# ITERATIVE TOTAL VARIATION REGULARIZATION WITH NON-QUADRATIC FIDELITY

LIN HE *, MARTIN BURGER †, AND STANLEY J. OSHER ‡

**Abstract.** A generalized iterative regularization procedure based on the total variation penalization is introduced for image denoising models with non-quadratic convex fidelity terms. By using a suitable sequence of penalty parameters we solve the issue of solvability of minimization problems arising in each step of the iterative procedure, which has been encountered in a recently developed iterative total variation procedure Furthermore, we obtain rigorous convergence results for exact and noisy data.

We test the behaviour of the algorithm on real images in several numerical experiments using $L^1$ and $L^2$ fitting terms. Moreover, we compare the results with other state-of-the art multiscale techniques for total variation based image restoration.

**Key words.** Iterative Regularization, Total Variation Methods, Image denoising, Multiscale methods, Bregman distances.

**1. Introduction.** The aim of this paper is to generalize the iterative regularization procedure (cf. [10]) for variational models with convex fidelity functionals based on the use of the Bregman distance (cf. [2, 5]). In particular we focus on variational models arising in image recovery. Among different ways to recover a distorted image, one of the best known and most influential methods is the total variation based model of Rudin, Osher, and Fatemi(ROF) (cf.[12]). The idea behind the model is to exhibit the reconstructed image as the minimizer of an energy functional:

$$u = \arg \min_{u \in BV(\Omega)} \left\{ |u|_{BV} + \lambda ||u - f||^2_{L^2} \right\}. \tag{1.1}$$

for a suitable parameter $\lambda > 0$. Here $\Omega$ is a domain in $R^N$ with Lipschitz boundary modeling the image region, e.g. a computer screen.

The function $f$ represents the observed and possibly noisy image, which an element of $L^2(\Omega)$. The regularization functional is the $BV$-seminorm, defined via

$$|u|_{BV} = \sup_{|g|_\infty <= 1, g \in C_c^1(\Omega)^2} \int_\Omega u(\nabla \cdot g) dx., \tag{1.2}$$

where $|g| = \sqrt{g_1^2 + g_2^2}$ and $C_c^1(\Omega)$ denotes the class of continuously differentiable functions of compact support in $\Omega$. The key feature of total variation regularization is the fact that it allows for (and even favours) discontinuous solutions, i.e., images with sharp edges. Nevertheless, this regularization suppresses oscillations and can still eliminate high-frequency noise.

However, the ROF model (1.1) has certain limitations. Meyer has performed a general analysis for standard ROF in [9]. Defining the (dual) space $G$ as the distributional closure of the set

$$\left\{ w = \partial_x g_1 + \partial_y g_2 = \nabla \cdot g \mid g \in C_c^1(\Omega)^2 \right\},$$

*UCLA Mathematics Department, Box 951555, Los Angeles, CA 90095-1555, USA. email: helin@math.ucla.edu

†Institut für Industriemathematik, Johannes Kepler Universität, Altenbergerstr. 69, A 4040 Linz, Austria. email: martin.burger@jku.at

‡UCLA Mathematics Department, Box 951555, Los Angeles, CA 90095-1555, USA. email: sjo@math.ucla.edu

equipped with the norm

$$||w||_* = \inf_{g, w=\nabla\cdot g} \left( \operatorname{ess\,sup}_x |g(x)| \right),$$

Meyer provided arguments in favour of considering elements of this dual space $G$ can be regarded as textures. He also showed that for the decompostion $f = u + v$, with $u$ being the minimizer of (1.1),

$$||f||_* < \tfrac{1}{2\lambda} \quad \Rightarrow u = 0, v = f;$$

$$||f||_* \geq \tfrac{1}{2\lambda} \quad \Rightarrow ||v||_* = \tfrac{1}{2\lambda}, \textstyle\int uv = \tfrac{1}{2\lambda}|u|_{BV}.$$

The term $v$ that represents noise is usually ignored; however, it often contains textures. In order to preserve textures, Meyer suggested the modified variational problem

$$u = \arg \min_{u \in BV(\Omega)} \left\{ |u|_{BV} + \lambda ||f - u||_* \right\}, \tag{1.3}$$

whose computational solution is a rather difficult task because of the nature of the norm $||\cdot||_*$ (cf. [1, 6] for a detailed discussion).

We also mention here another version of the ROF model that has been studied recently by Chan and Esedoglu [4] (see also the references therein), who used the $L^1$ norm instead the square of $L^2$ norm as in ROF, as a measure of fidelity between the observed and denoised images. Given an observed image $f \in L^1(\Omega)$, this model is based on the following variational problem:

$$u = \arg \min_{u \in BV(\Omega)} \left\{ |u|_{BV} + \lambda ||u - f||_{L^1} \right\}. \tag{1.4}$$

Even though this minimization problem may have a lack of uniqueness and continuous dependence on data is not clear, it has many desirable and some unexpected consequences in applications such as reconstruction of binary images, multiscale image decompositon (cf. [13]), and data driven parameter selection.

Both these generalizations of ROF with respect to the fidelity functional can yield an improvement with respect to some aspects, but they share the systematic error yielding a decrease in the total variation and therefore a loss of information in any case. In order to overcome this issue an iterative version of Tikhonov regularization has been introduced in [10], the Bregman distance (cf. [5]) is applied to iteratively refine a degraded image. The procedure of the iterative scheme is to start with $u_0 = 0$ and then to obtain an iterative improved reconstructed image $u_n$ by the following minimization

$$u_k = \arg \min_{u \in BV} \left\{ H(u, f) + \lambda^{-1} D_J(u, u_{k-1}) \right\}, \tag{1.5}$$

where $H(u, f) = \frac{1}{2}\|u - f\|^2_{L^2}$, $J(u) = |u|_{BV}$ and

$$D_J(x, y) = J(x) - J(y) - \langle x - y, \partial J(y) \rangle, \tag{1.6}$$

is the so-called *Bregman distance* related to a differentiable functional $J : \mathbb{R}^n \to \mathbb{R}^n$. Here $\langle \cdot, \cdot \rangle$ denotes the inner product in $R^n$ and $\partial J(y)$ is the gradient of $J$ at the point $y$. If $J$ is non-differentiable but still convex (like the total variation functional) an

analogous algorithm can still be used, but now with the generalized Bregman distance (cf .[7])

$$D_J(x, y) = J(x) - J(y) - \langle x - y, p\rangle, \qquad p \in \partial J(y), \tag{1.7}$$

now with $\partial J$ denoting the subgradient.

Using (1.5) it has been shown that the sequence $\{u_k\}$ converges monotonically in $L^2$ to the noisy image $f$. More importantly, as $k$ increases, for $k \leq \bar{k}$ and sufficiently small $\lambda$, the Bregman distance between $u_k$ and the true noise free image $\tilde{u}$ is decreasing. Here $\bar{k}$ is defined as

$$\max\{\, k \in \mathbb{N} | H(u_k, f) > \tau H(\tilde{u}, f)\,\} \tag{1.8}$$

where $\tau > 1$. We refer to [10] for more details of the analysis. Comparing with the denoised image from the ROF model, we see that this iterative algorithm improves the results significantly. Moreover, this algorithm is extremely easy to implement numerically. We just add back the noise plus texture term $v$ at every iteration (see [10]). Thus, recently attention has been paid to applying the iterative regularization procedure based on the Bregman distance for different fidelity terms or regularization terms. In [8], He, Marquina and Osher have applied the iterative idea to the blind deconvolution problem to recover finer scales. In that paper, the fidelity term was taken to $H(u, f) = ||K * u - f||^2_{L^2}$, where $K$ is a convolution operator.

A key observation in the analysis of the iterative scheme is a rewritten version as the generalization of *proximal point* or *Bregman iterations* (cf. e.g. [2, 5]). These methods are usually used to solve problems of the form

$$\min\{g(u) : u \in \mathbb{R}^n\},$$

where $g : \mathbb{R}^n \longmapsto (-\infty, \infty]$ is a proper, lower semi-continuous convex function. Proximal point methods generate a sequence $\{u_k\}$ from the minimization problems

$$u_k = \arg\min_{u\in\mathbb{R}^n} \left\{ g(u) + \lambda_k^{-1} D_J(u, u_{k-1}) \right\}. \tag{1.9}$$

In general, the convergence analysis of proximal point algorithms has two main requirements:

- $\lim\ \inf_{k\to\infty}\{\lambda_k :\geq 0\} > 0$;
- $f$ is bounded below, and the iterative scheme generates a sequence $\{u_k\}$ such that $u_k \in \mathbb{R}^n$ for all $k$.

The second assumption means that a solution of the minimization problem (1.9) exists, which is however not straight-forward to show for many important cases. We shall introduce our a modified model with suitable decay of the penalty parameters which allows to prove the existence of the minimizers in Section 2. Furthermore, following the theoretical proof in [10], monotonicity and convergence theorems are also obtained in the same section. In Section 3, some numerical results are given using various fidelity functionals $H$ while the regularization term is kept as $BV$ term. Comparisons with other methods of similar spirit are also presented.

