

Dual norm based iterative methods for image restoration*

Miyoun Jung*, Elena Resmerita[†] and Luminita A. Vese*

UCLA C.A.M. Report 09-88, November 2009

Abstract

A convergent iterative regularization procedure based on the square of a dual norm is introduced for image restoration models with general (quadratic or non-quadratic) convex fidelity terms. Iterative regularization methods have been previously employed for image deblurring or denoising in the presence of Gaussian noise, which use L^2 [38], [27], [39], and L^1 [15] data fidelity terms, with rigorous convergence results. Recently, Iusem-Resmerita [16] proposed a proximal point method using inexact Bregman distance for minimizing a convex function defined on a non-reflexive Banach space (e.g. $BV(\Omega)$), which is the dual of a separable Banach space. Based on this method, we investigate several approaches for image restoration such as image deblurring in the presence of noise or image deblurring via (cartoon+texture) decomposition. We show that the resulting algorithms approximate stably a true image. For image denoising-deblurring we consider Gaussian, Laplace, and Poisson noise models with the corresponding convex fidelity terms derived from the Bayesian approach. We test the behavior of proposed algorithms on synthetic and real images in several numerical experiments and compare the results with other state-of-the-art iterative procedures based on the total variation penalization as well as the corresponding existing one-step gradient descent implementations. The numerical experiments indicate that the iterative procedure yields high quality reconstructions and superior results to those obtained by one-step gradient descent, with faster computational time.

Key words: Proximal point method, iterative regularization, inexact Bregman distance, inverse problem, image restoration, bounded variation.

1 Introduction

Proximal point methods have been employed to stabilize ill-posed problems in infinite dimensional settings during the last decades, using L^2 (quadratic) [27] and L^1 data-fitting terms [15], respectively. Recently, [16] proposed a proximal point method for minimizing a general convex function defined on a non-reflexive Banach space which is the dual of a separable Banach space. Our aim here is to propose, based on that method, several iterative approaches for image restoration.

In the work of Tadmor et al [38], an iterative procedure for computing hierarchical (BV, L^2) decompositions has been proposed for image denoising, and this was extended to image restoration and segmentation in [39].

Osher et al [27] proposed another iterative procedure for approximating minimizers of quadratic objective functions, with the aim of image denoising or deblurring, providing significant improvements over the standard model introduced by Rudin, Osher, Fatemi (ROF) [34]. This turned out to be equivalent to a proximal point algorithm on a nonreflexive Banach space as well as to an augmented Lagrangian method for a convex minimization problem subject to linear constraints (see Yin et al. [46]).

*Research supported by a UC Dissertation Year Fellowship, by Austrian Science Fund Elise Richter Scholarship V82-N18 FWF, by NSF-DMS Award 0714945, and by NSF Expeditions in Computing Award CCF-0926127.

*Department of Mathematics, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 (gontaeng, lvese@math.ucla.edu)

[†]Industrial Mathematics Institute, Johannes Aeppler University, Altenbergerstraße 69, A-4040 Linz, Austria (elena.resmerita@jku.at)

In addition, He et al [15] extended the Bregman distance based iterative algorithm [27] to L^1 fidelity term by using a suitable sequence of penalty parameters. The authors of [15] also proved the well-definedness and the convergence of the algorithm with L^1 fidelity term, which is an iterative version of L^1 -TV considered by Chan and Esedoglu [8], and presented denoising results in the presence of Gaussian noise.

Benning and Burger [3] derived basic error estimates in the symmetric Bregman distance between the exact solution and the estimated solution satisfying an optimality condition, for general convex variational regularization methods. Furthermore, they investigated specific error estimates for several noise models in imaging such as Gaussian, Laplace, Poisson, and multiplicative with the corresponding quadratic or nonquadratic convex fidelity terms that were derived in the framework of MAP estimation.

Recently, Iusem and Resmerita [16] combine the idea of [27] with a surjectivity result, shown in [14] and [23], in order to obtain a proximal point method for minimizing more general convex functions, with interesting convergence properties. For the optimization case where the objective function is not necessarily quadratic, they use a positive multiple of an inexact Bregman distance associated with the square of the norm as the regularizing term; a solution is approached by a sequence of approximate minimizers of an auxiliary problem. Regarding the condition of being the dual of a Banach space, we recall that nonreflexive Banach spaces which are duals of other spaces include the cases of l^∞ and $L^\infty(\Omega)$, l^1 and $BV(\Omega)$ (the space of functions of bounded variation) which appear quite frequently in a large range of applications [25].

In Section 2, we first review the proximal point method proposed in [16]. In Section 3, we apply the proximal point method presented in Section 2 to general ill-posed operator equations, that are particularized in Section 4 to several image restoration problems. Thus, we show that the proximal point method combined either with an a priori or with an a posteriori stopping rule provides stable approximation of the true image. In the introduction of Section 4, we briefly mention preliminary work and the related models in image processing that we consider in this paper. Furthermore, in Section 4.1, we present several algorithms for image deblurring with Gaussian, Laplace, or Poisson noise models with corresponding convex fidelity terms, and in Section 4.2 we extend the iterative idea to image restoration via cartoon + texture model. Finally, in Section 5, several numerical results are presented for each image restoration model. Comparisons with other methods of similar spirit or one-step gradient descent models are also presented.

We mention that a very preliminary version of this work has been accepted for presentation and conference proceedings publication in ECCV 2010 [18].

2 Preliminaries

We recall the proximal point method and convergence results of Iusem-Resmerita from [16].

Let X be a nonreflexive Banach space and X^* its topological dual. For $u^* \in X^*$ and $u \in X$, we denote by $\langle u^*, u \rangle = u^*(u)$ the duality pairing. Denote by $h(u) = \frac{1}{2}\|u\|^2$, for $u \in X$.

For $\varepsilon > 0$, the ε -subdifferential of h at a point $u \in X$ is [12]

$$\partial_\varepsilon h(u) = \{u^* \in X^* : h(v) - h(u) - \langle u^*, v - u \rangle \geq -\varepsilon, \forall v \in X\}.$$

The normalized ε -duality mapping of X , introduced by Gossez [14], extends the notion of duality mapping as follows

$$J_\varepsilon(u) = \{u^* \in X^* : \langle u^*, u \rangle + \varepsilon \geq \frac{1}{2}\|u^*\|^2 + \frac{1}{2}\|u\|^2\}. \quad (1)$$

An equivalent definition for the ε -duality mapping is

$$J_\varepsilon(u) = \partial_\varepsilon \left(\frac{1}{2}\|u\|^2 \right).$$

The inexact Bregman distances with respect to the convex function h and to an ε -subgradient ξ of h were defined in [16] as follows:

$$D^\varepsilon(v, u) = h(v) - h(u) - \langle \xi, v - u \rangle + \varepsilon. \quad (2)$$

Note that when h is Fréchet differentiable, in which case $\xi = h'(u)$, we have $D^0(v, u) = D(v, u)$, where D denotes the standard Bregman distance related to h (see, e.g., [4]). Also, $D^\varepsilon(v, u) \geq 0$ for any $u, v \in X$ and

$$D^\varepsilon(u, u) = \varepsilon > 0, \quad \forall u \in X.$$

Given $\varepsilon \geq 0$ and a function $g : X \rightarrow \mathbb{R} \cup \{+\infty\}$, we say that $\bar{u} \in \text{dom } g = \{u \in X : g(u) < \infty\}$ is an ε -minimizer of g when

$$g(\bar{u}) \leq g(u) + \varepsilon \quad (3)$$

for all $u \in \text{dom } g$.

Consider exogenous sequences $\{\varepsilon_k\}$, $\{\lambda_k\}$ of positive numbers satisfying the following two assumptions:

H1) The sequence $\{\varepsilon_k\}$ is summable, i.e.,

$$\sum_{k=0}^{\infty} \varepsilon_k < \infty, \quad (4)$$

H2) The sequence $\{\lambda_k\}$ is bounded above.

The number ε_k is some sort of error bound for the inexact minimization performed at the k -th iteration of the algorithm, while $\{\lambda_k\}$ is the regularization parameter used in the same iteration.

The following proximal point algorithm is proposed in [16]:

Initialization

Take $u_0 \in \text{dom } g$ and $\xi_0 \in J_{\varepsilon_0}(u_0)$.

Iterative step

Let $k \in \mathbb{N}$. Assume that $u_k \in \text{dom } g$ and $\xi_k \in J_{\varepsilon_k}(u_k)$ are given. We proceed to define u_{k+1} , ξ_{k+1} . Define $D^{\varepsilon_k}(u, u_k) = h(u) - h(u_k) - \langle \xi_k, u - u_k \rangle + \varepsilon_k$ and $\bar{\varepsilon}_k = \lambda_k \varepsilon_{k+1}$.

Determine $u_{k+1} \in \text{dom } g$ as an $\bar{\varepsilon}_k$ -minimizer of the function $g_k(u)$ defined as

$$g_k(u) = g(u) + \lambda_k D^{\varepsilon_k}(u, u_k), \quad (5)$$

that is to say, in view of (3),

$$g(u_{k+1}) + \lambda_k D^{\varepsilon_k}(u_{k+1}, u_k) \leq g(u) + \lambda_k D^{\varepsilon_k}(u, u_k) + \bar{\varepsilon}_k \quad (6)$$

for all $u \in \text{dom } g$.

Let $\eta_{k+1} \in \partial g(u_{k+1})$ and $\xi_{k+1} \in J_{\varepsilon_{k+1}}(u_{k+1})$ such that

$$\eta_{k+1} + \lambda_k (\xi_{k+1} - \xi_k) = 0. \quad (7)$$

The results of well-definedness and convergence of the algorithm are recalled below.

Proposition 2.1. *Let X be a Banach space and $g : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function. Then the sequence $\{u_k\}$ generated by the above algorithm is well defined.*

Proposition 2.2. *Let X be a Banach space and $g : X \rightarrow \mathbb{R} \cup \{+\infty\}$ a proper, convex and lower semicontinuous function. Assume that z is a minimizer of g . Define $\beta_k = D^{\varepsilon_k}(z, u_k)$, $\gamma_k = D^{\varepsilon_k}(u_{k+1}, u_k)$, with D^{ε_k} as in (2). If H1 and H2 hold, then the sequence $\{u_k\}$ generated by the above algorithm has the following properties:*

i)

$$g(u_{k+1}) \leq g(u_k) + \lambda_k(\varepsilon_k + \varepsilon_{k+1}), \quad (8)$$

ii)

$$\beta_{k+1} - \beta_k + \gamma_k + \frac{g(u_{k+1}) - g(z)}{\lambda_k} \leq \varepsilon_{k+1}, \quad (9)$$

iii) The sequence $\{\beta_k\}$ is bounded,

iv) The sequence $\{\gamma_k\}$ is summable,

v) The sequence $\{g(u_k) - g(z)\}$ is summable.

Theorem 2.3. *Let X be a Banach space which is the dual of a separable Banach space and $g : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, convex and weakly* lower semicontinuous function. Assume that g has minimizers. If H1 and H2 hold, then the sequence $\{u_k\}$ generated by the above algorithm is bounded, $\lim_{k \rightarrow \infty} g(u_k) = \min_{u \in X} g(u)$, and all cluster points of $\{u_k\}$ in the weak* topology of X are minimizers of g .*

In the next section, we apply this general Iusem-Resmerita's algorithm [16] to linear ill-posed inverse problems.

3 Applications to ill-posed operator equations

Large classes of inverse problems can be formulated as operator equations

$$Ku = y.$$

Define the residual $g(u) = S(y, Ku)$ for any $u \in X$, where S is a similarity measure (see, e.g., [30], [3]). The iterative method described in the previous section can be applied to this exact data case setting and provides weakly* approximations for the solutions of the equation, provided that at least a solution exists.

Usually, the above equations are ill-posed, in the sense that the operator K may not be continuously invertible which means that small perturbations in the data y lead to high oscillations in the solutions.

Consider that only noisy data y^δ are given, such that

$$S(y^\delta, y) \leq r(\delta), \quad \delta > 0, \quad (10)$$

where $r = r(\delta)$ is a function of δ with

$$\lim_{\delta \rightarrow 0^+} r(\delta) = 0. \quad (11)$$

Denote

$$g^\delta(u) = S(y^\delta, Ku).$$

In this section, we show that the iterative method presented in the previous section yields a regularization method for such problems.

We will use the following

Assumptions (A)

- The operator $K : X \rightarrow Y$ is linear and bounded, and yields an ill-posed problem.
- X and Y are Banach spaces. In addition, X is the topological dual of a separable Banach space.
- The similarity measure S is such that

1. The function $g^\delta(u) = S(y^\delta, Ku)$ is convex and weakly* lower semicontinuous.
- 2.

$$\lim_{\delta \rightarrow 0_+} g^\delta(u_\delta) = 0 \quad \Rightarrow \quad \lim_{\delta \rightarrow 0_+} Ku_\delta = y, \quad (12)$$

whenever $\{u_\delta\}_{\delta > 0}$ is a net in X , the last limit being understood with respect to the norm of Y .

We consider only constant parameter $\lambda_k = \lambda$ for any $k \in \mathbb{N}$, where λ is a positive number. Then, the algorithm described in the previous section reads as follows in the current setting:

Algorithm 3.1. Take $u_0 \in \text{dom } g^\delta$ and $\xi_0 \in J_{\varepsilon_0}(u_0)$.

Iterative step

Let $k \in \mathbb{N}$. Assume that $u_k \in \text{dom } g^\delta$ and $\xi_k \in J_{\varepsilon_k}(u_k)$ are given. We proceed to define u_{k+1} , ξ_{k+1} . Define $D^{\varepsilon_k}(u, u_k) = h(u) - h(u_k) - \langle \xi_k, u - u_k \rangle + \varepsilon_k$ and $\bar{\varepsilon}_k = \lambda \varepsilon_{k+1}$.

