ON THE CHOICE OF THE REGULARIZATION PARAMETER: THE CASE OF BINARY IMAGES IN THE BAYESIAN RESTORATION FRAMEWORK

J. M. Dinten* X. Guyon† J. F. Yao†

ABSTRACT

We study the problem of the influence and of the choice of the regularization parameter β in the Bayesian image restoration framework. Binary and geometrically regular images are examined. The noise degradation process which leads to the observed record can be either additive Gaussian or a binary symmetric channel noise of transmission. MAP is not robust with respect to β, MPM and ICM are more robust. For the three methods, a good choice of β depends strongly on the noise level. On the basis of the observed record, two possible choices of β are examined: if the statistical one seems reasonable at a low noise level, it isn't the case for higher noise for which the cross-validation criterion still gives good results.

* C.N.R.S Statistique Appliquée UA 743, Université Paris-Sud, Bât. 425, 91405 ORSAY Cedex, France and C.E.A Centre d'Etude de Vaujours, BP N^o 7, 77181 COURTRY, France

† C.N.R.S Statistique Appliquée UA 743, Université Paris-Sud,Bât. 425, 91405 ORSAY Cedex, France

Keys words: Bayesian restoration of images,Regularization methods, Choice of regularization parameter

A.M.S subject classification: Primary, 60 G 35, 62 M 20; Secondary, 62 M 30, 62 M 99

1. Introduction: the context of our study

Let $x = \{x_s, s \in S\}, S = \{1, 2 \ldots, n\}^2$ an image to be reconstructed from an observation $y = \{y_s, s \in S\} = \Phi(x, \eta), \Phi$, the noise mechanism, know as well as the noise η. Suppose that $E(x)$ is an appropriate energy on the configuration x which summarizes the prior information about it and $D(y, x)$ is some fidelity distance between y and x. Then classical regularization methods estimate x by $\hat{x}(\beta) = \arg \text{Min}_x H(x : y; \beta)$, where the posterior energy H depends on a so-called regularization parameter β as following

$$H(x : y; \beta) = D(y, x) + \beta E(x)$$

Standard M.A.P restoration ([6]) are exactly of this kind with $(-H)$ the logarithm of the posterior probability $\Pr(x : y, \beta)$, β being the parameter of a prior law on x. M.P.M method ([11]) is also defined from such a scheme by choosing, in each pixel $s : \hat{x}_s = \arg \text{Max}_{x_s} \Pr(x_s : y; \beta)$, where this marginal probability in x_s is deduced from H. The prior E can be chosen as a Markovian energy, an entropy or a regularization function as for example, curvature in the spline smoothing context. The distance D is always directly derived from both the degradation process Φ and the noise η.

A crucial question is: *how to choose the regularization parameter β on the basis of the record y ?* Our paper will give some hints for such a choice based on experimental results in a specific and well defined context. We first examine the dependency of the reconstruction $\hat{x}(\beta)$ in β (Section 2) as well as some choice for β like statistical estimation , cross-validation choice and joint M.A.P estimate on (x, θ)—θ is the parameter of the y–model (Section 3). This will be studied for two kinds of binary images: the first one being the realization of some Markov random field (M.R.F); the other one a hand-drawn picture. The noise degradation can be a binary symmetric channel (B.S.C) noise of transmission, an additive Gaussian one or a textural variance noise.

By *well defined context* we mean that we are able to give some answers to the following questions:

1. What particular family is being studied in the large botanic of images?
2. For a known degradation model (Φ,η), what is the level of the noise: low, medium or high?
3. What is the quality criterion for restoration?
4. What is the regularization method used?

Without reasonable answers to such questions, there can be no reasonable results, experimental or theoretical, to the problem of choice of the regularization parameter.

Test images are chosen to be binary and regular in a natural geometric sense. In particular we will discuss the influence of the second and the fourth questions above on the problem. The noise level remain high giving from 20% to 40% error rate in the pixel by pixel maximum likelihood restoration. The mean percentage of pixel classification error rate is our restoration criterion (we note it by PEC through all the paper) and M.A.P, M.P.M and I.C.M methods will be experimented.

In the literature, theoretical results are obtained in the following two situations:

- regularization by spline functions ([4],[14],[18]) or approximate solutions for integral equations of the first kind ([9],[16],[17])
- smoothing techniques in image restoration in a $\mathcal{L}^2$ framework ([8]).

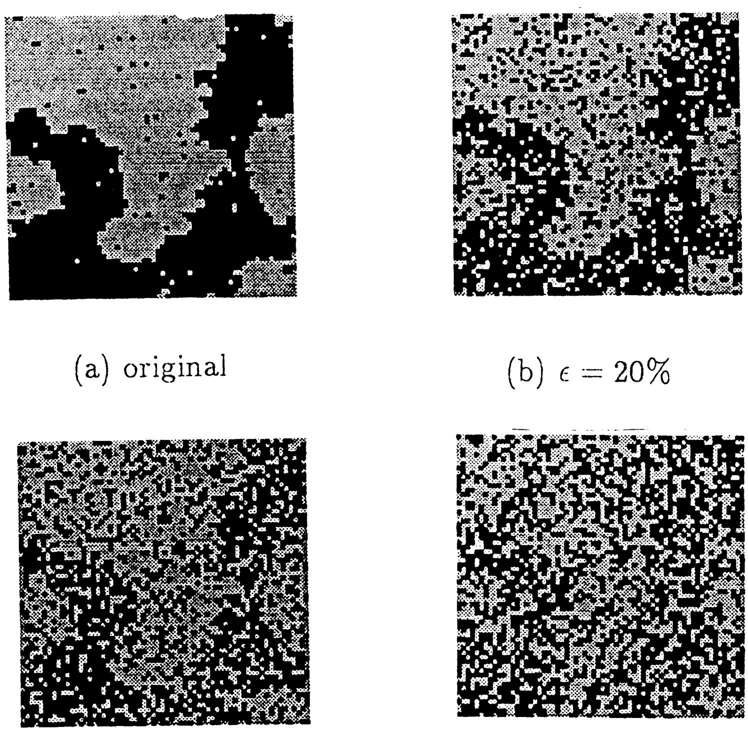
In each of these situations, the mathematical context is well defined and theoretical answers can be derived to help us in the choice of the regularization parameter: cross-validation choice (or a more easily computable variant) in the first context; for the second one where it is assumed that the variance σ^2 of the noise is low, one must choose the regularization parameter to be proportional to σ^2 whereas all classical choices suggest to take it proportional to σ which is too regularizing.

Let us describe now more precisely the object x and its noisy observation y. Two *true* images x are to be reconstructed:

I1 The first one is the realization of a binary isotropic M.R.F (with $H(x) = \beta E(x)$):

$$H(x) = -\beta \sum_{<s,t>} \delta(x_s, x_t) \tag{1}$$

Here the neighborhood system $\{< s,t >\}$ is the four nearest one.



(a) original (b) $\epsilon = 20\%$

(c) $\epsilon = 30\%$ (d) $\epsilon = 40\%$

Figure 1. Original picture 64 × 64 realization of a M.R.F. with β_0 = 1.149 and B.S.C. noisy image.

Figure (1-a) gives such a realization on a 64×64 lattice, using free boundary conditions with $\beta = \beta_0 = 1.149$ and the Gibbs sampler ([6]). This one is run during 80 raster scans starting at a 0–1 uniform i.i.d configuration. Figure (3-a) gives another 128×128 realization for $\beta_0 = 1.40$.

