VARIATIONAL IMAGE DENOISING WHILE CONSTRAINING THE DISTRIBUTION OF THE RESIDUAL[*]

ALESSANDRO LANZA[†], SERENA MORIGI[†], FIORELLA SGALLARI[†], AND ANTHONY J. YEZZI[‡]

Dedicated to Lothar Reichel on the occasion of his 60th birthday

Abstract. We present a denoising method aimed at restoring images corrupted by additive noise based on the assumption that the distribution of the noise process is known. The proposed variational model uses Total Variation (TV) regularization (chosen simply for its popularity; any other regularizer could be substituted as well) but constrains the distribution of the residual to fit a given target noise distribution. The *residual distribution constraint* constitutes the key novelty behind our approach. The restored image is efficiently computed by the constrained minimization of an energy functional using an Alternating Directions Methods of Multipliers (ADMM) procedure. Numerical examples show that the novel residual constraint indeed improves the quality of the image restorations.

Key words. image denoising, variational models, image residual, probability distribution function

AMS subject classifications. 68U10, 65K10, 65F22, 62E17

1. Introduction. A real captured image may be distorted by many expected or unexpected factors among which random noise is a typical and often unavoidable example. Hence, image denoising is a fundamental task in the field of image processing and a plethora of noise removal approaches have been proposed throughout the last few decades. Basically, there are three standard noise models in imaging systems. These are additive noise, multiplicative noise, and impulse noise. Typical image noise models are further characterized by the shape of their probability density function which in the discrete setting is represented by the noise histogram. In this paper, we focus on the restoration of images corrupted by additive noise, which we assume to be sampled from a known a-priori distribution. Note that previous work has been done to impose constraints on the histogram of the restored image itself. In this work, however, the focus is on the histogram of the residual. To our knowledge, this is the first attempt to impose such constraints.

Representing gray-scale two-dimensional images by real-valued functions defined on a rectangular domain $\Omega \subset \mathbb{R}^2$, the available observed noise-contaminated image u_0 is related to the unknown noise-free image $\bar{u}$ by the following degradation model

$$(1.1) \qquad u_0(x) = \bar{u}(x) + \bar{n}(x), \qquad x \in \Omega,$$

where $\bar{n}$ is an unknown realization of the random noise process, which we assume to be identically distributed with known probability density function. The goal of a denoising algorithm is to obtain an estimate u of the unknown noise-free image $\bar{u}$. This, from (1.1), allows us to define the *residual* as $n = u_0 - u$ which represents an estimate of the unknown noise realization $\bar{n}$. In principle, the greater the a priori information available on the noise-free image and the noise is, the better the chance for a successful denoising process will be. More general, in image restoration it is often beneficial to impose known properties of the noise-free image such as smoothness [20] and nonnegativity [15] during the solution process. On the other hand, most image restoration methods find the approximation u of $\bar{u}$ by using the a

priori information on the mean and the variance of the noise, that is, they mainly exploit only the norm of the residual. Among these methods, many including the discrepancy principle, the generalized cross validation, the Truncated Singular Value Decomposition [16], or the L-curve method [2, 5], exploit this limited amount of residual information in the regularization parameter choice.

One aspect generally missing from state-of-the-art image denoising algorithms is a full exploitation of all the available a-priori information on the noise. The main purpose of this work is to go beyond the use of the residual norm and to propose a new method that takes advantage of a higher amount of information present in the components of the residual vector since in many cases the noise distribution is known[1].

This concept is better illustrated in Figure 1.1. Three synthetic examples of additive image noise all with the same mean $\mu = 127.5$ and standard deviation $\sigma = 27.5$ are depicted in the first row, nevertheless the first noise example is completely different from the other two. In particular, the first is a simple step function, the second and the third are specific realizations of Gaussian and uniform noise distributions, respectively. In the second row, the associated histograms are shown which reflect the visual difference we can notice in the illustrations. Even if the second and the third illustrations appear visually similar, the associated histograms are significantly different from each other. Therefore, by exploiting the complete information on the noise distribution, we expect to obtain a residual which better fits the probabilistic model of the noise process, and, as a consequence, we expect an improvement in the restored image.

To the best of our knowledge, no previous attempt has been made to explicitly favor a target distribution *for the residual* during the denoising process. The histogram of the residual has been proposed for choosing the regularization parameter in [22] to make the residuals as close as possible to white noise. In [14] a novel fidelity functional was proposed in a variational framework in order to enforce whiteness of the residual. The histogram *of the image itself* has been recently proposed in [24] for image restoration, however it is based on the strong assumption that the histogram of the original image is known.

While our formulation is quite different due to our focus on the residual, it can nevertheless be related on a conceptual level with numerous methods that have been proposed for image contrast enhancement or segmentation problems which are based on the modification of the histogram of an input image toward a target histogram. The simplest such method is histogram linear stretching [17]. Sapiro and Caselles in [23] proposed histogram modification via image evolution equations and R. Chan et al. in [8] proposed a general variational framework for histogram modification.

Over the last two decades, a variety of PDE-based and variational methods have been developed to deal with the image denoising problem. A good review can be found in [4]. In these approaches, the use of variational methods [7, 10, 13, 18, 21, 26] and nonlinear partial differential equations (PDEs) [1, 6, 11, 19, 27] have significantly grown.

Our idea is to develop a variational denoising model which integrates soft or hard constraints to fit the distribution of the noise.

The paper is organized as follows. In Section 2 the variational approach to image denoising is illustrated. Section 3 describes how we defined the novel constraints based on the residual cumulative density function. In Section 5 the ADMM optimization technique is proposed to solve the constrained variational model. In Section 6 preliminary experimental results are presented and conclusions are drawn in Section 7.

---

[1]Technically, the distribution of the "noise process" is known with the actual noise being a realization of this process. If the process is spatially ergodic, then the histogram of the realized noise can be expected to resemble the distribution of the noise process itself.
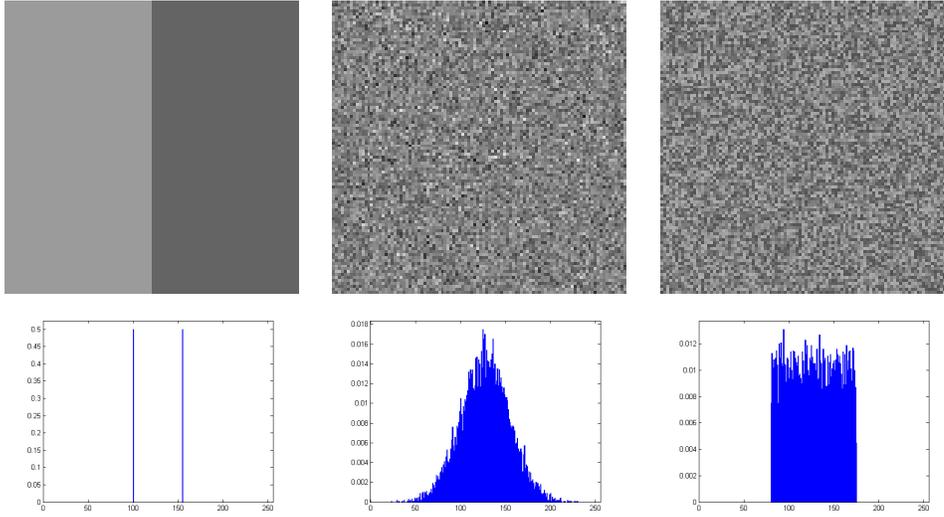
FIG. 1.1. *Top row: step image (left), realization of Gaussian (center) and uniform (right) noise distributions; bottom row: associated histograms.*

2. The variational approach to image denoising.

Variational approaches for image denoising commonly rely on the following energy functional minimization:

$$(2.1) \qquad \min_u \left\{ J(u) = \mathcal{R}(u) + \lambda \mathcal{F}(u_0 - u) \right\}.$$

The energy functional J minimized in (2.1) is thus the sum of a regularization term $\mathcal{R}$ and a so-called fidelity term $\mathcal{F}$ with the regularization parameter $\lambda > 0$ controlling the trade-off between "regularity" of the solution u and fidelity of the solution to the observed data u_0. In particular, the regularization functional in (2.1) encodes prior information on the smoothness of the unknown noise-free image $\bar{u}$ while the fidelity functional is based on the assumption on the residual $n = u_0 - u$, that is, on the additive noise $\bar{n}$ in (1.1).

A standard choice for the fidelity term is

$$(2.2) \qquad \mathcal{F}(u_0 - u) = \frac{1}{2} \int_\Omega (u_0 - u)^2 \, dx,$$

which indirectly encodes the prior knowledge on the noise standard deviation σ.

