

# A Class of Quasi-Variational Inequalities for Adaptive Image Denoising and Decomposition

Frank Lenzen · Florian Becker · Jan Lellmann ·  
Stefania Petra · Christoph Schnörr

Received: date / Accepted: date

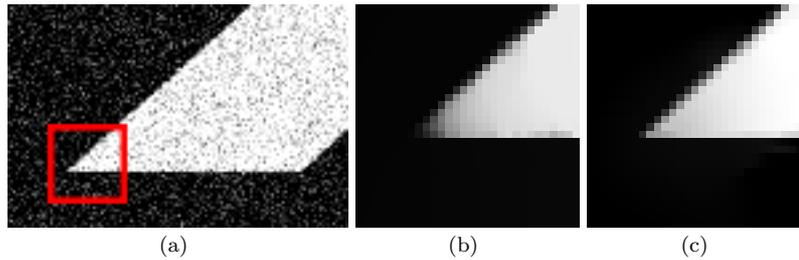
**Abstract** We introduce a class of adaptive non-smooth convex variational problems for image denoising in terms of a common data fitting term and a support functional as regularizer. Adaptivity is modeled by a set-valued mapping with closed, compact and convex values, that defines and steers the regularizer depending on the variational solution. This extension gives rise to a class of quasi-variational inequalities. We provide sufficient conditions for the existence of fixed points as solutions, and an algorithm based on solving a sequence of variational problems. Denoising experiments with spatial and spatio-temporal image data and an adaptive total variation regularizer illustrate our approach.

**Keywords** Quasi-variational inequalities · adaptive image denoising · total variation regularization · solution-dependent adaptivity

## 1 Introduction

Variational methods define the state of the art in image restoration [22]. The total variation measure [16, 21], for instance, has spurred tremendous research activities on non-smooth convex regularizers for image denoising and deblurring, together with numerous extensions, e.g., to image motion and to non-locally defined operators [4, 9, 10, 13, 14, 18]. The corresponding perspective on image signal representation by additive decomposition into geometric structure and texture and noise, respectively, led to novel ways to represent complex local image structure by simple high-dimensional signal geometries [2]. As a consequence, there has also been a renewed interest in large-scale convex programming [8] as a sound basis for implementing variational approaches on modern parallel hardware like GPUs.

Figure 1(b) displays a simple denoising example using the total variation measure. As is well known, this basic approach does an almost perfect job for piecewise smooth image signals as encountered in images of man-made scenes. Yet, for particular local signal geometries one often would like to adapt the denoising process as illustrated for a particular example in Fig. 1 (c). Such adaptive approaches are well-known for PDE based denoising methods [5, 11, 25]. In the field of variational denoising, there has been corresponding recent work [24] introducing valuable



**Fig. 1** Adaptive total variation (TV) filtering: (a) noisy test image, (b) filtering with the standard ROF model [21] (close-up of the region indicated in (a)), (c) adaptive TV filtering (close-up), see Sect. 5.2. The feasible set adapts to the local signal geometry and thus enables to remove noise more effectively.

ideas for adaptation. So far, a theoretical underpinning from both the variational viewpoint and concerning convex programming has been lacking in the image processing literature, however.

The objective of the present paper is to fill this gap. We show that the desired adaptivity can be conveniently modeled by resorting to the broader class of quasi-variational inequalities [6]. This leads to existence of solutions in terms of fixed points, and additionally to sound algorithms formulated in terms of a sequence of convex non-smooth variational problems, to each of which any established method can safely be applied.

While we motivate and introduce our approach using the total variation measure as the most prominent non-smooth convex regularizer, we sufficiently abstract from this particular approach by considering general support functions as regularizers. As a result, our approach is applicable to many of the extensions referred to above, and we expect that our result helps to provide solid ground for related existing and novel applications. In the second part of the present paper, we will exemplify the application of our approach to total variation based denoising of both spatial and spatio-temporal (motion) imaging scenarios. The latter, in particular, are highly relevant for industrial applications. A sketch of our approach was already announced in the conference paper [15]. We refer to [12] for a discussion on the continuous formulation of adaptive convex regularizers.

This paper is structured as follows. Section 2 introduces the variational formulation and the novel adaptive formulation in terms of a set-valued mapping that models the dependency of the set of feasible solutions on the solution itself. Such extensions of the variational formulation are covered by a corresponding class of quasi-variational inequalities detailed in Section 3, together with an existence theorem and sufficient conditions relevant to applications in image analysis. Section 4 specifies a basic algorithmic framework [17] and indicates aspects that are relevant for implementations. In Section 5 we rigorously work out the application of the variational framework to several major models for image denoising. Corresponding numerical experiments are presented in Section 6.

## 2 Problem

We introduce the basic variational formulation and specify the adaptivity illustrated by Fig. 1 and discussed in Section 1.

## 2.1 Primal Formulation

Let  $\Omega \subset \mathbb{R}^d$  be a bounded open domain with Lipschitz continuous boundary. The total variation of  $u \in L^2(\Omega)$  is defined by [27]

$$\text{TV}(u) := \sup \{ \langle u, \text{div } p \rangle_{L^2} : p \in C_c^1(\Omega; \mathbb{R}^d), \|p(x)\|_2 \leq 1, \forall x \in \Omega \}. \quad (1)$$

The subspace of functions  $u \in L^1(\Omega)$  with bounded variation  $\text{TV}(u) < \infty$  plays a major role in variational image analysis [1, 22]. The basic variational approach to denoising a given image function  $f$  reads [21]

$$\min_u \left\{ \frac{1}{2} \|u - f\|_2^2 + \alpha \text{TV}(u) \right\}, \quad \alpha > 0, \quad (2)$$

and provides the starting point of the present section. In what follows, we assume functions to be discretized on a regular grid covering  $\Omega$ , that is  $f, u \in \mathbb{R}^n$ , the divergence operator  $\text{div}$  in (1) is represented by a matrix  $L \in \mathbb{R}^{n \times nd}$ , and vector fields  $p(x) \in \mathbb{R}^d$ ,  $x \in \Omega$ , are given as vectors  $p \in \mathbb{R}^{nd}$ .

Note that  $\alpha \text{TV}(u)$  in (2) is a particular instance of a support function from convex analysis [20],

$$\sigma_{\mathcal{C}}(u) = \sup_{v \in \mathcal{C}} \langle u, v \rangle, \quad (3)$$

with  $\mathcal{C} = \text{div } \mathcal{D}$  and  $\mathcal{D} = \{p : \|p(x)\|_2 \leq \alpha\}$ . Accordingly, we start out from the more abstract variational problem in discretized form

$$\min_{u \in \mathbb{R}^n} \left\{ \frac{1}{2} \|u - f\|_2^2 + \sigma_{\mathcal{C}}(u) \right\}. \quad (4)$$

where

$$\mathcal{C} = L\mathcal{D} = \{v : v = Lp, p \in \mathcal{D}\}, \quad \mathcal{D} = \{p \in \mathbb{R}^{nd} : p_i \in \mathcal{D}_{loc}^i \subset \mathbb{R}^d, i = 1, \dots, n\}, \quad (5)$$

and  $\mathcal{D}_{loc}^i$  being *local constraint sets* for each  $p_i \in \mathbb{R}^d$ ,  $p = (p_1, \dots, p_n)^\top$ . Note that (5) covers the specific case  $\mathcal{D}_{loc}^i := B_\alpha(0) \subset \mathbb{R}^d$ , with  $B_\alpha(0)$  being a ball of radius  $\alpha$  centered at 0, corresponding to (1) after discretization.

## 2.2 Dual Formulation

We compute the dual variational problem of (2). Using the sub-differential  $\partial\sigma_{\mathcal{C}}(\bar{u}) = \{v : \bar{u} \in N_{\mathcal{C}}(v)\} = \{v : \langle \bar{u}, v - u \rangle \geq 0, \forall u \in \mathcal{C}\}$ , where  $N_{\mathcal{C}}(v)$  is the normal cone of  $\mathcal{C}$  at  $v \in \mathcal{C}$ , the optimality condition for  $\bar{u}$  minimizing (4) reads as

$$f - \bar{u} \in \partial\sigma_{\mathcal{C}}(\bar{u}) \Leftrightarrow \langle \bar{u}, f - \bar{u} - u \rangle \geq 0, \quad \forall u \in \mathcal{C}.$$

Using the additive decomposition  $f = \bar{u} + \bar{v}$ , we obtain

$$\langle \bar{v} - f, u - \bar{v} \rangle \geq 0, \quad \forall u \in \mathcal{C} \Leftrightarrow \bar{v} = \Pi_{\mathcal{C}}(f),$$

where  $\Pi_{\mathcal{C}}$  denotes the orthogonal projection onto  $\mathcal{C}$ , and with  $\bar{v} = L\bar{p}$  such that  $\bar{p}$  solves the dual problem

$$\min_{p \in \mathcal{D}} F(p), \quad F(p) = \frac{1}{2} \|f - Lp\|_2^2. \quad (6)$$

### 2.3 Solution-Dependent Adaptivity

We now generalize the variational formulation by additionally taking the variation of the constraints sets into account. That is,

- the dual constraint sets  $\mathcal{D}$  and  $\mathcal{D}_{loc}^i$  depend on the dual variable  $p$ , and
- the constraint set  $\mathcal{C}$  defining the primal formulation depends on  $u = f - Lp$ .

In what follows, we will be primarily concerned with the set-valued mapping

$$\mathcal{D}: \bar{p} \quad \Rightarrow \quad \mathcal{D}(\bar{p}) := \{p \in \mathbb{R}^{nd} : p_i \in \mathcal{D}_{loc}^i(\bar{p}) \subset \mathbb{R}^d, i = 1, \dots, n\}, \quad (7)$$

which in turn defines  $\mathcal{C}$  by (5). Accordingly, we focus on the variational problem

$$\min_{p \in \mathcal{D}(p)} F(p), \quad F(p) := \frac{1}{2} \|f - Lp\|_2^2, \quad (8)$$

that considerably generalizes (6) due to the dependency (7). We refer to this generalized concept as *solution-dependent adaptivity*.

## 3 Approach

For fixed  $\bar{p}$ , the problem

$$\operatorname{argmin}_{p \in \mathcal{D}(\bar{p})} F(p) \quad (9)$$

is convex and thus the mapping

$$\bar{p} \mapsto \operatorname{argmin}_{p \in \mathcal{D}(\bar{p})} F(p) \quad (10)$$

is well-defined. We are interested in a fixed point of this mapping.

### 3.1 A Quasi-Variational Inequality

In order to apply theoretical results from literature, we formulate our approach as a generalization of the variational inequality corresponding to the dual problem (9): For fixed  $\bar{p}$ , a minimizer  $p^*$  of (9) satisfies the following optimality condition:

$$-\nabla F(p^*) \in N_{\mathcal{D}(\bar{p})}(p^*) \quad \Leftrightarrow \quad \langle \nabla F(p^*), p - p^* \rangle \geq 0, \quad \forall p \in \mathcal{D}(\bar{p}). \quad (11)$$

A fixed point  $\bar{p} \in \mathcal{D}(\bar{p})$  of the mapping  $\bar{p} \rightarrow p^*$  therefore satisfies the quasi-variational inequality

$$-\nabla F(\bar{p}) \in N_{\mathcal{D}(\bar{p})}(\bar{p}) \quad \Leftrightarrow \quad \langle \nabla F(\bar{p}), p - \bar{p} \rangle \geq 0, \quad \forall p \in \mathcal{D}(\bar{p}). \quad (12)$$

### 3.2 Existence of Solutions

**Assumption 1** Each  $\mathcal{D}_{loc}^i : \mathbb{R}^{nd} \rightrightarrows \mathbb{R}^d, i = 1, \dots, n$  has the following properties:

- (i) For fixed  $p$  the set  $\mathcal{D}_{loc}^i(p)$  is a closed convex subset of  $\mathbb{R}^d$ .
- (ii) There exists  $C > 0$ , such that for all  $i, p: \mathcal{D}_{loc}^i(p) \subset \overline{B_C(0)}$ , where  $\overline{B_C(0)}$  is the closed ball with radius  $C$  centered at 0.
- (iii) For every  $i, p$  there exists  $0 < c := c(p)$ , such that  $B_c(0) \subset \mathcal{D}_{loc}^i(p)$ , where  $B_c(0)$  is the open ball with radius  $c$  centered at 0. In particular,  $\mathcal{D}_{loc}^i(p)$  is non-empty.
- (iv) The projection  $\Pi_{\mathcal{D}_{loc}^i(p)}(q)$  of  $q$  onto  $\mathcal{D}_{loc}^i(p)$  for a fixed  $q$  is continuous w.r.t.  $p$ .