Determine $u_{k+1} \in \text{dom } g^\delta$ as an $\bar{\varepsilon}_k$ -minimizer of the function $g_k^\delta(u)$ defined as

$$g_k^\delta(u) = g^\delta(u) + \lambda D^{\varepsilon_k}(u, u_k),$$

that is to say, in view of (3),

$$g^\delta(u_{k+1}) + \lambda D^{\varepsilon_k}(u_{k+1}, u_k) \leq g^\delta(u) + \lambda D^{\varepsilon_k}(u, u_k) + \bar{\varepsilon}_k$$

for all $u \in \text{dom } g^\delta$.

Let $\eta_{k+1} \in \partial g^\delta(u_{k+1})$ and $\xi_{k+1} \in J_{\varepsilon_{k+1}}(u_{k+1})$ such that

$$\eta_{k+1} + \lambda(\xi_{k+1} - \xi_k) = 0.$$

A posteriori strategy. We choose the stopping index based on a discrepancy type principle, similarly to the one in [27]:

$$k_* = \max\{k \in \mathbb{N} : g^\delta(u_k) \geq \tau r(\delta)\}, \quad (13)$$

for some $\tau > 1$.

We show below that the stopping index is finite and that Algorithm 3.1 together with the stopping rule stably approximate solutions of the equation.

Proposition 3.2. Let $\tilde{u} \in X$ verify $K\tilde{u} = y$, assume that inequality (10) is satisfied, assumptions (A) hold and that the sequence $\{\varepsilon_k\}$ is such that

$$\sum_{k=1}^{\infty} k\varepsilon_k < \infty. \quad (14)$$

Moreover, let the stopping index k_* be chosen according to (13). Then k_* is finite, the sequence $\{\|u_{k_*(\delta)}\|\}_\delta$ is bounded and hence, as $\delta \rightarrow 0$, there exists a weakly*-convergent subsequence $\{u_{k_*(\delta_n)}\}_n$ in X . If the following conditions hold, then the limit of each weakly* convergent subsequence is a solution of $Ku = y$:

- i) $\{k_*(\delta)\}_{\delta > 0}$ is unbounded;
- ii) Weak*-convergence of $\{u_{k_*(\delta_n)}\}_n$ to some $u \in X$ implies convergence of $\{Ku_{k_*(\delta_n)}\}_n$ to Ku , as $n \rightarrow \infty$ with respect to the weak topology of Y .

Proof: First, we show that the stopping index k_* is finite. Denote $\bar{\theta} = \sum_{k=1}^{\infty} k\varepsilon_k$. Finiteness of $\bar{\theta}$ implies $\theta = \sum_{k=0}^{\infty} \varepsilon_k < \infty$. One can check that an inequality similar to (9) holds when g is replaced by g^δ (see the proof of Proposition 3.2 in [16]), for any $k \in \mathbb{N}$, $k \geq 1$:

$$D^{\varepsilon_k}(\tilde{u}, u_k) + D^{\varepsilon_{k-1}}(u_k, u_{k-1}) + \frac{g^\delta(u_k) - g^\delta(\tilde{u})}{\lambda} \leq D^{\varepsilon_{k-1}}(\tilde{u}, u_{k-1}) + \varepsilon_k.$$

By summing up and using $g^\delta(\tilde{u}) = S(y^\delta, y)$ and (10), it follows

$$D^{\varepsilon_k}(\tilde{u}, u_k) + \sum_{j=1}^k D^{\varepsilon_{j-1}}(u_j, u_{j-1}) + \frac{1}{\lambda} \sum_{j=1}^k g^\delta(u_j) \leq \frac{kr(\delta)}{\lambda} + D^{\varepsilon_0}(\tilde{u}, u_0) + \sum_{j=1}^k \varepsilon_j. \quad (15)$$

Now (8) written for g^δ implies

$$g^\delta(u_k) \leq g^\delta(u_{k-1}) + \lambda(\varepsilon_{k-1} + \varepsilon_k) \leq g^\delta(u_j) + \lambda(\varepsilon_k + \varepsilon_j) + 2\lambda \sum_{i=j+1}^{k-1} \varepsilon_i,$$

$\forall k \geq 2, j \leq k-1$.

By summing up the last inequalities upon j , one obtains

$$\begin{aligned} kg^\delta(u_k) &\leq \sum_{j=1}^k g^\delta(u_j) + \lambda \sum_{j=1}^{k-1} \varepsilon_j + 2\lambda \sum_{j=2}^{k-1} (j-1)\varepsilon_j + \lambda(k-1)\varepsilon_k \\ &\leq \sum_{j=1}^k g^\delta(u_j) + \tilde{M}, \end{aligned}$$

where \tilde{M} depends on λ, θ and $\bar{\theta}$.

We combine the last inequality with (15) and get, based on the nonnegativity of the inexact Bregman distances $D^{\varepsilon_k}(\tilde{u}, u_k)$ and $D^{\varepsilon_{j-1}}(u_j, u_{j-1})$,

$$kg^\delta(u_k) \leq kr(\delta) + M, \quad (16)$$

where

$$M := \lambda D^{\varepsilon_0}(\tilde{u}, u_0) + \tilde{M}.$$

Inequality (16) written for $k = k_*$ together with (13) yield

$$\tau r(\delta) \leq g^\delta(u_{k_*}) \leq r(\delta) + \frac{M}{k_*}. \quad (17)$$

This implies

$$k_* \leq \frac{M}{(\tau-1)r(\delta)}, \quad (18)$$

which means that the stopping index k_* is finite.

Since $\sum_{j=1}^k D^{\varepsilon_{j-1}}(u_j, u_{j-1}) + \frac{1}{\lambda} \sum_{j=1}^k g^\delta(u_j) \geq 0$, it follows from (15) that

$$D^{\varepsilon_{k_*}}(\tilde{u}, u_{k_*}) \leq \frac{r(\delta)k_*}{\lambda} + D^{\varepsilon_0}(\tilde{u}, u_0) + \theta \leq \frac{M}{\lambda(\tau-1)} + D^{\varepsilon_0}(\tilde{u}, u_0) + \theta. \quad (19)$$

Therefore, the sequence $\{D^{\varepsilon_{k_*}(\delta)}(\tilde{u}, u_{k_*}(\delta))\}_{\delta>0}$ is bounded. By proceeding similarly as in [16], one concludes that the sequence $\{u_{k_*}(\delta)\}_{\delta>0}$ is also bounded. We show this below, for the sake of completeness:

$$\delta \geq D^{\varepsilon_{k_*}(\delta)}(\tilde{u}, u_{k_*}(\delta)) = \frac{\|\tilde{u}\|^2}{2} - \frac{\|u_{k_*}(\delta)\|^2}{2} - \langle \xi_{k_*}(\delta), \tilde{u} - u_{k_*}(\delta) \rangle + \varepsilon_{k_*}(\delta) \geq \frac{\|\tilde{u}\|^2}{2} - \langle \xi_{k_*}(\delta), \tilde{u} \rangle + \frac{\|\xi_{k_*}(\delta)\|^2}{2},$$

using (1) in the last inequality. Thus,

$$\frac{\|\xi_{k_*}(\delta)\|^2}{2} \leq \langle \xi_{k_*}(\delta), \tilde{u} \rangle + \delta \leq \|\xi_{k_*}(\delta)\| \|\tilde{u}\| + \delta, \quad (20)$$

which shows that the sequence $\{\xi_{k_*(\delta)}\}$ is bounded. Using now (1) and (20), one obtains that $\{u_{k_*(\delta)}\}_{\delta>0}$ is also bounded, as

$$\frac{\|u_{k_*(\delta)}\|^2}{2} \leq \langle \xi_{k_*(\delta)}, u_{k_*(\delta)} \rangle + \varepsilon_{k_*(\delta)} \leq \|\xi_{k_*(\delta)}\| \|u_{k_*(\delta)}\| + \theta.$$

Taking into account that X is provided with a weak* topology (see Assumption (A)), this implies that $\{u_{k_*(\delta)}\}_{\delta>0}$ has a weakly*-convergent subsequence, denoted the same, to some $u \in X$.

If $\{k_*(\delta)\}_{\delta>0}$ is unbounded, then $\{g^\delta(u_{k_*(\delta)})\}_{\delta>0}$ converges to zero as $\delta \rightarrow 0$, due to (17). That is, $Ku_{k_*(\delta)} \rightarrow y$, as $\delta \rightarrow 0$ (see (12)). Now hypothesis ii) implies that $Ku_{k_*(\delta)} \rightarrow K\tilde{u}$ weakly, as $\delta \rightarrow 0$ and thus, $K\tilde{u} = y$.

□

A priori strategy. One could stop Algorithm 3.1 by using a stopping index which depends on the noise level only, by contrast to the previously chosen k_* which depends also on the noisy data y^δ . More precisely, one chooses

$$k(\delta) \sim \frac{1}{r(\delta)}. \quad (21)$$

One can show that the sequence $\{u_{k(\delta)}\}_{\delta>0}$ converges weakly* to solutions of the equation as $\delta \rightarrow 0$. Indeed, inequality (19) written for $k(\delta)$ instead of k_* implies that the sequence $\{D^{\varepsilon_{k(\delta)}}(\tilde{u}, u_{k(\delta)})\}_{\delta>0}$ is bounded and, as above, the sequence $\{u_{k(\delta)}\}_{\delta>0}$ is bounded. Hence, a subsequence of it, denoted also by $\{u_{k(\delta)}\}_{\delta>0}$ converges weakly* to some $u \in X$. Now (16) together with (21) show that $\{g^\delta(u_{k(\delta)})\}_{\delta>0}$ converges to zero as $\delta \rightarrow 0$. The rest of the proof is similar to the one given above. Thus, the following proposition holds true:

Proposition 3.3. *Let $\tilde{u} \in X$ verify $K\tilde{u} = y$, assume that inequality (10) is satisfied, assumptions (A) hold and that the sequence $\{\varepsilon_k\}$ obeys (14). Moreover, let the stopping index $k(\delta)$ be chosen according to (21). Then the sequence $\{\|u_{k(\delta)}\|\}_\delta$ is bounded and hence, as $\delta \rightarrow 0$, there exists a weakly*-convergent subsequence $\{u_{k(\delta_n)}\}_n$ in X . If the following condition holds, then the limit of each weak* convergent subsequence is a solution of $Ku = y$: Weak*-convergence of $\{u_{k(\delta_n)}\}_n$ to some $u \in X$ implies convergence of $\{Ku_{k(\delta_n)}\}_n$ to Ku , as $n \rightarrow \infty$ with respect to the weak topology of Y .*

4 Several proximal point based approaches for image restoration

We present a few image restoration settings which fit the theoretical framework investigated in the previous section. First, we briefly mention prior relevant work. Also, note that we use simpler notations f , g , and g_k instead of y^δ , g^δ , and g_k^δ respectively used in the previous section. All image functions are defined on an open and bounded domain Ω of \mathbb{R}^N and take real values.

In Tadmor et al [38], [39] an iterative procedure for computing hierarchical (BV, L^2) decompositions has been proposed for image restoration. For image deblurring in the presence of Gaussian noise, assuming the degradation model $f = Ku + n$, the iterative method from [39] computes a sequence u_k , such that each u_{k+1} is the minimizer of $\lambda_0 2^k \|v_k - Ku_{k+1}\|_2^2 + \int_\Omega |Du_{k+1}|$, where $v_{-1} = f$, $k = 0, 1, \dots$ and $v_k = Ku_{k+1} + v_{k+1}$. The partial sum $\sum_{j=0}^k u_j$ is a denoised-deblurred version of f , and converges to f as $k \rightarrow \infty$.

Osher et al [27] proposed an iterative algorithm with quadratic fidelity term S and a convex regularization functional h (e.g. TV-regularizer $h(u) = |u|_{BV(\Omega)} = \int_\Omega |Du| dx \approx \int_\Omega |\nabla u| dx$): starting with u_0 and ξ_0 , $u_{k+1} \in BV(\Omega)$ is a minimizer of the functional g_k defined on $BV(\Omega)$

$$g_k(u) = S(f, Ku) + \lambda D(u, u_k) = \frac{1}{2} \|f - Ku\|_2^2 + \lambda [h(u) - h(u_k) - \langle \xi_k, u - u_k \rangle], \quad (22)$$

with $\xi_{k+1} = \xi_k + \frac{1}{\lambda} K^*(f - Ku_{k+1}) \in \partial h(u_{k+1})$, and a parameter $\lambda > 0$. They proved the well-definedness and the convergence of iterates u_k , and presented some applications to denoising or

deblurring in the presence of Gaussian noise, giving significant improvements over standard ROF model [34], [35] which is

$$u = \arg \min_u \left\{ \frac{1}{2} \|f - Ku\|_2^2 + \lambda |u|_{BV(\Omega)} \right\}. \quad (23)$$

The reader is referred also to [6] where convergence rates for method (22) are established.

He et al [15] generalized the regularization procedure [27] for image denoising models with non-quadratic convex fidelity terms, by using the varying parameter $\frac{\lambda}{2^k}$ with $\lambda > 0$ instead of a fixed parameter $\lambda > 0$, inspired by [36] and [38]: starting with u_0 and ξ_0 , $u_{k+1} \in BV(\Omega)$ is a minimizer of the functional g_k defined on $BV(\Omega)$,

$$g_k(u) = S(f, u) + \frac{\lambda}{2^k} D(u, u_k) = S(f, u) + \frac{\lambda}{2^k} [h(u) - h(u_k) - \langle \xi_k, u - u_k \rangle], \quad (24)$$

where $\xi_{k+1} = \xi_k - \frac{2^k}{\lambda} \partial_u S(f, u_{k+1})$, and $S(f, u) = s(f - u)$ with s being a nonnegative, convex, and positively homogeneous functional, which is continuous with respect to weak* convergence in BV . For example, $s(f - u) = \frac{1}{2} \|f - u\|_2^2$ or $s(f - u) = \|f - u\|_1$. The authors in [15] proved the well-definedness and the convergence of the iterative algorithm with L^1 fidelity term (as an iterative version of the L^1 -TV model considered by Chan and Esedoglu [8]), and presented denoising results in the presence of Gaussian noise.

