Image Restoration with Faithful Residuals

Håvard Rue*
Department of Mathematical Sciences
NTNU, Norway

Abstract

It is common in image restoration to obtain residuals that does not agree with our knowledge about the noise process, even if we accept a small deviation due to the bias in the restored image. We discuss how this information can be further used, leading to more faithful residuals. Based on the regularization approach, we introduce penalty terms so the sample covariance of the residuals match the covariance structure in the noise. This is tested on artificial noisy images with and without blur. The simulation results shows sharper edges and a better restoration in difficult areas. We achieve best results with a weak model for the local spatial smoothness. Actually, a spatially smooth restoration can be obtained without any model for the local spatial smoothness of the true image.

1 Introduction

The problem of image restoration is one of many inverse problems that occurs in imaging. The most popular methods for trying to solve such ill–posed or ill–conditioned problems, are by using the regularization approach (Hunt, 1973; Burch, Gull and Skilling, 1983; Skilling and Gull, 1985) or the Bayesian approach (Geman and Geman, 1984; Besag, 1986). Both approaches typically try to model the local spatial smoothness of the image and the restored image is a smoothed version of the least squares estimate in some sense. This paper discusses the residuals that these methods produces and how the information about the noise process could be further used within the regularization framework, leading to more faithful residuals.

Let the observed degraded image be represented by the vector $\mathbf{y}$ of dimension N. Corresponding to $\mathbf{y}$ there is a true image represented by the vector $\mathbf{x}$ of dimension N. The true image is the measured image under ideal conditions. Depending on the application and the goal for the restoration, the pixel values could be discrete or real. In the discrete case, the finite state

*Address for correspondence: Department of Mathematical Sciences, Norwegian University of Science and Technology, N–7034 Trondheim, Norway.
E–mail: havard.rue@imf.unit.no
This report was written in 1992, so I need to be excused for the old references. It is not published apart from a version that appears in my ph.d thesis (Rue 1993). The basic idea is nice, but lacks of theoretical justification. I still feel it's something more general in there, i just can't figure out!!!

space is typically $0 \ldots L-1$, where L is the number of different colors. If the pixel values are real numbers, they typically represent a positive quantity. We assume a linear degradation model with additive noise,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon. \tag{1}$$

The matrix $\mathbf{H}$ represent the linear distortion. Usually, this distortion is spatial invariant and described by the point spread function h as $H_{ij} = h_{i-j}$. In this case, $\mathbf{H}$ is called the blur matrix and we have a deblurring problem. The noise term ϵ is assumed throughout to be independent of $\mathbf{x}$ and to be normally distributed with zero mean and known diagonal covariance matrix, $\sigma^2\mathbf{I}$. The problem of *restoration* is to infer $\mathbf{x}$ from $\mathbf{y}$.

The outline of the report is as follows: In Section 2, we discuss restoration with regularization methods using the minimum penalized distance approach. Models for local spatial smoothness used in the experiments, will also be presented. The bias in the residuals that the methods produces and how the information could be further used, is discussed in Section 3. In Section 4, we present the algorithms used to calculate the restoration. Section 5 discuss the selection of the parameters. Finally, Section 6 present the numerical experiments.

2 Restoration by Regularization Methods

This section describes restoration methods based on regularization and models for smoothness.

2.1 The Minimum Penalized Distance Approach

The regularization approach can be formulated using the Minimum Penalized Distance (MPD) approach of Titterington (1985a) and Titterington (1985b). The goal is to define a general recipe or prescription, $\hat{\mathbf{x}}(\mathbf{y}, \alpha)$, that for a given value of the non–negative parameter α and the data $\mathbf{y}$ provide us with a restored image. As we let α tends to zero, $\hat{\mathbf{x}}$ tends to the least squares estimate $\mathbf{x}_0 = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$. The least squares estimate gives best possible fit to the data in the least squares sense. It is unbiased, but can have high covariance if the problem is ill–conditioned. As we let α tends to infinity, $\hat{\mathbf{x}}$ tends to a "ultra–smooth" estimate $\mathbf{x}_\infty$. This estimate has low or zero variance but high bias. For a finite value of α, the estimate will be somewhere between the limiting estimates $\mathbf{x}_\infty$ and $\mathbf{x}_0$. The parameter α determines the tradeoff between bias and variance and is often called the smoothing parameter. The two distance measures Δ_1 and Δ_2 formalizes the approach, as they define the distance from any proposal $\mathbf{x}'$ to $\mathbf{x}_0$ and $\mathbf{x}_\infty$. Each distance measure takes two possible images as arguments and is not a distance measure in the usual sense, e.g. it may not be symmetric in the arguments. It will however be true that $\Delta(\mathbf{x}', \mathbf{x}'') > 0$ if $\mathbf{x}'$ and $\mathbf{x}''$ differ and $\Delta(\mathbf{x}', \mathbf{x}') = 0$ for any $\mathbf{x}'$. We find an estimate of $\mathbf{x}$ that fits the data when the distance Δ_1 from $\mathbf{x}$ to $\mathbf{x}_0$ is zero. Fidelity with ultra–smooth $\mathbf{x}_\infty$, is found when the distance Δ_2 between $\mathbf{x}$ and $\mathbf{x}_\infty$ is zero. A compromise estimate that try to achieve smoothness and still be faithful with the data, is found by minimizing the total distance

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left(\Delta_1(\mathbf{x}, \mathbf{x}_0) + \alpha\Delta_2(\mathbf{x}, \mathbf{x}_\infty)\right). \tag{2}$$

This estimate is called the minimum penalized distance (MPD) estimator.

It is natural to use our knowledge about the statistical properties of the noise in our degradation model, in order to determine the distance measure Δ_1. Using that $\mathbf{x}_0$ is an unbiased estimate with covariance $(\mathbf{H}^T\mathbf{H})^{-1}\sigma^2$, a natural distance measure is

$$\Delta_1(\mathbf{x}, \mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x})^T \mathrm{Cov}^{-1}(\mathbf{x}_0)(\mathbf{x}_0 - \mathbf{x}) = \frac{1}{2\sigma^2}\mathbf{e}^T\mathbf{e}, \tag{3}$$

where $\mathbf{e}$ is the error vector defined for any vector $\mathbf{x}$ as

$$\mathbf{e} = \mathbf{y} - \mathbf{H}\mathbf{x}. \tag{4}$$

Note that $\Delta_1(\mathbf{x}, \mathbf{x}_0)$ is the minus log likelihood for $\mathbf{y}$ given $\mathbf{x}$.

Let $U(\mathbf{x})$ be a function that defines smoothness properties. The value of $U(\mathbf{x})$ is decreasing for more smoother versions of $\mathbf{x}$ and has a minimum value for the "ultra–smooth" $\mathbf{x}_\infty$. We can then define the distance $\Delta_2(\mathbf{x}, \mathbf{x}_\infty)$ as

$$\Delta_2(\mathbf{x}, \mathbf{x}_\infty) = U(\mathbf{x}) - U(\mathbf{x}_\infty). \tag{5}$$

With these distance measures, the MPD estimate becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left(\alpha U(\mathbf{x}) + \frac{1}{2\sigma^2}\mathbf{e}^T\mathbf{e}\right), \tag{6}$$

which is the common form. This MPD estimator has also a Bayesian interpretation as the maximum *a posteriori* (MAP) estimator, where the prior distribution for $\mathbf{x}$ is $\pi(\mathbf{x}) \propto \exp(-U(\mathbf{x}))$, see Geman and Geman (1984) or Besag (1986) for details.

2.2 Models for Smoothness

This part describes the different models for smoothness used in this study. The first model has a discrete state space for $\mathbf{x}$ and is common in the Bayesian framework, i.e. as a prior distribution. The second model has a more natural interpretation as a distance measure and treats $\mathbf{x}$ as a continuous variable.

2.2.1 The Geman and McClure model

The Geman and McClure (1987) model for the local smoothness are usually presented within the Bayesian framework as a Markov Random Field (MRF) prior model. We will however omit the MRF interpretation in this context.

