A duality-based splitting method for ℓ^1-TV image restoration with automatic regularization parameter choice[*]

Christian Clason[†]   Bangti Jin[‡]   Karl Kunisch[†]

November 30, 2009

A novel splitting method is presented for the ℓ^1-TV restoration of degraded images subject to impulsive noise. The functional is split into an ℓ^2-TV denoising and an ℓ^1-ℓ^2 deblurring part. The dual problem of the relaxed functional is smooth with convex constraints, and can be solved efficiently by applying an Arrow-Hurwicz type algorithm to the augmented Lagrangian formulation. The regularization parameter is chosen automatically based on a balancing principle. The accuracy, the fast convergence and robustness of the algorithm, and the use of the parameter choice rule are illustrated on some benchmark images and compared with an existing method.

1. Introduction

In this work we consider restoring images degraded by blurring and impulsive noise (e.g. salt-and-pepper noise). Specifically, the observed image f is the convolution of a blurring operator with the true image and then corrupted by noise. The discrete image formation process can be written as

$$(1.1) \qquad f = N_{imp}(Ku_t),$$

where $u_t \in X = \mathbb{R}^{n \times n}$ denotes the unknown true image (n the number of pixels in one direction), $K : X \to X$ the blurring matrix, and the notation N_{imp} denotes the impulsive noise formation mechanism. The special case $K = I$, the identity matrix, is known as denoising.

Image restoration problems are frequently ill-conditioned, and thus a direct solution of the linear system (1.1) often does not yield meaningful solutions. Special techniques are required for its stable and accurate solution. The now standard approach to overcome ill-conditioning are regularization methods due to Tikhonov [27], which consist of minimizing a regularized functional. In this paper, we consider a regularization functional of the form

$$(\mathcal{P}) \qquad \min_u \left\{ \mathcal{J}_\alpha(u) := \|Ku - f\|_{\ell^1} + \alpha \|\nabla u\|_{\ell^1} \right\},$$

where α is the regularization parameter determining the trade-off between the data fitting term $\|Ku - f\|_{\ell^1}$ and the regularization term $\|\nabla u\|_{\ell^1}$. The term $\|\nabla u\|_{\ell^1}$ denotes the discrete total variation semi-norm. The total variation regularization can take care of edges in the restored image, and thus usually delivers much better results than the standard smoothness regularization [24]. We refer to the monographs [8, 5] for an up-to-date account of the mathematical theory of and computational techniques in image processing.

The ℓ^1 data fitting in the functional ($\mathcal{P}$) is motivated by the non-Gaussian nature of the noise under consideration. For instance, the noise may follow a Laplace distribution as in certain inverse problems arising in signal processing [3]. Noise models of impulsive type, e.g. salt-and-pepper or random valued noise, arise in image processing because of malfunctioning pixels in camera sensors, faulty memory locations in hardware, or transmission in noisy channels [6]. Statistically, ℓ^1 data fitting is more robust to outliers than the more classical ℓ^2 data fitting in the sense that a small number of outliers has less influence on the solution [19, 9]. Consequently, functionals involving ℓ^1 data fitting have received growing interest in image processing, and various interesting theoretical properties [2, 21, 7, 29, 1, 11] of the functional ($\mathcal{P}$) have been established, mostly in the context of image denoising.

The minimization of the functional ($\mathcal{P}$) has been the topic of several recent investigations. In [14], the problem was reformulated as a linear programming problem which is solved by an interior point method, and [23] generalized the iteratively reweighted least-squares method. More recently, a primal dual active set algorithm for minimizing the functional ($\mathcal{P}$) has been investigated [10]. Also, several splitting methods were proposed [28, 16], which introduce two auxiliary variables in order to replace the ℓ^1-deblurring problem with a ℓ^2-deblurring (sub-)problem that can be solved either in closed form or via fast transforms.

In this paper, we combine a splitting approach with convex duality techniques to obtain a new and efficient algorithm for ℓ^1-TV image restoration. Specifically, we introduce only one splitting term to separate the ℓ^1-fitting and the TV penalty term. In contrast to existing splitting methods [28, 16], our approach does not require modifying the structure of the residual and the penalty term. The resulting relaxed problem can be transformed using Fenchel duality into a quadratic programming problem with a linear equality and simple box constraints, which can be solved very efficiently using Arrow-Hurwicz type methods [4, 22, 17]. Our numerical results show that the new algorithm can outperform existing algorithms in terms of computing time and accuracy. In addition, we propose a heuristic parameter choice rule based on a balancing principle for choosing an appropriate regularization parameter α in the functional ($\mathcal{P}$), and derive an efficient fixed point algorithm for computing the parameter. The authors are not aware of any existing parameter choice rule for ℓ^1-TV deblurring in the literature.

This paper is organized as follows. In the remainder of this section, we fix some notation. In Section 2, we introduce our splitting, and investigate the properties of the relaxed functional, including existence and its dual formulation. An Arrow-Hurwicz type algorithm for computing saddle points of the augmented Lagrangian formulation of the dual problem is considered, and its convergence is established. In Section 4, we describe a heuristic parameter choice rule based on a balancing principle and investigate its efficient numerical computation and its convergence. Finally, numerical results for some benchmark images are presented to illustrate the efficiency of the algorithm and utility of the parameter choice rule. The new method is compared with one state-of-the-art algorithm, FTVd [28], for which the source code is available online, and the new choice rule is evaluated against the optimal choice.

NOTATION

We consider images as elements of the vector space $X = \mathbb{R}^{n \times n}$, with n being the number of pixels in one direction, respectively. For $u \in X$, we denote by $u(i,j)$ the intensity value of the image u on pixel (i,j). The discrete gradient $\nabla : X \to Y := \mathbb{R}^{2 \times n \times n}$ with periodic boundary conditions is defined as

$$(\nabla u)(i,j) = \big((\partial_x u)(i,j), (\partial_y u)(i,j)\big),$$

with $(\partial_x u)(i,j)$ and $(\partial_y u)(i,j)$ given by

$$(\partial_x u)(i,j) = \begin{cases} u(i+1,j) - u(i,j), & \text{if } i < n, \\ u(1,j) - u(n,j), & \text{if } i = n, \end{cases}$$

$$(\partial_y u)(i,j) = \begin{cases} u(i,j+1) - u(i,j), & \text{if } j < n, \\ u(i,1) - u(i,n), & \text{if } j = n. \end{cases}$$

To define the dual problem for the total variation restoration problem, we need the discrete divergence operator div. The operator $\text{div} : Y \to X$ is defined by duality, i.e. imposing

$$\langle \nabla u, q \rangle_Y = -\langle u, \text{div}\, q \rangle_X, \ \forall\, u \in X, q = (q^1, q^2) \in Y.$$

A simple computation gives

$$(\partial_x^T q)(i,j) = \begin{cases} q^1(i,j) - q^1(i-1,j), & \text{if } 1 < i \le n, \\ q^1(n,j) - q^1(1,j), & \text{if } i = 1, \end{cases}$$

$$(\partial_y^T q)(i,j) = \begin{cases} q^2(i,j) - q^2(i,j-1), & \text{if } 1 < j \le n, \\ q^2(i,n) - q^2(i,1), & \text{if } j = 1, \end{cases}$$

$$(\text{div}\, q)(i,j) = (\partial_x^T q)(i,j) + (\partial_y^T q)(i,j).$$

The isotropic discrete total variation semi-norm $\|\nabla u\|_{\ell^1}$ is defined by

$$\|\nabla u\|_{\ell^1} = \sum_{i,j} |(\nabla u)(i,j)|_2 = \sum_{i,j} \sqrt{(\partial_x u)(i,j)^2 + (\partial_y u)(i,j)^2},$$

and the anisotropic total variation semi-norm is defined by

$$\|\nabla u\|_{\ell^1} = \sum_{i,j} |(\nabla u)(i,j)|_1 = \sum_{i,j} \left[|(\partial_x u)(i,j)| + |(\partial_y u)(i,j)| \right].$$

The algorithm developed in this paper can be applied to both the isotropic and the anisotropic total variation semi-norm. Because the derivations and results for these two cases are almost identical, we consider only the isotropic case unless otherwise noted, and avoid differentiating by notation.

Finally, we shall need the pointwise maximum norm for vectors $q = (q^1, q^2) \in Y$:

$$\|q\|_{\ell^\infty} = \max_{i,j} \left(|q(i,j)|_2 \right) = \max_{i,j} \left(|(q^1(i,j), q^2(i,j))|_2 \right)$$

in the isotropic case and

$$\|q\|_{\ell^\infty} = \max_{i,j} \left(|q(i,j)|_\infty \right)$$

with $|q(i,j)|_\infty = \max \left(|q^1(i,j)|, |q^2(i,j)| \right)$ in the anisotropic case.

Throughout the paper, C denotes some generic constant whose value may differ at each appearance.

2. Problem relaxation

We consider for $\mu > 0$ the following splitting of the functional ($\mathcal{P}$):

$$(\mathcal{P}_\mu) \qquad \min_{u,v \in X} \left\{ \mathcal{J}_{\alpha,\mu}(u,v) := \|Ku - f\|_{\ell^1} + \alpha \|\nabla v\|_{\ell^1} + \frac{1}{2\mu} \|u - v\|_{\ell^2}^2 \right\}.$$

In this section, we will study the existence of a minimizer to the functional $\mathcal{J}_{\alpha,\mu}$, the convergence of the minimizer as μ tends to zero and its dual formulation.

2.1. Primal problem

Theorem 2.1. *For each $\alpha > 0$ and $\mu > 0$, the functional $\mathcal{J}_{\alpha,\mu}$ is convex. Assume that $\ker K \cap \ker \nabla = \{0\}$, then problem ($\mathcal{P}_\mu$) has at least one minimizer (u_μ, v_μ).*

Proof. First we show the convexity of the functional. Observe that the first and second terms of the functional are obviously convex. The Hessian of $\|u - v\|_{\ell^2}^2$ has the non-negative eigenvalues 0 and 2, which implies convexity.