**Proposition 1** Let  $F(p) := \frac{1}{2}\|f - Lp\|_2^2$ , where  $L : \mathbb{R}^{nd} \rightarrow \mathbb{R}^n$  is a linear operator. Moreover, let  $\mathcal{D}(p)$  be defined as in (7), such that  $\mathcal{D}_{loc}^i(p), i = 1, \dots, n$  satisfy Assumption 1. Then, the problem

$$\text{Find } \bar{p} \in \mathbb{R}^{nd} \text{ such that } \quad \langle \nabla F(\bar{p}), p - \bar{p} \rangle \geq 0, \quad \forall p \in \mathcal{D}(\bar{p}) \quad (13)$$

(cf. (12)) has a solution.

The proof of Proposition 1 utilizes the following theorem and lemma.

**Theorem 1** (cf. Theorem 5.2 in [6]) Let  $G : \mathbb{R}^m \rightarrow \mathbb{R}^m$  be a continuous point-valued mapping and  $\mathcal{D} : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$  be a set-valued mapping. Suppose that there exists a nonempty compact convex set  $P$  such that

- (i)  $\mathcal{D}(P) := \cup_{p \in P} \mathcal{D}(p) \subseteq P$ ;
- (ii)  $\mathcal{D}$  takes nonempty closed convex sets as values;
- (iii)  $\mathcal{D}$  is continuous, that is  $\mathcal{D}(p^k) \rightarrow \mathcal{D}(\bar{p})$  whenever  $p^k \rightarrow \bar{p}$  in the sense that for the projection  $\Pi_{\mathcal{D}(p)}$  onto  $\mathcal{D}(p)$ 

$$\Pi_{\mathcal{D}(p^k)}(p) \rightarrow \Pi_{\mathcal{D}(\bar{p})}(p) \text{ for all } p. \quad (14)$$

Then, there exists  $\bar{p} \in \mathbb{R}^m$  such that  $\langle G(\bar{p}), p - \bar{p} \rangle \geq 0$  for all  $p \in \mathcal{D}(\bar{p})$ .

The proof can be found in [6, Theorem 5.2].

**Lemma 1** Let  $\mathcal{D}$  be defined as in (7). Assume for every  $i = 1, \dots, n$  and  $q \in \mathbb{R}^d$  the projection  $\Pi_{\mathcal{D}_{loc}^i(p)}(q)$  to be continuous w.r.t.  $p$ . Then,  $\Pi_{\mathcal{D}(p)}(q)$  is continuous in  $p$  for fixed  $q \in \mathbb{R}^{nd}$ .

*Proof*  $\Pi_{\mathcal{D}}(q)$  can be written as  $\Pi_{\mathcal{D}(p)}(q) = (\Pi_{\mathcal{D}_{loc}^1(p)}(q_1), \dots, \Pi_{\mathcal{D}_{loc}^n(p)}(q_n))^\top$ . Thus, each component of  $\Pi_{\mathcal{D}(p)}$  is continuous, from which the continuity of  $\Pi_{\mathcal{D}(p)}$  follows immediately.  $\square$

Proof of Proposition 1: We apply Theorem 1. Conditions (i) and (ii) follow from Assumption 1, that, in turn, has to be verified for each specific application later, see Prop. 2, 3 and 4. Lemma 1 shows that also condition (iii) holds.  $\square$

## 4 Algorithm

In this section, we consider the problem of numerically computing a solution to the QVI (12): find  $\bar{p} \in \mathcal{D}(\bar{p})$  such that

$$\langle \nabla F(\bar{p}), p - \bar{p} \rangle \geq 0, \quad \forall p \in \mathcal{D}(\bar{p}). \quad (15)$$

**Algorithm 1:** Fixed point iteration with projected gradient steps.

---

**Input:**  $N, M \geq 1$ , point  $p^0 \in \mathcal{D}(p^0)$ ,  $\lambda \in (0, 1]$  sufficiently large.  
**Output:**  $p = p^M$ .  
Choose  $0 < \tau < 2/\nu$ , where  $\nu$  denotes the Lipschitz-constant of  $\nabla F$ . // initialization  
**begin**  
  **for**  $k = 1$  **to**  $M$  **do**  
     $\mathcal{D}^k = \mathcal{D}(p^{k-1})$ ,  
     $q^0 = p^{k-1}$ .  
    **for**  $l = 1$  **to**  $N$  **do**  
       $q^l = q^{l-1} - \frac{1}{\lambda} (q^{l-1} - \Pi_{\mathcal{D}^k}(q^{l-1} - \tau \nabla F(q^{l-1})))$ .  
     $p^k = q^N$ .

---

## 4.1 Prior Work

We comment on prior work that is relevant to the algorithm proposed below for solving (15).

- For the specific class of convex-valued mappings

$$\mathcal{D}: \mathbb{R}^{nd} \rightrightarrows \mathbb{R}^{nd}, \quad \mathcal{D}(p) = M(p) + \mathcal{D}_0, \quad (16)$$

with fixed closed convex set  $\mathcal{D}_0$  and Lipschitz continuous, strongly monotone mappings  $\nabla F, M: \mathbb{R}^{nd} \rightarrow \mathbb{R}^{nd}$ , convergence of the basic fixed point iteration

$$p^{k+1} = \Pi_{\mathcal{D}(p^k)}(p^k - \tau \nabla F(p^k)), \quad k = 0, 1, 2, \dots \quad (17)$$

to a solution of (15) was shown in [6, Thm. 5.3], for sufficiently small  $\tau$  and arbitrary  $p^0$ . We proposed a corresponding algorithm in [15], which we recall here for the sake of completeness, see Algorithm 1.

- Nesterov [17, Thm. 5] showed that iteration (17) only converges if the rate of variation of the feasible set  $\mathcal{D}(p)$  with  $p$  is much smaller than the condition number (Lipschitz constant divided by the monotonicity parameter) of the operator  $\nabla F$ . Assuming  $\nabla F$  to be strongly monotone, the author developed a more efficient modification of (17) together with an estimate of the improved convergence rate.

In our case (15), the mapping  $\mathcal{D}(p)$  is more general than (16). Moreover the mapping

$$\nabla F(p) = L^\top(Lp - f) \quad (18)$$

is not strongly monotone, due to the kernel  $\mathcal{N}(L) \neq \{0\}$  of  $L$ . The following lemma however shows that we may restrict  $\nabla F$  to the subspace  $\mathcal{N}(L)^\perp \subset \mathbb{R}^{nd}$ .

**Lemma 2** *Let the image data  $f$  be non-constant and  $\bar{p}$  solve (15) with  $F(p)$  as in (8). Then,  $L\bar{p} \neq 0$ .*

*Proof* Assume  $L\bar{p} = 0$ . Then,

$$\langle L^\top(L\bar{p} - f), p - \bar{p} \rangle = -\langle f, Lp \rangle = \langle -L^\top f, p \rangle \not\geq 0, \quad \text{for some } p \in \mathcal{D}(\bar{p}),$$

because

- $-L^\top f \neq 0$  corresponds to the gradient of  $f$ , that does not vanish by the hypothesis,
- and  $\bar{B}_c(0) \subset \mathcal{D}(p)$ , for every  $p$  and some  $c > 0$ , due to Assumption 1(iii).

This contradicts that  $\bar{p}$  solves (15). □

As a consequence, by restriction to the subspace  $\mathcal{N}(L)^\perp \subset \mathbb{R}^{nd}$ , gradient (18) becomes

$$\widetilde{\nabla}F(p) := \Pi_{\mathcal{N}(L)^\perp} L^\top (L \Pi_{\mathcal{N}(L)^\perp} p - f) \quad (19)$$

and strongly monotone,

$$\langle \widetilde{\nabla}F(p) - \widetilde{\nabla}F(q), p - q \rangle \geq \mu \|p - q\|^2, \quad \forall p, q \in \mathcal{N}(L)^\perp, \quad (20)$$

with monotonicity parameter  $\mu$  equal to the smallest non-vanishing eigenvalue of  $L^\top L$ . As a result, we may adopt Nesterov's approach [17].

## 4.2 Applying Nesterov's Approach

The algorithm detailed below is based on

### Assumption 2

(i)  $\widetilde{\nabla}F(p)$  defined by (19) is Lipschitz continuous with constant  $\nu$  and strongly monotone with parameter  $\mu$ . We define the condition number

$$\gamma := \frac{\nu}{\mu}. \quad (21)$$

(ii) There exists  $\eta > 0$  such that

$$\|\Pi_{\mathcal{D}(p)}(r) - \Pi_{\mathcal{D}(q)}(r)\|_2 \leq \eta \|p - q\|_2, \quad \forall p, q, r \in \mathbb{R}^{nd}. \quad (22)$$

(iii) We define the contraction gap

$$\delta := 1 - \eta\gamma > 0 \quad (23)$$

and assume it to be positive.

*Remark 1* Note that the operator  $\widetilde{\nabla}F(p)$  is still ill-conditioned, as  $\mu$  can become small. On the other hand, the constant  $\eta$  is induced by the dependency of the constraint set  $\mathcal{D}$  on  $p$  and thus is fixed. However, the *contraction gap* can be controlled by preconditioning, as we will see in Sect. 4.3.

Based on Assumption 2, Nesterov showed

**Theorem 2** *If  $\delta > 0$  as defined by (23), then, there exists a unique solution to problem (15).*

*Proof* See [17], Theorem 6 and Corollary 2.

A few more quantities need to be introduced in order to specify the algorithm.

**Definition 1** For some  $\beta > 0$ , we set

$$\psi_q^\beta(p) := \langle \widetilde{\nabla}F(q), q - p \rangle - \frac{1}{2}\beta \|p - q\|_2^2, \quad (24)$$

$$\Psi_k(p) := \sum_{i=0}^k \tau^i \psi_{q^i}^\mu(p), \quad \tau_i > 0, \quad \forall i, \quad (25)$$

$$S^N := \sum_{i=0}^N \tau^i. \quad (26)$$

**Algorithm 2:** Inner Iteration

---

**Input:**  $N \geq 1$ , convex set  $\mathcal{D}$ , point  $\hat{p} \in \mathcal{D}$ .  
**Output:**  $q(\mathcal{D}, \hat{p}, N) := \frac{1}{S^N} \sum_{i=0}^N \tau^i \tilde{q}^i$ .  
Set  $\tau^0 = 1$  and  $\tilde{q}^0 = \operatorname{argmax}_{p \in \mathcal{D}} \psi_{\hat{p}}^\nu(p)$ . // initialization

**begin**

**for**  $k = 0$  **to**  $N - 1$  **do**

$p^k = \operatorname{argmax}_{p \in \mathcal{D}} \Psi_k(p)$ ,

$\tilde{q}^{k+1} = \operatorname{argmax}_{p \in \mathcal{D}} \psi_{p^k}^\nu(p)$ ,

$\tau^{k+1} = \frac{1}{\gamma} S^k$ .

The algorithm for solving (15) is a two-stage iteration, comprising an inner iteration (Algorithm 2) and an outer iteration (Algorithm 3).

The outer iteration, defined by Algorithm 3 below, calls Algorithm 2 with input data  $\hat{N}$ ,  $\mathcal{D}(q^k)$  and  $\hat{p}^k$ , where  $\hat{N} = \hat{N}(\eta, \gamma)$  is the minimal number satisfying the condition

$$3\gamma \exp\left(-\frac{\hat{N}}{2(\gamma+1)}\right) \leq \frac{\delta}{4} \quad \Rightarrow \quad \hat{N} = \left\lceil 2(\gamma+1) \log \frac{12\gamma}{1-\eta\gamma} \right\rceil + 1, \quad (27)$$

and  $\hat{p}^k$  and  $q^k$  are computed by Algorithm 3.

