(at least nearly) the optimality properties of the minimum l_1-rule in the nearly black case and perhaps allow as good or better behaviour when the object is not nearly black. One possible class of priors would be a mixture of a point mass at 0 with probability to be estimated from the data and a gamma prior with parameters also to be estimated. Here the probability of the point mass at 0 would play the role of the tuning parameter λ for the maximum entropy and minimum l_1-methods and would be adaptively determined. The potential advantage of such a method should come in its ability to adapt at least somewhat to the non-black part of the image. As such empirical Bayes methods are well known to the authors I presume that there are computational or other reasons for not using some variation on the empirical Bayes theme.

Didier Dacunha-Castelle (Université Paris Sud, Orsay): I am pleased to congratulate the authors for their fascinating paper. For their model (1) $Y_i = (Kx)_i + \epsilon_i$, x is chosen by maximizing $S(x)$ subject to the relaxed constraint. Model (2) is $\| Y - Kx \|_2 < \epsilon$; of course the procedure can be thought of as a deterministic fitting procedure. There is a certain kind of optimality if we choose $S(x) = \|x\|_{l_1}$. This optimality depends on

(a) the normality of (ϵ_i) and
(b) on the asymptotics chosen, i.e. a level of noise tending to 0.

What happens if we change condition (a)? What is the link between this result with the classical asymptotic optimality in robustness theory (as in P. Huber's work)? From this deterministic point of view, instead of model (2) we can regularize a problem maximizing $S(x)$ under constraint (3), $m_k(x) = m_k(y)$, $1 \leqslant k \leqslant m$, when

$$m_k(x) = \frac{1}{n} \sum_{k=1}^{n} x^k .$$

This point of view is linked to the use of a generalization of the maximum entropy principle as in Dacunha-Castelle and Gamboa (1990) and Gamboa and Gassiat (1990) and in some statistical problems. It seems well adapted when we work with the asymptotics of discretization, with (x_i) thought of as a finite number of values of a continuous function x_t. In this asymptotic, there is a Bayesian interpretation of the choice of S (Gamboa and Gassiat, 1991a, b).

Elisabeth Gassiat (Université Paris-Sud, Orsay): I would first like to congratulate the authors for their clear exposition of the properties of the so-called 'maximum entropy' methods in general, underlying the role of positiveness and 'nearly blackness' of the object to be recovered. I shall focus my contribution on the second part of the paper, where the operator K is not easily invertible. The authors show a superresolution property in the particular case where K is the finite Fourier transformation: $\mathbf{y} = K\mathbf{x} + \boldsymbol{\epsilon}$, $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{y} \in \mathbf{R}^m$, $m << n$. Here $\boldsymbol{\epsilon}$ is a deterministic error.

In this approach, and looking at their proofs, the role of the discretization and of the geometric structure induced by K does not demonstrate clearly why superresolution occurs.

The problem may be rephrased in a more general form: if U is a compact metrizable space, $\mathbf{x}$ a positive measure, K an m-dimensional continuous function on U, the observation is

$$\mathbf{y} = \int_U K \, \mathrm{d}\mathbf{x} + \boldsymbol{\epsilon}.$$

The problem is now a generalized moment problem. In Gamboa and Gassiat (1991b), it is shown that superresolution occurs if and only if $\int_U K \, \mathrm{d}\mathbf{x}$ is a determinate point, i.e. a point for which the set of positive measures solutions to the moment problem is reduced to a singleton. In that paper criteria are also given to show whether a point is determinate or not. All the results rely on developments on 'maximum entropy methods on the mean' developed in Gamboa and Gassiat (1991a). Another point of view of near-blackness has also been studied in Gassiat (1990), relating it to some concentration property of the support of any positive measure solution of the moment problem.

Andrew Gelman (University of California, Berkeley): The authors present an enlightening discussion of where and why maximum entropy methods are effective. Many of their results become even clearer from a Bayesian perspective: better prior distributions yield better estimates. Jaynes (1987) makes a similar point in his discussion of the non-linear shrinkage that results from Bayesian spectral estimation using a nearly black model (without maximum entropy).