**2. Analysis of a Modified Iterative Regularization Procedure.** In general it is quite easy to generalize the iterative regularization algorithm to variational models with a different regularization functional $J$, as long as this functional has suitable lower-semicontinuity and compactness properties in some topology (cf. [10]). The analysis of the algorithm (with respect to well-definedness, convergence, regularization

properties) can be transfered in a one-to-one way, only computational schemes have to be adapted to the specific form of the functional.

By far more challenging is the generalization of the procedure with respect to the fidelity term $H$. Of course, one can just write down (1.5) and try to solve the minimization problems in each step. However, the well-definedness of this procedure is not clear if $H$ is not quadratic, since only in the quadratic each minimization problem in (1.5) can be rewritten in the same form as (1.1) and corresponding existence and uniqueness results for minimizers can be carried over.

**2.1. Generalization of the Fidelity Term: A Modified Model.** Inspired by Scherzer and Groetsch (cf.[14]) and Tadmor, Nezzar and Vese (cf.[15]), who multiplied the parameter $\lambda$ by two after each iteration step, we formulate a new iterative total variation regularization as

$$u_k \in \arg\min_{u\in BV} E(u, u_{k-1}, f) := \arg\min_{u\in BV} \left\{ H(u, f) + \frac{1}{2^{k-1}\lambda} D_J(u, u_{k-1}) \right\} \qquad (2.1)$$

where $J(u)$ satisfies the conditions assumed above, and $H(u, f) := h(u - f)$ with $h$ being a nonnegative, convex, and positively homogeneous functional, which is continuous with respect to weak-* convergence in $BV$. Moreover, we assume that $h(c)$ does not vanish for constant functions $c \neq 0$, so that $h(u) + |u|_{BV}$ is indeed an equivalent norm on $BV(\Omega)$. For convenience we set $u_{-1} = 0$ and $p_{-1} = 0$ so that $u_0$ is defined as the minimizer of $H(u, f) + 2\frac{J(u)}{\lambda}$. Using this new construction, we are able to overcome the problem of lower boundedness and thus obtain well-definedness in Section 2.2 and prove some monotonicity and convergence theorems in Section 2.3.

The models in [15, 14], which inspire our setting, did not use a proximal point iteration but instead used Tikhonov-Morozov iteration for which the following convergence analysis does not apply. The difference is that one replaces $D_J(u, u_{k-1})$ in (2.1) by $J(u - u_{k-1})$. We call this Tikhonov-Morozov iteration with the multiscale coefficient alone a T-N-V method even if $H(u, f)$ is not the $L^2$ square fidelity term as used in the original papers.

We mention that the multiscale decomposition of T-N-V concerned the function $u$ directly, in [15] the method was even set up as a sequence of minimization problems for the update $v_k = u_k - u_{k-1}$. For choosing $\lambda$ constant, one would obtain $v_k \equiv 0$ for $k \geq 1$ and hence, $\lambda$ has to be decreased in order to access small scale features that have been eliminated in the previous steps. Our model (2.1) can rather be considered as a multiscale decomposition for the dual variable $p_k$. Note that the optimality condition for the variational problem in (2.1) reads

$$p_k - p_{k-1} \in -2^{k-1}\lambda \partial_u H(u_k, f),$$

and hence yields the update for $k$. In this case the dyadic choice is not as obvious. For an arbitrary variable choice $\lambda_k$ we obtain (after summation of the optimality conditions and using the notation $q_j \in \partial_u H(u_j, f)$)

$$p_k = p_0 + \sum_{j=1}^{k-1} \lambda_j q_j.$$

In the quadratic case $H(u_j - f) = \|u_j - f\|^2$, the subgradient is single valued and given by $q_j = 2(u_j - f)$, and in particular it scales with the residual. Since the residual tends to zero, the sum $\sum_{j=1}^{k-1} \lambda_j q_j$ can also be bounded if $\lambda$ is constant (and indeed is

as can be deduced from the results in [10]). In the case of positively one-homogeneous fitting functionals such the ones in (1.3) and (1.4), the norm of the subgradients of $H$ will be of order one in the dual space to the fitting norm. E.g. for the $L^1$-functional, we obtain that $\|q_j\|_\infty = 1$ for $q_j \in \partial_u H(u_j, f)$ (unless $u_j \equiv f$, one can think of $q_j$ as the sign of $u_j - f$ almost everywhere). Thus, the update in the dual variable will actually increase during the iteration in these spaces (except $u_k \equiv f$, which leads to no subsequent changes in the iteration). In the worst case one has to expect that the norm of $p_k$ is given by

$$\|p_k\|_\infty \sim \sum_{j=1}^{k} \lambda_j \|q_j\|_\infty = \sum_{j=1}^{k} \lambda_j.$$

In the iteration step from $k$ to $k+1$, the term $\frac{1}{\lambda_{k+1}} p_k$ appears in the Bregman distance and subsequently in the optimality condition. In order to obtain a reasonable scaling with the subgradient of $H$, also $p_k$ should be of order one and this is not true for constant $\lambda$ since from the above argument we would expect that

$$\frac{1}{\lambda_{k+1}} \|p_k\|_\infty \sim k.$$

On the other hand, the scaling to order one can be achieved directly by the dyadic choice $\lambda_k = 2^{k-1}\lambda$, for which we can expect

$$\frac{1}{\lambda_{k+1}} \|p_k\|_\infty \sim \frac{1}{2^k \lambda} \sum_{j=1}^{k} 2^{k-1}\lambda = 1.$$

We mention that the same type of scaling argument is possible for other positively one-homogeneous functionals, but . As we shall prove below, the dyadic choice leads to a well-defined model for rather general fidelity functionals, but from the scaling one might argue that it is not the optimal one if $H$ is not positively one-homogeneous (and one could find the optimal one from the scaling of subgradients). However, positively one-homogeneous fidelity terms are by far the most interesting examples in total variation based image restoration (except the previously analyzed quadratic ones), so that we focus on the dyadic choice in this paper. Possible extensions for other functionals can then be carried out along the lines of our analysis.

**2.2. Well-Definedness of the New Iteration Model.** We now derive the existence of minimizers for the modified iterative method (2.1):

**Theorem 2.1.** *Under the above conditions, the iteration scheme (2.1) yields a well-defined sequence $u_k \in BV(\Omega)$.*

With $u_k$ defined via (2.1) we obtain the Euler-Lagrange equations

$$q_k + \frac{1}{2^{k-1}\lambda}(p_k - p_{k-1}) = 0, \qquad q_k \in \partial_u H(u_k, f),\ p_j \in \partial J(u_j), \tag{2.2}$$

Using (2.2), first we express $p_k \in \partial J(u_k)$ in terms of the subgradients of the

functionals $H(u_j, f)$, $0 \le j < k$, which are denoted by $q_j \in \partial_u H(u_j, f)$:

$$\begin{aligned}\frac{1}{2^{k-1}\lambda} p_k &= \frac{1}{2^{k-1}\lambda} p_{k-1} - q_k \\ &= \frac{1}{2^{k-1}\lambda} p_{k-2} - \frac{1}{2} q_{k-1} - q_k \\ &= \dots \\ &= \frac{1}{2^{k-1}\lambda} p_0 - \sum_{j=1}^{k} \frac{1}{2^{k-j}} q_j = -\sum_{j=0}^{k} \frac{1}{2^{k-j}} q_j. \end{aligned} \tag{2.3}$$

Next we define $v := f - u$, thus $H(u, f) = h(v)$. Due to the convexity, nonnegativity, and homogeneity of the functional $h$, we have for any $w \in BV$, $q \in \partial h(w)$, and $t > 0$,

$$\begin{aligned} t\langle v, q\rangle &= \langle (tv + w) - w, q\rangle \le h(tv + w) - h(w) \\ &\le h(tv) = t h(v). \end{aligned}$$

Since $t > 0$, we obtain the inequality

$$\langle v, q\rangle \;\le\; h(v), \qquad \forall\; v, w, q \in \partial h(w). \tag{2.4}$$

In particular, if we choose $w = f - u_j$, then (2.4) becomes

$$\langle u - f, q_j\rangle \;\ge\; -H(u, f). \tag{2.5}$$

Using (2.3) and (2.5) and defining $\{w_n\}$ as a minimizing sequence to $u_k$, we know there exist $M > 0$ such that

$$\begin{aligned} M \ge\;& H(w_n, f) + \tfrac{1}{2^{k-1}\lambda} J(w_n) - \tfrac{1}{2^{k-1}\lambda} J(u_{k-1}) - \langle w_n - u_{k-1}, \tfrac{1}{2^{k-1}\lambda} p_{k-1}\rangle \\ =\;& H(w_n, f) + \tfrac{1}{2^{k-1}\lambda} J(w_n) - \tfrac{1}{2^{k-1}\lambda} J(u_{k-1}) + \langle u_{k-1} - f, \tfrac{1}{2^{k-1}\lambda} p_{k-1}\rangle \\ &+ \tfrac{1}{2}\langle w_n - f, \textstyle\sum_{j=1}^{k-1} \frac{1}{2^{k-1-j}} q_0\rangle \\ \ge\;& \tfrac{1}{2^{k-1}\lambda} J(w_n) - \tfrac{1}{2^{k-1}\lambda} J(u_{k-1}) + \langle u_{k-1} - f, \tfrac{1}{2^{k-1}\lambda} p_{k-1}\rangle \\ &+ H(w_n, f) - H(w_n, f) \textstyle\sum_{j=0}^{k-1} \frac{1}{2^{k-j}} \\ \ge\;& \tfrac{1}{2^{k-1}\lambda} J(w_n) - \tfrac{1}{2^{k-1}\lambda} J(u_{k-1}) + \langle u_{k-1} - f, \tfrac{1}{2^{k-1}\lambda} p_{k-1}\rangle. \end{aligned}$$

Since both $u_{k-1}$ and $f$ are fixed and bounded independent of $n$, the sequence $\{J(w_n)\}$ is bounded. From the compactness and lower semi-continuous properties of the functional $J$, we can conclude that the iteration model (2.1) is well-defined.