**Algorithm 3:** Outer Iteration

---

**Output:** Sequence  $(q^k)$  converging to the unique solution  $\bar{q}$  of (15).  
Choose  $q^0$  arbitrary. // initialization

**begin**

**for**  $k \geq 0$  **do**

$\hat{p}^k = \Pi_{\mathcal{D}(q^k)}(q^k)$ ,

$q^{k+1} = q(\mathcal{D}(q^k), \hat{p}^k, \hat{N})$ . //  $\rightarrow$  Algorithm 2

For a proof of convergence of Algorithm 3 we refer to [17], Theorem 8, which also provides the following convergence rate:

$$\|q^k - \bar{q}\|_2 \leq \frac{1}{\delta} \exp\left(-\frac{\delta}{2}k\right) \|q^0 - \bar{q}^0\|_2, \quad (28)$$

where  $\delta$  is the contraction gap,  $\bar{q}$  the unique solution to (12) and  $\bar{q}^0$  is the solution for the fixed constraint set  $\mathcal{D} = \mathcal{D}(q^0)$ .

### 4.3 Practical Aspects

We illustrate and discuss two further issues relevant for implementations.

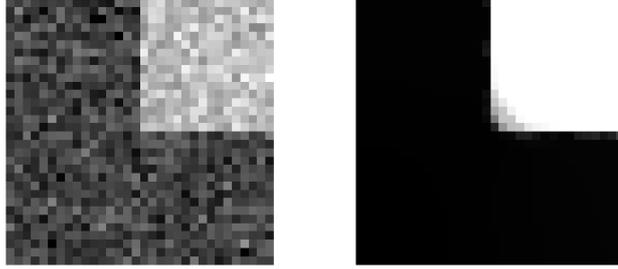
**Equivalence of  $\nabla F$  and  $\widetilde{\nabla F}$ :** To illustrate the assertion stated by Lemma 2, we numerically solved the variational inequality

$$\langle \nabla F(\bar{p}), p - \bar{p} \rangle \geq 0, \quad \forall p \in \mathcal{D} = (B_c(0))^n, \quad (29)$$

for some  $c > 0$  ( $d = 2$ )– see Figure 2. We used a basic primal-dual iteration [8] to compute a sequence  $(u^k, p^k)$  of primal and dual variables. The iteration was terminated if the gap-condition  $\|f - (u^k + v^k)\|_\infty < 10^{-5}$  with  $v^k = Lp^k$ , was satisfied. Next we solved

$$\langle \widetilde{\nabla F}(\bar{p}), p - \bar{p} \rangle \geq 0, \quad \forall p \in \mathcal{N}(L)^\perp \cap \mathcal{D}, \quad (30)$$

using the same termination criterion. Let  $(\bar{u}, \bar{p})$  denote the numerical solution to (29) and  $(\bar{u}', \bar{p}')$  the solution to (30). We observed  $\|\bar{u} - \bar{u}'\|_\infty < 10^{-12}$  and  $\|\bar{p} - \bar{p}'\|_\infty < 10^{-12}$ .



**Fig. 2** A denoising example to illustrate the equivalence of (29) and (30) – see text.

**Preconditioning:** Condition (23) couples the variation rate (22)  $\eta$  and the condition number  $\gamma$  defined by (21). In fact, the number  $0 < \eta\gamma < 1$  governs the contraction properties (convergence speed) of the fixed point iteration defined by Algorithm 3, cf. [17, Thm. 6]. As a consequence, if the condition number is poor, the mapping  $\mathcal{D}: p \mapsto \mathcal{D}(p)$  must be designed such that  $\mathcal{D}(p)$  varies slowly.

This situation, unfortunately, happens in practice. For the problem illustrated by Fig. 2, for instance, the largest and smallest eigenvalue of the linear mapping in (19) are  $\lambda_{\max} \approx 7.98$  and  $\lambda_{\min} \approx 0.00963$ , respectively, requiring  $\eta < 0.0012$  according to (23). And these numbers become worse as the data size  $n$  increases!

Using the notations

$$VV^\top := \Pi_{\mathcal{N}(L)^\perp}, \quad (31a)$$

$$A := V^\top L^\top LV, \quad (31b)$$

and (19), we rewrite the QVI (30),

$$\langle \widetilde{\nabla F}(\bar{p}), p - \bar{p} \rangle = \langle VV^\top L^\top (LVV^\top \bar{p} - f), p - \bar{p} \rangle \quad (32a)$$

$$= \langle AV^\top \bar{p} - V^\top L^\top f, V^\top (p - \bar{p}) \rangle \quad (32b)$$

$$= \langle P^{-1}AP^{-1}\bar{q} + g_f, q - \bar{q} \rangle \quad (32c)$$

$$\geq 0, \quad \forall q \in PV^\top \mathcal{D}(P^{-1}\bar{q}), \quad (32d)$$

where we introduced a symmetric, positive definite *preconditioning matrix*  $P$  and the shorthands

$$q := PV^\top p, \quad \bar{q} := PV^\top \bar{p}, \quad g_f := -V^\top L^\top f. \quad (33)$$

Note that using  $P$  in (32c) does not affect the unique solution  $\bar{p} = P^{-1}V\bar{q}$ . On the other hand,  $P$  does affect both the condition number of the mapping  $A$  by multiplication with  $P^{-1}$ , and the variation rate of the mapping

$$\mathcal{Q}: q \quad \Rightarrow \quad \mathcal{Q}(q) := PV^\top \mathcal{D}(P^{-1}q) \quad (34)$$

defined by (32c) and (32d). Let  $U$  be the orthogonal matrix diagonalizing  $A$ , such that  $A = UD_AU^\top$ , and choose  $P = UD_PU^\top$ . Then, (32c) reads

$$\langle UD_P^{-1}D_AD_P^{-1}U^\top \bar{q} + g_f, q - \bar{q} \rangle. \quad (35)$$

Thus, the preconditioning changes the condition number (21) to

$$\gamma \quad \rightarrow \quad \gamma'(P) := \frac{\max_i \{\lambda_i(A)/\lambda_i^2(P)\}}{\min_i \{\lambda_i(A)/\lambda_i^2(P)\}}. \quad (36)$$

Moreover, in view of (34), we have

$$\|PV^\top \Pi_{\mathcal{D}(p_1)}q - PV^\top \Pi_{\mathcal{D}(p_2)}q\| \leq \eta \lambda_{\max}(P) \|p_1 - p_2\|, \quad (37)$$

(cf. [23, Lemma 1.8.9]) and

$$\|p_1 - p_2\| = \|VP^{-1}q_1 - VP^{-1}q_2\| \leq \lambda_{\min}^{-1}(P) \|q_1 - q_2\|. \quad (38)$$

From (37) and (38) we conclude that the variation rate (22) changes to

$$\eta \quad \rightarrow \quad \eta'(P) := \gamma(P)\eta. \quad (39)$$

Choosing  $P$  amounts to a trade-off between reducing  $\gamma$  and increasing  $\eta$ . To illustrate this, we specifically choose  $P = A^{1/2}$  to obtain  $\gamma'(P) = \gamma'(A) = 1$  and  $\eta'(P) = \eta'(A) = \gamma(P)\eta = \gamma^{1/2}(A)\eta$ . The corresponding product defining the contraction gap (23) changes to

$$\eta\gamma = \eta \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} \quad \rightarrow \quad \eta'(A)\gamma'(A) = \eta\gamma^{1/2}(A). \quad (40)$$

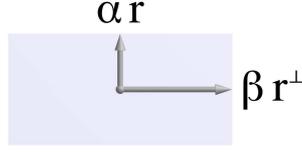
This is a considerable reduction and consequently we can choose a higher adaptation rate  $\eta$  satisfying condition (23). However, this advantage is compensated to some extent by the need to replace – as a part of the overall algorithm – the trivial orthogonal projection onto  $\mathcal{D}$  by the non-trivial orthogonal projection onto  $\mathcal{Q}$  defined by (34).

Taking into account this latter aspect and working out an optimal preconditioner  $P$  is left for future work.

## 5 Applications

In Sect. 5.2 we present three applications of the proposed adaptive total variation (TV) regularization, where the adaptivity depends on the solution of the optimization problem. These applications differ in the shape of the local constraint sets  $\mathcal{D}_{loc}^i(p)$  used. For each case we show that the global constraint set  $\mathcal{D}(p)$  fits into the concept presented above. In particular, we show that Assumption 1 in Sect. 3 is satisfied and that the projection  $\Pi_{\mathcal{D}(p)}(q)$  onto the constraint set is locally Lipschitz continuous w.r.t.  $p$ . In Sect. 5.3 we then comment on how the global Lipschitz property required for the convergence of Algorithm 3 (cf. Sect. 4) can be derived.

For the reader's convenience, we start with the simpler model of an adaptivity based on the (noisy) input data rather than the unknown solution. We refer to such a model as *data-dependent* adaptivity.



**Fig. 3** Local shape of  $\mathcal{D}_{loc}$  for the model of single directions: a rectangle with sides parallel/orthogonal to  $r$  and side length  $2\alpha$  and  $2\beta$ .

### 5.1 Data-Dependent Adaptivity

In order to derive an adaptive TV measure, we define the convex set  $\mathcal{D}$  (cf. (7)) based on edge information obtained from the noisy input data. In particular, in this section, the set  $\mathcal{D}$  does not depend on  $p$ . In our approach, we are interested in the local description of the set  $\mathcal{D}$  via sets  $\mathcal{D}_{loc}^i$  for each *pixel location*  $i = 1, \dots, n$ . Moreover, since translation invariance is desirable,  $\mathcal{D}_{loc}^i$  should not depend explicitly on  $i$ . Therefore, it suffices to consider one particular  $i$ . For simplicity of notation, we omit the index  $i$  of the pixel location in the following.

A local constraint set  $\mathcal{D}_{loc}$  of *anisotropic* shape can be motivated as follows. Let us assume for a moment, that, given a noisy input image  $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ , we know the location and orientation of image edges of the noise free data  $u$ . Moreover, we assume that image edges separate regions with homogeneous intensities. Therefore, we would like to define an anisotropic TV regularization term, which smoothes inside homogeneous regions, but preserves the borders between the regions. Thus, at edges, the regularization term should penalize jumps parallel to the edge stronger than jumps across the edge. In order to formulate such anisotropic regularization, let  $g(x) : \Omega \rightarrow \mathbb{R}_0^+$  be an edge-indicating function, i.e.  $g \approx 1$  is assumed near edges and  $g \approx 0$  in homogeneous regions. Moreover, let  $r : \Omega \rightarrow \mathbb{R}^2, \|r\| = 1$  be a vector field providing the edge normals.

One possible anisotropic regularization functional, which penalizes jumps parallel to edges stronger than across edges, then is

$$\phi(u) = \alpha \|r^\top \nabla u\| + \beta \|(r^\perp)^\top \nabla u\|, \quad (41)$$

where  $\beta \gg \alpha > 0$  near edges (where  $g \approx 1$ ) and  $\alpha = \beta$  in homogeneous regions (where  $g \approx 0$ ), e.g.

$$\alpha := g \alpha_0 + (1 - g) \beta_0, \quad \beta \equiv \beta_0, \quad (42)$$

for fixed  $0 < \alpha_0 \ll \beta_0$ . In our framework of dual constraint sets, this regularizer can be expressed as

$$\phi(u) = \sup_{v \in \mathcal{C}} \langle u, v \rangle = \sup_{p \in \mathcal{D}} \langle u, p \rangle, \quad (43)$$

with  $\mathcal{C}$  and  $\mathcal{D}$  being defined as in (5), and the local constraint set  $\mathcal{D}_{loc}$  defined as a rectangle:

$$\mathcal{D}_{loc} := \mathcal{R}(r, \alpha, \beta) = \{p \in \mathbb{R}^2 : |r^\top p| \leq \alpha, |(r^\perp)^\top p| \leq \beta\}. \quad (44)$$

Fig. 3 illustrates the parametrization of  $\mathcal{D}_{loc}$ .

