal{K}x^{k+1}, y^{k+1} \rangle.$$

Adding both expressions and employing the definition of the $\mathcal{M}$ scalar product (10), we get the desired statement

$$\langle x^{k+1} - \bar{x}^{k+1}, x_{\text{test}}^{k+1} \rangle + \langle y^{k+1} - \bar{y}^{k+1}, y_{\text{test}}^{k+1} \rangle = \langle u^k - u^{k+1}, u^{k+1} \rangle_{\mathcal{M}}. \qquad \square$$

Lemma 2 *For $x \in \operatorname{dom} \mathcal{F}$, $y \in \operatorname{dom} \mathcal{G}$ and $u = (x, y, x + \sigma\mathcal{K}^* y, y - \sigma\mathcal{K}x)$ it holds, for all k, that*

$$\mathcal{L}(x_{\text{test}}^{k+1}, y) - \mathcal{L}(x, y_{\text{test}}^{k+1}) \leq \frac{1}{2\sigma}\big(\|u - u^k\|_{\mathcal{M}}^2 - \|u - u^{k+1}\|_{\mathcal{M}}^2 - \|u^k - u^{k+1}\|_{\mathcal{M}}^2\big). \tag{14}$$

Proof By (11), we see

$$2x^{k+1} - \bar{x}^k \in x_{\text{test}}^{k+1} + \sigma\partial\mathcal{F}(x_{\text{test}}^{k+1}), \qquad 2y^{k+1} - \bar{y}^k \in y_{\text{test}}^{k+1} + \sigma\partial\mathcal{G}(y_{\text{test}}^{k+1})$$

which is equivalent to

$$\frac{x^{k+1} - \bar{x}^{k+1}}{\sigma} \in \partial\mathcal{F}(x_{\text{test}}^{k+1}), \qquad \frac{y^{k+1} - \bar{y}^{k+1}}{\sigma} \in \partial\mathcal{G}(y_{\text{test}}^{k+1}).$$

By the definition of subgradient, we have, for $(x, y) \in X \times Y$,

$$\mathcal{F}(x_{\text{test}}^{k+1}) - \mathcal{F}(x) \leq \frac{1}{\sigma}\langle x^{k+1} - \bar{x}^{k+1}, x_{\text{test}}^{k+1} - x \rangle, \qquad \mathcal{G}(y_{\text{test}}^{k+1}) - \mathcal{G}(y) \leq \frac{1}{\sigma}\langle y^{k+1} - \bar{y}^{k+1}, y_{\text{test}}^{k+1} - y \rangle.$$

Hence, employing Lemma 1 yields

$$
\begin{aligned}
&\mathcal{L}(x_{\mathrm{test}}^{k+1}, y) - \mathcal{L}(x, y_{\mathrm{test}}^{k+1}) \\
&\qquad \leq \langle y, \mathcal{K}x_{\mathrm{test}}^{k+1}\rangle - \langle x, \mathcal{K}^* y_{\mathrm{test}}^{k+1}\rangle + \frac{1}{\sigma}\langle x^{k+1} - \bar{x}^{k+1}, x_{\mathrm{test}}^{k+1} - x\rangle + \frac{1}{\sigma}\langle y^{k+1} - \bar{y}^{k+1}, y_{\mathrm{test}}^{k+1} - y\rangle \\
&\qquad = \langle \mathcal{K}^* y, \bar{x}^{k+1} - \bar{x}^k\rangle + \langle y, \mathcal{K}x^{k+1}\rangle - \langle \mathcal{K}x, \bar{y}^{k+1} - \bar{y}^k\rangle - \langle x, \mathcal{K}^* y^{k+1}\rangle \\
&\qquad\quad + \frac{1}{\sigma}\big(\langle u^k - u^{k+1}, u^{k+1}\rangle_{\mathcal{M}} - \langle x^{k+1} - \bar{x}^{k+1}, x\rangle - \langle y^{k+1} - \bar{y}^{k+1}, y\rangle\big)
\end{aligned}
\tag{15}
$$

The identities (13) allow to reformulate parts of the right-hand side as

$$
\begin{aligned}
&\langle y, \mathcal{K}x^{k+1}\rangle - \langle x, \mathcal{K}^* y^{k+1}\rangle + \frac{1}{\sigma}\big(\langle \bar{x}^{k+1} - x^{k+1}, x\rangle + \langle \bar{y}^{k+1} - y^{k+1}, y\rangle\big) \\
&\qquad = \frac{1}{\sigma}\big(\langle x, \bar{x}^{k+1} - \bar{x}^k + (N_1 - I)(x^{k+1} - x^k)\rangle + \langle y, \bar{y}^{k+1} - \bar{y}^k + (N_2 - I)(y^{k+1} - y^k)\rangle\big).
\end{aligned}
$$

Plugged into (15), this yields, letting $u = (x, y, x + \sigma\mathcal{K}^* y, y - \sigma\mathcal{K}x)$ and recalling (10),

$$
\begin{aligned}
\mathcal{L}(x_{\mathrm{test}}^{k+1}, y) - \mathcal{L}(x, y_{\mathrm{test}}^{k+1}) \leq \frac{1}{\sigma}\big(&\langle x + \sigma\mathcal{K}^* y, \bar{x}^{k+1} - \bar{x}^k\rangle + \langle y - \sigma\mathcal{K}x, \bar{y}^{k+1} - \bar{y}^k\rangle \\
&+ \langle (N_1 - I)x, x^{k+1} - x^k\rangle + \langle (N_2 - I)y, y^{k+1} - y^k\rangle \\
&+ \langle u^k - u^{k+1}, u^{k+1}\rangle_{\mathcal{M}}\big) = \frac{1}{\sigma}\langle u^k - u^{k+1}, u^{k+1} - u\rangle_{\mathcal{M}}.
\end{aligned}
$$

Finally, the polarization identity gives $\langle u^k - u^{k+1}, u^{k+1} - u\rangle_{\mathcal{M}} = \frac{1}{2}\big(\|u^k - u\|_{\mathcal{M}}^2 - \|u^{k+1} - u\|_{\mathcal{M}}^2 - \|u^k - u^{k+1}\|_{\mathcal{M}}^2\big)$ and, consequently, the result. □

Remark 1 With the prerequisites of Theorem 2, one immediately gets from (14) that

$$
0 \leq \mathfrak{G}_{X_0 \times Y_0}(x_{\mathrm{test}}^{k+1}, y_{\mathrm{test}}^{k+1}) \leq \frac{1}{\sigma}\Big[\sup_{(x', y') \in X_0 \times Y_0} \|u^{k+1} - u'\|_{\mathcal{M}}\Big]\|u^k - u^{k+1}\|_{\mathcal{M}} \to 0
$$

as $k \to \infty$ since $\{u^k\}$ is bounded (due to weak convergence), $u' = (x', y', x' + \sigma\mathcal{K}^* y', y' - \sigma\mathcal{K}x')$ is bounded for $(x', y') \in X_0 \times Y_0$ and the iteration satisfies $\|u^k - u^{k+1}\|_{\mathcal{M}} \to 0$ as $k \to \infty$, see [5, Lemma 2.8]. Hence, the restricted primal-dual gap converges for the test sequences, however, without a rate.

Let us eventually prove Theorem 2.

Proof of Theorem 2. Let $(x, y) \in X_0 \times Y_0$. Since $\mathcal{F}$ and $\mathcal{G}$ are convex, $\mathcal{L}$ is convex in the first argument and concave in the second argument. Thus, with the help of Lemma 2 and $u = (x, y, x + \sigma\mathcal{K}^* y, y - \sigma\mathcal{K}x)$, we get

$$
\begin{aligned}
\mathcal{L}(x_{\mathrm{erg}}^k, y) - \mathcal{L}(x, y_{\mathrm{erg}}^k) &\leq \frac{1}{k}\Big(\sum_{k'=0}^{k-1} \mathcal{L}(x_{\mathrm{test}}^{k+1}, y) - \mathcal{L}(x, y_{\mathrm{test}}^{k+1})\Big) \\
&\leq \frac{1}{k}\frac{1}{2\sigma}\Big(\|u^0 - u\|_{\mathcal{M}}^2 - \|u^k - u\|_{\mathcal{M}}^2 - \sum_{k'=0}^{k-1} \|u^{k'} - u^{k'+1}\|_{\mathcal{M}}^2\Big) \leq \frac{1}{k}\frac{\|u - u^0\|_{\mathcal{M}}^2}{2\sigma}.
\end{aligned}
$$

Taking the supremum then gives the result. □

Remark 2 Let one of the functions $\mathcal{F}$ or $\mathcal{G}$ be strongly convex, say $\mathcal{F}$, i.e., there exists a $\gamma > 0$ with

$$\mathcal{F}(x) + \langle x'', x' - x\rangle + \frac{\gamma}{2}\|x' - x\|^2 \leq \mathcal{F}(x') \qquad \text{for all } x, x' \in X, x'' \in \partial\mathcal{F}(x).$$

Then, for $(x^*, y^*) \in X_0 \times Y_0$ a saddle point of $\mathcal{L}$, we can estimate $\frac{\gamma}{2}\|x - x^*\|^2 \leq \mathcal{L}(x, y^*) - \mathcal{L}(x^*, y) \leq \mathfrak{G}_{X_0\times Y_0}(x, y)$ for $\gamma > 0$ and all $(x, y) \in \operatorname{dom}\mathcal{F} \times \operatorname{dom}\mathcal{G}$, see [9, Equation (36)]. Consequently, $\lim_{k\to\infty} x^k_{\mathrm{erg}} = x^*$ with rate

$$\|x^k_{\mathrm{erg}} - x^*\| \leq \sqrt{\frac{2}{k\gamma}}\Big[\sup_{(x,y)\in X_0\times Y_0} \frac{\|u - u^0\|_{\mathcal{M}}}{\sqrt{2\sigma}}\Big] = \mathcal{O}(1/\sqrt{k}).$$

Proposition 2 *If $\mathcal{L}$ possesses a saddle point (x^*, y^*) and $\mathcal{F}$, $\mathcal{G}$ are strongly coercive, i.e.,*

$$\frac{\mathcal{F}(x')}{\|x'\|} \to \infty \quad \text{for } \|x'\| \to \infty, \qquad \frac{\mathcal{G}(y')}{\|y'\|} \to \infty \quad \text{for } \|y'\| \to \infty,$$

then there exist bounded $X_0 \subset X$ and $Y_0 \subset Y$ such that for all k,

$$\mathfrak{G}(x^k_{\mathrm{test}}, y^k_{\mathrm{test}}) = \mathfrak{G}_{X_0\times Y_0}(x^k_{\mathrm{test}}, y^k_{\mathrm{test}}) \qquad \text{and} \qquad \mathfrak{G}(x^k_{\mathrm{erg}}, y^k_{\mathrm{erg}}) = \mathfrak{G}_{X_0\times Y_0}(x^k_{\mathrm{erg}}, y^k_{\mathrm{erg}}).$$

Proof Denote by $(\{x^k_t\}, \{y^k_t\})$ either $(\{x^k_{\mathrm{test}}\}, \{y^k_{\mathrm{test}}\})$ or $(\{x^k_{\mathrm{erg}}\}, \{y^k_{\mathrm{erg}}\})$. In both cases, the $\{x^k_t\}$, $\{y^k_t\}$ as well as the functional value sequences $\{\mathcal{F}(x^k_t)\}$, $\{\mathcal{G}(y^k_t)\}$ are bounded: This follows, on the one hand, from $\{x^k_{\mathrm{test}}\}$, $\{y^k_{\mathrm{test}}\}$ being bounded due to weak convergence (see [5, Theorem 2.3] and (11)) and, consequently, from the definition of x^k_{erg} and y^k_{erg}. On the other hand, Remark 1 and Theorem 2, respectively, give the boundedness of $\mathfrak{G}_{\{x^*\}\times\{y^*\}}(x^k_t, y^k_t)$ for which the the boundedness of $\{\mathcal{F}(x^k_t)\}$, $\{\mathcal{G}(y^k_t)\}$ can easily be deduced (using the definition, the continuity of $\mathcal{K}$ as well as the fact that $\{\mathcal{F}(x^k_t)\}$, $\{\mathcal{G}(y^k_t)\}$ are bounded). Now,

$$\langle \mathcal{K}x^k_t, y'\rangle + \mathcal{F}(x^k_t) - \langle x', \mathcal{K}^* y^k_t\rangle + \mathcal{G}(y^k_t) \leq \|\mathcal{K}\|\|x^k_t\|\|y'\| + \mathcal{F}(x^k_t) + \|\mathcal{K}\|\|y^k_t\|\|x'\| + \mathcal{G}(y^k_t)$$

so boundedness as well as strong coercivity allows us to choose a $C > 0$ independent of k such that $\|\mathcal{K}\|\|y^k_t\|\|x'\| + \mathcal{F}(x^k_t) < \mathcal{F}(x')$ for all $\|x'\| > C$ as well as $\|\mathcal{K}\|\|x^k_t\|\|y'\| + \mathcal{G}(y^k_t) < \mathcal{G}(y')$ for $\|y'\| > C$. This implies $\mathcal{L}(x^k_t, y') - \mathcal{L}(x', y^k_t) < 0$ for all $\|x'\| > C$, $\|y'\| > C$ and as $\mathfrak{G}(x^k_t, y^k_t) \geq 0$, we can conclude, letting $X_0 = \{x' \in \operatorname{dom}\mathcal{F} \,\big|\, \|x'\| \leq C\}$, $Y_0 = \{y' \in \operatorname{dom}\mathcal{G} \,\big|\, \|y'\| \leq C\}$, that

$$\mathfrak{G}(x^k_t, y^k_t) = \sup_{\substack{x'\in\operatorname{dom}\mathcal{F}\\ y'\in\operatorname{dom}\mathcal{G}}} \mathcal{L}(x^k_t, y') - \mathcal{L}(x', y^k_t) = \sup_{(x',y')\in X_0\times Y_0} \mathcal{L}(x^k_t, y') - \mathcal{L}(x', y^k_t) = \mathfrak{G}_{X_0\times Y_0}(x^k_t, y^k_t).$$

As C is independent from k, this holds for the whole sequence. Furthermore, one can choose $C > 0$ such that the above identity is true for both choices of $\{x^k_t\}$, $\{y^k_t\}$. □

Hence, in the situation where this proposition is applicable, the primal-dual gap according to (7) converges for the test sequences $\{x^k_{\mathrm{test}}\}$, $\{y^k_{\mathrm{test}}\}$ and we also have the rate $\mathcal{O}(1/k)$ for the ergodic sequences $\{x^k_{\mathrm{erg}}\}$, $\{y^k_{\mathrm{erg}}\}$.

4 TV-Regularized Denoising and Deblurring

We start the application of the preconditioned Douglas-Rachford iteration to imaging problems with preconditioning strategies for discrete TV-regularized problems, i.e.,

$$\min_{u \in U} \ F(u) + \alpha \, \mathrm{TV}(u) = \min_{u \in U} \ F(u) + \alpha \|\nabla u\|_1.$$

Here, U is a discrete image space and ∇ a discrete gradient operator which will be described in the following. Following essentially the presentation in [3, 7], consider the image domain $\Omega \subset \mathbf{Z}^2$ as the discretized 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. In order not having to distinguish cases in the discretization, we assume $N_x \ge 3$ and $N_y \ge 3$. We define the image space as $U = \{u : \Omega \to \mathbf{R}\}$ with the standard L^2 scalar product. 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 $V = U^2$ with the standard product scalar product, this gives a linear operator $\nabla : U \to V$. The discrete divergence is then the negative adjoint of ∇, i.e., the unique linear mapping $\operatorname{div} : V \to U$ which satisfies

$$\langle \nabla u, p \rangle_V = \langle u, \nabla^* p \rangle_U = -\langle u, \operatorname{div} p \rangle_U, \quad \forall u \in U, \ p \in V.$$

It can be computed to read as

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

involving the backward difference operators

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

In order to define the discrete version of the TV (and later the TGV) functionals, we still need the discrete versions of the L^1, L^2 and L^∞ norms: For $u \in U$, $p = (p^1, p^2) \in V$, $1 \le t < \infty$,

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

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

Further, problem-specific discrete concepts will be introduced later in this section.