From the proof one observes that the choice $2^{k-1}$ is crucial but not the only choice for the penalization parameters, but it suffices to choose a sequence $\lambda_k$ such that $\sum_{j=1}^{\infty} \frac{1}{\lambda_j} \le 1$.

**2.3. Convergence Analysis.** We now study some convergence properties of the new iterative regularization process. Our analysis below basically follows the lines in [10]. In particular we shall see that all desirable properties like monotonicity of residual and convergence for the exact image and for the noisy image still hold for our new model.

**Proposition 2.2 (Monotonicity).** *Let $u_k \in BV(\Omega)$ be a sequence defined by (2.1). Then the sequence $H(u_k, f)$ is monotonically non-increasing and satisfies:*

$$H(u_k, f) \le H(u_k, f) + \frac{1}{2^{k-1}\lambda} D(u_k, u_{k-1}) \le H(u_{k-1}, f). \tag{2.6}$$

*Moreover, if $J(u) < \infty$, then we have:*

$$D(u, u_k) + D(u_k, u_{k-1}) + 2^{k-1}\lambda H(u_k, f) \leq 2^{k-1}\lambda H(u, f) + D(u, u_{k-1}). \tag{2.7}$$

Using the definition of subgradient and the fact that $u_k$ minimizes $E(u, u_{k-1}, f)$, we have

$$\begin{aligned} H(u_k, f) \leq & \quad H(u_k, f) + \tfrac{1}{2^{k-1}\lambda} D(u_k, u_{k-1}) = E(u_k, u_{k-1}, f) \\ \leq & \quad E(u_{k-1}, u_{k-1}, f) = H(u_{k-1}, f). \end{aligned}$$

To prove (2.7), we use the convexity of the function $H(u, f)$ and a standard decomposition of the Bregman distance (with the same notation for subgradients as above)

$$\begin{aligned} D(u, u_k) - D(u, u_{k-1}) + D(u_k, u_{k-1}) = & \quad J(u) - J(u_k) - \langle u - u_k, p_k \rangle \\ & -J(u) + J(u_{k-1}) + \langle u - u_{k-1}, p_{k-1} \rangle \\ & +J(u_k) - J(u_{k-1}) - \langle u_k - u_{k-1}, p_{k-1} \rangle \\ = & \quad \langle u - u_k, p_{k-1} - p_k \rangle = \langle u - u_k, 2^{k-1}\lambda q_k \rangle \\ \leq & \quad 2^{k-1}\lambda [H(u, f) - H(u_k, f)]. \square \end{aligned}$$

The above proposition states that the fidelity term $H(u_k, f)$ is decreasing. Furthermore, if we choose $u = f$ in (2.7) and use the fact that $H(f, f) = 0$, then we obtain for "exact data" satisfying $J(f) < \infty$ that

$$\begin{aligned} D(f, u_k) \leq & \quad D(f, u_k) + D(u_k, u_{k-1}) \\ \leq & \quad D(f, u_{k-1}) + 2^{k-1}\lambda (H(f, f) - H(u_k, f)) \\ \leq & \quad D(f, u_{k-1}). \end{aligned} \tag{2.8}$$

Hence, the Bregman distance between the solution $u_k$ of (2.1) at the $k$-th iteration step and the image $f$ is decreasing, too. In fact, as illustrated by computations in Section 4, both the Bregman distance $D(u_k, f)$ and the fidelity term $H(u_k, f)$ decrease as $k$ increases.

**Theorem 2.3 (Exact Data).** *Let $f$ satisfy $J(f) < \infty$ and let $u_k$ be a sequence generated by (2.1) with data $f$. Then*

$$H(u_k, f) \leq \frac{J(f)}{2^k \lambda} \tag{2.9}$$

*and in particular $\{u_k\}$ is a minimizing sequence of $H(\cdot, f)$.*

*Moreover, $u_k \rightarrow f$ in the weak-* topology of $BV(\Omega)$.*

Summing all inequalities (2.7) from 1 to $k$, we have:

$$\sum_{j=1}^{k} [D(u_j, u_{j-1}) + 2^{j-1}\lambda H(u_j, f)] \leq D(f, u_0) - D(f, u_k). \tag{2.10}$$

From the fact that $H(u, f)$ is a convex function, we also have the following inequality:

$$J(u_0) + \lambda H(u_0, f) \leq J(f) - D(f, u_0). \tag{2.11}$$

Now adding (2.11) to (2.10) and using both $D(u_j, u_{j-1}) \geq 0$ and the monotonicity property of $H(u_j, f)$, we conclude

$$2^k \lambda H(u_k, f) \leq J(f).$$

Based on (2.10) and (2.11), we also obtain

$$\begin{aligned}
J(f) \geq & \sum_{j=1}^{k} D(u_j, u_{j-1}) + J(u_0) \\
= & J(u_k) - \sum_{j=1}^{k} \langle p_{j-1}, u_j - u_{j-1} \rangle \\
= & J(u_k) - \langle p_{k-1}, u_k - f \rangle + \sum_{j=1}^{k-1} \langle p_j - p_{j-1}, u_j - f \rangle \\
= & J(u_k) + \lambda \langle H_u(u_0, f) + \sum_{j=1}^{k-1} 2^{j-1} H_u(u_j, f), u_k - f \rangle \\
& - \sum_{j=1}^{k-1} \langle 2^{j-1} \lambda H_u(u_j, f), u_j - f \rangle \\
\geq & J(u_k) - \lambda 2^{k-1} H(u_k, f) - \sum_{j=1}^{k-1} 2^{j-1} \lambda H(u_j, f) \\
\geq & J(u_k) - \tfrac{3}{2} J(f).
\end{aligned}$$

Therefore $J(u_k) \leq \frac{5}{2} J(f)$. Because the level sets of $\{u \in \mathcal{U} \mid J(u) \leq M\}$ are compact, the further assertions then follow by standard weak-* convergence techniques analogous to the arguments in [10]. □

The above result is important from a theoretical point of view since it verifies convergence of the method. In practice however, the given data do not represent the exact but rather a noisy version of the image, since otherwise one would not need to denoise it. Therefore we consider the case where $f$ contains noise in the following. It is well-known for iterative methods that a regularizing effect is obtained only via appropriate stopping in dependence of the noise level, which is given for the fidelity functional $H$ as

$$H(u_*, f) \leq \delta. \tag{2.12}$$

We again inspect the decrease of the distance between the iterations and the noisy image, which is now guaranteed only until the residual becomes to small:

**Proposition 2.4.** *Let $u_*$ be the true noise free image and let $f$ be a given noisy version satisfying* (2.12)*. Then, as long as $H(u_k, f) > \delta$, the Bregman distance between $u_k$ and the true solution $u_*$ decreases, i.e.,*

$$D(u_*, u_k) \leq D(u_*, u_k) + D(u_k, u_{k-1}) < D(u_*, u_{k-1}). \tag{2.13}$$

Plugging the noise free image $u_*$ into (2.7), we obtain

$$D(u_*, u_k) + D(u_k, u_{k-1}) + 2^{k-1} \lambda H(u_k, f) \leq D(u_*, u_{k-1}) + 2^{k-1} \lambda H(u_*, f)$$

With the assumption $H(u_k, f) > \delta$ and (2.12), we can conclude (2.13).□

From Proposition 2.4 we can deduce that the *generalized discrepancy principle* (cf. [11]) is a good candidate as a stopping rule for (2.1), i.e., the iteration is stopped at the index

$$k_* = \max\{k \in N | H(u_k, f) \geq \tau\delta\} \tag{2.14}$$

where $\tau > 1$. Important features of $k_*$ are studied. For given $\delta > 0$, we shall consider $u_{k_*}$ as the regularized solution, i.e., the result of our iterative scheme. With such a stopping criterion, we can obtain a so-called *semi-convergence property*, i.e., the regularized solutions converge to $u_*$ as $\delta \to 0$, more precisely:

**Theorem 2.5 (Noisy Data).** *The stopping index $k_*$ is well-defined by* (2.14) *for any $\delta > 0$, and $k_*(\delta) = \mathcal{O}(\log \delta)$.*