Le et al [9] proposed a total variation model for denoising in the presence of Poisson noise:

$$\min_u \left\{ \int_{\Omega} (u - f \log u) dx + \lambda |u|_{BV(\Omega)} \right\}, \quad (25)$$

while Benning and Burger in a recent, parallel work [3] with ours, investigated a more general fidelity term. In particular, they consider a fidelity term that we will also use here,

$$S(f, Ku) = \int_{\Omega} \left[f \log \left(\frac{f}{Ku} \right) - f + Ku \right] dx.$$

Kim and Vese [20] proposed an image decomposition and restoration model in the presence of blur and noise using the Sobolev spaces $W^{s,p}(\Omega)$ for $s \in \mathbb{R}$, $1 \leq p \leq \infty$, by considering the following degradation model

$$f = K(u + v) + r$$

where u is the cartoon part, $v = \Delta q$ for some $q \in W^{-\alpha+2,p}$ is the texture part, and r is a small residual noise. They recovered the deblurred image $u + v = u + \Delta q$ by minimizing the functional

$$\min_{u,q} \left\{ \mu \int_{\Omega} (f - K(u + \Delta q))^2 dx + |u|_{BV(\Omega)} + \lambda \|q\|_{W^{s,p}} \right\} \quad (26)$$

where $\mu, \lambda > 0$, $s \geq 0$, $s = -\alpha + 2$, $\alpha > 0$ and $\|\cdot\|_{W^{s,p}}$ is a norm on $W^{s,p}$ [20].

We set below the general iterative algorithm for image deblurring in the presence of noise, and we consider the Gaussian, Laplace, or Poisson noise models with the corresponding (convex) fidelity terms. We also extend the iterative idea to the image deblurring model via decomposition [20] by considering the Sobolev space $H^{-1}(\Omega)$ for v , which is a dual space. Improved restoration results with faster convergence are obtained.

4.1 Image deblurring in the presence of noise

Let X, Y be Banach spaces, $X \subset Y$, where X is the dual of a separable Banach space. We consider degradation models of the form

$$f = \mathcal{F}(Ku, n)$$

where $f \in Y$ is the observed noisy data, $K : Y \rightarrow Y$ is the convolution operator with a blurring kernel K (i.e. $Ku := K * u$), $u \in X$ is the ideal image we want to recover, and n is noise of some known probability distribution function; we notice that the data f is expressed as a function of Ku and n (in a linear or nonlinear way).

Here, we present three noise models in infinite dimension prompted by the corresponding finite dimensional models based on the conditional probability $p(f|u)$: the Gaussian model, the Laplace model, the Poisson model. In finite dimensional spaces, the conditional probability $p(f|u)$ of the data f with given image u is the component of the Bayesian model that is influenced by the type of distribution of the noise (and hence the noisy data f).

Assuming $X = BV(\Omega)$ and $Y = L^p(\Omega)$ with $p = 1$ or 2 , we define

$$h(u) = \frac{1}{2} \|u\|_{BV}^2 = \frac{1}{2} \left(\int_{\Omega} |u| dx + \int_{\Omega} |Du| dx \right)^2.$$

In addition, we consider convex functions of the form $g(u) = S(f, Ku)$ for any $u \in X$, where S is convex with respect to u for a fixed f . Then, we propose the following general iterative algorithm to recover u :

Algorithm 4.1. Let $u_0 = 0$, $\xi_0 = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$.

- Given (u_k, ξ_k) , define $\bar{\varepsilon}_k = \lambda \varepsilon_{k+1}$, and compute u_{k+1} as an $\bar{\varepsilon}_k$ -*minimizer* of the functional below

$$g_k(u) = S(f, Ku) + \lambda [h(u) - h(u_k) - \langle \xi_k, u - u_k \rangle + \varepsilon_k].$$

- Determine

$$\eta_{k+1} \in \partial_u S(f, Ku_{k+1}), \quad \xi_{k+1} \in J_{\varepsilon_{k+1}}(u_{k+1})$$

such that

$$\eta_{k+1} + \lambda(\xi_{k+1} - \xi_k) = 0.$$

Remark 4.1. We refer to [1, Section 3.4.1] for the relation between Gateaux differentiability and $\bar{\varepsilon}_k$ -*minimizers*. If u is an $\bar{\varepsilon}_k$ -*minimizer* of the Gateaux-differentiable function $g_k(u)$, then we must have $\|\partial g_k(u)\| \leq \bar{\varepsilon}_k$ (note that this is a necessary condition). In practice, we obtain $u_{k+1} \in X$ as an $\bar{\varepsilon}_k$ -*minimizer* of $g_k(u)$ by solving the usual Euler-Lagrange equation for g_k given by

$$0 = \partial g_k(\tilde{u}) = \partial_u S(f, K\tilde{u}) + \lambda \partial h(\tilde{u}), \quad (27)$$

because $g_k(\tilde{u}) = \inf_{u \in X} g_k(u) \leq g_k(u) + \bar{\varepsilon}_k$ for any $u \in \text{dom}(g_k)$ (any 0-*minimizer* is also an $\bar{\varepsilon}_k$ -*minimizer*). We use time-dependent gradient descent to approximate the solution \tilde{u} by solving

$$\frac{\partial \tilde{u}}{\partial t} = -\partial g_k(\tilde{u}) \quad (28)$$

to steady state. We have also computed u_{k+1} as the solution \tilde{u} of one of the two possible equations,

$$\frac{\partial \tilde{u}}{\partial t} = -\partial g_k(\tilde{u}) \pm \bar{\varepsilon}_k. \quad (29)$$

Similar numerical results are obtained if we use (28) or (29).

Remark 4.2. For the numerical calculations, we assume that we work, in practice, with functions $u \in W^{1,1}(\Omega) \subset BV(\Omega)$. Also, we make the functional $h(u)$ differentiable by substituting it with $h(u) \approx \frac{1}{2} \left(\int_{\Omega} \sqrt{\varepsilon^2 + u^2} dx + \int_{\Omega} \sqrt{\varepsilon^2 + |\nabla u|^2} dx \right)^2$ for a small number $\varepsilon > 0$. The subgradient in this case becomes

$$\partial h(u) \approx \left(\int_{\Omega} \sqrt{\varepsilon^2 + u^2} + \sqrt{\varepsilon^2 + |\nabla u|^2} dx \right) \left[\frac{u}{\sqrt{\varepsilon^2 + u^2}} - \nabla \cdot \frac{\nabla u}{\sqrt{\varepsilon^2 + |\nabla u|^2}} \right].$$

Remark 4.3. We can start with $u_0 = 0$, $\xi_0 = 0$, $\varepsilon_0 = 0$ (except for the Poisson noise mode, where we need $u_0 > 0$). Although our theory considers positive parameters ε_k in order to ensure existence of the iterates u_k , one could still initialize the algorithm with $u_0 = 0$, $\xi_0 = 0$, $\varepsilon_0 = 0$ in many situations, including the particular ones investigated below. In such cases, existence of u_1 and ξ_1 is not based on the surjectivity result employed in [16], but rather on direct analysis of the function $S(f, Ku) + \lambda h(u)$ to be minimized.

4.1.1 Gaussian noise

If the degradation model is $f = Ku + n \in Y = L^2(\Omega)$ with Gaussian distributed noise and with the expectation Ku , the conditional probability $p(f|Ku)$ is described by

$$p(f|Ku) \sim e^{-\frac{\|f - Ku\|_2^2}{2\delta^2}},$$

where δ^2 is the variance of the noise n . Maximizing $p(f|Ku)$ with respect to u , is equivalent to minimizing $-\ln p(f|Ku)$, thus we obtain a convex fidelity term to be minimized for $u \in BV(\Omega)$,

$$S(f, Ku) = \frac{1}{2}\|f - Ku\|_2^2.$$

The function $g(u) = S(f, Ku)$ satisfies the conditions enforced in Assumptions (A) in dimension one and two. Moreover, let $r(\delta) = \delta^2/2 = \frac{1}{2}\|f - Ku_*\|_2^2$ with true image u_* - see (10).

Since such a quadratic S is Gâteaux-differentiable, its subgradient is given by

$$\partial_u S(f, Ku) = K^*(Ku - f)$$

which leads to

$$\xi_{k+1} = \xi_k - \frac{1}{\lambda}K^*(Ku_{k+1} - f).$$

Numerical Algorithm We have the following numerical algorithm:

I. Let $u_0 = 0$, $\xi_0 = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$ until $\|f - Ku_{k+1}\|_2 \leq \delta$:

- $u = u_{k+1}$: $\frac{\partial u}{\partial t} = K^*(f - Ku) - \lambda[\partial h(u) - \xi_k]$.
- ξ_{k+1} : $\xi_{k+1} = \xi_k + \frac{1}{\lambda}K^*(f - Ku_{k+1})$.

In addition, following [27], we let $\xi_k = \frac{K^*v_k}{\lambda}$ so that we have

$$v_{k+1} = v_k + (f - Ku_{k+1}).$$

With $v_0 = 0$, since $\lambda\xi_0 = 0 = K^*0 = K^*v_0$, we may conclude inductively that $\lambda\xi_k \in R(K^*)$, and hence there exists $v_k \in Y^* = L^2(\Omega)$ such that $\lambda\xi_k = K^*v_k$. Hence, we can have the following alternative numerical algorithm:

II. Let $u_0 = 0$, $v_0 = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$ until $\|f - Ku_{k+1}\|_2 \leq \delta$:

- $u = u_{k+1}$: $\frac{\partial u}{\partial t} = K^*(f + v_k - Ku) - \lambda\partial h(u)$.
- v_{k+1} : $v_{k+1} = v_k + (f - Ku_{k+1})$.

Table 1: CORRESPONDING VALUES (u_* : ORIGINAL IMAGE)

GAUSSIAN NOISE, SHAPE IMAGE, $\delta = \|f - Ku_*\|_2 = 15$, $\lambda = 0.1$ (FIG. 1)

(1) $\ f - Ku_{k+1}\ _2$, (2) $\ u_* - u_{k+1}\ _2$, (3) $g_k(u_{k+1})$, (4) $g_k(u_k) + \bar{\varepsilon}_k$, $\bar{\varepsilon}_k = \lambda \cdot \frac{20}{2^{k+1}}$							
$k + 1$		1	2	3	4	5	6
$\varepsilon_{k+1} = 0$	(1)	28.9304	16.3110	14.9658	14.3968	13.9388	13.5462
	(2)	38.4751	27.4330	21.8608	19.9682	<i>19.4057</i>	19.5140
	(3)/ 10^3	2.7519	0.1441	0.0759	0.0854	0.0826	0.0771
	(4)/ 10^3	27.4454	0.4185	0.1330	0.1120	0.1036	0.0971
$\varepsilon_{k+1} = +\frac{20}{2^{k+1}}$	(1)	28.2685	16.3597	14.9721	14.3981	13.9391	13.5465
	(2)	37.9876	27.4497	21.8568	19.9700	<i>19.4047</i>	19.5141
	(3)/ 10^3	2.7523	0.1612	0.0776	0.0858	0.0827	0.0771
	(4)/ 10^3	27.4464	0.4011	0.1346	0.1125	0.1038	0.0972

LAPLACE NOISE, RECTANGLES IMAGE, $\delta = \|f - Ku_*\|_1 = 10$, $\lambda = 0.05$ (FIG. 5)

(1) $\ f - Ku_{k+1}\ _1$, (2)-(4) are given above, $\bar{\varepsilon}_k = \lambda \cdot \frac{10}{2^{k+1}}$						
$k + 1$		1	2	3	4	5
$\varepsilon_{k+1} = -\frac{10}{2^{k+1}}$	(1)	19.2793	10.3728	9.9617	9.9699	9.9550
	(2)	34.2209	6.2486	2.4155	<i>2.1612</i>	2.3347
	(3)	23.9261	11.4119	9.5010	9.7609	9.9989
	(4)	31.1721	19.6543	10.5603	10.0555	10.0168
$\varepsilon_{k+1} = 0$	(1)	15.9363	10.2188	9.9628	9.9585	9.9544
	(2)	25.8151	5.3234	2.4314	<i>2.1524</i>	2.4106
	(3)	23.7543	11.7663	9.6135	9.7623	9.9969
	(4)	30.9221	16.1863	10.3438	10.0253	9.9998
$\varepsilon_{k+1} = +\frac{10}{2^{k+1}}$	(1)	13.3385	10.1475	9.9690	9.9498	9.9340
	(2)	18.0126	4.8230	2.5858	<i>2.4072</i>	2.5466
	(3)	24.2908	11.4858	9.6829	10.0388	9.9914
	(4)	31.1721	13.7135	10.3350	10.0627	9.9967

POISSON NOISE, SIMPLE IMAGE, $\delta = S(f, Ku_*) = 0.5020$, $\lambda = 0.005$ (FIG. 8)

