

# Low Rank Priors for Color Image Regularization

Thomas Möllenhoff, Evgeny Strekalovskiy, Michael Möller, and Daniels Cremers

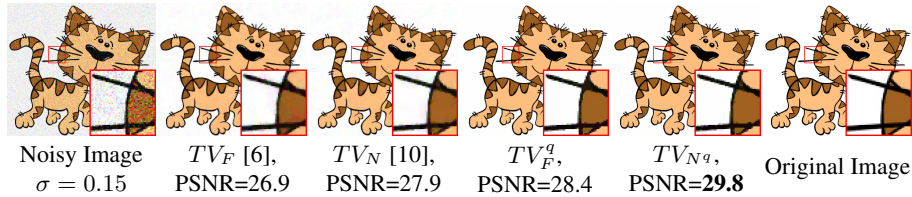
Department of Computer Science, Technical University of Munich, Germany

**Abstract.** In this work we consider the regularization of vectorial data such as color images. Based on the observation that edge alignment across image channels is a desirable prior for multichannel image restoration, we propose a novel scheme of minimizing the rank of the image Jacobian and extend this idea to second derivatives in the framework of total generalized variation. We compare the proposed convex and nonconvex relaxations of the rank function based on the Schatten- $q$  norm to previous color image regularizers and show in our numerical experiments that they have several desirable properties. In particular, the nonconvex relaxations lead to better preservation of discontinuities. The efficient minimization of energies involving nonconvex and nonsmooth regularizers is still an important open question. We extend a recently proposed primal-dual splitting approach for nonconvex optimization and show that it can be effectively used to minimize such energies. Furthermore, we propose a novel algorithm for efficiently evaluating the proximal mapping of the  $\ell^q$  norm appearing during optimization. We experimentally verify convergence of the proposed optimization method and show that it performs comparably to sequential convex programming.

## 1 Introduction

Developing effective image regularization priors is of central importance for variational image reconstruction methods and inverse problems. The *total variation (TV)* pioneered as a discontinuity-preserving regularizer [1], and still ranges among the most popular and versatile regularizers [2]. Since the classical total variation was proposed for grayscale images, a lot of recent research has focused on extending the TV to color images. Among these works are straightforward extensions of using TV regularization on each color channel separately [3], using a global coupling of the color channels by penalizing the  $\ell^2$  norm of the total variations of the channels [4], as well as using the Frobenius norm of the derivative matrix at each pixel [5, 6]. Additionally, it has been proposed to incorporate a change of color space [7], as well as to couple the color channels with an  $\ell^\infty$  norm [8].

Based on the class of methods presented by Sapiro and Ringach [5], the authors of [9] proposed the penalization of the Schatten- $\infty$  norm of the derivative matrix at each pixel, i.e. the penalization of the largest singular value of the Jacobian. One approach we are particularly interested in was also motivated by [5]: The authors of [10] suggested to penalize the Schatten-1 norm, also known as the nuclear norm, of the Jacobian at



**Fig. 1.** We propose a novel regularizer based on the nonconvex relaxation of the rank norm ( $TV_{N^q}$ ). The above comparison shows that the nonconvex regularizers (for values of  $q < 1$ , here  $q = 0.5$ ) outperform the convex ones, as they are able to better preserve discontinuities. The proposed regularizer has significantly less color artefacts at discontinuities as it favors coherent jumps of the color channels.

each pixel, i.e. they suggested to penalize

$$TV_N(u) := \|\nabla u\|_{N,1} := \int_{\Omega} \left\| \begin{pmatrix} \partial_x u_1(x, y) & \partial_x u_2(x, y) & \partial_x u_3(x, y) \\ \partial_y u_1(x, y) & \partial_y u_2(x, y) & \partial_y u_3(x, y) \end{pmatrix} \right\|_N dx dy, \quad (1)$$

for an image  $u : \Omega \rightarrow \mathbb{R}^3$ , where  $\|\cdot\|_N$  denotes the nuclear norm. Since the nuclear norm of the derivative is a convex relaxation for minimizing the rank of a matrix, we can interpret this approach as the rank minimization of the Jacobian. Note that the Jacobian being of rank one means that all gradient vectors are linearly dependent and thus point in the same (or opposite) direction. The latter is an interesting regularization property which has been exploited in other contexts such as nonlinear diffusion [11] or color Bregman iteration [12]. In this paper we propose a novel rank minimization of the derivative matrix through nonconvex relaxation by considering the penalization with the Schatten- $q$  norm for  $0 < q < 1$ .

Another motivation for such nonconvex relaxations comes from studies on the statistics of natural images. Filter responses are more faithfully represented by heavy-tailed distributions giving rise to nonconvex regularizers [13, 14]. This led to the work of Krishnan et al. [14], who demonstrated that standard  $TV$  denoising and deblurring results can indeed be improved by replacing the usual  $\ell^1$  norm of the gradient with the nonconvex  $\ell^q$  norm for  $q < 1$ .

While penalizing the nuclear norm instead of the Frobenius norm of the Jacobian yields an improvement as shown by Lefkimiatis et al. [10], and replacing the usual  $TV$ - $\ell^1$  norm by a  $TV$ - $\ell^q$  norm with  $q < 1$  yields another improvement [14], we will demonstrate that combining both ideas by replacing the nuclear norm with a Schatten- $q$  norm for  $q < 1$  leads to a regularization method superior to both previously mentioned approaches, as illustrated in Fig. 1.