4.1 TV-Regularized Denoising Problems

Let us first 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)

$$\min_{u \in U} \ F(u) + \alpha \|\nabla u\|_1, \qquad F(u) = \begin{cases} \frac{1}{2}\|u - f\|_2^2 & \text{for } L^2\text{-denoising,} \\ \|u - f\|_1 & \text{for } L^1\text{-denoising,} \end{cases} \tag{16}$$

with $f : \Omega \to \mathbf{R}$ a given 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 a equivalent saddle-point problem [11, 17] of type (2) with

$$X = U, \quad Y = V, \quad \mathcal{F} = F, \quad \mathcal{K} = \nabla \quad \text{and} \quad \mathcal{G} = \mathcal{I}_{\{\|v\|_\infty \leq \alpha\}}$$

while $\mathcal{I}_C$ denotes the indicator function of the set C, i.e.,

$$\mathcal{I}_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{else.} \end{cases} \tag{17}$$

With these prerequisites, preconditioned Douglas-Rachford methods can be derived.

4.1.1 The L^2-Denoising Case

Following [5], in the case of quadratic discrepancy, F is purely quadratic-linear (up to a constant), so one can use the iteration (PDRQ), i.e., the data

$$\mathcal{K} = \nabla, \quad Q = I, \quad f_0 = f, \quad \mathcal{G} = \mathcal{I}_{\{\|p\|_\infty \leq \alpha\}}.$$

The operator M should be a feasible preconditioner for $\sigma I - \sigma^2 \operatorname{div} \nabla = \sigma I - \sigma^2 \Delta$ (where $\Delta = \operatorname{div} \nabla$) and will be discussed in Subsection 4.1.3. The iteration then needs the resolvent $(I + \sigma\partial\mathcal{G})^{-1} = \mathcal{P}_\alpha$ which reads as [3]

$$\mathcal{P}_\alpha(p) = (I + \sigma\partial\mathcal{G})^{-1}(p) = \operatorname*{arg\,min}_{p' \in V} \ \frac{1}{2}\|p' - p\|_2^2 + \mathcal{I}_{\{\|p\|_\infty \leq \alpha\}}(p') = \frac{p}{\max(1, |p|/\alpha)} \tag{18}$$

with $|p| = \sqrt{(p^1)^2 + (p^2)^2}$ in the pointwise sense.

4.1.2 The L^1-Denoising Case

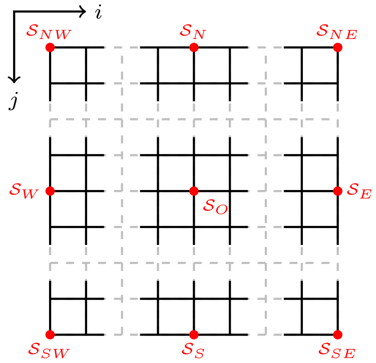
Here, one may use the iteration (PDR). It turns out to be beneficial to scale the gradient with a factor $\tau > 0$, leading to the following data:

$$\mathcal{K} = \tau\nabla, \quad \mathcal{F} = \|\cdot - f\|_1, \quad \mathcal{G} = \mathcal{I}_{\{\|p\|_\infty \leq \frac{\alpha}{\tau}\}}.$$

Analogously, M should be a preconditioner for $I - \sigma^2\Delta$. The resolvent of ∂F is moreover involved which is known to realize a soft-thresholding (or shrinkage) operation around f, i.e., $(I + \sigma\partial\mathcal{F})^{-1} = \mathcal{S}_\sigma(\cdot, f)$ with

$$\begin{aligned} (I + \sigma\partial\mathcal{F})^{-1}(u) &= \operatorname*{arg\,min}_{u' \in U} \ \frac{1}{2}\|u' - u\|_2^2 + \sigma\|u' - f\|_1 \\ &= \mathcal{S}_\sigma(u, f) = f + \operatorname{sign}(u - f)\max(0, |u - f| - \sigma). \end{aligned} \tag{19}$$

The respective resolvent for $\partial\mathcal{G}$ is again given by (18).



$$\underbrace{\begin{bmatrix} \lambda+2\mu & -\mu \\ -\mu & \end{bmatrix}}_{=\mathcal{S}_{NW}} \quad \underbrace{\begin{bmatrix} -\mu & \lambda+3\mu & -\mu \\ & -\mu & \end{bmatrix}}_{=\mathcal{S}_{N}} \quad \underbrace{\begin{bmatrix} -\mu & \lambda+2\mu \\ & -\mu \end{bmatrix}}_{=\mathcal{S}_{NE}}$$

$$\underbrace{\begin{bmatrix} -\mu & \\ \lambda+3\mu & -\mu \\ -\mu & \end{bmatrix}}_{=\mathcal{S}_{W}} \quad \underbrace{\begin{bmatrix} & -\mu & \\ -\mu & \lambda+4\mu & -\mu \\ & -\mu & \end{bmatrix}}_{=\mathcal{S}_{O}} \quad \underbrace{\begin{bmatrix} & -\mu \\ -\mu & \lambda+3\mu \\ & -\mu \end{bmatrix}}_{=\mathcal{S}_{E}}$$

$$\underbrace{\begin{bmatrix} -\mu & \\ \lambda+2\mu & -\mu \end{bmatrix}}_{=\mathcal{S}_{SW}} \quad \underbrace{\begin{bmatrix} & -\mu & \\ -\mu & \lambda+3\mu & -\mu \end{bmatrix}}_{=\mathcal{S}_{S}} \quad \underbrace{\begin{bmatrix} & -\mu \\ -\mu & \lambda+2\mu \end{bmatrix}}_{=\mathcal{S}_{SE}}$$

Table 4: Finite-difference stencils for $\lambda u - \mu \Delta u$ with homogeneous Neumann boundary conditions. The highlighted entry denotes the center element and u is assumed to be extended by arbitrary values outside of Ω.

Remark 3 Alternatively, (PDRQ) may also be employed by introducing an additional dual variable $v \in U$ and writing $\|u - f\|_1 = \max_{v \in U} \langle u, v\rangle - \big(\mathcal{I}_{\{\|v\|_\infty \leq 1\}}(v) + \langle f, v\rangle\big)$, resulting in the dual space $Y = V \times U$, operator $\mathcal{K}u = (\nabla u, u)$, $\mathcal{F} = 0$ and

$$\mathcal{G}(y) = \mathcal{G}(p, v) = G_1(p) + G_2(v), \quad G_1 = \mathcal{I}_{\{\|p\|_\infty \leq \alpha\}}, \quad G_2 = \mathcal{I}_{\{\|v\|_\infty \leq 1\}} + \langle f, \cdot\rangle.$$

The proximal mapping for the dual functional can be computed separately for p and v. For G_1, this is again given by (18), while for G_2, we have

$$\begin{aligned} (I + \sigma \partial G_2)^{-1}(v) &= \operatorname*{arg\,min}_{v' \in U} \frac{1}{2}\|v' - v\|_2^2 + \sigma \langle f, v'\rangle + \mathcal{I}_{\{\|v\|_\infty \leq 1\}}(v') \\ &= \max\big(-1, \min(1, v - \sigma f)\big). \end{aligned} \tag{20}$$

We will compare, in Section 7, the performance of both variants.

4.1.3 The Preconditioner

In both the L^2 and L^1 case, M is required to be a feasible preconditioner for operators of type $T = \lambda I - \mu \Delta$ for $\lambda, \mu > 0$ where $\Delta = \operatorname{div} \nabla$ can be interpreted as a discrete Laplace operator with homogeneous Neumann boundary conditions [25]. In other words: $Tu = b$ corresponds to a discrete version of the boundary value problem

$$\begin{cases} \lambda u - \mu \Delta u = b & \text{in } \Omega, \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\Omega. \end{cases} \tag{21}$$

With the choice of ∇u, we get a finite-difference equation with five-point stencils according to Table 4. As in [5], we will use a symmetric Gauss-Seidel update in conjunction with a Red-Black enumeration scheme as follows. Let

$$\Omega_{\text{red}} = \{(i,j) \in \Omega \mid i + j \text{ even}\}, \qquad \Omega_{\text{black}} = \{(i,j) \in \Omega \mid i + j \text{ odd}\},$$

and write $u = (u_{\text{red}}, u_{\text{black}})$ where u_{red} and u_{black} represent the image u on Ω_{red} and Ω_{black}, respectively. A Gauss-Seidel update then performs an update according to red $\to$ black while its

PDRQ Objective: L^2-TV denoising		$\min_{u\in U}\ \frac{1}{2}\|u-f\|_2^2 + \alpha\|\nabla u\|_1$
Initialization:	$(u^0, \bar{p}^0) \in U \times V$ initial guess, $\sigma > 0$ step-size, $n \geq 1$ inner iterations for symmetric Gauss-Seidel	
Iteration:	$u^{k+1} = \mathrm{SRBGS}^n_{\sigma,\sigma^2}\big(u^k, \sigma(f + \operatorname{div}\bar{p}^k)\big)$	according to (22)
	$p^{k+1} = \bar{p}^k + \sigma\nabla u^{k+1}$	
	$p^{k+1}_{\mathrm{test}} = \mathcal{P}_\alpha(2p^{k+1} - \bar{p}^k)$	according to (18)
	$\bar{p}^{k+1} = \bar{p}^k + p^{k+1}_{\mathrm{test}} - p^{k+1}$	

Table 5: The preconditioned Douglas-Rachford iteration for L^2-TV denoising.

adjoint corresponds to the reverse order black → red. One can easily see that a update of red and black pixels, respectively, is idempotent. Thus, n steps of symmetric Red-Black Gauss-Seidel, denoted by $\mathrm{SRBGS}^n_{\lambda,\mu}$, corresponds to the following procedure. Denoting $\mathcal{N}(i,j) = \{(i',j') \in \Omega \,\big|\, |i-i'| + |j-j'| = 1\}$ the set of neighbor points of (i,j) in Ω and $c_{i,j} = \#\mathcal{N}(i,j)$ (which is either 2, 3 or 4, depending on whether (i,j) is a corner, edge or interior point), we have

$$\begin{cases} \mathrm{SRBGS}^n_{\lambda,\mu}(u^k, b^k) = (u^{k+1}_{\mathrm{red}}, u^{k+1}_{\mathrm{black}}), \\ (u^{k+(\nu+1/2)/n}_{\mathrm{red}})_{i,j} = \dfrac{1}{\lambda + c_{i,j}\mu}\Big(b^k_{i,j} + \mu \displaystyle\sum_{(i',j')\in\mathcal{N}(i,j)} (u^{k+\nu/n}_{\mathrm{black}})_{i',j'}\Big) & (i,j) \in \Omega_{\mathrm{red}}, \\ (u^{k+(\nu+1)/n}_{\mathrm{black}})_{i,j} = \dfrac{1}{\lambda + c_{i,j}\mu}\Big(b^k_{i,j} + \mu \displaystyle\sum_{(i',j')\in\mathcal{N}(i,j)} (u^{k+(\nu+1/2)/n}_{\mathrm{red}})_{i',j'}\Big) & (i,j) \in \Omega_{\mathrm{black}}, \\ & \nu = 0,\ldots,n-1, \\ (u^{k+1}_{\mathrm{red}})_{i,j} = \dfrac{1}{\lambda + c_{i,j}\mu}\Big(b^k_{i,j} + \mu \displaystyle\sum_{(i',j')\in\mathcal{N}(i,j)} (u^{k+1}_{\mathrm{black}})_{i',j'}\Big) & (i,j) \in \Omega_{\mathrm{red}}. \end{cases} \tag{22}$$

According to Table 3, $\mathrm{SRBGS}^1_{\lambda,\mu}$ corresponds to a feasible preconditioner as it realizes a symmetric Gauss-Seidel step and of course, n-fold application $\mathrm{SRBGS}^n_{\lambda,\mu}$, is also feasible, see Proposition 1.

4.1.4 The Algorithms

Putting the building blocks together, we get a PDRQ method for L^2-TV denoising according to Table 5 (which was already derived in [5]) as well as a PDR and PDRQ method for L^1-TV denoising according to Tables 6 and 7, respectively. As U and V are finite-dimensional, all algorithms converge to a respective solution by virtue of Theorem 1 for any choice of $\sigma > 0$ and $\tau > 0$.

Remark 4 For a specific choice of the step-size σ and scaling τ, note that T is corresponding to the Neumann problem (21), which possesses a unique solution as $\lambda > 0$. However, for large μ/λ, the problem approximates a pure Neumann problem which requires compatibility conditions on b in order to be solvable. Numerically, the problem is ill-posed in this situation making the preconditioner less effective. Therefore, μ/λ should not be chosen too large. On the other hand, if μ/λ is too small, solving (21) is close to the inversion of the identity (up to a factor), hence the update step for u does not contribute much to the solution of the optimization problem. Therefore, μ/λ should not be chosen too low.

PDR Objective: L^1-TV denoising	$\min_{u\in U} \ \|u-f\|_1 + \alpha\|\nabla u\|_1$	
Initialization:	$(u^0,\bar u^0,\bar p^0) \in U\times U\times V$ initial guess, $\sigma>0$ step-size, $\tau>0$ gradient scaling, $n\geq 1$ inner iterations for symmetric Gauss-Seidel	
Iteration:	$u^{k+1} = \mathrm{SRBGS}^n_{1,(\sigma\tau)^2}\big(u^k, \bar u^k + \sigma\tau \operatorname{div} \bar p^k\big)$	according to (22)
	$p^{k+1} = \bar p^k + \sigma\tau\nabla u^{k+1}$	
	$u^{k+1}_{\mathrm{test}} = \mathcal{S}_\sigma(2u^{k+1} - \bar u^k, f)$	according to (19)
	$\bar u^{k+1} = \bar u^k + u^{k+1}_{\mathrm{test}} - u^{k+1}$	
	$p^{k+1}_{\mathrm{test}} = \mathcal{P}_{\alpha/\tau}(2p^{k+1} - \bar p^k)$	according to (18)
	$\bar p^{k+1} = \bar p^k + p^{k+1}_{\mathrm{test}} - p^{k+1}$	

Table 6: The scaled preconditioned Douglas-Rachford iteration for L^1-TV denoising.

PDRQ Objective: L^1-TV denoising	$\min_{u\in U} \ \|u-f\|_1 + \alpha\|\nabla u\|_1$	
Initialization:	$(u^0,\bar p^0,\bar v^0) \in U\times V\times U$ initial guess, $\sigma>0$ step-size, $n\geq 1$ inner iterations for symmetric Gauss-Seidel	
Iteration:	$u^{k+1} = \mathrm{SRBGS}^n_{\sigma^2,\sigma^2}\big(u^k, -\sigma\bar v^k + \sigma \operatorname{div} \bar p^k\big)$	according to (22)
	$p^{k+1} = \bar p^k + \sigma\nabla u^{k+1}$	
	$v^{k+1} = \bar v^k + \sigma u^{k+1}$	
	$p^{k+1}_{\mathrm{test}} = \mathcal{P}_\alpha(2p^{k+1} - \bar p^k)$	according to (18)
	$\bar p^{k+1} = \bar p^k + p^{k+1}_{\mathrm{test}} - p^{k+1}$	
	$v^{k+1}_{\mathrm{test}} = \max\big(-1, \min(1, 2v^{k+1} - \bar v^k - \sigma f)\big)$	according to (20)
	$\bar v^{k+1} = \bar v^k + v^{k+1}_{\mathrm{test}} - v^{k+1}$	

Table 7: The PDRQ variant for L^1-TV denoising.

For the L^2-TV-denoising problem these considerations suggest to take σ such that $1+\sigma\|\Delta\|$ is moderate. Likewise, for L^1-TV-denoising, it is reasonable to choose $\tau = c/\sigma$ with moderate $1+c^2\|\Delta\|$ for the PDR method and, in view of (18) and (19), to choose σ neither too large nor too small. For examples on how σ and τ might be chosen in order to make the methods efficient, we refer to Section 7, where numerical experiments are presented.