(1) $\int_{\Omega} [f \log \left(\frac{f}{Ku_{k+1}} \right) - f + Ku_{k+1}] dx$, (2)-(4) are given above, $\bar{\varepsilon}_k = \lambda \cdot \frac{10}{2^{k+1}}$							
$k + 1$		1	2	3	4	5	6
$\varepsilon_{k+1} = -\frac{10}{2^{k+1}}$	(1)	2.2478	0.5592	0.4953	0.4919	0.4898	0.4878
	(2)	18.0863	5.3421	3.7770	<i>3.7674</i>	3.8510	3.9571
	(3)	8.2437	0.6579	0.4975	0.4968	0.4920	0.4881
	(4)	79.7468	2.2853	0.5780	0.5047	0.4966	0.4921
$\varepsilon_{k+1} = 0$	(1)	2.0330	0.5737	0.4974	0.4908	0.4886	0.4868
	(2)	17.1028	5.5338	3.8835	<i>3.7856</i>	3.8462	3.9366
	(3)	8.2443	0.8847	0.5324	0.4996	0.4913	0.4878
	(4)	79.7468	2.0705	0.5925	0.5068	0.4955	0.4909
$\varepsilon_{k+1} = +\frac{10}{2^{k+1}}$	(1)	1.8291	0.5929	0.5031	0.4920	0.4888	0.4869
	(2)	16.1027	5.8568	4.0336	3.8203	3.8542	3.9384
	(3)	8.2575	1.0563	0.5662	0.5059	0.4931	0.4881
	(4)	79.7468	1.8666	0.6117	0.5125	0.4967	0.4911

Table 2: STOPPING CRITERIA AND COMPARISONS

Model	$r(\delta)$	Comparison with
Gaussian	$\frac{\delta^2}{2} = \frac{1}{2} \ f - Ku_*\ _2^2$	iterative algorithm using TV (22) or RO (23)
Laplace	$\delta = \ f - Ku_*\ _1$	one-step L^1 -TV model
Poisson	$\delta = \int_{\Omega} [f \log \left(\frac{f}{Ku_*} \right) - f + Ku_*] dx$	one-step $S(f, Ku) + TV$ (25)
Decomposition (without noise)	no stopping criteria	$S(f, Ku, Kv) + \frac{\lambda}{2} \ u\ _{BV}^2 + \frac{\mu}{2} \ v\ _{H^{-1}}^2$

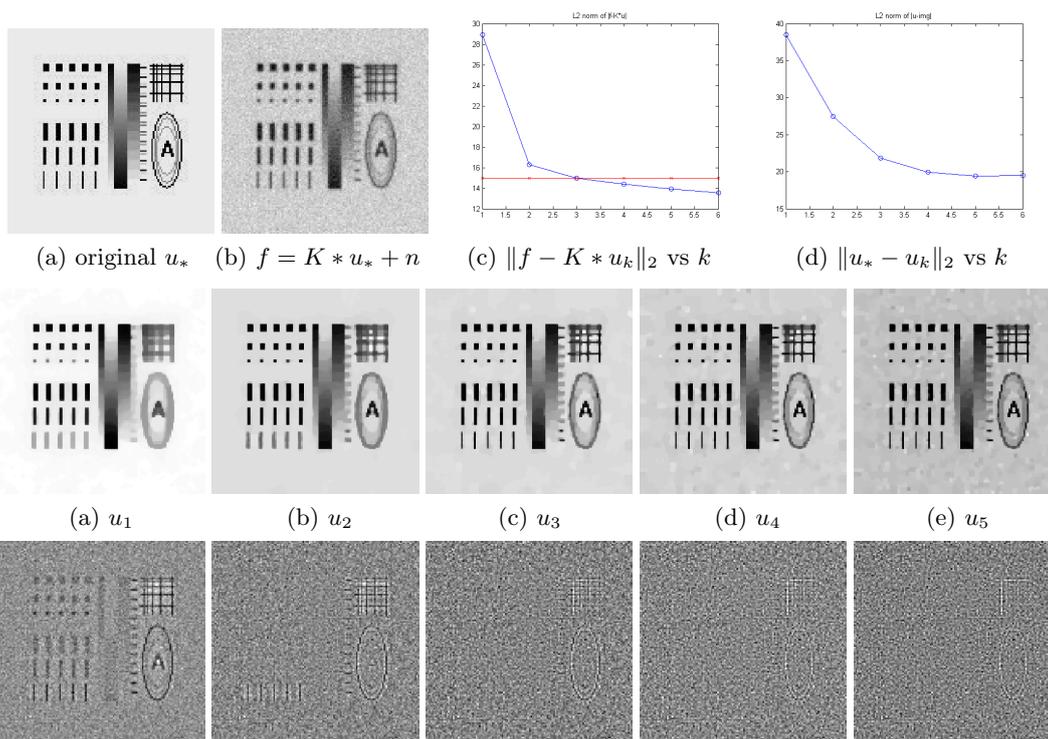


Figure 1: Gaussian noise model using our iterative method. 2nd and 3rd row: recovered images u_k and the corresponding residuals $f - K * u_k$. Data: Gaussian blur kernel K with standard deviation $\sigma_b = 0.7$, and Gaussian noise with $\delta = 15$. Parameters: $\lambda = 0.1$. $\|f - K * u_3\|_2 = 14.9658$. u_3 is the best recovered image (RMSE=21.8608).

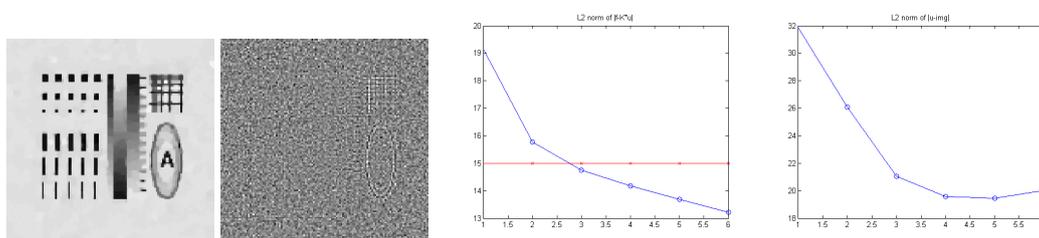


Figure 2: Results of the iterative algorithm (22) proposed by Osher et al with the same data from Fig. 1. The best recovered image u_3 ($\|f - K * u_3\|_2 = 14.7594$, RMSE=21.0500), residual $f - K * u_3$, and energies $\|f - K * u_k\|_2$, $\|u_* - u_k\|_2$ vs k .

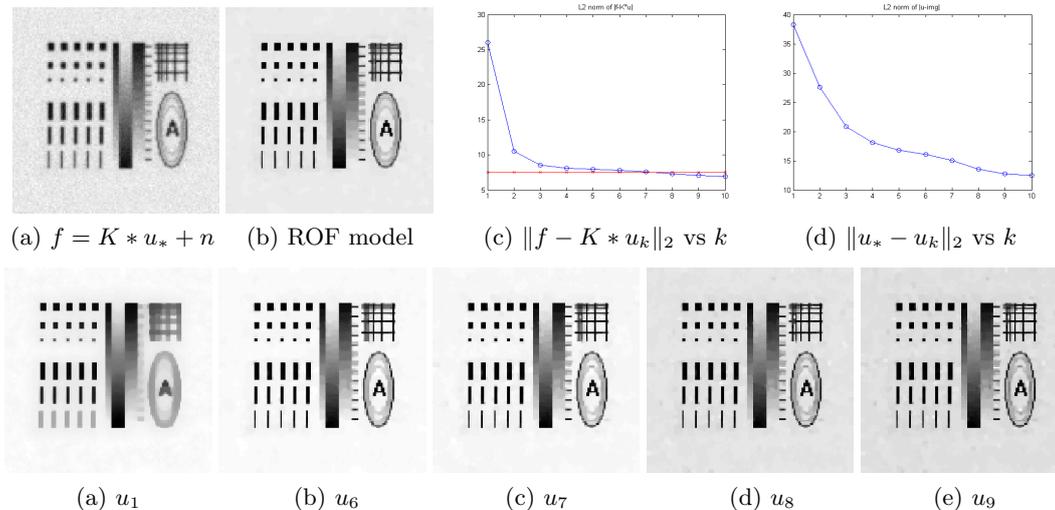


Figure 3: A priori strategy $k(\delta) \sim 1/r(\delta)$, and comparison with RO model (RMSE=16.5007). Data: same blur kernel K and parameter $\lambda = 0.1$ as in Fig. 1, but different Gaussian noise level with $\delta = 7.5$. u_8 is the best recovered image (RMSE=13.5407).

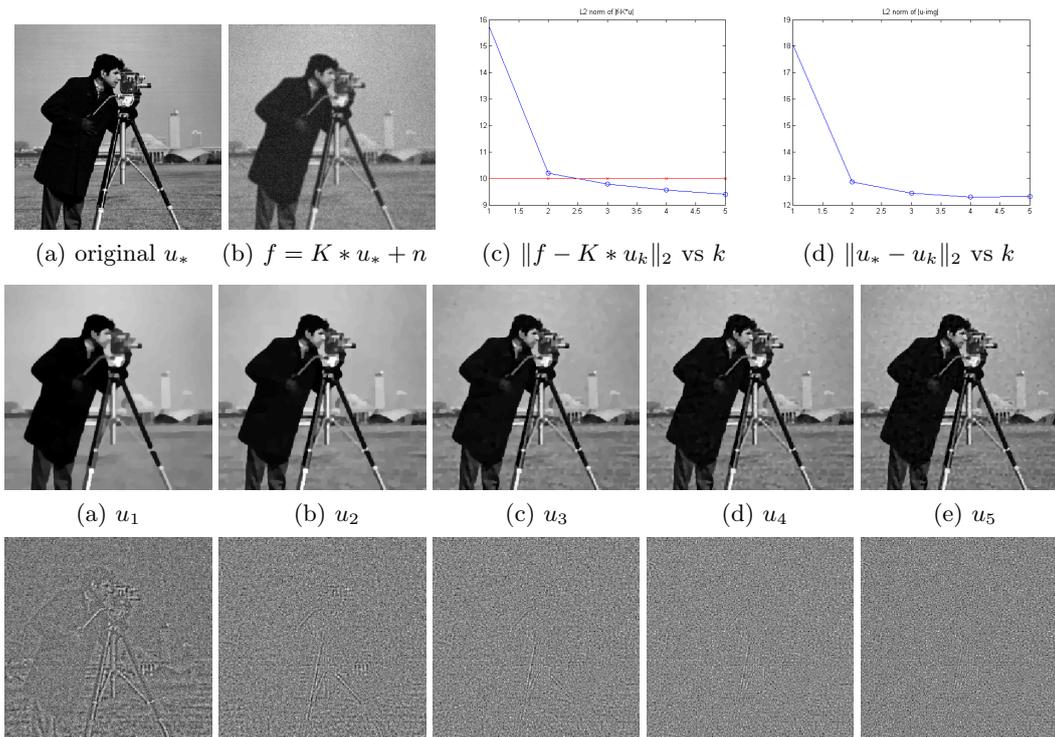


Figure 4: Gaussian noise model. Data: Gaussian blur kernel K with standard deviation $\sigma_b = 1$, and Gaussian noise with $\delta = 10$. Parameters: $\lambda = 0.1$. u_3 is the best recovered image (RMSE=12.2217).

4.1.2 Laplace noise

If the degradation model is $f = Ku + n \in Y = L^1(\Omega)$ with n being a Laplace distributed random variable with mean zero and variance $2\delta^2$, we have

$$p(f|Ku) \sim e^{-\frac{\|f - Ku\|_1}{\delta}}.$$

Then, similarly, we minimize with respect to u the quantity $-\ln p(f|Ku)$, thus we are led to consider the convex fidelity term

$$S(f, Ku) = \int_{\Omega} |f - Ku| dx.$$

Moreover, let $r(\delta) = \delta = \|f - Ku_*\|_1$ with true image u_* . Again, the function $g(u) = S(f, Ku)$ satisfies the conditions in Assumptions (A) in dimension one and two.

Unless $Ku \equiv f$, one can think of $\partial_u S(f, Ku) = K^* \text{sign}(Ku - f)$ almost everywhere, and moreover we have

$$\xi_{k+1} = \xi_k - \frac{1}{\lambda} K^* \text{sign}(Ku_{k+1} - f) \quad a.e.$$

Numerical algorithm We have the following numerical algorithm:

I. Let $u_0 = 0$, $\xi_0 = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$ until $\|f - Ku_{k+1}\|_1 \leq \delta$:

- $u = u_{k+1}$: $\frac{\partial u}{\partial t} = K^* \text{sign}(f - Ku) - \lambda[\partial h(u) - \xi_k]$.
- ξ_{k+1} : $\xi_{k+1} = \xi_k + \frac{1}{\lambda} K^* \text{sign}(f - Ku_{k+1})$.

Now again letting $\xi_k = \frac{K^* v_k}{\lambda}$, we can have

$$v_{k+1} = v_k + \text{sign}(f - Ku_{k+1}) \quad a.e.$$

With $v_0 = 0$, since $\lambda \xi_0 = 0 = K^* 0 = K^* v_0$, we may conclude inductively that $\lambda \xi_k \in R(K^*)$, and hence there exists $v_k \in Y^* = L^\infty(\Omega)$ such that $\lambda \xi_k = K^* v_k$. Hence, we have the alternative numerical algorithm:

II. Let $u_0 = 0$, $v_0 = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$ until $\|f - Ku_{k+1}\|_1 \leq \delta$:

- $u = u_{k+1}$: $\frac{\partial u}{\partial t} = K^* [\text{sign}(f - Ku) + v_k] - \lambda \partial h(u)$.
- v_{k+1} : $v_{k+1} = v_k + \text{sign}(f - Ku_{k+1})$.