Geman and McClure (1987) modeled the function $U(\mathbf{x})$ that defines the smoothness properties a sum of local contributions from each pixel and its eight nearest neighbors,

$$U(\mathbf{x}) = \sum_{\langle i,j \rangle} \Upsilon(x_i - x_j) + \frac{1}{\sqrt{2}} \sum_{[i,j]} \Upsilon(x_i - x_j). \tag{7}$$

The sum $\langle i,j \rangle$ are over all nearest neighbors and the sum $[i,j]$ are over all nearest diagonal neighbors. Note that $U(\mathbf{x})$ is a function of the difference between neighbor pixels only. The function Υ is given as

$$\Upsilon(\kappa) = \frac{-1}{1+(\kappa/\delta)^2}, \tag{8}$$

and δ is a scale parameter for the pixel differences. This model favor nearby pixels to be equal but also incorporate the edges in an implicit way; The derivative of Υ tends to zero as $|\kappa/\delta|$ increase and hence U hardly change with $|\kappa|$ for large $|\kappa|$.

2.2.2 Kullback–Leibler Directed Divergence

If all elements in the true image $\mathbf{x}$ are positive by definition, we can introduce a normalized image which is independent of the measurement units, $f_i = x_i/\sum_j x_j$. This leads to the interpretation of $\mathbf{f}$ as a probability density and Δ_2 could be a distance measure between probability densities. One such measure is the Kullback–Leibler directed divergence,

$$\Delta_2(\mathbf{x}, \mathbf{x}_\infty) = \Delta_2(\mathbf{x}, \mathbf{m}) = \sum_i f_i \ln \frac{f_i}{m_c}, \tag{9}$$

and measures the distance between $\mathbf{f}$ and a given reference $m_i = m_c$ for all i. The Kullback–Leibler distance measure is also known as the (minus) Shannon entropy and have been used in several types of problems known as maximum entropy methods (Skilling and Gull, 1985).

It is important to notice that the Kullback–Leibler distance measure does not model any local spatial smoothness of $\mathbf{f}$, as the ordering of $\mathbf{f}$ does not change the value of Δ_2. It does however model global smoothness, as if we let f_i and f_j tends to each other, the distance measure decrease.

3 The Residuals

It is of importance to use the information available about the noise process in the restoration. Unless the residuals follow (approximately) the noise distribution, they contain information which could be used to improve the restoration. We will discuss the reason for biased residuals and how we could make further use of this information.

3.1 Biased Residuals

We define the residuals $\mathbf{g}$ for any restored image $\hat{\mathbf{x}}$, as

$$\mathbf{g} = (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}})/\sigma. \tag{10}$$

The residuals are the noise estimated using the restored image. We should always check the residuals against the noise distribution for a given restoration. If the restoration $\hat{\mathbf{x}}$ happens to be the true image $\mathbf{x}$, the residuals will be equal to the observed sample realization from the noise distribution. The residuals achieved from $\hat{\mathbf{x}}$ does often not look like such a random sample. When we smooth the unbiased estimate $\mathbf{x}_0$ towards $\mathbf{x}_\infty$ to reduce the variance, we are introducing some bias unless the true image equals $\hat{\mathbf{x}}$. This is because we are smoothing against $\mathbf{x}_\infty$ and not the true image $\mathbf{x}$. The residuals are biased in subdomains, typically because the smoothing term $U(\mathbf{x})$ smooth off the humps and troughs are filled in. Quite frequently different patterns appear in the residuals, typically in regions with edges or discontinuities. By studying the residuals, we can often recognize some of the edges or discontinuities. The bias in the residuals will depend locally on the true image $\mathbf{x}$. These effects are more marked with a higher degree of smoothing and lower noise level. Figure (1) shows a simple example. In the neighborhood of the discontinuities, the residuals are clearly biased and there are systematic errors in the restored image.

Our goal is to achieve faithful residuals in the restoration. To do this in general, we will modify the restoration method. We allow the restoration to smooth and thereby accept a small amount of bias. It is the highly biased residuals and the systematic structure we want to avoid. We feel that the most critical property of the residuals apart from the range, is the local independence. In the case of no blur, there is a one–to–one connection between errors in the residuals and the restored image. Let ϵ_i be the value of the true noise in y_i. With no blur, we have

$$g_i - \epsilon_i = -(\hat{x}_i - x_i)/\sigma. \tag{11}$$

Bias in $\hat{x}_i$ only effect g_i, refer also to Figure (1). Residuals that are clearly dependent (e.g. high correlation) in a local region, indicate that we have systematic errors in the restored image. When blur is present, there is no such one–to–one connection. However, locally dependent residuals still indicate systematic errors in this region.

As an first approximation, we will restrict our self to consider only the linear dependency in the residuals by studying the correlation function. To control the sample moments of the residuals up to second order, we introduce a vector of K control variables or control statistics,

$$\mathbf{z} = (z_1, \ldots, z_K), \qquad z_k = z_k(\mathbf{g}). \tag{12}$$

Following the MPD approach, we must construct a recipe $\hat{\mathbf{x}}(\mathbf{y}, \alpha, \beta)$ that also incorporate our knowledge about the statistical properties about $\mathbf{z}$. The parameter β controls the weight we give to the $\mathbf{z}$ part. To handle local variations in the residuals, we allow each z_k to be a local function or a function defined on a subdomain. The first order moment of the residuals (the mean) is for most cases satisfactory. Myrheim and Rue (1992) discussed this in more detail. In this report we will focus on the second order moments (the covariance) structure in the residuals. We will use the correlation function of the residuals as the control variables. Since the noise has diagonal covariance matrix, our aim is near uncorrelated residuals. If the residuals are faithful, then $\hat{\mathbf{x}}$ is still a smoothed version of $\mathbf{x}_0$ but in a way that allow us to rely more on the restoration.

3.2 Restoration with Control Variables

As the control variables $\mathbf{z}$, we use the correlation function of the residuals. Let K be the dimension of $\mathbf{z}$. The different lags are refereed to as l_k, so l_k is lag number k. For later use we define l_0 as zero. In one dimension, l_k is a lag (e.g. (1)). In two dimensions, l_k is a two–dimensional vector, $l_k = (i, j)$. The first element i is the lag in horizontal direction. The second element j is the lag in vertical direction, e.g. $(2, 1)$. In this case the index k is a numbering of two–dimensional lags. Figure (2) shows an example of this numbering in a 7×7 square around $l_0 = (0, 0)$. Lag number 6 is there the lag $(2, 1)$. Without assuming any symmetric property about the correlation structure (e.g. rotational symmetry), the number of lags is 24.

The k'th component of the vector $\mathbf{z}$, z_k is proportional to the sample correlation function of $\mathbf{g}$ at lag l_k in subdomain Ω_k. The diagonal matrix $\mathbf{D}_k$ defines Ω_k with 1's at the diagonal in the subdomain and zero elsewhere,

$$\Omega_k = \cup_i \left((\mathbf{D}_k)_{ii} = 1 \right) \tag{13}$$

For later use define Ω_0 as the whole image, $\mathbf{D}_0 = \mathbf{I}_N$ (the identity matrix with dimension $N \times N$).

Since the mean of the residuals is assumed to be zero or near zero, we do not correct for the mean in the autocovariance. We define the autocovariance at lag l_k in subdomain k, $R_{l_k,k}$ using the shift matrix $\mathbf{S}_{l_k}$ as

$$R_{l_k,k} = \mathbf{e}^T \mathbf{D}_k \mathbf{S}_{l_k} \mathbf{D}_k \mathbf{e}. \tag{14}$$

The shift matrix shifts a vector according to lag l_k. In the one dimensional case, the shift matrix $\mathbf{S}_{l_k}$ is of Toeplitz form,

$$(\mathbf{S}_{l_k})_{ij} = \begin{cases} \frac{1}{2} & \text{if } |i - j| = l_k \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

In the two dimensional case, the matrix $\mathbf{S}_{l_k}$ is a block Toeplitz where each block is of Toeplitz form. For lag zero, the shift matrix equals the identity matrix, $\mathbf{S}_0 = \mathbf{I}_N$. The correlation function $r_{l_k,k}$ is the ratio between the autocovariance function at lag l_k and lag zero (l_0) in Ω_k, $r_{l_k,k} = R_{l_k,k}/R_{0,k}$. If $R_{0,k}$ is zero we define $r_{l_k,k}$ as zero. Let the k'th component of $\mathbf{z}$ be

$$z_k = \sqrt{N_k} r_{l_k,k} \tag{16}$$

where N_k are the number of observed data in Ω_k. We will in the experiments define a set of lags in the same subdomain, meaning that some of the subdomains are equal.