4.2 TV-Deblurring Problems

Next, the preconditioned Douglas-Rachford framework is applied to ill-posed linear inverse problems. We aim at applying it to deblurring problems of the type

$$\min_{u\in U}\ F(u) + \alpha\|\nabla u\|_1, \qquad F(u) = \begin{cases} \frac12\|u*\kappa - f\|_2^2 & \text{for } L^2\text{-deblurring}, \\ \|u*\kappa - f\|_1 & \text{for } L^1\text{-deblurring}. \end{cases} \tag{23}$$

Here, $\kappa : \Omega \to \mathbf{R}$ is a given convolution kernel satisfying $\sum_{i,j}\kappa_{i,j} \neq 0$. For sake of simplicity, we choose to use the *periodic convolution*, i.e., $u*\kappa$ denotes

$$(u*\kappa)_{i,j} = \sum_{i'=0}^{N_x-1}\sum_{j'=0}^{N_y-1} u_{i-i' \bmod N_x, j-j' \bmod N_y}\kappa_{i',j'}$$

with $a \bmod b$ being the unique number $c \in \{0, \ldots, b-1\}$ such that there is a $d \in \mathbf{Z}$ with $a = db+c$. We will refer to the periodic convolution operator with κ as L and also introduce the periodic gradient $\nabla_p : U \to V$ according to

$$\nabla_p u = \begin{pmatrix} \partial_x^{p+} u \\ \partial_y^{p+} u \end{pmatrix}, \quad (\partial_x^{p+} u)_{i,j} = u_{i+1 \bmod N_x, j} - u_{i,j}, \quad (\partial_y^{p+} u)_{i,j} = u_{i,j+1 \bmod N_y} - u_{i,j}$$

for $(i,j) \in \Omega$ as well as the periodic divergence $\operatorname{div}_p = -\nabla_p^*$ and Laplacian $\Delta_p = \operatorname{div}_p \nabla_p$. Let us again discuss the preconditioned Douglas-Rachford iterations for L^2- and L^1-discrepancies separately.

4.2.1 The L^2-Deblurring Case

The discrepancy functional $\mathcal{F}(u) = \frac{1}{2}\|u * \kappa - f\|_2^2$ in (23) can be rewritten to be quadratic-linear, hence, we employ again (PDRQ). Using the scaled gradient with a factor $\tau > 0$ data then reads as

$$\mathcal{K} = \tau \nabla, \quad Q = L^* L, \quad f_0 = L^* f, \quad \mathcal{G} = \mathcal{I}_{\{\|p\|_\infty \leq \alpha/\tau\}}.$$

Recall that L denotes the convolution operator with respect to κ, i.e., $Lu = u * \kappa$ for all $u \in U$. Let us choose a feasible preconditioner for $T = \sigma L^* L - (\sigma\tau)^2 \Delta$. One easily sees that $M = \sigma L^* L - (\sigma\tau)^2 \Delta_p$ satisfies $M - T \geq 0$ as

$$\langle (M-T)u, u \rangle = \langle -\Delta_p u, u \rangle - \langle -\Delta u, u \rangle = \|\nabla_p u\|^2 - \|\nabla u\|^2 \geq 0 \quad \text{for each} \quad u \in U,$$

where the latter inequality is due to the fact that $\partial_x^+ u$ is just $\partial_x^{p+} u$ with some entries replaced by 0 (and the analog for $\partial_y^+ u$). Since L^* is the periodic convolution with the mirrored kernel, M is the periodic convolution with the kernel κ_M according to

$$\kappa_M = \sigma \kappa' * \kappa + (\sigma\tau)^2 \kappa_\Delta, \qquad \begin{cases} \kappa'_{i,j} = \kappa_{N_x-i-1, N_y-j-1}, \\ (i,j) \in \Omega \end{cases} \qquad \kappa_\Delta = \begin{bmatrix} 0 & -1 & 0 \\ -1 & \boxed{4} & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{24}$$

where the highlighted entry corresponds to $i = j = 0$. Hence, M is diagonalized by the discrete Fourier transform $\mathcal{F}$ and can be inverted by

$$M^{-1} u = \mathcal{F}^{-1}\Big(\frac{1}{\mathcal{F}\kappa_M}(\mathcal{F}u)\Big) = \mathcal{F}^{-1}\Big(\frac{\mathcal{F}u}{\sigma|\mathcal{F}\kappa|^2 + (\sigma\tau)^2 \mathcal{F}\kappa_\Delta}\Big) \tag{25}$$

provided that the denominator is positive in each point. This is however the case as $(\mathcal{F}\kappa_\Delta)_{i,j} = 8\sin^2(i\pi/N_x) + 8\sin^2(j\pi/N_y)$ vanishes only for $i = j = 0$ and since $(\mathcal{F}\kappa)_{0,0} = \sum_{i,j} \kappa_{i,j} \neq 0$ by assumption. In particular, M is positive definite, leading to its feasibility for T.

Putting all together yields the PDRQ method for the solution of the L^2-TV-deblurring problem, see Table 8. Of course, an implementation should precompute and store $f * \kappa'$ as well as $|\mathcal{F}\kappa|^2 + \sigma\tau^2 \mathcal{F}\kappa_\Delta$ and utilize the Fast Fourier transform (FFT) [3, 9] in order to gain maximal efficiency.

PDRQ Objective: L^2-TV deblurring	$\min_{u\in U}\ \frac{1}{2}\|u*\kappa - f\|_2^2 + \alpha\|\nabla u\|_1$	
Initialization:	$(u^0,\bar p^0)\in U\times V$ initial guess, $\sigma>0$ step-size, $\tau>0$ gradient scaling	
Iteration:	$u^{k+1} = \mathcal{F}^{-1}\Big(\dfrac{\mathcal{F}(f*\kappa') + \mathcal{F}(\tau\operatorname{div}\bar p^k + \sigma\tau^2(\Delta-\Delta_p)u^k)}{\lvert\mathcal{F}\kappa\rvert^2 + \sigma\tau^2\mathcal{F}\kappa_\Delta}\Big)$	according to (24)
	$p^{k+1} = \bar p^k + \sigma\tau\nabla u^{k+1}$	
	$p_{\text{test}}^{k+1} = \mathcal{P}_{\alpha/\tau}(2p^{k+1} - \bar p^k)$	according to (18)
	$\bar p^{k+1} = \bar p^k + p_{\text{test}}^{k+1} - p^{k+1}$	

Table 8: The scaled preconditioned Douglas-Rachford iteration for the solution of the L^2-TV deblurring problem.

4.2.2 The L^1-Deblurring Case

This case is slightly different from L^1-TV-denoising as the proximal mapping associated with $\|L\cdot - f\|_1$ is not easy to compute. We therefore introduce another dual variable $v\in U$ and obtain (again, with a gradient scaling $\tau>0$) the saddle-point formulation

$$\min_{u\in U}\max_{\substack{p\in V\\ v\in U}}\ \langle\tau\nabla u, p\rangle + \langle Lu, v\rangle - G_1(p) - G_2(v). \tag{26}$$

$G_1 = \mathcal{I}_{\{\|p\|_\infty\le\alpha/\tau\}}$ and $G_2 = \mathcal{I}_{\{\|v\|_\infty\le 1\}} + \langle f,\cdot\rangle$ (also see Remark 3). For (PDRQ), this leads to the data

$$X = U,\quad Y = V\times U,\quad \mathcal{K} = \begin{bmatrix}\nabla\\ L\end{bmatrix},\quad Q = 0,\quad f_0 = 0,\quad \mathcal{G}(y) = \mathcal{G}(p,v) = G_1(p) + G_2(v),$$

and we have to choose a preconditioner for $T = \sigma^2\mathcal{K}^*\mathcal{K} = \sigma^2 L^*L - (\sigma\tau)^2\Delta$. Again, $M = \sigma^2 L^*L - (\sigma\tau)^2\Delta_p$ is feasible and can be computed as in (25) with a straightforward modification.

This results in the PDRQ algorithm for L^1-TV-deblurring which can be found in Table 9. This version assumes that it is beneficial to compute the convolution via pointwise multiplication on the Fourier side, hence the transformed variable $\hat u^k$ has been introduced. Again, one should consider precomputing $\mathcal{F}\kappa$ and $\sigma\lvert\mathcal{F}\kappa\rvert^2 + \sigma\tau^2\mathcal{F}\kappa_\Delta$ as well as using the Fast Fourier transform in order to obtain maximal efficiency.

5 TGV-Regularized Denoising Problems

Having discussed the derivation of preconditioned Douglas-Rachford algorithms for TV regularization, we now turn to developing preconditioned iterative methods for total generalized variation (TGV) regularization, for simplicity of second order. The concept of TGV, introduced in [7], realizes a convex regularizing functional which automatically adapts to higher-order smoothness while sharing the convenient properties of TV, in particular, the ability to model images with edges or, more generally, discontinuities on hypersurfaces. Especially, TGV is able to significantly reduce the so-called *staircasing effect* in the solutions, i.e., the undesired appearance of edges. It works convincingly well in various fields, e.g., for linear inverse problems [6], for diffusing tensor imaging [28], for JPEG decompression [4], for multichannel images [3] and in magnetic resonance imaging (MRI) [18].

The second-order total generalized variation TGV_α^2 is defined as in [6, 7].

PDRQ Objective: L^1-TV deblurring	$\min_{u\in U}\ \|u*\kappa - f\|_1 + \alpha\|\nabla u\|_1$	
Initialization:	$(u^0, \bar p^0, \bar v^0) \in U \times V \times U$ initial guess, $\hat u^0 = \mathcal{F}u^0$, $\sigma > 0$ step-size, $\tau > 0$ gradient scaling	
Iteration:	$\hat u^{k+1} = \dfrac{\mathcal{F}\big(\tau \operatorname{div} \bar p^k + \sigma\tau^2(\Delta - \Delta_p)u^k\big) - \mathcal{F}\bar v^k \overline{\mathcal{F}\kappa}}{\sigma\lvert\mathcal{F}\kappa\rvert^2 + \sigma\tau^2\mathcal{F}\kappa_\Delta}$	
	$u^{k+1} = \mathcal{F}^{-1}\hat u^{k+1}$	
	$p^{k+1} = \bar p^k + \sigma\tau\nabla u^{k+1}$	
	$p^{k+1}_{\text{test}} = \mathcal{P}_{\alpha/\tau}(2p^{k+1} - \bar p^k)$	according to (18)
	$\bar p^{k+1} = \bar p^k + p^{k+1}_{\text{test}} - p^{k+1}$	
	$v^{k+1} = \bar v^k + \sigma\mathcal{F}^{-1}(\hat u^{k+1}\mathcal{F}\kappa)$	
	$v^{k+1}_{\text{test}} = \max\big(-1, \min(1, 2v^{k+1} - \bar v^k - \sigma f)\big)$	according to (20)
	$\bar v^{k+1} = \bar v^k + v^{k+1}_{\text{test}} - v^{k+1}$	

Table 9: The scaled preconditioned Douglas-Rachford iteration for L^1-TV deblurring.

Definition 2 Let $\Omega \subset \mathbf{R}^d$ be a bounded Lipschitz domain and $\alpha = (\alpha_0, \alpha_1) > 0$ regularization parameters. The *Total Generalized Variation* of second order of $u \in L^1_{loc}(\Omega)$ is the value assigned by the following functional

$$\mathrm{TGV}^2_\alpha(u) = \sup\Big\{\int_\Omega u \operatorname{div}^2 q \,\mathrm{d}x \;\Big|\; q \in \mathcal{C}^2_c(\Omega, S^{d\times d}),\ \|q\|_\infty \le \alpha_0,\ \|\operatorname{div} q\|_\infty \le \alpha_1\Big\}. \tag{27}$$

Here, $S^{d\times d}$ is the set of symmetric matrices and $\mathcal{C}^2_c(\Omega, S^{d\times d})$ denotes the vector space of compactly supported, twice continuously differentiable $S^{d\times d}$-valued mappings. The divergences $\operatorname{div} q \in \mathcal{C}^1_c(\Omega, \mathbf{R}^d)$ and $\operatorname{div}^2 q \in \mathcal{C}_c(\Omega)$ are defined by

$$(\operatorname{div} q)_i = \sum_{j=1}^d \frac{\partial q_{ij}}{\partial x_j}, \quad \operatorname{div}^2 q = \sum_{i=1}^d \frac{\partial^2 q_{ii}}{\partial x_i^2} + 2\sum_{i<j}\frac{\partial q_{ij}}{\partial x_i \partial x_j}$$

while the norms of $q \in \mathcal{C}_c(\Omega, S^{d\times d})$, $p \in \mathcal{C}_c(\Omega, \mathbb{R}^d)$ are given by

$$\|v\|_\infty = \sup_{x\in\Omega}\Big(\sum_{i=1}^d |q_{ii}(x)|^2 + 2\sum_{i<j}|q_{ij}(x)|^2\Big)^{1/2}, \quad \|p\|_\infty = \sup_{x\in\Omega}\Big(\sum_{i=1}^d |p_i(x)|^2\Big)^{1/2}. \tag{28}$$

For our purposes, we will utilize that TGV^2_α can be expressed as a minimum [6]:

$$\mathrm{TGV}^2_\alpha(u) = \min_{w\in\mathrm{BD}(\Omega)} \alpha_1\|\nabla u - w\|_{\mathcal{M}} + \alpha_0\|\mathcal{E}w\|_{\mathcal{M}} \tag{29}$$

where $\mathrm{BD}(\Omega)$ denotes the space of vector fields of *Bounded Deformation* [24] which is the set of vector fields whose weak symmetrized derivative $\mathcal{E}w = (\nabla w + \nabla w^T)/2$ is a matrix-valued Radon measure. Moreover, $\|\cdot\|_{\mathcal{M}}$ denotes the Radon norm for the corresponding vector-valued and matrix-valued Radon measures (the L^1-type norm for measures).

We are therefore concerned with discrete denoising problems of the type

$$\min_{\substack{u\in U\\ w\in V}} F(u) + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1, \qquad F(u) = \begin{cases}\frac{1}{2}\|u-f\|_2^2 & \text{for } L^2\text{-denoising},\\ \|u-f\|_1 & \text{for } L^1\text{-denoising},\end{cases} \tag{30}$$

with the notation of Section 4 and $\mathcal{E}$ being a discretization of the symmetrized derivative which will be described in the following. For this purpose, consider the spaces

$$U = \{u : \Omega \to \mathbf{R}\}, \quad V = \{u : \Omega \to \mathbf{R}^2\}, \quad W = \{u : \Omega \to S^{2\times 2}\}.$$

where U and V have already been introduced in Section 4. We identify elements in $q \in W$ with $q = (q^1, q^2, q^3) \in U^3$ and employ the scalar product

$$\langle q, q' \rangle_W = \langle q^1, (q')^1\rangle_U + \langle q^2, (q')^2\rangle_U + 2\langle q^3, (q')^3\rangle_U.$$

Utilizing forward differences, the symmetrized derivative can be defined as follows

$$\mathcal{E}w = \begin{bmatrix} \partial_x^+ w^1 & \frac{1}{2}(\partial_y^+ w^1 + \partial_x^+ w^2) \\ \frac{1}{2}(\partial_y^+ w^1 + \partial_x^+ w^2) & \partial_y^+ w^2 \end{bmatrix} = \begin{bmatrix} \partial_x^+ w^1 \\ \partial_y^+ w^2 \\ \frac{1}{2}(\partial_y^+ w^1 + \partial_x^+ w^2) \end{bmatrix}$$

where the second equation has to be understood in terms of the identification $W = U^3$. Consequently, the negative adjoint realizes a discrete negative divergence operator according to $\langle \mathcal{E}w, q\rangle_W = -\langle w, \operatorname{div} q\rangle_V$ for all $w \in V$, $q \in W$, leading to

$$\operatorname{div} q = \begin{bmatrix} \partial_x^- q^1 + \partial_y^- q^3 \\ \partial_x^- q^3 + \partial_y^- q_2 \end{bmatrix}.$$

With $\Delta w = (\Delta w^1, \Delta w^2)$, we can express the operator $\mathcal{E}^*\mathcal{E}w = -\frac{1}{2}\Delta w - \frac{1}{2}\nabla^- \operatorname{div}^+ w$ where div^+ is the discrete divergence using forward differences and ∇^- is the discrete gradient using backward differences. Finally, we need to define meaningful norms: For $1 \leq t < \infty$ and $q \in W$, let

$$\|q\|_t = \Big(\sum_{(i,j)\in\Omega} \big((q^1_{i,j})^2 + (q^2_{i,j})^2 + 2(q^3_{i,j})^2\big)^{t/2}\Big)^{1/t}, \quad \|q\|_\infty = \max_{(i,j)\in\Omega} \sqrt{(q^1_{i,j})^2 + (q^2_{i,j})^2 + 2(q^3_{i,j})^2}.$$

With these definitions, we are able to give a meaning to the discrete denoising problems (30). Proceeding in the lines of Subsection 4.1, we derive the preconditioned Douglas-Rachford algorithms separately for L^2- and L^1-denoising and develop a suitable preconditioner afterwards. For TGV^2_α, the latter turns out to require more effort than for the TV case.

