Bayesian Image Restoration by Markov Chain Monte Carlo Methods

Karl-Rudolf Koch

Summary

To smooth digital images while preserving the edges of an image, the Bayesian analysis is applied. The prior information that the image is smooth except for the edges is introduced by a Gibbs distribution based on a modification of the density of Huber (1964) for robust parameter estimation. The analytical solution for estimating the intensity level of one pixel leads to the very fast estimation by iterated conditional modes (ICM) of Besag (1986). Applying it to the smoothing of a digital image, the results show an excellent edge preserving quality. They agree numerically with the estimates of Markov Chain Monte Carlo methods, i. e. the Gibbs sampler and the Metropolis algorithm. The method ICM therefore gives a maximum a posteriori (MAP) estimate.

Zusammenfassung

Um digitale Bilder zu glätten, wobei die Kanten im Bild zu erhalten sind, wird das Bayes-Verfahren angewendet. Die Vorinformation, dass ein Bild glatt bis auf die Kanten ist, wird durch eine Gibbs-Verteilung eingeführt, die auf einer Modifizierung der Dichte beruht, die Huber (1964) für die robuste Parameterschätzung einführte. Die analytische Lösung für die Schätzung des Intensitätsniveaus eines Pixels führt auf das sehr schnelle ICM (iterated conditional modes)-Verfahren von Besag (1986). Bei der Anwendung auf die Glättung eines digitalen Bildes zeigt es eine exzellente Eigenschaft, die Kanten im Bild zu erhalten. Die numerischen Ergebnisse stimmen mit den Schätzungen von Monte-Carlo-Methoden mit Markoff-Ketten überein, und zwar dem Gibbs-Verfahren und dem Metropolis-Algorithmus. Das ICM-Verfahren führt also auf eine MAP (maximum a posteriori)-Schätzung.

1 Introduction

Bayesian analysis in contrast to traditional statistics is founded on Bayes' theorem which gives the probability density function for the unknown parameters. By this density function the unknown parameters can be estimated, confidence regions for the parameters be established and hypotheses for the parameters be tested. In the linear model, e. g., analytical solutions are derived for these tasks. If this is not possible, one has to rely on numerical procedures, especially Monte Carlo integration.

In the last two decades, Markov Chain Monte Carlo methods, which sample from the probability distribution for the unknown parmeters, have dominated the numerical procedures. The Metropolis algorithm was the first one to be developed and goes back to Metropolis et al. (1953). It does not need a special distribution to sample from. A variant of the Metropolis algorithm is the technique of simulated annealing by Kirkpatrick et al. (1983), although it is an optimization instead of a simulation method. A scale parameter, called temperature, is introduced into the target distribution and gradually decreased to avoid being trapped in local minima.

Especially the Gibbs sampler of the Markov Chain Monte Calo methods turned out to be a very versatile and efficient procedure. It was applied together with the simulated annealing by Geman and Geman (1984) for the Bayesian restoration of digital images by a maximum a posteriori (MAP) estimation. Gelfand and Smith (1990) showed that it could be used for a variety of problems of Bayesian inference. It generates random values from conditional distributions for the unknown parameters. Monte Carlo simulations take computing time, but they can be readily applied in parallel computing, cf. Koch et al. (2004).

In digital image restoration, e. g. in computer tomography, there is the prior information that locally the image is smooth except for edges, sudden changes of the intensity values of the pixels of the image. Prior information in Bayesian analysis is readily expressed by the prior distribution. The intensity levels of the pixels of a digital image represent a random field for which the Markov property can be assumed because the intensity value of a pixel is mainly influenced by the ones of the pixels of the neighbourhood, cf. Koch and Schmidt (1994, p. 299). The prior information can therefore be expressed by the Gibbs distribution which may be defined such that large density values follow for smooth images and small ones for rough images so that a smooth image results from the prior information. However, the smoothing effect has to stop at the edges of the image.

This has been accomplished by introducing a random field of line elements, cf. Busch and Koch (1990). The line elements are positioned between the pixels and represented by discrete random variables which obtain the value zero if no element is present and the value one if there is an element. However, introducing prior information for the random field of edges is not quite a simple task.