It is easy to see that the functional $\mathcal{J}_{\alpha,\mu}$ is continuous and proper. To establish the existence of a minimizer, we need only to show that it is also coercive, i.e. $\mathcal{J}_{\alpha,\mu}(u,v) \to +\infty$ as $\|u\|_{\ell^2} + \|v\|_{\ell^2} \to \infty$. We argue by contraposition. Let $\{(u_k, v_k)\} \subset X \times X$ be a sequence such that

$$(2.1) \qquad \|Ku_k - f\|_{\ell^1} + \alpha \|\nabla v_k\|_{\ell^1} + \frac{1}{2\mu} \|u_k - v_k\|_{\ell^2}^2 \leq C.$$

4

In particular, we deduce that

$$(2.2) \qquad \|\nabla v_k\|_{\ell^1} \leq C.$$

Now by the discrete Poincaré inequality [5, Proposition 5.2.2, p.236], there exists a $C > 0$ such that

$$\|v_k - \overline{v}_k\|_{\ell^2} \leq C \|\nabla v_k\|_{\ell^1} \leq C$$

holds, where $\overline{v}_k$ is the mean of the vector v_k, i.e. $\overline{v}_k = \frac{1}{n^2} \sum_{i,j} v_k(i,j)$. Decompose also $u_k = \overline{u}_k + (u_k - \overline{u}_k)$, then appealing to inequality (2.1) again gives

$$\|u_k - v_k\|_{\ell^2}^2 = \|\overline{u}_k - \overline{v}_k + (u_k - \overline{u}_k) - (v_k - \overline{v}_k)\|_{\ell^2}^2 \leq C,$$

which together with the identity $\sum_{i,j} (u_k(i,j) - \overline{u}_k) = 0$ implies

$$\|\overline{u}_k - \overline{v}_k\|_{\ell^2}^2 + \|(u_k - \overline{u}_k) - (v_k - \overline{v}_k)\|_{\ell^2}^2 \leq C.$$

Therefore, we deduce that

$$(2.3) \qquad \|\overline{u}_k - \overline{v}_k\|_{\ell^2}^2 \leq C$$

and

$$(2.4) \qquad \|u_k - \overline{u}_k\|_{\ell^2}^2 \leq C.$$

Now the triangle inequality implies

$$\begin{aligned} \|K\overline{u}_k\|_{\ell^1} &\leq \|K(u_k - \overline{u}_k)\|_{\ell^1} + \|f\|_{\ell^1} + \|Ku_k - f\|_{\ell^1} \\ &\leq \|K\|_{\ell^1,\ell^2} \|u_k - \overline{u}\|_{\ell^2} + \|f\|_{\ell^1} + \|Ku_k - f\|_{\ell^1} \leq C \end{aligned}$$

by appealing again to inequalities (2.1) and (2.4). Here, $\|K\|_{\ell^1,\ell^2}$ denotes the operator norm of K between the normed spaces $(X, \|\cdot\|_{\ell^1})$ and $(X, \|\cdot\|_{\ell^2})$. By assumption, constant functions are not in the kernel of K, and thus the above inequality implies $|\overline{u}_k| \leq C$. Combining this inequality together with inequalities (2.2) and (2.3), we deduce

$$\|u_k\|_{\ell^2} + \|v_k\|_{\ell^2} \leq C,$$

for all k. This shows the coercivity of the functional $\mathcal{J}_{\alpha,\mu}$. The existence of a minimizer now follows directly from coercivity. $\qquad \square$

The functional $\mathcal{J}_{\alpha,\mu}$ is not strictly convex, and thus in general uniqueness of a minimizer (u_μ, v_μ) does not hold. However, we can show that the difference $u_\mu - v_\mu$ is uniquely determined.

Proposition 2.2. *Let $\alpha, \mu > 0$ be fixed and (u_1, v_1) and (u_2, v_2) be two minimizers of the functional $\mathcal{J}_{\alpha,\mu}$. Then $u_1 - v_1 = u_2 - v_2$ holds.*

Proof. By Theorem 2.1, $\mathcal{J}_{\alpha,\mu}$ is convex, and thus for any $t \in [0,1]$, we have

$$\mathcal{J}_{\alpha,\mu}(tu_1 + (1-t)u_2, tv_1 + (1-t)v_2) = t\mathcal{J}_{\alpha,\mu}(u_1, v_1) + (1-t)\mathcal{J}_{\alpha,\mu}(u_2, v_2).$$

Now by the convexity of the functional $\|Ku - f\|_{\ell^1}$ and $\|\nabla v\|_{\ell^1}$, we have

$$\|K(tu_1 + (1-t)u_2) - f\|_{\ell^1} \le t\|Ku_1 - f\|_{\ell^1} + (1-t)\|Ku_2 - f\|_{\ell^1},$$
$$\|\nabla(tv_1 + (1-t)v_2)\|_{\ell^1} \le t\|\nabla v_1\|_{\ell^1} + (1-t)\|\nabla v_2\|_{\ell^1}.$$

Consequently the above identity yields

$$\|t(u_1 - v_1) + (1-t)(u_2 - v_2)\|_{\ell^2}^2 \ge t\|u_1 - v_1\|_{\ell^2}^2 + (1-t)\|u_2 - v_2\|_{\ell^2}^2,$$

for any $t \in [0,1]$. However, by the convexity of the functional $\| \cdot \|_{\ell^2}^2$, we have

$$\|t(u_1 - v_1) + (1-t)(u_2 - v_2)\|_{\ell^2}^2 \le t\|u_1 - v_1\|_{\ell^2}^2 + (1-t)\|u_2 - v_2\|_{\ell^2}^2.$$

By combining the above two inequalities, we deduce by appealing again to strict convexity of the functional $\| \cdot \|_{\ell^2}^2$ that $u_1 - v_1 = u_2 - v_2$. This concludes the proof. $\square$

The next result shows the convergence of (u_μ, v_μ) as μ tends to zero.

Theorem 2.3. *Let $\{\mu_k\}$ be a sequence tending to zero. Then the sequence of minimizers $\{(u_{\mu_k}, v_{\mu_k})\}$ to the functional $\mathcal{J}_{\alpha,\mu_k}$ has a subsequence converging to (u^*, v^*) with $u^* = v^*$ being a minimizer of the functional $\mathcal{J}_\alpha$.*

Proof. Let $\tilde{u}$ be a global minimizer of the functional $\mathcal{J}_a(u)$. By the minimizing properties of the pair $(u_k, v_k) \equiv (u_{\mu_k}, v_{\mu_k})$ to the functional $\mathcal{J}_{\alpha,\mu_k}$, we have

$$(2.5) \qquad \mathcal{J}_{\alpha,\mu_k}(u_k, v_k) = \|Ku_k - f\|_{\ell^1} + \alpha \|\nabla v_k\|_{\ell^1} + \frac{1}{2\mu_k}\|u_k - v_k\|_{\ell^2}^2$$
$$\le \|K\tilde{u} - f\|_{\ell^1} + \alpha \|\nabla \tilde{u}\|_{\ell^1}.$$

From this and the coercivity of the functional $\mathcal{J}_{\alpha,\mu_k}$, we deduce that $\{(u_k, v_k)\}$ is uniformly bounded independently of k. Therefore there exists a subsequence of $\{(u_k, v_k)\}$, also denoted by $\{(u_k, v_k)\}$ and some (u^*, v^*), such that

$$(u_k, v_k) \to (u^*, v^*).$$

By rearranging (2.5), we have

$$\|u_k - v_k\|_{\ell^2}^2 \le 2\mu_k[\|K\tilde{u} - f\|_{\ell^1} + \alpha \|\nabla \tilde{u}\|_{\ell^1} - (\|Ku_k - f\|_{\ell^1} + \alpha \|\nabla v_k\|_{\ell^1})].$$

Taking the limit as $k \to \infty$, we obtain that

$$\|u^* - v^*\|_{\ell^2}^2 \le \liminf_k \|u_k - v_k\|_{\ell^2}^2$$
$$\le \lim_{k \to \infty} 2\mu_n[\|K\tilde{u} - f\|_{\ell^1} + \alpha \|\nabla \tilde{u}\|_{\ell^1} - (\|Ku_k - f\|_{\ell^1} + \alpha \|\nabla v_k\|_{\ell^1})] = 0$$

by noting $\mu_k \to 0$. Therefore, we deduce that $v^* = u^*$. Moreover, by taking the limit in (2.5) as $k \to \infty$, we obtain

$$
\|Ku^* - f\|_{\ell^1} + \alpha \|\nabla u^*\|_{\ell^1} \leq \lim_{k \to \infty} \|Ku_k - f\|_{\ell^1} + \lim_{k \to \infty} \alpha \|\nabla v_k\|_{\ell^1}
$$
$$
+ \lim_{k \to \infty} \frac{1}{2\mu_k} \|u_k - v_k\|_{\ell^2}^2 \leq \|K\tilde{u} - f\|_{\ell^1} + \alpha \|\nabla \tilde{u}\|_{\ell^1} .
$$

Therefore u^* is a minimizer of the functional $\mathcal{J}_\alpha$ since $\tilde{u}$ is a global minimizer. □

The preceding result suggests a continuation strategy in μ, where we solve the problem $(\mathcal{P}_\mu)$ for fixed μ_k and take these solutions as an initial guess for solving problem $(\mathcal{P}_\mu)$ for a $\mu_{k+1} < \mu_k$.

2.2. Dual problem

While problem $(\mathcal{P}_\mu)$ is challenging to solve directly due to the non-smoothness of the two ℓ^1 terms, we will show that the corresponding dual problem is a quadratic minimization problem over a convex set:

$(\mathcal{P}_\mu^*)$
$$
\begin{cases}
\displaystyle \min_{p \in X, q \in Y} \frac{\mu}{4} \|K^* p\|_{\ell^2}^2 + \frac{\mu}{4} \|\operatorname{div} q\|_{\ell^2}^2 - \langle p, f \rangle \\[2mm]
\text{s.t. } \|p\|_{\ell^\infty} \leq 1, \|q\|_{\ell^\infty} \leq \alpha, K^* p - \operatorname{div} q = 0.
\end{cases}
$$

Theorem 2.4. *The dual problem of $(\mathcal{P}_\mu)$ is $(\mathcal{P}_\mu^*)$. Problem $(\mathcal{P}_\mu^*)$ has at least one minimizer $(p_\mu, q_\mu) \in X \times Y$. Moreover, any minimizer (p_μ, q_μ) of $(\mathcal{P}_\mu^*)$ is related to minimizers (u_μ, v_μ) of $(\mathcal{P}_\mu)$ by the following extremality relations*

(2.6)
$$
\begin{cases}
\mu K^* p_\mu = u_\mu - v_\mu, \\
-\mu \operatorname{div} q_\mu = -u_\mu + v_\mu, \\
0 \leq \langle Ku_\mu - f, p - p_\mu \rangle, \\
0 \leq \langle \nabla v_\mu, q - q_\mu \rangle,
\end{cases}
$$

for all $p \in X$ with $\|p\|_{\ell^\infty} \leq 1$ and all $q \in Y$ with $\|q\|_{\ell^\infty} \leq \alpha$.