5.1 L^2- and L^1-Denoising

We follow the lines of, e.g., [3] and write (30) as a saddle-point problem according to

$$\min_{\substack{u\in U\\ w\in V}} \max_{\substack{p\in V\\ q\in W}} \langle \nabla u - w, p\rangle_V + \langle \mathcal{E}w, q\rangle_W + F(u) - \mathcal{I}_{\{\|p\|_\infty \leq \alpha_1\}}(p) - \mathcal{I}_{\{\|q\|_\infty \leq \alpha_0\}}(q). \tag{31}$$

Therefore, $X = U \times V$, $Y = V \times W$ and, denoting $x = (u, w)$, $y = (p, q)$,

$$\mathcal{K} = \begin{bmatrix} \nabla & -I \\ 0 & \mathcal{E} \end{bmatrix}, \quad \mathcal{F}(u, w) = F(u), \quad \mathcal{G}(p, q) = \mathcal{I}_{\{\|p\|_\infty \leq \alpha_1\}}(p) + \mathcal{I}_{\{\|q\|_\infty \leq \alpha_0\}}(q) \tag{32}$$

for which the resolvent of $\partial\mathcal{G}$ is easy to compute: It is given by (18), i.e., the pair $(\mathcal{P}_{\alpha_1}(p), \mathcal{P}_{\alpha_0}(q))$ where

$$\mathcal{P}_{\alpha_1}(p) = \frac{p}{\max(1, \frac{|p|}{\alpha_1})}, \qquad \mathcal{P}_{\alpha_0}(q) = \frac{q}{\max(1, \frac{|q|}{\alpha_0})} \tag{33}$$

and $|p| = \sqrt{(p^1)^2 + (p^2)^2}$ as well as $|q| = \sqrt{(q^1)^2 + (q^2)^2 + 2(q^3)^2}$ for $p \in V$ and $q \in W$ (also compare with (18)). In case of L^2-denoising, (PDRQ) will be employed with

$$Q = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \qquad f_0 = \begin{bmatrix} f \\ 0 \end{bmatrix}.$$

In case of L^1-denoising, (PDR) will be used with a scaling of $\mathcal{K}$ by a positive factor $\tau > 0$ as already utilized in Section 4. This leads to the following operators to be preconditioned:

$$T_{L^2\text{-}\mathrm{TGV}} = \begin{bmatrix} \sigma I - \sigma^2 \Delta & \sigma^2 \operatorname{div} \\ -\sigma^2 \nabla & \sigma^2 (I - \operatorname{div} \mathcal{E}) \end{bmatrix}, \quad T_{L^1\text{-}\mathrm{TGV}} = \begin{bmatrix} I - (\sigma\tau)^2 \Delta & (\sigma\tau)^2 \operatorname{div} \\ -(\sigma\tau)^2 \nabla & (1 + (\sigma\tau)^2) I - (\sigma\tau)^2 \operatorname{div} \mathcal{E} \end{bmatrix}. \tag{34}$$

5.2 The Preconditioner

Choosing $\lambda_u, \lambda_w, \mu > 0$, our goal is now to find preconditioners for

$$T = \begin{bmatrix} \lambda_u I - \mu \Delta & \mu \operatorname{div} \\ -\mu \nabla & \lambda_w I - \mu \operatorname{div} \nabla \end{bmatrix}$$

which corresponds to a discrete solution of the system of partial differential equations

$$\begin{cases} \lambda_u u - \mu \Delta u + \mu \operatorname{div} w = b_u, \\ -\mu \nabla u + \lambda_w w - \mu \operatorname{div} \mathcal{E} w = b_w \qquad \text{in } \Omega, \\ \dfrac{\partial u}{\partial \nu} = 0, \qquad (\nabla w + \nabla w^T)\nu = 0 \qquad \text{on } \partial\Omega. \end{cases} \tag{35}$$

By definition, feasible preconditioners require T to be invertible, which can easily be verified.

Lemma 3 *If $\lambda_w \geq \mu$, then T is a positive definite operator.*

Proof For $x = (u, w)$ we have

$$\begin{aligned} \langle Tx, x \rangle_X &= \lambda_u \|u\|_2^2 + \mu \|\nabla u\|_2^2 - \mu \langle u, \nabla^* w \rangle - \mu \langle w, \nabla u \rangle + \lambda_w \|w\|_2^2 + \mu \|\mathcal{E} w\|_2^2 \\ &\geq \mu \|\nabla u - w\|_2^2 + \lambda_u \|u\|_2^2. \end{aligned}$$

It can be seen that if $\langle Tx, x \rangle_X = 0$, then $u = 0$ and $w = 0$ leading to $x = 0$. □

Like in the TV case, we would like to employ a suitable Red-Black Gauss-Seidel method. However, as $-\operatorname{div} \mathcal{E} w = -\frac{1}{2} \Delta w - \frac{1}{2} \nabla^- \operatorname{div}^+ w$ contains discrete mixed derivatives, the corresponding finite-difference stencil will no longer depend only the 4-neighborhood of a pixel but also involve points from the 8-neighborhood. This problem can be circumvented by performing a suitable Red-Black Gauss-Seidel preconditioning on the equation $T'(u, w) = (b_u, w_u)$ according to

$$\underbrace{\begin{bmatrix} \lambda_u I - \mu \Delta & \mu \operatorname{div} \\ -\mu \nabla & \lambda_w I - \mu \Delta \end{bmatrix}}_{=T'} \begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} b_u \\ b_w \end{bmatrix} \quad \text{or} \quad \begin{cases} \lambda_u u - \mu \Delta + \mu \operatorname{div} w = b_u, \\ -\mu \nabla u + \lambda_w w - \mu \Delta w = b_w \qquad \text{in } \Omega, \\ \dfrac{\partial u}{\partial \nu} = 0, \quad \dfrac{\partial w}{\partial \nu} = 0 \qquad \text{on } \partial\Omega. \end{cases} \tag{36}$$

This is justified by the following lemma.

Lemma 4 *It holds that $T' - T \geq 0$.*

Proof For $x = (u, w)$ one computes

$$\langle (T' - T)x, x\rangle_X = \mu\|\nabla w\|^2 - \mu\|\mathcal{E}w\|^2 = \mu\big(\|\partial_y^+ w^1\|^2 + \|\partial_x^+ w^2\|^2 - 2\underbrace{\|\tfrac{1}{2}(\partial_y^+ w^1 + \partial_x^+ w^2)\|^2}_{\leq \frac{1}{2}\|\partial_y^+ w^1\|^2 + \frac{1}{2}\|\partial_x^+ w^2\|^2}\big) \geq 0.$$

□

It is then clear that feasible preconditioners M for T' are also feasible for T. Now, as (36) is a finite-difference equation for $x = (u, w^1, w^2)$, it can be represented, as in the case of (21), by finite-difference stencils with the difference that each entry corresponds to a 3×3 matrix. Here, each grid point is associated with a triple $x_{i,j} = (u_{i,j}, w^1_{i,j}, w^2_{i,j})$. Again, we have to distinguish whether a grid point is in the interior, on a boundary edge or a corner. Adopting the notation of Table 4, the center element can be computed to be

$$\underbrace{\begin{bmatrix} \lambda_u + 2\mu & \mu & \mu \\ \mu & \lambda_w + 2\mu & 0 \\ \mu & 0 & \lambda_w + 2\mu \end{bmatrix}}_{=\mathcal{C}_{NW}} \underbrace{\begin{bmatrix} \lambda_u + 3\mu & \mu & \mu \\ \mu & \lambda_w + 3\mu & 0 \\ \mu & 0 & \lambda_w + 3\mu \end{bmatrix}}_{=\mathcal{C}_{N}} \underbrace{\begin{bmatrix} \lambda_u + 2\mu & 0 & \mu \\ 0 & \lambda_w + 2\mu & 0 \\ \mu & 0 & \lambda_w + 2\mu \end{bmatrix}}_{=\mathcal{C}_{NE}}$$

$$\underbrace{\begin{bmatrix} \lambda_u + 3\mu & \mu & \mu \\ \mu & \lambda_w + 3\mu & 0 \\ \mu & 0 & \lambda_w + 3\mu \end{bmatrix}}_{=\mathcal{C}_{W}} \underbrace{\begin{bmatrix} \lambda_u + 4\mu & \mu & \mu \\ \mu & \lambda_w + 4\mu & 0 \\ \mu & 0 & \lambda_w + 4\mu \end{bmatrix}}_{=\mathcal{C}_{O}} \underbrace{\begin{bmatrix} \lambda_u + 3\mu & 0 & \mu \\ 0 & \lambda_w + 3\mu & 0 \\ \mu & 0 & \lambda_w + 3\mu \end{bmatrix}}_{=\mathcal{C}_{E}}$$

$$\underbrace{\begin{bmatrix} \lambda_u + 2\mu & \mu & 0 \\ \mu & \lambda_w + 2\mu & 0 \\ 0 & 0 & \lambda_w + 2\mu \end{bmatrix}}_{=\mathcal{C}_{SW}} \underbrace{\begin{bmatrix} \lambda_u + 3\mu & \mu & 0 \\ \mu & \lambda_w + 3\mu & 0 \\ 0 & 0 & \lambda_w + 3\mu \end{bmatrix}}_{=\mathcal{C}_{S}} \underbrace{\begin{bmatrix} \lambda_u + 2\mu & 0 & 0 \\ 0 & \lambda_w + 2\mu & 0 \\ 0 & 0 & \lambda_w + 2\mu \end{bmatrix}}_{=\mathcal{C}_{SE}}$$

The off-center entries are all the same for each pixel and only depend on the position relative to the center. Denoting

$$\mathcal{R} = \begin{bmatrix} -\mu & 0 & 0 \\ -\mu & -\mu & 0 \\ 0 & 0 & -\mu \end{bmatrix}, \qquad \mathcal{U} = \begin{bmatrix} -\mu & 0 & -\mu \\ 0 & -\mu & 0 \\ 0 & 0 & -\mu \end{bmatrix}.$$

and their adjoints by $\mathcal{R}^*$ and $\mathcal{U}^*$, respectively, the block stencil for (36) is given by

$$\begin{aligned}
&\underbrace{\begin{bmatrix} \boxed{\mathcal{C}_{NW}} & \mathcal{R} \\ \mathcal{U}^* & \end{bmatrix}}_{=\mathcal{S}_{NW}} \underbrace{\begin{bmatrix} \mathcal{R}^* & \boxed{\mathcal{C}_{N}} & \mathcal{R} \\ & \mathcal{U}^* & \end{bmatrix}}_{=\mathcal{S}_{N}} \underbrace{\begin{bmatrix} \mathcal{R}^* & \boxed{\mathcal{C}_{NE}} \\ & \mathcal{U}^* \end{bmatrix}}_{=\mathcal{S}_{NE}} \\
&\underbrace{\begin{bmatrix} \mathcal{U} & \\ \boxed{\mathcal{C}_{W}} & \mathcal{R} \\ \mathcal{U}^* & \end{bmatrix}}_{=\mathcal{S}_{W}} \underbrace{\begin{bmatrix} & \mathcal{U} & \\ \mathcal{R}^* & \boxed{\mathcal{C}_{O}} & \mathcal{R} \\ & \mathcal{U}^* & \end{bmatrix}}_{=\mathcal{S}_{O}} \underbrace{\begin{bmatrix} & \mathcal{U} \\ \mathcal{R}^* & \boxed{\mathcal{C}_{E}} \\ & \mathcal{U}^* \end{bmatrix}}_{=\mathcal{S}_{E}} \\
&\underbrace{\begin{bmatrix} \mathcal{U} & \\ \boxed{\mathcal{C}_{SW}} & \mathcal{R} \end{bmatrix}}_{=\mathcal{S}_{SW}} \underbrace{\begin{bmatrix} & \mathcal{U} & \\ \mathcal{R}^* & \boxed{\mathcal{C}_{S}} & \mathcal{R} \end{bmatrix}}_{=\mathcal{S}_{S}} \underbrace{\begin{bmatrix} & \mathcal{U} \\ \mathcal{R}^* & \boxed{\mathcal{C}_{SE}} \end{bmatrix}}_{=\mathcal{S}_{SE}}
\end{aligned} \tag{37}$$

adopting again the notation for the different types of grid points from Table 4 and highlighting the center element. Based on this, a block Red-Black symmetric Gauss-Seidel iteration can be performed, needing only forward evaluation of $\mathcal{U}, \mathcal{U}^*, \mathcal{R}, \mathcal{R}^*$ and the inverses of the center matrices $\mathcal{C}_{\{\cdot\}}$ which can be computed explicitly. Denoting by $b_{i,j} = \big((b_u)_{i,j}, (b^1_w)_{i,j}, (b^2_w)_{i,j}\big)$ for $(i, j) \in \Omega$, $\mathcal{C}_{i,j}$ the corresponding center element for the grid point (i, j), introducing x_{red} and x_{black} on Ω_{red}

and Ω_{black}, respectively and extending them outside their domain by 0, n-fold application of the block Red-Black symmetric Gauss-Seidel method, denoted $\mathrm{BSRBGS}^n_{\lambda_u,\lambda_w,\mu}$, corresponds to

$$\begin{cases} \mathrm{BSRBGS}^n_{\lambda_u,\lambda_w,\mu}(x^k,b^k) = (x^{k+1}_{\mathrm{red}}, x^{k+1}_{\mathrm{black}}) \\ (x^{k+(\nu+1/2)/n}_{\mathrm{red}})_{i,j} = \mathcal{C}^{-1}_{i,j}\big(b_{i,j} - \mathcal{R}^*(x^{k+\nu/n}_{\mathrm{black}})_{i-1,j} - \mathcal{R}(x^{k+\nu/n}_{\mathrm{black}})_{i+1,j} \\ \qquad - \mathcal{U}(x^{k+\nu/n}_{\mathrm{black}})_{i,j-1} - \mathcal{U}^*(x^{k+\nu/n}_{\mathrm{black}})_{i,j+1}\big) \qquad (i,j)\in\Omega_{\mathrm{red}}, \\ (x^{k+(\nu+1)/n}_{\mathrm{black}})_{i,j} = \mathcal{C}^{-1}_{i,j}\big(b_{i,j} - \mathcal{R}^*(x^{k+(\nu+1/2)/n}_{\mathrm{red}})_{i-1,j} - \mathcal{R}(x^{k+(\nu+1/2)/n}_{\mathrm{red}})_{i+1,j} \\ \qquad - \mathcal{U}(x^{k+(\nu+1/2)/n}_{\mathrm{red}})_{i,j-1} - \mathcal{U}^*(x^{k+(\nu+1/2)/n}_{\mathrm{red}})_{i,j+1}\big) \qquad (i,j)\in\Omega_{\mathrm{black}}, \\ \qquad \nu = 0,\ldots,n-1, \\ (x^{k+1}_{\mathrm{red}})_{i,j} = \mathcal{C}^{-1}_{i,j}\big(b_{i,j} - \mathcal{R}^*(x^{k+1}_{\mathrm{black}})_{i-1,j} - \mathcal{R}(x^{k+1}_{\mathrm{black}})_{i+1,j} \\ \qquad - \mathcal{U}(x^{k+1}_{\mathrm{black}})_{i,j-1} - \mathcal{U}^*(x^{k+1}_{\mathrm{black}})_{i,j+1}\big). \end{cases} \tag{38}$$

Let us show feasibility. This is implied by the following general theorem about block symmetric Gauss-Seidel and block SSOR:

Proposition 3 *Let $N, m \geq$ and $T \in \mathbf{R}^{mN\times mN}$ be a symmetric, positive definite block matrix with blocks $T_{i,j} \in \mathbf{R}^{m\times m}$ for $i,j = 1,\ldots,N$ and $\omega \in \left]0,2\right[$. Then, block SSOR, i.e.,*

$$\begin{cases} x^{k+1/2}_i = (1-\omega)x^k_i + \omega T^{-1}_{i,i}\Big(b_i - \sum_{j<i} T_{i,j}x^{k+1/2}_j - \sum_{j>i} T_{i,j}x^k\Big) & i = 1,\ldots,N, \\ x^{k+1}_i = (1-\omega)x^{k+1/2}_i + \omega T^{-1}_{i,i}\Big(b_i - \sum_{j>i} T_{j,i}x^{k+1}_i - \sum_{j<i} T_{j,i}x^{k+1/2}_i\Big) & i = N,\ldots,1, \end{cases} \tag{39}$$

is well-defined and feasible for the solution of $Tx = b$.