One well known property of total variation regularization is the preference of piecewise constant images which can lead to so called staircasing effects. To avoid these artifacts, higher order methods such as the total generalized variation (TGV) have been proposed [15]. The  $TGV_2^\alpha$  model on a grayscale image  $u : \Omega \rightarrow \mathbb{R}$  can be interpreted as a particular type of infimal convolution written as

$$TGV(u) = \inf_{\nabla u = v + z} \alpha \|v\|_{2,1} + (1 - \alpha) \|\nabla z\|_{2,1} \quad (2)$$

where  $\|v\|_{2,1} = \int_{\Omega} \sqrt{v_1(x)^2 + v_2(x)^2} dx$  and  $\alpha \in [0, 1]$  is a weighting parameter between first and second order penalization. Extensions for the TGV model include replacing the  $\|\cdot\|_{2,1}$  norms by nonconvex  $\|\cdot\|_{2,q}^q$  penalty functions on grayscale images [16], as well as extending the TGV model to color images by considering the Frobenius norms of the derivative matrices arising from having different color channels [17].

In this paper we propose a novel extension of the TGV approach to color images by considering the nuclear norm of the derivative matrices. We will demonstrate in the numerical results that our convex nuclear norm TGV approach outperforms the Frobenius norm TGV for color image denoising. Moreover, we show that, again, replacing the nuclear norm by a Schatten  $q$ -norm with  $q < 1$  can improve the denoising performance even further.

The minimization of the resulting nonsmooth and nonconvex energy is a challenging task. We will use a recent reformulation of the primal-dual hybrid gradient method [18–20], which makes it applicable to nonconvex energies [21]. Although a full convergence theory has not yet been established, we demonstrate that one obtains a very efficient numerical scheme for finding low energies, comparably to methods which rely on sequential convex programming such as [16].

The rest of this paper is organized as follows. In the next section we will further motivate the idea of penalizing the Schatten- $q$  norm of the derivative matrices in the TV as well as in the TGV case in greater detail. Section 3 discusses the numerical method for minimizing the proposed energies in detail. Particular emphasis is put on the efficient evaluation of the proximity operators of the  $\ell^q$  seminorms for  $q < 1$ . The numerical results in Section 4 demonstrate the superior behavior of derivative matrix rank minimization in the TV as well as in the TGV case and demonstrate the advantages of the nonconvex regularizations. Finally, we draw conclusions and point out directions of future research in Section 5.

## 2 TV and TGV Rank Minimization Approaches

In this section we will give more details on the idea and motivation for considering certain Schatten- $q$  norms for  $q < 1$ . We define the Schatten- $q$  “norm” as

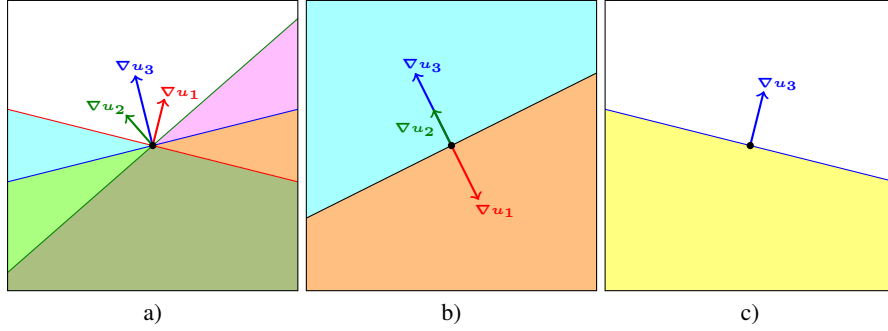
$$\|A\|_{N_q} := (\sigma_1^q + \dots + \sigma_n^q)^{1/q}, \quad (3)$$

where  $\sigma_i$  denotes the  $i$ -th singular value of  $A$ . Note that as a special case we obtain the rank function for  $q = 0$  (using the convention  $0^0 = 1$ ) and the nuclear norm for  $q = 1$ .

As pointed out in the introduction, the nuclear norm TV penalty (1) can be interpreted as a convex relaxation for encouraging a low rank of the Jacobian at each pixel. Our proposed Schatten- $q$  norm approximates the rank minimization, i.e. the penalty of the number of nonzero singular values, more closely.

$$TV_{N_q}(u) := \int_{\Omega} \left\| \begin{pmatrix} \partial_x u_1(x, y) & \partial_x u_2(x, y) & \partial_x u_3(x, y) \\ \partial_y u_1(x, y) & \partial_y u_2(x, y) & \partial_y u_3(x, y) \end{pmatrix} \right\|_{N_q}^q dx dy, \quad (4)$$

But why does it make sense to minimize the rank of this matrix? Note that the derivative matrix has at most rank two. A reduction of the rank could lead to a rank



**Fig. 2.** Illustration of a point with gradient matrix of rank two **a)** and two different matrices with rank one in **b)** and **c)**. A Jacobian of rank zero would correspond to a locally constant region.

zero, which has the simple interpretation of all derivatives being zero, i.e. none of the channels changing. We therefore still expect the regularization to prefer piecewise constant images. A derivative matrix with rank one on the other hand has the interpretation that all gradient vectors are linearly dependent and hence parallel (or antiparallel).

This is illustrated in Fig. 2, where on the left we show a rank two Jacobian and on the right two different rank one Jacobians. Note that the gradients always point in the normal direction to the level lines of each channel, such that the lines in Fig. 2 can be interpreted as particular level lines of the channels. The alignment of the normal lines in all channels seems to be a reasonable regularity assumption for natural images and leads to a reduction of color artifacts as we will see in the numerical results on color image denoising. As illustrated in the right image in Fig. 2, a derivative matrix with two derivative vectors being zero and one derivative vector being arbitrary also has rank one such that color edges are not necessarily forced to be aligned as in the middle image. We expect that the data term decides whether a full alignment as in the middle or a pointwise alignment as in the right image of Fig. 2 are to be preferred, such that we avoid overregularization or the introduction of artificial edges.