4.1.3 Poisson noise

We consider the degradation model $f = P(Ku) \in Y = L^1(\Omega)$ perturbed with Poisson noise P and positive almost everywhere, where the operator K has positive values. Then the conditional probability $p(f|Ku)$ is modeled in discrete terms as (where i corresponds to a pixel),

$$p(f|Ku) = \prod_{i=1}^m \frac{(Ku)_i^{f_i}}{f_i!} e^{-(Ku)_i}.$$

Thus, we are led to consider the convex fidelity term

$$S(f, Ku) = \int_{\Omega} \left[f \log \left(\frac{f}{Ku} \right) - f + Ku \right] dx,$$

as a natural extension to deblurring of model [9]. Note that $g(u) = S(f, Ku) = KL(f, Ku)$ (where KL is the Kullback-Leibler divergence) might not be finite at any $u \in BV(\Omega)$, so its domain is possibly smaller than $BV(\Omega)$, by contrast to the Gaussian and Laplace noise cases. Therefore, care must be taken when analyzing this case. Moreover, let $r(\delta) = \delta = \int_{\Omega} \left[f \log \left(\frac{f}{Ku_*} \right) - f + Ku_* \right] dx$ with true image u_* . The function $g(u) = KL(f, Ku)$ is convex and weakly* lower semicontinuous. The latter property is true since the function g is lower semicontinuous with respect to the L^1 -norm (see, e.g., [31]) and since weak* convergence in $BV(\Omega)$ implies strong convergence in $L^1(\Omega)$. In fact, assumption (A) is satisfied for this particular function g - see for instance the proof of Proposition 5.2 in [31].

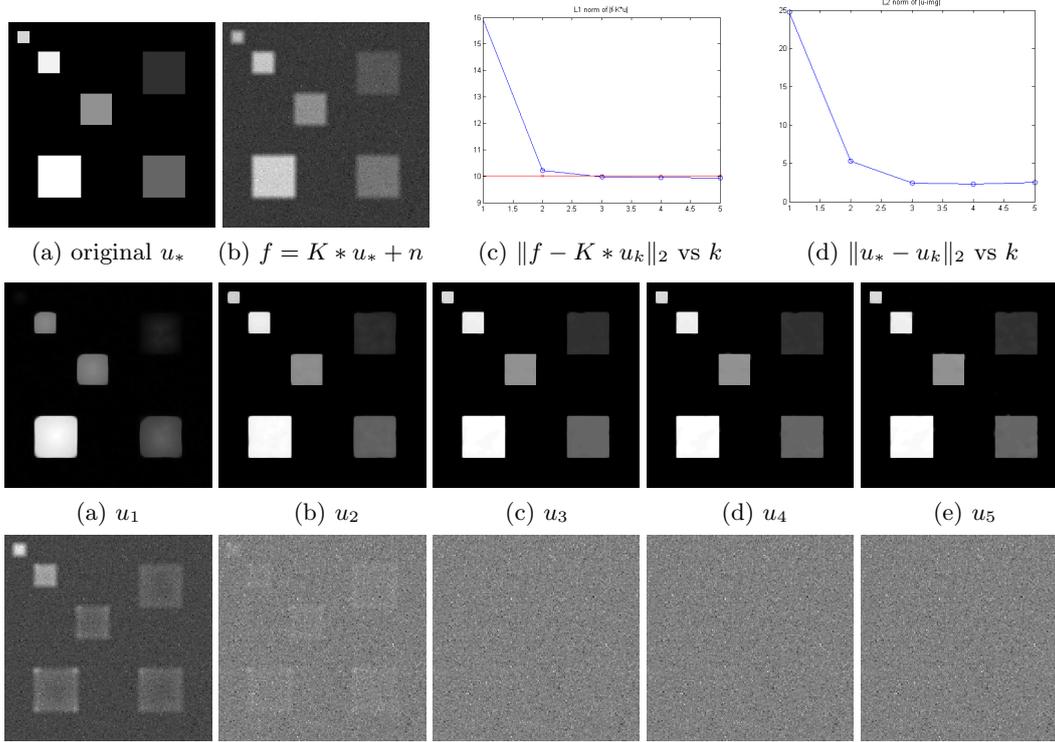


Figure 5: Laplace noise model. Data: Gaussian blur kernel K with with standard deviation $\sigma_b = 3$, and Laplace noise with $\delta = 10$. Parameters: $\lambda = 0.05$. $\|f - K * u_3\|_1 = 9.9629$. u_3 is the best recovered image (RMSE=2.4417).

If K is the identity operator, the function $u \mapsto \bar{g}(u) = \int_{\Omega} \left[f(x) \log\left(\frac{f(x)}{u(x)}\right) - f(x) + u(x) \right] dx$ is finite when $u \in L^1(\Omega)$, $u > 0$ a.e. and the integral is finite, and $\bar{g}(u) = +\infty$ otherwise. The function has subgradients at any $u > 0$ a.e. such that $f/u \in L^\infty(\Omega)$. Moreover, such subgradient is unique and given by $\xi = 1 - \frac{f}{u}$ - see, for instance, [2], Proposition 2.7, page 117.

In the case when K is not the identity operator, we assume that K satisfies $\text{Range}(K) \subset E$, where $E = \{v : 0 < \text{essinf}(v) \leq \text{esssup}(v) < \infty\}$ (that is, the values of K are pointwise bounded a.e. and pointwise bounded away from zero a.e.). Then we can use the following proposition, according to [47]:

Proposition 4.4. *Let $f \in L^\infty(\Omega)$, $f > 0$ a.e. and such that $\int_{\Omega} f(x) \log f(x) dx < \infty$. Then,*

$$\partial_u S(f, Ku) = \left\{ K^* \left(1 - \frac{f}{Ku} \right) \right\},$$

whenever $Ku \in L^\infty(\Omega)$ and $\text{essinf}(Ku) > 0$, that is, whenever $0 < b_1 < Ku \leq b_2 < \infty$ a.e. for some positive numbers b_1, b_2 .

In what follows, it is additionally assumed that $f \in L^\infty(\Omega)$, $f > 0$ a.e. and such that $\int_{\Omega} f(x) \log f(x) dx < \infty$. Thus Ku_{k+1} is pointwise bounded a.e. and pointwise bounded away from zero a.e. for any k . Consequently, the iterative formula becomes

$$\xi_{k+1} = \xi_k - \frac{1}{\lambda} K^* \left(1 - \frac{f}{Ku_{k+1}} \right).$$

Numerical algorithm We have the following numerical algorithm:

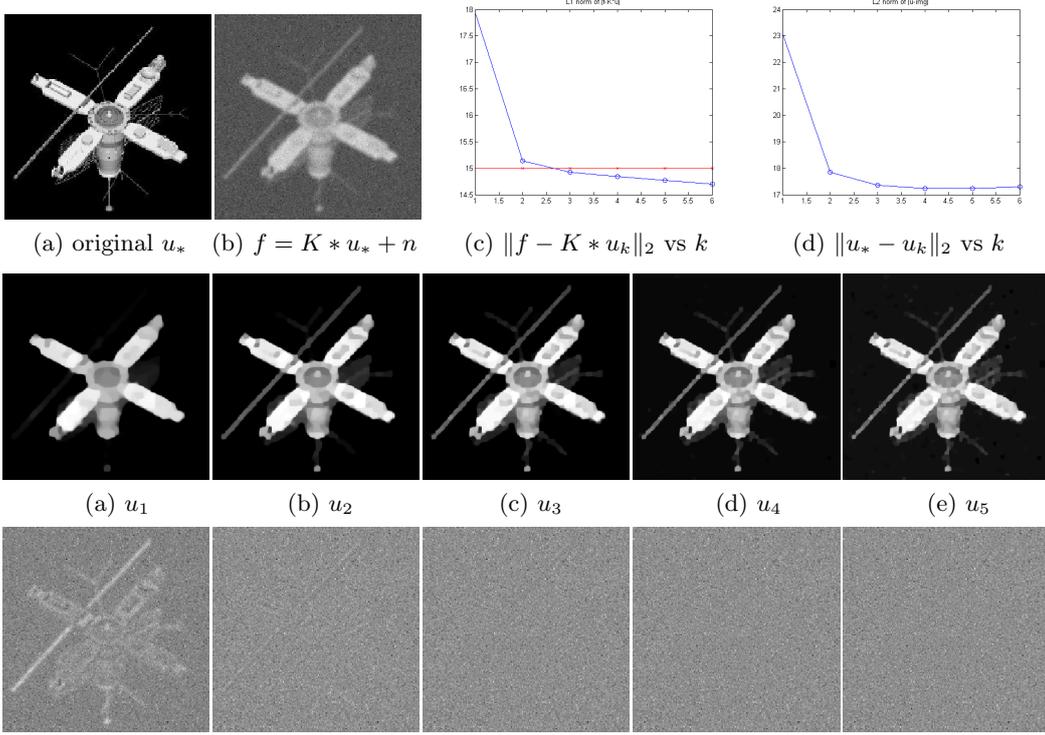


Figure 6: Laplace noise model. Data: Gaussian blur kernel K with $\sigma_b = 2$, and Laplace noise with $\delta = 15$. Parameters: $\lambda = 0.02$. $\|f - K * u_3\|_1 = 14.9234$. u_3 is the best recovered image (RMSE=17.3498).

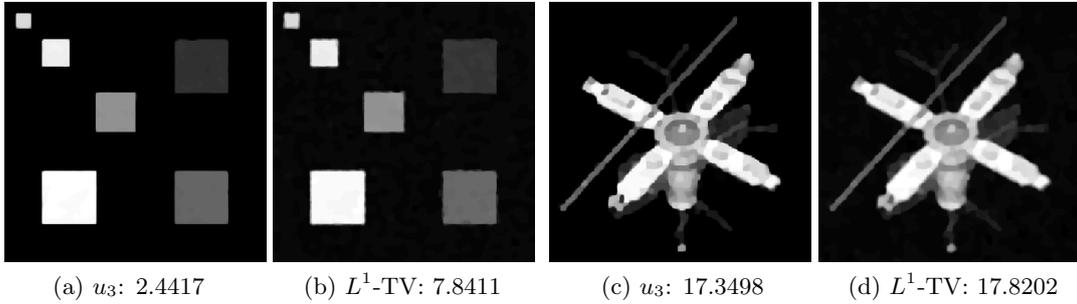


Figure 7: Comparison with one-step denoising-deblurring L^1 -TV. (a), (c): our iterative method. (b), (d): one-step L^1 -TV ($\|f - K * u\|_1$: (b) 9.8649 , (d) 14.9650). Recovered images u and RMSE values.

I. Let $u_0 > 0$ with $Ku_0 \in L^\infty(\Omega)$ and $\text{ess\,inf}(Ku_0) > 0$, $\xi_0 = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$:

- $u = u_{k+1}$: $\frac{\partial u}{\partial t} = K^* \left(\frac{f}{Ku} - 1 \right) - \lambda[\partial h(u) - \xi_k]$.
- ξ_{k+1} : $\xi_{k+1} = \xi_k + \frac{1}{\lambda} K^* \left(\frac{f}{Ku_{k+1}} - 1 \right)$.

Now letting $\xi_k = \frac{K^* v_k}{\lambda}$, we can have

$$v_{k+1} = v_k + \left(\frac{f}{Ku_{k+1}} - 1 \right).$$

With $v_0 = 0$, since $c\xi_0 = 0 = K^*0 = K^*v_0$, we may conclude inductively that $\lambda\xi_k \in R(K^*)$,

and hence there exists $v_k \in Y^* = L^\infty(\Omega)$ such that $\lambda \xi_k = \frac{K^* v_k}{\lambda}$. Hence, we have the alternative numerical algorithm:

II. Let $u_0 > 0$ with $Ku_0 \in L^\infty(\Omega)$ and $\text{essinf}(Ku_0) > 0$, $v_0 = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$:

- $u = u_k$: $\frac{\partial u}{\partial t} = K^* \left(\left[\frac{f}{Ku} + v_{k-1} \right] - 1 \right) - \lambda \partial h(u)$.
- v_{k+1} : $v_{k+1} = v_k + \left(\frac{f}{Ku_{k+1}} - 1 \right)$.