In the MPD approach, we include the control variables in the distance measure Δ_1 between $\mathbf{x}$ and $\mathbf{x}_0$. The asymptotic distribution for the correlation function of one–dimensional white noise is Gaussian distributed with zero mean and covariance $\mathbf{I}_N$ (Brockwell and Davis, 1987). A two–dimensional image with dimension $N_1 \times N_2$ can be written as a vector with length $N = N_1 N_2$. Apart from the border region, a two–dimensional lag can be written as a one–dimensional lag. For example, if we store the image row–by–row, then lag $(1, 2)$ corresponds to the one–dimensional lag $(1 + 2N_1)$. The natural distance measures for $\mathbf{z}$ is then,

$$\begin{aligned} \Delta_1^z(\mathbf{x}, \mathbf{x}_0) &= \frac{1}{2} \sum_{k=1}^{K} N_k \left(\frac{(\mathbf{x}_0 - \mathbf{x})^T \mathbf{H}^T \mathbf{D}_k \mathbf{S}_{l_k} \mathbf{D}_k \mathbf{H} (\mathbf{x}_0 - \mathbf{x})}{(\mathbf{x}_0 - \mathbf{x})^T \mathbf{H}^T \mathbf{D}_k \mathbf{H} (\mathbf{x}_0 - \mathbf{x}) + \varsigma} \right)^2 \\ &= \frac{1}{2} \sum_{k=1}^{K} z_k^2. \end{aligned} \tag{17}$$

Here, ς is an infinite small positive value so $\Delta_1^z(\mathbf{x}_0, \mathbf{x}_0) = 0$. We write the modified distance measure between $\mathbf{x}$ and $\mathbf{x}_0$, as a weighted sum of Δ_1 and Δ_1^z. The MPD estimate is then found by minimizing the total distance $Y(\mathbf{x})$,

$$\begin{aligned}
\hat{\mathbf{x}}(\mathbf{y}, \alpha, \beta) &= \arg\min_{\mathbf{x}} \left(\alpha\Delta_2(\mathbf{x}, \mathbf{x}_\infty) + \Delta_1(\mathbf{x}, \mathbf{x}_0) + \beta\Delta_1^z(\mathbf{x}, \mathbf{x}_0)\right) \\
&= \arg\min_{\mathbf{x}} \left(\alpha U(\mathbf{x}) + \frac{1}{2\sigma^2}\mathbf{e}^T\mathbf{e} + \frac{1}{2}\beta\sum_k z_k^2\right) \\
&= \arg\min_{\mathbf{x}} Y(\mathbf{x}). \qquad (18)
\end{aligned}$$

The parameter β is non–negative. An increasing value of β will put more weight to the $\mathbf{z}$–term.

4 Algorithms

The computation of the exact MPD estimate is a hard problem. Local minima in $Y(\mathbf{x})$ may occur because $Y(\mathbf{x})$ is a complicated function. The finite state space in the MRF model requires different algorithms than the continuous case. We use Monte Carlo type algorithms in the MRF model and an iterative conjugate gradient algorithm in the continuous Kullback–Leibler model to estimate the mode.

4.1 Monte Carlo Type Algorithms

To use Monte Carlo type algorithms to obtain the MPD estimate, we need to interpret $\mathbf{v}$ as a random variable with Gibbs distribution

$$\pi(\mathbf{v}) \propto \exp(-Y(\mathbf{v})). \qquad (19)$$

The total distance $Y(\mathbf{v})$ is then the energy function and defines $\mathbf{v}$ as a MRF where the neighborhood system depends on Ω_k and $\mathbf{H}$. The MPD estimate is now the mode in $\pi(\mathbf{v})$.

The basic operation in this type of algorithms, is to use a sampling algorithm to generate sample realizations from $\pi(\mathbf{v})$. The sampling algorithm generates a Markov chain with transition probabilities such that the equilibrium distribution is the required distribution. One popular sampling algorithm is the Gibbs Sampler (Geman and Geman, 1984); Choose a sequence of pixels so that each pixel is visited infinity often. At each pixel s in the sequence, replace v_s with a sample from the conditional distribution,

$$\pi(v_s|\mathbf{v}_{\setminus\{s\}}). \qquad (20)$$

There exists many such chains that has the required equilibrium distribution and any of them will do (Geman, 1990). The MPD estimate or the mode can then be found by annealing. To simulate the annealing process, an artificial temperature (or control parameter) $T(t)$ is introduced into the Gibbs distribution

$$\pi_{T(t)}(\mathbf{v}) \propto \pi(\mathbf{v})^{1/T(t)}, \qquad (21)$$

and the temperature change while we do sampling. If the temperature tends to zero sufficient slowly, $T(t) \geq C/(1+\ln t)$, for a small constant C, the nonstationary Markov chain will converge weakly to a distribution supported on the minima of $Y(\mathbf{v})$ (Geman and Geman, 1984). This schedule is highly unrealistic for real problems and some faster schedule is usually used, but then the minimum is no longer guaranteed. In practice we could use algorithms that compromise between the theoretical and the practical, such as low-temperature sampling ($T = \epsilon$) (Geman, Geman, Graffigne and Dong, 1990) or Iterated Conditional Modes (ICM) (Besag, 1986) which correspond to instantaneous freezing ($T = 0$).

4.2 An Iterative Numerical Algorithm

This section describes the algorithm used for computing the MPD estimate for the Kullback–Leibler distance measure. The presented algorithm is a generalization of the conjugate gradient based algorithm suggested by Myrheim and Rue (1992). We will therefore present the algorithm only briefly. The problem is to find $\mathbf{x}$ that minimize $Y(\mathbf{x})$ together with the positivity constraint. The minimization problem is nonlinear so we have to use an iterative algorithm. Let $\mathbf{u}$ be the search direction. The new estimate $\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{old}} + \xi\mathbf{u}$ is found such that ξ minimize $Y(\mathbf{x}_{\text{old}} + \xi\mathbf{u})$,

$$\mathbf{u}^T \nabla Y(\mathbf{x}_{\text{old}} + \xi\mathbf{u}) = 0, \tag{22}$$

where $\nabla = (\ldots, \frac{\partial}{\partial x_i}, \ldots)$. To find the search directions, we first compute the scaled gradient $\mathbf{r}_J = (\text{diag}\mathbf{A})^{-1}\nabla Y(\mathbf{x}_{\text{old}})$, where $\mathbf{A}$ is the symmetric hessian matrix $\nabla^2 Y(\mathbf{x}_{\text{old}})$. As our new search direction $\mathbf{u}$, we use the part of $\mathbf{r}_J$ that is $\mathbf{A}$–orthogonal to the previous search direction. The positivity constraints are controlled by rescaling critical components in $\mathbf{u}$, see Myrheim and Rue (1992) for details. If the matrix $\mathbf{H}$ is spatial invariant, we can use Fast Fourier Transform to calculate some of the matrix–vector products needed. The computational cost will then be independent of the size of the point spread function.

5 Selection of Parameters

It is of importance to have good values of the parameters α and β. There exist several strategies to estimate the smoothing parameter α. A numerical comparison of different strategies is given in Thompson, Brown, Kay and Titterington (1991). The classical χ^2–approach is based on fit to the data. The restored image should fit the data for the given degradation model and noise level (Hunt, 1973; Burch et al., 1983), the residual variance should be equal to the noise variance $R_{0,0}/N_0 = \sigma^2$. The experience with this, is that it smoothes a little too much. The χ^2–approach is however well–known and defines a certain level of smoothing. At this level the results obtained with and without the $\mathbf{z}$–term is comparable. We will therefore use this approach in the following experiments. In the Monte Carlo algorithms we estimate α by just trial and failure until the residual variance match the noise variance. In the conjugate gradient algorithm we update α in the same way as in Myrheim and Rue (1992).