For reconstructing images of computer tomography it has been proposed to model the Gibbs distribution for the prior information by the density of Huber (1964) for the robust parameter estimation, cf. Fessler et al. (2000). This idea is quite appealing since an intensity value of a pixel beyond an edge, which should not contribute to the smoothing process, is considered an outlier. For a better edge preserving quality the density of Huber (1964) has been modified here such that pixels beyond edges are not used for the smoothing. The resulting approach is investigated in the following for a digital image given by grey values.

Assuming a normal distribution for the measured grey values of the pixels and a Gibbs distribution for the prior information based on a modified density function of Huber (1964) leads to an analytical solution for the estimate of the unknown grey value of one pixel. By iteratively applying it, the estimation by iterated conditional modes (ICM) of Besag (1986) is obtained, which is very fast. It gives a smooth image with an excellent edge-preserving quality. To find out whether the ICM algorithm gives only approximate results, the Gibbs sampler and the Metropolis algorithm are used to estimate the unknown grey values of the restored image. The numerical solutions agree except for minor deviations. The method ICM therefore gives the MAP estimate.

In the following section a short review is given for Bayes estimation in linear models. In addition, the estimations by Monte Carlo integration and by the Markov Chain Monte Carlo methods are outlined. Section 3 formulates the problem of image restoration using a modified distribution for the robust parameter estimation as prior information, while section 4 presents a numerical example. The paper ends with conclusions.

2 Bayes Estimation by Monte Carlo Methods

Let $\boldsymbol{\beta}$ be the $u \times 1$ random vector of unknown parameters and $\boldsymbol{y}$ the $n \times 1$ random vector of observations, $p(\boldsymbol{\beta}|C)$ the prior density function for $\boldsymbol{\beta}$ given the background information C, $p(\boldsymbol{y}|\boldsymbol{\beta}, C)$ the likelihood function where $\boldsymbol{y}$ is given, then the posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y}, C)$ for $\boldsymbol{\beta}$ is obtained from Bayes' theorem, cf. Koch (2000, p. 36),

$$p(\boldsymbol{\beta}|\boldsymbol{y}, C) = \frac{1}{c} p(\boldsymbol{\beta}|C) p(\boldsymbol{y}|\boldsymbol{\beta}, C) \tag{2.1}$$

with c being the normalizing constant

$$c = \int_B p(\boldsymbol{\beta}|C) p(\boldsymbol{y}|\boldsymbol{\beta}, C) d\boldsymbol{\beta} \tag{2.2}$$

where the integral is extended over the parameter space B of $\boldsymbol{\beta}$. Usually the constant c is omitted in (2.1), thus

$$p(\boldsymbol{\beta}|\boldsymbol{y}, C) \propto p(\boldsymbol{\beta}|C) p(\boldsymbol{y}|\boldsymbol{\beta}, C) \tag{2.3}$$

where $\propto$ denotes proportionality. For a simpler notation the conditioning on the background information C is omitted in the following.

Estimating the unknown parameters $\boldsymbol{\beta}$ is considered a decision problem and by introducing a quadratic loss function we obtain the Bayes estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \int_B \boldsymbol{\beta} p(\boldsymbol{\beta}|\boldsymbol{y}) d\boldsymbol{\beta} \, . \tag{2.4}$$

Using the zero-one loss we find the maximum a posteriori (MAP) estimate. The same loss function leads to the tests of hypotheses. Also confidence regions can be derived by (2.1). In the following, however, we will concentrate on the Bayes estimate (2.4).

A linear model for the unknown parameters $\boldsymbol{\beta}$ is defined by

$$\boldsymbol{X\beta} = E(\boldsymbol{y}|\boldsymbol{\beta}) \quad \text{with} \quad D(\boldsymbol{y}|\sigma^2) = \sigma^2 \boldsymbol{P}^{-1} \tag{2.5}$$

with $\boldsymbol{X}$ being an $n \times u$ matrix of given coefficients, which for simplicity shall possess full column rank, σ^2 the variance factor and $\boldsymbol{P}$ the known positive definite weight matrix of the observations $\boldsymbol{y}$, which are assumed as normally distributed, thus

$$\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2 \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{P}^{-1}) \, . \tag{2.6}$$

