Journal of Global Optimization **10:** 91–103, 1997.

Restoration of Degraded Images with Maximum Entropy

DOMINIKUS NOLL
Université Paul Sabatier, Laboratoire Approximation et Optimisation, 118 route de Narbonne, 31062 Toulouse Cedex, France (e-mail: noll@dumbo.ups-tlse.fr)

(Received: 13 November 1995; accepted 19 June 1996)

Abstract. We present a Maximum Entropy based approach to the restoration of degraded images as an alternative to restoration techniques using inverse Wiener filtering. The method we discuss applies in particular to images corrupted by a relatively high system noise. A variety of experimental results supporting our imaging model are included.

Key words: Image restoration, Maximum Entropy, Wiener filtering, deblurring, deconvolution.

1. Introduction

Images recorded in electronic or photographic media are often degraded due to system imperfections such as diffraction effects, optical system aberrations, camera and/or object motions, defocussing, or atmospheric turbulences. In addition to these blurring effects, the recorded image is usually corrupted by various random noise effects such as the randomness of the film grain, photoelectric effects, but also random measurement or transmission errors and a further degradation caused by digitization. The general goal of an image restoration is then to model these degradations in an appropriate way which, given the recorded data, allows for estimating the original signal.

Our recording model described below comprises both spatial degradation caused by blur, and random pixel degradations responsible for the recording noise. We assume a linear system transmission of the form

$$y(i,j) \;=\; \sum_{k=1}^{N} \sum_{\ell=1}^{M} a(i,j;k,\ell)\, x(k,\ell) \;+\; e(i,j), \tag{1.1}$$

$i = 1, \ldots, N$, $j = 1, \ldots, M$, where $y(i,j) \geq 0$ are the gray levels of the observed degraded image, and the $x(i,j) \geq 0$ the gray levels of the unknown original image of size $N \times M$. The linear operator $A = \{a(i,j;k,\ell)\}$ is called the blur, and the $e(i,j)$ represent a signal-independent pointwise mean zero noise process.

In addition to the linearity of the recording system (1.1), we further assume that the total energy of the original object x be preserved during the recording, and that the recording device neither absorbs nor generates optical energy. This leads

to the conditions $a(i,j;k,\ell) \geq 0$ and $\sum_{k=1}^{N}\sum_{\ell=1}^{M} a(i,j;k,\ell) = 1$ for every pixel (i,j). We mention that in practical applications, the blur is often spatially invariant: $a(i,j;k,\ell) =: d(i-k,j-\ell)$, and the function $d(i,j)$ is then referred to as the *point spread function* (PSF), or the mask of the system (1.1), which then has the convolutional form

$$y(i,j) \;=\; \sum_{k=1}^{N}\sum_{\ell=1}^{M} d(i-k,j-\ell)\,x(k,\ell) \;+\; e(i,j). \tag{1.2}$$

Some standard masks encountered in practical situations are *out-of-focus blur*, *linear motion blur*, and various types of *pillbox blurs*, which are typical, e.g. for photographic images, and *atmospheric turbulence*, which is often apparent in satellite images. For a description of the corresponding masks and other types of blurrings, see e.g. [1, 2, 7, 12, 14, 23, 31].

In practice the additive noise $e(i,j)$ in (1.1) is often modeled as a white noise process with spatially invariant variance σ_e^2, i.e., $\mathrm{cov}(e) = \sigma_e^2 I$, but our theoretical considerations apply to more general situations involving colored noise. In the following, we shall refer to $d(i,j)$ and σ_e^2, (or more generally $\mathrm{cov}(e)$), as the model parameters. The process of restoring a blurred-and-noisy image now involves two principal goals, to be achieved either simultaneously or in two stages.

Problem 1. Blur identification
Given only the observed image $y(i,j)$, and assuming the linear model (1.4), identify (or rather estimate) the model parameters $d(r,s)$ and σ_e^2.

Problem 2. Restoration
Given the observed image $y(i,j)$, and assuming that the mask $d(r,s)$ and the noise variance σ_e^2 are known, restore the noisy-blurred image such that it is as close as possible to its undegraded version.

In practical situations, Problem 1 may be hard to solve, or even intractable, unless some a priori information is available which allows e.g. to select the unknown parameters from a list of possible blurrings. A fairly general pattern for attacking Problem 1 is the following. In a first step, we estimate the noise variance σ_e^2. This is usually possible by directly inspecting parts of the image with a relatively homogeneous gray tone and calculating the variance of y over such a region.