We can use the χ^2–approach to estimate β as well, by require that also $\mathbf{z}$ should fit the data, $\sum_{k=1}^{K} z_k^2 = K$. In the algorithms we do this simply by increasing β until this is satisfied.

With these choices of the parameters α and β, we are trying to match the residual sample covariance to the covariance for the true noise.

There are two basic strategies to select the subdomains. The first one is most intuitive; Select the subdomains Ω_k on the basis of the residual pattern achieved with the ordinary method without the $\mathbf{z}$–term, so the subregions cover regions of unsatisfactory residuals. Using the iterative numerical algorithm this approach is preferred as it lead to few subregions. Using the Monte Carlo algorithm it can be preferred to use another strategy; If a pixel s is within one of the subdomains Ω_k, say, then the conditional distribution (20) depends on all v_i in Ω_k in addition to the dependencies introduced by H. The neighborhood for pixel s can therefore be large for large subdomains. To preserve the local neighborhood structure, we can use local subdomains, so Ω_s is defined to be a local neighborhood of pixel s. For example, all pixels within a radius of 3 pixels from s. The number of subdomains can now be large and of order N_0, but locally it is few as the conditional distribution (20) depends only of a few nearby subdomains. We can use any set of subdomains and lags that gives satisfactory residuals. The actual choice will depend also on the problem and is not unique.

6 Experiments

This section present some results using simulated data. The two test–images have a great potential to be improved using control variables, as they have a lot of edges and difficult areas. We used the whole image (except for a border region in the second image) as the only subdomain and all the 24 different lags within a square with dimension 7×7 as shown in Figure (2). This seems to be sufficient for the presented examples.

We present the images in a linear grey scale between the minimum and the maximum value in the true image, and the residuals in a linear scale between -3 and 3.

The first example is a big chess board like image with dimension 64×64 and pixel values 1 for black and 255 for white. We blurred the image with a Gaussian point spread function with variance 4 and added Gaussian noise with zero mean and variance 0.25. Figure (3) and Figure (4) shows the true and noisy image. This restoration problem is hard because of the blur and the image type. We did a restoration with both the Kullback–Leibler and the Geman and McClure distance measure without any success, as shown in Figure (5) for the Kullback–Leibler measure. Figure (6) shows the residuals from the restored image in Figure (5) and the correlation function is tabulated in Table (1). The variance of the estimates of the correlation function is about $1/64^2 = 0.016^2$. Using the conjugate gradient algorithm, we did the restoration with the $\mathbf{z}$–term and the Kullback–Leibler measure. Figure (7) shows the result and Figure (8) shows the residuals with the correlation function tabulated in Table (1). The structure in the residuals is recovered and the high frequency components in the true image are restored.

The second example is a 64×64 image with pixel values between 0 and 127. The image is piecewise constant and consists of superimposed randomly placed squares and lines. We added Gaussian noise with zero mean and standard deviation 10. Figure (9) and Figure (10) shows the true and noisy image. We restore the image using the Geman and McClure model with values of the parameter δ between 5 and 60, using low–temperature sampling. Figure (11) and Figure (13) shows the restored image with and without the $\mathbf{z}$–term $\delta = 50$. Figure (12) and Figure (14) shows the residuals and Table (2) tabulate the correlation function.

We evaluate the results for this example by two error measures. The first measures the global error Λ_g as the mean square value of the error image $\mathbf{x}^e = \hat{\mathbf{x}} - \mathbf{x}$, $\Lambda_g = \sum_i (x_i^e)^2/N$. The second measures how well the edges are recovered. Let η_{ij}^e, be the difference between pixel i and j in the error image, $\eta_{ij}^e = x_i^e - x_j^e$. Godtliebsen (1991) defines the edge error as

$$\Lambda_e = \frac{1}{8N} \sum_i \sum_{j \in \partial i} (\eta_{ij}^e - \overline{\eta^e})^2, \tag{23}$$

where ∂i means all the 8 nearest neighbors of pixel i and $\overline{\eta^e}$ is the mean of η_{ij}^e over all $i, \partial i$.

Figure (15) and Figure (16) shows the error measures calculated over the whole image (except for a border region) as a function of δ. Including the control variables gives a better restoration for $\delta > 10$, as seen in both error measures. The results shows that the restoration using control variables recover the edges better and do a better job in general for $\delta > 10$. The residuals are nearly satisfactory without the $\mathbf{z}$–term for low values of δ. This is because a low value of δ prevents smoothing over the edges and the difference adding the $\mathbf{z}$–term is then small. Higher values of δ increase the smoothing of the edges and this increase the effect of adding the $\mathbf{z}$–term.

When we include the $\mathbf{z}$–term, the results shows that a weaker distance measure Δ_2 (higher value of δ) makes both the global error and the edge error decrease when we increase δ. The $\mathbf{z}$–term seems to do better when it can work more freely and is not working against a strong distance measure Δ_2.

An interesting results shows up when the Kullback–Leibler distance measure is used to restore the noisy image in example two. Without the $\mathbf{z}$–term the MPD estimate is just a pixel transform independent of the neighbors,

$$\hat{x}_i = m_c \exp\left(\frac{1}{\alpha\sigma^2}(y_i - \hat{x}_i) - 1\right) \tag{24}$$

So high pixel values are pulled down and low values are pulled up. The residuals will always be positive for high pixel values and negative for low pixel values. The restored image will then be very similar to the noisy image and there will be no local spatial smoothing. By including the control variables, we "unsmooth" in the residual space and this leads to a spatial smoothing in the pixel space as show in Figure (17) and Figure (18), where the MPD estimate is clearly locally smooth. The correlation function is tabulated in Table (3). The error measures have values $\Lambda_g = 63$ and $\Lambda_e = 113$. With respect to these error measures, this method performs better than the Geman and McClure model (refer to Figure (15) and Figure (16)). A smoothing technique like this, does not assume any model for the local spatial smoothness in the true image. The result, will however depend on the true image.

7 Discussion and Conclusion

We have discussed the residuals that common image restoration methods estimate. Such methods can estimate highly biased residuals that will depend locally on the true underlying image. We have discussed how this information can be used further within the regularization framework, leading to more faithful residuals. Only moments of the residuals up to second order were considered, so higher order moments and long range correlation are not controlled by the control variables in the presented experiments. The results so far shows a better restoration of the edges and in difficult areas. In the rather special "chess–board" case, the structure in the residuals is resolved and high frequency components in the true image are restored.

The presented error measures should be used critically and should be used together with the subjective evaluation. With the term "a better" restoration, we mean with respect to the selected error measures and the statements are not absolute. The error measures will at least give an indication of the quality in the restoration.

We achieve best results with a weak model for the local spatial smoothness, together with the $\mathbf{z}$–term. On the basis of the selected global and edge error measures, the value of these measure decrease when the parameter δ in the Geman and McClure model increase. The experiments shows that a locally smooth restoration can be obtained without any model for the local spatial smoothness of the true image in the case of additive noise. The smoothing is done by unsmoothing in the residual space, which leads to smoothing in the pixel space.

The presented method of using control variables to control the residuals in the restoration, can also be looked upon as a generalization of common regularization methods (Hunt, 1973; Burch et al., 1983). We use local χ^2–based constraints on the correlation function of the residuals in addition to the constraint used on the residual variance.

Acknowledgments

Special thanks to Professor Emil Spjøtvoll for helpful comments.

References

BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion), *Journal of the Royal Statistical Society, Series B* 48: 259–302.

BROCKWELL, P. J. AND DAVIS, R. A. (1987). *Time Series: Theory and Methods*, Springer, Berlin.

BURCH, S. F., GULL, S. F. AND SKILLING, J. (1983). Image restoration by a powerful maximum entropy method, *Computer Vision, Graphics and Image Processing* 23: 113–128.

GEMAN, D. (1990). Random fields and inverse problems in imaging, *in* P. L. Hennequin (ed.), *Ecole d'ete de probabilites de Saint-Flour XVIII, 1988*, Springer, Berlin. Lecture notes in mathematics; 1427.

GEMAN, S. AND GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–741.

GEMAN, S. AND MCCLURE, D. E. (1987). Statistical methods for tomographic image reconstruction, *Bulletin of the International Statistical Institute* BK. 4: 5–21.