Proof Denoting $T = D - E - E^*$ with $D = \mathrm{diag}(T_{1,1},\ldots,T_{N,N})$ the block diagonal matrix of T, and E, E^* the negative block lower and upper triangular matrix, respectively, the two steps in (39) correspond to the application of the preconditioner $M_0 = \frac{1}{\omega}D - E$ followed by M_0^*. By Proposition 1 (i), we only need to check the positive definiteness of $M_0 - \frac{1}{2}T$. Note that since T is positive definite, each $T_{i,i}$ is positive definite, hence $\langle T_{i,i}\xi,\xi\rangle \geq c\|\xi\|^2$ for $c > 0$ and each $\xi \in \mathbf{R}^m$, $i = 1,\ldots,N$. Hence, for $x \in \mathbf{R}^{mN}$,

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

Hence, by Proposition 1 (ii) and Lemmas 3 and 4, $\mathrm{BSRBGS}^n_{\lambda_u,\lambda_w,\mu}$ as described in (38) is feasible for the solution of (36) provided that $\lambda_w \geq \mu$. It can therefore been written as

$$x^{k+1} = x^k + M^{-1}(b^k - T'x^k)$$

with $M - T' \geq 0$. To apply the preconditioner for T, i.e., the discrete solution of (35), we have replace T' by T which means to correct the data according to

$$x^{k+1} = x^k + M^{-1}\big(b^k + (T' - T)x^k - T'x^k\big).$$

The corrected data then reads as

$$(b')^k = b^k + (T' - T)x^k = \begin{bmatrix} b^k_u \\ b^k_w \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \frac{\mu}{2}(\nabla^- \operatorname{div}^+ - \Delta) \end{bmatrix}\begin{bmatrix} u^k \\ w^k \end{bmatrix} = \begin{bmatrix} b^k_u \\ b^k_w + \frac{\mu}{2}(\nabla^- \operatorname{div}^+ - \Delta)w^k \end{bmatrix}. \tag{40}$$

In the algorithms, the block symmetric Gauss-Seidel preconditioner according to (38) is then applied to the data x^k and $(b')^k$.

PDRQ Objective: L^2-TGV denoising	$\min_{u\in U}\ \frac{1}{2}\|u-f\|_2^2 + \mathrm{TGV}_\alpha^2(u)$	
Initialization:	$(u^0, w^0, \bar p^0, \bar q^0) \in U\times V\times V\times W$ initial guess, $\sigma > 0$ step-size, $n \geq 1$ inner iterations for block symmetric Gauss-Seidel	
Iteration:	$b_u^k = \sigma(f + \operatorname{div}\bar p^k)$	
	$b_w^k = \sigma\big((\bar p^k + \operatorname{div}\bar q^k) + \frac{\sigma}{2}(\nabla^- \operatorname{div}^+ - \Delta)w^k\big)$	
	$(u^{k+1}, w^{k+1}) = \mathrm{BSRBGS}^n_{\sigma,\sigma^2,\sigma^2}(u^k, w^k, b_u^k, b_w^k)$	according to (38)
	$p^{k+1} = \bar p^k + \sigma(\nabla u^{k+1} - w^{k+1})$	
	$q^{k+1} = \bar q^k + \sigma\mathcal{E}w^{k+1}$	
	$p_{\mathrm{test}}^{k+1} = \mathcal{P}_{\alpha_1}(2p^{k+1} - \bar p^k)$	according to (33)
	$\bar p^{k+1} = \bar p^k + p_{\mathrm{test}}^{k+1} - p^{k+1}$	
	$q_{\mathrm{test}}^{k+1} = \mathcal{P}_{\alpha_0}(2q^{k+1} - \bar q^k)$	according to (33)
	$\bar q^{k+1} = \bar q^k + q_{\mathrm{test}}^{k+1} - q^{k+1}$	

Table 10: The preconditioned Douglas-Rachford iteration for L^2-TGV denoising.

Remark 5 An alternative to this approach would be to apply one block symmetric Gauss-Seidel step according to (38) for T and iterate n times. This results in a preconditioner which performs, n times, a correction according to (40) and a red $\to$ black $\to$ red update according to (38). Compared to (38), n more corrections and red updates have to be performed.

5.3 The Algorithms

The PDRQ algorithm for the solution of the L^2-TGV_α^2 denoising problem with the block symmetric Gauss-Seidel preconditioner as previously described is summarized in Table 10. Likewise, the PDR algorithm for L^1-TGV_α^2 denoising can be found in Table 11. For the latter, it can be noted that $\mathcal{F}$ does not depend on w, hence $(I + \sigma\partial\mathcal{F})^{-1}$ is the identity with respect to the w-component leading to $\bar w^k = w^k$ for each k if $w^0 = \bar w^0$. Hence, $\bar w^k$ can be replaced by w^k, a step which has already been incorporated in the algorithm. It is then clear by Theorem 1 that both algorithms converge. The u-part then gives a solution to the original denoising problem.

6 The Primal-Dual Gap as Stopping Criterion

For convex-concave saddle-point problems (2) which correspond to the solution of the primal and dual problem (3), the primal-dual gap according to (7) can be used to estimate both the primal and dual error in energy, see (8), in case a saddle point exists. Thus, if the PDR iteration is stopped if $\mathfrak{G}(x, y) \leq \varepsilon$ for some $\varepsilon > 0$ and some iterate (x, y), we can conclude that the pair (x, y) is ε-optimal for both the primal and dual problem. However, in general, $\mathfrak{G}(x, y) = \infty$ might occur for the iterates or might not converge to 0, hence, the "full" gap is not suitable as a stopping criterion.

Nevertheless, from Section 3 we know that restricted gaps associated with bounded sets and evaluated for the test sequences according to (11) converge, i.e., $\mathfrak{G}_{X_0\times Y_0}(x_{\mathrm{test}}^k, y_{\mathrm{test}}^k) \to 0$ as $k \to \infty$, see Remark 1, so these may always be used for a stopping criterion. If $\mathcal{F}$ and $\mathcal{G}$ are strongly coercive, the restricted gap associated with sufficiently large bounded sets coincides with the "full" gap, see Proposition 2. Thus, in this special case, $\mathfrak{G}(x_{\mathrm{test}}^k, y_{\mathrm{test}}^k) \to 0$ as $k \to \infty$ and hence, the primal-dual gap can be used as a stopping criterion.

PDR Objective:	L^1-TGV denoising	$\min_{u\in U}\ \|u-f\|_1 + \mathrm{TGV}^2_\alpha(u)$
Initialization:	$(u^0, w^0, \bar u^0, \bar p^0, \bar q^0) \in U \times V \times U \times V \times W$ initial guess, $\sigma > 0$ step-size, $\tau > 0$ gradient scaling, $n \geq 1$ inner iterations for block symmetric Gauss-Seidel	
Iteration:	$b_u^k = \bar u^k + \sigma\tau \operatorname{div} \bar p^k$	
	$b_w^k = w^k + \sigma\tau\big(\bar p^k + \operatorname{div} \bar q^k + \frac{\sigma\tau}{2}(\nabla^- \operatorname{div}^+ - \Delta) w^k\big)$	
	$(u^{k+1}, w^{k+1}) = \mathrm{BSRBGS}^n_{1,1+(\sigma\tau)^2,(\sigma\tau)^2}(u^k, w^k, b_u^k, b_w^k)$	according to (38)
	$p^{k+1} = \bar p^k + \sigma\tau(\nabla u^{k+1} - w^{k+1})$	
	$q^{k+1} = \bar q^k + \sigma\tau \mathcal{E} w^{k+1}$	
	$u_{\mathrm{test}}^{k+1} = \mathcal{S}_\sigma(2u^{k+1} - \bar u^k, f)$	according to (19)
	$\bar u^{k+1} = \bar u^k + u_{\mathrm{test}}^{k+1} - u^{k+1}$	
	$p_{\mathrm{test}}^{k+1} = \mathcal{P}_{\alpha_1/\tau}(2p^{k+1} - \bar p^k)$	according to (33)
	$\bar p^{k+1} = \bar p^k + p_{\mathrm{test}}^{k+1} - p^{k+1}$	
	$q_{\mathrm{test}}^{k+1} = \mathcal{P}_{\alpha_0/\tau}(2q^{k+1} - \bar q^k)$	according to (33)
	$\bar q^{k+1} = \bar q^k + q_{\mathrm{test}}^{k+1} - q^{k+1}$	

Table 11: The scaled preconditioned Douglas-Rachford iteration for L^1-TGV denoising.

In the following, we discuss situations in which the gap can be guaranteed to converge or in which the problem can be reformulated such that convergence of the gap may be established, namely the L^2-TV and L^2-TGV denoising problem, respectively.

6.1 L^2-TV Denoising

For the L^2-TV denoising problem, the primal-dual gap is well-known, see [9], for instance. It reads as

$$\mathfrak{G}_{L^2\text{-}\mathrm{TV}}(u,p) = \frac{\|u-f\|_2^2}{2} + \alpha\|\nabla u\|_1 + \frac{\|\operatorname{div} p + f\|_2^2}{2} - \frac{\|f\|_2^2}{2} + \mathcal{I}_{\{\|p\|_\infty \leq \alpha\}}(p) \tag{41}$$

which corresponds to (7) with the data $K = \nabla$, $\mathcal{F}(u) = \frac{1}{2}\|u-f\|_2^2$ and $\mathcal{G}(p) = \mathcal{I}_{\{\|p\|_\infty \leq \alpha\}}(p)$. Both $\mathcal{F}$ and $\mathcal{G}$ are strongly coercive, so one can conclude that $\mathfrak{G}_{L^2\text{-}\mathrm{TV}}$ converges for the test sequence $(u_{\mathrm{test}}^k, p_{\mathrm{test}}^k)$. Furthermore, $\mathfrak{G}_{L^2\text{-}\mathrm{TV}}$ is continuous on $U \times \{\|p\|_\infty \leq \alpha\}$ hence, as $u_{\mathrm{test}}^k - u^k \to 0$ as $k \to \infty$, we can also use $\mathfrak{G}_{L^2\text{-}\mathrm{TV}}(u^k, p_{\mathrm{test}}^k) \leq \varepsilon$ as stopping criterion. The latter may directly be incorporated into the algorithm in Table 5.

Moreover, recall that $\mathcal{F}$ is strongly convex with $\gamma = 1$ as $\partial\mathcal{F}(u) = \{u - f\}$ and

$$\frac{\|u'-f\|_2^2}{2} - \frac{\|u-f\|_2^2}{2} - \langle u - f, u' - u\rangle \geq \frac{\|u'-u\|_2^2}{2}.$$

Consequently, $\frac{1}{2}\|u^k - u^*\|_2^2 \leq \mathfrak{G}_{L^2\text{-}\mathrm{TV}}(u^k, p_{\mathrm{test}}^k)$, see Remark 2, meaning that the primal-dual gap can also be used to control the approximation error with respect to the solution in norm.

6.2 L^2-TGV Denoising

If $\alpha\,\mathrm{TV}$ is replaced by TGV_α^2, the data (32) has to be used for the saddle-point problem. One computes the primal-dual gap function as

$$\begin{aligned}\mathfrak{G}_{L^2\text{-}\mathrm{TGV}}(u,w,p,q) = \frac{\|u-f\|_2^2}{2} + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 + \frac{\|\operatorname{div} p + f\|_2^2}{2} - \frac{\|f\|_2^2}{2} \\ + \mathcal{I}_{\{0\}}(p + \operatorname{div} q) + \mathcal{I}_{\{\|p\|_\infty \le \alpha_1\}}(p) + \mathcal{I}_{\{\|q\|_\infty \le \alpha_0\}}(q).\end{aligned} \tag{42}$$

Here, the algorithm according to Table 10 generates $(p_{\mathrm{test}}^{k+1}, q_{\mathrm{test}}^{k+1})$ for which $\mathcal{I}_{\{0\}}(p_{\mathrm{test}}^{k+1} + \operatorname{div} q_{\mathrm{test}}^{k+1})$ might be infinite, rendering this primal-dual gap practically useless as stopping criterion.

To circumvent this problem, first suppose that we know that $\frac{1}{2}\|u-u^*\|_2^2 \le C_{L^2}$ for $u \in U$ and some u^* belonging to a solution pair (u^*, w^*) of (30). Then,

$$\frac{\|u-f\|_2^2}{2} - \frac{\|u^*-f\|_2^2}{2} = \langle u-f, u-u^*\rangle - \frac{\|u-u^*\|_2^2}{2} \le \sqrt{2C_{L^2}}\|u-f\|_2.$$

Thus, by optimality, we have for any w that

$$\begin{aligned}\alpha_1\|\nabla u^* - w^*\|_1 + \alpha_0\|\mathcal{E}w^*\|_1 &\le \Big(\frac{\|u-f\|_2^2}{2} - \frac{\|u^*-f\|_2^2}{2}\Big) + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \\ &\le \sqrt{2C_{L^2}}\|u-f\|_2 + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 = C_{\mathrm{TGV}}.\end{aligned} \tag{43}$$

In this case, we may restrict the problem to (u,w) satisfying $\alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \le C_{\mathrm{TGV}}$ with the minimizers staying the same:

Lemma 5 *Let, for a u^* associated with a solution pair (u^*, w^*) of* (30)*, $C_{\mathrm{TGV}} \ge 0$ such that $\alpha_1\|\nabla u^* - w^*\|_1 + \alpha_0\|\mathcal{E}w^*\|_1 \le C_{\mathrm{TGV}}$. Then, the solutions of*

$$\min_{\substack{u\in U\\ w\in V}} \frac{\|u-f\|_2^2}{2} + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \quad \text{subject to} \quad \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \le C_{\mathrm{TGV}} \tag{44}$$

coincide with the solutions of (30).

Proof As $u \mapsto \frac{1}{2}\|u-f\|_2^2$ is strictly convex, the u^* belonging to a solution pair (u^*, w^*) of the unrestricted problem (30) must be unique. By assumption, $\alpha_1\|\nabla u^* - w^*\|_1 + \alpha_0\|\mathcal{E}w^*\|_1 \le C_{\mathrm{TGV}}$ and since the optimal value is also unique, $\alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \le C_{\mathrm{TGV}}$ for any solution pair (u,w). Hence, any solution of the unrestricted problem is also a solution of the restricted problem (44).

Conversely, if (u^{**}, w^{**}) is a solution of the restricted problem, then for each $(u,w) \in U \times V$ we have, as (u^*, w^*) is a solution of (30) with $\alpha_1\|\nabla u^* - w^*\|_1 + \alpha_0\|\mathcal{E}w^*\|_1 \le C_{\mathrm{TGV}}$,

$$\begin{aligned}\frac{\|u^{**}-f\|_2^2}{2} + \alpha_1\|\nabla u^{**} - w^{**}\|_1 + \alpha_0\|\mathcal{E}w^{**}\|_1 &\le \frac{\|u^*-f\|_2^2}{2} + \alpha_1\|\nabla u^* - w^*\|_1 + \alpha_0\|\mathcal{E}w^*\|_1 \\ &\le \frac{\|u-f\|_2^2}{2} + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1.\end{aligned}$$

Hence, (u^{**}, w^{**}) is a solution of the unrestricted problem. □

The restricted problem (44) can be written in terms of (3) with data similar to (32):

$$\mathcal{K} = \begin{bmatrix} \nabla & -I \\ 0 & \mathcal{E} \end{bmatrix}, \quad \mathcal{F}(u,w) = \frac{\|u-f\|_2^2}{2}, \quad \mathcal{G}(p,q) = C_{\mathrm{TGV}} \max\Bigl(0, \max\Bigl(\frac{\|p\|_\infty}{\alpha_1}, \frac{\|q\|_\infty}{\alpha_0}\Bigr) - 1\Bigr). \tag{45}$$

This leads, in particular, to the dual problem

$$\max_{\substack{p\in V\\ q\in W}} \frac{\|f\|_2^2}{2} - \frac{\|\operatorname{div} p + f\|_2^2}{2} - C_{\mathrm{TGV}} \max\Bigl(0, \max\Bigl(\frac{\|p\|_\infty}{\alpha_1}, \frac{\|q\|_\infty}{\alpha_0}\Bigr) - 1\Bigr) \text{ subject to } p + \operatorname{div} q = 0 \tag{46}$$

hence, the associated primal-dual gap thus becomes, for $\alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \leq C_{\mathrm{TGV}}$,

$$\begin{aligned}\mathfrak{G}^{C_{\mathrm{TGV}}}_{L^2\text{-}\mathrm{TGV}}(u,w,p,q) = {}& \frac{\|u-f\|_2^2}{2} + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 + \frac{\|\operatorname{div} p + f\|_2^2}{2} - \frac{\|f\|_2^2}{2} \\ &+ \mathcal{I}_{\{0\}}(p + \operatorname{div} q) + C_{\mathrm{TGV}} \max\Bigl(0, \max\Bigl(\frac{\|p\|_\infty}{\alpha_1}, \frac{\|q\|_\infty}{\alpha_0}\Bigr) - 1\Bigr)\end{aligned}$$

and $\mathfrak{G}^{C_{\mathrm{TGV}}}_{L^2\text{-}\mathrm{TGV}}(u,w,p,q) = \infty$ otherwise. Plugging in $(u, w, -\operatorname{div} q, q)$ the modified gap simplifies, if $\alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \leq C_{\mathrm{TGV}}$ and $\|q\|_\infty \leq \alpha_0$, to

$$\begin{aligned}\mathfrak{G}^{C_{\mathrm{TGV}}}_{L^2\text{-}\mathrm{TGV}}(u,w,-\operatorname{div} q,q) = {}& \frac{\|u-f\|_2^2}{2} + \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \\ &+ \frac{\|f - \operatorname{div}^2 q\|_2^2}{2} - \frac{\|f\|_2^2}{2} + C_{\mathrm{TGV}} \max\Bigl(0, \frac{\|\operatorname{div} q\|_\infty}{\alpha_1} - 1\Bigr).\end{aligned} \tag{47}$$

Note that the PDRQ iteration yields $\|q^k_{\mathrm{test}}\|_\infty \leq \alpha_0$, hence, one can achieve the prerequisites for (47) by replacing C_{TGV} by $\alpha_1\|\nabla u^k - w^k\|_1 + \alpha_0\|\mathcal{E}w^k\|_1$ if necessary and evaluating at $(u^k, w^k, -\operatorname{div} q^k_{\mathrm{test}}, q^k_{\mathrm{test}})$. However, an estimate C_{TGV} still has to be obtained which can be done by (43) and an estimate C_{L^2} based on the following lemma.