Once the noise variance has been estimated, identifying the blur, which is the really difficult part, may be treated in the following way. We specify a certain family $\mathcal{D}$ of possible masks, and consider an optimization problem of the form

$$\text{minimize} \qquad f(x(d)) \quad \text{subject to} \quad d \in \mathcal{D}, \tag{1.3}$$

where $f(x)$ is some optimization criterion whose evaluation requires knowledge of the solution $x(d)$ to the restoration Problem 2, given the mask $d(r,s)$ in $\mathcal{D}$ and the

noise variance σ_e^2 estimated before. For instance, in [23, 1, 31], an approach of type (1.3) lead to a maximum likelihood parameter estimation, with $f(x)$ denoting the corresponding log-likelihood function, the family $\mathcal{D}$ represented possible blurs of known size, e.g. all masks of a given size, and where function evaluation $f(x(d))$ was performed using Kalman digital filtering.

It should be clear that the optimization problem (1.3) is difficult to solve since function evaluation by itself is relatively costly and possibly lacking in precision, and since derivative information is not available. Consequently, in order to pursue the approach (1.3), a highly effective method for solving Problem 2 is needed, which serves as a black box in (1.3), and in the present paper we shall focus on this aspect. We discuss a Maximum Entropy based approach to solving Problem 2, and we present numerical results based on mathematical programming duality.

2. Noise Model

Assuming that the statistical law of the additive mean zero noise process $e(i,j)$ is known, we are led to consider the following noise model. Arrange the 2D sequence $e(i,j)$ as a 1D sequence e_k in the usual left-to-right top-to-bottom way. If we assume that the e_k are independent and second moments exist, the strong law of large numbers implies

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 \;\to\; \sigma_e^2 \qquad a.s. \qquad (n \to \infty) \tag{2.1}$$

which for large n justifies the asymptotic estimate $\sum_{i=1}^{n} e_i^2 \;\approx\; n\sigma_e^2$. Our image model (1.1) therefore yields the estimate

$$\|Ax - y\|_2^2 \;=\; \|e\|_2^2 \;=\; \sum_{i=1}^{N}\sum_{j=1}^{M} e(i,j)^2 \;\approx\; NM\sigma_e^2. \tag{2.2}$$

It is clear how to modify this type of argument in the case of a colored noise with known covariance matrix (see e.g. [2]). In the case of white noise, solving the restoration problem 2 now leads to the following model:

$$(R_{\text{tol}}) \qquad \begin{array}{c} \text{Given } y,\ A \text{ rsp. } d,\ \text{ and } \sigma_e^2,\ \text{ find } x(i,j) \geq 0 \text{ satisfying:} \\ \|Ax - y\|_2 = \sigma_e\sqrt{NM} =: \epsilon \\ \sum_{i,j} x(i,j) = T := \sum_{i,j} y(i,j). \end{array}$$

We mention two other approaches to choose the tolerance level. The spectrum matching method proposed in [2], and an approach proposed by Gull [16].

3. Maximum Entropy Reconstruction

In order to solve the under-determined reconstruction problem (R_{tol}), we replace it by the following Maximum Entropy optimization program:

$$(P_{\text{tol}}) \qquad \begin{aligned} &\text{minimize} \quad I[x] = -H[x] = \sum_{i=1}^{N}\sum_{j=1}^{M} x(i,j)\log x(i,j) \\ &\text{subject to} \quad \sum_{i=1}^{N}\sum_{j=1}^{M} x(i,j) = T := \sum_{i=1}^{N}\sum_{j=1}^{M} y(i,j),\ x(i,j) \geq 0 \\ &\qquad\qquad \|Ax - y\|_2 \leqslant \epsilon = \sigma_e\sqrt{NM} \end{aligned}$$

The Maximum Entropy (ME) approach for solving under-determined or unstable problems has been applied in various fields (see [2, 4, 8, 10, 18, 25, 30]), and several theoretical models supporting its use are known (see [20, 21, 6]). For the image restoration problem, an entropy based model has been discussed by Frieden (see [14] or [18]). We do not discuss its validity here, but mention that the choice of the Boltzmann–Shannon entropy $\phi(x) = x\log x$ has been challenged. For instance, for a related inversion problem in speech processing, the authors of [22] propose the use of the Burg entropy $\phi(x) = -\log x$ instead. At equal rights, Burg's entropy has been used in image reconstruction, see e.g. [2]. For other choices of entropy type objectives see [4, 5, 9, 10, 26, 27, 28].