*Moreover, let $f_m$ denotes a sequence of noisy data satisfying* (2.12) *with noise level $\delta_m \to 0$. If we denote by $u^m = u_{k_*(\delta_m)}$ the regularized solutions obtained for data $f_m$, then there exists a subsequence $u^{m_\ell}$ that converges in the weak-* topology of $BV$, and the limit of each convergent subsequence is a minimizer of $H(., u_*)$. Furthermore, if $u_*$ is the unique minimizer of $H(., u_*)$ then the whole sequence $u^m$ converges to $u_*$ in the weak-* topology.*

It is easy to see that $k_*$ is well-defined because $H(u_k, f)$ is monotonically decreasing and $H(u_k, f) \to 0$ as $k$ increases. Furthermore, if we sum the inequalities (2.7) from 1 to $k$, and then add the following inequality derived from the convexity of the functional $H(u, f)$

$$J(u_0) + D(u_*, u_0) + \lambda H(u_0, f) \leq J(u_*) + \lambda H(u_*, f), \tag{2.15}$$

we obtain:

$$2^{k_*}\lambda H(u_{k_*}, f) \leq J(u_*) + \sum_{j=0}^{j=k_*} 2^{j-1}\lambda H(u_j, f) \leq 2^{k_*}\lambda\delta + J(u_*). \tag{2.16}$$

This means

$$\tau\delta \leq H(u_{k_*}, f) \leq \delta + \tfrac{J(u_*)}{2^{k_*}\lambda},$$

i.e. $2^{k_*} \leq \frac{J(u_*)}{\lambda(\tau-1)\delta}$.

In order to prove the weak-* convergence of subsequences, it suffices to prove that the BV-norm of $u^m$ is uniformly bounded with respect to $m$, which can be obtained by analogous arguments to the proofs of Theorem 2.3 and the semiconvergence result in [10]. □

**3. Numerical Results.** In this section, we will present some numerical results obtained from two different fidelity term models by using our iterative regularization procedure. We shall also compare them with the results from the original iterative regularization procedure (cf. [10]) and the hierarchical decomposition algorithm by Tadmor-Nezzar-Vese(T-N-V)(cf. [15, 14]). For the sake of appropriate comparison, we change the coefficient $\frac{1}{2^{k-1}\lambda}$ to $\frac{1}{2^k\lambda}$ in the model (2.1), which is equivalent to doubling $\lambda$, and finally allows to choose the same parameter $\lambda$ in all three methods.

**3.1. $L^2$ Fidelity Model.** We start with the iterative total variation regularization using an $L^2$ fidelity term (but not the square of the $L^2$ norm as in [10]), i.e., we

$$u_k \in \arg\min_{u \in BV(\Omega)} \left\{ ||u - f||_{L^2} + \frac{1}{2^k\lambda} D(u, u_{k-1}) \right\}. \tag{3.1}$$

We start with a noisy satellite image with Gaussian white noise with $\delta = 20.0$ and $SNR = 11.5$. In Figures 3.1 3.2 and 3.3 provide a comparison of the three methods mentioned above with $\lambda = 10.0$. In Figure 3.1, (d)-(f) and (j)-(l) display the recovered image $u_k$ resulted from our new Bregman iteration, and (g)-(i) and

(m)-(o) display the corresponding residual image $u_k - f$. The choice of a small value of $\lambda$ yields an oversmoothed iteration $u_0$, and $u_k$ gradually improves as $k$ increases until the stopping criterion becomes valued at $k_* = 3$. The plot of the residual in (p) confirms the monotonicity theorem, $H(u_k, f)$ simply decreases for all $k$. Moreover, at the $k$th step, $H(u_k, f)$ is bounded above by $\frac{J(f)}{2^k\lambda}$. the residual $H(u_k, f) = \|f - u_k\|$ The plots in (q) and (r) showing thes error $\|u_* - u_k\|$ and $D(u_*, u_k$ as a function of the iteration index are in good agreement with the noisy data theorem, which states that both the $L^2$ distance and the Bregman distance between $u_k$ and the true solution $u_*$ decrease as long as $H(u_k, f) > \delta$. Figure 3.2 is obtained from the iterative total variation regularization model proposed in [10], i.e, we do not multiply 2 to $\lambda$ after each iteration. We observe similar but slightly worse results as in Figure 3.1. From (m) one observes that the residual $H(u_k, f)$ is decreasing slower than in the modified version, which is not too surprising since the residual is multiplied by increasing parameters in the modified minimization problems.

Figure 3.3 displays the results obtained from the T-N-V algorithm (i.e., Tikhonov-Morozov) by using the $BV + L^2$ model. The recovered $u_3$ in (g) seems to be the visually best one. From the rescaled version of $u_3 - f$ in (j) one observes that almost all the important features of the satellite including the antenna are removed from the residual. The plot of the $L^2$-error in (n) shows a very similar behaviour to the iterative total variation regularization; the closest distance between $u_k$ and the exact image $u_*$, $||u_3 - u_*||_{L^2} = 12.3$, is even better than $||u_{k_*} - u_*||_{L^2} = 14.9$ in Figure 3.1. However, at the moment there are no rigorous convergence and monotonicity results for the T-N-V algorithm using the $BV + L^2$ norm. Moreover, it is not clear how the T-N-V algorithm behaves with respect to Bregman distances, which might provide more information about the convergence speed of certain features such as edges (cf. [3]). Unfortunately one cannot even compute suitable Bregman distances during the T-N-V algorithm since it does not provide any subgradients.

In the next three pictures we reconstruct a noisy image of a face with Gaussian white noise $\delta = 20.0$ and $SNR = 8.89$. The penalty parameter is chosen as $\lambda = 20.0$. One observes obvious improvements in the solutions as $k$ increases in all these three cases. Figures 3.4, 3.5 and 3.6 correspond again to our new iterative total variation method, the iterative total variation method, and the T-N-V decomposition algorithm, respectively. When we compare the distance in the $L^2$-norm between the best recovered $u_{k_*}$ and the exact image $u_*$ ( $||u_{k_*} - u_*||_{L^2}$ ) for the three methods, we find that the T-N-V decomposition algorithm $||u_3 - u_*||_{L^2} = 9.3$ still yields the lowest among all the three methods, compared to $||u_4 - u_*||_{L^2} = 10.4$ for the new and $||u_5 - u_*||_{L^2} = 11.1$ for the original iterative total variation regularization.

**3.2. $L^1$ Fidelity Model.** In this section we apply the new iterative regularization procedure using the $L^1$ fidelity term model

$$u_k \in \arg\min_{u \in BV(\Omega} \left\{ ||u - f||_{L^1} + \frac{1}{2^k\lambda} D(u, u_{k-1}) \right\}. \tag{3.2}$$

Corresponding to the motivation of the $L^1$-model as a suitable fidelity for binary images, we consider the denoising of a black-and white finger print image as a test case

Figures 3.7, 3.8 and 3.9 show restored images with the three different methods (same order as before), for Gaussian white noise $\delta = 10.0$ and $SNR = 14.8$. The results show similar properties as noticed in the $L^2$-case in the previous subsection. At the earlier stage, $u_k$ is over-smoothed. As $k$ increases, $u_k$ is getting closer to the

exact image $u_*$ in terms of $L^1$ distance, until $k_*$ is reached, and subsequently noise is added to $u_k$. Comparing the value of $||u_{k_*} - u_*||_{L^1}$ among the three models we see that the differences are very small this time, less than 0.1. However, visually the result of the T-N-V model seems more noisy than the ones obtained from the iterative total variation method.

A difference to the $L^2$-case is the behaviour of the original iterative total variation regularization. In the case of the $L^2$ fidelity term (see (m) in in both Figure 3.2 and Figure 3.5), $H(u_k, f)$ was decreasing rather slow, which is not the case for the $L^1$ fidelity term model (see (m) in Figure 3.8). In fact, the fidelity term $||u_k - f||_{L^1}$ seems to converge to zero even though we cannot prove that the original iteration model based on the $L^1$ fidelity term is well defined. A detailed analysis of this effect might be an interesting problem for future study.

Finally we draw attention to possible cartoon-texture decompositions (cf. [4, 13]) for the $BV + L^1$ model, which corresponds to (d), (g) and (e), (h) of Figure 3.7 for the first and second iteration. One observes that the main cartoon is already incorporated in the image after the first step, while the texture remains in the residual $v = f - u$. In the later steps the texture is gradually incorporated into the image. Hence, as a by-product of our algorithm we obtain cartoon-texture decompositions at different scales.

**4. Conclusion.** In this paper, we have proposed a modified iterative total variation regularization procedure (2.1) that can be applied to any variational model given by $H(u, f) + \frac{1}{\lambda} J(u)$, with a rather wide possibility of choosing the fidelity term $H$ (e.g. any Banach space norm) By doubling the penalty parameter $\lambda$ at each iteration step, we are able to show that the iteration procedure is well-defined, and subsequently following the approach in [10] we can also obtain convergence results.