Proof. We apply Fenchel duality [12], setting

$$
\mathcal{F} : X \times X \to \mathbb{R}, \qquad \mathcal{F}(u, v) = \frac{1}{2\mu} \|u - v\|_{\ell^2}^2 ,
$$
$$
\mathcal{G} : X \times Y \to \mathbb{R}, \qquad \mathcal{G}(u, v) = \|u - f\|_{\ell^1} + \alpha \|v\|_{\ell^1} ,
$$
$$
\Lambda : X \times X \to X \times Y, \qquad \Lambda(u, v) = (Ku, \nabla v).
$$

We first calculate the Fenchel conjugate $\mathcal{F}^* : X \times X \to \mathbb{R} \cup \{\infty\}$,

$$
\mathcal{F}^*(p, q) = \sup_{(u, v)} \langle u, p \rangle + \langle v, q \rangle - \frac{1}{2\mu} \|u - v\|_{\ell^2}^2 .
$$

If $p = -q$, we have

$$\mathcal{F}^*(p,q) = \sup_{(u,v)} \langle u - v, p \rangle - \frac{1}{2\mu} \| u - v \|_{\ell^2}^2 ,$$

where the supremum is attained for $p = \frac{1}{\mu}(u - v)$, yielding

$$\mathcal{F}^*(p,q) = \frac{\mu}{2} \| p \|_{\ell^2}^2 = \frac{\mu}{4} \| p \|_{\ell^2}^2 + \frac{\mu}{4} \| q \|_{\ell^2}^2 .$$

On the other hand, if $p \neq -q$, we can choose a sequence (u_k, v_k) for which $u_k = v_k$, so that $\mathcal{F}^*(p,q)$ is unbounded. Hence,

$$\mathcal{F}^*(p,q) = \begin{cases} \frac{\mu}{4} \| p \|_{\ell^2}^2 + \frac{\mu}{4} \| q \|_{\ell^2}^2 & \text{if } p + q = 0, \\ \infty & \text{otherwise.} \end{cases}$$

The Fenchel conjugate of $\mathcal{G}$ can be computed term by term and is given by

$$\mathcal{G}^* : X \times Y \to \mathbb{R} \cup \{\infty\}, \qquad \mathcal{G}^*(p,q) = \langle p, f \rangle + I_{\{\| p \|_{\ell^\infty} \leq 1\}} + I_{\{\| q \|_{\ell^\infty} \leq \alpha\}},$$

where I_S denotes the indicator function of the set S.

The adjoint of the operator Λ is

$$\Lambda^* : X \times Y \to X \times X, \qquad \Lambda^*(p,q) = (K^* p, - \operatorname{div} q).$$

Since the functionals $\mathcal{F}$ and $\mathcal{G}$ are convex lower semicontinuous, proper and continuous at 0, and K is a continuous linear operator, the Fenchel duality theorem states that

$$(2.7) \qquad \inf_{(u,v)} \mathcal{F}(u,v) + \mathcal{G}(\Lambda(u,v)) = \sup_{(p,q)} -\mathcal{F}^*(\Lambda^*(p,q)) - \mathcal{G}^*(-p,-q)$$

holds, and that the right-hand side of (2.7) has at least one solution.

Furthermore, the equality in (2.7) is attained at (u_μ, v_μ), (p_μ, q_μ) if and only if

$$\begin{cases} \Lambda^*(p_\mu, q_\mu) \in \partial \mathcal{F}(u_\mu, v_\mu), \\ -(p_\mu, q_\mu) \in \partial \mathcal{G}(\Lambda(u_\mu, v_\mu)). \end{cases}$$

Since $\mathcal{F}$ is Fréchet-differentiable, the first relation of (2.7) follows by direct calculation. From the definition of the subgradient, we immediately have that

$$-(p_\mu, q_\mu) \in \partial \mathcal{G}(\Lambda(u_\mu, v_\mu)) \Leftrightarrow \Lambda(u_\mu, v_\mu) \in \partial \mathcal{G}^*(-p_\mu, -q_\mu).$$

Subdifferential calculus then yields

$$K u_\mu - f \in \partial I_{\{\| -p_\mu \|_{\ell^\infty} \leq 1\}},$$
$$\nabla v_\mu \in \partial I_{\{\| -q_\mu \|_{\ell^\infty} \leq \alpha\}}.$$

Recall that the subdifferential of the indicator function I_S of a set S coincides with the normal cone at S (cf., e.g., [18, Ex. 4.21]). We thus obtain that

$$0 \geq \langle K u_\mu - f, p + p_\mu \rangle,$$
$$0 \geq \langle \nabla v_\mu, q + q_\mu \rangle,$$

for all p with $\| p \|_{\ell^\infty} \leq 1$ and all q with $\| q \|_{\ell^\infty} \leq \alpha$, from which the second set of relations follows. $\qquad \square$

Regarding uniqueness of minimizers for the dual problem, we have the following:

Proposition 2.5. *Assume that $f \in$ range K and that the solution (p_μ, q_μ) to Problem $(\mathcal{P}_\mu^*)$ is such that q_μ is bi-active. Then q_μ is unique.*

Proof. Let $\overline{f}$ be such that $K\overline{f} = f$. Then the dual problem $(\mathcal{P}_\mu^*)$ can equivalently be expressed as

$$
\begin{cases}
\min\limits_{p \in X, q \in Y} \dfrac{1}{2} \left\| \sqrt{\mu}\, \mathrm{div}\, q - \dfrac{1}{\sqrt{\mu}} \overline{f} \right\|_{\ell^2} - \dfrac{1}{2\mu} \left\| \overline{f} \right\|_{\ell^2}^2 \\
\text{s.t. } \|p\|_{\ell^\infty} \le 1, \|q\|_{\ell^\infty} \le \alpha, K^*p - \mathrm{div}\, q = 0.
\end{cases}
$$

This is an optimization problem with convex cost and convex constraint, so the q component of the minimizer must be of the form $q_\mu = \tilde{q} + c$ for some $\tilde{q} \in Y$, where c is a constant. Since q_μ is bi-active by assumption, this constant must necessarily be unique. $\qquad\square$

3. COMPUTATION OF MINIMIZERS

To compute a minimizer of the dual problem $(\mathcal{P}_\mu^*)$ for fixed $\mu > 0$, we use a projected gradient descent-ascent method of Arrow-Hurwicz type. While the projection onto the set given by the box constraints is easy to compute, this is more difficult for the linear equality constraint $K^*p - \mathrm{div}\, q = 0$. Therefore, we consider an augmented Lagrangian formulation of this constraint. The augmented Lagrangian for any $c > 0$ is defined as

$$
\mathcal{L}(p, q, \lambda) := \frac{\mu}{4} \|K^*p\|_{\ell^2}^2 + \frac{\mu}{4} \|\mathrm{div}\, q\|_{\ell^2}^2 - \langle p, f \rangle + \langle \lambda, K^*p - \mathrm{div}\, q \rangle + \frac{c}{2} \|K^*p - \mathrm{div}\, q\|_{\ell^2}^2.
$$

By setting

$$
B(1) := \{ p \in X : \|p\|_{\ell^\infty} \le 1 \}, \quad B(\alpha) := \{ q \in Y : \|q\|_{\ell^\infty} \le \alpha \},
$$

we therefore have to solve the saddle point problem

$$
(3.1) \qquad \min_{p \in B(1), q \in B(\alpha)} \sup_{\lambda \in X} \mathcal{L}(p, q, \lambda).
$$

The existence of a solution of the constrained problem $(\mathcal{P}_\mu^*)$ immediately ensures existence of a saddle point (p^*, q^*, λ^*) for (3.1), which can be computed using a variant of the classical Arrow-Hurwicz method proposed by Popov [22]. Popov's variant was designed for finding a saddle point of general convexoconcave functions on a closed convex domain. The main idea is to compute at each iteration a "leading" point $(\overline{p}^k, \overline{q}^k, \overline{\lambda}^k)$, which can ensure the convergence of the algorithm. We employ a modified variant, which uses different step lengths for the leading point in order to allow more aggressive updates for the Lagrange multiplier.

To give the algorithm, we denote by $P_{B(c)}$ for $c > 0$ the projection on the ball $B(c)$ in X (respectively Y), which is given pointwise via

$$
P_{B(c)}(x)(i, j) = c \frac{x(i, j)}{\max(c, |x(i, j)|_r)},
$$

where $r = 2$ in the isotropic and $r = \infty$ in the anisotropic case. We denote by $\mathcal{L}_p$ the Fréchet derivative of $\mathcal{L}$ with respect to p, and similarly $\mathcal{L}_q$ and $\mathcal{L}_\lambda$.

The complete procedure is given in Algorithm 1, where we have set $c = \mu$ for computational simplicity.

Algorithm 1 Extrapolated gradient algorithm

1: Set $p^0 = \lambda^0 = \overline{p}^0 = \overline{\lambda}^0 = 0$ and $q^0 = \overline{q}^0 = 0$, set step sizes $\gamma_p, \gamma_q, \gamma_\lambda, \rho$
2: **for** $k = 1, \ldots, N$ **do**
3:      $\eta_p^k = \mathcal{L}_p(\overline{p}^{k-1}, \overline{q}^{k-1}, \overline{\lambda}^{k-1})$
4:      $\eta_q^k = \mathcal{L}_q(\overline{p}^{k-1}, \overline{q}^{k-1}, \overline{\lambda}^{k-1})$
5:      $\eta_\lambda^k = \mathcal{L}_\lambda(\overline{p}^{k-1}, \overline{q}^{k-1}, \overline{\lambda}^{k-1})$
6:      $p^k = P_{B(1)}(p^{k-1} - \gamma_p \eta_p^k)$
7:      $q^k = P_{B(\alpha)}(q^{k-1} - \gamma_q \eta_q^k)$
8:      $\lambda^k = \lambda^{k-1} + \gamma_\lambda \eta_\lambda^k$
9:      $\overline{p}^k = P_{B(1)}(p^k - \gamma_p \eta_p^k)$
10:      $\overline{q}^k = P_{B(\alpha)}(q^k - \gamma_q \eta_q^k)$
11:      $\overline{\lambda}^k = \lambda^k + \rho \gamma_\lambda \eta_\lambda^k$
12: **end for**
13: Set $u^N = \lambda^N + \frac{\mu}{2} K^* p^N$, $v^N = \lambda^N - \frac{\mu}{2} \operatorname{div} q^N$,
14: **OUTPUT** Approximate minimizer u^N, v^N

We have the following convergence result. The (technical) proof of this theorem can be found in Appendix A.

Theorem 3.1. *For each $\rho \in (\frac{1}{2}, +\infty)$, there exists a positive number $\gamma_0 > 0$ such that for all γ_p, $\gamma_q, \gamma_\lambda \in (0, \gamma_0)$, the sequence $\{(p^k, q^k, \lambda^k)\}$ generated by Algorithm 1 converges to a saddle point of the functional $\mathcal{L}(p, q, \lambda)$.*

Once a saddle point (p^*, q^*, λ^*) has been computed, the corresponding primal variable u^* (or v^*) can be obtained using the extremality relations for the augmented Lagrangian: Setting

$$\mathcal{F}(p, q) = \frac{\mu}{4} \|p\|_{\ell^2}^2 + \frac{\mu}{4} \|q\|_{\ell^2}^2 + \langle \lambda^*, p + q \rangle + \frac{c}{2} \|p + q\|_{\ell^2}^2,$$

we have that $(u^*, v^*) = \partial \mathcal{F}(K^* p^*, -\operatorname{div} q^*)$ and $K^* p^* = \operatorname{div} q^*$, from which it follows that

$$u^* = \lambda^* + \frac{\mu}{2} K^* p^*,$$
$$v^* = \lambda^* + \frac{\mu}{2}(-\operatorname{div} q^*).$$

To see that this indeed yields the desired solution, we subtract these equalities to obtain

$$u^* - v^* = \frac{\mu}{2}(K^* p^* + \operatorname{div} q^*).$$

Using again the fact that $K^* p^* = \operatorname{div} q^*$, we have that

$$u^* - v^* = \mu K^* p^* = \mu \operatorname{div} q^*,$$

i.e., $(u^*, v^*), (p^*, q^*)$ satisfies the extremality relations (2.6), and (u^*, v^*) is therefore a solution of problem ($\mathcal{P}_\mu$).