Let us now examine the tolerance program (P_{tol}). Observe that the equality $\|Ax-y\|_2 = \epsilon$ ranging in (R_{tol}) has been replaced by an inequality $\|Ax-y\|_2 \leqslant \epsilon$ in order to obtain a convex program. We argue that this is justified in all situations of practical relevance.

Indeed, notice that program (P_{tol}) consists in minimizing $I[x]$ over the intersection $\mathcal{M}\cap\mathcal{Q}$ of the affine manifold $\mathcal{M} = \{x : \sum_{i,j} x(i,j) = T\}$ and the ellipsoid $\mathcal{Q} = \{x : \|Ax - y\|_2 \leqslant \epsilon\}$. Since $I[x]$ is strictly convex on $\mathcal{M}$, the minimum is attained at a boundary point of the $(n-1)$-dimensional ellipsoid $\mathcal{M}\cap\mathcal{Q}$ in $\mathcal{M}$ – unless the global minimum $\tilde{x}$ of $I[x]$ over $\mathcal{M}$ happens to be an interior point of $\mathcal{M}\cap\mathcal{Q}$ relative to $\mathcal{M}$. But notice that $\tilde{x}(i,j) = T/NM$ is a constant image without structure, which by conservation of energy satisfies $A\tilde{x} = \tilde{x}$. Hence $\|A\tilde{x} - y\|_2 = \|\tilde{x} - y\|_2 \leqslant \epsilon$ meant that the observed image y was close to $\tilde{x}$ – a situation which may be excluded in practice. Therefore, the maximum entropy solution $\bar{x}$ of (P_{tol}) practically always lies on the relative boundary of $\mathcal{Q}$ and therefore satisfies the equality $\|A\bar{x} - y\|_2 = \epsilon$, as claimed. In particular, $\bar{x}$ is then automatically a solution to (R_{tol}).

Observe next that the convexity of (P_{tol}) allows us to apply the well-known techniques of convex programming duality. The details of the Lagrangian duality could be found in [3, 4]. The associated dual program (P^*_{tol}), written in convex form, is then to minimize

$$f_{\text{tol}}(\lambda,\mu) = e^{-\mu-1}\sum_{i,j} e^{-(A^*\lambda)_{i,j}} + \lambda^* y + \epsilon\|\lambda\|_2 + \mu T$$

over all λ, μ. This is an unconstrained minimization program, due to the fact that the Boltzmann-Shannon entropy $\phi(x) = x \log x,\ x \geq 0$ has a globally defined Fenchel conjugate $\phi^*(y) = e^{y-1}$. The link between the primal and the dual is the following return formula:

$$\bar{x}(i,j) = e^{-\bar{\mu}-1} e^{-(A^*\bar{\lambda})_{i,j}}, \qquad i = 1, \ldots, N,\ j = 1, \ldots, M, \tag{3.1}$$

which gives an explicit representation of the optimal solution $\bar{x}$ of (P_{tol}) in terms of the optimal multipliers $(\bar{\lambda}, \bar{\mu})$.

We mention that the following penalty model might be discussed as an alternative to (P_{tol}) and its dual:

$$(P_{\text{pen}}) \qquad \begin{array}{ll} \text{minimize} & I[x] + \frac{K}{2}\|Ax - y\|_2^2 \\ \text{subject to} & \sum_{i,j} x(i,j) = T. \end{array}$$

Using the same dual approach, the associated convex dual program (P^*_{pen}) here consists in an unconstrained minimization of the function

$$f_{\text{pen}}(\lambda, \mu) = e^{-\mu-1} \sum_{i,j} e^{-(A^*\lambda)_{i,j}} + \lambda^* y + \frac{1}{2K}\|\lambda\|_2^2 + \mu T, \tag{3.2}$$

with the return formula being again (3.1). For a discussion of a penalty model to a Fourier inversion problem in crystallography we refer to [10].

The penalty model bears some similarity with Frieden's approach (cf. [14, 18]):

$$(P_{\text{fried}}) \qquad \begin{array}{ll} \text{minimize} & I[x] + \alpha \sum_{i,j} (e(i,j) + M) \log(e(i,j) + M) \\ \text{subject to} & \sum_{i,j} x(i,j) = T \\ & Ax - y = e \end{array}$$

Here $M > 0$ is a constant to be chosen so that the terms $e(i,j) + M$ for the relevant errors are nonnegative. The penalty constant is now α. The associated dual is again an unconstrained minimization problem, linked to its primal via the same return formula (3.1), and whose objective is

$$f_{\text{fried}}(\lambda, \mu) = e^{-\mu-1} \sum_{i,j} e^{-(A^*\lambda)_{i,j}} + \alpha \sum_{i,j} e^{\lambda_{i,j}/\alpha - 1} + \lambda^*(y - M\mathbf{1}) + \mu T. \tag{3.3}$$

Notice that in contrast with the tolerance model (P_{tol}), where ϵ could be specified using theoretical considerations, the penalty type models (P_{pen}) and (P_{fried}) suffer from the fact that no good estimates for the constants K rsp. α, M seem to be available, so that these have to be found experimentally (see also [10] for this).