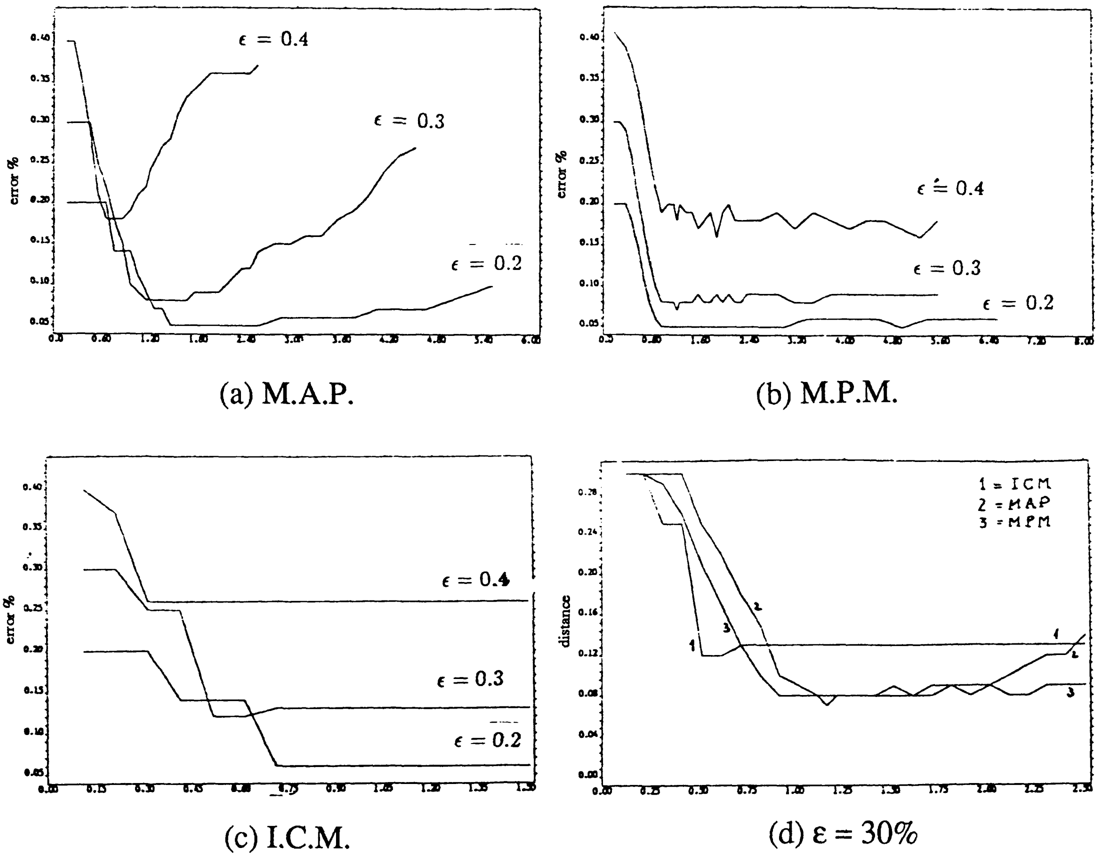


Figure 2. The curves of $PEC(\beta)$ corresponding to several levels of noise, for MAP, MPM, and ICM.

I2 The second one is a hand-drawn (looking like a β) 32×32 image given in figure (4-a)[1]

The noise degradation is assumed pixel independent:

$$\Pr(y : x) = \prod_{s \in S} \Pr(y_s : x_s)$$

---

[1] For reconstruction, the same energy (1)) is used.

We will examine the following three kinds of noise:

N1 Binary symmetrical channel noise (B.S.C.,[11]):

$$\forall s \in S, \quad \Pr(y_s : x_s) = \begin{cases} \varepsilon & \text{if } y_s \neq x_s \\ 1-\varepsilon & \text{if not} \end{cases} \tag{2}$$

Experiments are realized with $\epsilon = 0.2$, 0.3 and 0.4.

N2 Additive Gaussian noise:

$$y = x + \eta$$

Here η is a Gaussian white noise with mean zero and variance σ^2. σ is chose as $0.594, 0.953$ and 1.974 which corresponds to error rate $20\%, 30\%$ and 40% in maximum likelihood classification.

N3 Texture of variance noise :

$$\forall s \in S, \quad \Pr(y_s : x_s) \sim \mathcal{N}(0, \sigma_{x_s}^2) \tag{4}$$

Let us define $u = \sigma_1^2/\sigma_0^2$ and suppose that it is greater than 1. For a fixed σ_0^2, the smaller u, the greater the noise. With $\sigma_0^2 = 1.0$ and $u = 2, 4, 8$, the M.L classification error rate are respectively $46\%, 38\%$ and 30%.

2. Influence of the regularization parameter β on the restored image

Using the prior (1) for x and the noise degradation (2)–(3)–(4) we obtain the three posterior energies $H_i(x : y)$, $i = 1, 3$:

$$\begin{cases} H_i(x : y) = -\alpha_i \sum_s (2y_s - 1)x_s - \beta \sum_{<s,t>} \delta(x_s, x_t), \quad i = 1, 2 \\ \alpha_1 = \ln \dfrac{(1-\epsilon)}{\epsilon}, \quad \alpha_2 = \dfrac{1}{2\sigma^2} \\ H_3(x : y) = -\dfrac{1}{2} \sum_s x_s \left[\ln \dfrac{\sigma_0^2}{\sigma_1^2} + y_s^2 \left(\dfrac{1}{\sigma_0^2} - \dfrac{1}{\sigma_1^2} \right) \right] - \beta \sum_{<s,t>} \delta(x_s, x_t) \end{cases} \tag{5}$$

On the basis of such posterior probabilities, MAP, MPM and ICM are used to reconstruct the image x. For the MAP and to avoid the dependency on a particular cooling schedule, we use the exact maximization algorithm proposed in [12] for binary images. Referring to this work, we can see that a practical simulated annealing can differ strongly from the exact MAP and that it is also strongly dependent on β (see the test image **A** of [12]). The MPM solution is obtained via the Gibbs sampler after fifty raster scans (with the first ten sweeps run to reach stationarity). The ICM restoration is obtained after 8 sweeps.

2.1. Markov random field image

2.1.1. M.R.F image 64×64 for $\beta_0 = 1.149$ (see figure (1-a))

Experiments carried out in ([11]) prove that for low noise level, MAP and MPM for $\beta = \beta_0$ give good restorations[2]. For higher noise level, it seems that such property

[2] In [11] MAP restorations are obtained via simulated annealing with a logarithmic schedule.

is not preserved and we have performed experimentations to examine the subjective criterion PEC (β) in β.

First note that, based on the true original image (1-a), pseudo-maximum likelihood estimation ([1],[7]) for β gives $\hat{\beta}_{PML} = 1.01$ and the stochastic algorithm for maximum likelihood estimation proposed by ([21],[22]) gives $\hat{\beta}_{ML} = 0.92$. From a B.S.C noisy record with the error rate $\epsilon = 30\%$, the Gibbsian E.M. algorithm developed by B. Chalmond([3]) gives $\hat{\beta}_{EM} = 1.00$ (and $\hat{\epsilon} = 0.28$). Such results shows that x is in accordance with a Markovian realization of model (1) for $\beta_0 = 1.149$.

Figure 1.b-c-d are noisy records y for B.S.C noise levels $20\%, 30\%$ and 40%.

The curves shown in Figure 2.a-b-c illustrates the variations of the PEC in β for the M.A.P, the M.P.M and the I.C.M respectively. Each point of the curve was a mean point based on five independent realizations of the noise.

For the Figure 2-b (M.P.M method) a stochastic behavior of PEC (β) appears: this is a consequence of the Monte-Carlo algorithm used to compute the marginal mode and based on a relatively little number of iterations (50 in fact). Such behavior fluctuation doesn't exist in Figure 2-a or 2-c, because for the M.A.P and the I.C.M the performed algorithms are deterministic. Figure 2-d gives, for $\epsilon = 30\%$, PEC for the three methods.

For the I.C.M reconstruction, a plateau phenomenon appears (See Appendix A): if β exceeds some threshold $\beta(\epsilon)$ depending on the noise level ϵ, the I.C.M and the Iterated Modal Filtering are equivalent. Such a behavior can also be observed for the M.P.M.

Figures 3.a-b-c give the M.A.P, M.P.M and I.C.M reconstructions of the noisy image 1-d with the (subjective) optimal choice of β whereas Figures 3.d-e-f give such reconstructions for $\beta = \beta_0$. Table 1 gives the (subjective) optimal choices for each method and each noise

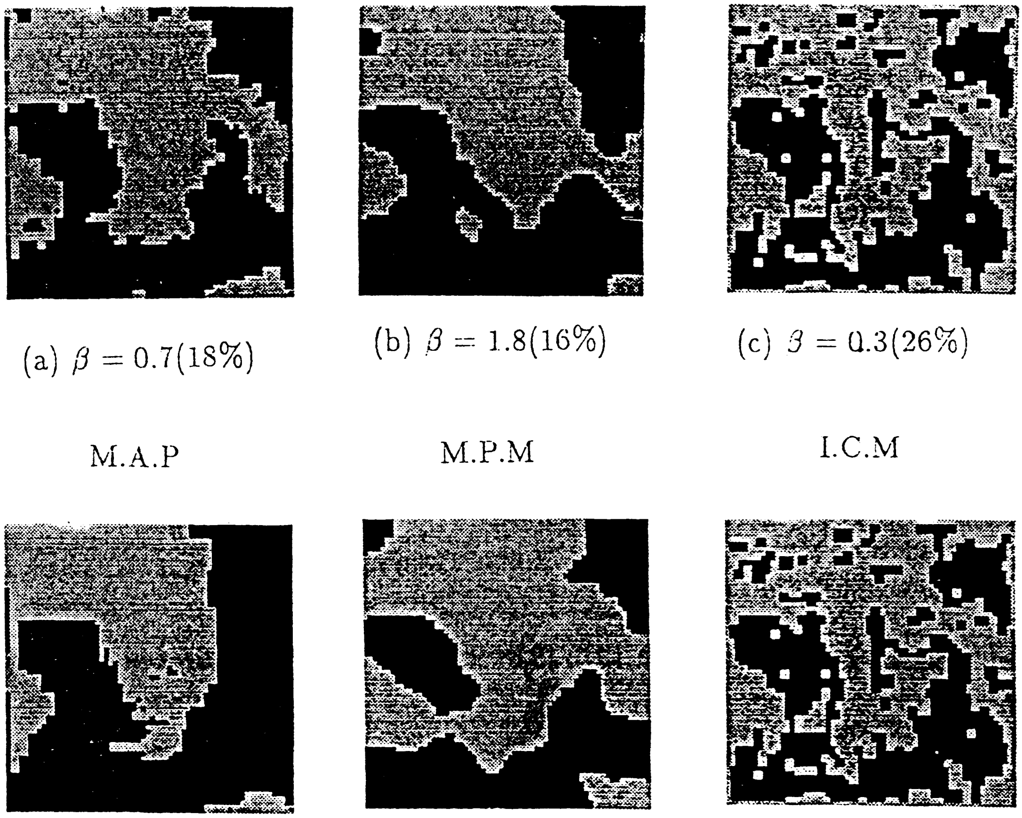
Method	$\varepsilon = 0.2$	$\varepsilon = 0.3$	$\varepsilon = 0.4$
ICM	$[0.7, \infty)$	$[0.5, \infty)$	$[0.3, \infty)$
MAP	$[1.4, 2.5]$	$[1.1, 1.6]$	$[0.6, 0.8]$
MPM	$[1.1, 1.7]$	$[1.1, 1.7]$	$[1.5, 1.6]$