The likelihood function $p(\boldsymbol{y}|\boldsymbol{\beta})$ following from (2.6) together with the noninformative prior for $\boldsymbol{\beta}$, which is a constant, lead – with Bayes' theorem (2.3) – to the posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$ for $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}|\boldsymbol{y}) \propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X\beta})\right\} \, . \tag{2.7}$$

This is the density function of the normal distribution

$$\boldsymbol{\beta}|\boldsymbol{y} \sim N(\hat{\boldsymbol{\beta}}, \sigma^2(\boldsymbol{X}'\boldsymbol{PX})^{-1}) \tag{2.8}$$

with $\hat{\boldsymbol{\beta}}$ being the Bayes estimate (2.4) of $\boldsymbol{\beta}$. It is identical with the one obtained by the well known method of least squares, cf. Koch (2000, p. 90),

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{PX})^{-1}\boldsymbol{X}'\boldsymbol{Py} \, . \tag{2.9}$$

This Bayes estimate is also identical with the MAP estimate.

In case we cannot derive an analytical solution like (2.9), we have to rely on numerical methods. The Monte Carlo integration is – for higher dimensions – well suited to solve the integral (2.4). It is based on the assumption that a density function $u(\boldsymbol{\beta})$ is known which approximates the posterior distribution $p(\boldsymbol{\beta}|\boldsymbol{y})$ and from which random values for $\boldsymbol{\beta}$ can be generated. Let these random values be denoted by $\boldsymbol{\beta}_i$ with $i \in \{1 \ldots m\}$. The Bayes estimate $\hat{\boldsymbol{\beta}}$ then follows with, cf. Koch (2000, p. 191),

$$\hat{\boldsymbol{\beta}} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{\beta}_i p(\boldsymbol{\beta}_i|\boldsymbol{y}) / u(\boldsymbol{\beta}_i) \, . \tag{2.10}$$

This is called importance sampling because the random values $\boldsymbol{\beta}_i$ are generated at points which are important as $u(\boldsymbol{\beta})$ approximates $p(\boldsymbol{\beta}|\boldsymbol{y})$. A simpler estimate than (2.10) is found by the Markov Chain Monte Carlo methods which sample from the posterior distribution for $\boldsymbol{\beta}$ itself. Thus, with $p(\boldsymbol{\beta}|\boldsymbol{y}) = u(\boldsymbol{\beta})$ and with $\boldsymbol{\beta}_i$ now distributed like $p(\boldsymbol{\beta}|\boldsymbol{y})$ we obtain instead of (2.10) the Bayes estimate

$$\hat{\boldsymbol{\beta}} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{\beta}_i \, . \tag{2.11}$$

In addition, the covariance matrix of $\boldsymbol{\beta}$ can be determined by the random values $\boldsymbol{\beta}_i$ and hypotheses for $\boldsymbol{\beta}$ be tested, cf. Gundlich et al. (2003) and Koch (2005).

As mentioned already in section 1, the Metropolis algorithm was the first of the Markov Chain Monte Carlo methods to be developed. Its idea is to simulate a Markov Chain in the parameter space B of $\boldsymbol{\beta}$ such that the limiting distribution of the chain is the target distribution, in our case the posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$. The Metropolis algorithm can be applied to generate random samples from any target distribution whose normalizing constant does not have to be known so that (2.3) can be used. However, slow convergence in case of higher dimensions might be encountered. The algorithm runs as follows, cf. Gelman et al. (2004, p. 289).

One samples a proposal $\boldsymbol{\beta}^*$ from a jumping or proposal distribution $p_t(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{t-1})$ for $t \in \{1, 2, \ldots\}$ with $\boldsymbol{\beta}^{t-1}$ being the previously generated vector. The jumping distribution has to be symmetric for the Metropolis algorithm, however not for the Metropolis-Hastings procedure, which will not be considered here. Symmetry means that the probability of obtaining $\boldsymbol{\beta}'$ from perturbing $\boldsymbol{\beta}$ is equal to the one of obtaining $\boldsymbol{\beta}$ from perturbing $\boldsymbol{\beta}'$. The ratio r of the densities

$$r = \frac{p(\boldsymbol{\beta}^*|\boldsymbol{y})}{p(\boldsymbol{\beta}^{t-1}|\boldsymbol{y})} \tag{2.12}$$

is computed. One sets

$$\boldsymbol{\beta}^t = \begin{cases} \boldsymbol{\beta}^* & \text{with probability } \min(r, 1) \\ \boldsymbol{\beta}^{t-1} & \text{otherwise.} \end{cases} \tag{2.13}$$

The last step requires generating a random number v uniformly distributed in the interval $[0, 1]$ and $\boldsymbol{\beta}^*$ is accepted, if $v \le r$ since $P(v) = v$.