The type of the regularization functional in (2.1) is important for the success of the denoising process. A very popular choice for it is the Total Variation semi-norm since it has the desirable property to allow for sharp edges in the solution. The popular Rudin-Osher-Fatemi (ROF) denoising algorithm [21] considers the TV-L^2 functional,

$$(2.3) \qquad \min_u \left\{ \int_\Omega |\nabla u| dx + \frac{\lambda}{2} \int_\Omega (u_0 - u)^2 dx \right\}.$$

This model yields very satisfactory results for removing image noise while preserving edges and contours of objects.

The unconstrained TV-L^2 model in (2.3) can be equivalently reformulated as the following constrained optimization problem:

$$\min_{u \in \mathcal{V}} \int_\Omega |\nabla u| dx,$$

with the admissible set defined as

$$\mathcal{V} = \left\{ u : \|u_0 - u\|_2^2 \leq |\Omega|\,\sigma^2 \right\} ,$$

where $|\Omega|$ is the area of the image domain.

Instead of the standard L_2 fidelity term in (2.3), in order to explicitly exploit the assumption that the noise distribution is known, we introduce a novel *residual fidelity term* that enforces the similarity between the residual distribution and the target noise distribution, thus obtaining the penalty formulation,

$$(2.4) \qquad \min_u \left\{ \int_\Omega |\nabla u|\, d\Omega + \frac{1}{2} \int_\mathbb{R} \lambda(z) \left(\overline{P}(z) - P_n(z) \right)^2 dz \right\}$$

where $\overline{P}$ denotes the known noise distribution in (1.1) and P_n is the distribution function of the residual $n = u_0 - u$.

For instance, if the noise is known to be zero-mean Gaussian with a certain standard deviation, $\overline{P}$ will be the associated cumulative distribution function. Details on these distributions will be provided in Section 4.

In this paper, we propose a hard-constrained version of the variational model in (2.4) aimed at forcing explicitly the constraints on the noise distribution function of the residual. We chose the TV regularization simply for its popularity; any other regularizers could be substituted as well. The proposed model is

$$(2.5) \qquad \min_{u \in \mathcal{H}} \int_\Omega |\nabla u|\, dx ,$$

where the new admissible set $\mathcal{H}$ contains solutions u such that the corresponding residual $u_0 - u$ has a distribution close to the theoretical noise distribution function. We will characterize the set $\mathcal{H}$ and, accordingly, we will present an efficient minimization algorithm based on the Alternating Directions Method of Multipliers (ADMM) optimization technique. The ADMM method and its variants are largely used to solve minimization problems in image processing. We refer the reader to [3] for a general dissertation on optimization techniques such as ADMM methods or others and their applications to image processing.

Let us remark how the proposed residual fidelity term in (2.4) and the related constraint $u \in \mathcal{H}$ can be seen as an infinite dimensional extension of the classical ones in (2.2). Instead of constraining the mean and variance of the residual only, we force the entire distribution $P_n(z)$ (i.e., for any real intensity $z \in \mathbb{R}$) toward the known noise distribution $\overline{P}$. As such, the proposed *residual distribution term* is infinitely more constraining than the classical mean and variance penalties.

3. Continuous distribution constraints. In this section we present a few preliminary equations that will be exploited in the following sections when we develop models that constrain (either in the form of a penalty or in the form of a hard constraint) the distribution of the residual of the reconstruction. The basic idea, which will be explained in more detail in the following sections, is to force the residual to exhibit some of the known a-priori properties of the distribution of the additive noise embedded in the original observed image. Since our models are developed in the variational framework, we will need to address constraints on the distribution of the residual in this framework as well. Here we present the necessary preliminaries. Note that we will focus our attention on the cumulative distribution function rather than the standard probability distribution function as it exhibits an extra degree of regularity while still containing the same amount of information about the global distribution of the values of a given function.

Let $p(z, t)$ denote the probability density function (pdf),

$$(3.1) \qquad p(z,t) = \frac{1}{|\Omega|} \int_\Omega \delta\big(z - f(x,t)\big)\, dx \, ,$$

where δ represents the standard Dirac measure. Then, with $P_f(z,t)$, we denote the time-varying cumulative distribution function (cdf) of a time-varying function $f(x,t)$ of a 2D space variable $x \in \Omega$ as

$$(3.2) \qquad P_f(z,t) = \frac{1}{|\Omega|} \int_\Omega H\big(z - f(x,t)\big)\, dx$$

where $\Omega \subset \mathbb{R}^2$ is a fixed, compact domain and $H \colon \mathbb{R} \to \{0,1\}$ represents the standard Heaviside function. The time derivative of P_f is

$$\frac{\partial P_f}{\partial t} = -\frac{1}{|\Omega|} \int_\Omega \dot{H}\big(z - f\big) \frac{\partial f}{\partial t}\, dx = -\frac{1}{|\Omega|} \int_\Omega \delta\big(z - f\big) \frac{\partial f}{\partial t}\, dx$$

where δ represents the standard Dirac measure.

Now consider a weighted penalty term $\mathcal{F}_{cdf}\big(f(t)\big)$ that measures the difference between $P_f(z,t)$ and a given target cdf $\bar{P}(z)$,

$$(3.3) \qquad \mathcal{F}_{cdf}\big(f(t)\big) = \frac{1}{2} \int_D \lambda(z) \big(\bar{P}(z) - P_f(z,t)\big)^2 dz \, ,$$

where the interval $D \subset \mathbb{R}$ contains the range of values of f and where the weighting parameter $\lambda > 0$ is allowed to vary within this interval. We may interpret such a term as a weighted fidelity term between the target cdf and the actual cdf of the function f. Note that in the upcoming sections we will be utilizing such a term but applied to the residual rather than the reconstructed image, and therefore the function f presented here in these preliminary equations serves as a placeholder for the difference between the reconstructed image and the observed noisy image. As such, the interpretation of this as a fidelity term is to be understood in the sense of the residual having a similar distribution as the additive noise within the observed image. As the function f varies in time, the cdf penalty varies as follows:

$$\begin{aligned}
\frac{d}{dt} \mathcal{F}_{cdf}\big(f(t)\big) &= -\int_D \lambda(z)\big(\bar{P}(z) - P_f(z,t)\big) \frac{\partial P_f}{\partial t}(z,t)\, dz \\
&= \frac{1}{|\Omega|} \int_D \lambda(z)\big(\bar{P}(z) - P_f(z,t)\big) \left(\int_\Omega \delta\big(z - f(x,t)\big) \frac{\partial f(x,t)}{\partial t}\, dx \right) dz \\
&= \frac{1}{|\Omega|} \int_\Omega \frac{\partial f(x,t)}{\partial t} \left(\int_D \lambda(z)\big(\bar{P}(z) - P_f(z,t)\big) \delta\big(z - f(x,t)\big)\, dz \right) dx \\
&= \frac{1}{|\Omega|} \int_\Omega \frac{\partial f(x,t)}{\partial t} \lambda(f)\big(\bar{P}(f) - P_f(f,t)\big)\, dx \, .
\end{aligned}$$

As such, the gradient of $\mathcal{F}_{cdf}$ is given by

$$(3.4) \qquad D_{\mathcal{F}_{cdf}} = \frac{1}{|\Omega|} \lambda(f)\big(\bar{P}(f) - P_f(f)\big) \, .$$

Finally, it is worth noting that in the continuum (where our variational models are developed), standard a-priori noise distributions (such as the Gaussian) are often supported over unbounded intervals (including the case $D = \mathbb{R}$) whereas in the discretized and quantized

observed image, the range of the image values as well as the additive noise is truncated. As such, the residual will also be bounded. In this case, it makes sense to consider a target distribution supported over a finite interval D. We will present a Monte Carlo based approach to develop such bounded noise models in a later section.

We notice that the distribution penalty in (3.3) is the fidelity functional proposed in (2.4) with the generic function f replaced by the residual $n = u_0 - u$ and with $\overline{P}$ denoting the known cumulative distribution of noise.

The admissible set $\mathcal{H}$ which defines the hard constraint for the proposed variational problem (2.5) is defined as follows:

$$(3.5) \qquad \mathcal{H} = \left\{ u \colon\ P_{u_0-u}(z) = \bar{P}(z) \quad \forall z \in D \right\}.$$

In fact, if a residual $u_0 - u$ represents a realization of a noise with known distribution $\bar{P}(z)$, its cumulative distribution function $P_{u_0-u}(z)$ will fit perfectly $\bar{P}(z)$.

4. Discrete distribution constraints. In this section we illustrate how the distribution constraints on the residual $n = u_0 - u$ can be applied in the case of discretized and quantized images. Without loss of generality, we will consider square $d \times d$ images.