The numerical experiments indicate that the iteration procedure yields high-quality reconstructions and has a similar multiscale nature as the previously known T-N-V Tikhonov-Morozov iteration. In some cases the T-N-V method seems to perfrom slightly better than the iterative total variation approach, but it might also be more sensitive to noise. Moreover, the new iterative total variation approach seems to be only one allowing for a complete convergence analysis so far.

An interesting future study is to apply our new procedure to the $BV + G$ model (1.3) suggested by Meyer. Since the $G$ norm is a Banach norm satisfying the assumptions nonnegativity, convexity and positive homogeneity, it seems possible that our approach could improve the denoising performance of the model.

**Acknowledgements.** The work of and L.H. and S.O. has been supported by the NIH through grant U54RR021813, and the NSF through grants DMS-0312222, ACI-0321917 and DMI-0327077. The work of M.B. has been supported by the Austrian National Science Foundation FWF through project SFB F 013 / 08 and by the Johann Radon Institute for Computational and Applied Mathematics (Austrian Academy of Sciences ÖAW).

REFERENCES

[1] J. F. Aujol, G. Aubert, L. Blanc-Feraud, and A. Chambolle, *Decomposing an image: Application to textured images and SAR images*, J. Math. Imaging Vision, (2005).

[2] L. M. Bregman, *The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. and Math. Phys., 7 (1967), pp. 200–217.

[3] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1422.
[4] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized $L^1$ function approximation*, CAM-Report 04-07, UCLA, Los Angeles, CA, (2004).
[5] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using bregman functions*, SIAM J. Optim., 3 (1993), pp. 538–543.
[6] D. Goldfarb and W. Yin, *Second order cone programming methods for total variation-based image restoration*, CORC Report TR-2004-05, Columbia University, New York, (2004).
[7] K. C. Kiwiel, *Proximal minimization methods with generalized bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142–1168.
[8] H. Lin, A. Marquina, and S. Osher, *Blind deconvolution using TV regularization and bregman iteration*, UCLA CAM report, (2004), pp. 04–51.
[9] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, AMS, Providence, RI, 2001.
[10] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation based image restoration*, Multiscale Modeling and Simulation. to appear 2005.
[11] R. Plato, *The discrepancy principle for iterative and parametric methods to solve linear ill-posed problems*, Numer. Math., 75 (1996), pp. 99–120.
[12] L. I. Rudin, S. J. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[13] O. Scherzer and C. Groetsch, *Inverse Scale Space Theory for Inverse Problems*, Springer-Verlag, New York, 2001.
[14] O. Scherzer and C. Groetsch, *Inverse scale space theory for inverse problems*, in Scale-Space and Morphology in Computer Vision, Lecture Notes in Comput. Sci. 2106, M.Kerckhove, ed., Springer, New York, 2001, pp. 317–325.
[15] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical $(BV, L^2)$ decompositions*, Multiscale Model. Simul., 2 (2004), pp. 554–579.

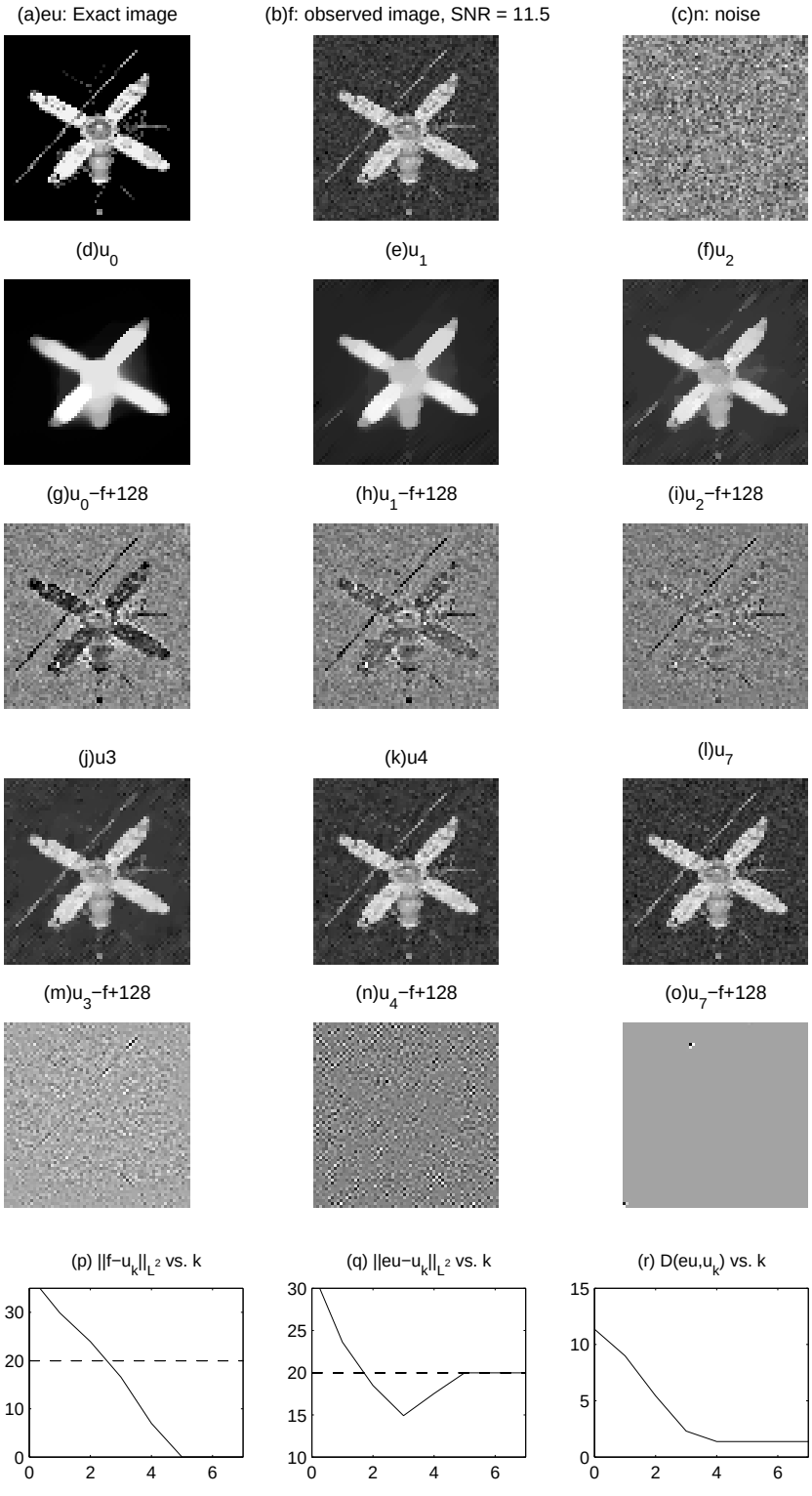


FIG. 3.1. *Satellite reconstructions with the new iterative total variation method, Gaussian noise* $\delta = 20$, $\lambda = 10$.

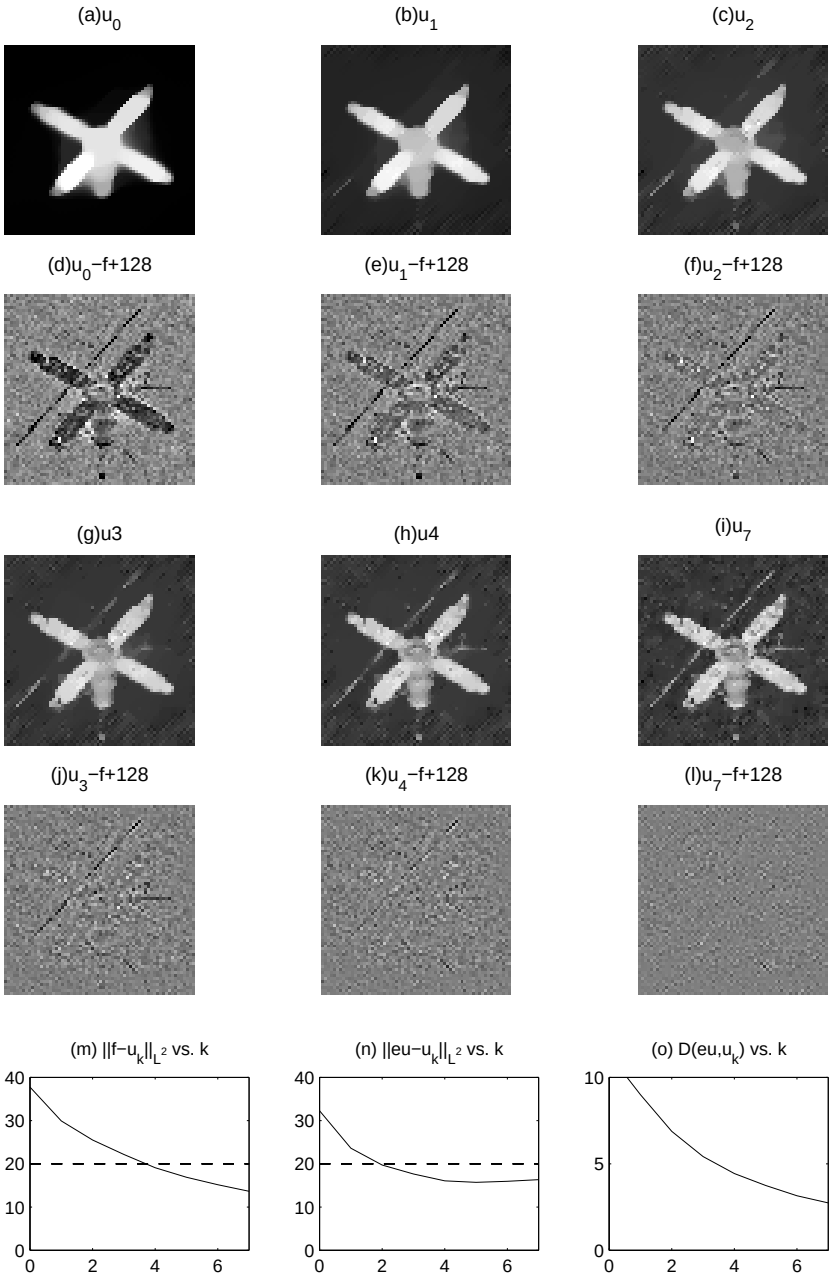


FIG. 3.2. *Satellite reconstructions with the original iterative total variation method, Gaussian noise* $\delta = 20$, $\lambda = 10$.