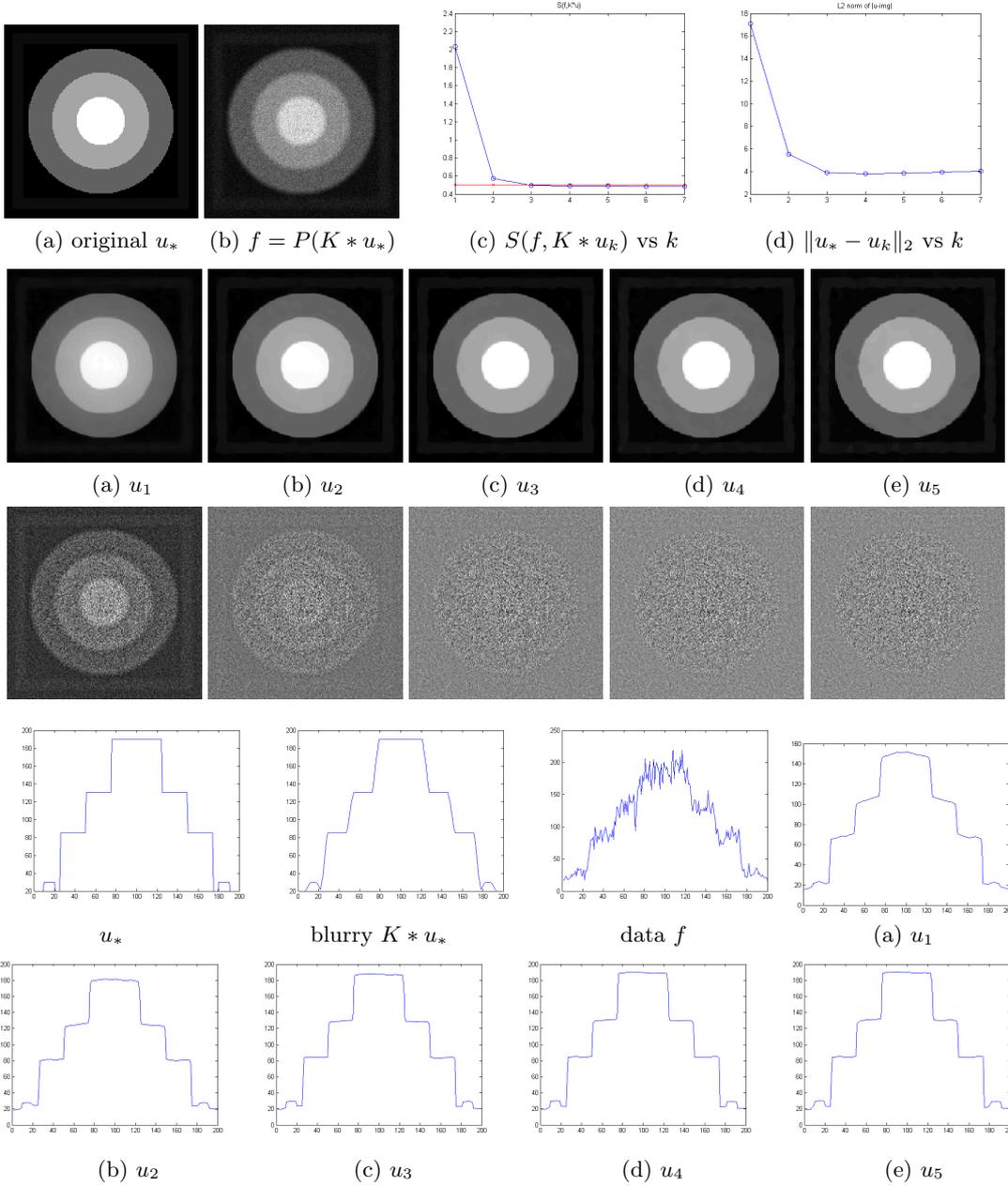
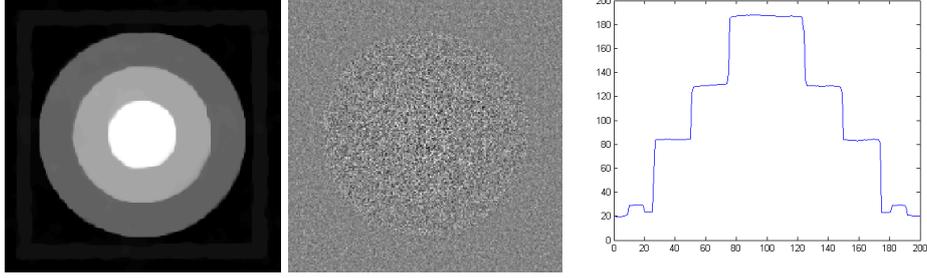
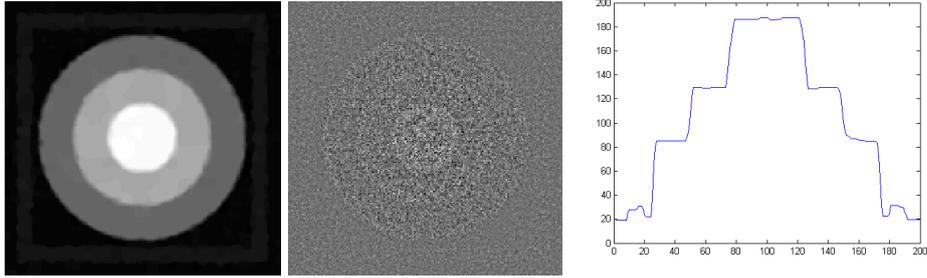


Figure 8: Poisson noise model. Data: Gaussian blur kernel K with $\sigma_b = 3$. Parameters: $\lambda = 0.005$. $S(f, K * u_3) = 0.4974 < \delta = S(f, K * u_*) = 0.5020$. u_3 is the best recovered image (RMSE= 3.8835)



Recovered u_3 using our iterative algorithm: RMSE=3.8835



Recovered u using the one-step model: RMSE=5.2019

Figure 9: Comparison with the one-step model (25) proposed by Le et al. Recovered images u (1st column), corresponding residuals $f - K * u$ (2nd), and recovered signals (3rd): $S(f, K * u) = 0.4948$, RMSE=5.2019.

4.2 Image restoration using cartoon+texture decomposition

Let $X = X_1 \times X_2$ with $X_1 \subset Y$, $X_2 \subset Y$, where X_1 and X_2 are Banach spaces as well as the duals of two separable Banach spaces and Y is a Banach space. Thus, X is also a Banach space as well as the dual of a separable Banach space, according to [42, p. 259]. We consider the standard linear degradation model

$$f = A\tilde{u} + n$$

where $f \in Y$ is the observed data, $A : X \rightarrow Y$ is a linear, compact operator defined by $A(u, v) = Ku + Kv$ for a convolution operator $K : Y \rightarrow Y$ ($Ku := K * u$). Here, we want to recover a sharp image \tilde{u} , and moreover we decompose \tilde{u} into the cartoon and texture parts, which will be denoted by $u \in X_1$ and $v \in X_2$.

Hence, we consider the minimization of the convex function

$$g(u, v) = S(f, Ku, Kv) = \int_{\Omega} (f - K * (u + v))^2 dx$$

with two variables u and v .

Based on the functional (26) from [20], we assume that $X_1 = BV(\Omega)$, $X_2 = H^{-1}(\Omega)$ (more precisely $X_2 = H^{-1}(\Omega) \cap L^2(\Omega)$), $Y = L^2(\Omega)$ and

$$h(u, v) = \frac{1}{2} \|(u, v)\|^2 = \lambda h_1(u) + \mu h_2(v),$$

where λ and μ are positive parameters, and

$$\begin{aligned} h_1(u) &= \frac{1}{2} \|u\|_{BV}^2 = \frac{1}{2} \left(\int_{\Omega} |u| dx + \int_{\Omega} |Du| \right)^2, \\ h_2(v) &= \frac{1}{2} \|v\|_{H^{-1}}^2 = \frac{1}{2} \left(\int_{\Omega} |\Delta^{-1}v|^2 + |\nabla \Delta^{-1}v|^2 dx \right). \end{aligned}$$

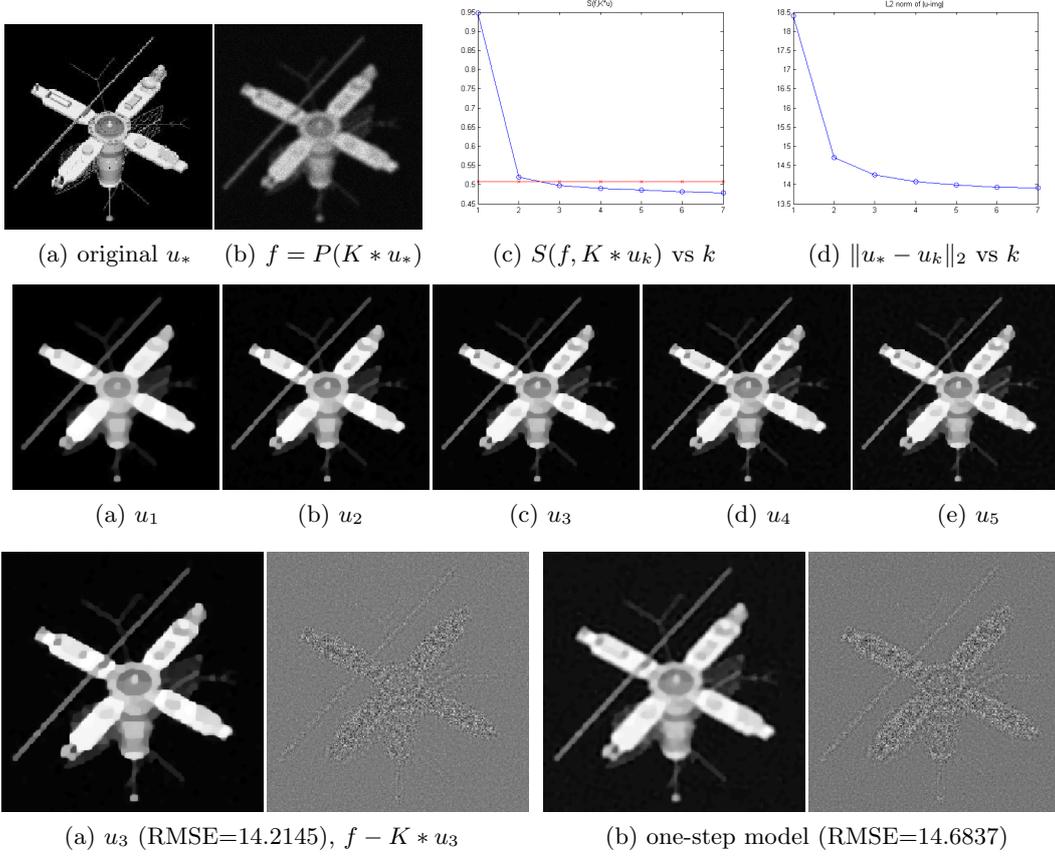


Figure 10: Poisson noise model. Data: Gaussian blur K with $\sigma_b = 3$. Parameters: $\lambda = 0.002$. $S(f, K * u_3) = 0.4978 < \delta = S(f, K * u_*) = 0.5082$. u_3 is the best recovered image (RMSE=14.2145). Bottom row: (a) recovered image u_3 using our iterative method, (b) recovered image u using one-step model (25), $S(f, K * u) = 0.4966$, RMSE=14.6837.

Furthermore, we can define the inexact Bregman distance with respect to h and $\xi_k \in \partial_{\varepsilon_k} h(u_k, v_k)$ as before:

$$\begin{aligned}
 D^k(u, u_k, v, v_k) &= h(u, v) - h(u_k, v_k) - \langle \xi_k, (u, v) - (u_k, v_k) \rangle + \varepsilon_k \\
 &= \lambda[h_1(u) - h_1(u_k) - \langle \xi_k^u, u - u_k \rangle + \varepsilon_{k,1}] \\
 &\quad + \mu[h_2(v) - h_2(v_k) - \langle \xi_k^v, v - v_k \rangle + \varepsilon_{k,2}],
 \end{aligned}$$

for some $\varepsilon_{k,1}, \varepsilon_{k,2} > 0$ such that $\lambda\varepsilon_{k,1} + \mu\varepsilon_{k,2} = \varepsilon_k$.

The last equality in the chain holds because

$$\begin{aligned}
 J_{\varepsilon_k}(u_k, v_k) &= \cup_{\varepsilon_{k,1} + \varepsilon_{k,2} = \varepsilon_k} \partial_{\varepsilon_{k,1}}(\lambda h_1)(u_k) \times \partial_{\varepsilon_{k,2}}(\mu h_2)(v_k) \\
 &= \cup_{\varepsilon_{k,1} + \varepsilon_{k,2} = \varepsilon_k} \lambda \partial_{\varepsilon_{k,1}/\lambda} h_1(u_k) \times \mu \partial_{\varepsilon_{k,2}/\mu} h_2(v_k) \\
 &= \cup_{\varepsilon_{k,1} + \varepsilon_{k,2} = \varepsilon_k} \lambda J_{\varepsilon_{k,1}/\lambda}(u_k) \times \mu J_{\varepsilon_{k,2}/\mu}(v_k)
 \end{aligned}$$

and thus,

$$\xi_k \in J_{\varepsilon_k}(u_k, v_k) \leftrightarrow \xi_k = (\lambda \xi_k^u, \mu \xi_k^v),$$

with $\xi_k^u \in J_{\varepsilon_{k,1}}(u_k)$ and $\xi_k^v \in J_{\varepsilon_{k,2}}(v_k)$, $\lambda\varepsilon_{k,1} + \mu\varepsilon_{k,2} = \varepsilon_k$.

Then, given $(u_k, v_k, \xi_k^u, \xi_k^v)$, the $\bar{\varepsilon}_k = \varepsilon_{k+1} = \lambda\varepsilon_{k+1,1} + \mu\varepsilon_{k+1,2} - \text{minimizer}(u_{k+1}, v_{k+1})$ can

be obtained by minimizing the function below

$$\begin{aligned}
g_k(u, v) &= S(f, Ku, Kv) + D^k(u, u_k, v, v_k) \\
&= S(f, Ku, Kv) + \lambda[h_2(u) - h_2(u_k) - \langle \xi_k^u, u - u_k \rangle + \varepsilon_{k,1}] \\
&\quad + \mu[h_2(v) - h_2(v_k) - \langle \xi_k^v, v - v_k \rangle + \varepsilon_{k,2}]
\end{aligned} \tag{30}$$

with $\xi_k^u \in \partial_{\varepsilon_{k,1}} h_1(u_k)$, $\xi_k^v \in \partial_{\varepsilon_{k,2}} h_2(v_k)$.