In the following, we will define the discrete counterpart of the probability density function (3.1) and of the cumulative distribution function (3.2). To this purpose, first we need to quantize the range of possible image values $z \in D$ by introducing a partition of D into Q (non-overlapping) intervals, called *bins*, defined as $b_i := [z_{i-1}, z_i], z_{i-1} < z_i, i = 1, \dots, Q$, as illustrated in Figure 4.1.
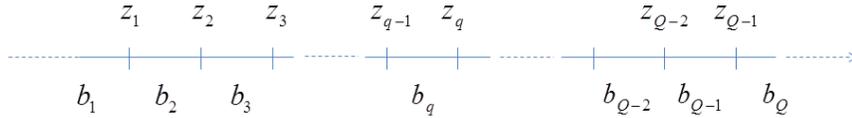
FIG. 4.1. *Partition of the image range into bins.*

The *normalized histogram* $h_n \in \mathbb{R}^Q$ of the residual n is defined as

$$h_n[q] = \frac{1}{d^2} \sum_{i,j=1}^{d} I_{b_q}(n_{i,j}), \quad q = 1, 2, \dots, Q$$

where $n_{i,j}$ denotes the value of n at pixel (i,j) and $I_S(\cdot)$ denotes the indicator function of the set S, i.e., the function having value 1 inside the set S and 0 outside.

Summing up, we get the *normalized cumulative histogram*,

$$H_n[q] = \sum_{i=1}^{q} h_n[i], \quad q = 1, 2, \dots, Q.$$

We can define *theoretical histograms* also in this discrete and quantized setting. Given the known noise probability density function $\bar{p}$ and cumulative distribution function $\bar{P}$ and a range partition, we define the *theoretical normalized histogram* $\bar{h}$ for the residual as

$$\bar{h}[q] = \bar{p}(z_q) - \bar{p}(z_{q-1}), \quad q = 1, 2, \dots, Q,$$

and the *theoretical normalized cumulative histogram* $\bar{H}$ as

$$\bar{H}[q] = \sum_{i=1}^{q} \bar{h}[i], \quad q = 1, 2, \dots, Q.$$
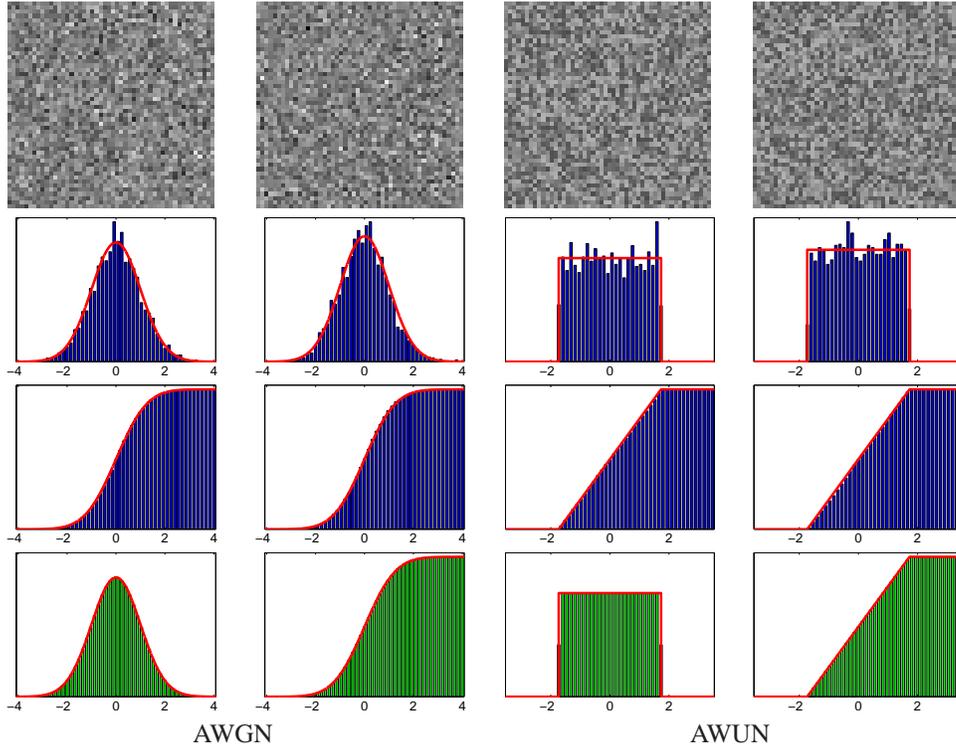
FIG. 4.2. *First row: pair of (50 × 50 pixels) different realizations from a standard identically distributed AWGN (first two columns) and AWUN (last two columns) process; second and third rows: associated normalized histograms and cumulative histograms; fourth row: theoretical histograms.*

The noise models we consider are the additive white Gaussian noise (AWGN) and the additive white uniform noise (AWUN), whose probability density functions are, respectively,

$$(4.1) \qquad p_G(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma} \right)^2 \right)$$

and

$$p_U(z) = \begin{cases} \dfrac{1}{2\sqrt{3}\,\sigma} & \text{for } |z-\mu| \le \sqrt{3}\,\sigma, \\[2mm] 0 & \text{otherwise,} \end{cases}$$

where μ and σ denote the mean and the standard deviation. To highlight the difference between the discrete and the continuous case, we show in Figure 4.2 four images (50 × 50 pixels) which represent different realizations from a standard (i.e., zero-mean and unit-variance) AWGN and AWUN (first row), the associated normalized histograms (second row), normalized cumulative histograms (third row), and theoretical histograms (fourth row). We notice that for both the AWGN and the AWUN, the two realizations yield quite different histograms and (though much less visible in the figures) cumulative histograms. Hence, in contrast with the continuous case, the distribution constraint can not be an equality constraint as in (3.5). Instead, we must constrain the histograms to reside within a band around the theoretical histograms.
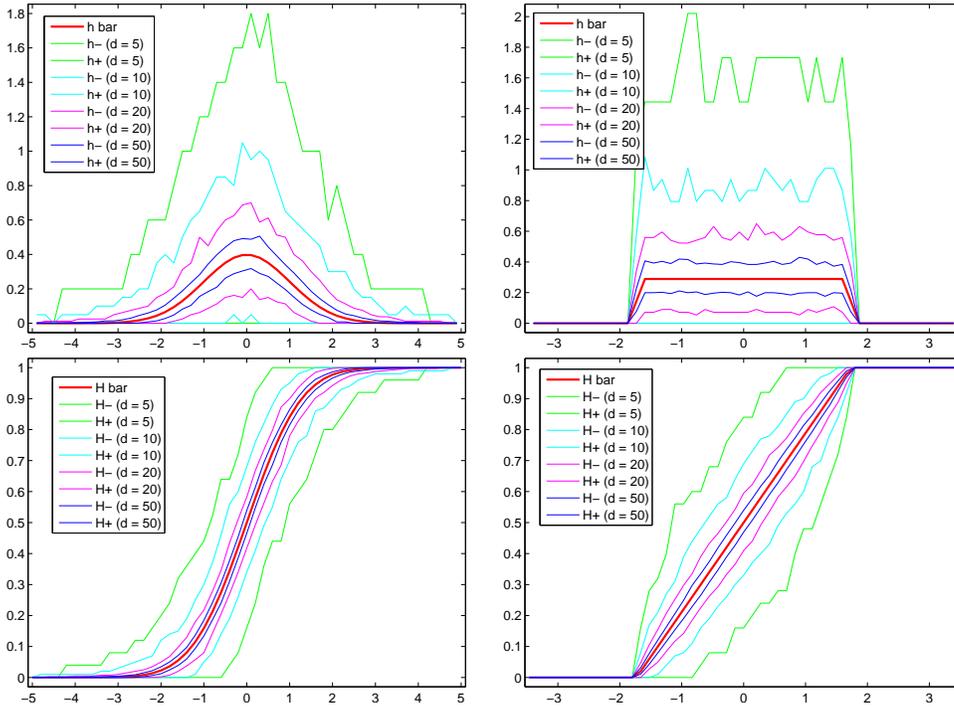
FIG. 4.3. *Upper and lower limits of the box-constraints on the residual histogram (top row) and cumulative histogram (bottom row) for standard AWGN (left column) and standard AWUN (right column) obtained for image dimensions $d = 5, 10, 20, 50$. We depict in red the theoretical histograms and cumulative histograms.*

We are ready to describe, in the discrete setting, the admissible set $\mathcal{H}$ which defines the hard constraint for the proposed variational problem (2.5) as

$$(4.2) \qquad \mathcal{H} = \left\{ u \in \mathbb{R}^{d^2} : \quad \bar{H}_L[i] \leq H_{u_0-u}[i] \leq \bar{H}_U[i] \quad \forall i = 1, \ldots, Q \right\}$$

where $\bar{H}_L$ and $\bar{H}_U$ are the lower and upper limits, respectively, of the band around the theoretical histograms.