In the following we define the functions  $g$  and  $r$ . In order to detect edges in the data  $f$ , we utilize the structure tensor  $J(f) \in \mathbb{R}^{2 \times 2}$  (definition below), and its eigenvectors  $v_i(f), i = 1, 2$  and eigenvalues  $\lambda_i(f), i = 1, 2$ . Let

$$J_0(f) := \nabla f_\sigma \nabla f_\sigma^\top, \quad (45)$$

where  $f_\sigma := f * K_\sigma$  is a smoothed version of  $f$ , obtained by convolution with a discrete Gaussian kernel  $K_\sigma$  with standard deviation  $\sigma > 0$ . The structure tensor  $J(f)$  is given as

$$J(f) := J_0(f) * K_\rho, \quad (46)$$

with  $\rho > 0$ . (Here the convolution is applied componentwise).

We now turn to the eigenvalues and eigenvectors of  $J(f)$ . Without loss of generality we assume that the eigenvalues of  $J(f)$  are ordered, i.e.  $\lambda_1(f) \geq \lambda_2(f) \geq 0$ , with corresponding eigenvectors  $v_1(f)$  and  $v_2(f)$ . From the structure tensor, we derive an edge normal by setting  $r = r(f) = v_1(f)$ . Moreover, as edge indicator, we choose

$$g(f) := g(\lambda_1(f) - \lambda_2(f)), \quad (47)$$

where  $g : \mathbb{R}_0^+ \rightarrow [0, 1]$  is a continuous and increasing function, such that  $g(0) = 0$  and  $g(s) = 1$  for  $s$  sufficiently large, e.g.  $g(x) = \frac{1}{1+x^2/K^2}, K > 0$ .

To conclude, we have derived a closed convex set  $\mathcal{D} = \mathcal{D}(f)$  by defining local sets  $\mathcal{D}_{loc}$  as rotated rectangles. Note that, since  $\mathcal{D}(f)$  is convex and not depending on  $u$ , the optimization problem (4) is convex.

## 5.2 Solution-Dependent Adaptivity

As we have seen in the previous section, for defining an adaptive regularization, information about image edges is required. When such edge structure is estimated from the noisy input data, even with a pre-smoothing step applied, the noise still has an impact on the estimation. In addition, edge estimation is decoupled from the actual image denoising task.

In contrast our approach derives more accurate edge information from the denoised image  $u = f - Lp$  by introducing the dependency of the constraint set  $\mathcal{D}_{loc}(p)$  on  $p$ . In particular, the interdependent problem of denoising the image and retrieving edge information is handled in a joint manner, which, however, leads to a non-convex optimization problem.

In the following sections, we describe three different kinds of anisotropic constraint sets  $\mathcal{D}_{loc}(p)$ , which differ in shape and edge structure used for the anisotropy.

### 5.2.1 Anisotropic TV with a Single Direction

In our first application, similar to Sect. 5.1, we will define the set  $\mathcal{D}_{loc}(p)$  as a rotated rectangle with one side parallel and one side orthogonal to a given edge.

To this end, we need to determine the normals of edges in  $u$  (one single direction  $r$  per pixel location). We again utilize the structure tensor, but now depending on  $u$ . Note that the entries of  $J(u)$  depend locally Lipschitz continuously on  $u$ , since  $J$  is obtained by applying linear operations and convolutions. Again, we assume that the eigenvalues of  $J(u)$  are ordered, i.e.  $\lambda_1(u) \geq \lambda_2(u) \geq 0$ , with the corresponding eigenvectors denoted by  $v_1(u)$  and  $v_2(u)$ .

A straightforward choice for an edge adaptive regularization is

$$\mathcal{D}_{loc}(p) := \mathcal{R}(r(u), \alpha, \beta) = \mathcal{R}(r(f - Lp), \alpha, \beta), \quad (48)$$

where  $\mathcal{R}(r, \alpha, \beta)$ , is defined as in (44) and  $r(u) := v_1(u)$ . For the moment, we consider  $\alpha, \beta > 0$  to be fixed. Then, however, the projection  $\Pi_{D(p)}$  is not continuous w.r.t.  $p$ , since the eigenvector  $v_1(u)$  in general does not depend continuously on the entries of  $J(u)$ . On the other hand, for  $\mathcal{R}(r, \alpha, \beta)$  as defined in (48),  $\Pi_{\mathcal{R}(r, \alpha, \beta)}$  is locally Lipschitz continuous w.r.t.  $r$ , as Lemma 5 (cf. appendix) shows. Moreover,  $p \rightarrow u = f - Lp$  is Lipschitz continuous. Thus, ensuring the Lipschitz continuity of  $r(u)$  is sufficient to guarantee the Lipschitz continuity of  $\Pi_{D(p)}$ .

In the following, we describe the construction of a vector  $r(u)$ , which is locally Lipschitz continuous w.r.t.  $u$ , such that  $r(u) = v_1(u)$ , if  $\lambda_1(u) \gg \lambda_2(u)$ . Note that the eigenvectors of  $J(u) \in \mathbb{R}^{2 \times 2}$  are locally Lipschitz continuous w.r.t.  $u$ , as long as the eigenvalues  $\lambda_1(u)$  and  $\lambda_2(u)$  differ (cf. Theorem 3 in [19]). We denote the difference between the two eigenvalues by

$$\text{coh}(u) := \lambda_1(u) - \lambda_2(u) \geq 0. \quad (49)$$

In contrast to  $v_1(u)$ ,  $\text{coh}(u)$  depends Lipschitz continuously on the entries of  $J$ , since the eigenvalues do (cf. e.g. Theorem of Wielandt-Hoffman in [26]). Since  $J(u)$  is locally Lipschitz continuous, the local Lipschitz continuity of  $\text{coh}(u)$  follows.

We recall the function  $g : \mathbb{R}_0^+ \rightarrow [0, 1]$  with the above mentioned properties. Let  $I(p, q, t) : S^1 \times S^1 \times [0, 1] \rightarrow S^1$  be an interpolation from  $p$  to  $q$  on the sphere  $S^1$  as described in the appendix (cf. Definition 2). Lemma 3 in the appendix then shows, that

$$r(u) := I(v_1(u), (0, 1)^\top, g(\text{coh}(u))) \quad (50)$$

is locally Lipschitz continuous.

For the side lengths  $\alpha$  and  $\beta$  of the rectangle we would like to have  $\alpha \ll \beta$  near edges and  $\alpha = \beta$  in homogeneous regions. To this end, for fixed  $0 < \alpha_0 \leq \beta_0$  we set

$$\alpha(u) := (\alpha_0 - \beta_0)g(\text{coh}(u)) + \beta_0, \quad \beta(u) = \beta_0,$$

where function  $g$  is defined as above. Note that  $\text{coh}(u)$  depends Lipschitz-continuously on  $u$ , thus also  $\alpha(u)$  is Lipschitz-continuous. With these particular  $r(u)$ ,  $\alpha(u)$  and  $\beta(u) = \beta_0$  we set

$$\boxed{\mathcal{D}_{loc}(p) := \mathcal{R}(r(f - Lp), \alpha(f - Lp), \beta_0).} \quad (51)$$

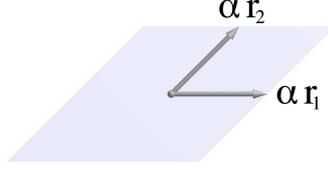
Then,  $\mathcal{D}_{loc}(p)$  satisfies Assumption 1:

**Proposition 2** *Let  $\mathcal{D}_{loc}(p)$  be defined as in (51). Then*

- (i)  $\mathcal{D}_{loc}(p)$  is closed, convex and satisfies  $B_\alpha(0) \subset \mathcal{D}_{loc}(p) \subset \overline{B_{\sqrt{2}\beta_0}}(0)$ .
- (ii) For fixed  $q \in \mathbb{R}^2$ ,  $u \rightarrow \Pi_{\mathcal{D}_{loc}(p)}(q)$  is locally Lipschitz continuous, with the Lipschitz constant depending on  $\|q\|_2$ .

*Proof* (i) The set  $\mathcal{D}_{loc}(p)$  is a closed rectangle with center 0 and side lengths  $2\beta_0$  and  $2\alpha \leq 2\beta_0$  and contains the ball  $B_{\alpha_0}(0)$ . (ii) Note that  $u = f - Lp$  is Lipschitz continuous w.r.t.  $p$ . Since  $J(u)$  is a composition of convolutions and differentiation,  $J(u)$  is locally Lipschitz continuous. Lemma 4 then provides the local Lipschitz continuity of  $r(u)$ . Together with Lemma 5, the local Lipschitz continuity of  $\Pi_{\mathcal{D}_{loc}(p)}$  follows. □

Alternatives to (51) are e.g. to use rotated squares or two-dimensional ellipsoids with one half-axis  $r(u)$  of length  $\alpha$  and the other half-axis of length  $\beta \gg \alpha$ . We discuss elliptical constraint sets for the 3D case in detail in Sect. 5.2.3.



**Fig. 4** Local shape  $\mathcal{D}_{loc}$  for the model of Steidl & Teuber [24]: a parallelogram with sides parallel to  $r_1$  and  $r_2$  with side length  $2\alpha$ .

### 5.2.2 Anisotropic TV with Double Directions

In order to preserve edges as well as corners, Steidl & Teuber [24] proposed an anisotropic TV method based on the estimation of two orientations  $r_1, r_2 : \Omega \rightarrow \mathbb{R}^2$ . They consider the variational problem

$$\min_u \frac{1}{2} \|u - f\|_2^2 + \alpha (\|r_1^\top \nabla u\|_2 + \|r_2^\top \nabla u\|_2), \quad (52)$$

where the directions  $r_i, i = 1, 2$  are retrieved from the data  $f$ . In the dual formulation of (52) the set  $\mathcal{D}_{loc} = \mathcal{D}_{loc}(f)$  is a parallelogram with sides  $r_i, i = 1, 2$ :

$$\boxed{\mathcal{D}_{loc}(p) = \mathcal{P}(r_1, r_2, \alpha) := \{p \in \mathbb{R}^2 : |r_1^\top p| \leq \alpha, |r_2^\top p| \leq \alpha\}.} \quad (53)$$

Figure 4 illustrates  $\mathcal{P}(r_1, r_2, \alpha)$ .

Steidl & Teuber [24] discuss two related models to obtain  $r_i, i = 1, 2$ , the *transparent model* and the *occlusion model*. In our considerations, we concentrate on the occlusion model. Moreover, we consider  $r_i = r_i(u)$  depending on the unknown  $u = f - Lp$  and, by introducing slight changes of the original approach, guarantee the applicability of the theoretical results of Sect. 3. The orientations  $r_i(u)$  are obtained as follows. Let

$$\nu(u) := ((\partial_x u_\sigma)^2, \partial_x u_\sigma \partial_y u_\sigma, (\partial_y u_\sigma)^2)^\top, \quad (54)$$

where  $u_\sigma$  is obtained by convolution of  $u$  with a discrete Gaussian kernel  $K_\sigma, \sigma > 0$  as in the previous section. For the occlusion model, the following structure tensor is utilized:

$$J_0(u) := \nu(u)\nu^\top(u), \quad J(u) := J_0(u) * K_\rho, \quad (55)$$

where the convolution is applied componentwise.

Now let  $\lambda_1(u) \geq \lambda_2(u) \geq \lambda_3(u) \geq 0$  denote the eigenvalues of  $J(u)$  and  $v_1(u), v_2(u)$  and  $v_3(u)$  the corresponding eigenvectors. Analogously to the previous section, in view of the continuity of  $v_i(u)$ , we have to deal with non-isolated eigenvalues. To this end, we define

$$\text{coh}_1(u) := \lambda_1(u) - \lambda_2(u), \quad \text{coh}_2(u) := \lambda_2(u) - \lambda_3(u). \quad (56)$$

In order to define  $r_1(u), r_2(u)$ , we consider the following cases:

1. **Corners** ( $\text{coh}_2(u) > 0$ ): Steidl & Teuber propose to use the unit vectors

$$r_1^{(1)}(u) \parallel (v_{3,1}(u), y_1(u))^\top, \quad r_2^{(1)}(u) \parallel (v_{3,1}(u), y_2(u))^\top, \quad (57)$$

where  $a \parallel b$  denotes that  $a$  and  $b$  are parallel and  $y_1(u), y_2(u)$  are the solutions of the quadratic equation

$$y^2 + v_{3,2}(u)y + v_{3,1}(u)v_{3,3}(u) = 0. \quad (58)$$

The local constraint set degenerates, if  $r_1^{(1)}(u)$  and  $r_2^{(1)}(u)$  become parallel. This is the case if

- $(v_{3,1}(u), y_1(u)) \parallel (v_{3,1}(u), y_2(u))$ , or, equivalently  $y_1(u) = y_2(u)$ , or
- $v_{3,1} = 0$ .