A simple algorithm, which is frequently applied, is the random-walk Metropolis, cf. Liu (2001, p. 114). The last generated vector $\boldsymbol{\beta}^{t-1}$ is perturbed by a random vector $\boldsymbol{\epsilon}^{t-1}$ to obtain the proposal $\boldsymbol{\beta}^* = \boldsymbol{\beta}^{t-1} + \boldsymbol{\epsilon}^{t-1}$. The components of $\boldsymbol{\epsilon}^{t-1}$ are assumed as being independent and identically distributed usually like the normal distribution which is symmetric. Thus, $\boldsymbol{\epsilon}^{t-1}$ can be easily generated. Random-walk Metropolis together with the normal distribution will be used in the example of section 4.

As mentioned in the introduction, the Gibbs sampler is the algorithm frequently applied to sample from a posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$ by a Markov Chain Monte Carlo method. To apply it, one has to sample, if we define $\boldsymbol{\beta} = |\beta_1, \beta_2, \ldots, \beta_u|'$, from the conditional density functions

$$p(\beta_i|\beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_u) \quad \text{for} \quad i \in \{1, \ldots, u\} \tag{2.14}$$

of the posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$. Highly correlated components of $\boldsymbol{\beta}$ should be collected in subvectors and random values should be generated for these subvectors, cf. Gundlich et al. (2003). In case of (2.14) the Gibbs sampler begins with arbitrary starting values

$$\beta_1^{(0)}, \ldots, \beta_u^{(0)} . \tag{2.15}$$

Then, random values

$$\begin{array}{lll} \beta_1^{(1)} & \text{from} & p(\beta_1|\beta_2^{(0)}, \ldots, \beta_u^{(0)}, \boldsymbol{y}) \\ \beta_2^{(1)} & \text{from} & p(\beta_2|\beta_1^{(1)}, \beta_3^{(0)}, \ldots, \beta_u^{(0)}, \boldsymbol{y}) \\ \beta_3^{(1)} & \text{from} & p(\beta_3|\beta_1^{(1)}, \beta_2^{(1)}, \beta_4^{(0)}, \ldots, \beta_u^{(0)}, \boldsymbol{y}) \\ \ldots & \ldots & \ldots \\ \beta_u^{(1)} & \text{from} & p(\beta_u|\beta_1^{(1)}, \ldots, \beta_{u-1}^{(1)}, \boldsymbol{y}) \end{array} \tag{2.16}$$

are sequentially generated to complete the first step of an iteration. After a burn-in phase of let say o iterations the Gibbs sampler can be shown to converge to the posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$ so that the generated samples $\boldsymbol{\beta}_i$ with $i > o$ are distributed like $p(\boldsymbol{\beta}|\boldsymbol{y})$.

3 Image Restoration

We are dealing with rectangular arrays of pixels whose intensity values have been measured e. g. by counts of photon emissions of a computer tomography. The measurements are distorted by noise so that the image has to be reconstructed from noisy measurements, cf. Geman and McClure (1987). The image shall contain discontinuities or edges which means sudden changes of the intensity levels. In the following we will concentrate on the problem of smoothing the image except for the edges. We therefore consider the simple problem of a digital photography of one colour where a reconstruction of the image like in computer tomography is not needed and where the intensity is measured by grey values. The point spread function of the imaging process, which describes the functional relation between measurements and unknown parameters, is then linear with the coefficient matrix being a unit matrix.