4. Practical Aspects

The overall advantage of the dual technique is that the constrained primal optimization program is transformed into a dual which we solve using techniques for unconstrained problems. The approach works since the original program is convex, the Fenchel conjugate can be calculated explicitly, and the return formula (3.1) is explicit. On the other hand, the challenge is of course the large scale, the number of variables λ_{ij} being NM. For instance, the 200×320 image used for the experiments already gives a total of 64 000 variables, and an image of size 1000×1000 involves a million variables.

Let us now discuss the algorithmic aspects of the dual. We found that it is advantageous to break the minimization into two steps:

$$\inf_{\mu\in\mathbf{R}} \quad \inf_{\lambda\in\mathbf{R}^{NM}} \quad f_{\text{tol}}(\lambda,\mu),$$

the constraints $\|Ax - y\|_2 \leqslant \epsilon$ and $\sum_{i,j} x(i,j) = T$ being different in nature. Indeed, notice that for a relatively small tolerance, the constraint $\sum x(i,j) = T$ will become redundant, since the operator A preserves total energy. Even for a relatively large ϵ one may still hope that $\|Ax - y\|_2 \leqslant \epsilon$ implies $\sum x(i,j) \approx \sum y(i,j)$. Let $\bar\lambda(\mu)$ denote the solution of the inner minimization $\inf_\lambda f_{\text{tol}}(\lambda,\mu)$, and define $\bar x(\mu)$ via the return formula, i.e., $\bar x(\mu;i,j) = \exp\{-\mu - 1\}\exp\{-(A^*\bar\lambda(\mu))_{i,j}\}$. Then the optimal $\bar x$ for (P^*_{tol}) is just $\bar x = \bar x(\bar\mu)$, which may therefore be calculated by solving the equation $\sum_{i,j} \bar x(\mu;i,j) = T$ for μ. We therefore always start the inner minimization over λ with the default value $\mu = 0$. If the solution $\bar x$ so obtained fits the constraint $\sum_{i,j} \bar x(i,j) = T$ up to some tolerance, we stop. Otherwise we adjust this constraint doing a line search over μ. This strategy turns out to be successful since in most cases the default value $\mu = 0$ is satisfactory, and even when this is not the case, adjusting μ often does not significantly improve the quality of the reconstruction, so that the line search in μ can be limited to a few steps.

A drawback of the tolerance approach (P^*_{tol}) seems to be that f_{tol} is not differentiable at 0, due to the term $\epsilon\|\lambda\|_2$. The latter practically excludes the use of any Newton type methods since the Hessian is a full matrix of size $NM \times NM$. On the other hand, the penalty models (P_{pen}) and (P_{fried}) seem to enable Newton type methods combined with line search techniques, since the matrices A in (1.1) are typically sparse band matrices. However, our experiments with Newton's method and a truncated Newton method, when applied to the penalty model, indicate that the use of Newton's method is inefficient, since generally too much time is spent on calculating Newton directions not leading to a reasonable decrease in function value.

Using secant methods like the BFGS update does not present a real alternative, since the band structure of H is not preserved by the update, and since special structure preserving updates are not considered as truly efficient (see [11]). Alternatively, the inverse BFGS update avoids solving a large scale linear system, but requires storing and updating a full $NM \times NM$ matrix, which is hardly expected to

be efficient. On the other hand, recent experiments with a limited memory inverse BFGS method in the spirit of Liu and Nocedal [24] are more encouraging. Even though algorithmically more demanding, this approach seems to improve on the present technique. We refer the discussion to a forthcoming report.

The method which worked best for both the tolerance and the penalty models was conjugate gradients, which needs only few storage locations, and which we implemented in the Pollak–Ribière variant (see [13]), combined with various line search techniques (including a non-monotonous line search [15]). For comparison, we also tested a gradient type method which to some degree uses second order information, known as conjugate gradient alternative (see [29]). While the conjugate gradient method worked surprisingly well and efficient, the gradient based method could not really compete with the former, but turned out to be quite reliable.