Table 1. Optimal values for β

2.1.2. M.R.F. Image 128×128 for $\beta_0 = 1.4$ (See Figure 4-a)

A similar study was done for a 128×128 realization of the M.R.F with energy function (1) and $\beta_0 = 1.4$. Figure 4.b is the noisy image y with B.S.C noise level $\epsilon = 40\%$. Figures 4.c-d give the optimal M.A.P and M.P.M reconstruction whereas Figures 4.e-f are the same restorations for $\beta_0 = 1.4$. Estimations on the basis of x give $\hat{\beta}_{PMV} = 1.42$, $\hat{\beta}_{ML} = 1.19$.

Experiments on y based on other independent realizations of the noise lead to the same results. The PEC curves in β are given in Figure 5.



(a) $\beta = 0.7(18\%)$ (b) $\beta = 1.8(16\%)$ (c) $\beta = 0.3(26\%)$

M.A.P M.P.M I.C.M

(d) $\beta = \beta_0(24\%)$ (e) $\beta = \beta_0(18\%)$ (f) $\beta = \beta_0(26\%)$

Figure 3. Restorations of image 1-d ($\epsilon = 40\%$): *(a-c)* with $\beta_{optimal}$; *(d-f)* with β_0. For each restoration, the *PEC* is given in brackets.

On this graphic, we have also show the variation of another (more contextual) criterion, the percentage of misclassified windows:

$$\mathrm{PEC}_2(\beta) \stackrel{\text{def}}{=} \frac{1}{c}\sum_{s\in S}\delta(x_{W_s}, \hat{x}_{W_s})$$

where $\{W_s\}$ is the nearest neighbor system and c a normalizing constant.

As can be seen here, the two optimal choices relative to both criteria are very close.

From these observed results,one can note that

- The M.A.P restoration is strongly sensible to β, all the more if the noise is high. The (subjective) optimal value of β (see Table 1) is decreasing when the noise increases. At the "error rate" $\epsilon = 40\%$, this value becomes significantly less than to β_0.

- For the MPM, optimal value β_{MPM} of β are nearer to β_0 and for $\beta \geq \beta_{MPM}$, there is a plateau phenomenon giving good robustness in β in the following sense: too large choices for β are not very dangerous.

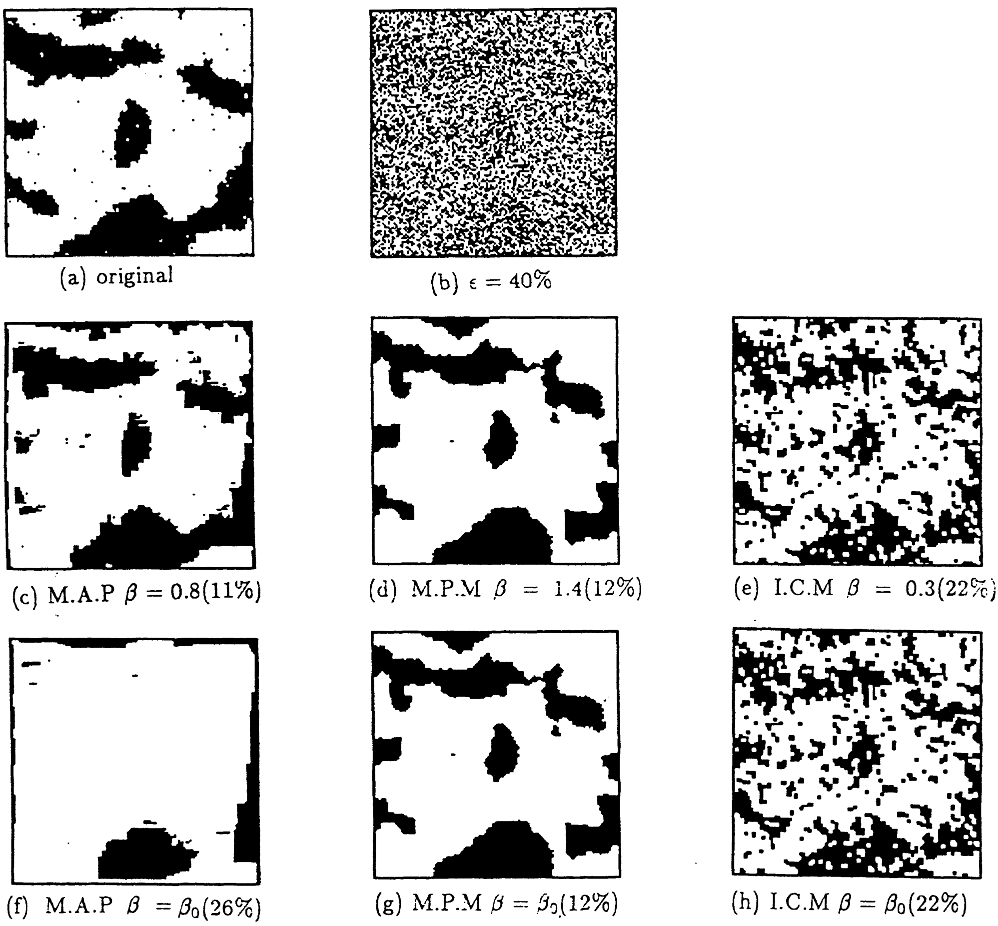


Figure 4. *(a)* Original image: 128 x 128 M.R.F. with $\beta_0 = 1.4$. *(b)* B.S.C. record with $\epsilon = 40\%$; *(c-e)* with optimal choice of β; *(f-h)* with β_0

- An accentuation of this phenomenon appears for the ICM, the plateau for the PEC function begin at a threshold β_{ICM} which is independent of β_0 for B.S.C noise ($\beta_{ICM} = \alpha/2$, see Appendix A).

- For the three methods(M), the optimal value $\beta_M(\epsilon)$ is decreasing in ϵ. We have also noted (see fig.5), using two distinct criteria, that the dependency on these criteria is not strong.

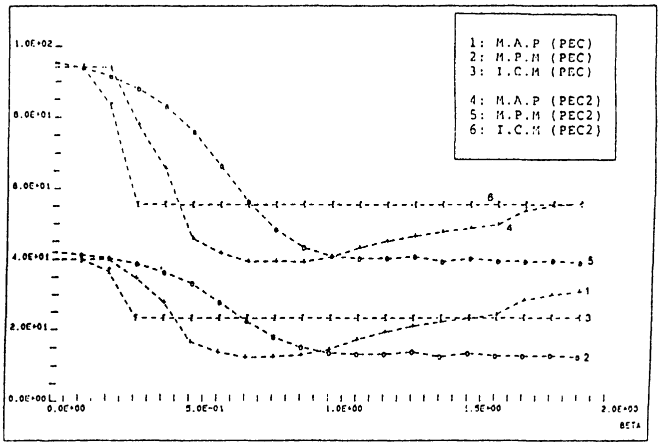


Figure 5. Restoration of the 128 × 128 record 4(b). The curves of $PEC(\beta)$ and $PEC_2(\beta) + 0.25$.

2.2. Hand-drawn image 32 × 32 (see Fig.6-a)

We have performed various estimations for the parameter β of the a priori energy(1) for this image: from x itself we obtained:

$$\hat{\beta}_{\mathrm{PML}} = 2.74 \quad \hat{\beta}_{\mathrm{ML}} = 0.86 \quad \hat{\beta}_{\mathrm{POS}} = 1.58$$

where $\hat{\beta}_{\mathrm{POS}}$ is the logistic estimation proposed by A.Possolo for binary images([13]). From the record y and the Gibbsian algorithm([3]) we obtain:

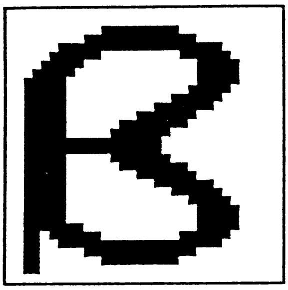
- For a B.S.C at $\epsilon = 20\%$: $\hat{\beta}_{\mathrm{EM}} = 1.15$, $(\hat{\epsilon}_{\mathrm{EM}} = 0.20)$.
- For an additive Gaussian noise,$\sigma = 0.594$: $\hat{\beta}_{\mathrm{EM}} = 0.99$, $(\hat{\sigma}_{\mathrm{EM}} = 0.565)$.