From (30), $\bar{\varepsilon}_k$ -*minimizer* (u_{k+1}, v_{k+1}) satisfies

$$\begin{aligned}
0 &= \eta_{k+1}^u + \lambda[\xi_{k+1}^u - \xi_k^u], & \zeta_{k+1}^u &\in \partial_u S(f, Ku_{k+1}, Kv_{k+1}), & \xi_{k+1}^u &\in \partial_{\varepsilon_{k+1,1}} h_1(u_{k+1}), \\
0 &= \eta_{k+1}^v + \mu[\xi_{k+1}^v - \xi_k^v], & \eta_{k+1}^v &\in \partial_v S(f, Ku_{k+1}, Kv_{k+1}), & \xi_{k+1}^v &\in \partial_{\varepsilon_{k+1,2}} h_2(v_{k+1}),
\end{aligned}$$

which can be rewritten as

$$\xi_{k+1}^u = \xi_k^u - \frac{1}{\lambda} \eta_{k+1}^u, \quad \xi_{k+1}^v = \xi_k^v - \frac{1}{\mu} \eta_{k+1}^v.$$

Algorithm 4.2. Let $u_0 = v_0 = 0$, $\xi_0^u = \xi_0^v = 0$, $\varepsilon_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$:

- Given $(u_k, v_k, \xi_k^u, \xi_k^v)$, define $\bar{\varepsilon}_k = \lambda\varepsilon_{k+1,1} + \mu\varepsilon_{k+1,2}$, and compute (u_{k+1}, v_{k+1}) as an $\bar{\varepsilon}_k$ -*minimizer* of the function below

$$\begin{aligned}
g_k(u, v) &= S(f, Ku, Kv) + D^k(u, u_k, v, v_k) \\
&= S(f, Ku, Kv) + \lambda[h_1(u) - h_1(u_k) - \langle \xi_k^u, u - u_k \rangle + \varepsilon_{k,1}] \\
&\quad + \mu[h_2(v) - h_2(v_k) - \langle \xi_k^v, v - v_k \rangle + \varepsilon_{k,2}].
\end{aligned}$$

- Update

$$\xi_{k+1}^u = \xi_k^u - \frac{1}{\lambda} \eta_{k+1}^u, \quad \xi_{k+1}^v = \xi_k^v - \frac{1}{\mu} \eta_{k+1}^v.$$

Numerical algorithm For $v \in H^{-1}(\Omega) \cap L^2(\Omega)$, we can define a new variable $p = \Delta^{-1}v \in \{p \in H^1(\Omega) : \int_{\Omega} p = 0\}$ due to the fact that for a bounded domain Ω , $\Delta : \{p \in H^1(\Omega) : \int_{\Omega} p = 0\} \rightarrow \{v \in H^{-1}(\Omega) : \int_{\Omega} v = 0\}$ is an isomorphism (one to one and onto) with Neumann boundary condition. Since $S(f, Ku, Kv) = S(f, Ku, K(\Delta p))$ and $h_2(v) = \frac{1}{2} \|v\|_{H^{-1}}^2 = \frac{1}{2} (\int_{\Omega} |\Delta^{-1}v|^2 + |\nabla \Delta^{-1}v|^2 dx) = \frac{1}{2} \int_{\Omega} p^2 + |\nabla p|^2 dx = h_2'(p)$, then we have

$$\begin{aligned}
\xi_{k+1}^p &= \xi_k^p - \frac{1}{\mu} \partial_p S(f, Ku_{k+1}, K(\Delta p_{k+1})) \\
&= \xi_k^p - \frac{1}{\mu} \Delta^* K^* (K(u_{k+1} + \Delta p_{k+1}) - f).
\end{aligned}$$

Moreover, since we have $\partial_u S(f, Ku, K(\Delta p)) = K^*(K(u + \Delta p) - f)$, similarly we can let

$$\xi_{k+1}^u = \frac{K^* w_{k+1}}{\lambda}, \quad \xi_{k+1}^p = \frac{\Delta^* K^* z_{k+1}}{\mu},$$

which leads to

$$w_{k+1} = w_k - (K(u_{k+1} + \Delta p_{k+1}) - f), \quad z_{k+1} = z_k - (K(u_{k+1} + \Delta p_{k+1}) - f).$$

With $w_0 = z_0 = 0$, since $\lambda \xi_0^u = 0 = K^* 0 = K^* w_0$ and $\mu \xi_0^p = 0 = \Delta^* K^* 0 = \Delta^* K^* z_0$, we may conclude inductively that $\lambda \xi_k^u \in R(K^*)$ and $\mu \xi_k^p \in R(\Delta^* K^*)$, and hence there exist $w_k, z_k \in Y^* = L^2(\Omega)$ such that $\lambda \xi_k^u = K^* w_k$ and $\mu \xi_k^p = \Delta^* K^* z_k$. Hence, we have the following numerical algorithm.

Let $u_0 = v_0 = 0$, $w_0 = 0$ and iterate for $k \in \mathbb{Z}$, $k \geq 0$:

Table 3: RESTORATION VIA DECOMPOSITION: ITERATIVE VS ONE-STEP METHOD

Image	Method	λ	Time (or iteration #)	RMSE	SNR
Group	Iterative	0.001	307 sec (k=12)	3.3347	20.2292
	One-step		1976 sec (iter=15000)	3.8023	19.0818
	Iterative	0.0005	519 sec (k=12)	3.0358	21.0584
	One-step		1979 sec (iter=15000)	3.5909	19.5764
Barbara	Iterative	0.001	744 sec (k=12)	5.7470	19.8125
	One-step		2109 sec (iter=15000)	6.6679	18.3643
	Iterative	0.0002	832 sec (k=6)	5.7408	19.8324
	One-step		2157 sec (iter=15000)	6.4039	18.7123

- Given (u_k, p_k, w_k) , $\bar{\varepsilon}_k - \text{minimizer}$ (u_{k+1}, p_{k+1}) is obtained by evolving the following equations:

$$\begin{aligned}\frac{\partial u}{\partial t} &= K^*(f + w_k - K(u + \Delta p)) - \lambda \partial h_1(u) \\ \frac{\partial p}{\partial t} &= \Delta K^*(f + w_k - K(u + \Delta p)) - \mu \partial h_2(p)\end{aligned}$$

- Update

$$w_{k+1} = w_k + (f - K(u_{k+1} + \Delta p_{k+1}))$$

where $\partial h_2(p) = p - \Delta p$.

5 Numerical results

We assume that $|\Omega| = 1$ and the function $r(\delta) = S(f, Ku_*)$ with true (or exact) image u_* is known. However, the estimation for the noise level $r(\delta)$ is possible from a data f , which is briefly mentioned in [27] for the Gaussian noise model. We could also estimate $r(\delta)$ by restricting the image to a square region which is uniform and contains no edges, taking the mean value (M) of the region (assuming $Ku_* = M$), and computing the fidelity term $r(\delta) = S(f, Ku_*)$ for each noise model.

First, we mention that, based on the property $\|\partial g_k(u)\| \leq \bar{\varepsilon}_k$ for an $\bar{\varepsilon}_k - \text{minimizer}$ u , with a fixed $\bar{\varepsilon}_k > 0$ and λ , we obtain three different $\bar{\varepsilon}_k - \text{minimizers}$ u_{k+1} by solving one of the following three equations:

$$\frac{\partial u}{\partial t} = \partial g_k(u) \pm \bar{\varepsilon}_k, \quad \text{or} \quad \frac{\partial u}{\partial t} = \partial g_k(u).$$

We compare the fidelity and error values of each $\bar{\varepsilon}_k - \text{minimizer}$, and compute $g_k(u_{k+1})$ and $g_k(u_k) + \bar{\varepsilon}_k$ in order to show that $g_k(u_{k+1}) \leq g_k(u_k) + \bar{\varepsilon}_k$. All these $\bar{\varepsilon}_k - \text{minimizers}$ provide similar fidelity and error values, justifying our simple algorithm to obtain an $\bar{\varepsilon}_k - \text{minimizer}$ by solving the usual Euler-Lagrange equation: $0 = \partial g_k(u)$ (as shown in Table 4.1).

Now, we consider the results of deblurring in the presence of noise (as explained in Table 4.1). As k increases, the image u_k recovers more details and fine scales, and eventually gets noise back. Thus, in practice, the residual $g(u_k) = S(f, Ku_k)$ keeps decreasing, while $\|u_* - u_k\|_2$ (Root Mean Square Error or RMSE) has a minimum value at some k' . But, note that k' does not correspond to the optimal $k_* = \min\{k : g(u_k) = S(f, Ku_k) \leq r(\delta)\}$ (i.e. usually $k' > k_*$), which is not surprising: in the presence of blur and noise, $u_{k'}$ can have lower RMSE since $u_{k'}$ might get sharper than u_{k_*} even though $u_{k'}$ gets noisier than u_{k_*} . However, the visual quality is also the best at the optimal k_* . For example, in Fig. 1 with Gaussian noise, u_3 ($k_* = 3$) recovers the details well enough leading to the best visual quality, while $\|u_* - u_k\|_2$ has a minimum at $k' = 4$ where u_k starts to become noisier. Thus the optimal k_* is a reasonable choice for the proposed noise models.

In Figures 1-4, we test the Gaussian noise model using L^2 fidelity term, and moreover we compare our result with the iterative algorithm (22) proposed by Osher et al. In Figures 1 and



original u_*

blurry data $f = K * u_*$



(a) $u_1 + v_1$

(b) cartoon u_1

(c) texture $v_1 = \Delta p_1$



(a) $u_2 + v_2$

(b) cartoon u_2

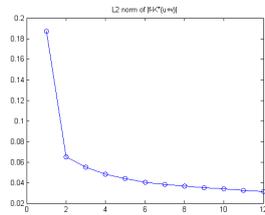
(c) texture $v_2 = \Delta p_2$



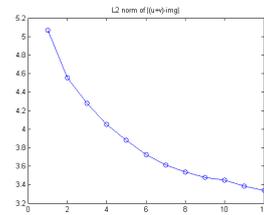
(a) $u_{12} + v_{12}$

(b) cartoon u_{12}

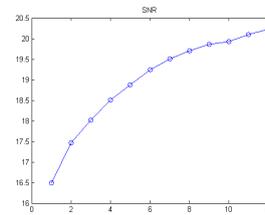
(c) texture $v_{12} = \Delta p_{12}$



$\|f - K * (u_k + v_k)\|_2$



$\|u_* - (u_k + v_k)\|_2$



$SNR(u_k + v_k)$

Figure 11: Restoration via decomposition using our iterative method. 2nd to 4th row: (a) recovered image $\tilde{u}_k = u_k + v_k$, (b) cartoon part u_k , (c) texture part v_k for $k = 1, 2, 12$. Parameters: $\lambda = 0.001$, $\mu = 0.00005$.

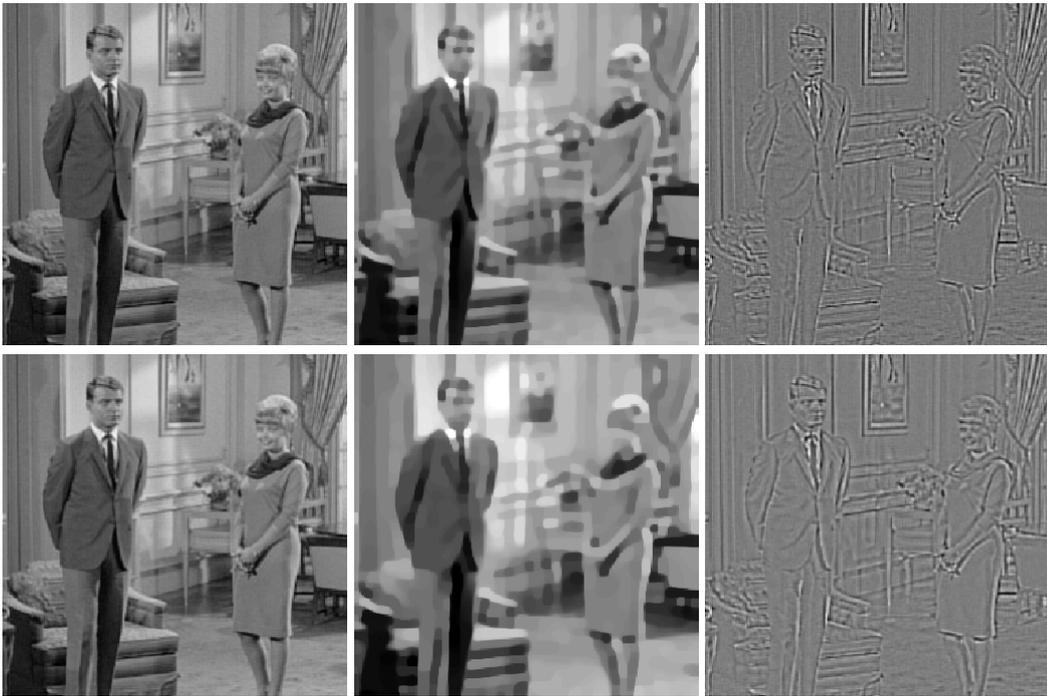


Figure 12: Comparison with one-step model. Top row: $\tilde{u}_{12} = u_{12} + v_{12}$, u_{12} , v_{12} using our iterative algorithm, RMSE=3.0358, SNR=21.0584, Time=519 sec. Bottom row: one-step model, RMSE=3.5909, SNR=19.5764, Time=1979 sec ($\|f - K(u+v)\|_2 = 0.1602$ (2000th) 0.1452 (5000th) 0.1393 (10000th) 0.1371 (15000th)). Parameters: $\lambda = 0.0005$, $\mu = 0.00005$.