Lemma 6 *Let (u^*, w^*) be a solution of (30) and $C_{\mathrm{TGV}} \geq 0$ such that $\alpha_1\|\nabla u^* - w^*\|_1 + \alpha_0\|\mathcal{E}w^*\|_1 \leq C_{\mathrm{TGV}}$. Then, for $(u,w,q) \in U \times V \times W$ with $\alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \leq C_{\mathrm{TGV}}$ and $\|q\|_\infty \leq \alpha_0$, we have*

$$\frac{\|u-u^*\|_2^2}{2} \leq \mathfrak{G}^{C_{\mathrm{TGV}}}_{L^2\text{-}\mathrm{TGV}}(u,w,-\operatorname{div} q,q)$$

for $\mathfrak{G}^{C_{\mathrm{TGV}}}_{L^2\text{-}\mathrm{TGV}}(u,w,-\operatorname{div} q,q)$ according to (47).

Proof By Lemma 5, the solutions of the restricted and unrestricted problems coincide, hence (u^*, w^*) is optimal for the restricted problem. A dual solution pair (p^*, q^*) for (46) obviously has to exist and we furthermore see that Fenchel-Rockafellar duality holds (since, e.g., $\mathcal{F}$ is continuous and $\mathcal{G}^*$ is continuous on a set with non-empty interior). Consequently, the primal-dual pair $(x^*, y^*) = ((u^*, w^*), (p^*, q^*))$ is a saddle point of the associated Lagrangian and optimality can equivalently be expressed by the primal-dual system $-\mathcal{K}^*y^* \in \partial\mathcal{F}(x^*)$ and $\mathcal{K}x^* \in \partial\mathcal{G}(y^*)$ with the data (45). The last inclusion is equivalent to $y^* \in \partial\mathcal{G}^*(Kx^*)$, thus, for $x = (u,w) \in U \times V$ and $y = (p,q) \in V \times W$, the associated primal-dual gap satisfies

$$\begin{aligned}\mathfrak{G}(x,y) &\geq \mathcal{F}(x^*) + \mathcal{G}^*(\mathcal{K}x^*) - \mathcal{F}(x^*) - \mathcal{G}(\mathcal{K}x^*) \\ &= \mathcal{F}(x) - \mathcal{F}(x^*) + \langle \mathcal{K}^*y^*, x - x^*\rangle + \mathcal{G}^*(\mathcal{K}x) - \mathcal{G}^*(\mathcal{K}x^*) - \langle y^*, \mathcal{K}x - \mathcal{K}x^*\rangle \\ &\geq \mathcal{F}(x) - \mathcal{F}(x^*) + \langle \mathcal{K}^*y^*, x - x^*\rangle.\end{aligned}$$

Now, $-\mathcal{K}^*y^* \in \partial\mathcal{F}(x^*)$ means $-\mathcal{K}^*y^* = (u^* - f, 0)$ as the subgradient is a singleton, see the definition of $\mathcal{F}$ in (45). Plugging in $y = (-\operatorname{div} q, q)$ for some $q \in W$ with $\|q\|_\infty \leq \alpha_0$ and assuming $\alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1 \leq C_{\mathrm{TGV}}$, the above estimate becomes, by virtue of (47) as well as bilinear computations,

$$\mathfrak{G}^{C_{\mathrm{TGV}}}_{L^2\text{-}\mathrm{TGV}}(u, w, \operatorname{div} q, q) \geq \frac{\|u-f\|_2^2}{2} - \frac{\|u^*-f\|_2^2}{2} + \langle u^* - f, u^* - u\rangle = \frac{\|u-u^*\|_2^2}{2}. \qquad \square$$

Consequently, once a bound C_{TGV} has been found, we can use Lemma 6 to estimate C_{L^2} and (43) to obtain a possibly tighter bound for C_{TGV}. Therefore, the PDRQ method may be amended as follows. Starting with $u^0 = 0$, $w^0 = 0$, $\bar{p}^0 = 0$ and $\bar{q}^0 = 0$, compute additionally

$$\begin{cases} \tilde{C}_0 = \alpha_1\|\nabla f\|_1 \\ (C_{\mathrm{TGV}})_{k+1} = \max\big(\tilde{C}_k, \alpha_1\|\nabla u^{k+1} - w^{k+1}\|_1 + \alpha_0\|\mathcal{E}w^{k+1}\|_1\big) \\ \mathfrak{G}_{k+1} = \mathfrak{G}^{(C_{\mathrm{TGV}})_{k+1}}_{L^2\text{-}\mathrm{TGV}}(u^{k+1}, w^{k+1}, -\operatorname{div} q^{k+1}_{\mathrm{test}}, q^{k+1}_{\mathrm{test}}) \\ \tilde{C}_{k+1} = \min\big(\tilde{C}_k, \sqrt{2\mathfrak{G}_{k+1}}\|u^{k+1} - f\|_2 + \alpha_1\|\nabla u^{k+1} - w^{k+1}\|_1 + \alpha_0\|\mathcal{E}w^{k+1}\|_1\big) \end{cases} \tag{48}$$

and stop as soon as $\mathfrak{G}_k \leq \varepsilon$ for some given tolerance ε.

Note that $\{\tilde{C}_k\}$ is non-increasing and since the iterates $(u^k, w^k, p^k_{\mathrm{test}}, q^k_{\mathrm{test}})$ converge to a primal-dual solution as $k \to \infty$, we can find a bound $C_{\mathrm{TGV}} \geq 0$ with $C^k_{\mathrm{TGV}} \leq C_{\mathrm{TGV}}$ for each k. With such a constant, $0 \leq \mathfrak{G}_k \leq \mathfrak{G}^{C_{\mathrm{TGV}}}_{L^2\text{-}\mathrm{TGV}}(u^k, w^k, -\operatorname{div} q^k_{\mathrm{test}}, q^k_{\mathrm{test}})$ for each k and as the right-hand side can be expressed by (47), continuity and the limit of the iterates being a solution imply convergence to zero. Consequently, $\mathfrak{G}_k \to 0$ as $k \to \infty$ meaning that the stopping criterion is eventually fulfilled.

7 Numerical Experiments

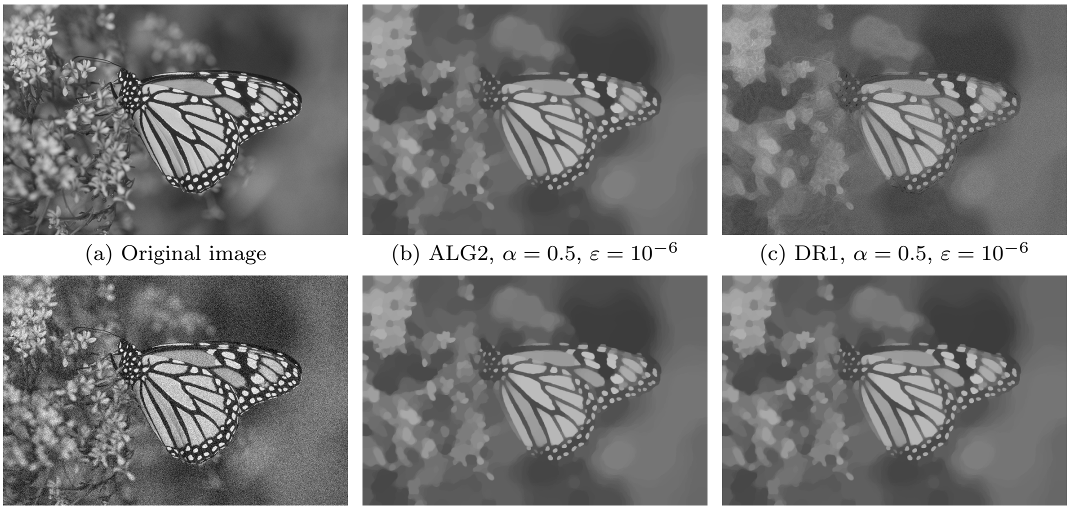
We implemented and performed numerical computations for each of the proposed algorithms using Matlab (MATLAB and Image Processing Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States) with 8 cores each at 3.40GHz. Comparisons with general-purpose primal-dual algorithms for the solution of (2) were performed, using the methods in [9] as a reference. Additionally, numerical tests with ADMM, split Bregman [12, 16] as well as the methods in [1, 2, 8, 27] were performed where applicable.

7.1 TV-Denoising Problems

7.1.1 L^2-Denoising

For L^2-TV denoising problems, besides the proposed PDRQ iteration, various first order primal-dual algorithm were chosen for comparison. Their parameter settings are as follows.

- ALG2: $\mathcal{O}(1/k^2)$ accelerated primal-dual algorithm introduced in [9] 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 [9] throughout this section.
- DR: The original Douglas-Rachford splitting method as in Remark 2.9 of [8]. Here, the discrete cosine transform (DCT) is used for calculating the inverse of the elliptic operator $I - \gamma^2\Delta$. The step size γ is chosen as 3 or 12 and $\lambda_n = 1$ which is corresponding to the PDR case with $M = T$.
- DR1/DR2: Douglas-Rachford type primal-dual methods (Algorithm 3.1 & 3.2, respectively) as in [1]. The parameters are chosen as $\sigma = \tau = 1.5\sqrt{2}$, $\lambda_n = 1.5$, $n \geq 1$ for DR1 and $\tau = \sigma = 1/\sqrt{8}$, and $\lambda_n = 1.5$, $n \geq 1$ for DR2.



(a) Original image (b) ALG2, $\alpha = 0.5$, $\varepsilon = 10^{-6}$ (c) DR1, $\alpha = 0.5$, $\varepsilon = 10^{-6}$

(d) Noisy image, noise level: 10% (e) VU, $\alpha = 0.5$, $\varepsilon = 10^{-6}$ (f) PDRQ, $\alpha = 0.5$, $\varepsilon = 10^{-6}$

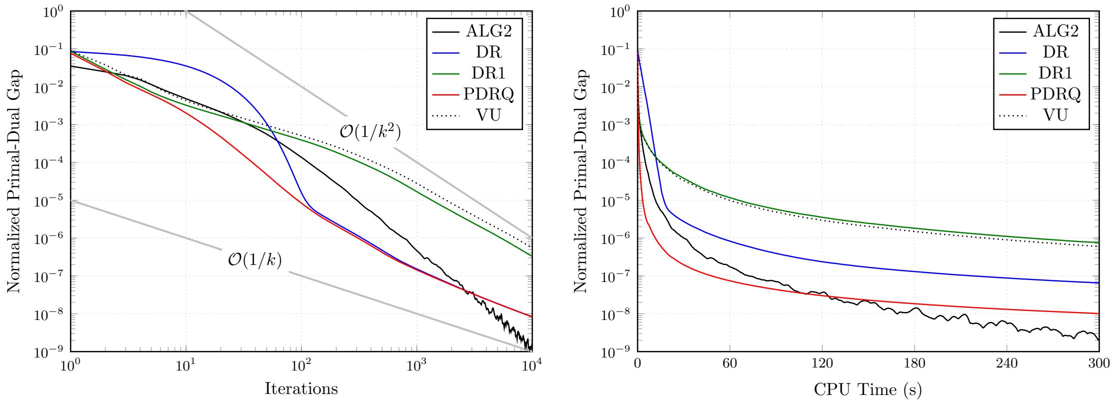
	$\alpha = 0.1$		$\alpha = 0.5$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
ALG2	19 (0.48s)	96 (2.14s)	115 (2.34s)	766 (15.86s)
DR	(3) 15 (2.32s)	(3) 69 (10.70s)	(12) 77 (11.84s)	(12) 334 (46.28s)
DR1	13 (0.60s)	269 (12.01s)	323 (14.29s)	5352 (241.39s)
DR2	17 (0.67s)	360 (12.51s)	430 (14.87s)	7135 (244.39s)
M+S	25 (1.14s)	553 (24.93s)	662 (29.09s)	10988 (558.75s)
PDRQ	(3,2) 9 (0.38s)	(3,2) 66 (2.36s)	(12,2) 38 (1.33s)	(12,2) 309 (10.78s)
VU	17 (0.51s)	360 (11.81s)	430 (13.31s)	7135 (220.99s)

Fig. 1: L^2-TV denoising: Results and performance. (a) shows the original 768×512 butterfly image while (d) is a noisy version which has been corrupted by 10% Gaussian noise. (b), (c), (e) and (f) are the denoised images given by the ALG2, DR1, VU and PDRQ algorithm, respectively, with regularization parameter α and such that the normalized primal-dual gap is less than 10^{-6}. The performance of the algorithms used for comparison can be seen in the table. Iteration number and computation time are shown in the format $k(t)$. Additionally, the parameters (σ, n) are given for PDRQ and the parameter (σ) is for given for DR.

– M+S: A primal-dual type algorithm for composite monotone inclusions as in [8]. Here the step size $\gamma_n = (1-\varepsilon)/\sqrt{8}$, $n \geq 1$, with $\varepsilon = 1/(10(\sqrt{8}+1))$.
– PDRQ: Preconditioned Douglas-Rachford method for pure quadratic-linear primal (or dual) functionals in (2). A symmetric Red-Black Gauss-Seidel preconditioner is employed (see Table 5). The notation PDRQ(σ, n) represents the step size σ and n iterations for the preconditioner.
– VU: A forward-backward splitting algorithm for monotone inclusions proposed in [27]. The parameters are chosen as $\tau = \sigma = 1/\sqrt{8}$, $\lambda_n = 1.5$, $n \geq 1$.

The stopping criterion is chosen as the normalized primal-dual gap $\mathfrak{G}_{L^2\text{-}\mathrm{TV}}/(N_x N_y)$ given by (41) and $N_x N_y$ being the number of pixels in the image, see also [5, 9].

For the computations, a butterfly image with size 768×512, see Figure 1(a), was corrupted by 10% Gaussian noise, see Figure 1(d) and denoising was performed with different regularization parameters and stopping tolerances, see Figure 1 for some of the results. Additionally, Figure 1



(a) Numerical convergence rate of normalized primal-dual gap compared with iteration number.

(b) Numerical convergence rate of normalized primal-dual gap compared with iteration time.

Fig. 2: L^2-TV denoising: Numerical convergence rates. The normalized primal-dual gap is compared in terms of iteration number and computation time for Figure 1(d) with $\alpha = 0.5$. Note the double-logarithmic and semi-logarithmic scale, respectively, which is used in the plots.

contains a table which shows the respective iteration numbers and CPU timings. It turns out that for this experiment, the methods ALG2 and PDRQ are the most competitive in terms of both iteration numbers and CPU time. In a moderate range of the primal-dual gap, PDRQ can be seen to be faster than ALG2, in particular for higher regularization parameters. The detailed comparison for $\alpha = 0.5$ in Figure 2 shows that this range roughly spans 10^{-3} and 10^{-7} both in terms of iteration number and computation time. Below 10^{-7}, the optimal asymptotic convergence speed of $\mathcal{O}(1/k^2)$ of ALG2 becomes visible which PDRQ apparently does not admit, making ALG2 faster there. Additionally, Figure 2(a) shows that the speed of PDRQ and DR method coincides in an asymptotic way when compared by iteration numbers since they are both variants of the Douglas-Rachford splitting method. However, Figure 2(b) tells that PDRQ is indeed an improvement of the original Douglas-Rachford method with proposed preconditioners with respect to computational speed.