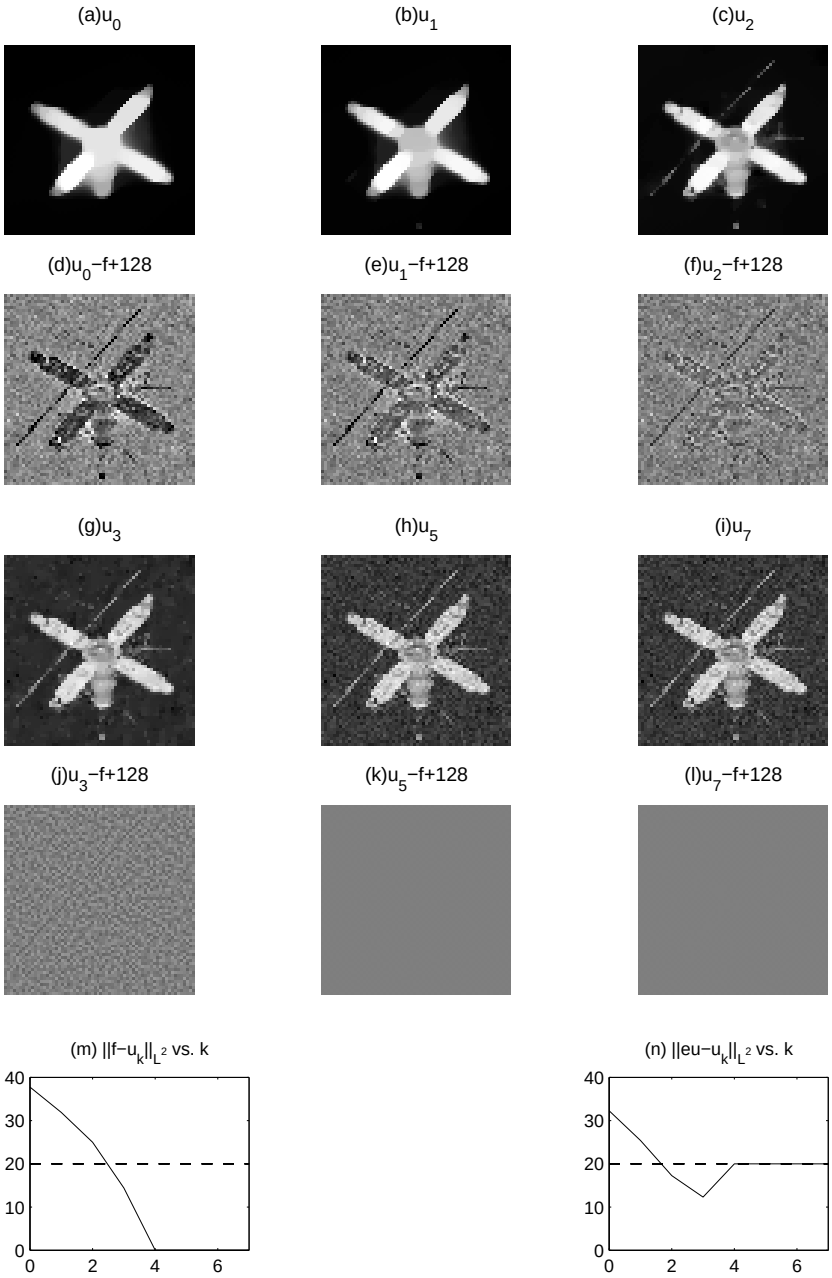


FIG. 3.3. *Satellite reconstructions with the $BV + L^2$ T-N-V method, Gaussian noise $\delta = 20$, $\lambda = 10$.*

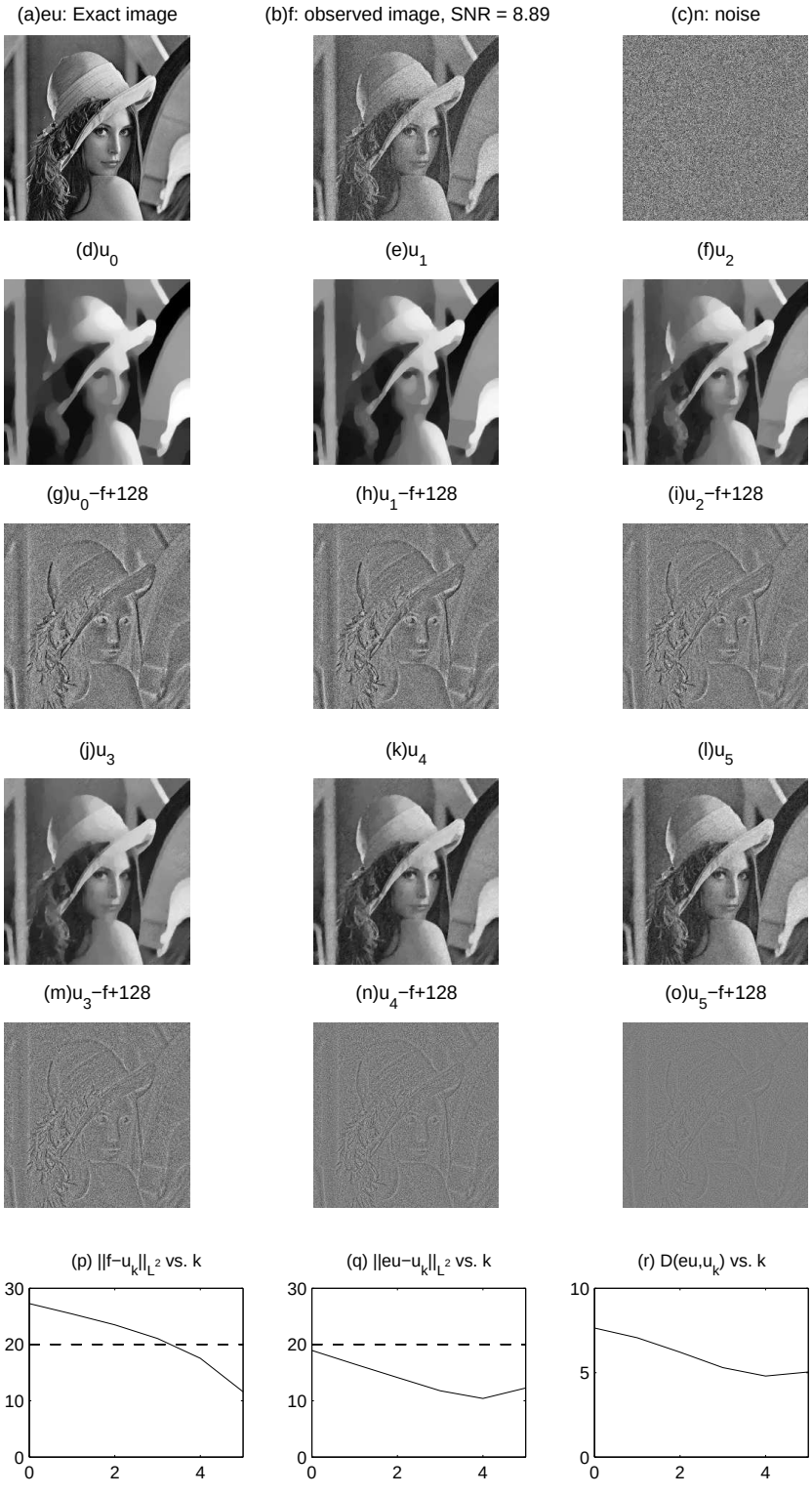


FIG. 3.4. *Restorations of a face obtained with the new iterative TV model, Gaussian noise,* $\delta = 20$, $\lambda = 20$.