As it was observed for M.R.F Images, the structure of the noise is very well recovered by the E.M estimation, but here, the four estimations proposed for β are widely scattered. Two reasons can explain this: the small size of the image, but also some inadequacy of this image to be the realization of a M.R.F driven by (1).

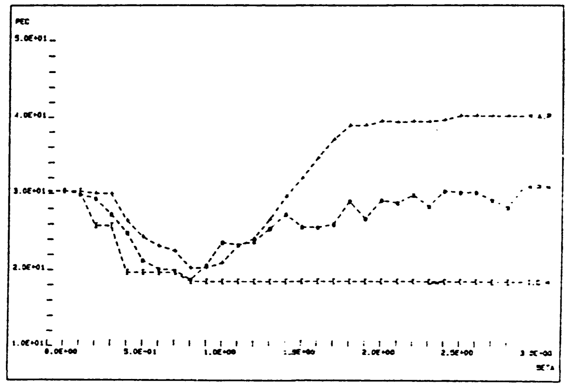
2.2.1. B.S.C Noise degradation

Figures 6.b-c-d give PEC (β) at the three levels 20%, 30%, 40% . For $\beta \geq \beta_{\mathrm{ICM}}(\epsilon)$, ICM is equivalent to Iterative Modal Filtering and superior to any MAP or MPM. This phenomenon is specific to this kind of noise as we shall see later for additive Gaussian or variance texture noise. As we have explained before, the fluctuation of the MPM-curves results from the small number of iterations in the Monte Carlo algorithm.

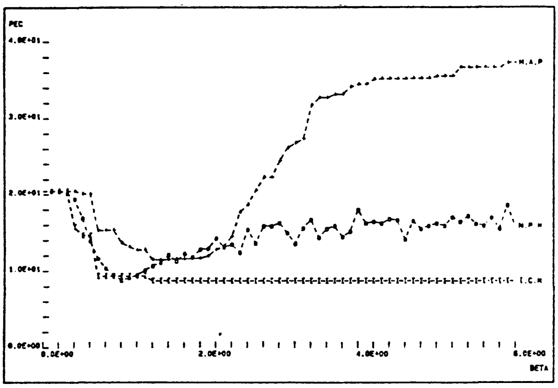
$\beta_{\mathrm{MPM}}(\epsilon)$, the optimal choice of β , has little sensitivity to ϵ whereas $\beta_{\mathrm{MAP}}(\epsilon)$ has a great sensibility in ϵ (Table 2)



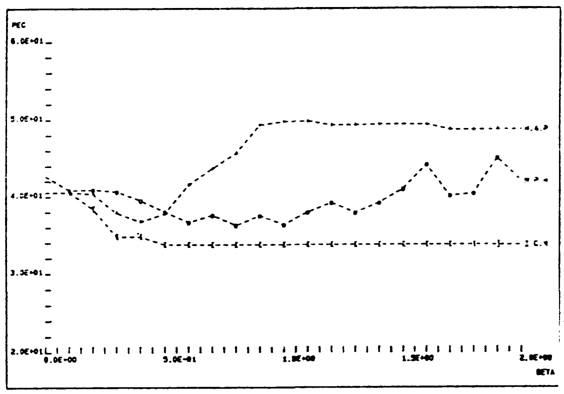
(a) original



(c) $\epsilon = 30\%$



(b) $\epsilon = 20\%$



(d) $\epsilon = 40\%$

Figure 6. The original hand-drawn image and the curves of $PEC(\beta)$ (B.S.C. noise).

Method	B.S.C. Noise			Gaussian noise		
	$\varepsilon = 0.2$	$\varepsilon = 0.3$	$\varepsilon = 0.4$	$\sigma = 0.594$	$\sigma = 0.953$	$\sigma = 1.974$
MAP	1.7	0.9	0.4	1.6	1.3	0.5
MPM	1.0	0.9	0.8	1.1	0.9	0.7

Table 2. Optimal β

2.2.2. Additive Gaussian noise:

For additive Gaussian noise at level $20\%, 30\%, 40\%$ ($\sigma = 0.594, 0.953, 1.974$ respectively) figures 7.a-b-c give the PEC in β. Here, MPM at β_{MPM} is optimal, there is still

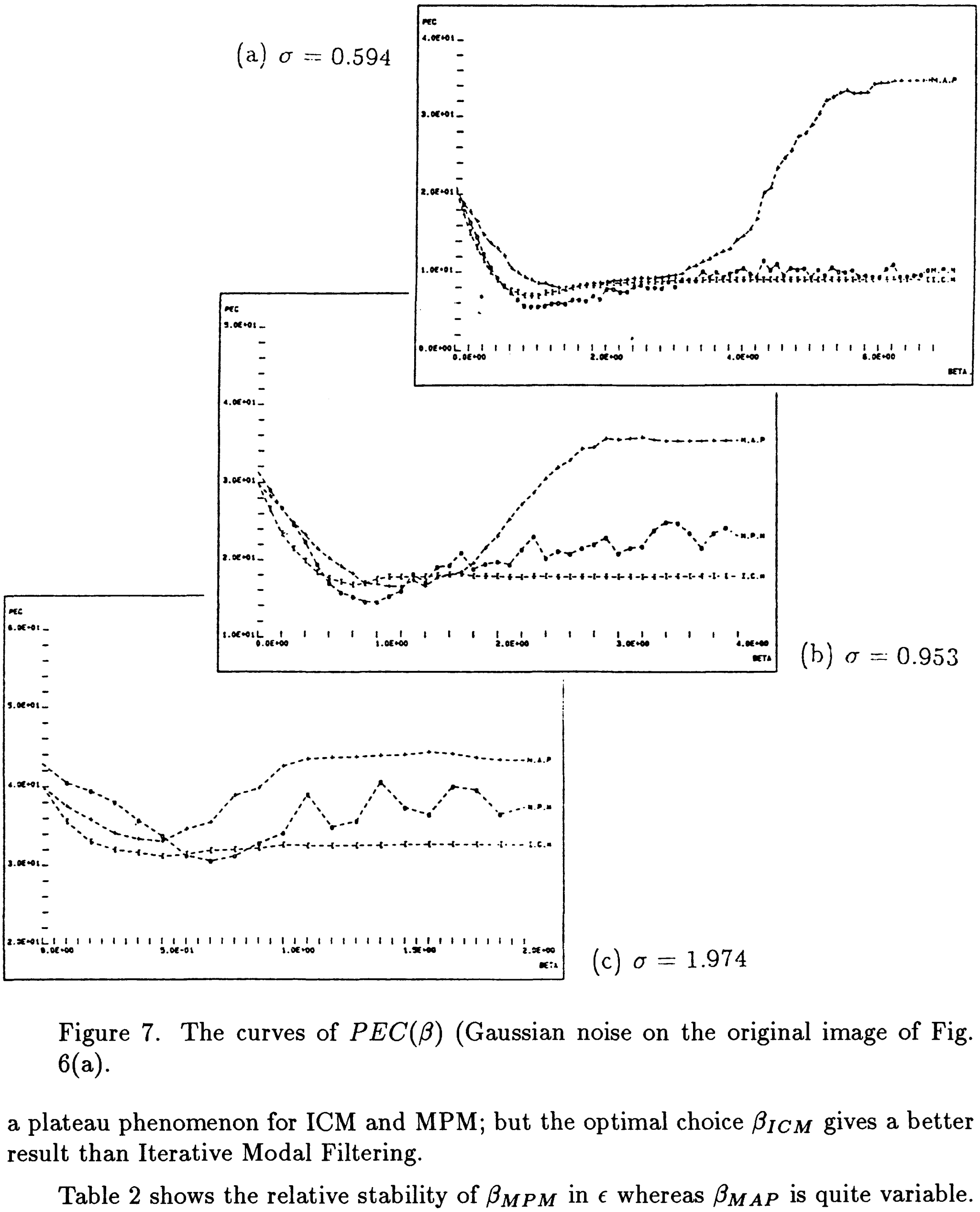


Figure 7. The curves of $PEC(\beta)$ (Gaussian noise on the original image of Fig. 6(a).

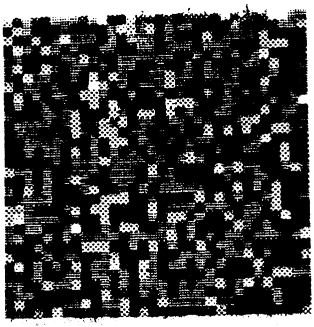
a plateau phenomenon for ICM and MPM; but the optimal choice β_{ICM} gives a better result than Iterative Modal Filtering.