Let Ω be the set of pixels forming a lattice with

$$\begin{aligned} \Omega &= \{r = (m, n),\ 0 \le m \le M,\ 0 \le n \le N\}, \\ u &= (M+1)(N+1) \end{aligned} \tag{3.1}$$

and let the measurement of the grey value of pixel r be y_r with $r \in \{1, \ldots, u\}$ and $\boldsymbol{y}$ be the $u \times 1$ vector of measurements. Let the unknown, restored grey value of pixel r be β_r with $r \in \{1, \ldots, u\}$ and $\boldsymbol{\beta}$ the $u \times 1$ vector of unknown parameters. Let the measurements be independent und have variances σ^2. We then obtain instead of (2.5) the simple linear model

$$\boldsymbol{\beta} = E(\boldsymbol{y}|\boldsymbol{\beta}) \quad \text{with} \quad D(\boldsymbol{y}|\sigma^2) = \sigma^2 \boldsymbol{I} . \tag{3.2}$$

As already mentioned in section 1, the measurements $\boldsymbol{y}$ as well as the unknown parameters $\boldsymbol{\beta}$ represent Markov random fields. We assume the observations as normally distributed so that the likelihood function follows from

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{r\in\Omega}(y_r - \beta_r)^2\right\} . \tag{3.3}$$

The restored image should be smooth except for the edges. This prior information may be introduced by a Gibbs distribution because of the equivalence of Markov random fields and neighbour Gibbs fields, cf. Besag (1974). The Gibbs distribution for the vector $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta}) = \frac{1}{Z} \exp\{-U(\boldsymbol{\beta})\} . \tag{3.4}$$

where Z denotes the normalizing constant and the potential $U(\boldsymbol{\beta})$ is introduced for so-called cliques which are configurations of neighbours and defined for a neighbourhood N_p of order p for a pixel r. For the numerical example of section 4 we introduce the neighbourhood N_3 of order $p = 3$ defined by the following indices s_i with $i \in \{1, \ldots, 6\}$, cf. Koch and Schmidt (1994, p. 277),

$$\begin{aligned} N_3 &= \{s_1, s_2, s_3, s_4, s_5, s_6\} \\ &= \{(1,0), (0,1), (1,-1), (1,1), (2,0), (0,2)\} . \end{aligned} \tag{3.5}$$

The neighbourhood of pixel r follows with $r + s_i$ and $r - s_i$ so that it consists of 12 surrounding pixels for the neighbourhood N_3. If we define potential functions only for cliques with two sites, we may formulate the prior information for $\boldsymbol{\beta}$ by, see Busch and Koch (1990),

$$p(\boldsymbol{\beta}) \propto \exp\left\{-\frac{c_\beta}{2\sigma^2} \sum_{r\in\Omega} \sum_{s\in N_p} (\beta_r - \beta_{r+s})^2\right\} \tag{3.6}$$

which is a normal distribution. The larger the difference $\beta_r - \beta_{r+s}$ between the unknown grey value of pixel r and the one of the neighbouring pixel $r + s$ the smaller is the density value. The estimation of a rough image is therefore less likely than the restoration of a smooth one. The contribution of the prior information is controlled by the coefficient c_β. A prior distribution with a similar property which, however, is not a normal distribution has been proposed e. g. by Geman and McClure (1987).

The sum in (3.6) has to be extended over the potentials of all cliques with two sites in the set Ω which is accomplished by summing over the cliques of half of the neighbourhood N_p. The conditional density function $p(\beta_r|\partial\beta_r, \sigma^2)$ for β_r given the unknown grey values $\partial\beta_r$ in the neighbourhood of r follows from summing over the potentials of the cliques of the whole neighbourhood N_p, cf. Koch and Schmidt (1994, p. 262),

$$p(\beta_r|\partial\beta_r, \sigma^2) \propto \exp\left\{-\frac{c_\beta}{2\sigma^2} \sum_{\pm s\in N_p} (\beta_r - \beta_{r+s})^2\right\} \tag{3.7}$$

which again is a normal distribution.

The posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2)$ for the unknown grey values $\boldsymbol{\beta}$ of the restored image is obtained from Bayes' theorem (2.3) with the prior density function (3.6) and the likelihood function (3.3) by

$$p(\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{r\in\Omega}(y_r - \beta_r)^2 - \frac{c_\beta}{2\sigma^2}\sum_{r\in\Omega}\sum_{s\in N_p}(\beta_r - \beta_{r+s})^2\right\} . \tag{3.8}$$

We are looking for iterative estimates of $\boldsymbol{\beta}$ by means of an analytical solution for the estimate of β_r of one pixel r. We use the conditional density function $p(\beta_r|\partial\beta_r, \boldsymbol{y}, \sigma^2)$ for β_r following with (3.7) from (3.8)

$$p(\beta_r|\partial\beta_r, \boldsymbol{y}, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(y_r - \beta_r)^2 - \frac{c_\beta}{2\sigma^2}\sum_{\pm s\in N_p}(\beta_r - \beta_{r+s})^2\right\} . \tag{3.9}$$

By comparing this density function with the posterior density function (2.7) we recognize that (3.9) results from the special linear model

$$\boldsymbol{X} = \begin{vmatrix} 1 \\ \ldots \\ 1 \end{vmatrix} , \quad \boldsymbol{\beta} = \beta_r , \quad \boldsymbol{y} = \begin{vmatrix} y_r \\ \beta_{r+s1} \\ \beta_{r-s1} \\ \ldots \\ \beta_{r+sl} \\ \beta_{r-sl} \end{vmatrix} ,$$

$$\boldsymbol{D}(\boldsymbol{y}|\sigma^2) = \sigma^2 \begin{vmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1/c_\beta & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1/c_\beta & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1/c_\beta & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1/c_\beta \end{vmatrix} \tag{3.10}$$

with s_l being the last index for the neighbourhood N_p. According to (2.8) the unknown grey value β_r of pixel r is therefore normally distributed like

$$\beta_r|\partial\beta_r, \boldsymbol{y}, \sigma^2 \sim N\left(\hat{\beta}_r, \sigma^2(1 + \sum_{\pm s\in N_p} c_\beta)^{-1}\right) \tag{3.11}$$

with the Bayes estimate $\hat{\beta}_r$ of β_r from (2.9)

$$\hat{\beta}_r = \left(1 + \sum_{\pm s\in N_p} c_\beta\right)^{-1} \left(y_r + \sum_{\pm s\in N_p} c_\beta \beta_{r+s}\right) . \tag{3.12}$$

Although large differences $\beta_r - \beta_{r+s}$ in grey values between pixel r and the pixel $r + s$ in half of the neighbourhood of r or in the whole neigborhood of r cause small density values according to (3.8) or (3.9), they still contribute to the estimate $\hat{\beta}_r$. If pixel $r + s$ lies with respect to pixel r beyond an edge, it must not effect the estimate $\hat{\beta}_r$. Since edges in images are formed by jumps in grey values, we modify the density of Huber (1964), see also Koch (1999, p. 259), for a robust parameter estimation such that we use in (3.6) and (3.7) for a given value of the index s

$$\begin{array}{llll} p(\beta_r) & \propto & \exp\,(\beta_r-\beta_{r+s})^2/2 & \text{for} \quad |\beta_r-\beta_{r+s}| \le c \\ p(\beta_r) & = & 0 & \text{for} \quad |\beta_r-\beta_{r+s}| > c \end{array} \tag{3.13}$$

where the constant c has to be set according to the jumps of the grey values representing the edges which one wants to preserve. Introducing (3.13) means that in (3.8), (3.9), (3.11) and (3.12) we set

$$\begin{array}{llll} c_\beta & \neq & 0 & \text{for} \quad |\beta_r-\beta_{r+s}| \le c \\ c_\beta & = & 0 & \text{for} \quad |\beta_r-\beta_{r+s}| > c\,. \end{array} \tag{3.14}$$

Thus, (3.14) prevents the grey values of pixels beyond edges specified by c to enter the estimate $\hat{\beta}_r$ in (3.12) so that an edge preserving quality is obtained.

The estimate (3.12) together with (3.14) is iteratively applied. In one cycle of an iteration the grey values of all pixels $r \in \Omega$ are estimated. Each estimate replaces immediately the estimate of the previous iteration. This gives the estimation by iterated conditional modes (ICM) of Besag (1986) since the Bayes estimate (3.12) is also a MAP estimate, as already mentioned in connection with (2.9). It is very fast, but one has to find out whether this algorithm gives only approximate results.