GEMAN, S., GEMAN, D., GRAFFIGNE, C. AND DONG, P. (1990). Boundary detection by constrained optimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12: 609–628.

GODTLIEBSEN, F. (1991). Noise reduction using markov random fields, *Journal of Magnetic Resonance* 92: 102–114.

HUNT, B. R. (1973). The application of constrained least squares estimation to image restoration by digital computer, *IEEE Transaction on Computers*.

MYRHEIM, J. AND RUE, H. (1992). New algorithms for maximum entropy image restoration, *Graphical Models and Image Processing* 54(3): 223–238.

RUE, H. (1993). *Topics in Image Analysis*, PhD thesis, The Norwegian Institute of Technology.

SKILLING, J. AND GULL, S. F. (1985). Algorithms and applications, *in* C. R. Smith and W. T. G. Jr. (eds), *Maximum–Entropy and Bayesian Methods in Inverse Problems*, D. Reidel Publ. Co., pp. 83–132.

THOMPSON, A. M., BROWN, J. C., KAY, J. W. AND TITTERINGTON, D. M. (1991). A study of methods of choosing the smoothing parameter in image restoration by regularization, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13: 326–339.

TITTERINGTON, D. M. (1985a). Common structure of smoothing techniques statistics, *International Statistics Review* 53: 141–170.

TITTERINGTON, D. M. (1985b). General structure of regularization procedures in image reconstruction, *Astron. Astrophys.* 144: 381–387.

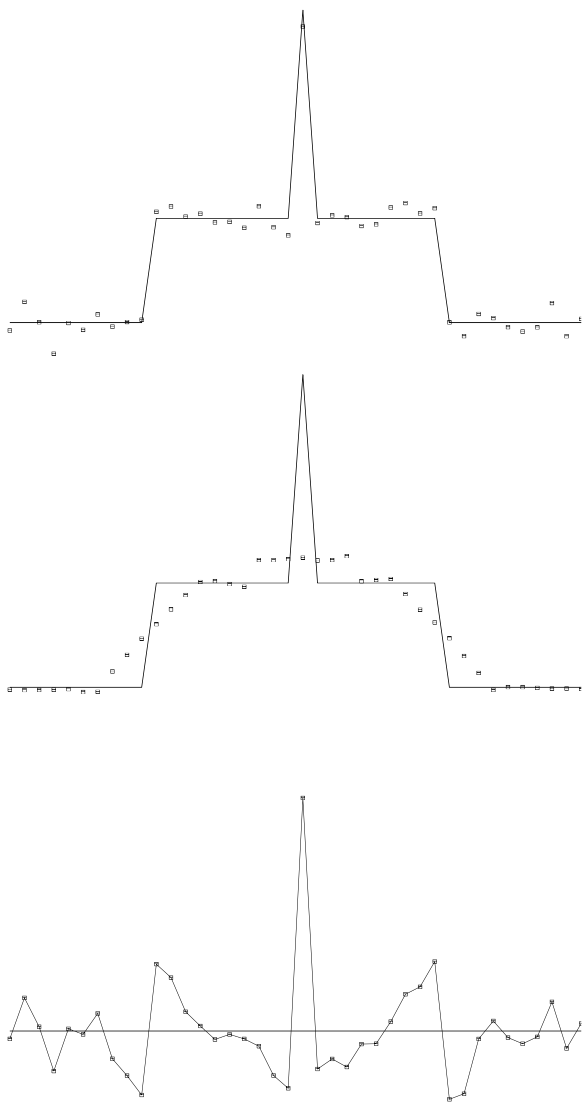

Figure 1: A simple example of biased residuals in the case of no blur and additive gaussian noise. On the top, the true (line) and the noisy curve. In the middle, the smoothed version of the noisy curve together with the true one is shown. The residuals obtained refereed to zero level are shown at bottom. The residuals are clearly biased in the neighborhood of the discontinuities, which could also be located as discontinuities in the residuals.

22	23	24	13	14	15	16
21	11	12	5	6	7	17
20	10	4	1	2	8	18
19	9	3	0	3	9	19
18	8	2	1	4	10	20
17	7	6	5	12	11	21
16	15	14	13	24	23	22

Figure 2: The 24 lags within a square with dimension 7×7.

−0.275	(−0.009)	0.288	(0.013)	−0.248	(−0.003)	0.287	(−0.031)
0.259	(−0.024)	−0.311	(−0.007)	0.293	(−0.003)	−0.257	(−0.040)
−0.396	(−0.019)	0.194	(−0.020)	−0.298	(−0.009)	0.303	(0.016)
1.000	(1.000)	−0.393	(−0.014)	0.239	(−0.016)	−0.265	(0.021)
−0.396	(−0.019)	0.231	(−0.008)	−0.301	(−0.006)	0.281	(0.010)
0.259	(−0.024)	−0.302	(−0.007)	0.285	(−0.010)	−0.248	(−0.022)
−0.275	(−0.009)	0.271	(0.005)	−0.264	(−0.033)	0.286	(−0.015)

Table 1: The correlation function of the residuals for the big chess board case using the Kullback–Leibler distance measure. The nearest 24 lags are shown with the value obtained with the **z**–term in parentheses. Only the lags for $x \geq 0$ are shown (refer to Figure (2)). In the first row starting from left, the lag number $13, 14, 15$ and 16 is shown. The second row contains the lag number $5, 6, 7$ and 17. The third row contains the lag number $1, 2, 8$ and 18 etc.

0.354	(0.005)	−0.116	(−0.002)	−0.028	(0.023)	0.106	(0.026)
0.258	(−0.033)	−0.199	(−0.013)	−0.122	(−0.021)	0.021	(−0.012)
0.196	(−0.002)	−0.314	(−0.018)	−0.164	(0.000)	−0.026	(0.030)
1.000	(1.000)	0.086	(0.004)	0.231	(0.021)	0.335	(0.020)
0.196	(−0.002)	−0.307	(−0.024)	−0.149	(0.005)	−0.011	(0.025)
0.258	(−0.033)	−0.238	(−0.033)	−0.112	(−0.015)	−0.007	(−0.025)
0.354	(0.005)	−0.108	(0.011)	−0.003	(0.018)	0.088	(0.023)

Table 2: The correlation function of the residuals for the second example using the Geman and McClure distance measure with the value of $\delta = 50$. The nearest 24 lags are shown with the value obtained with the **z**–term in parentheses. Only the lags for $x \geq 0$ are shown (refer to Figure (2)). In the first row starting from left, the lag number $13, 14, 15$ and 16 is shown. The second row contains the lag number $5, 6, 7$ and 17. The third row contains the lag number $1, 2, 8$ and 18 etc.

0.510	(−0.018)	0.303	(−0.026)	0.262	(−0.028)	0.281	(−0.024)
0.560	(0.001)	0.342	(−0.008)	0.296	(−0.027)	0.308	(0.024)
0.626	(−0.006)	0.384	(−0.007)	0.346	(0.007)	0.345	(0.002)
1.000	(1.000)	0.589	(0.028)	0.537	(0.017)	0.520	(−0.003)
0.626	(−0.006)	0.386	(0.000)	0.344	(−0.011)	0.348	(−0.003)
0.560	(0.001)	0.325	(−0.022)	0.287	(−0.024)	0.285	(0.029)
0.510	(−0.018)	0.288	(−0.009)	0.248	(−0.037)	0.245	(−0.032)

Table 3: The correlation function of the residuals for the second example using the Kullback–Leibler distance measure. The nearest 24 lags are shown with the value obtained with the **z**–term in parentheses. Only the lags for $x \geq 0$ are shown (refer to Figure (2)). In the first row starting from left, the lag number $13, 14, 15$ and 16 is shown. The second row contains the lag number $5, 6, 7$ and 17. The third row contains the lag number $1, 2, 8$ and 18 etc.