Table 2 shows the relative stability of β_{MPM} in ϵ whereas β_{MAP} is quite variable.

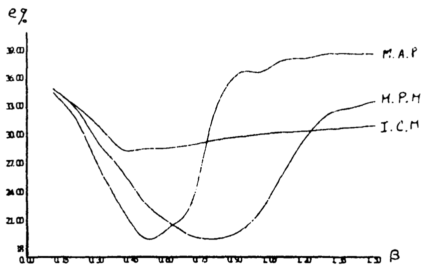
2.2.3. Variance texture noise:

Figure 8-a gives such a degradation at level $u = \sigma_1^2/\sigma_0^2 = 4$ ($\sigma_0^2 = 1$) and curves 8.b-c give respectively PEC in β for the MAP at various levels and PEC at level $u = 4$ (error rate 35%) for MAP, MPM and ICM.

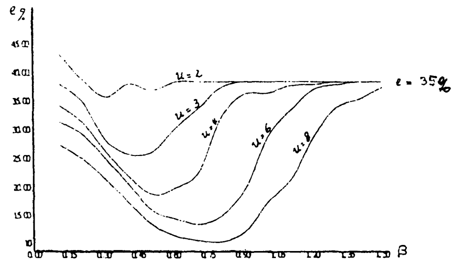
If $\beta > 1.50$ all MAP restorations are uniformly white: this is a negative effect of the strong spatial regularization property of the MAP restoration if β is too large. For MPM, robustness in β is better than for MAP but it is not very good indeed; ICM is still robust for $\beta > \beta_{ICM}$ but its quality is relatively poor. (See also [20] where comparison



(a) noisy image ($u = 4$)



(b) $PEC(\beta)$ curves of M.A.P



(c) $PEC(\beta)$ curves of M.A.P, M.P.M, I.C.M ($u = 4$)

Figure 8. *(a)* Variance texture record of the image of Figure 6(a), and *(b-c) PEC* curves for various reconstructions and various noise levels u.

between global and local Bayesian restoration is studied).

3. On the choice of the regularization parameter β

We shall examine here three possible choices for β on the basis of the record y: 1. the statistical choice; 2. the cross-validation choice; 3. the jointly MAP choice on (θ, x) for both θ (the model parameters : $\theta = (\epsilon, \beta)$ or (σ^2, β)) and x .

3.1. Statistical choice:

We have first used the Gibbsian E.M algorithm([3]): in this algorithm, the pseudo-likelihood([1],[7]) of the Gibbsian field is substituted to the exact likelihood and in the E-step, expectations are calculated by the Monte-Carlo method with simulations given through the Gibbs sampler. These estimation results are the following:

- M.R.F 64 x 64 image (see Fig. 1-a, $\beta_0 = 1.149$) and for B.S.C noise:

$$\begin{aligned} &\text{for} \quad \epsilon = 0.02 \quad (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.28, 0.03) \\ &\text{for} \quad \epsilon = 0.20 \quad (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.15, 0.20) \\ &\text{for} \quad \epsilon = 0.30 \quad (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.00, 0.28) \end{aligned}$$

- Hand-drawn Image 32 × 32 : (see Fig. 6-a)

B.S.C noise

$$\text{for} \quad \epsilon = 0.02 \quad (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.43, 0.02)$$
$$\text{for} \quad \epsilon = 0.20 \quad (\hat{\beta}_{EM}, \hat{\epsilon}) = (1.15, 0.20)$$

Additive Gaussian noise

$$\text{for} \quad \sigma = 0.594 \quad (\hat{\beta}_{EM}, \hat{\sigma}) = (0.99, 0.565)$$
$$\text{for} \quad \sigma = 0.953 \quad (\hat{\beta}_{EM}, \hat{\sigma}) = (0.56, 0.797)$$

For higher level of the noise ($\epsilon = 0.4$ or $\sigma = 1.974$), the Gibbsian E.M doesn't give satisfactory results.

Secondly, moment estimations proposed by A. Frigessi & M. Piccioni([5]) in the B.S.C noise case are used. We have approximated the involved elliptic integrals by polynomials of degree 5 (error uniformly bounded by 10^{-8})[3]. The estimation results are as following:

- M.R.F 64 × 64 image (see Fig. 1-a, $\beta_0 = 1.149$) and for B.S.C noise:

$$\text{for} \quad \epsilon = 0 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.98, 0.00)$$
$$\text{for} \quad \epsilon = 0.20 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.88, 0.16)$$
$$\text{for} \quad \epsilon = 0.30 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (1.02, 0.31)$$
$$\text{for} \quad \epsilon = 0.40 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.94, 0.38)$$

- Hand-drawn Image 32 × 32 : (see Fig. 6-a)

$$\text{for} \quad \epsilon = 0 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.84, -0.04)$$
$$\text{for} \quad \epsilon = 0.20 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.77, 0.11)$$
$$\text{for} \quad \epsilon = 0.30 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.99, 0.32)$$
$$\text{for} \quad \epsilon = 0.40 \quad (\hat{\beta}_{MT}, \hat{\epsilon}) = (0.38, 0.25)$$

Despite the fact that energy(1) is a good measure of geometric regularity, the remoteness of the object x from a realization of a M.R.F can lead , in real images (here the hand-drawn one) to important fluctuations in the various estimations of β. Further, such a statistical choice of β is independent of the restoration method: this is not satisfactory when we take into account the strong dependency of the subjective optimal choice β_M on the method M. Note that other methods of estimation using incomplete data are available (see L. Younes, [21],[22]).

Restoration experiments with these estimated values $\hat{\beta}_M$ and $\hat{\beta}_{MT}$ are given in the Figure 10 and 11.

3.2. Cross-validation choice:

In some sense, this non parametric choice takes into account all the factors defining the context of the problem: the method of reconstruction, the quality criterion, and the process of degradation.

---

[3] See Abramowitz & I. Stegun: Handbook of Mathematical Functions.

Given one pixel $s \in S$ let $\hat{x}^{(s)}$ be the restoration obtained when all the record but the value y_s is observed. The Cross-Validation (C.V) distance is defined by:

$$d_{CV}(\beta) \stackrel{\text{def}}{=} \frac{1}{a} \sum_s (y_s - \hat{x}_s^{(s)})^2$$

where a is a convenient normalizing factor depending only on S (see [15], [4], [9], [14], [16], [17]). The C.V choice for β is:

$$\beta_{CV} \stackrel{\text{def}}{=} \arg \ \mathrm{Min}_\beta \, d_{CV}(\beta)$$

Theoretical results are obtained in some linear and Hilbertian context for the object x (see for example [4]): if we note

$$d(\beta) \stackrel{\text{def}}{=} \frac{1}{a} \sum_s (x_s - \hat{x}_s)^2$$

the distance to the true object x, then under good conditions and when the density of observations increases to infinity, then $d_{CV}(\beta)$ and $d(\beta)$ reach their minimum at "the same value" $\beta = \beta_{CV}$. Note that for a binary image x, $d(\beta)$ and $PEC(\beta)$ are proportional.

When the context for x doesn't stay linear (convex or not convex constraints on x), no theoretical results are available, but for some problems, the C.V choice still seems reasonable (see [17]). The following experiments confirm this opinion: values of PEC (β) and $d_{\mathrm{CV}}(\beta)$ have been computed from $\beta = 0.2$ to $\beta = 2$ with an increments of 0.1 for MAP, MPM, ICM methods, for BSC or additive Gaussian noise at levels 20% and 30%. Table 3 gives values of β minimizing d_{CV} and PEC, showing the proximity of the two determinations.

Method	B.S.C. Noise		Gaussian noise	
ICM	0.70–0.70	0.50–0.50	1.13–1.11	0.76–0.75
MPM	1.21–0.89	0.92–0.69	1.27–1.04	0.95–0.61
MAP	1.66–1.74	1.00–1.10	1.39–1.28	0.82–0.95

Table 3. PEC and **V.C** choices for β

These minima values of β are deduced from the smoothing curves $PEC(\beta)$ and $d_{CV}(\beta)$, obtained themselves by polynomial regressions of degree 3. Figure 9 gives a example of smoothed PEC and d_{CV} curves after a convenient change of scale for d_{CV}.

The set of figure 10 and 11 allows us to say that overall there is a good quality of reconstruction with β_{CV}. Statistic selection lead to important fluctuations in this quality.

Nevertheless, at our stage, C.V. criterion cannot be used due to over lengthy computing time unless the restoration method is very rapid (for example ICM): here, one exact MAP restoration takes 30 seconds (CPU time) on a VAX 750 (less for small values of β), MPM with 50 loops in the Monte-Carlo estimation needs 7 seconds (the same for Simulated Annealing with the same characteristics), ICM needs only 0.2 second.