For the computation of the $\bar{H}_L$ and $\bar{H}_U$ limits, we propose the following simple Monte Carlo approach. Given a target probability density function for the noise and a selected image dimension d, we generate a large number of $d \times d$ different noise realizations by sampling from the target distribution. For each realization, we compute the normalized histogram and the cumulative histogram. Finally, we compute the minimum and the maximum values within all the realizations for all the histogram bins. In Figure 4.3 we report the results for the AWGN and AWUN distributions obtained by using four different image dimensions $d = 5, 10, 20, 50$. First, we notice that the bands for the cumulative histograms are narrower than those for the histograms. This means that the random fluctuations of the cumulative histogram values over different realizations are smaller than those for the histogram values. For this reason, we choose to constrain the distribution of the residual by means of the cumulative histogram. Second, the size of the bands gets smaller as the image dimension d increases. In other words, the bigger the image dimension d is, the narrower the band of the distribution constraint will be. The histograms converge to the theoretical ones as d tends to infinity.

5. ADMM for the constrained minimization problem. To solve the proposed constrained minimization problem in (2.5) with the admissible set $\mathcal{H}$ defined in (4.2), we adapt the well known ADMM optimization technique [9] to our framework. We first introduce two auxiliary variables t and r to reformulate the minimization problem into the equivalent form,

$$u^* \leftarrow \underset{u,t,r\in\mathcal{H}}{\operatorname{argmin}} \ \sum_{i=1}^{d^2} \| t_i \|_2$$

(5.1)
$$\text{s.t.} : \quad t = Du, \ r = u_0 - u \,,$$

where the auxiliary variable t is introduced to transfer the discrete gradient operator D out of the non-differentiable term $\| \cdot \|_2$ and the variable r plays the role of the residual $u_0 - u$ within the distribution constraint so that the constraint is now imposed on r instead of u.

To solve (5.1), we define the augmented Lagrangian functional and seek its stationary points

$$\mathcal{L}(u,t,r;\lambda_t,\lambda_r) = \sum_{i=1}^{d^2} \| t_i \|_2$$

$$- \langle\, \lambda_t \,,\, t - Du \,\rangle \ + \ \frac{\beta_t}{2} \, \| \, t - Du \, \|_2^2$$

$$- \langle\, \lambda_r \,,\, r - (u_0 - u) \,\rangle \ \ + \ \frac{\beta_r}{2} \, \| \, r - (u_0 - u) \, \|_2^2 \,,$$

where $\beta_t > 0, \beta_r > 0$ are the scalar penalty parameters and $\lambda_t \in \mathbb{R}^{2d^2}$, $\lambda_r \in \mathbb{R}^{d^2}$ are the vectors of Lagrangian multipliers.

Starting at $u = u^k$, $\lambda_t = \lambda_t^k$, $\lambda_r = \lambda_r^k$, the ADMM iterative scheme applied to the solution of (5.1) reads as follows:

(5.2)
$$t^{k+1} \quad \leftarrow \quad \underset{t}{\operatorname{argmin}} \ \mathcal{L}(u^k,t,r^k;\lambda_t^k,\lambda_r^k)$$

(5.3)
$$r^{k+1} \quad \leftarrow \quad \underset{r}{\operatorname{argmin}} \ \mathcal{L}(u^k,t^{k+1},r;\lambda_t^k,\lambda_r^k)$$

(5.4)
$$u^{k+1} \quad \leftarrow \quad \underset{u}{\operatorname{argmin}} \ \mathcal{L}(u,t^{k+1},r^{k+1};\lambda_t^k,\lambda_r^k)$$

$$\begin{pmatrix} \lambda_t^{k+1} \\ \lambda_r^{k+1} \end{pmatrix} \leftarrow \begin{pmatrix} \lambda_t^k \ - \ \gamma\,\beta_t\left(\, t^{k+1} \ - \ Du^{k+1}\,\right) \\ \lambda_r^k \ - \ \gamma\,\beta_r\left(\, r^{k+1} \ - \ (u_0 - u^{k+1})\,\right) \end{pmatrix} ,$$

where γ is a relaxation parameter chosen in the interval $(0, (\sqrt{5}+1)/2)$ as analyzed in [12]. The two minimization sub-problems (5.2) and (5.4) for t and u can be easily solved by d^2 two-dimensional shrinkage operations and by the fast solution of a $d^2 \times d^2$ linear system [9], respectively. More attention must be payed to the sub-problem (5.3) for r that, when made explicit, reads as

(5.5)
$$r^{k+1} \leftarrow \underset{r\in\mathcal{H}}{\operatorname{argmin}} \ \left\| r - \left(u_0 - u^k + \frac{1}{\beta_r}\lambda_r \right) \right\|_2^2 = \mathcal{P}_{\mathcal{H}}\left[u_0 - u^k + \frac{1}{\beta_r}\lambda_r \right]$$

where $\mathcal{P}_{\mathcal{H}}[\cdot]$ denotes the Euclidean projection onto the distribution set $\mathcal{H}$ defined in (4.2). Computing the solution of (5.5) is not as straightforward as for the sub-problems (5.2) and (5.4), due to the characteristics of the set $\mathcal{H}$ onto which we project. However, from the definitions of the distribution set in (4.2), we notice that the complicated constraint on the auxiliary

variable r can be turned into a simple box-constraint on its cumulative histogram $H(r)$, that is, $H(r)$ must belong to the set,

$$(5.6) \qquad \mathcal{B} = \left\{ H(r) \in \mathbb{R}^Q \; : \; \bar{H}_L \; \leq \; H(r) \; \leq \; \bar{H}_U \right\},$$

where $\bar{H}_L, \bar{H}_U$ represent the lower and upper limits of the band around $\bar{H}$. This simpler set allows us to solve the overall optimization problem using ADMM with a new auxiliary variable $v = H(r)$. In particular, the minimization problem in (5.1) is rewritten as

$$u^* \leftarrow \operatorname*{argmin}_{u,t,r,v \in \mathcal{B}} \sum_{i=1}^{d^2} \| t_i \|_2$$

$$(5.7) \qquad \text{s.t.} \; : \quad t = Du, \; r = u_0 - u, \; v = H(r).$$

The augmented Lagrangian functional associated with (5.7) is

$$\mathcal{L}(u,t,r,v; \lambda_t, \lambda_r, \lambda_v) = \sum_{i=1}^{d^2} \| t_i \|_2$$

$$- \langle \lambda_t , t - Du \rangle + \frac{\beta_t}{2} \| t - Du \|_2^2$$

$$- \langle \lambda_r , r - (u_0 - u) \rangle + \frac{\beta_r}{2} \| r - (u_0 - u) \|_2^2$$

$$(5.8) \qquad - \langle \lambda_v , v - H(r) \rangle + \frac{\beta_v}{2} \| v - H(r) \|_2^2,$$

where $\beta_t, \beta_r, \beta_v > 0$ are the scalar penalty parameters and $\lambda_t \in \mathbb{R}^{2d^2}$, $\lambda_r \in \mathbb{R}^{d^2}$, $\lambda_v \in \mathbb{R}^Q$ are the vectors of Lagrangian multipliers.

Solving (5.7) is thus equivalent to search for the solutions of the saddle point problem,

$$\text{Find } (x^*; \lambda^*) \; \in \; X \times \Lambda$$
$$\text{s.t. } \mathcal{L}(x^*; \lambda) \; \leq \; \mathcal{L}(x^*; \lambda^*) \; \leq \; \mathcal{L}(x; \lambda^*)$$
$$\forall \, (x; \lambda) \; \in \; X \times \Lambda,$$

with $\mathcal{L}$ defined in (5.8) and where, for simplicity of notation, we set $x = (u, t, r, v)$, $\lambda = (\lambda_t, \lambda_r, \lambda_v)$, $X = \mathbb{R}^{d^2} \times \mathbb{R}^{2d^2} \times \mathbb{R}^{d^2} \times \mathbb{R}^Q$, and $\Lambda = \mathbb{R}^{2d^2} \times \mathbb{R}^{d^2} \times \mathbb{R}^Q$.