7.1.2 L^1-Denoising

Using an L^1 discrepancy for TV-regularized denoising according to (16) is often beneficial to remove outliers as it is robust with respect to geometry perturbations. We performed experiments for the following algorithms.

- ADMM: Alternating direction method of multipliers [12, 16] based on part 4.4 of [12] with $\beta = 1.5$ and $\alpha = 30$ where β is the coefficient of the L^1 data fidelity term and $\alpha/2$ is the coefficient of the augmented term of ADMM. We use fast DCT methods to compute solutions of the linear elliptic equation in each iteration step.
- ALG1: $\mathcal{O}(1/k)$ primal-dual algorithm introduced in [9] with constant step sizes, the dual step size $\tau_0 = 0.02$, $\tau\sigma L^2 = 1$, $\theta = 1$ with $L = \sqrt{8}$ as proposed in [9].
- PDR: Preconditioned Douglas-Rachford method with symmetric Red-Black Gauss-Seidel preconditioner and scaled gradient according to Table 6. Here, $\sigma = 0.1$, $\tau = 1/\sigma$ and $n = 1$.

(a) Original image

(b) PDR, $\alpha = 2/3$, $\varepsilon = 10^{-4}$

(c) PDRQ, $\alpha = 2/3$, $\varepsilon = 10^{-5}$

(d) Noisy image, noise level: 25%

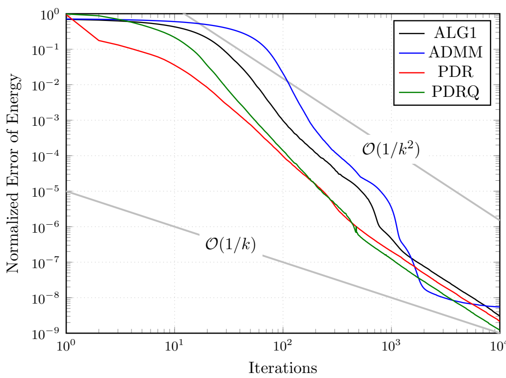
(e) ALG1, $\alpha = 2/3$, $\varepsilon = 10^{-4}$

(f) ALG1, $\alpha = 2/3$, $\varepsilon = 10^{-5}$

Fig. 3: L^1-TV denoising: Results. (a) shows the 256×256 input Lena image while (d) is a noisy image for which 25% of the pixels are corrupted by salt-and-pepper noise. (b) and (c) are the denoised images by the PDRQ algorithm with regularization parameter $\alpha = 2/3$ and different tolerances ε for the normalized error of energy. (e) and (f) are the respective outcomes of Chambolle-Pock's ALG1 algorithm. Note that the solutions are not unique.

	$\alpha = 2/3$		
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$
ADMM	250 (2.21s)	617 (5.73s)	903 (8.09s)
ALG1	214 (1.01s)	528 (2.55s)	779 (3.41s)
PDR	100 (0.62s)	235 (1.64s)	475 (3.12s)
PDRQ	111 (0.58s)	227 (1.23s)	450 (2.58s)

(a) Comparison of iteration numbers and CPU time cost.



(b) Numerical convergence rate compared with iteration number.

Fig. 4: L^1-TV denoising: Performance table and numerical convergence rates. (a) shows the performance of the four algorithms in terms of iteration numbers and CPU times. (b) depicts the numerical performance rate in a double-logarithmic scale.

- PDRQ: Preconditioned Douglas-Rachford method for the pure quadratic case with symmetric Red-Black Gauss-Seidel preconditioner according to Table 7. Here, $\sigma = 16$ and $n = 1$.

As a practical primal-dual gap is not available, the stopping criterion is chosen as the normalized error of the primal energy $(E^k - E^*)/E^*$, i.e., E^k is the functional value in (16) for u^k and E^* is obtained by a large number of iterations, we use 10^5 in this article. It is chosen as the minimal value with respect to all algorithms used for comparison.

Computations were performed on the 256×256 Lena image (Figure 3(a)) corrupted by 25% salt-and-pepper noise (Figure 3(d)) with regularization parameter $\alpha = 2/3$, see Figure 3 for the outcome for some of the algorithms. The table in Figure 4(a) shows the numerical evaluation with comparison of the iteration numbers and CPU time cost for the four algorithms. Figure 4(b) is the detailed comparison of iteration numbers. It could be seen that both PDR, PDRQ and ALG1 effectively admit the asymptotic convergence rate of $\mathcal{O}(1/k^2)$ which is good for L^1-type problems but not covered by the theory. In the example, both PDR and PDRQ admit the best performance, with PDRQ being faster for low tolerances.

7.2 TV-Deblurring Problems

7.2.1 L^2-Deblurring

We consider the model of L^2-TV deblurring as stated in (23). For comparison, we choose again ALG1 from [9], utilizing the following setup.

- ALG1: $\mathcal{O}(1/k)$ primal-dual algorithm introduced with constant steps, $\tau = 6$, $\tau\sigma L^2 = 1$, $\theta = 1$ with $L = \sqrt{8}$ as in [9].
- PDRQ: Preconditioned Douglas-Rachford method with scaled gradient and preconditioner $M = \sigma L^* L - (\sigma\tau)^2 \Delta_p$ according to Table 8. Here we choose, for step size, $\sigma = 15$ and the scaling parameter as $\tau = 1/\sigma$.

Again, stopping is performed based on the normalized error of energy.

Computations have been carried out for the 512×512 house image in Figure 5(a) which has been degraded by motion blur of approximately 40 pixels and Gaussian noise with standard deviation 0.01, see Figure 5(b). The regularization parameter α has been chosen as 0.0005, such that TV-regularization is effective, see Figure 5 for the results, in particular a comparison to the Wiener filter as well as a table with performance results. Figure 6 shows numerical convergence rates and CPU timings. Note that the proposed PDRQ outperforms ALG1 with appropriate step size. Also, both PDRQ and ALG1 converge fast, however, PDRQ needs less iterations.

7.2.2 L^1-Deblurring

Computations were performed for the TV-deblurring algorithm with L^1 discrepancy using the following algorithms and setup.

- ALG1: $\mathcal{O}(1/k)$ primal-dual algorithm with constant step sizes, dual step size $\tau = 0.05$, $\tau\sigma L^2 = 1$, $\theta = 1$ with $L = \sqrt{9}$ as proposed in [9].
- PDRQ: The scaled preconditioned Douglas-Rachford splitting method with preconditioner $M = \sigma^2 L^* L - (\sigma\tau)^2 \Delta_p$ according to Table 9, step-size $\sigma = 5$ and scaling parameter $\tau = 1/\sigma$.

The algorithm again stops once the normalized error of energy falls below a certain tolerance.

Computations were carried out for the 490×490 original image in Figure 7(a) subjected to motion blur of 30 pixels and corruption of 50% of the pixels by salt-and-pepper noise, see 7(d).

(a) Original image

(b) Degraded image

(c) Wiener filter

(d) PDRQ, $\alpha = 0.0005$, $\varepsilon = 10^{-4}$

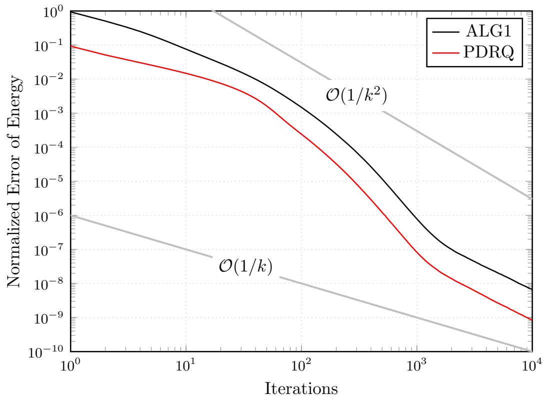
(e) PDRQ, $\alpha = 0.0005$, $\varepsilon = 10^{-5}$

(f) PDRQ, $\alpha = 0.0005$, $\varepsilon = 10^{-6}$

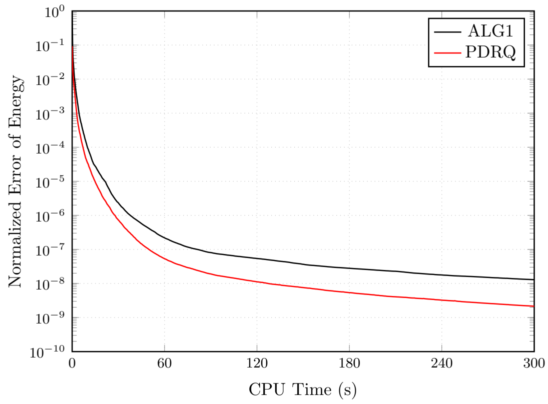
	$\alpha = 0.0005$		
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$
ALG1	270 (9.46s)	521 (17.87s)	939 (30.10s)
PDRQ	138 (7.30s)	285 (12.60s)	531 (23.40s)

Fig. 5: L^2-TV deblurring: Results and performance. (a) shows the 512×512 input image and (b) is the degraded image (≈ 40 pixels motion blur, Gaussian noise, standard deviation 0.01). (c) is the result of Wiener filtering and (d), (e) and (f) are the results of the PDRQ algorithm with normalized error of energy less than ε. The table shows the iteration number and CPU time needed for the normalized error of energy to get below ε.

With α chosen as 0.1, the PDRQ algorithm already produces visually appealing reconstructions after 50 iterations, see Figure 7 for an overview of the results. Performance values and plots can be found in Table 8 and represents the iteration numbers and iteration time cost for different stopping tolerances. It could also be seen that a significant speed-up with respect to time is achievable with PDRQ. The numerical convergence rate of PDRQ and ALG1 seems to be faster than $\mathcal{O}(1/k)$ and a little slower than $\mathcal{O}(1/k^2)$.



(a) Numerical convergence rate compared with iteration number.



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

Fig. 6: L^2-TV deblurring: Numerical convergence rates. The plots show a comparison of iteration numbers and CPU time. Again, a double-logarithmic and semi-logarithmic scale, respectively, is used in the plots.

7.3 TGV-Denoising Problems

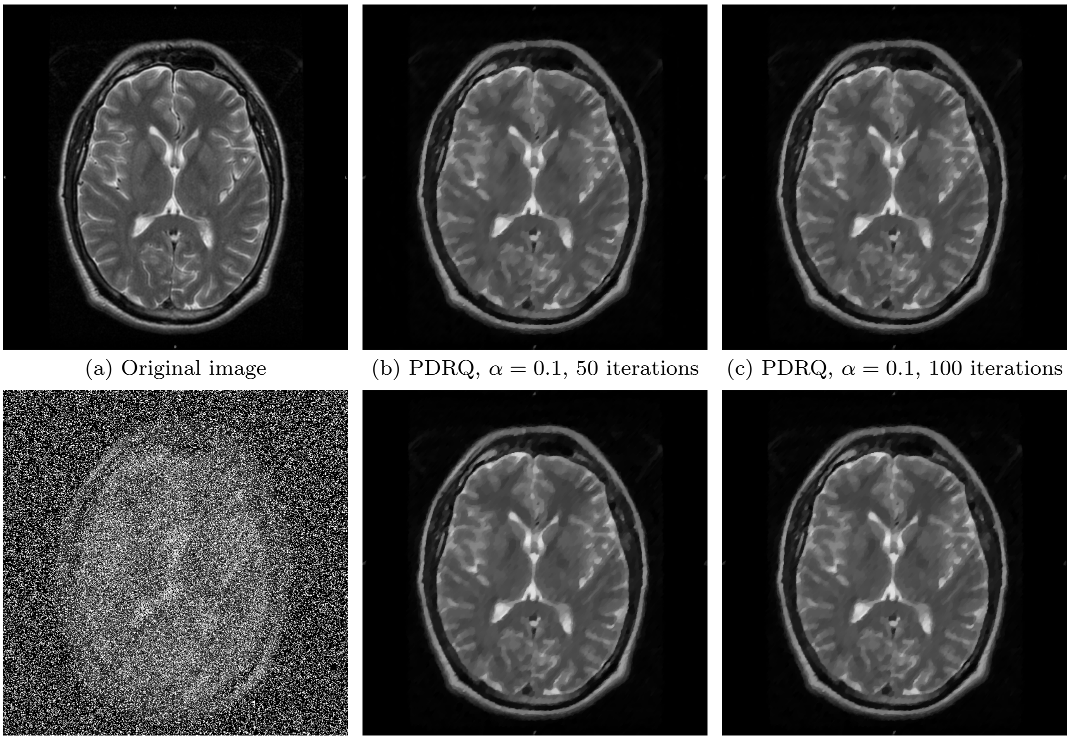
7.3.1 L^2-Denoising

The following algorithms and the parameter settings are used in our numerical experiments and comparison.

- ALG1: $\mathcal{O}(1/k)$ primal-dual algorithm discussed in [3, 9] with constant step sizes, the dual step size $\tau = 0.05$, $\tau\sigma L^2 = 1$, $\theta = 1$ with $L = \sqrt{12}$.
- FB/FBF: Forward-backward and forward-backward-forward type algorithms as discussed in [2]. The parameters are chosen as follows: $\sigma = \tau = 1/\sqrt{8}$, $\theta_1 = \gamma_1 = 1 - 1/50$, $\theta_2 = \gamma_2 = 1/\sqrt{8}$, $\lambda_n = 1.5$, $n \geq 1$ for FB and $\gamma_n = (2\sqrt{8}+1)/(2\sqrt{8}(\sqrt{8}+1))$, $n \geq 1$ for FBF.
- PDRQ: The preconditioned Douglas-Rachford splitting method for quadratic problem according to Table 10. Here, 1 inner iteration for the block symmetric Red-Black Gauss-Seidel preconditioner is performed. The step size is chosen as σ and the method is denoted by PDRQ(σ) in the evaluations.

Stopping criterion is the normalized modified primal-dual gap $\mathfrak{G}_k/(N_xN_y)$ according to (48).

We tested the algorithms on the 512×357 pixel image shown in Figure 9(a) subjected to 5% and 10% percent Gaussian noise, see Figure 9(d). Choosing $\alpha_0 = 2\alpha_1$ as in [7], we performed L^2-TGV denoising with parameters $\alpha_1 = 0.05$ and $\alpha_1 = 0.1$, respectively, depending on the noise level. Results and a numerical performance table are depicted in Figure 9. Figure 10 shows a detailed comparison with respected to iteration number and iteration time for the two algorithms. Here, ALG1 and PDRQ are competitive for all noise level and stopping tolerances. Furthermore, it may be seen that PDRQ is faster than ALG1, in particular if one aims at highly accurate numerical solutions. The plot also reveals that the normalized primal-dual gap admits oscillations. This numerical instability can be regarded as the price on has to pay for a provable upper bound for the distance to the minimizer as well as for the distance $\frac{1}{2N_xN_y}\|u^k - u^*\|_2^2$,



(a) Original image (b) PDRQ, $\alpha = 0.1$, 50 iterations (c) PDRQ, $\alpha = 0.1$, 100 iterations

(d) Degraded image (e) PDRQ, $\alpha = 0.1$, $\varepsilon = 10^{-4}$ (f) PDRQ, $\alpha = 0.1$, $\varepsilon = 10^{-5}$

Fig. 7: L^1-TV deblurring: Results. (a) shows the 490×490 input brain image which is gratefully taken from [13]. (d) is the degraded image with motion blur of 30 pixels and 50% of salt-and-pepper noise corruption. (b) and (c) are the results of by the PDRQ algorithm after 50 and 100 iterations. (e) and (f) show the reconstructions for the normalized error of energy less than ε.

7.3.2 L^1-Denoising

For the final denoising experiment, we again compare PDR and ALG1 with the following parameters.

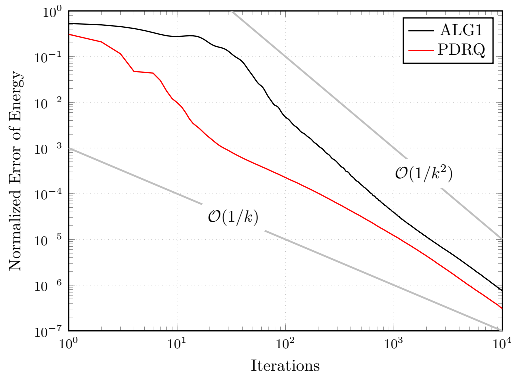
- ALG1: $\mathcal{O}(1/k)$ primal-dual algorithm discussed in [3, 9] with constant step sizes, dual step size $\tau = 0.01$, $\tau\sigma L^2 = 1$, $\theta = 1$ with $L = \sqrt{12}$.
- PDR: The scaled preconditioned Douglas-Rachford splitting method with symmetric block Red-Black Gauss-Seidel preconditioner according to Table 11. The step size is chosen as $\sigma = 0.1$, the scaling τ as $3/\sigma$ and 1 inner iteration for the preconditioner is performed.