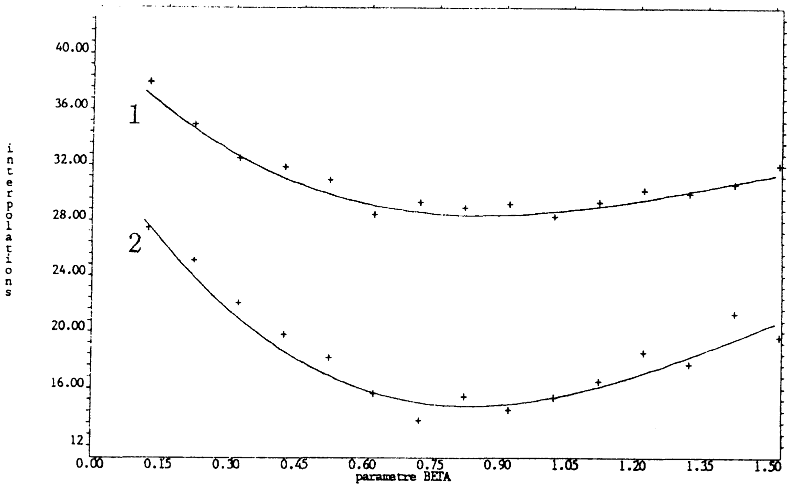


Figure 9. Exact and smoothed d_{CV} (curve 1), and PEC (curve 2). Additive Gaussian noise record ($\sigma = 0.953$) and M.P.M. restoration.

3.3. Joint MAP on both parameter (to be estimated) and object (to be restored):

It is known that, frequently, maximizing the conditional likelihood $\Pr(x : y, \beta)$ in both parameters β and object x leads to some degenerated solutions, see for example ([10],[19]). We shall briefly illustrate this behavior on three examples of filtering (see appendix B for the sketch of the proofs).

Example 1. Let $x = \{x_1, x_2, \ldots, x_n\}$ a binary signal taking values in $\{-1, 1\}$ and $y = \{y_1, y_2, \ldots, y_n\}$ the record: $y_i = x_i + \epsilon_i$ where $\{\epsilon_i\}$ are i.i.d $\mathcal{N}(0, \sigma^2)$ variables. Suppose x has the (one dimensional) prior given by (1). Then the likelihood of the complete data is:

$$\Pr(x, y; \sigma^2) = K(\sigma^2, \beta, y) \exp\left[\beta n(x) + \frac{1}{\sigma^2} < y, x >\right]$$

where $n(x) = \sum_{<i,j>} \delta(x_i, x_j)$ and $< x, y >$ is the scalar product in $\mathcal{R}^n$.

Let $T(y), T(|y|)$ be the sums of the components of y, $|y| = \{|y_1|, |y_2|, \ldots, |y_n|\}$ respectively, $\eta = \inf_i |y_i|$, and denote $s(z)$ the sign of a real z. Let P^y be the conditional likelihood. Then we have the following result:

- If $\sigma^2 \to 0$, $\beta \to \infty$ and $\beta\sigma^2 \geq C > 2T(|y|)$, then for $\hat{x} = \{s, s, \ldots, s\}$ with $s = s(T(y))$, $P^y \to 1$.

- If $\sigma^2 \to 0$ and $\beta \to \infty$ with $\beta\sigma^2 \leq C < 2\eta/n$, then with $\hat{x} = \{s(y_i), i = 1, n\}$, $P^y \to 1$.

In both case, a degeneration appears: the estimated parameter value goes to the boundary of the domain of definition and restoration will be driven uniquely by the regularization component ($\beta\sigma^2$ is large) or by the fidelity to the data y ($\beta\sigma^2$ is small).

The same behavior appears exactly if x is recorded via a B.S.C with parameter α_1 (see (5)): complete regularization if $\beta\alpha_1^{-1} \geq C > n/2$, complete fidelity to the record if $\beta\alpha_1^{-1} \leq C < 2/n$ (α_1^{-1} is increasing with level of noise).

Example 2. Examine now the following filtering of an $AR(1)$ process. The model is:

$$\begin{cases} y_i = x_i + \epsilon_i & \\ & i = 1, 2, \ldots, n \\ x_i = \rho x_{i-1} + \eta_i, & \rho \geq 0 \end{cases}$$

with $\{\epsilon_i\}$, $\{\eta_i\}$ i.i.d variables, and $\epsilon_i \sim \mathcal{N}(0, \tau^2)$, $\eta_i \sim \mathcal{N}(0, \sigma^2)$. The parameter is $\theta = (\tau^2, \sigma^2, \rho)$. We have:

$$P^y = \mathcal{L}(x : y, \theta) = \mathcal{N}_n\left((I + \tau^2\Sigma_x^{-1})^{-1}y, \Sigma_x(I - (I + \tau^2\Sigma_x^{-1})^{-1})\right)$$

from which it can be deduced:

- (ρ, σ^2) being fixed, $\tau^2 \to 0_+$ and $\hat{x} = y$ give the maximum of P^y.
- τ^2 and $\sigma_x^2 \stackrel{\text{def}}{=} \frac{\sigma^2}{1-\rho^2}$ being fixed, $\sigma \to 0_+$ and $\rho \to 1_-$, $\hat{x} = (\overline{y}, \ldots, \overline{y})$ gives the maximum of P^y.

Example 3. This third example is taken from an exercise in the book by Ch. Gourieroux.[4] Let $x = (x_1, x_2, \ldots, x_n)$, $x_i \in \overline{\mathcal{R}}$ with no prior knowledge and independent observations $y_i = \begin{pmatrix} y_{i_1} \\ y_{i_2} \end{pmatrix}$, $i = 1, n$, where $\{y_{i_1}\}$ and $\{y_{i_2}\}$ are independent Bernoulli variables:

$$\begin{cases} \Pr(y_{i_1} = 1) = F(x_i) \\ \Pr(y_{i_2} = 1) = F(x_i + \beta) \quad i = 1, n \\ \text{with } F(x) = (1 + e^{-x})^{-1} \text{ (Logistic distribution)} \end{cases}$$

Then the M.L (estimation $\times$ restoration) for β and x leads to: if $y_{i_1} + y_{i_2} = 0, \hat{x}_i = -\infty$ and if $y_{i_1} + y_{i_2} = 2, \hat{x}_i = +\infty$.

Let $\overline{n}$ be the cardinal of the set $\{i/1 \leq i \leq n, y_{i_1} + y_{i_2} = 1\}$ and $\overline{\overline{y_2}}$ the arithmetic mean of the $\{y_{i_2}\}$ for the preceding index set, then:

- If $y_{i_1} + y_{i_2} = 1, \hat{x}_i = -\hat{\beta}/2$ where $\hat{\beta} = 2\ln\left(\frac{1-\overline{\overline{y_2}}}{\overline{\overline{y_2}}}\right)$ is the joint M.L estimator of β . It is easy to check that $\hat{\beta} \xrightarrow{\text{a.s.}} 2\beta$. Then, if we may consider as satisfactory the restoration of x , we see that it is associated to a biased estimation of the unknown parameter.

[4] Econometrie des variables qualitatives (Ed. Economica, Paris 1984), Ex. 19, p. 109.