Starting at $u = u^k$, $\lambda_t = \lambda_t^k$, $\lambda_r = \lambda_r^k$, $\lambda_v = \lambda_v^k$, the ADMM iterative scheme applied to the solution of (5.7) reads as follows:

$$(5.9) \qquad t^{k+1} \quad \leftarrow \quad \operatorname*{argmin}_{t} \mathcal{L}(u^k, t, r^k, v^k; \lambda_t^k, \lambda_r^k, \lambda_v^k)$$

$$(5.10) \qquad r^{k+1} \quad \leftarrow \quad \operatorname*{argmin}_{r} \mathcal{L}(u^k, t^{k+1}, r, v^k; \lambda_t^k, \lambda_r^k, \lambda_v^k)$$

$$(5.11) \qquad v^{k+1} \quad \leftarrow \quad \operatorname*{argmin}_{v} \mathcal{L}(u^k, t^{k+1}, r^{k+1}, v; \lambda_t^k, \lambda_r^k, \lambda_v^k)$$

$$(5.12) \qquad u^{k+1} \quad \leftarrow \quad \operatorname*{argmin}_{u} \mathcal{L}(u, t^{k+1}, r^{k+1}, v^{k+1}; \lambda_t^k, \lambda_r^k, \lambda_v^k)$$

$$(5.13) \qquad \begin{pmatrix} \lambda_t^{k+1} \\ \lambda_r^{k+1} \\ \lambda_v^{k+1} \end{pmatrix} \leftarrow \begin{pmatrix} \lambda_t^k - \gamma \beta_t \left(t^{k+1} - Du^{k+1} \right) \\ \lambda_r^k - \gamma \beta_r \left(r^{k+1} - (u_0 - u^{k+1}) \right) \\ \lambda_v^k - \gamma \beta_v \left(v^{k+1} - H(r^{k+1}) \right) \end{pmatrix}.$$

In the following subsections, we show how to solve the four sub-problems (5.9)–(5.12) and then we present the iterative ADMM-based minimization algorithm.

5.1. Solving the sub-problem for t. Given the definition of the augmented Lagrangian functional in (5.8), the minimization sub-problem for t in (5.9) can be written as

$$
t^{k+1} \leftarrow \underset{t \in \mathbb{R}^{2d^2}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{d^2} \| t_i \|_2 - \langle \lambda_t^k, t - Du^k \rangle + \frac{\beta_t}{2} \| t - Du^k \|_2^2 \right\}
$$

$$
(5.14) \qquad = \underset{t \in \mathbb{R}^{2d^2}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{d^2} \| t_i \|_2 - \langle \lambda_t^k, t \rangle + \frac{\beta_t}{2} \| t - Du^k \|_2^2 \right\}
$$

$$
(5.15) \qquad = \underset{t \in \mathbb{R}^{2d^2}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{d^2} \left(\| t_i \|_2 - \langle \lambda_{t,i}^k, t_i \rangle + \frac{\beta_t}{2} \| t_i - D_i u^k \|_2^2 \right) \right\}.
$$

Note that in (5.14), we have omitted the constant terms while in (5.15) we have written the functional to be minimized in an explicit component-wise form. The minimization in (5.15) is equivalent to the d^2 two-dimensional problems

$$
t_i^{k+1} \leftarrow \underset{t_i \in \mathbb{R}^2}{\operatorname{argmin}} \left\{ \| t_i \|_2 - \langle \lambda_{t,i}^k, t_i \rangle + \frac{\beta_t}{2} \| t_i - D_i u^k \|_2^2 \right\}
$$

$$
(5.16) \qquad = \underset{t_i \in \mathbb{R}^2}{\operatorname{argmin}} \left\{ \| t_i \|_2 + \frac{\beta_t}{2} \left\| t_i - \left(D_i u^k + \frac{1}{\beta_t} \lambda_{t,i}^k \right) \right\|_2^2 \right\}, \quad i = 1, \ldots, d^2.
$$

Following [9] and setting

$$
(5.17) \qquad q_i^k := D_i u^k + \frac{1}{\beta_t} \lambda_{t,i}^k, \quad i = 1, \ldots, d^2,
$$

the solution of (5.16) is given explicitly by the d^2 two-dimensional shrinkages

$$
(5.18) \qquad t_i^{k+1} = \max \left\{ \| q_i^k \|_2 - \frac{1}{\beta_t}, 0 \right\} \frac{q_i^k}{\| q_i^k \|_2}, \quad i = 1, \ldots, d^2.
$$

where $0 \cdot (0/0) = 0$ is assumed. We notice that the computational cost of (5.17)–(5.18) is linear with respect to the number of pixels d^2.

5.2. Solving the sub-problem for r. The minimization sub-problem for r in (5.10) can be rewritten as

$$
r^{k+1} \leftarrow \underset{r \in \mathbb{R}^{d^2}}{\operatorname{argmin}} \left\{ - \langle \lambda_r^k, r - (u_0 - u^k) \rangle + \frac{\beta_r}{2} \| r - (u_0 - u^k) \|_2^2 \right.
$$

$$
\left. + \langle \lambda_v^k, v^k - H(r) \rangle + \frac{\beta_v}{2} \| v^k - H(r) \|_2^2 \right\}
$$

$$
= \underset{r \in \mathbb{R}^{d^2}}{\operatorname{argmin}} \left\{ - \langle \lambda_r^k, r \rangle + \frac{\beta_r}{2} \| r - (u_0 - u^k) \|_2^2 \right.
$$

$$
(5.19) \qquad \left. + \frac{\beta_v}{2} \left\| H(r) - \left(v^k - \frac{1}{\beta_v} \lambda_v^k \right) \right\|_2^2 \right\}.
$$

In order to compute the gradient of the objective function in (5.19), we can exploit the result we obtained for the continuous case in (3.4), where $H(r)$ is the discrete counterpart of

the cumulative distribution function $P_f(f)$, while the constant vector $v^k - \frac{1}{\beta_v}\lambda_v^k$ plays the role of the target cdf $\overline{P}(f)$. Stationary points of (5.19) are thus obtained by

$$-\lambda_r^k + \beta_r(r - u_0 + u^k) + \frac{\beta_v}{d^2}\left(v^k - \frac{1}{\beta_v}\lambda_v^k - H(r)\right) = 0.$$

The gradient vanishes for r-values obtained by a few steps of this simple fixed-point iteration

$$(5.20) \qquad r = \frac{1}{\beta_r}\left(\lambda_r^k + \beta_r(u_0 - u^k) - \frac{\beta_v}{d^2}\left(v^k - \frac{1}{\beta_v}\lambda_v^k - H(r)\right)\right).$$

5.3. Solving the sub-problem for v. The minimization sub-problem for v in (5.11) is

$$v^{k+1} \leftarrow \underset{v \in \mathcal{B}}{\operatorname{argmin}}\left\{-\langle\lambda_v^k, v - H(r^{k+1})\rangle + \frac{\beta_v}{2}\|v - H(r^{k+1})\|_2^2\right\}$$

$$(5.21) \qquad = \underset{v \in \mathcal{B}}{\operatorname{argmin}}\left\{\frac{\beta_v}{2}\left\|v - \left(H(r^{k+1}) + \frac{1}{\beta_v}\lambda_v^k\right)\right\|_2^2\right\}.$$

The solution of (5.21) is thus given by a simple Euclidean projection of the vector $H(r^{k+1}) + \frac{1}{\beta_v}\lambda_v^k$ onto the box-constraints defined by the set $\mathcal{B}$ in (5.6),

$$v^{k+1} = \mathcal{P}_\mathcal{B}\left[H(r^{k+1}) + \frac{1}{\beta_v}\lambda_v^k\right].$$

This projection can be obtained in a straightforward manner by computing the following Q component-wise projections, one for each histogram bin,

$$(5.22) \qquad v_i^{k+1} = \min\left\{\max\left\{\left((H(r^{k+1}))_i + \frac{1}{\beta_v}\lambda_{v,i}^k\right), (\overline{H}_L)_i\right\}, (\overline{H}_U)_i\right\},$$

$$i = 1, \ldots, Q.$$

The computational complexity of this sub-problem is clearly linear in the number of bins Q.