4. Parameter choice rule

Our parameter choice rule is based on the following balancing principle, recently derived by the authors using the model function approach [9]:

$$(4.1) \qquad (\sigma - 1) \, \|Ku_\alpha - f\|_{\ell^1} - \alpha \, \|\nabla u_\alpha\|_{\ell^1} = 0.$$

The underlying idea of the principle is to balance the data fitting term with the regularization term, and the parameter $\sigma > 1$ controls the trade-off between them. We point out that this principle does not require knowledge of the noise level. This kind of balancing idea underlies a number of existing parameter choice rules for ℓ^2-ℓ^2 formulations [13], e.g. the local minimum criterion and the L-curve criterion.

We shall denote by $r(\alpha) = (\sigma - 1) \, \|Ku_\alpha - f\|_{\ell^1} - \alpha \, \|\nabla u_\alpha\|_{\ell^1}$ the equation error. We propose to compute a solution α^* to the balancing equation (4.1) by the following simple fixed point algorithm

$$(4.2) \qquad \alpha_{k+1} = (\sigma - 1) \frac{\|Ku_{\alpha_k} - f\|_{\ell^1}}{\|\nabla u_{\alpha_k}\|_{\ell^1}}.$$

To show its convergence, the following monotonicity result is useful.

Lemma 4.1. *The functions* $\|Ku_\alpha - f\|_{\ell^1}$ *and* $\|\nabla u_\alpha\|_{\ell^1}$ *are monotonic in* α *in the sense that for* $\alpha_1, \alpha_2 > 0$, *there holds*

$$\left(\|Ku_{\alpha_1} - f\|_{\ell^1} - \|Ku_{\alpha_2} - f\|_{\ell^1} \right)(\alpha_1 - \alpha_2) \geq 0$$

and

$$\left(\|\nabla u_{\alpha_1}\|_{\ell^1} - \|\nabla u_{\alpha_2}\|_{\ell^1} \right)(\alpha_1 - \alpha_2) \leq 0.$$

Proof. The minimizing property of u_{α_1} and u_{α_2} yields

$$\|Ku_{\alpha_1} - f\|_{\ell^1} + \alpha_1 \, \|\nabla u_{\alpha_1}\|_{\ell^1} \leq \|Ku_{\alpha_2} - f\|_{\ell^1} + \alpha_1 \, \|\nabla u_{\alpha_2}\|_{\ell^1},$$
$$\|Ku_{\alpha_2} - f\|_{\ell^1} + \alpha_2 \, \|\nabla u_{\alpha_2}\|_{\ell^1} \leq \|Ku_{\alpha_1} - f\|_{\ell^1} + \alpha_2 \, \|\nabla u_{\alpha_1}\|_{\ell^1}.$$

Adding these two inequalities together gives the second estimate. The first one can be obtained by dividing the two inequalities by α_1 and α_2 and then adding them together. $\square$

The next lemma shows the theoretically interesting and practically desirable monotonicity of the iteration (4.2).

Lemma 4.2. *The sequence of regularization parameters $\{\alpha_k\}$ by the fixed point algorithm is monotonic. Moreover, it is monotonically increasing if $r(\alpha_0) > 0$ and monotonically decreasing if $r(\alpha_0) < 0$.*

Proof. By the definition of the iteration, we have

$$\alpha_{k+1} - \alpha_k = (\sigma - 1)\frac{\|Ku_{\alpha_k} - f\|_{\ell^1}}{\|\nabla u_{\alpha_k}\|_{\ell^1}} - (\sigma - 1)\frac{\|Ku_{\alpha_{k-1}} - f\|_{\ell^1}}{\|\nabla u_{\alpha_{k-1}}\|_{\ell^1}}$$

$$= \frac{(\sigma - 1)}{\|\nabla u_{\alpha_{k-1}}\|_{\ell^1}\|\nabla u_{\alpha_k}\|_{\ell^1}}\left[\|Ku_{\alpha_k} - f\|_{\ell^1}\left(\|\nabla u_{\alpha_{k-1}}\|_{\ell^1} - \|\nabla u_{\alpha_k}\|_{\ell^1}\right)\right.$$

$$\left. + \left(\|Ku_{\alpha_k} - f\|_{\ell^1} - \|Ku_{\alpha_{k-1}} - f\|_{\ell^1}\right)\|\nabla u_{\alpha_k}\|_{\ell^1}\right].$$

By Lemma 4.1, the two terms both have the sign of $(\alpha_k - \alpha_{k-1})$, and thus the sequence is monotonic. Now if $r(\alpha_0) > 0$, then by the definition of the iteration, we have

$$\alpha_1 = (\sigma - 1)\frac{\|Ku_{\alpha_0} - f\|_{\ell^1}}{\|\nabla u_{\alpha_0}\|_{\ell^1}} > \alpha_0.$$

This concludes the proof of the lemma. □

Theorem 4.3. *If the initial guess α_0 satisfies $r(\alpha_0) < 0$, the sequence $\{\alpha_k\}$ converges.*

Proof. By Lemma 4.2, if $r(\alpha_0) < 0$, the sequence $\{\alpha_k\}$ is monotonically decreasing, and it is also bounded from below by zero, from which convergence follows. □

Finally, we remark that the noise level can be estimated using the residual

$$\delta_{est} = \frac{\|Ku - f\|_{\ell^1}}{n^2}.$$

5. NUMERICAL RESULTS

We now illustrate the efficiency of the proposed method and the parameter choice rule (referred to as balancing principle hereafter), for image restoration problems. The algorithms described above were implemented in MATLAB. For the operator K, the standard Gaussian low-pass filter provided by the MATLAB Image Processing Toolbox (imfilter) was used. Unless stated otherwise, the window size of the blurring kernel was fixed at 7×7, and the standard deviation at 5. Although our approach applies to both isotropic and anisotropic TV penalty terms, we shall present only results for the former, as the restoration results for the latter are quite similar in all subsequent examples.

In addition the corruption percentage d defined below, we also measure the degradation of the blurred and noisy image using the noise level

$$\delta = \frac{\|Ku_t - f\|_{\ell^1}}{n^2}.$$

To facilitate comparison, the quality of the restoration u_α is measured by relative error e_{rel} and peak signal-to-noise ratio (PSNR), which are respectively defined by

$$e_{rel} = \frac{\|u_\alpha - u_t\|_{\ell^2}}{\|u_t\|_{\ell^2}}, \qquad \text{PSNR} = -20\log_{10}\frac{\|u_t - u_\alpha\|_{\ell^2}}{n}.$$

We consider two types of impulsive noise. A very common noise model is salt-and-pepper noise, where the noisy image f is given pixel-wise as

$$f(i,j) = \begin{cases} 1 & \text{with probability } d/2, \\ 0 & \text{with probability } d/2, \\ (Ku_t)(i,j), & \text{with probability } 1-d. \end{cases}$$

A more realistic noise model for many applications (e.g., CCD sensor pixels which are subject to possible cosmic ray bombardment of varying energy) is positive Gaussian random-valued additive noise. In this model, if a pixel is corrupted, the absolute value of a normally distributed random variable $\xi(i,j)$ with zero mean and standard deviation 0.5 is added to the data:

$$f(i,j) = \begin{cases} (Ku_t)(i,j) + |\xi(i,j)| & \text{with probability } d, \\ (Ku_t)(i,j), & \text{with probability } 1-d. \end{cases}$$

The step sizes in all experiments were fixed as follows: $\gamma_p = (\mu L_K)^{-1}$, $\gamma_q = (\mu L_{\text{div}})^{-1}$, $\gamma_\lambda = .9\mu$, and $\rho = \frac{1}{2}$, where $L_K = 1$ and $L_{\text{div}} = 8$ denote the Lipschitz constants of the discrete operators K and div. This choice of γ_λ is based on the Lagrange multiplier update in the decomposition approach for the augmented Lagrangian method (cf. [15, Chap. VI]), while the choice of γ_p and γ_q is motivated by their role in Nesterov's optimal first-order gradient descent method [20]. Since the algorithm is non-monotonic, it was terminated after a fixed number of iterations. In all our experiments, 40 gradient descent (ascent) iterations and 4 continuation iterations proved sufficient for the desired results. The continuation parameters were chosen as $\mu_0 = 1$, $\mu_{k+1} = \mu_k/10$. In this way, the final relaxation parameter μ is comparable with that used in [28]. In all experiments, the final iterates in the continuation scheme for μ satisfied $\|u_\mu - v_\mu\|_{\ell^2} / \|u_\mu\|_{\ell^2} \approx 2 \times 10^{-5}$.

The choice of the relative weight σ in the fixed point iteration (4.2) is important. The optimal choice depends on the noise type, i.e. for salt-and-pepper and impulsive random valued noise, different choices can be advantageous. However, it should be pointed out that in our experiments the choice was very robust with respect to the noise level and the strength of the blurring. For the results presented below, it was set as $\sigma = 1.01$ for salt-and-pepper noise, and $\sigma = 1.04$ for impulsive Gaussian noise. The fixed point iteration was started with $\alpha_0 = 1$, and stopped if the relative change was smaller than 1×10^{-2}.

All computations were performed with MATLAB version 2009a on a single core of a 2.4GHz workstation with 4 GByte RAM. MATLAB codes implementing the algorithm presented in this paper can be downloaded under http://www.uni-graz.at/~clason/codes/l1tvsplitting.zip. We compared the proposed method to the *fast total variation deconvolution* (FTVd) described in [28], using the freely available implementation from http://www.caam.rice.edu/~optimization/L1/ftvd/ (v3.01). For FTVd, the default parameters set by the developers were taken.

5.1. RESTORATION QUALITY AND PARAMETER CHOICE

To show the capabilities of the method, we first consider an academic test example for which a TV penalty term is appropriate. The image consists of a flat box and a linear ramp (cf. Fig. 1a).
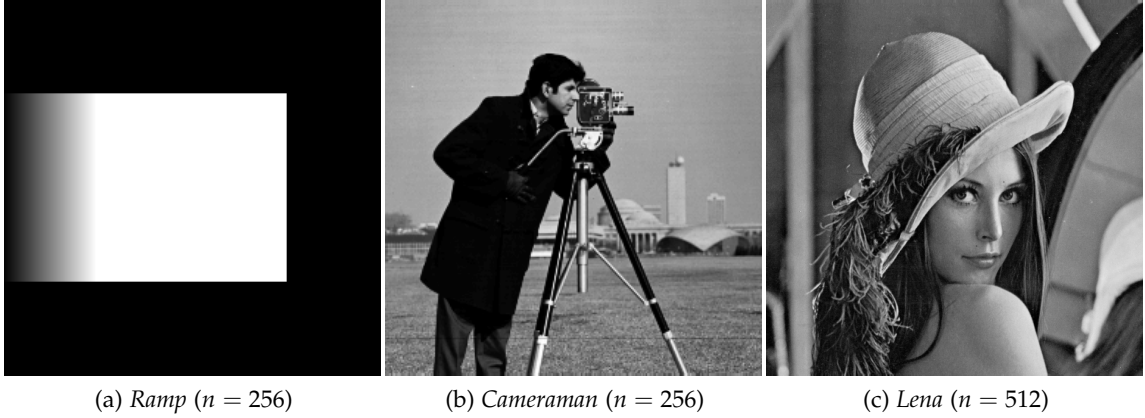
(a) *Ramp* ($n = 256$)   (b) *Cameraman* ($n = 256$)   (c) *Lena* ($n = 512$)

Figure 1: Test images.