Furthermore, we propose to extend the idea of rank penalization of the derivatives to the TGV framework by minimizing

$$\begin{aligned}
 TGV_{N_q}(u) &:= \inf_{\nabla u=v+z} \alpha \int_{\Omega} \left\| \begin{pmatrix} v_{1,1}(x,y) & v_{1,2}(x,y) & v_{1,3}(x,y) \\ v_{2,1}(x,y) & v_{2,2}(x,y) & v_{2,3}(x,y) \end{pmatrix} \right\|_{N_q}^q dx dy \\
 &+ (1-\alpha) \int_{\Omega} \left\| \begin{pmatrix} \partial_x z_{1,1}(x,y) & \partial_x z_{1,2}(x,y) & \partial_x z_{1,3}(x,y) \\ \partial_x z_{2,1}(x,y) & \partial_x z_{2,2}(x,y) & \partial_x z_{2,3}(x,y) \\ \partial_y z_{1,1}(x,y) & \partial_y z_{1,2}(x,y) & \partial_y z_{1,3}(x,y) \\ \partial_y z_{2,1}(x,y) & \partial_y z_{2,2}(x,y) & \partial_y z_{2,3}(x,y) \end{pmatrix} \right\|_{N_q}^q dx dy. \quad (5)
 \end{aligned}$$

The above penalization can be motivated as follows. The Jacobian  $\nabla u$  of an image is optimally divided into two parts. The first part corresponds to  $v$  where the Schatten- $q$  norm of  $v$  is penalized. Thus, the interpretation of  $v$  is similar to the plain TV case discussed above: This part of the gradient of  $u$  should point in the same direction for

all color channels. The second part of the functional penalizes the derivatives of  $z$  in the Schatten- $q$  norm and might be more difficult to interpret at first sight. The variable  $z$  contains parts of the Jacobian of  $u$ . For interpretation purposes let us assume that  $z = \nabla u$ . Then each column of the matrix in the second term of (5) is exactly the Hessian of one of the color channels. In this sense, the second term tries to align parts of the Hessian matrices of the color channels and therefore is the natural extension of aligning the first derivatives. Having the interpretation of a Hessian in mind, one could also motivate our approach by considering an image  $u$  whose color channels are twice continuously differentiable. In this case a second order Taylor expansion could describe the local behavior of each channel. Particularly, for color channels with parallel Hessians the second order behavior or the curvature of all color channels is the same up to a scaling and thus extends the coupling of different color channels from the first derivatives in the TV case to the second derivatives in the TGV case.

### 3 Application to Inverse Problems in Image Processing

We now consider inverse problems involving the proposed regularizers, given an input image  $f : \Omega \rightarrow \mathbb{R}^k$  with  $k$  channels on a  $d$ -dimensional discretized domain  $\Omega$ . For regularization of piecewise constant images we have the following variational problem

$$\min_u \frac{\lambda}{2} \|u - f\|^2 + \mathcal{R}(u), \quad (6)$$

where  $\mathcal{R}(u)$  is either  $TV_F^q(u) = \|\nabla u\|_{2,q}^q = \int_{\Omega} \|\nabla u(x)\|_2^q dx$  or  $TV_{N^q}(u)$  as defined in (4). For inverse problems involving *piecewise affine* and natural images we propose

$$\min_{u,v} \frac{\lambda}{2} \|u - f\|^2 + \mathcal{R}(u, v), \quad (7)$$

where  $\mathcal{R}(u, v)$  is  $TGV_F^q(u, v) = \alpha \|\nabla u - v\|_{2,q}^q + (1 - \alpha) \|\nabla_2 v\|_{2,q}^q$  or  $TGV_{N^q}(u, v)$  as defined in (5). Since for  $q < 1$  the regularizers are nonconvex and nonsmooth, their efficient numerical optimization is a challenging problem. In the next section we propose a minimization algorithm for energies involving the proposed regularizers.

#### 3.1 Splitting Methods in the Nonconvex Setting

Let us first introduce the *proximal mapping* associated with a proper, lower-semicontinuous function  $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ :

$$\text{prox}_{\tau, f}(y) := \arg \min_x f(x) + \frac{1}{2\tau} \|x - y\|^2. \quad (8)$$

Note that if  $f$  is *nonconvex*, this mapping is not necessarily single-valued.

It has recently been shown experimentally that primal-dual splitting methods for convex optimization are also often applicable in the nonconvex setting [21, 22]. Here we show how to generalize the recent approach [21] to our setting. In general, we aim to minimize cost functions of the form

$$\min_u G(u) + F(g) \quad \text{subject to} \quad Ku = g, \quad (9)$$

where  $G$  is convex,  $F$  possibly nonconvex and  $K$  a linear operator. The algorithm studied in [21, 23] is given as

$$\begin{aligned} g^{n+1} &\in \text{prox}_{\sigma^{-1}, F} (K\bar{u}^n + \sigma^{-1}q^n), \\ q^{n+1} &= q^n + \sigma(K\bar{u}^n - g^{n+1}), \\ u^{n+1} &= \text{prox}_{\tau, G} (u^n - \tau K^T q^n), \\ \bar{u}^{n+1} &= u^{n+1} + \theta(u^{n+1} - u^n), \end{aligned} \tag{10}$$

and reduces to the primal-dual hybrid gradient method (cf. [20]) for convex  $F$ . Interestingly, this update scheme can also be interpreted as gradient descent in the primal variables  $u$  and  $g$  and gradient ascent in the dual variable  $q$  on the following Lagrangian saddle-point formulation of (9):

$$\max_q \min_{u, g} G(u) + F(g) + \langle q, Ku - g \rangle. \tag{PD}$$

Since for nonconvex  $F$  it is generally not possible to interchange min and max, this is not the same as the primal-dual saddle point problem involving the Fenchel dual  $F^*$  from [2]. As observed in [23], a necessary condition on the dual step size for the algorithm to converge for *semiconvexity*  $F$ , i.e. for  $F$  with the property that  $F(u) + \frac{\omega}{2}\|u\|^2$  is convex, seems to be  $\sigma \geq 2\omega$ .

**Adaptive Step-Size Scheme** As the  $\ell^q$  seminorms are neither semiconvex nor differentiable for  $q < 1$  one possibility would be to approximate it by a regularized or smoothed variant. However, this turns out to be difficult for the nonconvex relaxations of the rank function.