For comparison, the iteration is stopped when the normalized error of energy $(E^k - E^*)/E^*$ falls below a tolerance where, again, E^* is precomputed.

The L^1-TGV denoising algorithms were tested on the 510×383 image in Figure 11(a) for which 30% of the pixels have been corrupted by salt-and-pepper noise (Figure 11(d)). With the regularization chosen as $\alpha_1 = 1.0$ and $\alpha_0 = 2\alpha_1$, results for the compared methods and different stopping tolerances are depicted in Figure 11 while a performance table and a plot showing numerical performance rates can be found in Figure 12. Here, it can be seen that PDR is significantly faster than ALG1 with respect to iteration number and iteration time with speed-ups ranging between 2 to 3.

	$\alpha = 0.1$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$
ALG1	626 (27.58s)	2164 (96.72s)
PDRQ	205 (9.67s)	1148 (52.80s)

(a) Comparison of iteration numbers and CPU time cost.



(b) Numerical convergence rate compared with iteration number.

Fig. 8: L^1-TV deblurring: Performance table and numerical convergence rates. (a) shows the performance of the compared algorithms in terms of iteration numbers and CPU times. (b) depicts the numerical performance rate in a double-logarithmic scale.

8 Discussion and Conclusions

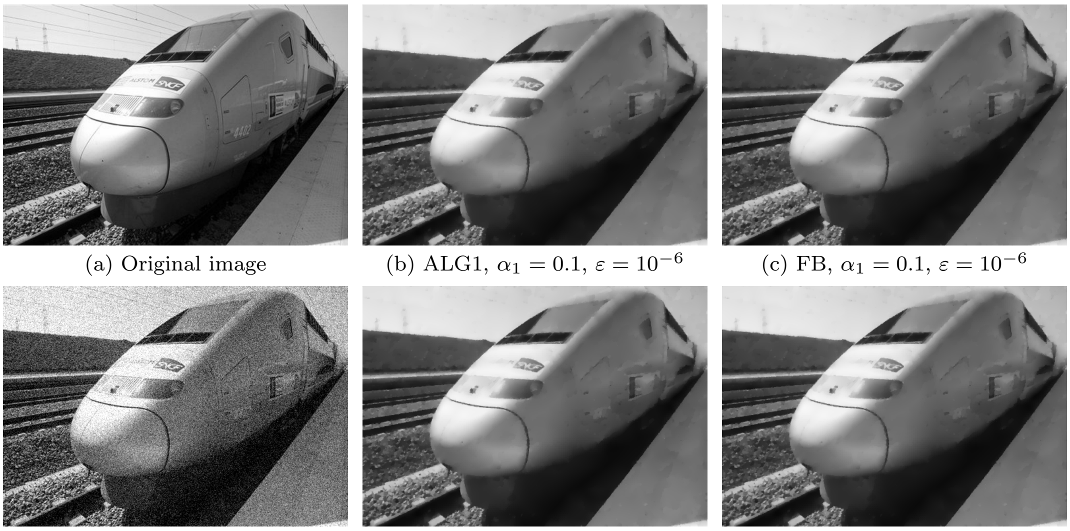
We derived an ergodic convergence rate of $\mathcal{O}(1/k)$ for restricted primal-dual gaps of the iterates generated by the preconditioned Douglas-Rachford method in [5]. Furthermore, these methods have been applied for various discrete variational imaging problems. Different preconditioning strategies for both TV and TGV regularized problems have been developed. While for TV-regularized denoising, the symmetric Red-Black Gauss-Seidel is an effective preconditioner, we proposed a respective symmetric block Red-Black Gauss-Seidel iteration for TGV-based denoising problems. In both cases, these preconditioners are tailored for the underlying differential structure which is an discrete elliptic equation in the TV-case and an discrete elliptic system of PDEs in the TGV-case. For L^2-TGV denoising problems, we moreover derived an effective primal-dual gap which gives a provable bound to the distance in functional values and which can also be used to estimate the L^2 norm to the optimal solution.

The numerical results show that PDR and its variants have the potential to bring out appealing benefits and fast algorithms, provided that efficient preconditioners are found. The algorithm are most beneficial in situations where accelerated primal-dual methods can not be applied which is the case for problems which are not strictly convex. This is the case for L^1 discrepancies as well as for TGV regularization. Indeed, significant speed-ups were achieved in these situations. Nevertheless, there are some open questions that need to be figured out. The possibility to accelerate the convergence speed of PDR and its variants to the optimal convergence speed of first-order methods, i.e., $\mathcal{O}(1/k^2)$, is, for instance, an important aspect to investigate in the future.

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. Boţ RI, Hendrich C (2013) A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. SIAM Journal on Optimization 23(4):2541–2565



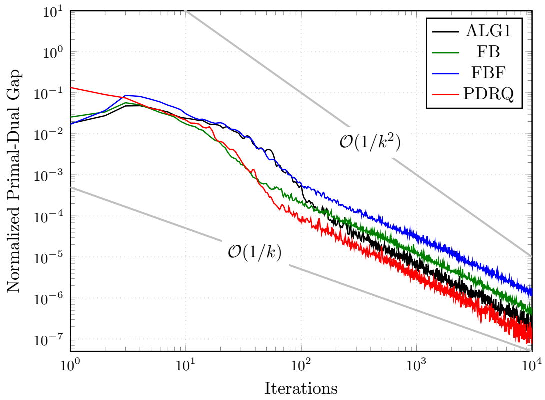
(a) Original image (b) ALG1, $\alpha_1 = 0.1$, $\varepsilon = 10^{-6}$ (c) FB, $\alpha_1 = 0.1$, $\varepsilon = 10^{-6}$

(d) Noisy image, noise level: 10% (e) FBF, $\alpha_1 = 0.1$, $\varepsilon = 10^{-6}$ (f) PDRQ, $\alpha_1 = 0.1$, $\varepsilon = 10^{-6}$

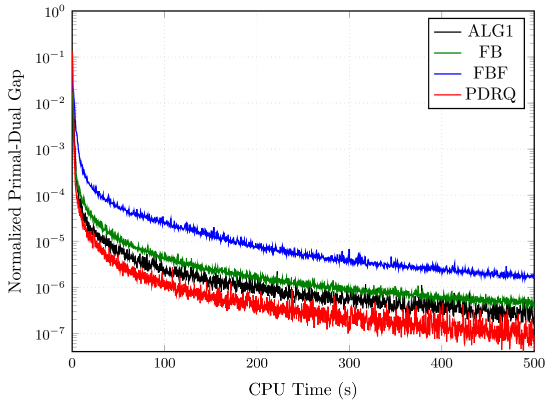
	$\alpha_1 = 0.05$, noise level: 5%		$\alpha_1 = 0.1$, noise level: 10%	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
ALG1	153 (4.66s)	1748 (72.16s)	171 (6.84s)	2900 (109.66s)
FB	79 (4.04s)	2288 (98.65s)	184 (7.99s)	5583 (253.86s)
FBF	165 (9.64s)	4764 (265.35s)	348 (19.27s)	10770 (795.56s)
PDRQ	(1.5,1) 63 (3.07s)	(1.5,1) 1419 (66.09s)	(3,1) 88 (4.10s)	(3,1) 1778 (82.30s)

Fig. 9: L^2-TGV denoising: Results and performance. (a) shows the 512×357 input train image and (d) is a noisy image which has been corrupted by 10% Gaussian noise. (b), (c), (e) and (f) are the denoised images by the ALG1, FB, FBF and PDRQ algorithm, respectively, with normalized primal-dual gap function less than ε. The table shows iteration number and computation time $k(t)$ for both algorithms and noise level of 5% and 10%. For PDRQ, step size and inner iterations for the preconditioner are denoted by (σ, n).

2. Boţ RI, Hendrich C (2013) Solving monotone inclusions involving parallel sums of linear composed maximal monotone operators. ArXiv:13063191
3. Bredies K (2014) Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty. In: Bruhn A, Pock T, Tai XC (eds) Efficient Algorithms for Global Optimization Methods in Computer Vision, Lecture Notes in Computer Science, Springer Berlin, pp 44–77
4. Bredies K, Holler M (2012) Artifact-free JPEG decompression with total generalized variation. Proceedings of VISAPP 2012 - International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications pp 12–21
5. Bredies K, Sun HP (2014) Preconditioned Douglas-Rachford splitting methods for convex-concave saddle-point problems. SFB-Report-2014-002, University of Graz
6. Bredies K, Valkonen T (2011) Inverse problems with second-order total generalized variation constraints. Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications, Singapore
7. Bredies K, Kunisch K, Pock T (2010) Total generalized variation. SIAM Journal on Imaging Sciences 3(3):492–526
8. Briceño-Arias L, Combettes PL (2011) A monotone+skew splitting model for composite monotone inclusions in duality. SIAM Journal on Optimization 21(4):1230–1250
9. Chambolle A, Pock T (2011) A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision 40(1):120–145
10. Eckstein J, Bertsekas DP (1992) On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming 55(1-3):293–318



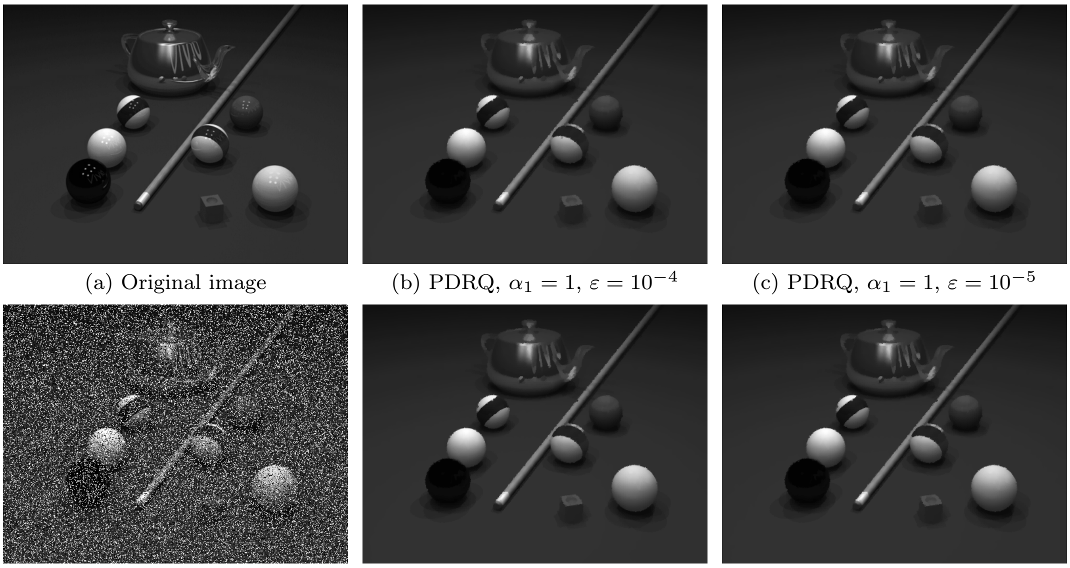
(a) Numerical convergence rate compared with iteration number.



(b) Numerical convergence rate compared with iteration time. Every algorithm runs for 500 seconds.

Fig. 10: L^2-TGV denoising: Numerical convergence rates. The plots show a comparison of iteration numbers and CPU time. Again, a double-logarithmic and semi-logarithmic scale, respectively, is used in the plots.

11. Ekeland I, Témam R (1999) Convex Analysis and Variational Problems. No. 1 in Classics in Applied Mathematics, Society for Industrial and Applied Mathematics
12. Esser E (2009) Applications of Lagrangian-based alternating direction methods and connections to split Bregman. Tech. rep., CAM report 09-2009, UCLA
13. Farrall A (2011) Human brain MRI. `http://www.anatomy.mvm.ed.ac.uk/museum/explore-anatomy.php`, [Online; accessed 07-February-2014]
14. Gabay D (1983) Applications of the method of multipliers to variational inequalities. In: Fortin M, Glowinski R (eds) Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, Studies in Mathematics and Its Applications, vol 15, Elsevier, pp 299 – 331
15. Gabay D, Mercier B (1976) A dual algorithm for the solution of nonlinear variational problems via finite elemxent approximation. Computers and Mathematics with Applications 2(1):17–40
16. Goldstein T, Osher S (2009) The split Bregman method for L^1-regularized problems. SIAM Journal on Imaging Sciences 2(2):323–343
17. Ito K, Kunisch K (2008) Lagrange Multiplier Approach to Variational Problems and Applications. Advances in Design and Control, Society for Industrial and Applied Mathematics
18. Knoll K, Bredies K, Pock T, Stollberger R (2011) Second order total generalized variation (TGV) for MRI. Magnetic Resonance in Medicine 65(2):480–491
19. Lions P, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis 16(6):964–979
20. Lorenz DA, Pock T (2014) An inertial forward-backward algorithm for monotone inclusions. ArXiv:14033522
21. Pock T, Chambolle A (2011) Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: Computer Vision (ICCV), 2011 IEEE International Conference on, pp 1762–1769
22. Rockafellar RT (1976) Monotone operators and the proximal point algorithm. SIAM Journal on Control and Optimization 14(5):877–898
23. Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60(1-4):259–268
24. Temam R (1985) Mathematical Problems in Plasticity. Bordas
25. Thomas JW (1999) Numerical Partial Differential Equations: Conservation Laws and Elliptic Equations, Texts in Applied Mathematics, vol 33. Springer New York
26. Uzawa H (1958) Iterative methods in concave programming. In: Arrow K, Hurwicz L, Uzawa H (eds) Studies in Linear and Nonlinear Programming, Stanford University Press, pp 154–165
27. Vũ BC (2013) A splitting algorithm for dual monotone inclusions involving cocoercive operators. Advances in Computational Mathematics 38(3):667–681



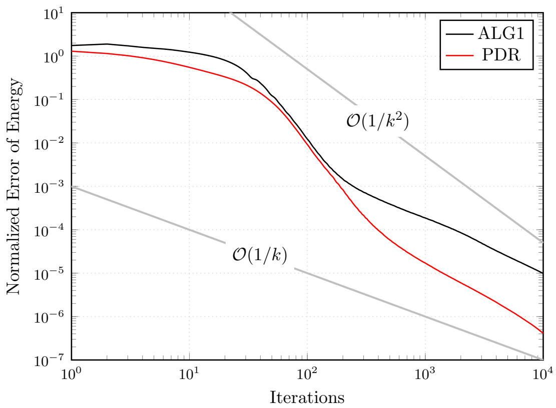
(a) Original image (b) PDRQ, $\alpha_1 = 1$, $\varepsilon = 10^{-4}$ (c) PDRQ, $\alpha_1 = 1$, $\varepsilon = 10^{-5}$

(d) Noisy image, noise level: 30% (e) ALG1, $\alpha_1 = 1$, $\varepsilon = 10^{-4}$ (f) ALG1, $\alpha_1 = 1$, $\varepsilon = 10^{-5}$

Fig. 11: L^1-TGV denoising: Results. (a) shows the 510×383 original image while (d) is its noisy version which has been corrupted by 30% salt-and-pepper noise. (b) and (c) are the denoised images by the PDR algorithm with the stopping criterion of the normalized error of primal energy less than ε. (e) and (f) are the respective outcomes of Chambolle-Pock's ALG1 algorithm.

	$\alpha_1 = 1.0$	
	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$
ALG1	1761 (49.10s)	9897 (264.17s)
PDR	396 (19.46s)	1408 (69.67s)

(a) Comparison of iteration numbers and CPU time cost.



(b) Numerical convergence rate compared with iteration number.

Fig. 12: L^1-TGV denoising: Performance table and numerical convergence rates. (a) shows the performance of the compared algorithms in terms of iteration numbers and CPU times. (b) depicts the numerical performance rate in a double-logarithmic scale.

28. Valkonen T, Bredies K, Knoll F (2013) Total generalized variation in diffusion tensor imaging. SIAM Journal on Imaging Sciences 6(1):487–525
29. Zhang XQ, Burger M, Osher S (2011) A unified primal-dual algorithm framework based on Bregman iteration. Journal of Scientific Computing 46(1):20–46