5.4. Solving the sub-problem for u. The minimization sub-problem for u in (5.12) can be rewritten as

$$u^{k+1} \leftarrow \underset{u \in \mathbb{R}^{d^2}}{\operatorname{argmin}}\left\{-\langle\lambda_t^k, t^{k+1} - Du\rangle - \langle\lambda_r^k, r^k - (u_0 - u)\rangle\right.$$

$$\left. + \frac{\beta_t}{2}\|t^{k+1} - Du\|_2^2 + \frac{\beta_r}{2}\|r^k - (u_0 - u)\|_2^2\right\}$$

$$= \underset{u \in \mathbb{R}^{d^2}}{\operatorname{argmin}}\left\{+\langle\lambda_t^k, Du\rangle - \langle\lambda_r^k, u\rangle\right.$$

$$(5.23) \qquad \left. + \frac{\beta_t}{2}\|t^{k+1} - Du\|_2^2 + \frac{\beta_r}{2}\|r^k - (u_0 - u)\|_2^2\right\}.$$

The problem (5.23) is a quadratic optimization problem whose optimality conditions are

$$(5.24) \qquad \left(D^TD + \frac{\beta_r}{\beta_t}I\right)u^{k+1} = D^T\left(t^{k+1} - \frac{1}{\beta_t}\lambda_t^k\right) + \frac{\beta_r}{\beta_t}\left(u_0 - r^k + \frac{1}{\beta_r}\lambda_r^k\right).$$

Under periodic boundary conditions for u, the coefficient matrix $D^TD + \frac{\beta_r}{\beta_t}I$ is block circulant with circulant blocks and thus it is diagonalizable by the 2D discrete Fourier transform (FFT implementation). Therefore, (5.24) can be solved by one forward FFT and one inverse FFT, each at a cost of $O(d^2 \log d)$.

5.5. ADMM iterative scheme. To solve the proposed constrained minimization problem (5.7), we use the ADMM iterative scheme reported in Algorithm 1. In general, choosing good values for the β penalty parameters in ADMM-based optimization is a difficult and sensitive problem. Some criteria for the automatic tuning of these parameters along iterations have been proposed in the literature; see, e.g., [3] and the references therein. In this work, as will be detailed in the next experimental section, we preferred to hand-tune these parameters.

Algorithm 1 ADMM for the proposed distribution constrained problem (5.7).

Input: $u_0, \beta_t > 0, \beta_r > 0, \beta_v > 0$

Output: approximate solution u^* of (5.7)

1. *Initialize*: $u^0 = u_0, r^0 = 0, \lambda_t^0 = 0, \lambda_r^0 = 0, \lambda_v^0 = 0$;

2. **For** $k = 0, 1, 2, \ldots$ *until convergence:*

      1) compute t^{k+1} according to (5.17)–(5.18)

      2) compute r^{k+1} according to (5.20)

      3) compute v^{k+1} according to (5.22)

      4) compute u^{k+1} by solving (5.24)

      7) compute $\lambda_t^{k+1}, \lambda_r^{k+1}, \lambda_v^{k+1}$ by (5.13)

  End For

---

6. Computed examples. In this section we demonstrate the usefulness of the proposed constraint on the distribution of the residual in image denoising by illustrating the performance of the proposed algorithm on both real and synthetic 2D images corrupted by additive zero-mean white noise with known distribution. Additive noise is a good model for the thermal noise within photoelectric sensors and the term "white" noise identifies a noise which is spatially uncorrelated: the noise for each pixel is independent and identically distributed. The additive zero-mean white noise models we considered are the Gaussian noise (AWGN) defined in (4.1) and the uniform noise (AWUN) in a given interval $[-\sigma\sqrt{3}, \sigma\sqrt{3}]$ whose probability density function is given in (4). Even if uniform noise is not often encountered in real-world imaging systems, it provides a useful comparison with Gaussian noise.

We compare the proposed algorithm, referred to as TV-CDF, with the well-known Rudin-Osher-Fatemi (ROF) model [21] based on the minimization of the TV-L_2 functional (2.3). The TV-L_2 approach is implemented by the Alternating Direction Method (ADM), a variant of the classic augmented Lagrangian method for structured optimization which reformulates a TV problem as a linear equality constrained problem. The ADMTV algorithm is stable, efficient, and, in particular, faster than most of the state-of-the-art denoising algorithms. The package for ADMTV-L_2 is freely available[2] and described in detail in [25].

The regularization parameter of the TV-L_2 model is adjusted so that the solution is guaranteed to satisfy the discrepancy principle, that is, the variance of the residual is equal to the variance of the noise. For the proposed TV-CDF model, we used the ADMM minimization procedure illustrated in Section 5 with the following parameters setting $\beta_t = 10$, $\beta_r = 30$,

---

[2] http://www.caam.rice.edu/~optimization/L1/ftvd/v4.0/

TABLE 6.1
RMSE values obtained by the compared denoising algorithms on the images `barbara`, `test`, `skyscraper` *corrupted by zero-mean AWGN with standard deviation σ.*

σ	barbara TV-L_2	barbara TV-CDF	test TV-L_2	test TV-CDF	skyscraper TV-L_2	skyscraper TV-CDF
10	8.09	**7.11**	7.41	**6.82**	8.48	**7.54**
20	12.51	**11.57**	11.00	**10.34**	13.85	**12.72**
30	15.13	**14.56**	13.55	**13.02**	17.80	**16.72**

TABLE 6.2
RMSE values obtained by the compared denoising algorithms on the images `barbara`, `test`, `skyscraper` *corrupted by zero-mean AWUN with standard deviation σ.*

σ	barbara TV-L_2	barbara TV-CDF	test TV-L_2	test TV-CDF	skyscraper TV-L_2	skyscraper TV-CDF
10	8.01	**6.98**	7.28	**6.60**	8.41	**7.43**
20	12.32	**11.27**	10.76	**10.00**	13.73	**12.48**
30	14.93	**14.17**	13.21	**12.56**	17.64	**16.34**

$\beta_v = 1, \gamma = 1$. These parameters have been hand-tuned so as to guarantee fast convergence for all the considered experiments. The iterations of the two algorithms are stopped as soon as the relative difference between two successive iterates satisfies the termination criterion $\|u^k - u^{k-1}\|_2^2 \,/\, \|u^{k-1}\|_2^2 < 10^{-5}$.

The accuracy of the methods is evaluated by the Root Mean Squared Error (RMSE) defined as

$$\mathrm{RMSE}\,(u, \bar{u}) := \sqrt{\|u - \bar{u}\|_2^2 \,/\, d^2}\,,$$

where $u \in \mathbb{R}^{d^2}$ is the computed approximation of the desired noise-free image $\bar{u} \in \mathbb{R}^{d^2}$. This quantity provides a quantitative measure of the quality of the restored image u. A small RMSE value indicates that u is an accurate approximation of $\bar{u}$; we recall, however, that the RMSE values are not always in agreement with visual perception.

We consider the restoration of three different images: `barbara` ($d = 512$), `test` ($d = 256$) and `skyscraper` ($d = 256$) which present interesting mixtures of textures, flat regions, and shaded areas. The noise-free versions of the images are depicted in Figure 6.1(a), Figure 6.2(a), and Figure 6.3(a), respectively.

Tables 6.1 and 6.2 show the quantitative results of the comparison for the two considered noise types, i.e., AWGN and AWUN, respectively. In particular, the tables report the RMSE values obtained by applying ROF (TV-L_2) and our (TV-CDF) algorithm on the test images corrupted by noise with different standard deviations $\sigma = 10, 20, 30$. The bold numbers in the tables indicate the better (lower) RMSE values obtained between the two methods.

The results in Tables 6.1 and 6.2 demonstrate that the proposed TV-CDF model outperforms the TV-L_2 approach in terms of obtained RMSE values on the three test images for all the considered noise levels.

The quantitative results shown in Tables 6.1 and 6.2 are validated by the visual inspection of the image restorations illustrated in Figures 6.1–6.3 for the AWGN corruption and in Figure 6.4 for the AWUN corruption.

The results for the first example, i.e., images `barbara`, `test`, and `skyscraper` corrupted by zero-mean AWGN with standard deviation $\sigma = 10$ are illustrated in Figures 6.1–6.3. In particular, the noise-free images are depicted in Figures 6.1–6.3(a), the noise-corrupted images are shown in Figures 6.1–6.3(b), the images restored by TV-L_2 and by TV-CDF are given in Figures 6.1–6.3(c) and Figures 6.1–6.3(d), respectively. In the last rows of Figures 6.1–6.3, we show the normalized histograms of the residuals associated with the restored images depicted in Figures 6.1–6.3(c) and Figures 6.1–6.3(d).

The effect of the proposed constraint is evident from the residual histograms. In fact, the residual histograms obtained by the proposed TV-CDF method and displayed in Figures 6.1–6.3(f) adhere more faithfully to the target Gaussian probability density function (red curve) than those reported in Figures 6.1–6.3(e) and computed by the TV-L_2 model.

The results for the second example, i.e., images `barbara`, `test`, and `skyscraper` corrupted by zero-mean AWUN with standard deviation $\sigma = 30$, are illustrated in Figure 6.4. In particular, in Figure 6.4(a) we show the noise-corrupted images while in Figures 6.4(b)–(e) we show zoomed details of the noisy images, of the original noise-free images, of the images restored by TV-L_2, and of the images restored by TV-CDF, respectively. This experiment demonstrates that the proposed constrained denoising approach is effective also for non-Gaussian noise models.