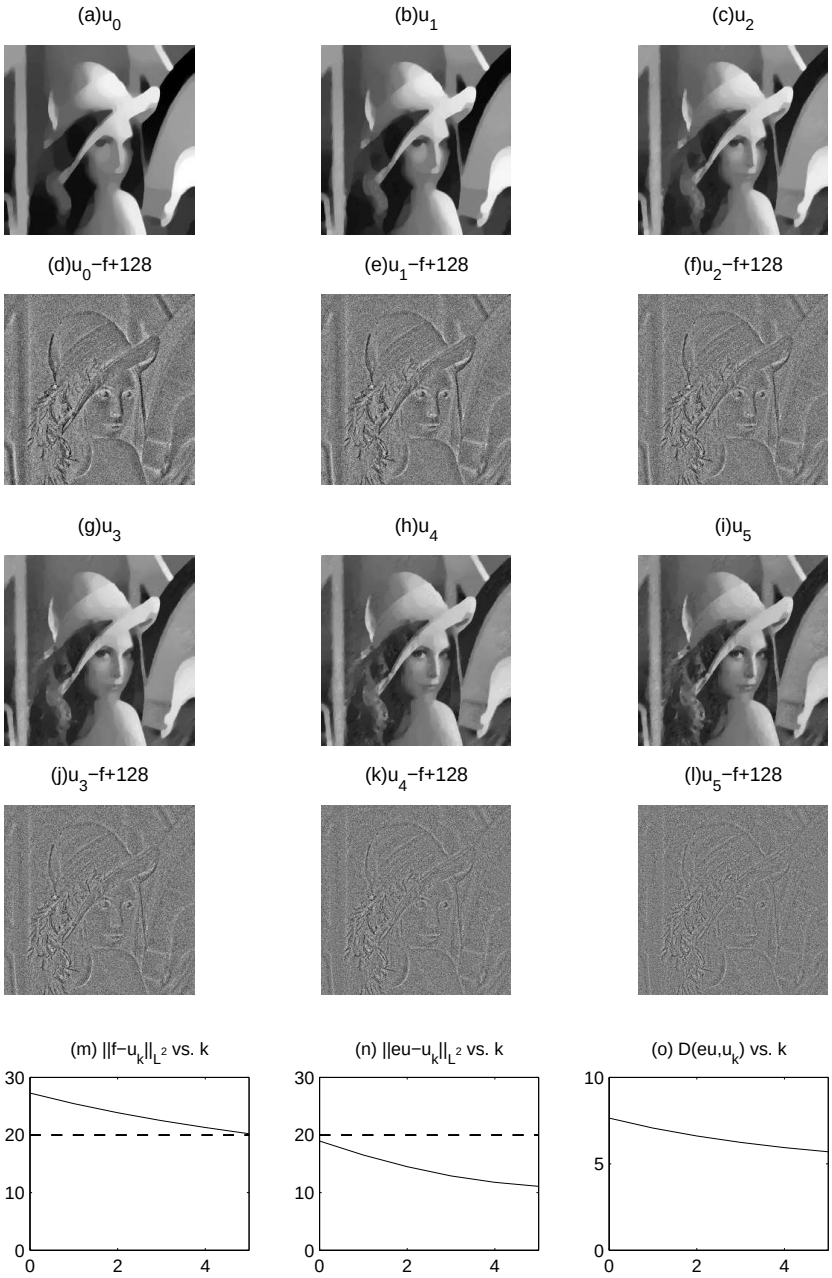


FIG. 3.5. *Restorations of a face obtained with the original iterative TV model, Gaussian noise,* $\delta = 20$, $\lambda = 20$.

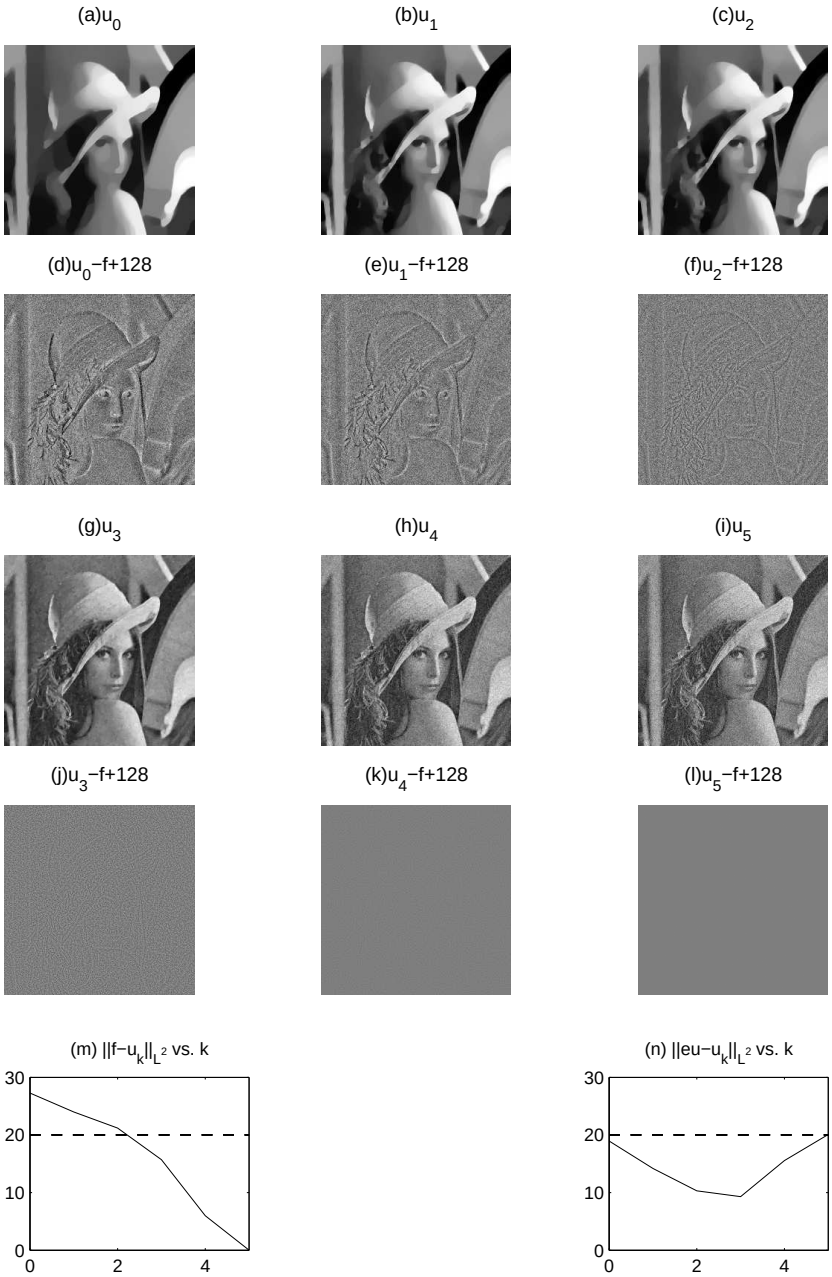


FIG. 3.6. *Restorations of a face obtained with the $BV + L^2$ T-N-V model , Gaussian noise, $\delta = 20$, $\lambda = 20$.*