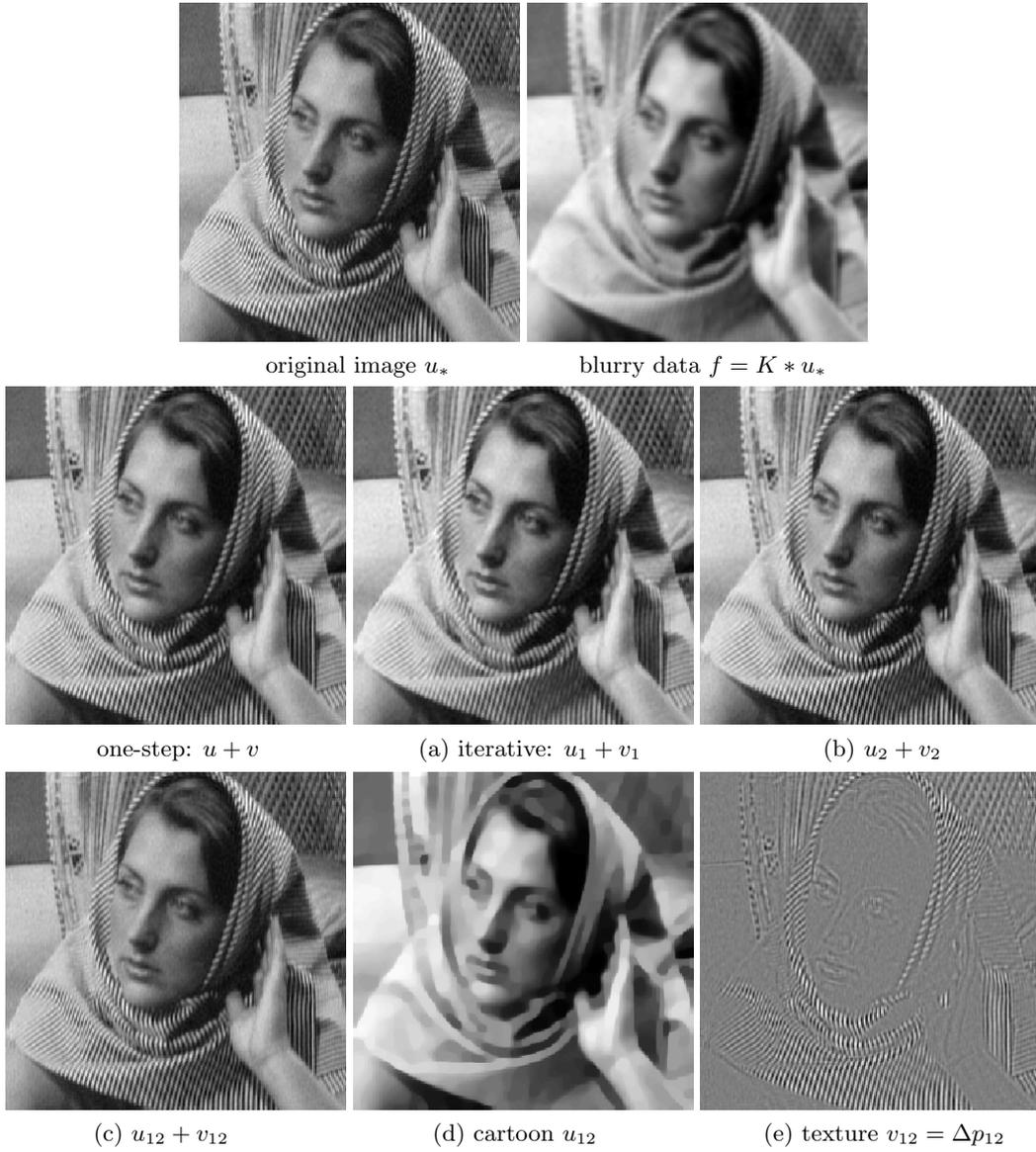


Figure 13: Restoration via decomposition using iterative algorithm. Data: Gaussian blur K with $\sigma_b = 1.5$. Second row: recovered image $\tilde{u} = u + v$ (RMSE=6.6679, SNR=18.3643, Time=2109 sec) using one-step proposed model, and (a)-(c) $\tilde{u}_k = u_k + v_k$ for $k = 1, 2, 12$ using iterative algorithm. (c) RMSE=5.7470, SNR=19.8125, Time=744 sec. Parameters: $\lambda = 0.001$, $\mu = 0.00005$.

4, we test our iterative algorithm. For both examples, u_3 recovers texture parts or details better than the previous iterates, and it is well denoised while the next iterate u_4 becomes noisier. In Figures 1 and 2, we observe that our iterative algorithm and Osher et al's model (22) provide similar numerical results (with similar best recovered images and similar behavior). Fig. 3 verifies the a-priori property for the stopping index (21); with smaller noise $\delta = 7.5$, the stopping index $k_* = 8$ is twice larger than the one ($k_* = 3$) with $\delta = 15$. Moreover, Fig. 3 shows that our iterative scheme provides superior result to the RO model [35] by recovering details or texture parts better.

In Figures 5-7, we show the recovered images u_k in the presence of Laplace noise with L^1 fidelity term, and we compare our results with one-step L^1 -TV deblurring-denoising model. In Figures 5 and 6, u_k restores fine scales and becomes sharper until the optimal $k_* = 3, 2$ respectively, and u_{k_*} gives cleaner (less noisier) images than u_k for $k > k_*$. In Fig. 7, we observe that our iterative method gives better visual quality images (cleaner and sharper images), and smaller RMSE than by the one-step L^1 -TV deblurring-denoising model.

For the Poisson noise model, in Figures 8-10, we obtain the same results as for the previous noise models: the best recovered images (u_3 in both Figures 8 and 10) provide sharper and cleaner images than any other iterates. In Figures 9 and 10, comparing our results with one-step model (25) proposed by Le et al, we also observe that the proposed iterative method gives much sharper and cleaner images.

In Figures 11-13, we apply the iterative algorithm to image deblurring via (cartoon + texture) decomposition with blurry data (as explained in Table 4.2). In this case, we do not have any stopping criteria either for the inner iteration to obtain (u_{k+1}, v_{k+1}) of $g_k(u, v)$ or for the outer iteration to choose the optimal $\tilde{u}_{k_*} = u_{k_*} + v_{k_*}$: the energy functional $g_k(u, v)$ keeps decreasing (see [20]). Thus, to obtain each (u_{k+1}, v_{k+1}) of $g_k(u, v)$, we stop at some (m -th) iterations (e.g. $m = 2000$ for (u_1, v_1) , $m = 1000$ for (u_2, v_2) , and so on). For the outer iteration, we plot the graphs of energies $\|f - K(u_k + v_k)\|_2$, $\|u - (u_k + v_k)\|_2$, $\text{SNR}(u_k + v_k)$ vs k , all of which keep decreasing or increasing as k increases (see Fig. 11). In addition, Fig. 11 shows the iterates $\tilde{u}_k = u_k + v_k$ with corresponding cartoon part u_k and texture part $v_k = \Delta p_k$ for $k = 1, 2, 12$. As k increases, especially the texture part $v_k = \Delta p_k$ gets sharper, resulting in sharper image \tilde{u}_k . In Figures 12 and 13, we compare our results with the one-step model (modified version of [20], by replacing the term $|u|_{BV(\Omega)}$ by $h_1(u)$ defined in our algorithm) using the same parameters λ, μ and time steps. Since the iterative algorithm adds residual ($v_k = f - K(u_k + \Delta p_k)$) to data f after some iterations, it enforces faster convergence of (u_k, v_k) than by the one-step model: indeed, in Table 4.2, with the same parameters, we see that the iterative method provides better restored images in much shorter time.

6 Conclusion

In this paper, we introduced a general iterative regularization method based on the square of the norm for image restoration models with general convex fidelity terms. We applied the proximal point method [16] using inexact Bregman distance to several ill-posed problems in image processing such as image deblurring in the presence of noise or image deblurring via (cartoon + texture) decomposition. The numerical experiments indicate that for deblurring in the presence of noise, the iterative procedure yields high quality reconstructions and superior results than by one-step gradient-descent models. For image deblurring via decomposition, the iterative algorithm enforces faster convergence of iterates \tilde{u}_k , thus it produces better restored images in a significantly shorter amount of time than by the one-step gradient descent model. Note that we have considered here the full norms in defining the regularization $h(u)$; since in most cases we work with quotient spaces (for example, $u \in BV(\Omega)$ such that $\int_{\Omega} = \int_{\Omega} f$), we could have also considered the square of the semi-norm (which becomes a norm on the quotient space). Such simplification and modification would lead to even faster implementations.

References

- [1] H. Attouch, G. Buttazzo, and G. Michaille, *Variational analysis in Sobolev and BV spaces*, MPS-SIAM Series on Optimization, 2006.
- [2] V. Barbu and T. Precupanu, *Convexity and optimization in Banach spaces*, D. Reidel Publishing Company, 1985.
- [3] M. Benning and M. Burger, *Error Estimates for Variational Models with Non-Gaussian Noise*, UCLA CAM Report 09-40, 2009.
- [4] L. M. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. Phys. 7, 200-217, 1967.
- [5] R. S. Burachik and S. Scheimberg, *A proximal point algorithm for the variational inequality problem in Banach space*, SIAM J. Control Optim. 39, 1615-1632, 2001.
- [6] M. Burger, E. Resmerita and L. He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing 81(2-3): 109-135, 2007.
- [7] D. Butnariu and A.N. Iusem, *On a proximal point method for convex optimization in Banach spaces*, Numer. Funct. Anal. Optim. 18, 723-744, 1997.
- [8] T. F. Chan and S. Esedoglu, *Aspects of Total Variation Regularized L1 Function Approximation*, SIAM J. Appl. Math. 65:5, pp. 1817-1837, 2005.
- [9] T. Le, R. Chartrand, and T. J. Asakiz, *A variational approach to reconstructing images corrupted by Poisson noise*, JMIV, vol. 27(3), pp. 257-263, 2007.
- [10] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., 3:538-543, 1993.
- [11] F. H. Clarke, *Optimization and Nonsmooth Analysis, Second edition*, Philadelphia, PA: SIAM, 1990.
- [12] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, SIAM 1999.
- [13] J. B. Garnett, P. W. Jones, T. M. Le, and L. A. Vese, *Modeling oscillatory components with the homogeneous spaces $BMO^{-\alpha}$ and $W^{-\alpha,p}$* , UCLA CAM Report 07-21, 2007 (to appear in Pure and Applied Mathematics Quarterly).
- [14] J. P. Gossez, *Opérateurs monotones nonlineaires dans les espaces de Banach nonreflexifs*, J. Math. Anal. Appl. 34, 371-395, 1971.
- [15] L. He, S. Osher, and M. Burger, *Iterative total variation regularization with non-quadratic fidelity*, J. Math. Imag. Vision, 26:167-184, 2005.
- [16] A. N. Iusem and E. Resmerita, *A proximal point method in nonreflexive Banach spaces*, Set-Valued and Variational Analysis, Volume 18, Number 1, pp. 109-120, 2010.
- [17] A. N. Iusem and R. Gárciga Otero, *Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces*, Numer. Funct. Anal. Optim. 22, 609-640, 2001.
- [18] M. Jung, E. Resmerita and L.A. Vese, *An Iterative Method with General Convex Fidelity Term for Image Restoration*, Proceedings of the 11th European Conference on Computer Vision (ECCV 2010), September 2010.
- [19] G. Kassay, *The proximal point algorithm for reflexive Banach spaces*, Studia. Univ. Babeş-Bolyai, Mathematica 30, 9-17, 1985.

- [20] Y. Kim and L. A. Vese, *Image recovery using functions of bounded variation and Sobolev spaces of negative differentiability*, Inverse Problems and Imaging, vol 3, 43-68, 2009.
- [21] S. Lintner and F. Malgouyres, *Solving a variational image restoration model which involves L^∞ constraints*, Inverse Problems 20, pp. 815-831, 2004.
- [22] F. Malgouyres, *Minimizing the total variation under a general convex constraint for image restoration*, IEEE Transactions on Image Processing, 11 (2002), 14501456.
- [23] A. M. Marques and B. Svaiter, *On the surjectivity properties of perturbations of maximal monotone operators in non-reflexive Banach spaces*, To be published.
- [24] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française d'Informatique et de Recherche Opérationnelle, série rouge, tome 4, No. 3, pp. 154-158, 1970.
- [25] Y. Meyer, *Oscillatory patterns in image processing and nonlinear evolution equations*, University Lecture series, vol. 22, American Mathematical Society, Providence, 2001.
- [26] J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93, 273-299, 1965.
- [27] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation based image restoration*, Multiscale Modelling and Simulation 4, 460-489, 2005.
- [28] S. Osher, A. Solé, and L. Vese, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Model. and Simul., 1:349-370, 2003.
- [29] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. PAMI, 12(7):629-639, 1990.
- [30] C. Pöschl, *Regularization with a similarity measure*, PhD Thesis, University of Innsbruck, 2008.
- [31] E. Resmerita and R.S. Anderssen, *Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems*, Mathematical Methods in the Applied Sciences, 30(13), 1527-1544, 2007.
- [32] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. 14, 877-898, 1976.
- [33] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1997.
- [34] L. I. Rudin, S. J. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60, pp. 259-268, 1992.
- [35] L. I. Rudin and S. Osher, *Total variation based image restoration with free local constraints*, in ICIP (1), pp. 31-35, 1994.
- [36] O. Scherzer and C. Groetsch, *Inverse Scale Space Theory for Inverse Problems*, In M.Kerckhove, editor, Scale-Space and Morphology in Computer Vision, Lecture Notes in Comput. Sci. 2106, pp. 317-325. Springer, New York, 2001.
- [37] J. L. Starck, M. Elad, and D. L. Donoho, *Image decomposition via the combination of sparse representations and a variational approach*, IEEE Transactions On Image Processing, 14 (2005), 1570-1582.
- [38] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical (BV, L^2) decompositions*, Multiscale Model. Simul., 2, pp. 554-579, 2004.

- [39] E. Tadmor, S. Nezzar, and L. Vese, *Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation*, Commun. Math. Sci. Vol. 6, No. 2, pp. 281-307, 2008.
- [40] A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math Dokl 4, 1035-1038 English translation of Dokl Akad Nauk SSSR, 151 (1963), 501-504.
- [41] A. N. Tikhonov and V. A. Arsenin, *Solution of Ill-posed Problems*, Winston and Sons, Washington, 1977.
- [42] F. Trèves, *Topological Vector Spaces Distributions, and Kernels*, Pure and Applied Mathematics 25, Academic Press, New York, 1967.
- [43] L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Applied Mathematics and Optimization, 44 (2001), 131-161.
- [44] L. Vese and S. Osher, *Modeling textures with total variation minimization and oscillatory patterns in image processing*, J. Sci. Comput., 19:553-572, 2003.
- [45] W. Yin, D. Goldfarb, and S. Osher, *Image cartoon-texture decomposition and feature selection using the total variation regularized L^1 functional*, Lecture notes in Computer Science, 3752:73-84, 2005.
- [46] W. Yin, S. Osher, D. Goldfarb, J. Darbon, *Bregman iterative algorithm for l_1 minimization with applications to compressed sensing*, SIAM J. in Imaging Sciences 1(1): 143-168, 2008.
- [47] C. Zalinescu, Personal communication, 2010.