Instead we opt to employ a variable step size scheme where the dual step size approaches infinity ( $\sigma \rightarrow \infty$ ) as suggested in [21]:

$$\theta_n = 1/\sqrt{1 + 2\gamma\tau_n}, \sigma_{n+1} = \sigma_n/\theta_n, \tau_{n+1} = \tau_n\theta_n, \tag{11}$$

with  $\tau_0\sigma_0\|K\|^2 < 1$ . Here  $\gamma$  is an additional parameter which is usually chosen according to the strong convexity constant of  $G$ , e.g.  $\gamma = \lambda$  for  $G(u) = \frac{\lambda}{2}\|u - f\|^2$ . In the case of  $TGV$  regularization, the function  $G$  is not strongly convex due to the additional primal variable. We still pick  $\gamma = \lambda$  as a heuristical choice, as it works well in practice.

A similar approach is suggested by Storath et al. [22] for minimizing the Potts model, based on the direct application of the Alternating Direction Method of Multipliers (ADMM) to the nonconvex  $\ell^0$  regularizer while having a similar step size scheme where the penalty parameter in the ADMM method approaches infinity. Here, the adaptive step size scheme for the above primal-dual algorithm comes with an immediate interpretation in the *convex* setting for strongly convex  $G$ .

As the adaptive step size scheme yields  $\sigma_n \rightarrow \infty$ , the following interpretation is interesting. By considering the optimality conditions of the iterates produced by Algorithm (10) we see that the inclusion  $q^n \in \partial F(g^n)$  holds in every iteration (cf. [23]).

For *differentiable*  $F$  the subdifferential is a singleton and we can thus eliminate the variable  $q$ , and retrieve a formulation in terms of primal variables:

$$\begin{aligned} g^{n+1} &\in \text{prox}_{\sigma_n^{-1}, F} (K\bar{u}^n + \sigma_n^{-1}\nabla F(g^n)), \\ u^{n+1} &= \text{prox}_{\tau_n, G} (u^n - \tau_n K^T \nabla F(g^n)), \\ \bar{u}^{n+1} &= u^{n+1} + \theta_n (u^{n+1} - u^n). \end{aligned} \quad (12)$$

As  $\sigma_n \rightarrow \infty$  the proximity operator in the update in  $g$  becomes the identity and it can be seen that we approach the forward-backward splitting algorithm in the limit, applied to the primal problem (9) with an additional inertial term on  $u$ . This relation is interesting as convergence of forward-backward methods in the nonconvex setting is theoretically proven [24].

In order to apply Algorithm (10) to the problems at hand we require the efficient evaluation of the proximal mappings coming from the nonconvex regularizers, which is a nonconvex optimization problem itself.

### 3.2 Evaluation of the $\|\cdot\|_{2,q}^q$ Proximal Mapping

First we note that due to the separability, the proximal mappings reduce to pointwise evaluations of  $\|\cdot\|_2^q$ . We will focus on the former case first. Given  $g_0 \in \mathbb{R}^{d \times k}$ , we will consider the efficient minimization of the following proximal mapping for  $0 \leq q < 1$ :

$$\text{prox}_{\tau, \|\cdot\|_2^q}(g_0) = \arg \min_{g \in \mathbb{R}^{d \times k}} \frac{\|g - g_0\|_2^2}{2\tau} + \|g\|_2^q. \quad (13)$$

**The case  $q = 0$ .** An important special case is  $q = 0$ , which corresponds to Potts regularization. In this case the minimization (13) can actually be solved explicitly via *hard thresholding*:

$$\text{prox}_{\tau, \|\cdot\|_2^0}(g_0) = \begin{cases} 0 & \text{if } \|g_0\|_2 \leq \sqrt{2\tau}, \\ g_0 & \text{otherwise.} \end{cases} \quad (14)$$

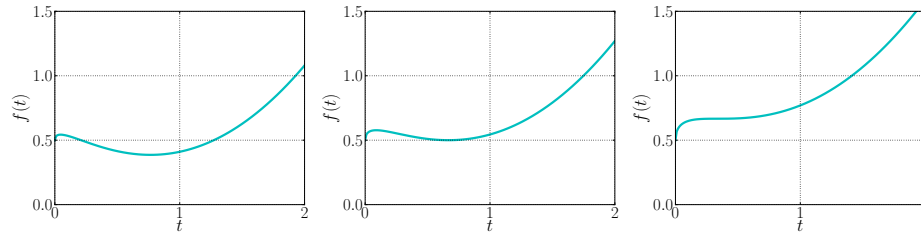
For the general case  $0 < q < 1$ , we first note that the evaluation of the proximal operator (13) can be reduced to a scalar problem.

**Proposition 1** *Given  $g_0 \in \mathbb{R}^{d \times k}$ ,  $\tau > 0$ ,  $q \in (0, 1)$  and  $\lambda > 0$ , the solution of the proximal operator*

$$\text{prox}_{\tau, \|\cdot\|_2^q}(g_0) = \arg \min_{g \in \mathbb{R}^{d \times k}} \frac{\|g - g_0\|_2^2}{2\tau} + \|g\|_2^q$$

*has the form  $\hat{g} = tg_0$  for some real  $t \geq 0$ .*

*Proof.* A proof is given in the appendix of [25].



**Fig. 3.** The function (15) for three different values of  $\alpha$  and  $q = 0.5$ . **From left to right:**  $\alpha \approx 0.41 < \frac{2\sqrt{6}}{9}$ ,  $\alpha = \frac{2\sqrt{6}}{9} \approx 0.54$ ,  $\alpha = \frac{4}{3\sqrt{3}} \approx 0.77$ . It can be seen that the desired stationary point is also the global minimum for  $\alpha < \frac{2\sqrt{6}}{9}$  (left). In the equality case, the value at the stationary point is the same as the boundary value (center). For  $\alpha > \frac{4}{3\sqrt{3}}$  the function is increasing (right).