In view of Assumption 1 (iii), we prevent the degeneration of the constraints sets by choosing

$$r_1^{(2)}(u) \parallel (v_{3,2}(u), v_{3,3}(u))^\top, \quad r_2^{(2)}(u) \parallel (-v_{3,3}(u), v_{3,2}(u))^\top, \quad (59)$$

if  $\omega(u) := |y_1(u) - y_2(u)|$  or  $|v_{3,1}|$  become zero.

2. **Edges** ( $\text{coh}_2(u) = 0$ ,  $\text{coh}_1(u) > 0$ ): Since we can only guarantee that eigenvalue  $\lambda_1(u)$  is isolated, we determine  $r_1, r_2$  depending on the eigenvector  $v_1(u)$ .

Along straight edges, the eigenvector  $v_1$  is parallel to the normal of the edge. Therefore,  $v_1$  and  $v_1^\perp$  are suitable for defining the orientation for anisotropic TV at edges. We set

$$r_1^{(3)}(u) \parallel (v_{1,1}(u), v_{1,2}(u))^\top, \quad r_2^{(3)}(u) \parallel (-v_{1,2}(u), v_{1,1}(u))^\top. \quad (60)$$

3. **Homogeneous regions** ( $\text{coh}_1(u) = \text{coh}_2(u) = 0$ ):

$$r_1^{(4)}(u) := (1, 0)^\top, \quad r_2^{(4)}(u) := (0, 1)^\top. \quad (61)$$

In general,  $r_1(u), r_2(u)$  have to be continuous interpolations between the cases above. Let

$$r_i(u) = I \left( I \left( r_i^{(1)}(u), r_i^{(2)}(u), g(\omega(u))g(|v_{3,1}(u)|) \right), \right. \\ \left. I \left( r_i^{(3)}(u), r_i^{(4)}(u), g(\text{coh}_1(u)) \right), g(\text{coh}_2(u)) \right), \quad i = 1, 2, \quad (62)$$

using  $g$  as defined in the previous section and  $I$  as defined above (see also Def. 2 in the appendix).

**Proposition 3** Let  $\mathcal{D}_{loc}(p) = \mathcal{P}(r_1(f - Lp), r_2(f - Lp), \alpha)$  with  $\mathcal{P}(r_1, r_2, \alpha)$  being the parallelogram defined in (53) and  $r_i(u)$ ,  $i = 1, 2$  defined as in (62).

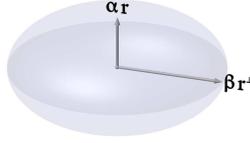
1.  $\mathcal{D}_{loc}(p)$  is closed, convex and satisfies  $\mathcal{D}_{loc}(p) \subset \overline{B_{2\alpha}(0)}$ .
2. For any  $p$  there exists  $c(p)$  such that  $B_{c(p)} \subset \mathcal{D}_{loc}(p)$ .
3.  $\Pi_{\mathcal{D}_{loc}(p)}(q)$  is Lipschitz continuous w.r.t.  $p$ , with the Lipschitz constant depending on  $\|q\|_2$ .

In particular,  $\mathcal{D}_{loc}(p)$  satisfies Assumption 1.

*Proof* The first claim follows from the fact that  $\mathcal{D}_{loc}(p)$  is a closed parallelogram with sides of length  $\alpha$ . The second claim follows from the fact that, by definition of  $r_1, r_2$ , the case  $r_1 \parallel r_2$  is excluded, i.e. the parallelogram can not degenerate. As a consequence, there exists  $c(p)$  such that  $B_{c(p)} \subset \mathcal{D}_{loc}(p)$ .

For the local Lipschitz continuity of  $\Pi_{\mathcal{D}_{loc}(p)}$ , we observe the following: First,  $u = f - Lp$  is Lipschitz continuous w.r.t.  $p$ . Since  $J(u)$  is a composition of convolutions and differentiation,  $J(u)$  is locally Lipschitz continuous. The vectors  $r_i^{(j)}(u)$ ,  $i = 1, 2$ ,  $j = 1, \dots, 3$  are defined in a way that they are locally Lipschitz continuous w.r.t.  $u$ , when restricted to the individual case. Note that also  $v_{3,1}(u)$  and  $\omega(u)$  are locally Lipschitz continuous, as long as  $\text{coh}_2(u) > 0$ . Moreover,  $r_i^{(4)}$  is constant w.r.t.  $u$ . The local Lipschitz continuity of  $r_i(u)$ ,  $i = 1, 2$  is guaranteed by smooth interpolation, c.f. Lemma 4 applied to the nested interpolations. The local Lipschitz continuity of  $\mathcal{P}(r_1, r_2, \alpha)$  is provided by Lemma 5.

□



**Fig. 5** Local shape  $\mathcal{D}_{loc}$  for spatio-temporal TV regularization. Prototype of an ellipsoid with one half axis parallel to  $r$  with length  $\alpha$  and the other half-axes of length  $\beta$ . In order to regularize surfaces in the spatio-temporal domain, one would choose  $r$  parallel to the surface normal and  $\alpha \ll \beta$  to achieve strong smoothing tangential to the surfaces, which results in shrunken ellipsoid.

### 5.2.3 Anisotropic Spatio-Temporal TV Minimization

In the following we describe a spatio-temporal TV minimization approach. We interpret time as third coordinate, thus  $u, f : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ ,  $p : \Omega \rightarrow \mathbb{R}^6$ .

To obtain directional information, we utilize the three-dimensional structure tensor  $J_\rho(u)$  defined analogously to (45) and (46). Let  $\lambda_1(u) \geq \lambda_2(u) \geq \lambda_3(u) > 0$  denote the eigenvalues and  $v_1(u), v_2(u), v_3(u)$  the eigenvectors of  $J_\rho(u)$ .

Let us assume that a spatio-temporal intensity gradient is present in  $u_\sigma(x)$ , which forms a two-dimensional iso-surface. Then,  $\lambda_1(u) \gg \lambda_2(u)$  and  $v_1(u)$  approximates the normal to this surface. The aim is to penalize variations mainly in directions tangential to the surface and allow variations in normal direction. To this end we set

$$\mathcal{D}_{loc}(p) := \mathcal{E}(v_1(f - Lp), \alpha, \beta), \quad (63)$$

where

$$\mathcal{E}(r, \alpha, \beta) := \left\{ q \in \mathbb{R}^3 : \frac{1}{\alpha^2} |r^\top q|^2 + \frac{1}{\beta^2} \|\text{Id} - rr^\top\|_2^2 \leq 1 \right\}, \quad 0 < \alpha \ll \beta. \quad (64)$$

is a three-dimensional ellipsoid (see Fig. 5 for an illustration).

In homogeneous regions, where a unique orientation  $r$  can not be estimated, we choose  $\mathcal{D}_{loc}(p) := B_\alpha(0)$ . A continuous transition between both cases is obtained by defining

$$\boxed{\mathcal{D}_{loc}(p) := \mathcal{E}(r(f - Lp), \tilde{\alpha}(f - Lp), \beta)}, \quad (65)$$

where

$$r(u) := I(v_1(u), (0, 0, 1)^\top, g(\text{coh}(u))), \quad (66)$$

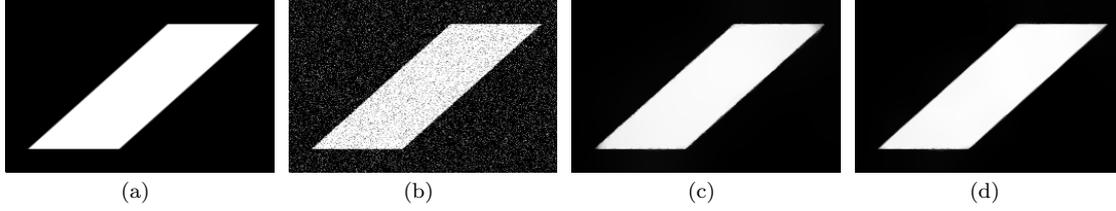
$$\tilde{\alpha}(u) := g(\text{coh}(u))\alpha + (1 - g(\text{coh}(u)))\beta, \quad (67)$$

$$\text{coh}(u) := \lambda_1(u) - \lambda_2(u) \gg 0. \quad (68)$$

In order to remove speckles and similar kinds of distortions, an adaptation of (65) is required. This is due to the fact that at speckles,  $v_1(u)$  points in direction of  $(0, 0, 1)^\top$ . Using (65) with the above  $\tilde{\alpha}$  then would lead to a penalization of  $\nabla u$  mainly in spatial directions, which is not suitable for removing distortions of medium/large scale in spatial directions. Instead, we propose to use (65) with

$$\tilde{\alpha}(u) = g(\text{coh}(u))g(\phi(u))\alpha + (1 - g(\text{coh}(u))g(\phi(u)))\beta, \quad (69)$$

where  $\phi(u)$  is the angle between  $v_1(u)$  and  $(0, 0, 1)^\top$ . The above modification leads to stronger smoothing of surfaces parallel to the  $x_1, x_2$ -axes.



**Fig. 6** Filtering of an artificial test image: (a) ground truth, (b) noisy test images, (c) filtering with the standard ROF model, (d) anisotropic filtering with double directions. Intensity scales are identical for each result. Undesirable smoothing effects at corners are considerably reduced by anisotropic filtering. See Fig. 7 for a close-up.

**Proposition 4** *The set  $\mathcal{D}_{loc}(p)$  defined in (65) with the above definitions of  $\tilde{\alpha}(u)$  satisfies Assumption 1. Moreover,  $(p, q) \rightarrow \Pi_{\mathcal{D}_{loc}(p)}(q)$  is locally Lipschitz continuous.*

*Proof* The set  $\mathcal{D}_{loc}(p)$  is a closed ellipsoid and therefore is convex. Its half-axes are bounded by  $c := \min(\alpha, \beta)$  from below and  $C := \max(\alpha, \beta)$  from above. Thus,  $B_c(0) \subset \mathcal{D}_{loc}(p) \subset B_C(0)$ .

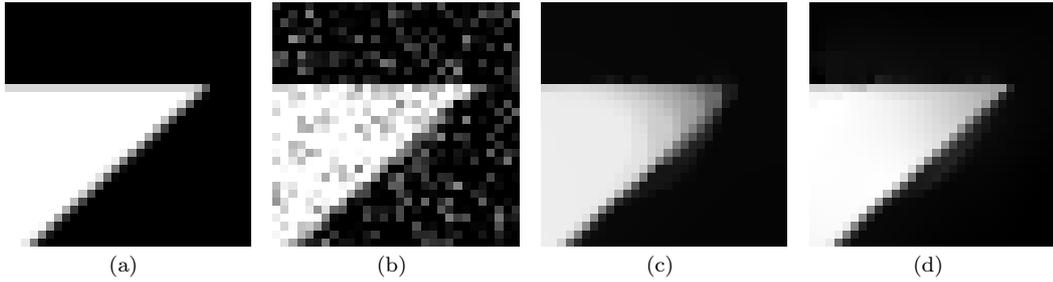
The local Lipschitz continuity of  $\Pi_{\mathcal{D}_{loc}(p)}$  can be shown as follows. First,  $u = f - Lp$  is Lipschitz continuous w.r.t.  $p$ . Since  $J(u)$  is a composition of discrete convolutions and differentiation,  $J(u)$  is locally Lipschitz continuous.  $\tilde{\alpha}$  is locally Lipschitz continuous w.r.t.  $u$ , since  $g$ ,  $\text{coh}(u)$  and  $\phi(u)$  are. Lemma 4 provides the local Lipschitz continuity of  $r(u)$ . By Lemma 6 the projection  $\Pi_{\mathcal{E}(r, \alpha, \beta)}(q)$  onto the ellipsoid  $\mathcal{E}(r, \alpha, \beta)$  is local Lipschitz continuous w.r.t.  $r$ ,  $\alpha$  and  $\beta$ .  $\square$

### 5.3 Global Lipschitz Continuity

In the above applications, we have considered three different examples for constraint sets  $\mathcal{D}(p)$ , defined via the local sets  $\mathcal{D}_{loc}(p) = \mathcal{R}(r, \alpha)$ ,  $\mathcal{D}_{loc}(p) = \mathcal{P}(r_1, r_2, \alpha)$  and  $\mathcal{D}_{loc}(p) = \mathcal{E}(r, \alpha, \beta)$ , respectively, (cf. (51), (53) and (65)) and directions  $r$ ,  $r_1$ , and  $r_2$  were estimated from structure tensors depending on  $u = f - Lp$ . (cf. (50), (62) and (66)). We have shown that the constraint sets satisfy Assumption 1 and that the projections are locally Lipschitz continuous w.r.t.  $p$ .