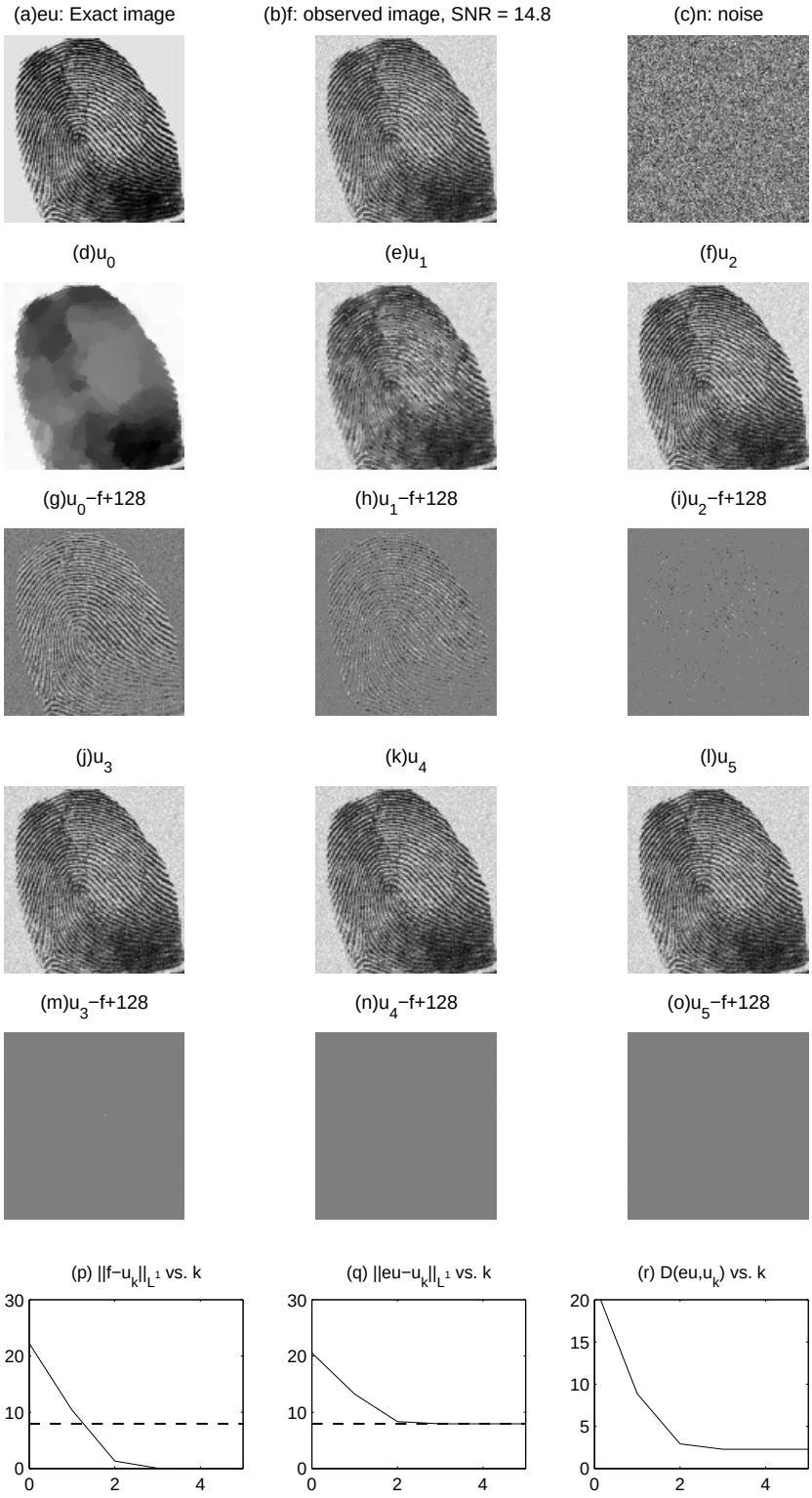


FIG. 3.7. *Restorations of a finger print obtained with the new iterative TV model, Gaussian noise,* $\delta = 10$, $\lambda = 1.0$.

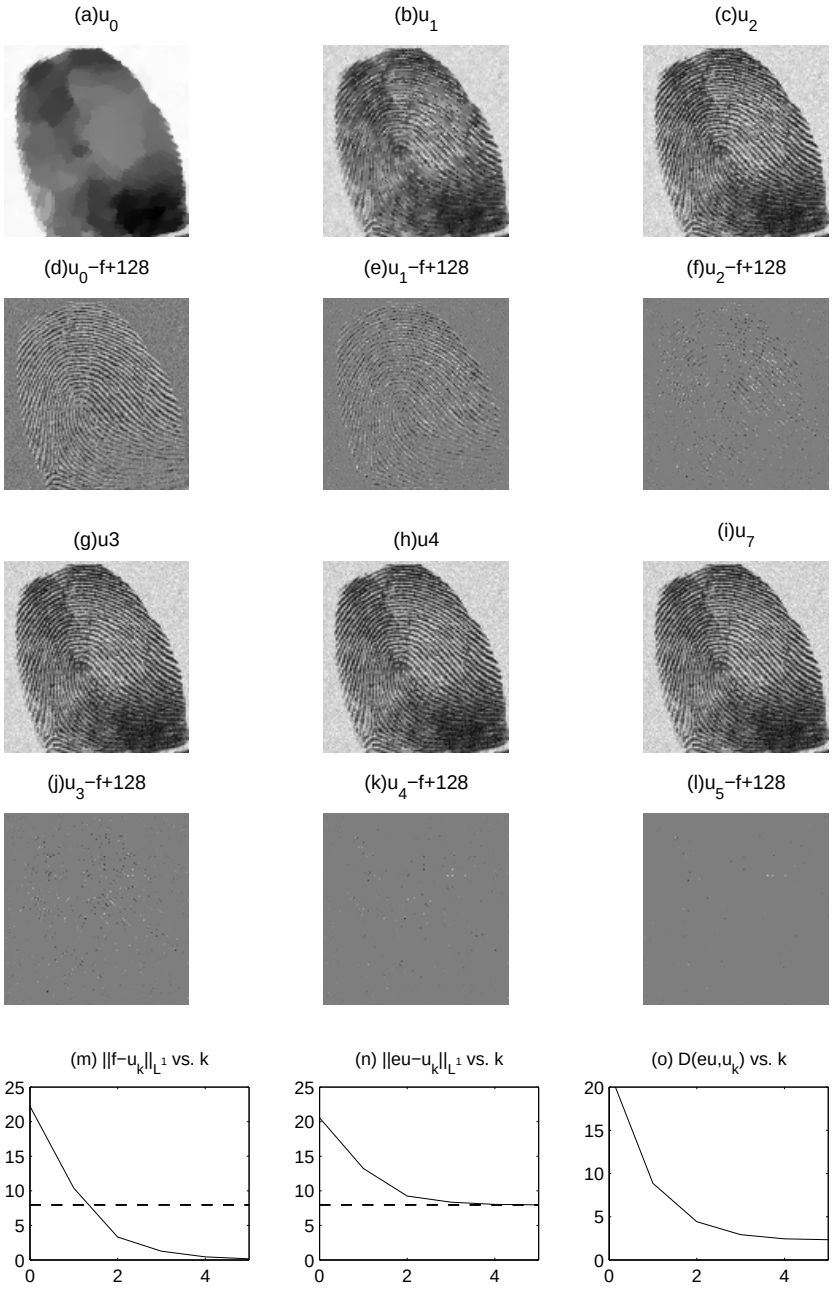


FIG. 3.8. *Restorations of a finger print obtained with the original iterative TV model, Gaussian noise,* $\delta = 10$, $\lambda = 1.0$.

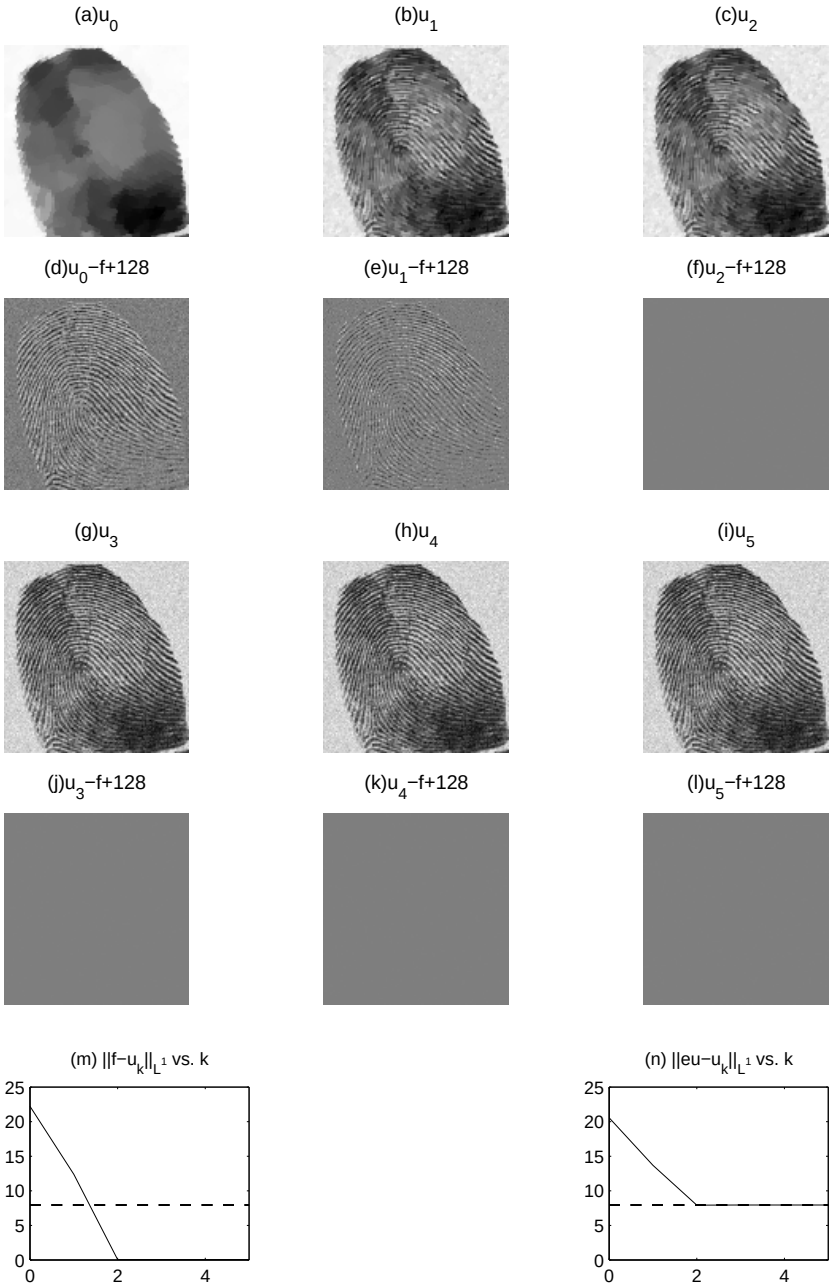


FIG. 3.9. *Restorations of a finger print obtained with the $BV+L^1$ T-N-V model, Gaussian noise, $\delta = 10$, $\lambda = 1.0$.*