Salt-and-pepper noise with corruption percentage $d \in \{0.3, \ldots, 0.6\}$ is added to the blurred image, and the reconstruction for the parameter α_b chosen by the balancing principle is compared to the optimal parameter α_o, which is chosen by sampling the interval $[0.01, 1]$ uniformly in a logarithmic scale at 100 points and then selecting the parameter with the lowest relative error e_{rel}. The results are given in Table 2a for a kernel size 7×7 and in Table 2b for a kernel size 15×15. The results for impulsive Gaussian noise are given in Tables 2c and 2d for a kernel size 7×7 and 15×15, respectively. In the table, t_{split} and t_{FTVd} refer to the computing time for the proposed method and FTVd, respectively. The noisy data, reconstructions using the proposed method and using FTVd are shown in Fig. 2 for the maximal noise level, $d = 0.6$. In the figure, α_b and α_o indicate the reconstructions by the proposed method with the regularization parameter given by the balancing principle and the optimal parameter, respectively, and FTVd refers to the restoration obtained by FTVd with the optimal parameter.

Several observations on the numerical results are in order. Firstly, as the noise level δ decreases, the accuracy of the reconstruction improves, which is the case for both automatically determined and optimal regularization parameters. Secondly, the automatically determined regularization parameter α_b is close to the optimal one α_o for all the cases under consideration, and thus the resulting restored images are comparable with those obtained with the optimal parameters. Therefore, the proposed choice rule is indeed reasonable and practically useful. Furthermore, the experiments indicate that one fixed value of σ can give reasonable restorations for noisy images with from 7×7 to 15×15 blur and $d = 0.3$ to $d = 0.6$ corruption percentage, and thus the choice of σ seems insensitive to noise levels and size of blurring kernels. In practice, the convergence of the fixed point iteration (4.2) is usually achieved within five iterations. Although not presented, the estimate δ_{est} in all cases was very close to the exact noise level δ. Thirdly, with the same α value, the accuracy of the proposed method is slightly better than that of FTVd for both salt-and-pepper and impulsive Gaussian noise. The differences can be clearly observed in the flat region. We point out that the case of 60% Gaussian random-valued noise is very challenging and requires a very large regularization parameter, rendering staircasing in the reconstruction inevitable.

14

Table 1: Results for *ramp*.

(a) salt-and-pepper noise, kernel size 7×7

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.150	0.165	2.66e-3	56.83	0.156	2.65e-3	56.86	4.14e-3	52.99	4.19	5.30
0.4	0.200	0.217	4.18e-3	52.91	0.148	3.92e-3	53.47	7.02e-3	48.40	4.18	5.53
0.5	0.248	0.266	1.29e-2	43.11	0.376	1.16e-2	44.08	2.46e-2	37.51	4.24	5.08
0.6	0.302	0.321	1.63e-2	41.09	0.376	1.58e-2	41.37	3.48e-2	34.50	4.59	7.69

(b) salt-and-pepper noise, kernel size 15×15

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.150	0.166	1.96e-2	39.50	0.272	1.65e-2	40.99	4.76e-2	31.79	4.84	8.39
0.4	0.198	0.217	2.00e-2	39.30	0.298	1.87e-2	39.91	5.17e-2	31.06	4.85	8.42
0.5	0.250	0.271	2.12e-2	38.81	0.313	2.09e-2	38.95	5.72e-2	30.18	4.85	8.59
0.6	0.298	0.320	2.61e-2	37.00	0.298	2.60e-2	37.05	5.57e-2	30.42	4.89	9.57

(c) Gaussian noise, kernel size 7×7

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.119	0.490	5.95e-3	49.84	0.260	4.98e-3	51.38	1.03e-2	45.09	4.57	5.79
0.4	0.160	0.639	2.26e-2	38.25	0.599	2.22e-2	38.40	3.55e-2	34.33	4.59	7.20
0.5	0.197	0.799	6.34e-2	29.29	0.599	6.08e-2	29.67	8.40e-2	26.85	4.38	10.72
0.6	0.241	0.969	2.12e-1	18.80	0.870	2.12e-1	18.81	2.20e-1	18.48	4.39	10.52

(d) Gaussian noise, kernel size 15×15

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.121	0.500	3.25e-2	35.10	0.260	1.84e-2	40.04	4.86e-2	31.60	4.85	9.48
0.4	0.159	0.646	6.82e-2	28.66	0.285	3.13e-2	35.42	5.07e-2	31.24	4.92	16.55
0.5	0.200	0.845	9.74e-2	25.57	0.327	8.20e-2	27.06	1.03e-1	25.05	5.02	27.10
0.6	0.239	0.987	2.29e-1	18.12	1.000	2.29e-1	18.12	2.34e-1	17.95	5.28	25.71

By comparing Table 2a with Table 2b, we observe that as the blurring window size increases, the restorations deteriorate accordingly. The computing time for our approach is relatively independent of the kernel size, as a consequence of fixed number of iterations, see Tables 2c and 2d.

5.2. Application to natural images

We next illustrate the performance of our method for two benchmark realistic images.

The first real image is *cameraman* ($n = 256$, cf. Fig. 1b). Reconstructions using our method and FTVd are given in Table 3a for salt-and-pepper noise and in Table 3b for impulsive Gaussian noise. Fig. 3 shows the results for a representative noise level for each noise model. Again we observe that the reconstructions by the automatically determined regularization parameter α_b and those by the optimal parameter α_o are fairly close.

The second image, *Lena*, has size $n = 512$ (cf. Fig. 1c). Here, we investigate the effect of the blurring kernel size on the reconstruction accuracy and speed. Table 4a summarizes
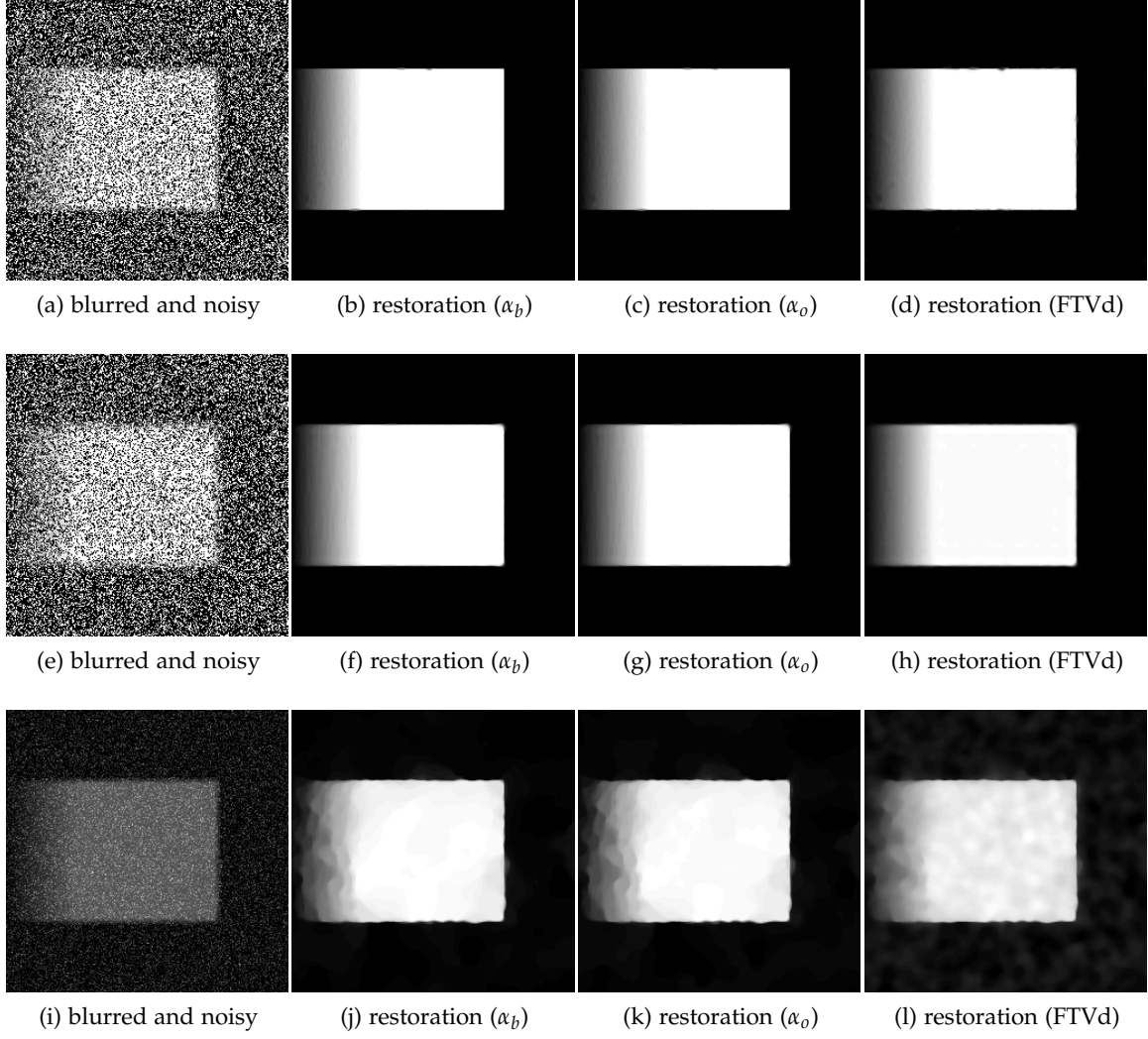
Figure 2: Restorations of *ramp* with $d = 0.6$. (a–d): salt and pepper noise, kernel size 7×7; (e–h): salt-and-pepper noise, kernel size 15×15; (i–l): Gaussian noise, kernel size 7×7.

Table 2: Results for *cameraman*.

(a) salt-and-pepper noise, kernel size 7×7

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.149	0.046	8.51e-2	27.08	0.032	8.26e-2	27.34	7.40e-2	28.29	4.48	11.79
0.4	0.199	0.074	9.80e-2	25.85	0.042	9.25e-2	26.35	8.78e-2	26.81	4.46	10.74
0.5	0.251	0.109	1.13e-1	24.59	0.056	1.07e-1	25.08	1.04e-1	25.30	4.48	12.81
0.6	0.298	0.139	1.23e-1	23.86	0.107	1.22e-1	23.96	1.24e-1	23.83	4.45	12.01

(b) Gaussian noise, kernel size 7×7

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.119	0.237	1.18e-1	24.21	0.049	9.71e-2	25.93	9.33e-2	26.28	4.38	9.30
0.4	0.158	0.374	1.41e-1	22.72	0.142	1.29e-1	23.47	1.33e-1	23.20	4.47	9.46
0.5	0.195	0.638	1.84e-1	20.37	0.359	1.76e-1	20.78	1.81e-1	20.51	4.40	12.63
0.6	0.238	0.996	2.92e-1	16.37	0.498	2.82e-1	16.67	2.87e-1	16.53	4.40	11.18

the results for a kernel size 7×7, and Table 4b for a kernel size 15×15. See Fig. 4 for the different reconstructions in case of $d = 0.5$. Note that while FTVd gives slightly more accurate results for the larger kernel size, the performance in terms of computing time is worse by a factor of three to five.