In order to fit these applications into the algorithms presented in Sect. 4, we require global Lipschitz continuity, where the Lipschitz constant is independent of  $q$ . This can be achieved by considering  $p, q$  only on a given compact subset of  $\mathbb{R}^{nd}$ , as we will see in the following.

In the case of problem (12), the solution  $p$  has to lie in  $\mathcal{D}(p)$ . Since  $\mathcal{D}(p)$  consists of the local sets  $\mathcal{D}_{loc}^i$ , which by Assumption 1, all lie in  $\overline{B_C(0)}$ , we find  $p \in \overline{B_C(0)}$ . Therefore, the optimization problem can be restricted to  $p \in B_C(0)$  and, consequently,  $q \in B_C(0)$ . We then obtain global Lipschitz continuity of  $\Pi_{\mathcal{D}(p)}(q)$  with a constant independent from  $q$  as follows: First we achieve Lipschitz continuity of the projections  $\Pi_{\mathcal{D}_{loc}^i}$ . For the first two cases based on rectangles and parallelograms, this Lipschitz continuity follows from the fact that the mapping  $p \rightarrow r(u)$  is locally Lipschitz continuous and that the Lipschitz constant for the projection onto parallelograms depends on  $\|q\|_2$ , see Lemma 5 in the appendix. In the third case, using ellipsoids, Lipschitz continuity follows from the local Lipschitz continuity of  $p \rightarrow (r(u), \alpha, \beta)$  and  $(r, \alpha, \beta, q) \rightarrow \Pi_{\mathcal{E}(r, \alpha, \beta)}(q)$ , see Lemma 6. Finally, we can prove similar to Lemma 1 that  $\Pi_{\mathcal{D}(p)}$  is Lipschitz continuous, whenever the projections  $\Pi_{\mathcal{D}_{loc}^i}$  to the local constraint sets are.



**Fig. 7** Zoom into a region of the images of Fig. 6: (a) ground truth, (b) noisy test images, (c) filtering with the standard ROF model, (d) anisotropic filtering with double directions. Undesirable smoothing effects at the corners are considerably reduced by anisotropic filtering.

## 6 Experiments

We present experimental results for the anisotropic TV models presented in Sect. 5. Up to now Algorithm 1 is used for solving the corresponding quasi-variational inequalities. A thorough evaluation of Algorithm 3 including dedicated convex solvers for Algorithm 2 is beyond the scope of this paper and will be reported in future work.

### 6.1 Anisotropic TV Minimization with Double Directions

Let us start with experimental results for the anisotropic TV model with *double directions*, where  $\mathcal{D}_{loc}(p) = \mathcal{P}(r_1, r_2, \alpha)$  with  $\mathcal{P}(r_1, r_2, \alpha)$  defined as in (53) and  $r_1, r_2$  as in (62). We compare our approach to standard ROF minimization [21] using the same regularization parameter  $\alpha$  and consider two different test images, both with artificially added Gaussian noise of mean zero.

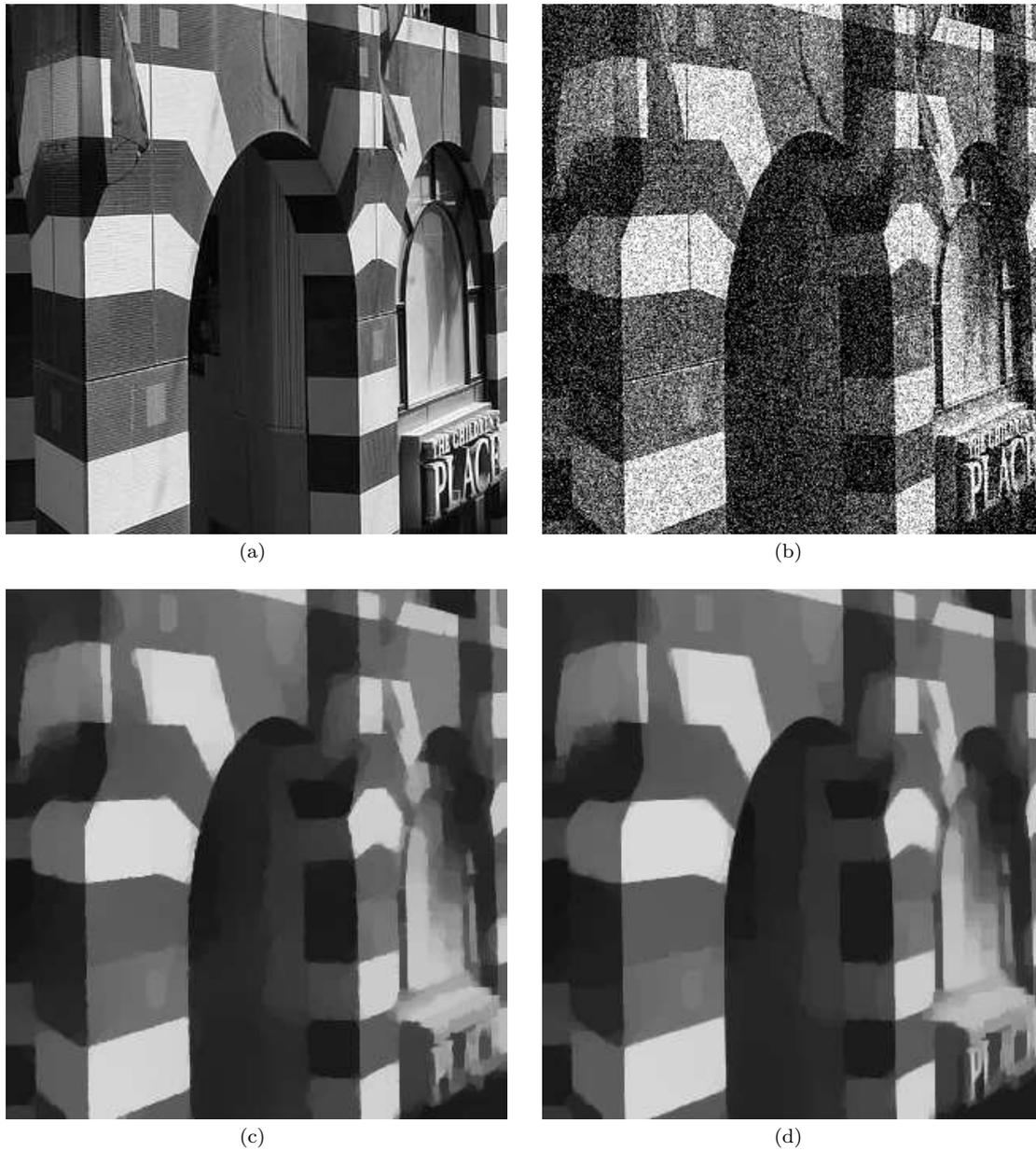
For the first test image (cf. Fig. 6, left) we use  $\alpha = 0.6$  and 10 outer iteration steps. The results of the standard and anisotropic TV models in Figs. 6 and 7 show that anisotropic TV minimization reconstructs corners of the parallelogram better and produces less smoothing at corners (as already demonstrated in [24]).

The second test image is a real world image with artificial noise. The results of standard ROF and anisotropic TV minimization for  $\alpha = 0.4$  and 10 outer iteration steps are depicted in Fig. 8. In order to highlight the differences, we zoom into a region of the image: Fig. 9 shows results for the standard ROF model (Fig. 9(b)), of applying anisotropic TV minimization with double directions, where the constraint set depends only on the data  $f$ , i.e.  $\mathcal{D} = \mathcal{D}(f)$  (Fig. 9(c)), and of anisotropic TV with the constraint set depending on the solution,  $\mathcal{D} = \mathcal{D}(p)$  (Fig. 9(d)).

Anisotropic filtering leads to an improved and more regular reconstruction of edges and less stair-casing. If the constraint sets depend on the solution itself, an adaptation to local structures can be observed during the iterations, see Fig. 9, right. Here, the reconstruction of the characters improves when using fully adaptive constraint sets. For a comparison with the non-local means algorithm, we refer to [24].

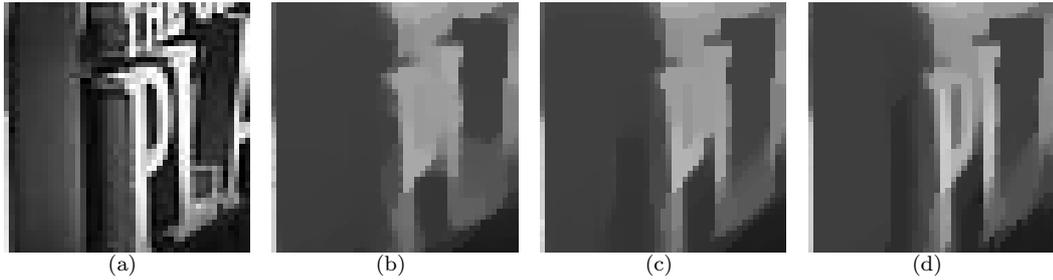
### 6.2 Adaptive Motion-Based TV Minimization for Image Sequences

In our example for spatio-temporal TV minimization, we consider an image sequence recorded with a time-of-flight (ToF) camera, see Fig. 10 (4 frames from of the whole sequence). ToF

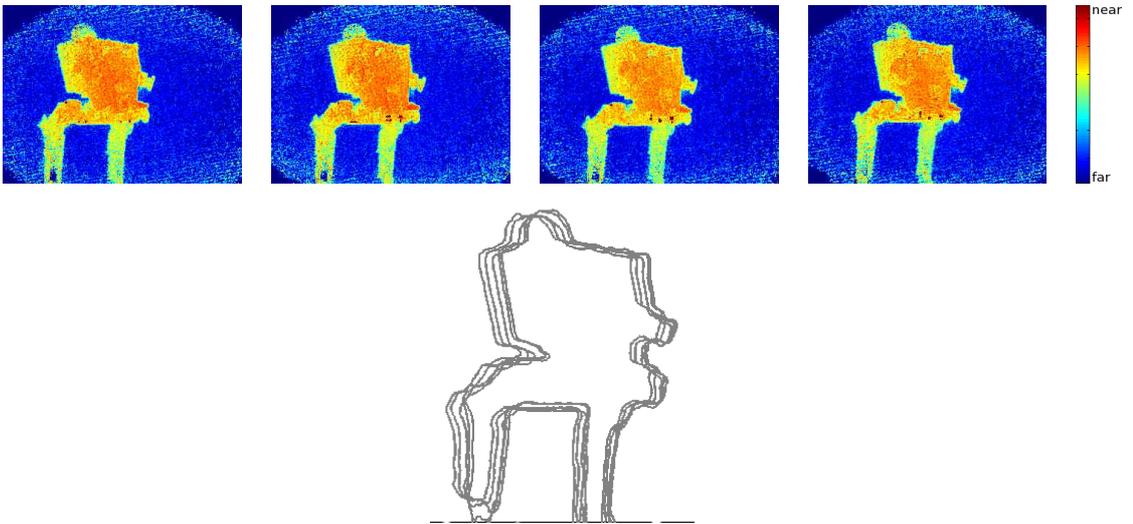


**Fig. 8** Filtering of real test image: (a) ground truth, (b) noisy test images, (c) result of standard ROF minimization, (d) result of anisotropic TV minimization with double directions. Undesirable smoothing effects at the corners are considerably reduced on the right. Detailed views can be found in Fig. 9.

cameras provide a depth map of the captured scene, which here consist of a foreground object in front of a background. The noise and speckles, which can be observed in the original data, are introduced by the camera system. Induced by the camera movement, the object's position in the frame sequence shifts to the right. Fig. 10, bottom, illustrates this shift by showing the level lines of one specific depth level for each of the four frames.