The restoration results illustrated in the examples obtained by the TV-CDF method together with the results in Table 6.1 and Table 6.2 allow us to conclude that by exploiting the information present in the components of the residual vectors, such as in this case the noise distribution, we can go beyond the use of just the residual norm or the variance of the noise model, thus providing better restorations.

A proof of convergence is beyond the scope of this paper and will be investigated in future work. However, in Figure 6.5 we provide empirical evidence of the numerical convergence of the proposed algorithm for the example illustrated in Figure 6.1. In particular, in Figure 6.5 the relative change of the approximate solution computed as $\|u^k - u^{k-1}\|_2 / \|u^{k-1}\|_2$ is shown as a function of the iteration count k. The same convergent behavior can be observed for all the other tested examples.

7. Conclusions. In this paper we have proposed a hard-constrained variational model aimed at denoising images corrupted by identically distributed additive noise with known probability density function. In particular, our model is obtained by enforcing explicitly the constraints on the similarity between the residual distribution and the target noise distribution. The classical TV-seminorm term is considered but any other regularizer could be substituted as well. The constrained variational problem is efficiently solved by using the popular ADMM procedure suitably adapted to our particular model. We have compared our model with the well-known ROF model by applying them to a few image restoration examples. The experiments show that the TV-CDF based method outperforms the TV-L_2 based method for preserving better contrast and more details in the denoising process. Robust preliminary estimation of the noise distribution, i.e., of the target residual histogram to be enforced within our approach, will be a matter of future work.

(a) noise-free image
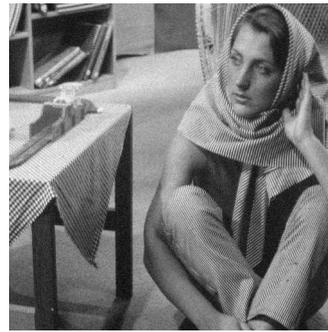
(b) noisy image

(c) TV-L_2: restored image

(d) TV-CDF: restored image
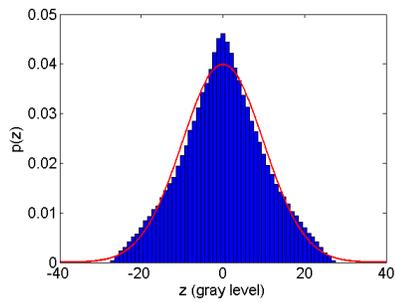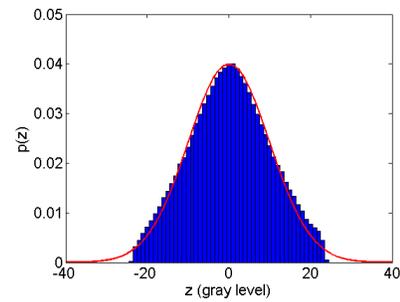
(e) TV-L_2: residual pdf

(f) TV-CDF: residual pdf

FIG. 6.1. *Restoration results on* barbara *image corrupted by zero-mean AWGN with standard deviation* $\sigma = 10$.
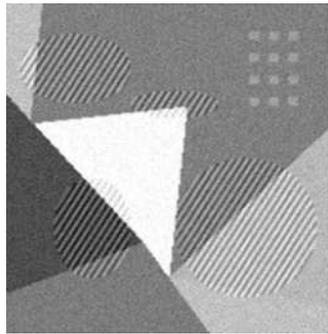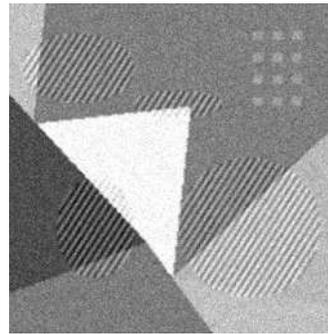
(a) noise-free image
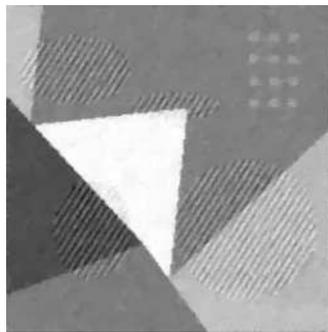
(b) noisy image

(c) TV-L_2: restored image
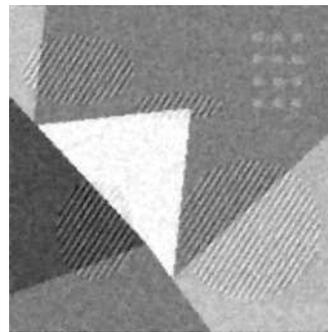
(d) TV-CDF: restored image
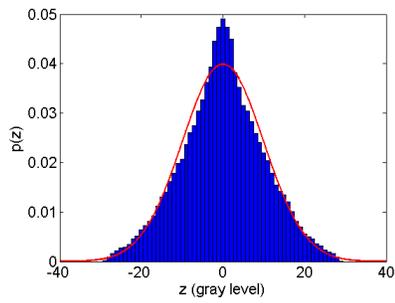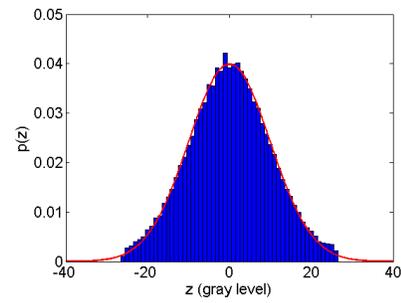
(e) TV-L_2: residual pdf

(f) TV-CDF: residual pdf

FIG. 6.2. *Restoration results on* test *image corrupted by zero-mean AWGN with standard deviation* $\sigma = 10$.

(a) noise-free image

(b) noisy image

(c) TV-L_2: restored image

(d) TV-CDF: restored image
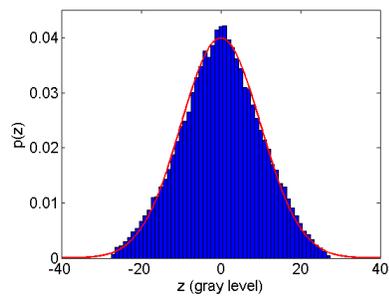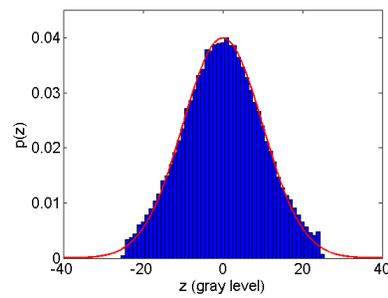
(e) TV-L_2: residual pdf

(f) TV-CDF: residual pdf

FIG. 6.3. *Restoration results on* skyscraper *image corrupted by zero-mean AWGN with standard deviation* $\sigma = 10$.

(a) noisy images

(b) noisy images (zoomed details)

(c) original images (zoomed details)

(d) images restored by TV-L_2 (zoomed details)