5.3. Restoration of large images

In many scientific and engineering imaging applications, relevant images are of very large size and call for fast algorithms. To show the feasibility of using the proposed method for large-scale image restoration problems, we consider denoising of scanning transmission electron microscopy (STEM) images. Fig. 5 shows a STEM image of a lingual nerve (size $n = 2048$), the same image corrupted by impulsive Gaussian noise ($d = 0.3$, $\delta = 0.120$), and its reconstruction using our method ($\alpha = 0.9$, cpu time 180 seconds). The restored image agrees well with the exact one, with a relative error $e_{rel} = 9.02 \times 10^{-2}$ and PSNR = 27.30. A close inspection of the restoration (Figs. 5d–5f) indicates that the details of the STEM image are excellently preserved, which is especially important in medical applications.

Remark 5.1. *After the preparation of the original manuscript, the authors of the package FTVd released a new version (version 4.0) [26], which replaces the penalization approach for treating the constraints with an augmented Lagrangian approach and applies the classical alternating direction method. This circumvents certain drawbacks of the algorithm, thereby improving its computational efficiency. In our tests (again, with the default parameters), the new version is faster than version 3.01, but the margin depends on the noise level. For low noise levels, the performance is much improved from version 3.01 to 4.0, and is largely comparable with our algorithm. For higher noise levels, especially for impulsive Gaussian noise, the results by the newer version remain close to version 3.01, and thus our method still compares favorably with version 4.0 in these situations.*

Table 3: Results for *Lena*.

(a) salt-and-pepper noise, kernel size 7×7

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.151	0.063	6.29e-2	31.01	0.037	6.05e-2	31.35	5.43e-2	32.29	19.65	58.76
0.4	0.199	0.091	7.03e-2	30.05	0.049	6.71e-2	30.45	6.28e-2	31.03	19.48	51.37
0.5	0.250	0.123	7.81e-2	29.13	0.074	7.60e-2	29.37	7.36e-2	29.64	19.73	57.54
0.6	0.299	0.157	8.80e-2	28.10	0.135	8.76e-2	28.14	8.86e-2	28.04	19.64	54.96

(b) salt-and-pepper noise, kernel size 15×15

noise		balancing parameter			optimal parameter			FTVd (optimal)		cpu time	
d	δ	α	e_{rel}	PSNR	α	e_{rel}	PSNR	e_{rel}	PSNR	t_{split}	t_{FTVd}
0.3	0.150	0.093	1.02e-1	26.84	0.024	9.59e-2	27.35	7.94e-2	28.99	22.09	72.50
0.4	0.201	0.131	1.07e-1	26.42	0.029	9.94e-2	27.03	8.40e-2	28.50	22.44	75.11
0.5	0.249	0.168	1.11e-1	26.06	0.035	1.04e-1	26.69	8.88e-2	28.02	22.13	85.13
0.6	0.300	0.210	1.16e-1	25.72	0.042	1.09e-1	26.26	9.54e-2	27.39	22.09	74.68

(a) blurred and noisy   (b) restoration (α_b) (c) restoration (α_o) (d) restoration (FTVd)

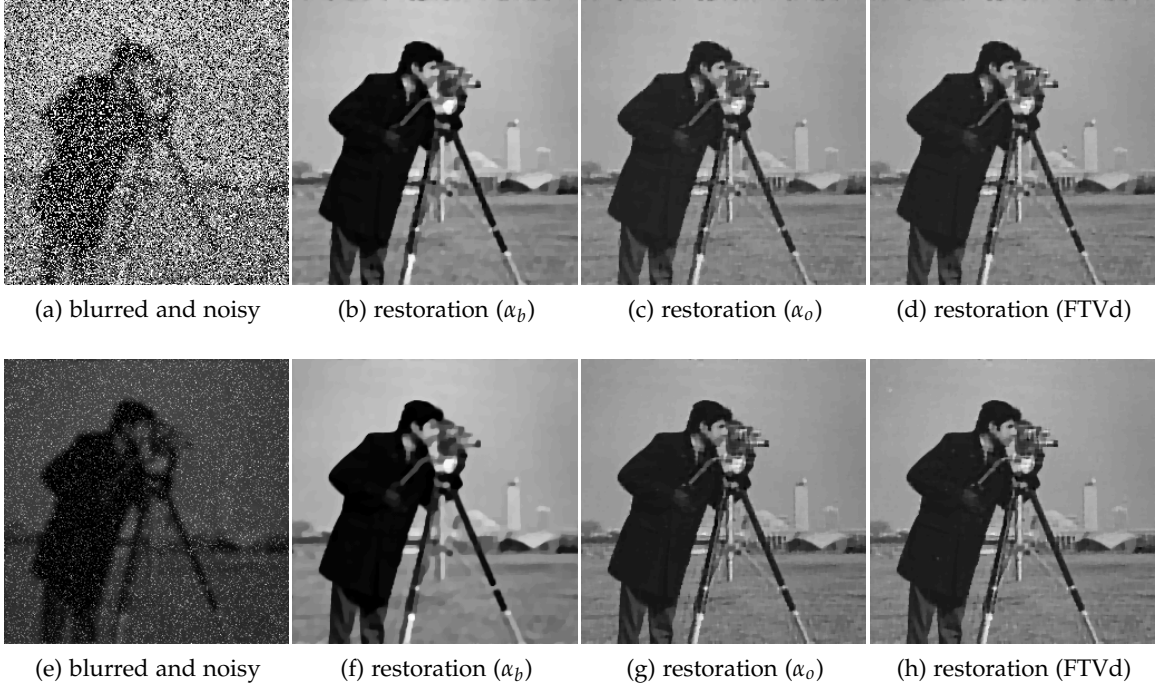(e) blurred and noisy   (f) restoration (α_b) (g) restoration (α_o) (h) restoration (FTVd)

Figure 3: Restorations of *cameraman* with kernel size 7×7. (a–d): salt and pepper noise, $d = 0.5$; (e–h): Gaussian noise, $d = 0.3$.

(a) blurred and noisy	(b) restoration (α_b)	(c) restoration (α_o)	(d) restoration (FTVd)
(e) blurred and noisy	(f) restoration (α_b)	(g) restoration (α_o)	(h) restoration (FTVd)

Figure 4: Restorations of *Lena* with salt-and-pepper noise and $d = 0.5$. (a–d): kernel size 7×7; (e–h): kernel size 15×15.

6. Conclusion

We have presented a novel splitting approach for ℓ^1-*TV* restoration of blurred and noisy images, together with an automatic parameter choice rule. Our numerical results indicate that the algorithm is very efficient and can deliver performance competitive with state-of-the-art methods such as FTVd. Since other recent methods [16, 10] are reported to yield results close to FTVd, our method compares favorably to those as well. The main advantage of the proposed approach is the good scaling with respect to problem size and difficulty, which makes it suitable for the restoration of very large images as occurring, for example, in microscopy applications.

To the authors' knowledge, the proposed parameter choice rule is the first such rule for ℓ^1-*TV* image restoration. Furthermore, it is very easy to implement, does not require knowledge of the noise level, and can yield results very close to or only slightly worse than the optimal one.

The combination of splitting and duality techniques seems promising for the efficient solution of a wide class of non-smooth problems in image restoration. Particularly, current work is concerned with its application to models with Poisson data.

Finally, it should be remarked that variants of the proposed algorithm such as using the leading point for the Lagrange multiplier update, or dispensing with leading points altogether, could in some cases give better accuracy or performance, but their theoretical convergence could not be ensured. It would be interesting to develop convergence results for these more aggressive algorithms.
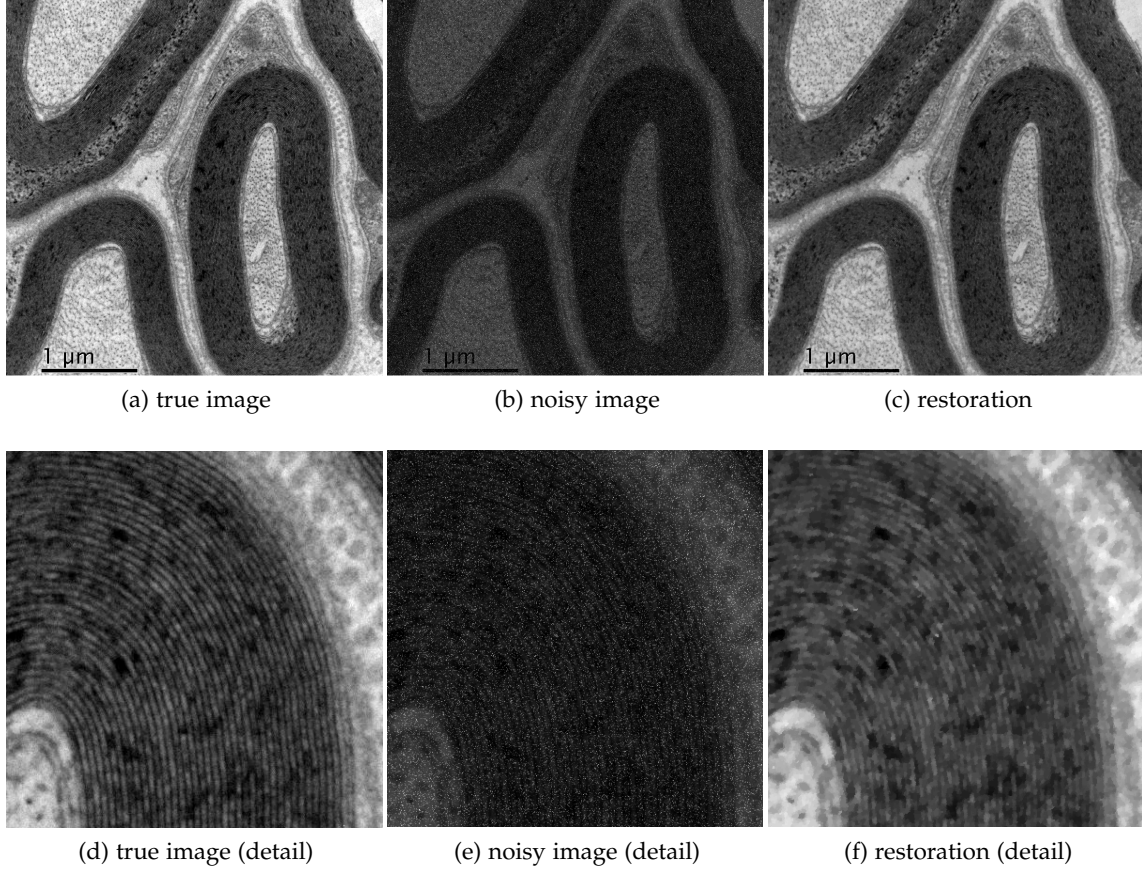
(a) true image	(b) noisy image	(c) restoration
(d) true image (detail)	(e) noisy image (detail)	(f) restoration (detail)

Figure 5: Denoising of STEM image.