For the implementation of the cg algorithm we simply ignored the singularity of f_{tol} at 0. As a stopping criterion we compare the relative change of f with the relative change in λ_i,

$$\frac{\dfrac{f(\lambda+\delta e_i)-f(\lambda)}{f(\lambda)}}{\dfrac{\|(\lambda+\delta e_i)-\lambda\|}{\lambda_i}} \approx \frac{\partial_i f(\lambda)\lambda_i}{f(\lambda)},$$

which leads to the test

$$\max_i \left|\frac{\partial_i f(\lambda)\lambda_i}{f(\lambda)}\right| \leqslant \text{eps}, \tag{4.1}$$

where we used $\text{eps} = 10^{-6}$ in our experiments. Notice that this criterion works well here since $f(\lambda)$ and λ are away from 0. As further halting criteria we tested the progress in λ and in function values as suggested in [11].

5. Experiments

Our experiments were performed for the 200×320 image shown in Figure 1, provided by the MATLAB image tool box, which presents a good simulation of black-and-white photographic pictures. This image was blurred using various 7×7 masks. In particular, we used the 7×7 pillbox blur, the 7×7 ring mask having 24 boundary entries $d(r, s)$ equal to $1/24$, and the 7×7 approximation of an out-of-focus blur. Independent mean zero random normal noise of different variance levels σ_e^2 was then added to further degrade the blurred images. In order to judge the influence of the noise, it is customary to compare the noise variance σ_e^2 with the variance σ_x^2 of the original image x, which gives the signal-to-noise ratio (S/N), usually measured in decibels (dB) via the formula

$$10 \log_{10} \frac{\sigma_x^2}{\sigma_e^2} = \eta \text{ dB}. \tag{5.1}$$

Figures 2, 3 and 4 show the blurred-and-noisy clown degraded by two types of 7×7 supported masks with noise of different levels added.

To compare the total CPU time for the reconstructions, the initial vector λ was chosen as constant ones. Notice, however, that a considerable gain in CPU time could be obtained during experiments with e.g. different tolerances ϵ, ϵ' or different penalization constants K, K' if the solution to a certain reconstruction with say ϵ was used as starting point for the reconstruction with ϵ' nearby.

Our experiments suggest that the tolerance model is best when the 'correct' $\epsilon = \sigma_e \sqrt{NM}$ is used. In contrast, it seems to be more difficult to determine a satisfactory default value for the penalty constant K in (P^*_{pen}). In [14], the default value $\alpha \approx 20$ for the model (P_{fried}) was suggested as satisfactory for a large class of realistic reconstructions.

EXAMPLE 1. In Figure 2 (left) the clown was blurred with the 7×7 pillbox blur and further degraded with additive random normal noise of variance $\sigma_e^2 = 3.99$, giving $\epsilon = 505.2$, and leading to a signal-to-noise level of $S/N = 22$dB. The reconstruction (right) was calculated via the model (P^*_{tol}), with $\mu = 0$, and initial $\lambda^{(1)} = \mathbf{1}$. The number of iterations was 150, with an average of 12 function evaluations during line search. No further search in μ was performed. The total CPU was 1175.7, and the stopping criterion (4.1) became active with eps $= 10^{-6}$. This corresponded with the final gradient $\|\nabla f(\lambda^{(150)})\|_2 = 23.06$.

As further criteria for the goodness of the numerical procedure we calculated the following criteria. The total energy of the blurred-and-noisy image was $T = T_{\text{dirty}} = 1998846$, and the total energy $\sum \bar{x}(i,j)$ of the reconstruction was found to be $T_{\text{rec}} = 1949513$, which gave the relative energy difference

$$\Delta T = \left| \frac{T_{\text{rec}} - T_{\text{dirty}}}{T_{\text{dirty}}} \right| = 0.0247.$$

As a second test for goodness we calculated the relative duality gap, whose theoretical value is 0. This compares the entropy $H[\bar{x}]$ of the reconstruction with the dual value $f(\bar{\lambda})$, and was found to be

$$\Delta I = \left| \frac{H[\bar{x}] - f(\bar{\lambda})}{H[\bar{x}]} \right| = 0.00095.$$

EXAMPLE 2. In Figure 3 the clown was blurred using the 7×7 pillbox blur with $S/N = 14$dB noise added (left). The reconstruction (right) was again obtained via model (P^*_{tol}), with $\sigma_e^2 = 24.87$, and $\epsilon = 1261$. The algorithm was started with $\mu = 0$ and $\lambda^{(1)} = \mathbf{1}$, and the number of iterations was 101, where again the criterion (4.1) with eps $= 10^{-6}$ became active. The total CPU was 1103.3, and the average number of function evaluations during line search was 13.2. The total energies $T_{\text{dirty}} = 2313191$ and $T_{\text{rec}} = 2183724$ led to a relative energy difference of $\Delta T = 0.056$. The relative duality gap was $\Delta I = 0.0012$.