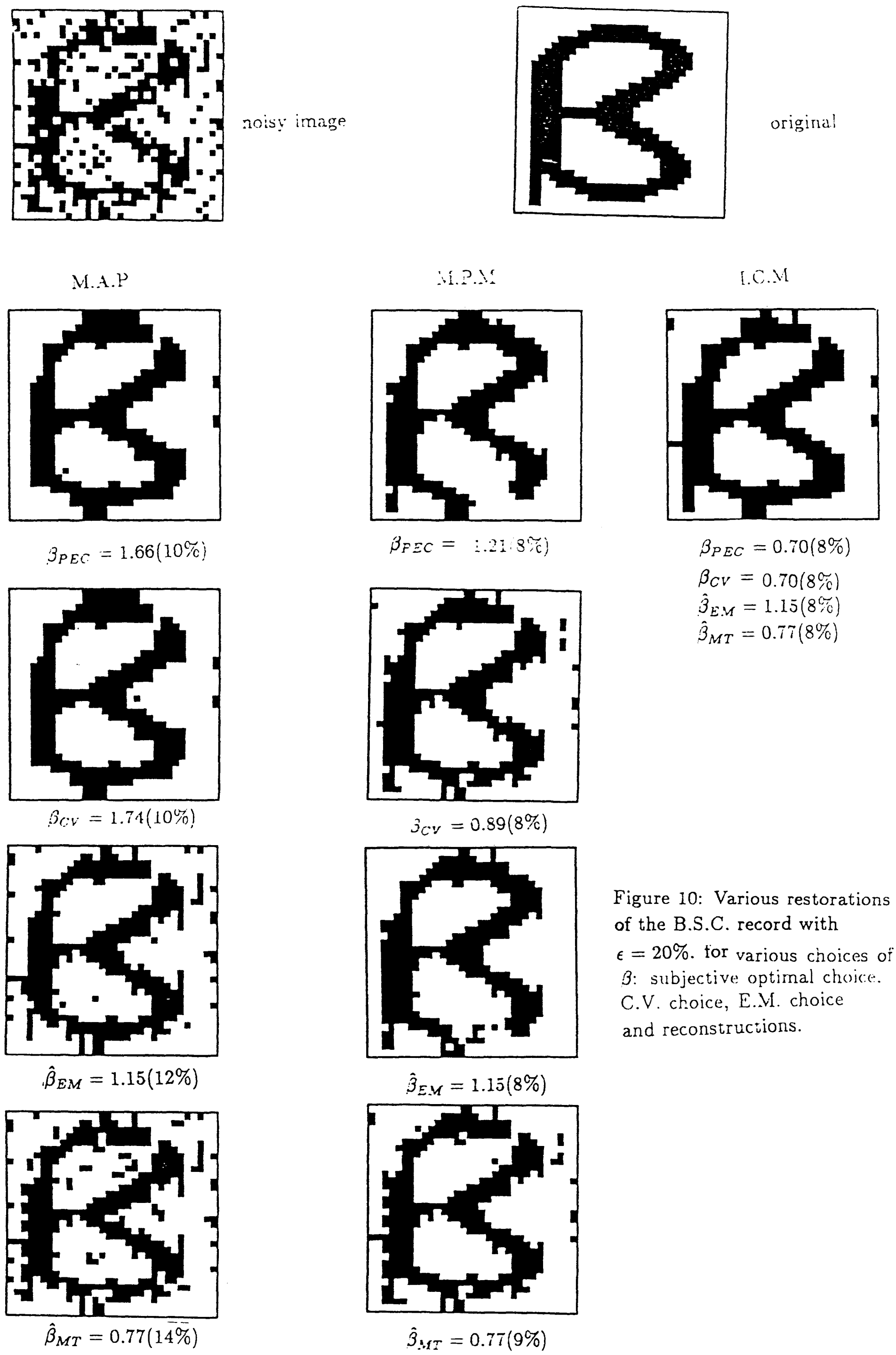


Figure 10. Various restorations of the B.S.C. record with $\epsilon = 20\%$ for various choices of β: subjective optimal choice, C.V. choice, E.M. choice.

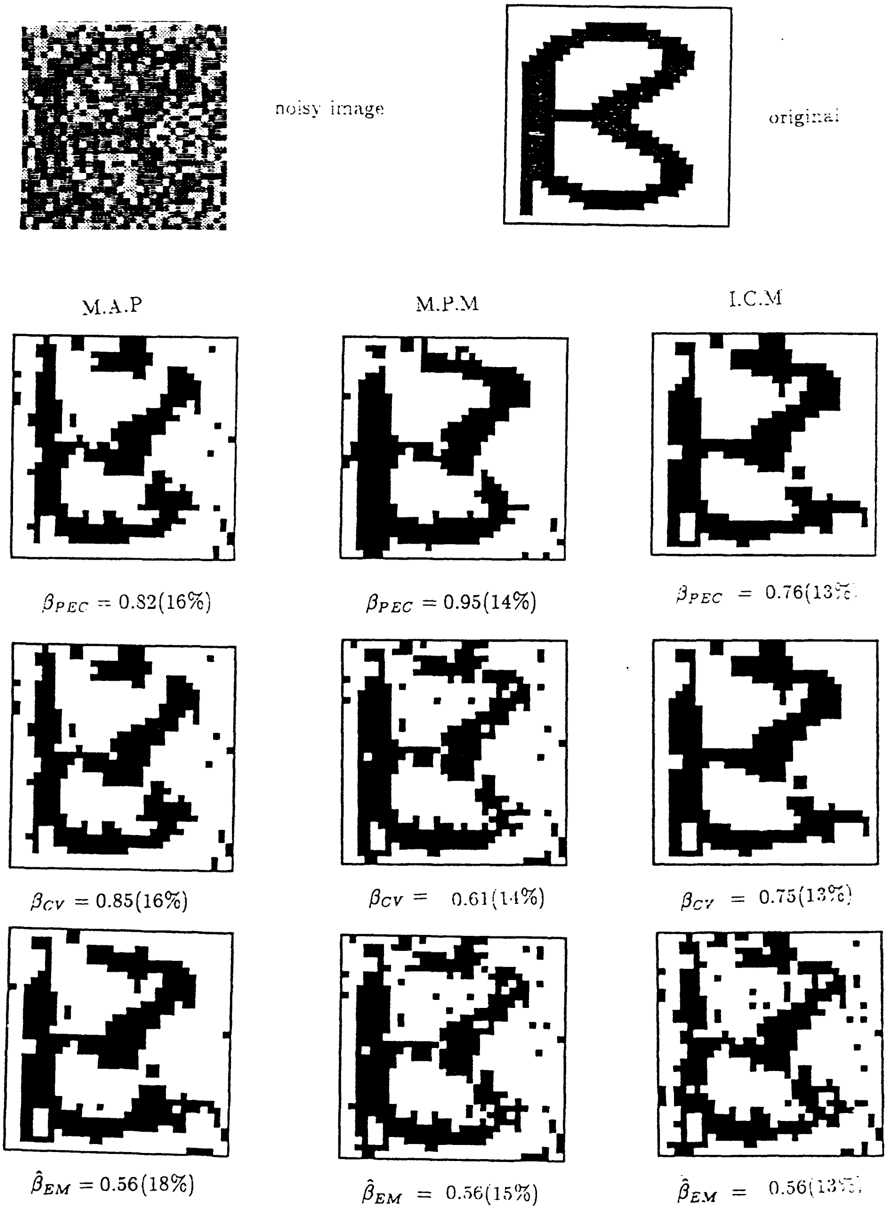


Figure 11: Similar experiments as Fig. 10 for additive gaussian noise with $\sigma = 0.953$.

Figure 11. Similar experiments as in Figure 10, for additive Gaussian noise with $\sigma = 0.953$..

4. Discussion and further work

We have seen that it is difficult to give a general answer to the crucial question: *How to choose the parameter (say β in a regularization reconstruction method?* First, the general context of the problem need to be conveniently defined: in which family is the object to be reconstructed?, what is the level of the noise degradation?, but also the choice of regularization parameter is dependent on quality criterion selected and on the reconstruction method. We have focussed this experimental study on geometrically regular binary image x and record y resulting from a noise degradation process of relatively high level. The robustness in β is decreasing from I.C.M to M.A.P via M.P.M, and it is observed that for high level of the noise, optimal values of β can differ strongly on true (M.R.F images) or estimated (modelling of a non M.R.F image by a M.R.F) β, particularly for M.A.P.

If a statistical criterion (for ex. E.M estimation) of choice of β seems reasonable at low noise level, this is not the case in medium or high noise level. In this case, cross-validation criterion gives satisfactory results, but, at our stage of study, its numerical implementation is too expensive to be used in practice. Next we note that joint M.L. on both parameter and object leads to degenerated solutions.

Future work arises naturally: how relevant are such results to other families of objects?, Are they some theoretical results that validate the C.V choice, in the context of constrained regularization? What are numerically computable variants of the C.V criterion?

APPENDIX

A. Equivalence for large β between I.C.M and the iterative modal filtering

Up to a constant, we have (see (5), $i = 1, 2$):

$$H_i(x : y) = -\alpha_i \sum_s (2y_s - 1)x_s - \beta \sum_{<s,t>} \delta(x_s, x_t)$$

which leads to the local conditional probabilities:

$$\ln P_s(x_s : x_t, t \neq s, y) = a_s + \alpha x_s(2y_s - 1) + \beta \sum_{t:<s,t>} \delta(x_s, x_t)$$

In the I.C.M context, let $x^{(k)}$ be the configuration after the k^{th} iteration and $s = s(k+1)$ be the pixel visited at time $k+1$, then:

$$x_s^{(k+1)} = \arg\ Max_{x_s}\ P_s(x_s : x_t^{(k)}, t \neq s, y)$$

So, if $v = v_s^{(k)} = \sum_{t:<s,t>} x_t^{(k)}$,

$$x_s^{(k+1)} = \begin{cases} 1 & \text{if } v > 2 - \alpha(2y_s - 1)/2\beta \\ 0 & \text{if not} \end{cases}$$

First case: B.S.C noise. Here the above updating rule becomes

$$\text{if } y_s = 1 x_s^{(k+1)} = \begin{cases} 1 & \text{if } v > 2 - \alpha/2\beta \\ 0 & \text{if not} \end{cases}$$

$$\text{if } y_s = 0 x_s^{(k+1)} = \begin{cases} 0 & \text{if } v < 2 + \alpha/2\beta \\ 1 & \text{if not} \end{cases}$$

Then suppose that $\alpha/2\beta < 1$,

$$\text{if } y_s = 1 x_s^{(k+1)} = \begin{cases} 1 & \text{if } v \geq 2 \\ 0 & \text{if not} \end{cases}$$
$$\text{if } y_s = 0 x_s^{(k+1)} = \begin{cases} 0 & \text{if } v \leq 2 \\ 1 & \text{if not} \end{cases}$$

$x_s^{(k+1)}$ is exactly obtained by modal choice. For ϵ =0.2, 0.3 and 0.4 the threshold β_{ICM} are respectively 0.693, 0.424 and 0.202.