**Fig. 9** Zoom into a region of the filtered images shown in Fig. 8: (a) ground truth (b) standard ROF, (c) anisotropic TV minimization with  $\mathcal{D} = \mathcal{D}(f)$ , (d) adaptive anisotropic TV minimization with  $\mathcal{D} = \mathcal{D}(p)$ . Solution-dependent adaptivity of the TV regularization improves details.

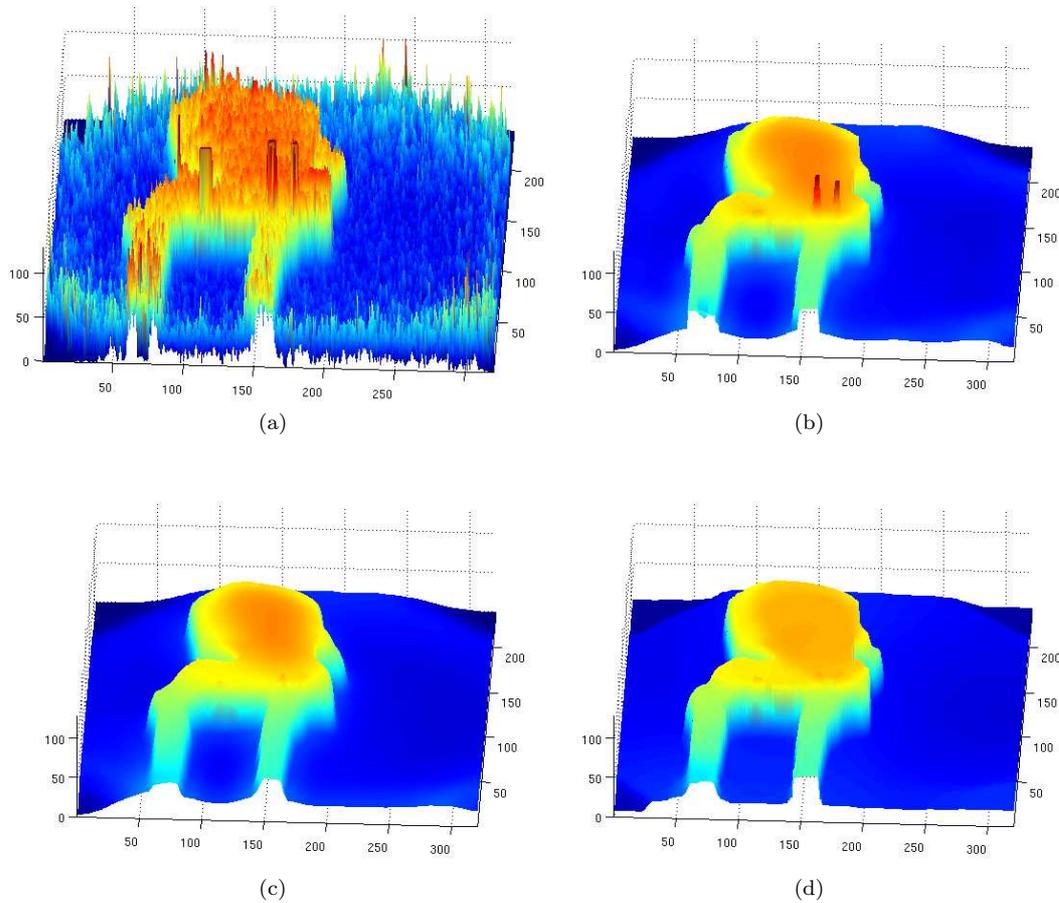


**Fig. 10** **Top row:** Four exemplarily selected frames of a sequence of depth maps taken with a time-of-flight camera. **Bottom:** Movement of one specific contour line over time.

For filtering, we propose to use spatio-temporal anisotropic TV with

$$\mathcal{D}_{loc}(p) = \mathcal{E}(r(f - Lp), \tilde{\alpha}(f - Lp), \beta), \quad (70)$$

where  $r(u)$  is defined as in (66) and  $\tilde{\alpha}(u)$  as in (69). As parameters, we chose  $\alpha = 0.3$ ,  $\beta = 0.001$  and 10 steps for the outer iteration. The result for one specific frame is depicted in Fig. 11(d). We compare this method with standard 2D ROF (Fig. 11(b)) and 3D ROF in the spatio-temporal domain (Fig. 11(c)), using the same parameter  $\alpha = 0.3$ . Additionally, we zoom into two image regions, see Fig. 12. We observe that standard 2D ROF filtering provides a good noise removal and edges preservation, but is not able to remove the speckles. 3D ROF filtering removes both noise and speckles, but introduces some blurring of edges, which is caused by the object's shift in time and the stair-casing effect of the TV model in 3D. The proposed adaptive anisotropic TV regularization combines the advantages of the 2D and 3D isotropic model: As for the 2D ROF model, noise is removed while the edges are kept sharp. In addition the smoothing now is coherent in time, as for the 3D ROF model, but without introducing motion blur. Due to the temporal coherence speckles are removed.

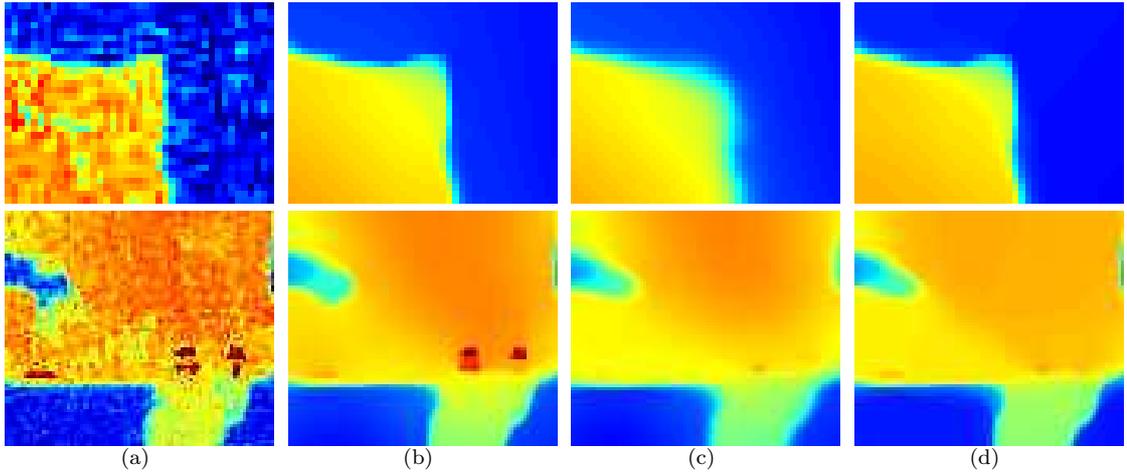


**Fig. 11** Spatio-temporal denoising of time-of-flight data. (a) one of the original frames taken with a time-of-flight camera with system-inherent noise, (b) frame filtered with standard 2D ROF, (c) frame sequence filtered with standard 3D ROF, (d) frame sequence filtered with the proposed adaptive TV minimization. Only the spatio-temporal methods are able to remove both noise and speckles. Anisotropic TV regularization keeps the result sharper than isotropic 3D TV minimization.

## 7 Conclusion

We presented a novel class of variational denoising approaches based on non-smooth convex regularizers with adaptive constraint sets. Sufficient conditions that hold in typical applications were specified for the existence of fixed points, along with an algorithm for computing them by solving a sequence of large convex programs. Our approach covers in particular spatial and spatio-temporal denoising with adaptive total variation regularization as special cases, that served to demonstrate our approach by a range of numerical experiments.

Our future work will focus on Nesterov's algorithm and the joint optimization of its components for specific applications: large-scale sparse convex solvers for the inner iterative loop (Algorithm 2) and preconditioning depending on the linear operators involved.



**Fig. 12** Zoom into two regions of the depth map shown in Fig. 11: (a) original data with system-inherent noise, (b) result of standard 2D ROF filtering, (c) 3D ROF filtering, (d) proposed adaptive TV minimization. 2D ROF filtering provides sharp edges (top), but can not remove speckles (bottom). Standard 3D ROF removes speckles, but blurs the image edges. Only spatio-temporal anisotropic TV is able to both remove noise and speckles *and* to keep edges sharp.

## A Appendix

### A.1 Smooth Interpolation on $S^{d-1}$

In this section we consider a smooth approximation of one specific eigenvector of a symmetric positive definite matrix. To this end, let  $J(u) \in \mathbb{R}^{d \times d}$  be a symmetric, positive semi-definite matrix, such that its entries depend Lipschitz continuously on  $u$ . We denote the eigenvalues of  $J(u)$  by  $\lambda_1(u), \dots, \lambda_n(u)$  and assume w.l.o.g. that they are ordered in decreasing order. Let  $v_1(u), \dots, v_2(u)$  denote the corresponding normalized eigenvectors.

We are interested in the Lipschitz continuous vector field  $r(u)$  parallel to  $v_1(u)$ . Unfortunately,  $v_1(u)$  is not Lipschitz continuously depending on  $u$ . Problems arise, when  $\lambda_1(u)$  is a multiple eigenvalue, in which case the eigenvector  $v_1(u)$  is not uniquely defined.

We therefore construct a vector field  $r(u)$  with the property that  $r(u) \parallel v_1(u)$  at least where  $\text{coh}(u) := \lambda_1(u) - \lambda_2(u) \geq 0$  is large enough. (Note that  $\text{coh}(u)$  does depend Lipschitz continuously on  $u$ , cf. e.g. Theorem of Wielandt-Hoffman in [26]).

To this end, we consider an interpolation  $I(p, q, t) : S^{d-1} \times S^{d-1} \times [0, 1] \rightarrow S^{d-1}$  between points  $p$  and  $q$  on the sphere, with the following properties:

- $I(p, q, t)$  is locally Lipschitz continuous w.r.t.  $p, q$  and  $t$ ,
- $I(p, q, 1) = p, I(p, q, 0) = q$ , and
- $|I(p, q, t) - q| \leq Ct$ .

For example, a steady rotation of vector  $p$  onto  $q$  suffices:

**Definition 2** For  $p, q \in S^{d-1}$  we define

$$I(p, q, t) := \cos(\alpha_0 t)q + \sin(\alpha_0 t)q^\perp, \quad (71)$$

where  $q^\perp$  is the orthogonal vector to  $q$  such that  $p, q, q^\perp$  lie on a plane,  $\langle p, q^\perp \rangle \geq 0$  (unique if  $p \not\parallel q$ ) and  $\alpha_0 = \alpha_0(p, q)$  is the angle between  $p$  and  $q$ .

**Lemma 3**  $I(p, q, t)$  as defined in (71) is locally Lipschitz continuous and satisfies  $|I(p, q, t) - q| \leq \pi|t|$ .

*Proof* Note that  $\alpha_0 = \alpha_0(p, q)$  can be calculated using arctan and thus is locally Lipschitz continuous w.r.t.  $p, q$  and  $t$ . Together with the smoothness of sin and cos the first claim follows. Moreover,  $|I(p, q, t) - q|$  is bounded by the arc length  $\alpha_0 t$  between  $I(p, q, t)$  and  $q$ , thus

$$|I(p, q, t) - q| \leq \alpha_0 t \leq \pi t. \quad (72)$$

□

**Lemma 4** Let  $q(u) : \mathbb{R}^d \rightarrow S^{d-1}$  be locally Lipschitz continuous. Moreover, let  $g : \mathbb{R}_0^+ \rightarrow [0, 1]$  be a locally Lipschitz continuous and increasing function, such that  $g(0) = 0$  and  $g(x) = 1$  for  $x \geq \nu > 0$ . Then,  $r(u) := I(v_1(u), q(u), g(\text{coh}(u)))$  is local Lipschitz continuous.

Note that we do not assume  $v_1(u)$  to be locally Lipschitz continuous.