EXAMPLE 3. Here the clown was blurred using the 7×7 ring mask, and white noise with $S/N = 8.1$dB was added, which corresponded to $\sigma_e^2 = 94.39$ (left). The reconstruction (right) was calculated via the tolerance model with the correct $\epsilon = 2457$. The number of iterations before the criterion (4.1) became active was 77, with a total of 11.6 function evaluations during line search, leading to a total CPU of 752.3. The final gradient was $\|\nabla f(\lambda^{(77)})\|_2 = 25.7$. We obtained $T_{\text{dirty}} = 2620246$ and $T_{\text{rec}} = 2845137$, giving $\Delta T = 0.0858$. The relative duality gap was $\Delta I = 0.00065$.

In order to test our algorithm, a greater variety of simulations was performed, and the results are presented in the following table:

Mask	S/N (dB)	♯ iter	♯ eval	CPU
M1	22.0	150	12	1175.7
M1	14.0	101	13.2	1103.0
M1	18.0	95	13	1095.7
M1	8.0	77	12	781.6
M2	8.1	77	11.6	752.3
M2	14.2	61	12	635.3
M2	6.5	24	13.8	406.0
M3	9.96	62	12.3	740.1

Here M1 refers to the pillbox mask, M2 means the ring mask, and M3 the 7×7 approximation of the out-of-focus blur. The restorations for the masks M1, M2 are shown in Figures 2, 3 and 4. Results for the out-of-focus M3 not displayed here lead to the same type of result. Our experiments were obtained in double precision on a RS 6000/320 H with 32 MB memory, but for comparison we also tested two faster processors, a Pentium 90 with 32 MB and a HP 9000/735 with 64 MB memory. While the Pentium 90 was approximately 1.2 times faster, the HP 9000/735 was considerably faster and needed only 60 per cent of the CPU displayed above.

As a general rule, it may be observed that a badly degraded image (having low signal-to-noise ratio), which according to the noise model requires a larger tolerance level ϵ, makes the restoration algorithm faster but, as would be expected, less performing. In these situations a further treatment of the images might be necessary (see the final section).

Our approach must be compared with other experiments based on ME type models. In [14], the model (P_{fried}) was solved using a Lagrange multiplier approach, but the convexity of the program was not exploited directly, i.e., (P^*_{fried}) was not discussed. In [7], an approach similar to (P_{tol}) was chosen, but the associated dual (P^*_{tol}) was not discussed, and the authors present a rather sophisticated approach

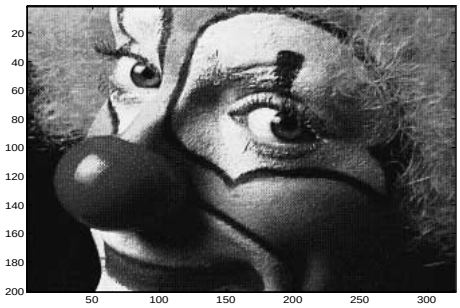

Figure 1. The true image

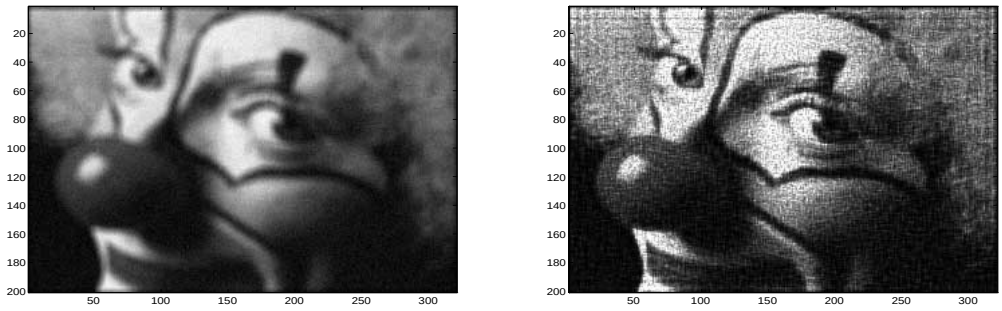

Figure 2. 7×7 pillbox blur; $S/N = 22$dB. Blurred-and-noisy image (left), Restoration (right).

to solving (P_{tol}) which is certainly algorithmically demanding. Similarly, in [18], a differential equation approach for solving the primal program (P_{tol}) is proposed, which is claimed to perform well but in our opinion is rather complicated. We suggest that the dual approach (P^*_{tol}) is conceptually easier and since it performs well should be given the preference over the discussed methods. On the other hand, in contrast with the duality approach (P^*_{tol}), the method in [18] might still be applied to nonlinear filters where program (P_{tol}) is not necessarily convex. We mention another approach to entropy type optimization problems via nonlinear interior point methods (given in [17]), which are reported to perform well for entyropy type problems.