A. Proof of Theorem 3.1

Using the property of the projection operator P_K on a convex set $K \subset X$, we have for all $h \in K$ and for all $z \in X$ that

$$(A.1) \qquad \langle z - P_K z, h - P_K z \rangle \leq 0$$

holds, from which it follows easily that

$$\|h - P_K z\|_{\ell^2}^2 \leq \|z - h\|_{\ell^2}^2 - \|z - P_K z\|_{\ell^2}^2.$$

20

Let (p^*, q^*, λ^*) be any saddle point of the functional $\mathcal{L}(p, q, \lambda)$. To simplify notation, let $\bar{z}^k = (\bar{p}^k, \bar{q}^k, \bar{\lambda}^k)$. By taking $h = p^*$ and $z = p^k - \gamma_p \mathcal{L}_p(\bar{z}^k)$, we find that

$$
\begin{aligned}
\|p^{k+1} - p^*\|_{\ell^2}^2 &\leq \|p^k - \gamma_p \mathcal{L}_p(\bar{z}^k) - p^*\|_{\ell^2}^2 - \|p^k - \gamma_p \mathcal{L}_p(\bar{z}^k) - p^{k+1}\|_{\ell^2}^2 \\
&= \|p^k - p^*\|_{\ell^2}^2 - \|p^k - p^{k+1}\|_{\ell^2}^2 - 2\gamma_p \langle \mathcal{L}_p(\bar{z}^k), p^{k+1} - p^* \rangle.
\end{aligned}
$$

Analogously, we can derive that

$$
\|q^{k+1} - q^*\|_{\ell^2}^2 \leq \|q^k - q^*\|_{\ell^2}^2 - \|q^k - q^{k+1}\|_{\ell^2}^2 - 2\gamma_q \langle \mathcal{L}_q(\bar{z}^k), q^{k+1} - q^* \rangle,
$$

and

$$
\|\lambda^{k+1} - \lambda^*\|_{\ell^2}^2 = \|\lambda^k - \lambda^*\|_{\ell^2}^2 - \|\lambda^{k+1} - \lambda^k\|_{\ell^2}^2 + 2\gamma_\lambda \langle \mathcal{L}_\lambda(\bar{z}^k), \lambda^{k+1} - \lambda^* \rangle.
$$

Let

$$
D(p, q, \lambda) = \gamma_q \gamma_\lambda \|p - p^*\|_{\ell^2}^2 + \gamma_p \gamma_\lambda \|q - q^*\|_{\ell^2}^2 + \gamma_p \gamma_q \|\lambda - \lambda^*\|_{\ell^2}^2.
$$

Then from the preceding three relations we have that

(A.2)
$$
\begin{aligned}
D(p^{k+1}, &q^{k+1}, \lambda^{k+1}) - D(p^k, q^k, \lambda^k) \\
\leq & - \gamma_q \gamma_\lambda \|p^{k+1} - p^k\|_{\ell^2}^2 - \gamma_p \gamma_\lambda \|q^{k+1} - q^k\|_{\ell^2}^2 - \gamma_p \gamma_q \|\lambda^{k+1} - \lambda^k\|_{\ell^2}^2 \\
& - \gamma_p \gamma_q \gamma_\lambda \left[\langle \mathcal{L}_p(\bar{z}^k), p^{k+1} - p^* \rangle + \langle \mathcal{L}_q(\bar{z}^k), q^{k+1} - q^* \rangle - \langle \mathcal{L}_\lambda(\bar{z}^k), \lambda^{k+1} - \lambda^* \rangle \right] \\
\leq & - \gamma_q \gamma_\lambda \|p^{k+1} - p^k\|_{\ell^2}^2 - \gamma_p \gamma_\lambda \|q^{k+1} - q^k\|_{\ell^2}^2 - \gamma_p \gamma_q \|\lambda^{k+1} - \lambda^k\|_{\ell^2}^2 \\
& - 2\gamma_p \gamma_q \gamma_\lambda \left[\langle \mathcal{L}_p(\bar{z}^k), p^{k+1} - \bar{p}^k \rangle + \langle \mathcal{L}_q(\bar{z}^k), q^{k+1} - \bar{q}^k \rangle - \langle \mathcal{L}_\lambda(\bar{z}^k), \lambda^{k+1} - \bar{\lambda}^k \rangle \right] \\
= & - \gamma_q \gamma_\lambda \left[\|p^{k+1} - \bar{p}^k\|_{\ell^2}^2 + \|\bar{p}^k - p^k\|_{\ell^2}^2 - 2\langle p^k - \gamma_p \mathcal{L}_p(\bar{z}^{k-1}) - \bar{p}^k, p^{k+1} - \bar{p}^k \rangle \right] \\
& - \gamma_p \gamma_\lambda \left[\|q^{k+1} - \bar{q}^k\|_{\ell^2}^2 + \|q^k - \bar{q}^k\|_{\ell^2}^2 - 2\langle q^k - \gamma_q \mathcal{L}_q(\bar{z}^{k-1}) - \bar{q}^k, q^{k+1} - \bar{q}^k \rangle \right] \\
& - \gamma_p \gamma_q \left[\|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2}^2 + \|\bar{\lambda}^k - \lambda^k\|_{\ell^2}^2 - \frac{2}{\rho} \langle \lambda^k + \rho \gamma_\lambda \mathcal{L}_\lambda(\bar{z}^{k-1}) - \bar{\lambda}^k, \lambda^{k+1} - \bar{\lambda}^k \rangle \right. \\
& \left. - (2 - \frac{2}{\rho})\langle \lambda^k - \bar{\lambda}^k, \lambda^{k+1} - \bar{\lambda}^k \rangle \right] \\
& + \gamma_p \gamma_q \gamma_\lambda \left[\langle \mathcal{L}_p(\bar{z}^{k-1}) - \mathcal{L}_p(\bar{z}^k), p^{k+1} - \bar{p}^k \rangle + \langle \mathcal{L}_q(\bar{z}^{k-1}) - \mathcal{L}_q(\bar{z}^k), q^{k+1} - \bar{q}^k \rangle \right. \\
& \left. - \langle \mathcal{L}_\lambda(\bar{z}^{k-1}) - \mathcal{L}_\lambda(\bar{z}^k), \lambda^{k+1} - \bar{\lambda}^k \rangle \right],
\end{aligned}
$$

where in deriving the second inequality we have made use of the inequality

$$
\langle \mathcal{L}_p(\bar{z}^k), \bar{p}^k - p^* \rangle + \langle \mathcal{L}_q(\bar{z}^k), \bar{q}^k - q^* \rangle - \langle \mathcal{L}_\lambda(\bar{z}^k), \bar{\lambda}^k - \lambda^* \rangle \geq 0.
$$

Appealing again to inequality (A.1) with $z = p^k - \gamma_p \mathcal{L}_p(\bar{z}^{k-1})$ and $h = p^{k+1}$, we obtain that

$$
\langle p^k - \gamma_p \mathcal{L}_p(\bar{z}^{k-1}) - \bar{p}^k, p^{k+1} - \bar{p}^k \rangle \leq 0.
$$

Similarly, we find that

$$\langle q^k - \gamma_q \mathcal{L}_q(\bar{z}^{k-1}) - \bar{q}^k, q^{k+1} - \bar{q}^k \rangle \leq 0,$$

and, by the definition of $\bar{\lambda}^k$,

$$\langle \lambda^k + \rho\gamma_\lambda \mathcal{L}_\lambda(\bar{z}^{k-1}) - \bar{\lambda}^k, \lambda^{k+1} - \bar{\lambda}^k \rangle = 0.$$

Note also the elementary inequality

$$2|\langle \lambda^k - \bar{\lambda}^k, \lambda^{k+1} - \bar{\lambda}^k \rangle| \leq \|\lambda^k - \bar{\lambda}^k\|_{\ell^2}^2 + \|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2}^2.$$

With the help these inequalities, the Cauchy-Schwarz inequality, the Lipschitz continuity of the Fréchet derivatives (with their Lipschitz constants denoted by L_p, L_q and L_λ, respectively, and $L = \max\{L_p, L_q, L_\lambda\}$) and Young's inequality $2ab \leq a^2 + b^2$, inequality (A.2) can be simplified to (with $c(\rho) = 1 - |1 - \frac{1}{\rho}| > 0$ for $\rho > 1/2$)

$$
\begin{aligned}
&D(p^{k+1}, q^{k+1}, \lambda^{k+1}) - D(p^k, q^k, \lambda^k) \\
&\leq -\gamma_q\gamma_\lambda \left[\|p^{k+1} - \bar{p}^k\|_{\ell^2}^2 + \|\bar{p}^k - p^k\|_{\ell^2}^2 \right] - \gamma_p\gamma_\lambda \left[\|q^{k+1} - \bar{q}^k\|_{\ell^2}^2 + \|q^k - \bar{q}^k\|_{\ell^2}^2 \right] \\
&\quad - c(\rho)\gamma_p\gamma_q \left[\|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2}^2 + \|\bar{\lambda}^k - \lambda^k\|_{\ell^2}^2 \right] \\
&\quad + 2\gamma_p\gamma_q\gamma_\lambda \left[\|\mathcal{L}_p(\bar{z}^{k-1}) - \mathcal{L}_p(\bar{z}^k)\|_{\ell^2} \|p^{k+1} - \bar{p}^k\|_{\ell^2} \right. \\
&\quad + \|\mathcal{L}_q(\bar{z}^{k-1}) - \mathcal{L}_q(\bar{z}^k)\|_{\ell^2} \|q^{k+1} - \bar{q}^k\|_{\ell^2} \\
&\quad + \|\mathcal{L}_\lambda(\bar{z}^k) - \mathcal{L}_\lambda(\bar{z}^{k+1})\|_{\ell^2} \|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2} \\
&\quad \left. + \|\mathcal{L}_\lambda(\bar{z}^{k-1}) - \mathcal{L}_\lambda(\bar{z}^k)\|_{\ell^2} \|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2} \right] \\
&\leq -\gamma_q\gamma_\lambda \left[\|p^{k+1} - \bar{p}^k\|_{\ell^2}^2 + \|\bar{p}^k - p^k\|_{\ell^2}^2 \right] - \gamma_p\gamma_\lambda \left[\|q^{k+1} - \bar{q}^k\|_{\ell^2}^2 + \|q^k - \bar{q}^k\|_{\ell^2}^2 \right] \\
&\quad - c(\rho)\gamma_p\gamma_q \left[\|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2}^2 + \|\bar{\lambda}^k - \lambda^k\|_{\ell^2}^2 \right] \\
&\quad + 2\gamma_p\gamma_q\gamma_\lambda \left[(\|p^k - \bar{p}^{k-1}\|_{\ell^2} + \|p^k - \bar{p}^k\|_{\ell^2} \right. \\
&\quad + \|q^k - \bar{q}^{k-1}\|_{\ell^2} + \|q^k - \bar{q}^k\|_{\ell^2} + \|\lambda^k - \bar{\lambda}^{k-1}\|_{\ell^2} + \|\lambda^k - \bar{\lambda}^k\|_{\ell^2}) \\
&\quad \left. \cdot (L_p\|p^{k+1} - \bar{p}^k\|_{\ell^2} + L_q\|q^{k+1} - \bar{q}^k\|_{\ell^2} + L_\lambda\|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2}) \right] \\
&\leq -\gamma_q\gamma_\lambda \left[(1 - 6\gamma_p L)\|p^{k+1} - \bar{p}^k\|_{\ell^2}^2 + (1 - 3\gamma_p L)\|\bar{p}^k - p^k\|_{\ell^2}^2 \right] \\
&\quad - \gamma_p\gamma_\lambda \left[(1 - 6\gamma_q L)\|q^{k+1} - \bar{q}^k\|_{\ell^2}^2 + (1 - 3\gamma_q L)\|q^k - \bar{q}^k\|_{\ell^2}^2 \right] \\
&\quad - c(\rho)\gamma_p\gamma_q \left[(1 - 6\gamma_\lambda \frac{L}{c(\rho)})\|\lambda^{k+1} - \bar{\lambda}^k\|_{\ell^2}^2 + (1 - 3\gamma_\lambda \frac{L}{c(\rho)})\|\bar{\lambda}^k - \lambda^k\|_{\ell^2}^2 \right] \\
&\quad + 3\gamma_p\gamma_q\gamma_\lambda L \left[\|p^k - \bar{p}^{k-1}\|_{\ell^2}^2 + \|q^k - \bar{q}^{k-1}\|_{\ell^2}^2 + \|\lambda^k - \bar{\lambda}^{k-1}\|_{\ell^2}^2 \right].
\end{aligned}
$$