*Proof* We distinguish between the cases  $\text{coh}(u) > 0$ , and  $\text{coh}(u) = 0$ .

In the first case,  $v_1(u)$  is an eigenvector to an isolated eigenvalue and thus depends locally Lipschitz continuously on  $J(u)$ , see [19]. Moreover  $\text{coh}(u)$  is Lipschitz continuous (cf. [26]). From the local Lipschitz continuity of  $I$ , the local Lipschitz continuity of  $r(u)$  follows.

In the second case,  $\text{coh}(u) = 0$ , for every  $v$  we find

$$\begin{aligned} |r(v) - r(u)| &= |I(v, q(v), g(\text{coh}(v))) - I(u, q(u), g(\text{coh}(u)))| \\ &= |I(v, q(v), g(\text{coh}(v))) - q(u)| \\ &\leq |I(v, q, g(\text{coh}(v))) - q(v)| + |q(v) - q(u)| \\ &\leq Cg(\text{coh}(v)) + |q(v) - q(u)| \\ &= C|g(\text{coh}(v) - g(\text{coh}(u)))| + |q(v) - q(u)|, \end{aligned} \quad (73)$$

and the local Lipschitz continuity of  $r(u)$  follows from the local Lipschitz continuity of  $g(\cdot)$ ,  $\text{coh}(\cdot)$  and  $q(\cdot)$ .  $\square$

*Remark 2* Recall that  $g(s) = 1$  for  $s \geq \nu$ . We have found a locally Lipschitz continuous approximation  $r(u)$  of  $v_1(u)$ , with the property, that  $r(u) = v_1(u)$ , if  $\text{coh}(u)$  becomes larger or equal to  $\nu$ .

## A.2 Projections

**Definition 3** We define the following closed and convex sets:

(i) **Rectangle:** For  $r \in S^1$ ,  $\alpha, \beta > 0$  we define

$$\mathcal{R}(r, \alpha, \beta) := \{p \in \mathbb{R}^2 : |r^\top p| \leq \alpha, |(r^\perp)^\top p| \leq \beta\}. \quad (74)$$

(ii) **Parallelogram:** For  $r_1, r_2 \in S^1$ ,  $r_1 \not\parallel r_2$ , and  $\alpha > 0$  let

$$\mathcal{P}(r_1, r_2, \alpha) := \{p \in \mathbb{R}^2 : |r_1^\top p| \leq \alpha, |r_2^\top p| \leq \alpha\}. \quad (75)$$

(iii) **Ellipsoid:** For  $r \in S^{d-1}$ ,  $d = 2, 3$  and  $\alpha, \beta > 0$  we define

$$\mathcal{E}(r, \alpha, \beta) := \left\{ q \in \mathbb{R}^d : \frac{1}{\beta^2} |r^\top q|^2 + \frac{1}{\alpha^2} \|(\text{Id} - rr^\top)q\|_2^2 \leq 1 \right\}. \quad (76)$$

**Lemma 5** Let  $\mathcal{R}(r, \alpha, \beta)$  and  $\mathcal{P}(r_1, r_2, \alpha)$  be defined as in Definition 3(i) and (ii).

Then,  $\Pi_{\mathcal{R}(r, \alpha, \beta)}(q)$  and  $\Pi_{\mathcal{P}(r_1, r_2, \alpha)}(q)$  are Lipschitz continuous w.r.t.  $r, \alpha$  and  $r_1, r_2, \alpha$ , respectively, with the Lipschitz constant depending on  $\alpha$  and  $q$ .

*Proof* We prove only the Lipschitz continuity of  $\Pi_{\mathcal{P}(r_1, r_2, \alpha)}(q)$ . The proof for  $\Pi_{\mathcal{R}(r, \alpha, \beta)}(q)$  is analogous.

On the local Lipschitz continuity of  $\Pi_{\mathcal{P}(r_1, r_2, \alpha)}(q)$  w.r.t.  $\alpha$ , we note that for fixed  $q$  and  $\alpha, \tilde{\alpha} > 0$ , the difference between  $\Pi_{\mathcal{P}(r_1, r_2, \alpha)}(q)$  and  $\Pi_{\mathcal{P}(r_1, r_2, \tilde{\alpha})}(q)$  is bounded by  $\sqrt{2}|\alpha - \tilde{\alpha}|$ .

Next, we show the Lipschitz continuity w.r.t.  $r_1, r_2$ . For  $q \in \mathcal{P}(r_1, r_2, \alpha)$ , we have  $\Pi_{\mathcal{P}(r_1, r_2, \alpha)}(q) = q$ , i.e. the projection is constant w.r.t.  $r_i, i = 1, 2$ . For  $q \notin \mathcal{P}(r_1, r_2, \alpha)$  the projection onto  $\mathcal{P}(r_1, r_2, \alpha)$  can be calculated as follows: Let  $j^* := \arg\min_{j=1, \dots, 4} \|q - \Pi_j(q)\|_2$ , where  $\Pi_j$  is the projection on the  $j$ -th side of the rectangle. Then,  $\Pi_{\mathcal{P}(r_1, r_2, \alpha)}(q) = \Pi_{j^*}(q)$ . If  $j^*$  changes to  $\tilde{j}^*$  by varying  $r_1, r_2$ , the transition from  $\Pi_{j^*}(q)$  to  $\Pi_{\tilde{j}^*}(q)$  is continuous. Thus, it remains to show that  $\Pi_j(q)$  is Lipschitz continuous w.r.t.  $r_1, r_2$ .

Each of the projections  $\Pi_j$  is a composition of the orthogonal projection onto a line and a projection from the line onto a line segment. E.g. for the side given by  $\{\alpha r_1 + t r_2 \mid t \in [-\alpha, \alpha]\}$ , we can calculate  $\Pi_j(q)$  as follows

$$\begin{aligned} q_0 &:= r_2^\top (q - \alpha r_1), \\ t_0 &:= \max(\min(q_0, \alpha), -\alpha), \\ \Pi_j(q) &:= \alpha r_1 + t_0 r_2. \end{aligned} \quad (77)$$

Thus, for two parameter sets  $r_1, r_2$  and  $\tilde{r}_1, \tilde{r}_2$

$$|\Pi_j(q) - \Pi_{\tilde{j}}(q)| \leq \alpha |r_1 - \tilde{r}_1| + \underbrace{|t_0|}_{\leq \alpha} |r_2 - \tilde{r}_2| + \underbrace{|\tilde{r}_2|}_{\leq 1} |t_0 - \tilde{t}_0| \quad (78)$$

and

$$\begin{aligned} |t_0 - \tilde{t}_0| &\leq |q_0 - \tilde{q}_0| = |r_2^\top (q - \alpha r_1) - \tilde{r}_2^\top (q - \alpha \tilde{r}_1)| \\ &\leq |r_2^\top (q - \alpha r_1) - \tilde{r}_2^\top (q - \alpha r_1)| + |\tilde{r}_2^\top (q - \alpha r_1) - \tilde{r}_2^\top (q - \alpha \tilde{r}_1)| \\ &\leq (|q| + \alpha) |r_2 - \tilde{r}_2| + \alpha |r_1 - \tilde{r}_1|. \end{aligned} \quad (79)$$

Inserting (79) into (78), we get

$$|\Pi_j(q) - \Pi_{\tilde{j}}(q)| \leq 2\alpha |r_1 - \tilde{r}_1| + (|q| + 2\alpha) |r_2 - \tilde{r}_2| \leq \sqrt{2}(|q| + 2\alpha) \|r - \tilde{r}\|_2. \quad (80)$$

This result generalizes to the three other sides of the parallelogram.

Finally, the transition between the cases  $q \in \mathcal{P}(r_1, r_2, \alpha)$  and  $q \notin \mathcal{P}(r_1, r_2, \alpha)$  is continuous and therefore  $\Pi_{\mathcal{P}(r_1, r_2, \alpha)}(q)$  is Lipschitz continuous.  $\square$

**Lemma 6** *Let  $\mathcal{E}(r, \alpha, \beta)$  be defined as in Definition 3 (iii). Then,  $\Pi_{\mathcal{E}(r, \alpha, \beta)}(q)$  is locally Lipschitz continuous w.r.t.  $r, \alpha, \beta$ .*

*Proof* The projection onto the ellipsoid  $\mathcal{E}(r, \alpha, \beta)$  can be expressed as a locally Lipschitz continuous function of  $r, \alpha, \beta$  and one distinct root of a polynomial function, see [7]. The coefficients of the polynomial are locally Lipschitz continuous w.r.t.  $r, \alpha, \beta$ . Thus, also the root is local Lipschitz continuous w.r.t.  $r, \alpha$  and  $\beta$ , see [3].  $\square$

## References

1. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing, 2nd edn. Springer (2006)
2. Aujol, J.F., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decomposition – modeling, algorithms, and parameter selection. *Int. J. Comp. Vision* **67**(1), 111–136 (2006)
3. Bronshtein, M.D.: Smoothness of roots of polynomials depending on parameters. *Siberian Mathematical Journal* **20**, 347–352 (1979)
4. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. *Simul* **4**, 490–530 (2005)
5. Catté, F., Lions, P.L., Morel, J.M., Coll, T.: Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.* **29**(1), 182–193 (1992)
6. Chan, D., Pang, T.: The generalized quasi-variational inequality problem. *Math. Operat. Res.* **7**(2), 211–222 (1982)
7. Eberly, D.: Distance from a point to an ellipse in 2D. Tech. rep., www.magic-software.com, Magic Software Inc. (2002)
8. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imag. Sci.* **3**(4), 1015–1046 (2010)
9. Gilboa, G., Darbon, J., Osher, S., Chan, T.: Nonlocal convex functionals for image regularization. Tech. Rep. 06-57, UCLA CAM report (2006)
10. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Model. Simul.* **7**(3), 1005–1028 (2008)
11. Gilboa, G., Sochen, N., Zeevi, Y.: Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *Image Processing, IEEE Transactions on* **11**(7), 689 – 703 (2002)
12. Grasmair, M., Lenzen, F.: Anisotropic total variation filtering. *Appl. Math. Optim.* **62**(3), 323–339 (2010)
13. Kervrann, C., Boulanger, J., Coupé, P.: Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In: Proceedings of the 1st international conference on Scale space and variational methods in computer vision, SSVM’07, pp. 520–532. Springer-Verlag (2007)
14. Kindermann, S., Osher, S., Jones, P.: Deblurring and denoising of images by nonlocal functionals. *Multiscale Model. Simul.* **4**(4), 1091–1115 (2005)
15. Lenzen, F., Becker, F., Lellmann, J., Petra, S., Schnörr, C.: Variational image denoising with adaptive constraint sets. In: Proceedings of the 3rd International Conference on Scale Space and Variational Methods in Computer Vision. Springer (2011)
16. Meyer, Y.: Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, *Univ. Lect. Series*, vol. 22. Amer. Math. Soc. (2001)

17. Nesterov, Y., Scrimali, L.: Solving strongly monotone variational and quasi-variational inequalities. *Discrete Contin. Dynam. Systems* **31**(4) (2006)
18. Peyr, G., Bougleux, S., Cohen, L.: Non-local regularization of inverse problems. In: D. Forsyth, P. Torr, A. Zisserman (eds.) *Computer Vision ECCV 2008, LNCS*, vol. 5304, pp. 57–68. Springer (2008)
19. Rellich, F.: Störungstheorie der Spektralzerlegung, I. *Math. Ann.* pp. 600–619 (1936)
20. Rockafellar, R., Wets, R.J.B.: *Variational Analysis*, 2nd edn. Springer (2004)
21. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)
22. Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F.: *Variational Methods in Imaging*. Springer (2009)
23. Schneider, R.: *Convex bodies: the Brunn-Minkowski theory*. Cambridge University Press (1993)
24. Steidl, G., Teuber, T.: Anisotropic smoothing using double orientations. In: *Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 477–489 (2009)
25. Weickert, J.: *Anisotropic Diffusion in Image Processing*. Teubner Verlag (1998)
26. Wilkinson, J.: *The algebraic eigenvalue problem*. Numerical Mathematics and Scientific Computation. Oxford University Press (1988)
27. Ziemer, W.: *Weakly Differentiable Functions*. Springer (1989)