(e) images restored by TV-CDF (zoomed details)
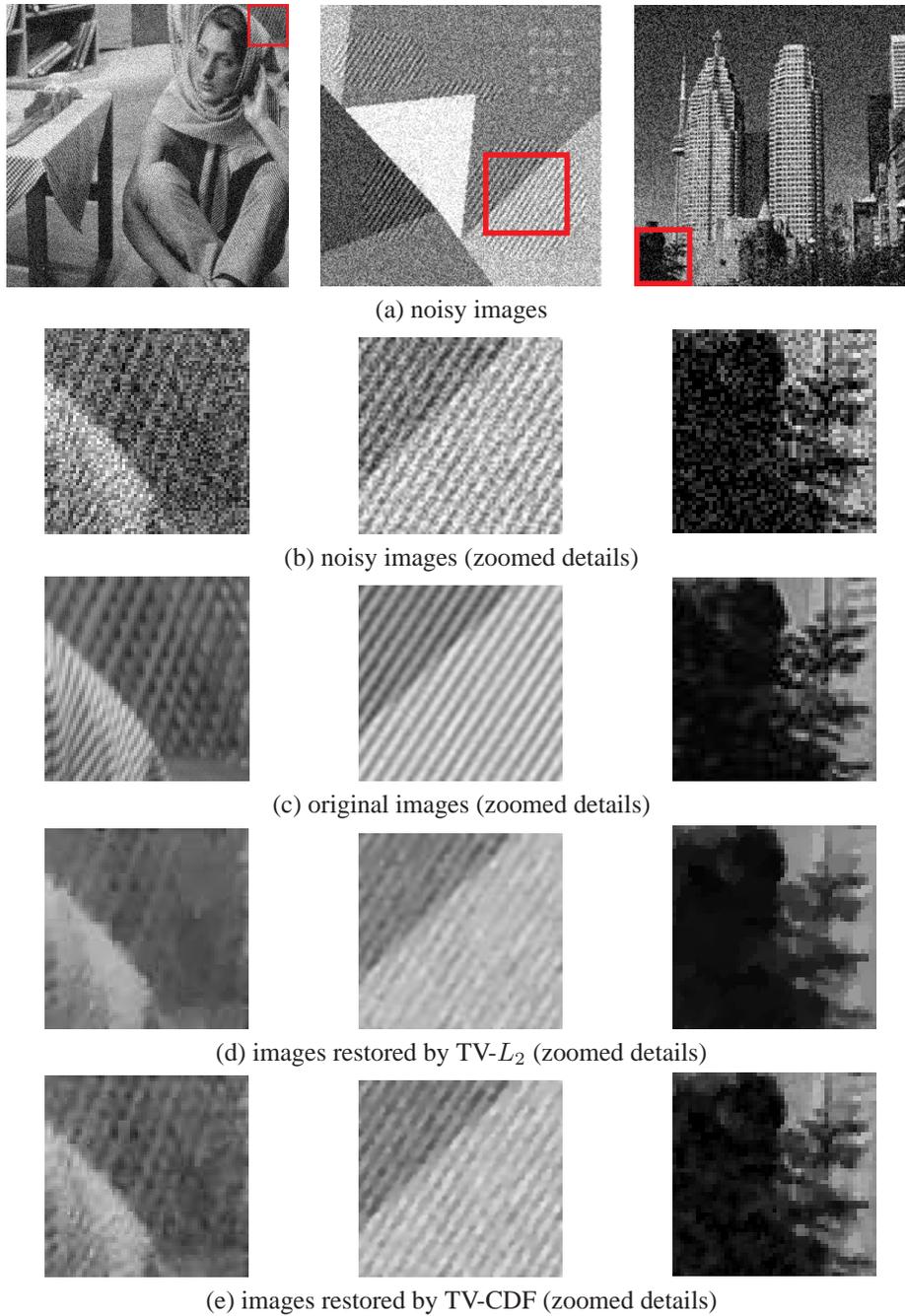
FIG. 6.4. *Restoration results on* barbara, test *and* skyscraper *images corrupted by zero-mean AWUN with standard deviation $\sigma = 30$.*
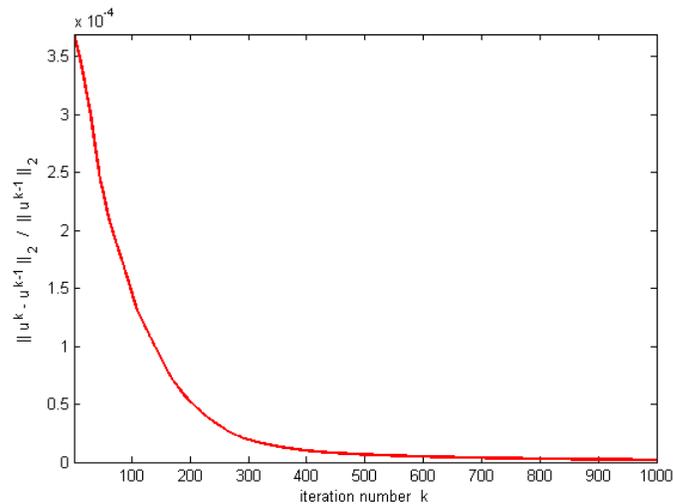
FIG. 6.5. *Plot of the relative change of the computed approximate solution as a function of iterations count for the example illustrated in Figure 6.1.*

REFERENCES

[1] J. BAI AND X. C. FENG, *Fractional-order anisotropic diffusion for image denoising*, IEEE Trans. Image Process., 16 (2007), pp. 2492–2502.

[2] F. BAUER AND M. A. LUKAS, *Comparing parameter choice methods for regularization of ill posed problems*, Math. Comput. Simulation, 81 (2011), pp. 1795–1841.

[3] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.

[4] A. BUADES, B. COLL, AND J. M. MOREL, *A review of image denoising algorithms, with a new one*, Multiscale Model. Sim., 4 (2005), pp. 490–530.

[5] D. CALVETTI, S. MORIGI, L. REICHEL, AND F. SGALLARI, *An L-ribbon for large underdetermined linear discrete ill-posed problems*, Numer. Algorithms, 25 (2000), pp. 89–107.

[6] V. CASELLES, R. KIMMEL, AND G. SAPIRO, *Geodesic Active Contours*, Int. J. Comput. Vis., 22 (1997), pp. 61–79.

[7] A. CHAMBOLLE AND P. L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.

[8] R. CHAN, M. NIKOLOVA, AND Y. W. WEN, *A variational approach for exact histogram specification*, in Scale Space and Variational Methods in Computer Vision, A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, M. M. Bronstein, eds., Lecture Notes in Computer Science, 6667, Springer, Heidelberg, 2012, pp. 86–97.

[9] R. CHAN, M. TAO, AND X. M. YUAN, *Constrained total variational deblurring models and fast algorithms based on alternating direction method of multipliers*, SIAM J. Imaging. Sci., 6 (2013), pp. 680–697.

[10] I. DAUBECHIES AND G. TESCHKE, *Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising*, Appl. Comput. Harmonic Anal., 19 (2005), pp. 1–16.

[11] S. DIDAS, J. WEICKERT, AND B. BURGETH, *Properties of higher order nonlinear diffusion filtering*, J. Math. Imaging Vision, 35 (2009), pp. 208–226.

[12] R. GLOWINSKI, *Numerical Methods for Nonlinear Variational Problems*, Springer, New York, 1984.

[13] L. JIANG, X. C. FENG, AND H. Q. YIN, *Variational Image Restoration and Decomposition with Curvelet Shrinkage*, J. Math. Imaging Vision, 30 (2008), pp. 125–132.

[14] A. LANZA, S. MORIGI, F. SGALLARI, AND A. YEZZI, *Variational image denoising based on auto-correlation whiteness*, SIAM J. Imaging. Sci., 4 (2013), pp. 1931–1955.

[15] S. MORIGI, R. PLEMMONS, L. REICHEL, AND F. SGALLARI, *A hybrid multilevel-active set method for large box-constrained discrete ill-posed problems*, Calcolo, 48 (2011), pp. 89–105.

[16] S. MORIGI, L. REICHEL, AND F. SGALLARI, *A truncated projected SVD method for linear discrete ill-posed problems*, Numer. Algorithms, 43 (2006), pp. 197–213.

[17] S. NARASIMHAN AND S. NAYAR, *Contrast restoration of weather degraded images*, IEEE Trans. Pattern

84 A. LANZA, S. MORIGI, F. SGALLARI, AND A. J. YEZZI

Anal. Mach. Intell., 25 (2003), pp. 713–724.

[18] M. NIKOLOVA, M. K. NG, AND C. P. TAM, *Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction*, IEEE Trans. Image Process., 19 (2010), pp. 3073–3088.

[19] P. PERONA AND J. MALIK, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 629–639.

[20] L. REICHEL, F. SGALLARI, AND Q. YE, *Tikhonov regularization based on generalized Krylov subspace methods*, Appl. Numer. Math., 62 (2012), pp. 1215-1228.

[21] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[22] B. W. RUST AND D. P. O'LEARY, *Residual periodograms for choosing regularization parameters for ill-posed problems*, Inverse Problems, 24 (2008), 034005 (30 pages).

[23] G. SAPIRO AND V. CASELLES, *Histogram modification via partial differential equations*, in Proc. IEEE International Conference on Image Processing vol. 3, IEEE Conference Proceedings, Los Alamitos, CA, 1995, pp. 632–635.

[24] P. SWOBODA AND C. SCHNORR, *Convex variational image restoration with histogram priors*, SIAM J. Imaging Sci., 6 (2013), pp. 1719–1735.

[25] M. TAO AND J. YANG, *Alternating direction algorithms for total variation deconvolution in image reconstruction*, Tech. Rep. TR0918, Department of Mathematics, Nanjing University, China, 2009.

[26] L. A. VESE AND S. J. OSHER, *Modeling textures with total variation minimization and oscillatory patterns in image processing*, J. Sci. Comput., 19 (2003), pp. 553–572.

[27] M. WELK, D. THEIS, T. BROX, AND J. WEICKERT, *PDE-based deconvolution with forward-backward diffusivities and diffusion tensors*, in Scale-Space and PDE Methods in Computer Vision, R. Kimmel, N. Sochen, J. Weickert, eds., Lecture Notes in Computer Science, 3459, Springer, Berlin, 2005, pp. 589–597.