Perhaps closest to our present approach is the dual method presented in [2]. In contrast with (P^*_{tol}) the authors of [2] use a ME model based on the Burg entropy expression, and they propose a slightly different approach to determine a tolerance level ϵ'. The major difference to our present experiments is, however, that their experiments are only for examples of academic scale $n = NM \approx 25$.

6. Prospective

The restoration process outlined here, while conceptually easy, was found to improve on several related approaches. Moreover, as has been reported in [27],

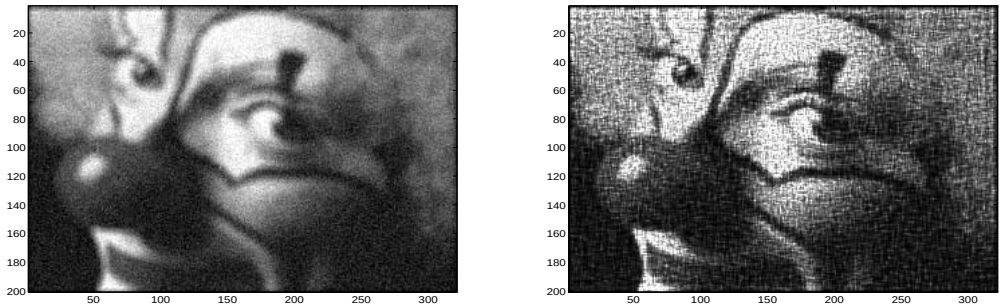

Figure 3. 7×7 pillbox blur; $S/N = 14$dB. Blurred-and-noisy image (left), Restoration (right).

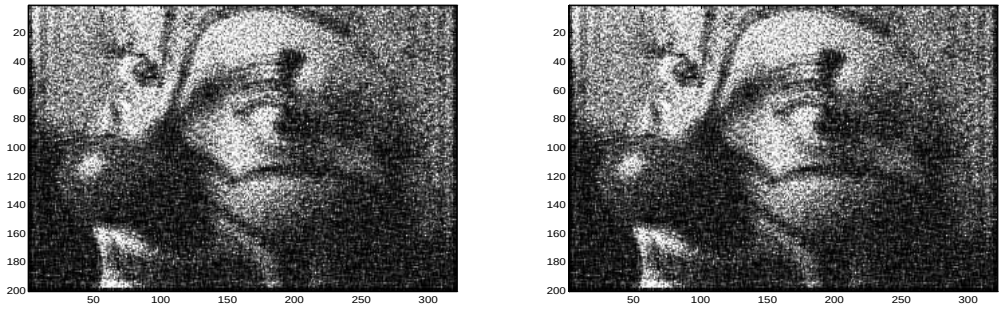

Figure 4. 7×7 ring mask; $S/N = 8.1$dB. Blurred-and-noisy image (left), Restoration (right).

it is clearly superior to the performance of linear inverse filters as for instance proposed by Hunt [19]. Nevertheless, a further improvement of the present technique could be obtained by combining the maximum entropy model with a priori and/or a posteriori smoothing techniques. This is most obvious in cases where the signal-to-noise ratio of the dirty image is low, and the present algorithm terminates faster, but with a less satisfactory result (cf. Figure 4). This combined approach has been discussed in [27], which also comments on how to avoid ringing effects in restorations (Figure 2).

Acknowledgement

We thank Jon Borwein and Mark Limber (Burnaby) and Martin Wursthorn and Jürgen Dippon (Stuttgart) for helpful discussions during the preparation of this paper.

References

1. D.L. Angwin and H. Kaufman (1991), *Non homogeneous image identification and restoration procedures*, Springer Series in Information Sciences, vol. 23, 177–208.
2. Ch. Auyeung and R.M. Mersereau (1991), *A dual approach to signal restoration*, Springer Series in Information Sciences, vol. 23, 21–56.