**Solving the Scalar Problem.** Since we now know that the optimal solution is a scalar multiple of  $g_0$  we substitute  $g = tg_0$  in (13) and arrive at the following problem

$$\arg \min_{t \geq 0} \frac{(t-1)^2}{2} + \alpha t^q =: \arg \min_{t \geq 0} f(t) \quad (15)$$

for  $\alpha = \tau \|g_0\|_2^{q-2} \geq 0$ . Thus, evaluating the proximal operator (13) reduces to solving the above problem (15) for  $t \geq 0$ .

The minimization problem (15) can be solved in closed form for certain values of  $q$  such as  $1/2$  or  $3/4$  as described in [14]. In the following, we provide a more concise analytic solution for the special case  $1/2$  and an efficient algorithm based on Newton's method for the general case.

*Concise Closed Form Solution for  $q = 1/2$ .* Setting the derivative of the cost function (15) to 0 and substituting  $t = s^2$  we arrive at the cubic equation

$$s^3 - s + \frac{\alpha}{2} = 0. \quad (16)$$

Following the work of [26] we arrive at the following closed form expression for the root which corresponds to the minimum of (15):

$$\hat{s} = \frac{2}{\sqrt{3}} \sin\left(\frac{1}{3}\left(\arccos\left(\frac{3\sqrt{3}}{4}\alpha\right) + \frac{\pi}{2}\right)\right). \quad (17)$$

Interestingly this solution based on trigonometric expressions is quite a bit shorter than the one proposed in [14]. As shown in Fig. 3, we see that for some values of  $\alpha$  the value at the boundary is the optimal value. This is precisely for all  $\alpha$  satisfying the condition

$$\alpha > \frac{2\sqrt{6}}{9}. \quad (18)$$

If (18) is satisfied we simply set  $\hat{t} = 0$ , otherwise we find the root  $\hat{s}$  of (16) that corresponds to the local minimum using formula (17) and set  $\hat{t} = \hat{s}^2$ . As seen in Fig. 3, for  $\alpha > \frac{4}{3\sqrt{3}}$ , the function is increasing and does not have a stationary point. This corresponds to the case where the root in (17) is not real anymore.



---

**Algorithm 1:** Newton's method for solving the nonconvex  $\ell^q$  proximal operator.
 

---

**Input:** Parameters  $g_0 \in \mathbb{R}^{d \times k}$ ,  $\tau > 0$  and  $q \in (0, 1)$ , machine precision  $\varepsilon > 0$   
**Output:** Minimizer  $\hat{g} \in \mathbb{R}^{d \times k}$  of (13)

```

if  $\|g_0\|_2 > 0$  then
   $\alpha = \tau \|g_0\|_2^{q-2}$ 
  if  $\alpha$  satisfies (19) then
     $\hat{t} = 0$ 
  else
    // Solve for optimum using Newton's method.
     $t_0 = 1$ 
    for  $k \geq 1$  until  $f'(t_k)/f''(t_k) < \varepsilon$  do
       $t_k = t_{k-1} - f'(t_{k-1})/f''(t_{k-1})$ 
     $\hat{t} = t_k$ 
     $\hat{g} = \hat{t} g_0$ 
  else
     $\hat{g} = 0$  // In the case  $g_0 = 0$  we can just set  $\hat{g} = 0$ .
  
```

---

Newton's method for general  $0 < q < 1$ . For general values of  $q$ , we solve the scalar  $\ell_q$  problem (15) using Newton's method. For all  $\alpha$  satisfying the condition

$$\alpha > \frac{1}{2-q} \left( 2 \frac{1-q}{2-q} \right)^{1-q} \quad (19)$$

the boundary value is lower than the value at the local minimum as shown for  $q = 0.5$  in Fig. 3, so we set  $\hat{t} = 0$  if (19) is satisfied and otherwise we use Newton's method. For that we note that for  $\alpha = 0$  the optimal point is at  $\hat{t} = 1$ , and for  $\alpha > 0$  we have  $\hat{t} < 1$ . So we pick the starting point for Newton's method  $t_0 = 1$ . We perform the iteration

$$t_{k+1} = t_k - f'(t_k)/f''(t_k) \quad (20)$$

where  $f'$  and  $f''$  denote the first and second derivatives of  $f$ . It can be shown that the derivative  $f'$  is convex and increasing on the closed interval  $[\hat{t}, 1]$ , so Newton's method always converges to the minimum. The final algorithm to evaluate the proximal operator (13) for  $0 < q < 1$  is given as Algorithm 1.

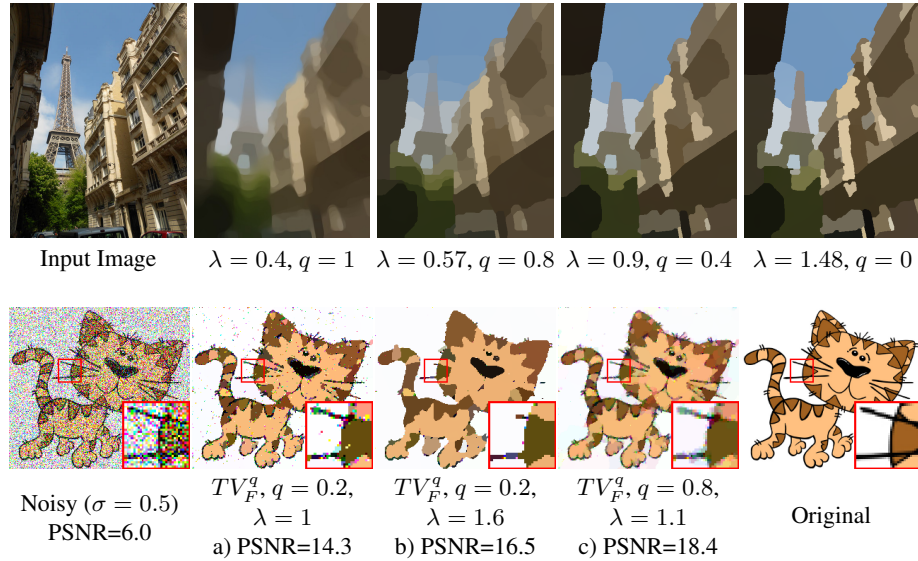
### 3.3 Evaluation of the $\|\cdot\|_{N_q}^q$ Proximal Mapping