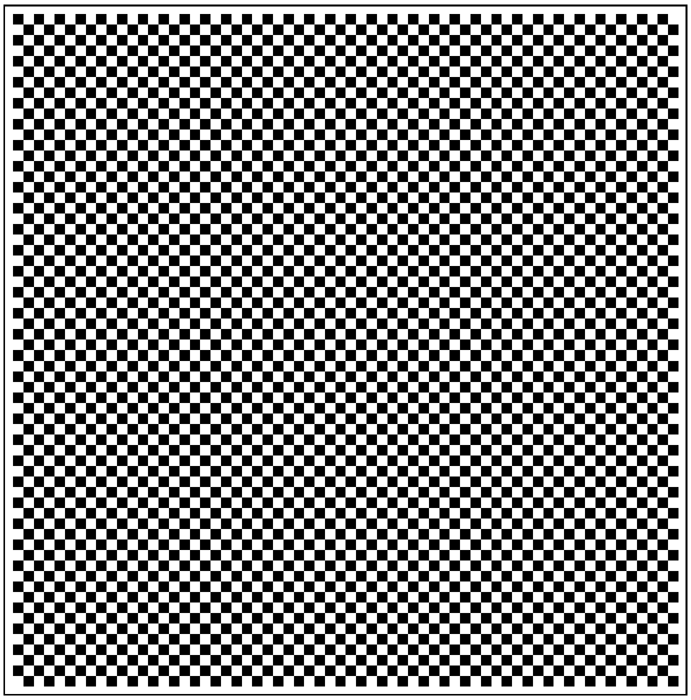

Figure 3: The original big chess–board image. The dimension is 64×64 and the two states have values 1 and 255.

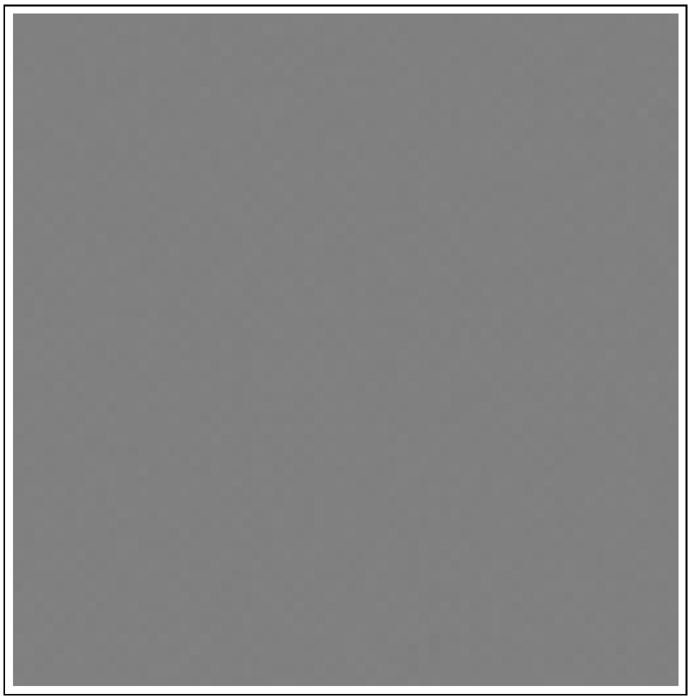

Figure 4: The big chess–board image with blur and added noise. The blur was a Gaussian probability density function with standard deviation equal 2. The noise was Gaussian with zero mean and standard deviation 0.5.

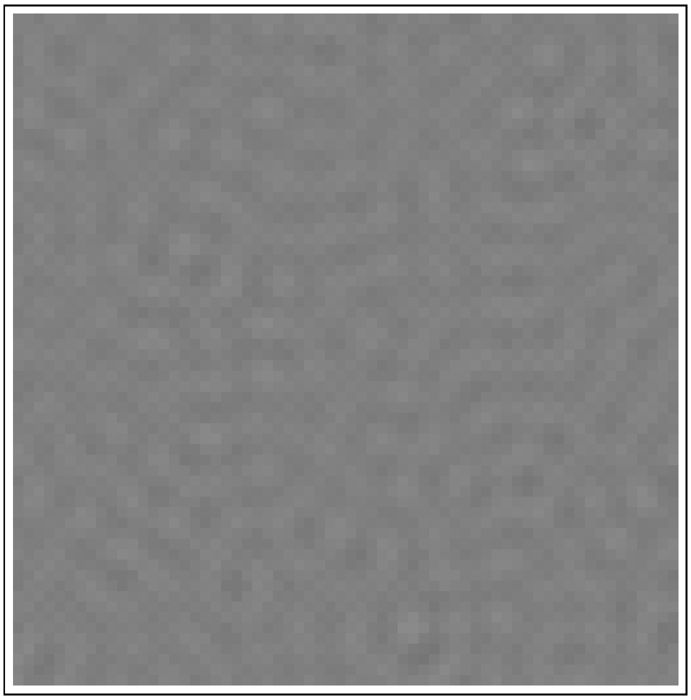

Figure 5: The restored image from Figure (4) with the Kullback–Leibler distance measure and no $\mathbf{z}$–term.

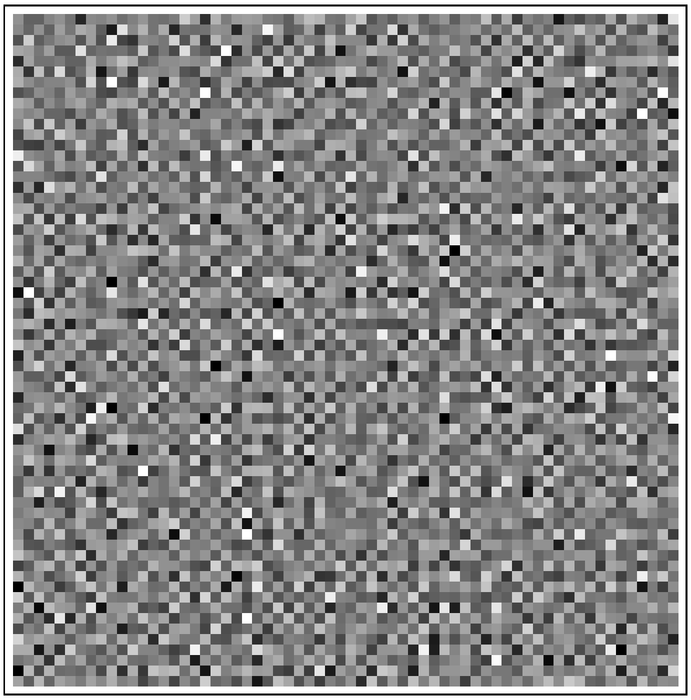

Figure 6: The residuals obtained from Figure (4) and Figure (5).

Figure 7: The restored image from Figure (4) with the Kullback–Leibler distance measure and the $\mathbf{z}$–term.

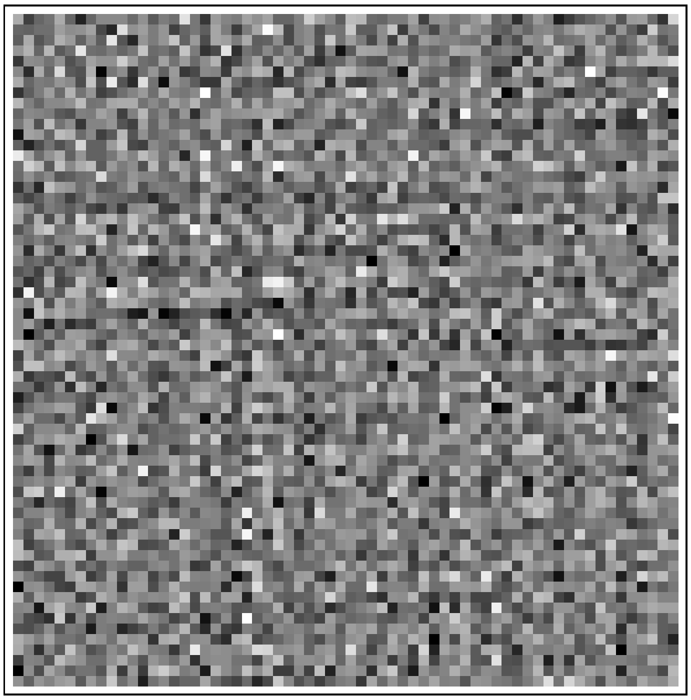

Figure 8: The residuals obtained from Figure (4) and Figure (7).