Summing the above inequality from N to $M > N$ gives

$$(A.3) \quad D(p^{M+1}, q^{M+1}, \lambda^{M+1}) - D(p^N, q^N, \lambda^N)$$

$$\leq -\gamma_q \gamma_\lambda \left[(1 - 9\gamma_p L) \sum_{k=N}^{M-1} \|p^{k+1} - \overline{p}^k\|_{\ell^2}^2 + (1 - 3\gamma_p L) \sum_{k=N}^{M} \|\overline{p}^k - p^k\|_{\ell^2}^2 \right]$$

$$- \gamma_p \gamma_\lambda \left[(1 - 9\gamma_q L) \sum_{k=N}^{M-1} \|q^{k+1} - \overline{q}^k\|_{\ell^2}^2 + (1 - 3\gamma_q L) \sum_{k=N}^{M} \|q^k - \overline{q}^k\|_{\ell^2}^2 \right]$$

$$- c(\rho) \gamma_p \gamma_q \left[(1 - 9\frac{\gamma_\lambda L}{c(\rho)}) \sum_{k=N}^{M-1} \|\lambda^{k+1} - \overline{\lambda}^k\|_{\ell^2}^2 + (1 - 3\frac{\gamma_\lambda L}{c(\rho)}) \sum_{k=N}^{M} \|\overline{\lambda}^k - \lambda^k\|_{\ell^2}^2 \right]$$

$$+ 3\gamma_p \gamma_q \gamma_\lambda L \left[\|p^N - \overline{p}^{N-1}\|_{\ell^2}^2 + \|q^N - \overline{q}^{N-1}\|_{\ell^2}^2 + \|\lambda^N - \overline{\lambda}^{N-1}\|_{\ell^2}^2 \right]$$

$$- \gamma_q \gamma_\lambda (1 - 6\gamma_p L) \|p^{M+1} - \overline{p}^M\|_{\ell^2}^2 - \gamma_p \gamma_\lambda (1 - 6\gamma_q L) \|q^{M+1} - \overline{q}^M\|_{\ell^2}^2$$

$$- c(\rho) \gamma_p \gamma_q (1 - 6\frac{\gamma_\lambda L}{c(\rho)}) \|\lambda^{M+1} - \overline{\lambda}^M\|_{\ell^2}^2.$$

If the step sizes γ_p, γ_q and γ_λ satisfy the condition $\frac{9L}{c(\rho)} \max\{\gamma_p, \gamma_q, \gamma_\lambda\} < 1$ (i.e., $\gamma_0 = \frac{c(\rho)}{9L}$) the sequence $\{D(p^{M+1}, q^{M+1}, \lambda^{M+1})\}$ is uniformly bounded. Thus the sequence of iterates $\{(p^{M+1}, q^{M+1}, \lambda^{M+1})\}$ is uniformly bounded, and it contains a convergent subsequence $\{(p^{k_i}, q^{k_i}, \lambda^{k_i})\} \to (\tilde{p}, \tilde{q}, \tilde{\lambda})$. We prove that $(\tilde{p}, \tilde{q}, \tilde{\lambda})$ is a saddle point. To this end, we first observe that it follows from inequality (A.3) that the sequence $\{(p^{k+1} - \overline{p}^k, q^{k+1} - \overline{q}^k, \lambda^{k+1} - \overline{\lambda}^k)\}$ converges to zero. Therefore, on taking the limit as $k_i \to +\infty$, we get

$$(A.4) \qquad\qquad \tilde{p} = P_{B(1)}(\tilde{p} - \gamma_p \mathcal{L}_p(\tilde{p}, \tilde{q}, \tilde{\lambda})),$$

$$(A.5) \qquad\qquad \tilde{q} = P_{B(\alpha)}(\tilde{q} - \gamma_q \mathcal{L}_q(\tilde{p}, \tilde{q}, \tilde{\lambda})),$$

$$(A.6) \qquad\qquad \tilde{\lambda} = \tilde{\lambda} + \rho \gamma_\lambda \mathcal{L}_\lambda(\tilde{p}, \tilde{q}, \tilde{\lambda}).$$

Since $z = P_K(x)$ if and only if $\langle x - z, v - z \rangle \leq 0$ for all $v \in K$ (cf. [25, Lemma 2.11]), we find that (A.4) is equivalent to $\langle \mathcal{L}_p(\tilde{p}, \tilde{q}, \tilde{\lambda}), p - \tilde{p} \rangle \geq 0$ for all $p \in B(1)$. Analogously, it follows from (A.5) and (A.6) that $\langle \mathcal{L}_q(\tilde{p}, \tilde{q}, \tilde{\lambda}), q - \tilde{q} \rangle \geq 0$ for all $q \in B(\alpha)$ and $\mathcal{L}_\lambda(\tilde{p}, \tilde{q}, \tilde{\lambda}) = 0$. These are the necessary and sufficient conditions for the point $(\tilde{p}, \tilde{q}, \tilde{\lambda})$ to be a saddle point of the functional $\mathcal{L}(p, q, \lambda)$. Since the point (p^*, q^*, λ^*) has been chosen from the set of saddle points arbitrarily, we can assume that $(p^*, q^*, \lambda^*) = (\tilde{p}, \tilde{q}, \tilde{\lambda})$. The whole sequence converges to this point. Indeed, for any $\varepsilon > 0$ there exists N_ε such that

$$D(p^{N_\varepsilon}, q^{N_\varepsilon}, \lambda^{N_\varepsilon}) \leq \frac{\varepsilon}{2},$$

$$3\gamma_p \gamma_q \gamma_\lambda L \left[\|p^{N_\varepsilon} - \overline{p}^{N_\varepsilon-1}\|_{\ell^2}^2 + \|q^{N_\varepsilon} - \overline{q}^{N_\varepsilon-1}\|_{\ell^2}^2 + \|\lambda^{N_\varepsilon} - \overline{\lambda}^{N_\varepsilon-1}\|_{\ell^2}^2 \right] \leq \frac{\varepsilon}{2}.$$

Consequently, it follows from inequality (A.3) that $D(p^{M+1}, q^{M+1}, \lambda^{M+1}) < \varepsilon$ for all $M > N_\varepsilon + 1$. This proves the theorem.

REFERENCES

[1] W. K. ALLARD, *Total variation regularization for image denoising. I. Geometric theory*, SIAM J. Math. Anal., 39 (2007/08), pp. 1150–1190.

[2] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Trans. Signal Process., 45 (1997), pp. 913–917.

[3] S. Alliney and S. A. Ruzinsky, *An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation*, IEEE Trans. Signal Process., 42 (1994), pp. 618–627.

[4] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*, Stanford Mathematical Studies in the Social Sciences, vol. II, Stanford University Press, Stanford, Calif., 1958.

[5] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations (2nd ed.)*, vol. 147 of Applied Mathematical Sciences, Springer-Verlag, New York, 2006.

[6] A. C. Bovik, *Handbook of Image and Video Processing (Communications, Networking and Multimedia)*, Academic Press, Inc., Orlando, FL, USA, 2005.

[7] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837 (electronic).

[8] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.

[9] C. Clason, B. Jin, and K. Kunisch, *A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration*, Tech. Rep. 2009-015, SFB Report, 2009.

[10] Y. Dong, M. Hintermüller, and M. Neri, *An efficient primal-dual method for $\ell^1 tv$ image restoration*, SIAM J. Imaging Sci., 2 (2009), pp. 1168–1189.

[11] V. Duval, J.-F. Aujol, and Y. Gousseau, *The TVL1 model: a geometric point of view*, Multiscale Model. Simul., 8 (2009), pp. 154–189.

[12] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[13] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996.

[14] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, *Efficient minimization methods of mixed l2-l1 and l1-l1 norms for image restoration*, SIAM J. Sci. Comput., 27 (2006), pp. 1881–1902 (electronic).

[15] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Scientific Computation, Springer-Verlag, Berlin, 2008. Reprint of the 1984 original.

[16] X. Guo, F. Li, and M. Ng, *A fast ℓ_1-TV algorithm for image restoration*, SIAM J. Sci. Comput., 31 (2009), pp. 2322–2341.

[17] Q. Hu and J. Zou, *An iterative method with variable relaxation parameters for saddle-point problems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 317–338 (electronic).

[18] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, vol. 15 of Advances in Design and Control, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[19] T. Kärkkäinen, K. Kunisch, and K. Majava, *Denoising of smooth images using L^1-fitting*, Computing, 74 (2005), pp. 353–376.

[20] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.

[21] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vision, 20 (2004), pp. 99–120.

[22] L. D. Popov, *A modification of the Arrow-Hurwicz method for search of saddle points*, Math. Notes, 28 (1980), pp. 845–848.

[23] P. Rodríguez and B. Wohlberg, *Efficient minimization method for a generalized total variation functional*, IEEE Trans. Image Process., 18 (2009), pp. 322–332.

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

[25] A. Ruszczyński, *Nonlinear Optimization*, Princeton University Press, Princeton, NJ, 2006.

[26] M. Tao and J. Yang, *Alternating direction algorithms for total variation deconvolution in image reconstruction*. TR0918, Department of Mathmatics, Nanjing University, November, 2009, http://www.optimization-online.org/DB_HTML/2009/11/2463.html.

[27] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977.

[28] J. Yang, Y. Zhang, and W. Yin, *An efficient tvl1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31 (2009), pp. 2842–2865.

[29] W. Yin, D. Goldfarb, and S. Osher, *The total variation regularized L^1 model for multiscale decomposition*, Multiscale Model. Simul., 6 (2007), pp. 190–211 (electronic).