Similar to the previous section, due to the separability we are only interested in the pointwise evaluation of  $\|\cdot\|_{N_q}^q$ . Given  $g_0 \in \mathbb{R}^{d \times k}$  we wish to evaluate the proximal mapping

$$\text{prox}_{\tau, \|\cdot\|_{N_q}^q}(g_0) = \arg \min_{g \in \mathbb{R}^{d \times k}} \|g\|_{N_q}^q + \frac{1}{2\tau} \|g - g_0\|_2^2. \quad (21)$$

In order to do so, we start with the singular value decomposition of the input argument  $g_0 = U \Sigma_{g_0} V^T$  and substitute that into (21):

$$\arg \min_{g \in \mathbb{R}^{d \times k}} \|g\|_{N_q}^q + \frac{1}{2\tau} \|g - U \Sigma_{g_0} V^T\|_2^2. \quad (22)$$



**Fig. 4.** Effect of the parameter  $q$  illustrated on a color image and a denoising example. Values of  $q < 1$  lead to piecewise constant results and smaller values of  $q$  lead to higher contrast between the regions. In the second row we show the effect of the parameter  $q$  for image denoising. **a), b)** While smaller values of  $q$  lead to sharp boundaries and clearer regions, large noise outliers are not being removed since big jumps get penalized less. **c)** For such high noise levels we found values around  $q \approx 0.8$  to give the highest PSNR values as it describes a good trade-off.

Since the functions  $\|\cdot\|_{N_q}^q$  and  $\|\cdot\|_2$  are unitarily invariant, the optimization problem can be reduced to the following:

$$\arg \min_{\Sigma \in \mathbb{R}^{d \times k}} \|\Sigma\|_{N_q}^q + \frac{1}{2\tau} \|\Sigma - \Sigma_{g_0}\|_2^2, \quad (23)$$

where  $\Sigma \in \mathbb{R}^{d \times k}$  is a diagonal matrix. We can restrict the optimization problem (23) to diagonal matrices due to a result by Mirsky [27, Theorem 5].

As this minimization problem is separable, we can compute the  $\ell^q$  proximal mapping for each singular value in  $\Sigma_{g_0}$ . Given the solutions  $\hat{\Sigma}$  to the previous problem (23), the final solution  $\hat{g} \in \mathbb{R}^{d \times k}$  is recovered as

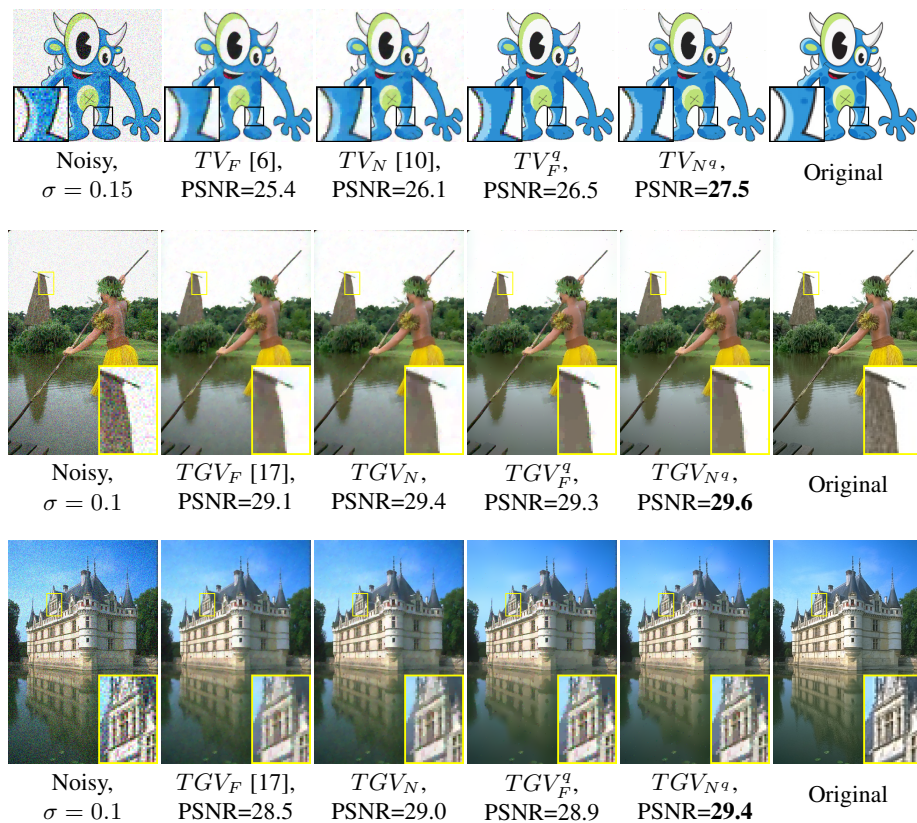
$$\hat{g} = U \hat{\Sigma} V^T = g_0 V \Sigma_{g_0}^+ \hat{\Sigma} V^T, \quad (24)$$

where  $\Sigma_{g_0}^+$  denotes the pseudoinverse of  $\Sigma_{g_0}$ . Note that it is not required to calculate a full singular value decomposition of  $g_0$ , but just the eigenvalue decomposition of  $g_0^T g_0$  to obtain  $V$ . In case of  $TGV_{N_q}$ , this is an eigenvalue decomposition of a  $3 \times 3$  matrix<sup>1</sup>.