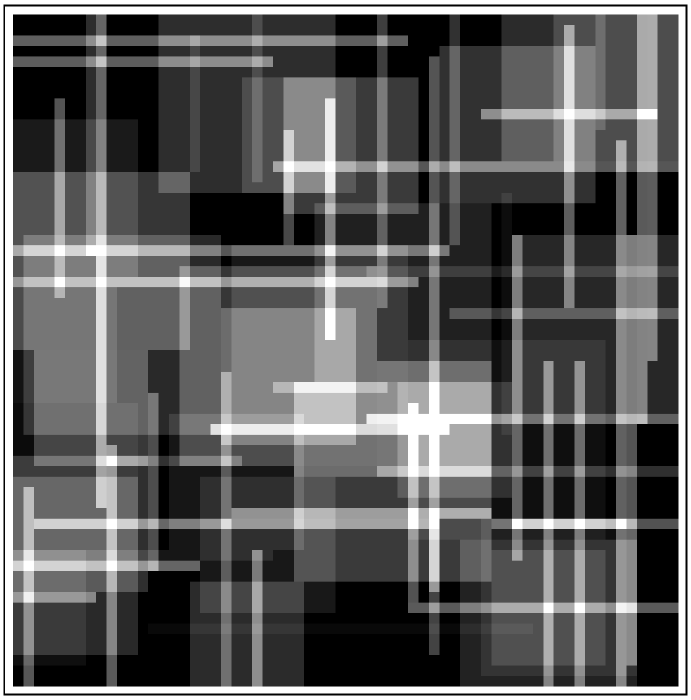

Figure 9: The original 64×64 image with pixel values among 0 and 127.

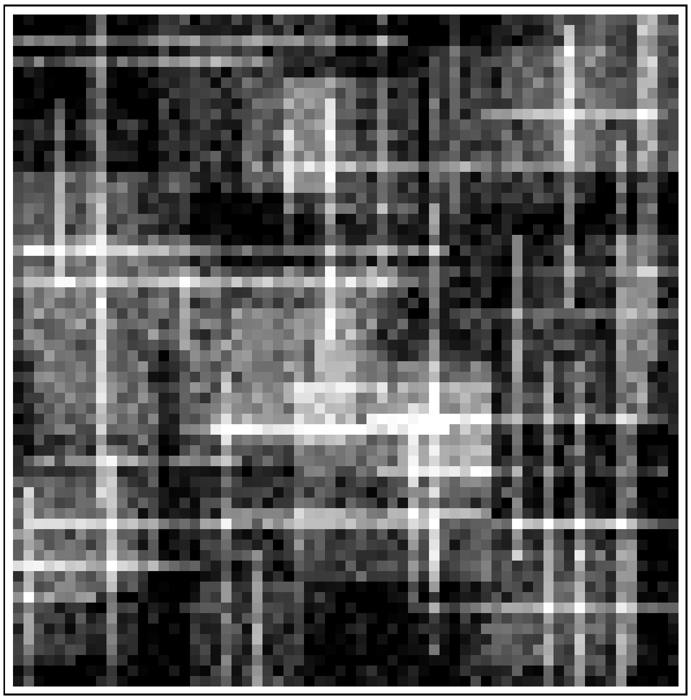

Figure 10: The true image Figure (9) with gaussian additive noise with zero mean and standard deviation 10.

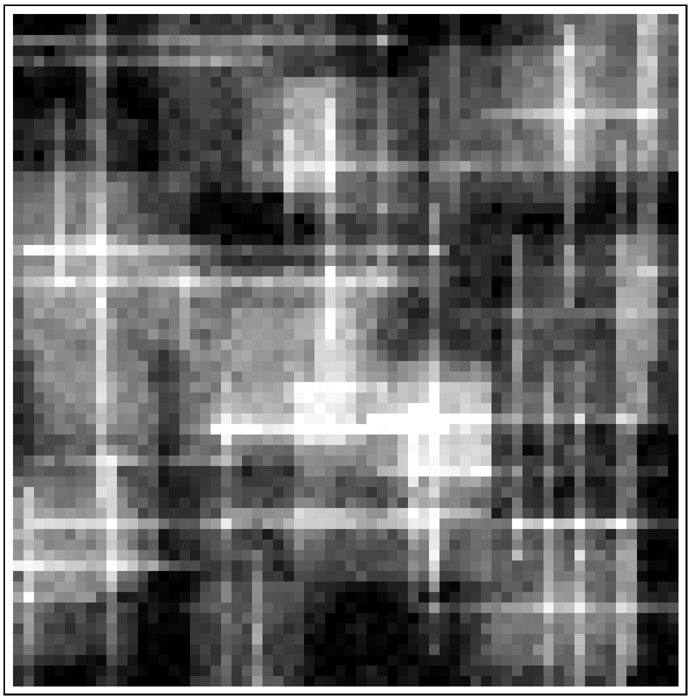

Figure 11: The restored image from Figure (10) with the Geman and McClure model with the parameter $\delta = 50$ and no $\mathbf{z}$–term.

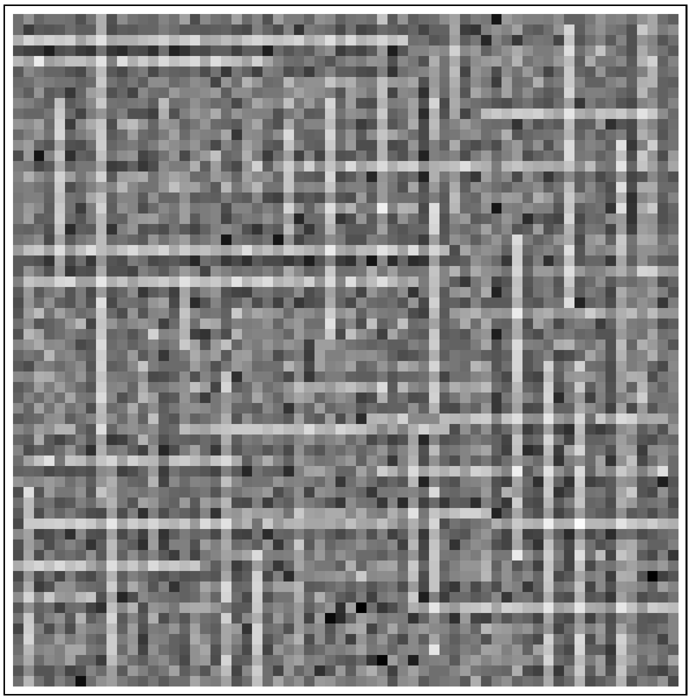

Figure 12: The residuals obtained from Figure (10) and Figure (11).

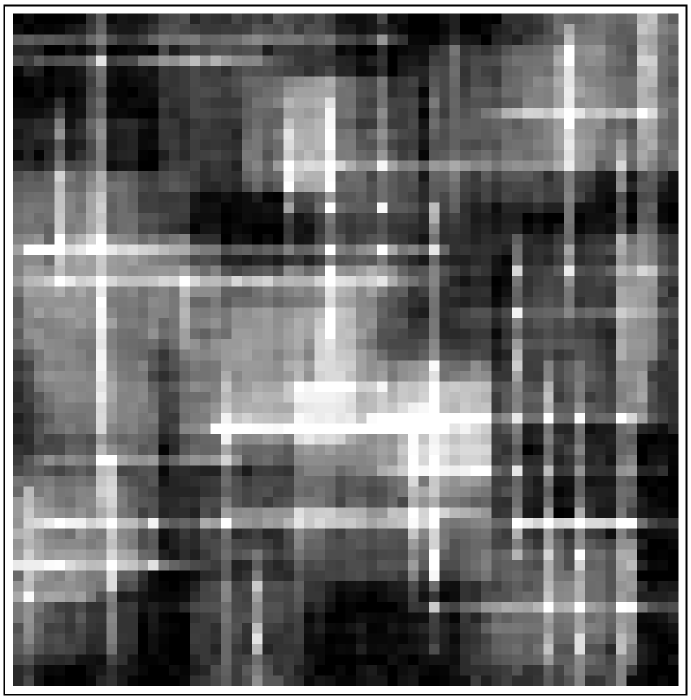

Figure 13: The restored image with the $\mathbf{z}$–term from Figure (10) with the Geman and McClure model with the parameter $\delta = 50$.