The grey values of the restored image are therefore also estimated by the Gibbs sampler. The conditional density function for the Gibbs sampler (2.16) is obtained by the distribution for $\beta_r|\partial\beta_r, \boldsymbol{y}, \sigma^2$ with $r \in \{1, \ldots, u\}$ from (3.11) together with (3.12) and (3.14). After a burn-in phase the Gibbs sampler gives random values for the vector $\boldsymbol{\beta}$ of grey values distributed like the posterior density function (3.8). The estimates $\hat{\boldsymbol{\beta}}$ of the grey values then follow from (2.11).

Finally, random-walk Metropolis is applied to estimate the grey values of the restored image. Densities of the posterior density function (3.8) with (3.14) are inserted into (2.12) to compute the ratio r. Because of (3.14) sudden changes of the densities from (3.8) occur. Local minima may therefore be encountered for $\boldsymbol{\beta}^{t-1}$ in (2.13) which cannot be left in case of $r \le 1$, if the proposal $\boldsymbol{\beta}^*$ differs too much from $\boldsymbol{\beta}^{t-1}$. Small variances in comparison to σ^2 in (3.8) have therefore to be applied to generate $\boldsymbol{\epsilon}^{t-1}$ from the normal distribution for the random walk. This results in a slow convergence of the Metropolis algorithm. For the following application we therefore start from the estimate of the Gibbs sampler thus checking in addition whether the Gibbs sampler has reached as limiting distribution the posterior density function (3.8).

4 Numerical Application

The derived method for restoring an image by smoothing has been applied to a digital photography of low contrast with the intensities of the pixels measured by grey values. The photograph comprising 525×446 pixels, see Fig. 1, has been taken from outside through a double window into a room showing a bookshelf and double reflections of bushes outside. Fig. 2 depicts a moderate smoothing by applying the method ICM with (3.12) and by setting $c = 10.0$ expressed in grey values and $c_\beta = 1.0$. The starting values were the measured grey values. For the convergence of the iterations the maximum difference between the grey values of successive iterations should be less than a few grey values. Here, the maximum difference was chosen to be less than 0.01 which was reached after 75 iterations. A stronger smoothing is shown in Fig. 3 with $c = 21.0$ and $c_\beta = 1.0$. More details than in Fig. 3 appear in Fig. 4 by setting $c = 21.0$ and $c_\beta = 0.5$. It is obvious by comparing Fig. 2, 3 and 4 with Fig. 1 that the edges of the image are well preserved. If a smoother image is needed than the one shown in Fig. 3, a larger neighbourhood than N_3 in (3.5) should be chosen.

The results of Fig. 2 to 4 have been obtained by estimating the grey values during one iteration systematically row by row. The pixels were also randomly selected for each iteration. This procedure is slower than the one for the systematic scan. The square root of the mean squared differences between both approaches equals 3.0 grey values. The maximum difference is 37.6 grey values and appears at the boundary of an edge. It means that the boundary of the discontinuity is shifted by one pixel between both results. This cannot be recognized by looking at the two images so that the results of both approaches can be assumed to agree.

The grey values shown in Fig. 2 have also been estimated using a systematic scan by the Gibbs sampler as described in section 3. The standard deviation $\sigma = \sqrt{\sigma^2}$ in (3.11) was set equal to 4.0 according to the variations of the grey values in the image. After a burn-in phase of 75 iterations 1000 samples $\boldsymbol{\beta}_i$ were generated by (2.16) with (3.11) for the vector $\boldsymbol{\beta}$ and lead to the estimate $\hat{\boldsymbol{\beta}}$ according to (2.11). The square root of the mean squared differences between these results and the ones of Fig. 2 equals 3.4 grey values while the maximum difference is 35.7. It appears at the boundary of an edge. This cannot be recognized by looking at the two images so that the results can be assumed to agree.

Finally, the Metropolis algorithm has been applied starting from the estimate of the Gibbs sampler as explained in section 3. The standard deviation for the normal distribution of the random walk was set to 0.0008 and 50 000 iterations were computed. Despite of the small standard deviation the rate for accepting a proposal to rejecting it was only 0.065. The acceptance rate for an efficient jumping rule for the Metropolis algorithm in high dimensions is about 0.23, cf. Gelman et al. (2004, p. 306), so that a smaller standard deviation could have been chosen which would result, however, in a slower convergence. During the 50 000 iterations the proposals were accepted for $r \ge 1$ in (2.12) in 21 130 cases and for $r < 1$ for the rest of the cases. This indicates that the Metropolis algorithm was in a state of equilibrium with the posterior density function (3.8) as limiting distribution. The square root of the mean squared differences between the estimates of the Gibbs sampler and of the Metropolis algorithm is 0.10 grey values with the maximum difference being 1.05. Thus, also the Gibbs sampler has converged to the posterior density function (3.8).