## 4 Numerical Experiments

For all experiments we initialized the primal and dual variables ( $u$ ,  $g$  and  $q$ ) with zero.

<sup>1</sup> Efficient evaluation: <http://www.mpi-hd.mpg.de/personalhomes/globes/3x3/>

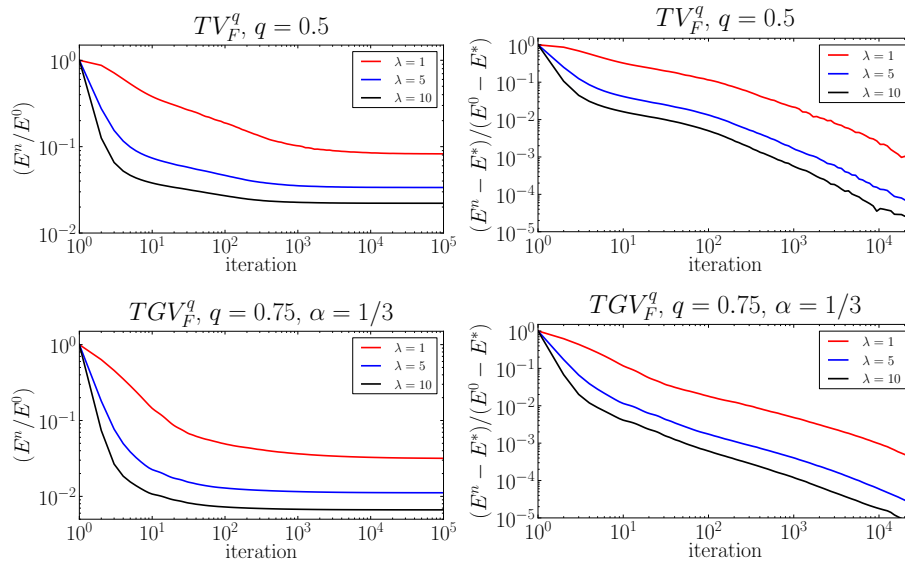


**Fig. 5.** Denoising of a piecewise constant image and natural images with  $TV^q$  ( $q = 0.5$ ) and  $TGV^q$  regularization ( $q = 0.75$ ) using the proposed primal-dual algorithm. We chose  $\alpha = 1/3$  for all the  $TGV$  experiments and the data fidelity parameters were optimized for maximal PSNR.

#### 4.1 Effect of the Parameter $q$ in the $TV_F^q$ Model

In Fig. 4 we show the effect of the parameter  $q$  on a natural image for the  $TV_F^q$  model. Values of  $q < 1$  lead to piecewise constant approximations and for smaller values of  $q$  we observe higher contrast between the regions. That is because for smaller values of  $q$ , bigger jumps are penalized less and less until for  $q = 0$  all jumps are penalized equally. Note that the proposed algorithm produces consistent results in a sense that smaller values of  $q$  systematically lead to a higher contrast between the regions.

In the second row in Fig. 4 we illustrate how the denoising performance of the algorithm depends on the parameter  $q$ . While smaller values of  $q$  lead to desirable sharper boundaries and higher contrast, strong noise outliers do not get removed anymore due to the lower penalization of large jumps. Finding the correct value of  $q$  means finding a good trade off and values of  $q \approx 0.8$  lead to the highest PSNR for this particular noise level.



**Fig. 6.** Experimental convergence of the proposed algorithm on the  $256 \times 256$  RGB lena image for  $TV_F^q$  and  $TGV_F^q$  regularization and different data fidelities  $\lambda$ . We observe convergence for both the normalized energy  $(E^n - E^*)/(E^0 - E^*)$  and the normalized energy  $E^n/E^0$ . Similar convergence results are to be expected for the  $TV_{N^q}$  and  $TGV_{N^q}$  cases.

## 4.2 Denoising of Piecewise Constant Images with $TV^q$ Regularization

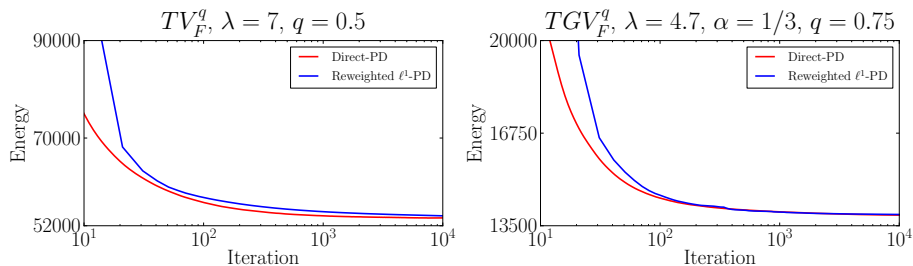
In the first row of Fig. 5 and in Fig. 1 the denoising performance of the different regularizers on a piecewise constant image is shown. We chose  $q = 0.5$  and the hyperparameter  $\lambda$  was chosen in order to obtain the highest PSNR values. It can be seen that the nuclear norm reduces color artifacts at the jumps significantly and the use of nonconvex norms leads to less contrast loss and yields sharper discontinuities. Combining both aspects yields the overall highest PSNR and an improvement of 2 – 3 PSNR values over the baseline approach [6] in Fig. 1 and Fig. 5.

## 4.3 Denoising of Natural Images with $TGV^q$ Regularization

In the second and third row of Fig.5 we show the result of the proposed algorithm applied to the  $TGV$ -denoising functional for the different variants of  $TGV$  regularization. The data fidelity parameter  $\lambda$  was tuned for maximal PSNR. Again we see that nonconvex  $TGV^q$  yields sharper discontinuities and higher PSNR values while the use of the nuclear norm reduces color artefacts. The combination yields an improvement of at least  $1/2$  PSNR over [17] in the experiments in Fig. 5.