For simplicity, I shall analyse the example of Section 2. Each of the estimates considered by the authors

TABLE 2
Shrinkage estimates and their corresponding prior densities

Estimate $\hat{x}_i$	*Prior density on* x_i
y_i	Uniform$(-\infty, \infty)$
$\max(y_i, 0)$	Uniform$[0, \infty)$
$\hat{x}_{\mathrm{RLS}}$	Normal$(0, \sigma^2/\lambda)$
$\max(\hat{x}_{\mathrm{RLS}}, 0)$	Normal$(0, \sigma^2/\lambda)$, truncated to be non-negative
δ_{ME}	Density proportional to $\exp\{(-\lambda/\sigma^2)x_i \log x_i\}$
δ_{l_1}	Exponential(σ^2/λ)
δ_{Thresh}	Mixture of point mass at 0 and uniform$[0, \infty]$

may be interpreted as a posterior mode under the appropriate family of proper or improper prior distributions on $x_1, \ldots, x_n$ (Table 2).

The regularization parameter λ is assumed known. In practice, the normal and maximum entropy prior distributions typically each take another parameter, fitting location for the normal and scale for the entropy distribution. In either case, fitting the additional parameter allows the fixed point of shrinkage to be fitted to the data, rather than be fixed at 0 for quadratic regularization and e^{-1} for maximum entropy. (Incidentally, fitting the second parameter eliminates the problem with maximum entropy described in the penultimate paragraph of Section 2.2.)

As the authors report, maximum entropy reconstruction shrinks large data values proportionately less than small data values, thus preserving peaks while suppressing noise. Least squares regularization performs worse for the nearly black object, even when restricted to positivity, because it shrinks all positive data by a common factor.

The shrinkage behaviour of the estimates makes perfect sense in light of the corresponding prior distributions. The prior density for the trivial estimate is constant, so the data are not pulled at all. The priors for least squares regularization have rapidly decaying $\exp(-x^2)$ tails, so large data points will be pulled strongly towards the fixed point. The priors for maximum entropy—and also for the l_1-estimate—decay only like $\exp(-x)$ and so large data values are shrunk less strongly. The threshold estimate δ_{Thresh} pulls points to the point mass at 0, but its uniform component fails by not shrinking the positive components x_i to their common mean.

In summary: under the 'nearly black' model, the normal prior is terrible, the entropy prior is better and the exponential prior is slightly better still. (An even better prior distribution for the nearly black model would combine the threshold and regularization ideas by mixing a point mass at 0 with a proper distribution on $[0, \infty]$.) Knowledge that an image is nearly black is strong prior information that is not included in the basic maximum entropy estimate.

The following contributions were received in writing after the meeting.

Bob Anderssen (CSIRO, Canberra): In making this important contribution, the authors, not surprisingly, find that maximum entropy has its limitations. The impact which this paper has will depend on the extent to which it removes the belief that maximum entropy is a panacea. The shortcomings of maximum entropy have already been noted by others (e.g. Engl and Landl (1991), Koch and Anderssen (1987) and Titterington (1984)). What the present authors have achieved is to have placed this earlier understanding on a much firmer foundation.

Certainly, strong support for maximum entropy can be based on physical (thermodynamical) and philosophical (minimum assumptions) considerations. But, it is one among many such principles like minimum energy in elasticity and minimum action in optics and seismology. The choice of methodology should be driven by the information flowing from the context in which the data have been collected rather than other considerations. The context also has a bearing on the statistical interpretation of the approach (O'Sullivan, 1986; Wahba, 1990). For example, Wahba (1990), section 1.5, gives a Bayesian estimation interpretation for regularization with a quadratic regularizor.

The success (when it occurs) of maximum entropy relates to its implementation via a regularization framework such as

$$\min_{\mathbf{x}}\{\|\mathbf{y}-K\mathbf{x}\|_2^2+2\lambda\,\Lambda(\mathbf{x})\}, \tag{25}$$