Second case: Additive Gaussian noise. Now the record values $\{y_s\}$ don't belong to a finite set as above. However, there is the same plateau phenomenon when $\alpha/2\beta$ is small.

For M.P.M restorations, similar behavior occurs: the reason is that the Gibbs sampler uses the same local conditional laws $\{P_s(x_s : x_t, t \neq s, y)\}$. Then if β is large, the Gibbs sampler leads to the modal choice with strong probability p. In the B.S.C noise case, simple calculations show that for $a \in (0,1)$,

$$\text{if } b \stackrel{\text{def}}{=} \max\left(1/2 \ln \frac{a(1-\epsilon)}{\epsilon(1-a)}, 1/2 \ln \frac{a\epsilon}{(1-\epsilon)(1-a)}\right), \text{ then } \beta \geq b \Longrightarrow p \geq a$$

As example for $\epsilon = 0.3$, $\beta \geq 1.90$ ensures that $p \geq 0.95$.

B. Degeneracy of joint M.L estimate on θ the parameter and x the object

Example 1. Choose β,σ^2 such that $\beta\sigma^2 > 2T(|y|)$: if we suppose that $T(y) > 0$, then $\hat{x} = (1, 1 \ldots, 1) = \hat{x}(\beta, \sigma^2)$ gives the maximum for $\Pr(x : y; \beta, \sigma^2)$. Looking at function ψ of β, σ^2: $(\beta, \sigma^2) \mapsto \Pr(\hat{x} : y; \beta, \sigma^2)$, we have the estimation

$$\psi \geq [1 + \exp(-2T(y)/\sigma^2) + 2^n \exp(-\beta + 2T(: y :)/\sigma^2)]^{-1}$$

Then if $\beta\sigma^2 = C > 2T(|y|)$ and $\sigma^2 \to 0_+$, this probability goes to 1.

Suppose now that $\beta\sigma^2 = C < 2\eta/n$. Then it's easy to see that $\hat{x} = \{s(y_i), i = 1, n\}$ is the (β, σ^2)-MAP estimate. Then for the function ψ

$$\psi \geq [1 + 2^n \exp(n\beta + 2\eta/\sigma^2)]^{-1}$$

So if $\beta\sigma^2 = c < 2\eta/n, \psi \to 1 : \text{as} : \sigma^2 \to 0_+$.

When we are in the presence of a B.S.C noise, we have

$$\Pr(x, y; \alpha, \beta) = K(\alpha, \beta) \exp\{\beta n(x) + \alpha n(x, y)\}$$

with

$$n(x) \stackrel{\text{def}}{=} \sum_{<i,j>} \delta(x_i, x_j) \text{ and } n(x,y) \stackrel{\text{def}}{=} \sum_i \delta(x_i, y_i)$$

If $2\beta\alpha^{-1} > n, \hat{x} = \{a, a \ldots, a\}$: where : $a = s(T(y))$ (the mode of the sequence $\{y_i, i = 1, n\}$) is the MAP in x and the estimation

$$\Pr(\hat{x} : y; \beta, \alpha) \geq [1 + \exp(\alpha[n(-\hat{x}, y) - n(\hat{x}, y)] + 2^n \exp(-\beta + n\delta/2)]^{-1}$$

shows that this probability goes to 1 if $\beta\alpha^{-1} = C > n/2$ and $\alpha \to \infty$.

If $n\beta\alpha < 2, \hat{x} = y$ is the MAP and

$$\Pr(\hat{x} : y; \beta, \alpha) \geq [1 + 2^n \exp(n\beta/2 - \delta)]^{-1}$$

So $\beta\alpha^{-1} = c < 2/n$ and $\alpha \to 0$ gives limit value of 1 to this probability.

Example 2. Straightforward calculation from the Gaussian form of the conditional law $\Pr(x : y)$.

Example 3.

$$\ln \ell(y_{i_1}, y_{i_2}; x_i, \beta) = x_i(y_{i_1} + y_{i_2} - 2) - \beta(1 - y_{i_2}) - \ln[(1 + e^{-x_i})(1 + e^{-(x_i+\beta)})]$$

and so we obtain the announced result if $y_{i_1} + y_{i_2} = 0$ or 2. If $y_{i_1} + y_{i_2} = 1$, maximization in x_i gives: $\exp(-\hat{x}_i) = \exp(\beta/2)$. Then, the informative part for β is:

$$\begin{aligned} \ell(\beta) &= \prod_{i:y_{i_1}+y_{i_2}=1} \ell(y_{i_1}, y_{i_2}; \hat{x}_i, \beta) \\ &= \exp[\beta\overline{n}(\overline{\overline{y_2}} - 1/2)][2 + \exp(\beta/2) + \exp(-\beta/2)]^{-\overline{n}} \end{aligned}$$

and maximization in β gives: $\hat{\beta} = 2\ln(\frac{\overline{\overline{y_2}}}{1-\overline{\overline{y_2}}})$

Examine now the conditional law $\Pr(y_{i_2} : y_{i_1} + y_{i_2} = 1)$

$$\Pr(y_{i_2} : y_{i_1} + y_{i_2} = 1) = \frac{1}{1 + e^{-\beta}}$$

Then, by the law of large number, $\overline{\overline{y_2}} \stackrel{a.s}{\longrightarrow} \frac{1}{1+e^{-\beta}}$ and so : $\hat{\beta} \stackrel{a.s}{\longrightarrow} 2\beta$.

Acknowledgements

The authors are grateful to B. Chalmond and L. Younes for helpful discussions in the use of the Gibbsian E.M. and stochastic M.L. algorithms, and to Sylvia and Marie Yvonne for the revision of the English version of the paper.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society* **B-36**, 192-236.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society* **B-48**, 259-302.

Chalmond, B. (1988). An iterative Gibbsian technique for reconstruction of M-ary images. *Pattern Recognition.*

Craven, P., & Wahaba, G. (1979). Smoothing noisy data by spline functions. *Numerische Mathematik* **31**, 377-403.

Frigessi, A., & Piccioni, M. (1988). Parameter estimation for two-dimensional Ising fields corrupted by a noise. *Quaderno 10*, Inst. Appl. del Calculo, Roma.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, 721-741.

Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblance conditionnelle: étude asymptotique et application au cas markovien. Droesbecke, F. (1987, ed.), *Spatial processes and spatial time series analysis*, Proceedings of the Sixth Franco-Belgian Meeting of Statistics, Bruxelles.

Hall, P., & Titterington, D. M. (1986). On some smoothing techniques in image restoration. *Journal of the Royal Statistical Society* **B-48**, 330-343.

Hilgers, J. W. (1976). On the equivalence of regularization and certain reproducing kernel Hilbert space approximation for solving first kind problems. *S.I.A.M Journal on Numerical Analysis* **13**, 172-174.

Little, R. J. A., & Rubin, D. B. (1983). On jointly estimating parameters and missing values by maximizing the complete data likelihood. *American Statistics* **37**, 218-220.

Marroquin, J., Mitter, S., & Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association* **79**, (387), 584-589.

Greig, D. M., Porteous, B. T., & Seheult, A. H. (1987). Exact Maximum A Posteriori estimation for binary images. *Journal of the Royal Statistical Society* **B-51** (2), 271–279.

Possolo, A. (1986). Estimation of binary Markov Random Fields. *Technical Report* **77**, Department of Statistics, University of Washington.

Silverman, B. W. (1984). A fast efficient cross-validation method for smoothing parameter choice in spline regression. *Journal of the American Statistical Association* **79** (387), 584-589.

Stone, M. (1974). Cross-validatory choice and assessment of statistical prediction. *Journal of the Royal Statistical Society* **B-36**, 111-147.

Wahba, G. (1977). Practical approximate solutions to linear operators equations when the data are noisy. *S.I.A.M, Journal on Numerical Analysis* **14**, 651-667.

Wahba, G. (1982). Constrained regularization for ill-posed linear operator equations with application in meteorology and medicine. Gupta, S., & Berger, J. D. (1982, eds.), *Statistical Design Theory and Related Topics: III*, Volume 2, Academic Press.

Wendelberger, J. (1981). The computation of Laplacian smoothing splines with examples. *Technical Report* **648**. Department of Statistics, University of Wisconsin.

Woodward, W. A, Parr, W. C, Schucany, W. R., & Lindsey, H. (1984). A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of the American Statistical Association* **79**, 590-598.

Yao, J. F. (1988). Segmentation bayésienne d'images: comparaison des methodes contextuelles et globales. *Cahiers du Centre d'études de Recherche Opérationnelle* **30** (4) (Université Libre de Bruxelles), 269–290.

Younès, L. (1988). Estimation and Annealing for Gibbsian Fields. *Annales de l'Institute Henri Poincaré*, 269-294.

Younès, L. (1988). *Problèmes d'estimation paramétriques pour des champs de Gibbs Markoviens. Applications en traitement d'images*, Thèse de Docteur en Sciences, Universite de Paris XI.