## 4.4 Convergence of the Energy

As the theoretical convergence of the algorithm is still an important open question we validated the convergence of the algorithm experimentally by precomputing a  $u^* = u^{10^5}$  as an approximation to the converged solution. It can be seen in Fig. 6 that the normalized energies  $(E(u^n) - E(u^*)) / (E(u^0) - E(u^*))$  and  $E(u^n)/E(u^0)$  converge.



**Fig. 7.** We show the energy decrease over iterations (total inner iterations for the iterative reweighted  $\ell_1$  algorithm) for the  $TV_F^q$  (left) and  $TGV_F^q$  (right) denoising examples in Fig. 1 and Fig. 5. Our proposed direct algorithm minimizes the energy functional comparably to the state-of-the-art iterative reweighted  $\ell^1$  algorithm [16].

We compare the energy decrease of the proposed method over iterations to iterative reweighted  $\ell_1$  (IRL1) optimization [16], and show the results in Fig. 7. For the iterative reweighting method we chose the smoothing parameter  $\varepsilon = 10^{-6}$  as a regularization parameter to make the  $\ell^q$  function Lipschitz continuous. The inner convex optimization problem is solved using the same primal-dual algorithm (but of course in the convex setting) and uses the same termination criterion for the inner iterations as detailed in [16]. We see that the direct application of the primal-dual method in the nonconvex setting performs overall comparably to the state-of-the-art iterative reweighted  $\ell^1$  method.

## 5 Conclusion

We proposed novel regularizers for vector valued images based on convex and nonconvex relaxations of a rank minimization prior. Numerous experiments on piecewise constant and natural images show that the proposed regularizers yield overall state-of-the-art performance.

Furthermore, to deal with the nonconvex and nonsmooth optimization problem an efficient optimization method for solving related inverse problems was presented. We have shown how to efficiently find globally optimal solutions to the arising nonconvex proximal mapping. Our numerical experiments indicated that the direct application of a primal-dual splitting method in the nonconvex setting performs comparably to sequential convex programming methods. For future work we mainly wish to study the convergence properties of convex splitting methods in the nonconvex setting.

## References

1. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60** (1992) 259–268
2. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. In: *Theoretical Foundations and Numerical Methods for Sparse Recovery*. De Gruyter (2010)
3. Attouch, H., Buttazzo, G., Michaille, G.: *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization* (Mps-Siam Series on Optimization 6). SIAM (2005)
4. Blomgren, P., Chan, T.F.: Color TV: Total variation methods for restoration of vector valued images. *IEEE Trans. Image Processing* **7** (1998) 304–309

5. Sapiro, G., Ringach, D.: Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. Img. Proc.* **5**(11) (1996) 1582–1586
6. Bresson, X., Chan, T.F.: Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging* **2**(4) (2008) 255–284
7. Condat, C.: Joint demosaicking and denoising by total variation minimization. In: *IEEE Conference on Image Processing*. (2012) 2781–2784
8. Miyata, T., Sakai, Y.: Vectorized total variation defined by weighted l infinity norm for utilizing inter channel dependency. In: *Image Processing (ICIP), 2012 19th IEEE International Conference on*. (Sept 2012) 3057–3060
9. Goldluecke, B., Cremers, D.: An approach to vectorial total variation based on geometric measure theory. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2010)
10. Lefkimmatis, S., Roussos, A., Unser, M., Maragos, P.: Convex generalizations of total variation based on the structure tensor with applications to inverse problems. In: *Scale Space and Variational Methods in Computer Vision*. Volume 7893 of LNCS. Springer (2013) 48–60
11. Ehrhardt, M.J., Arridge, S.: Vector-valued image processing by parallel level sets. *IEEE Trans. on Image Processing* **23** (2014) 9 – 18
12. Moeller, M., Brinkmann, E., Burger, M., Seybold, T.: Color bregman tv Preprint. On ArXiv <http://arxiv.org/abs/1310.3146>.
13. Huang, J., Mumford, D.: Statistics of natural images and models. In: *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*. (1999)
14. D. Krishnan and R. Fergus: Fast Image Deconvolution using Hyper-Laplacian Priors. In: *Proc. Neural Information Processing Systems*. (2009) 1033–1041
15. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Img. Sci.* **3**(3) (September 2010) 492–526
16. Ochs, P., Dosovitskiy, A., Pock, T., Brox, T.: An iterated L1 Algorithm for Non-smooth Non-convex Optimization in Computer Vision. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2013)
17. Bredies, K.: Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty. In: *Efficient Algorithms for Global Optimization Methods in Computer Vision*. Springer (2014) 44–77
18. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In: *IEEE Int. Conf. on Comp. Vis. (ICCV)*. (2009)
19. 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. Img. Sci.* **3**(4) (2010) 1015–1046
20. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40** (2011) 120–145
21. Strelakovsky, E., Cremers, D.: Real-Time Minimization of the Piecewise Smooth Mumford-Shah Functional. In: *Proceedings of the European Conference on Computer Vision*. (2014)
22. Storath, M., Weinmann, A., Demaret, L.: Jump-sparse and sparse recovery using potts functionals. *CoRR* (2013) –1–1
23. Möllenhoff, T., Strelakovsky, E., Möller M., Cremers, D.: The Primal-Dual Hybrid Gradient Method for Semiconvex Splittings. preprint <http://arxiv.org/abs/1407.1723> (2014)
24. Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: Inertial Proximal Algorithm for Non-convex Optimization. *SIAM Journal on Imaging Sciences (SIIMS)* (2014) Preprint.
25. Bouaziz, S., Tagliasacchi, A., Pauly, M.: Sparse Iterative Closest Point. *Computer Graphics Forum (Symposium on Geometry Processing)* **32**(5) (2013) 1–11
26. McKelvey, J.P.: Simple transcendental expressions for the roots of cubic equations. *Amer. J. Phys.* **52**(3) (March 1984) 269–270
27. Mirsky L.: Symmetric gauge functions and unitarily invariant norms. *Quart. J. Math. Oxford Ser. (2)* (1960) 50–59