5 Conclusions

By assuming a Gibbs distribution as prior based on a modified density of Huber for a robust parameter estimation the fast algorithm ICM is derived. Controlled by two parameters it gives a smooth image where the edges are well preserved. The numerical results of the method ICM agree – except for minor deviations – with the Bayes

Fig. 1: Original Image

Fig. 2: Image Smoothed with $c = 10.0$ and $c_\beta = 1.0$

Fig. 3: Image Smoothed with $c = 21.0$ and $c_\beta = 1.0$

Fig. 4: Image Smoothed with $c = 21.0$ and $c_\beta = 0.5$

estimates of the Gibbs sampler and the Metropolis algorithm. It can therefore be concluded that the algorithm ICM does not give an approximate solution but the Bayes estimate which for the linear model is identical with the MAP estimate.

Acknowledgement

The author is indebted to B. Gundlich and to an anonymous reviewer for their valuable comments.

References

Besag, J. E.: Spatial interaction and the statistical analysis of lattice systems. J Royal Statist Society, B 36: 192–236, 1974.

Besag, J. E.: On the statistical analysis of dirty pictures. J Royal Statist Society, B 48: 259–302, 1986.

Busch, A., Koch, K. R.: Reconstruction of digital images using Bayesian estimates. Z Photogrammetrie und Fernerkundung, 58: 173–181, 1990.

Fessler, J. A., Erdogan, H., Wu, W. B.: Exact distribution of edge-preserving map estimators for linear signal models with Gaussian measurement noise. IEEE Trans Im Proc, 9(6): 1049–56, 2000.

Gelfand, A. E., Smith, A. F. M.: Sampling-based approaches to calculating marginal densities. J American Statistical Association, 85: 398–409, 1990.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B.: Bayesian Data Analysis, 2nd Ed. Chapman and Hall, Boca Raton, 2004.

Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Machine Intell, PAMI–6: 721–741, 1984.

Geman, S., McClure, D. E.: Statistical methods for tomographic image reconstruction. Bull Int Statist Inst, 52-21.1: 5–21, 1987.

Gundlich, B., Koch, K. R., Kusche, J.: Gibbs sampler for computing and propagating large covariance matrices. J Geodesy, 77: 514–528, 2003.

Huber, P. J.: Robust estimation of a location parameter. Annals Mathematical Statistics, 35: 73–101, 1964.

Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P.: Optimization by simulated annealing. Science, 220: 671–680, 1983.

Koch, K. R.: Parameter Estimation and Hypothesis Testing in Linear Models, 2nd Ed. Springer, Berlin, 1999.

Koch, K. R.: Einführung in die Bayes-Statistik. Springer, Berlin, 2000.

Koch, K. R.: Determining the maximum degree of harmonic coefficients in geopotential models by Monte Carlo methods. Studia Geophysica et Geodaetica, 49: 259–275, 2005.

Koch, K. R., Kusche, J., Boxhammer, C., Gundlich, B.: Parallel Gibbs sampling for computing and propagating large covariance matrices. zfv – Z Geodäsie, Geoinformation und Landmanagement, 129: 2–42, 2004.

Koch, K. R., Schmidt, M.: Deterministische und stochastische Signale. Dümmler, Bonn, 1994.

Liu, J. S.: Monte Carlo Strategies in Scientific Computing. Springer, Berlin, 2001.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. L., Teller, A. H., Teller, E.: Equation of state calculations by fast computing machines. J Chem Phys, 21: 1087–1092, 1953.

Author's adress

Prof. Dr.-Ing., Dr.-Ing. E. h. mult. Karl-Rudolf Koch (em.)
Institute of Theoretical Geodesy, University of Bonn
Nussallee 17, D-53115 Bonn, Germany
koch@theor.geod.uni-bonn.de