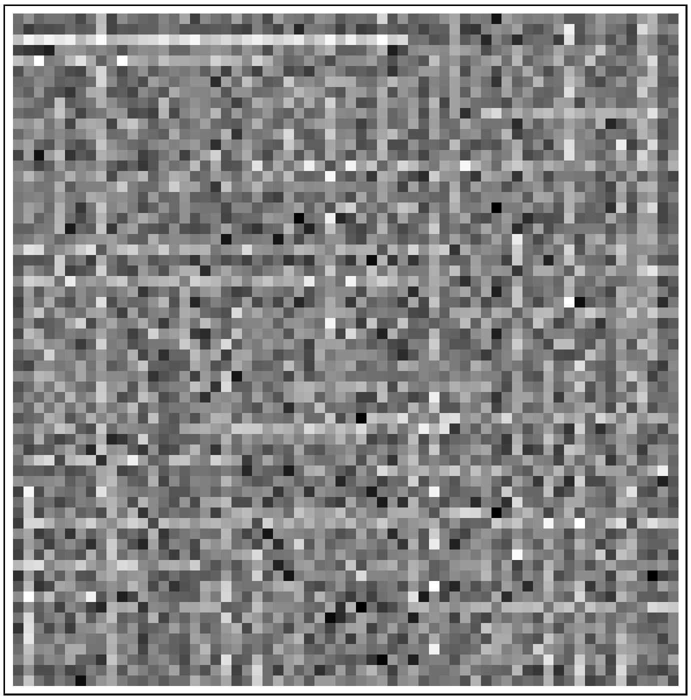

Figure 14: The residuals obtained from Figure (10) and Figure (13). Note that the border region is not included in Ω_1.

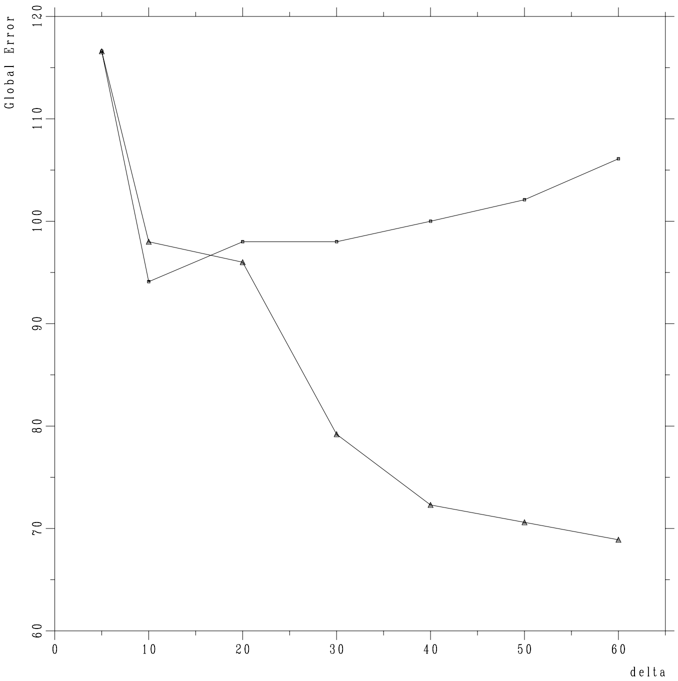

Figure 15: The global error for the restored images from Figure (10) with and without the $\mathbf{z}$–term. The Geman and McClure model was used with values of the parameter δ varying from 5 to 60. The upper curve is without the $\mathbf{z}$–term and the lower curve is with the $\mathbf{z}$–term.

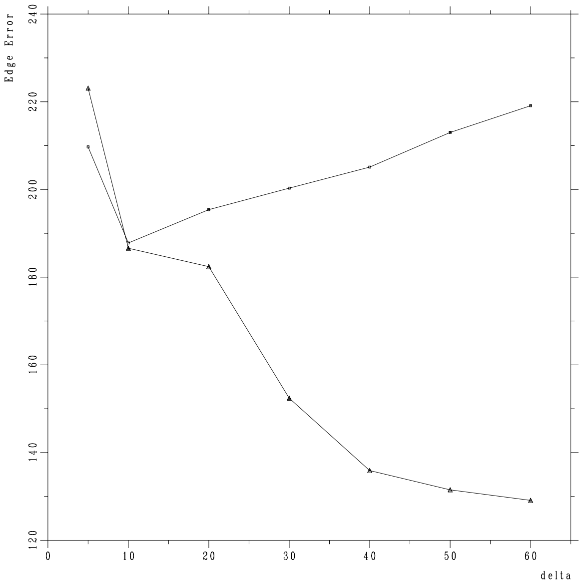

Figure 16: The edge error for the restored images from Figure (10) with and without the $\mathbf{z}$–term. The Geman and McClure model was used with values of the parameter δ varying from 5 to 60. The upper curve is without the $\mathbf{z}$–term and the lower curve is with the $\mathbf{z}$–term.

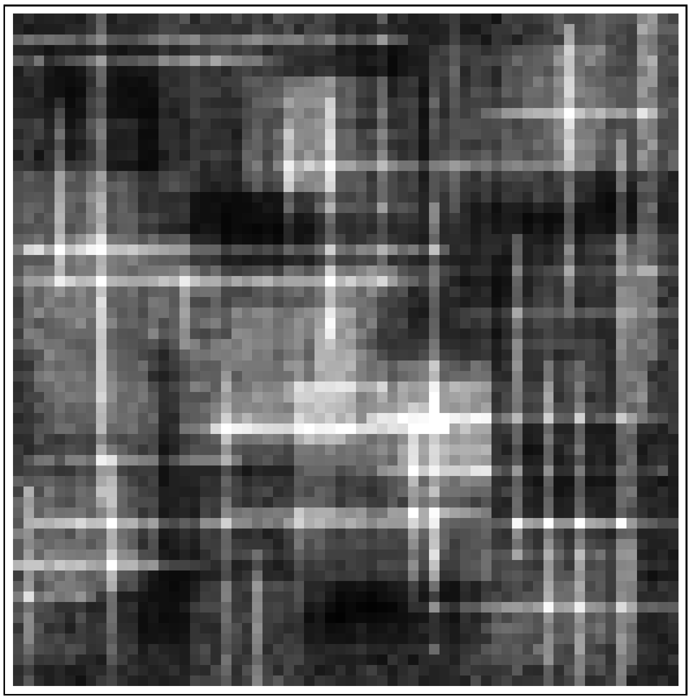

Figure 17: The restored image Figure (10) with the Kullback–Leibler distance measure and the $\mathbf{z}$–term. Note that the border region is not included in Ω_1.

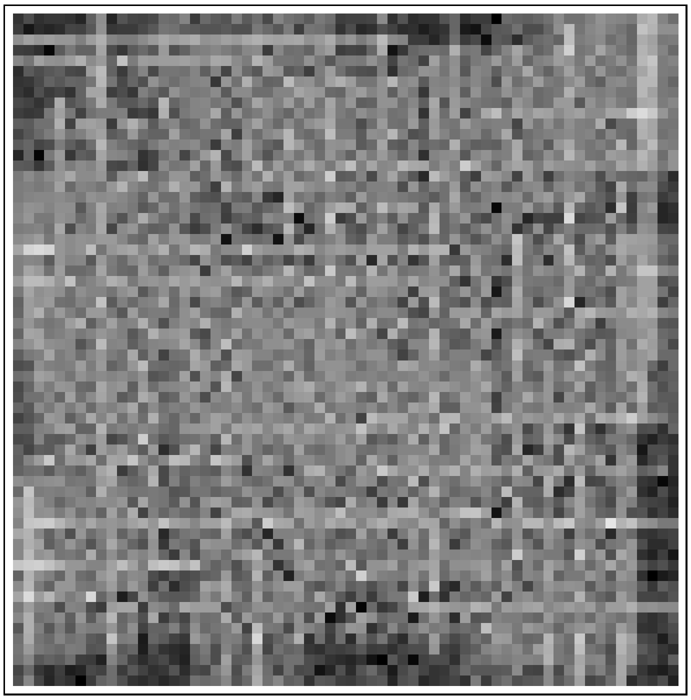

Figure 18: The residuals obtained from Figure (10) and Figure (17). Note that the border region is not included in Ω_1.