3. J.M. Borwein and A.S. Lewis (1991), Duality relationships for entropy-like minimization problems, *SIAM J. Control Optimiz*. 29 325–338.
4. J.M. Borwein, A.S. Lewis and D. Noll, Maximum entropy reconstruction using derivative information, part I: Fisher information and convex duality, *Math. of Operations Research*, 21, 1996, 442–468.
5. J.M. Borwein, A.S. Lewis, M.N. Limber and D. Noll, Maximum entropy reconstruction using derivative information, part II: Computational Results, *Numerische Mathematik*, 69, 1995, 243–256.
6. B. Buck and V.A. MacAuley (1991), *Maximum Entropy in Action*, Oxford Science Publ., Oxford.
7. S.F. Burch, S.F. Gull and J.K. Skilling (1983), Image restoration by a powerful maximum entropy method. *Computer Vision, Graphics, and Image Processing* 23, 113–128.
8. J.P. Burg (1967), *Maximum Entropy Spectral Analysis*, 37th meeting of the Society of Exploration Geophysicists, Oklahoma City.
9. D. Dacunha-Castelle and F. Gamboa (1990), Maximum d'entropie et problème des moments, *Ann. Inst. Henri Poincaré* 26, 567–596.
10. A. Decarreau, D. Hilhorst, C. Lemaréchal and J. Navaza (1992), Dual methods in entropy maximization: Application to some problems in crystallography, *SIAM J. Optimization* **2** 173–197.
11. J.E. Dennis jr. and R.B. Schnabel (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall Series in Computational Mathematics.
12. T.S. Durrani and C.E. Goutis (1990), Optimization techniques for digital images reconstructed from their projections, *IEEE Proc. Part E*, 161–169.
13. R. Fletcher (1987), *Practical Methods of Optimization*, John Wiley & Sons, New York.
14. B.R. Frieden (1972), Restoring with maximum likelihood and maximum entropy. *J. Opt. Soc. Amer.* **G2**, 511–518.
15. L. Grippo, F. Lampariello and S. Lucidi (1986), A non monotone line search technique for Newton's method, *SIAM J. Numer. Anal*. **23**, 707–716.
16. Gull, S.F. (1989), *Maximum Entropy and Bayesian Methods*, vol 13, pp. 53–171, ed. by J.K. Skilling, Kluwer.
17. C.G. Han, P.M. Pardalos and Y. Ye (1992), Implementation of interior point algorithms for some entropy optimization problems, *Optimization and Software* 1, 71–80.
18. R.M. Haralick, E. Østevold and X. Zhuang (1987), A differential equation approach to maximum entropy image reconstruction. *IEEE Trans. Acoust. Speech Signal Processing* 35: 208–218.
19. Hunt, B.R. (1972), Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory, *IEEE Trans. Autom. Contr.*, 703 – 705.
20. E.T. Jaynes (1968), Prior probabilities, *IEEE Transactions on Systems, Science and Cybernetics SSC-4*, 227–241.
21. E.T. Jaynes (1982), On the rationale of maximum entropy methods, *Proc. IEEE*, Special issue on spectral estimation, vol. 70, 939–952.
22. R.W. Johnson and J.E. Shore (1984), Which is the better entropy expression for speech processing: $-S \log S$ or $\log S$? *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-32, 129–137.
23. A.K. Katsaggelos and K.-T. Lay (1991), *Maximum Likelihood Identification and Restoration of Images Using the Expectation-Maximization Algorithm*, Springer Series in Information Sciences, vol. 23, 143–176.
24. D.C. Liu and J. Nocedal (1989), On the limited memory BFGS update for large scale optimization, *Math. Programming* **45**, 1989, 503 – 520.
25. J.H. McClellan and S.W. Lang (1982), Multi-dimensional MEM spectral estimation, *Spectral Analysis and Its Use in Underwater Acoustics*, 10.1–10.8, Imperial College, London, U.K., Institute of Acoustics.
26. D. Noll (1995), Rates of convergence for best entropy estimates, *Statistics and Decisions*, 13, 1995, 141–165
27. D. Noll, Consistency of a nonlinear deconvolution method with applications to image restoration, *Advances in Mathematical Sciences and Applications*, to appear.

28. D. Noll, *Reconstruction with Noisy Data: An Approach via Eigenvalue Optimization*, to appear in *SIAM J. Optimization*
29. M. Raydan, *The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem*, to appear.
30. J. Skilling (ed.) (1989), *Maximum-Entropy and Bayesian Methods*, Kluwer, Dordrecht, The Netherlands.
31. A.M. Tekalp and G. Pavlović (1991), *Restoration of Scanned Photographic Images*, Springer Series in Information Sciences, vol. 23, 209–238.