Deconvolution of Neutron Depth Profiling Spectra with Entropic Priors

Carlos C. Rodríguez
Department of Mathematics and Statistics
State University of New York at Albany
E-Mail: carlos@omega.albany.edu

March 18, 1992

Abstract

An iterative algorithm to deconvolve neutron depth profiling spectra is derived. The method is similar to the EM algorithm used by Coakley [1] but it aims at maximizing the entropic posterior probability instead of the likelihood. This allows to introduce objective and invariant prior information into the problem and it is therefore expected to improve the reconstructions. Basically the same algorithm can be used to reconstruct not just *depth* but the full three dimensional concentrations of impurities in a sample. General guidelines for computational experiments, to be designed in close contact with NIST scientists, are proposed to evaluate the operating characteristics of the new procedure.

1 Derivation of the Problem

I propose to incorporate the available objective prior information to the problem of deconvolution of neutron depth profiling (NDP) spectra by means of the associated family of entropic priors. Entropic priors (see Rodríguez [2]) have desirable theoretical properties and they have been successfully applied to deconvolution problems in image processing.

In NDP a sample (e.g. silicon) containing an unknown distribution of impurities (e.g. boron) is bombarded with a beam of neutrons. When the neutrons interact with the impurities radiation is emitted. The energy of this radiation depends on the position of the impurities in the sample in a way that allows to specify $p(b, d)$, the probability of observing intensity d when a particle is produced at position b, from the geometry of the experimental setting. If we assume that the emissions follow a spatial Poisson point process in the region of the sample and we discretize the positions ($b = 1, \ldots, B$) and energies ($d = 1, \ldots, D$)

we have that the probability of observing counts $n = (n(1), \ldots, n(D))$ when the true intensity rates are $\lambda = (\lambda(1), \ldots, \lambda(B))$ is given by:

$$\begin{aligned} P_\lambda(n) &= Pr\left(n(1), \ldots, n(D) \,|\, \lambda(1), \ldots, \lambda(B)\right) \\ &= \prod_{d=1}^{D} e^{-\lambda^*(d)} \frac{\lambda^*(d)^{n(d)}}{n(d)!} \end{aligned} \tag{1}$$

with

$$\lambda^*(d) = \sum_{b=1}^{B} \lambda(b) p(b,d). \tag{2}$$

The Poisson model was first used in this context by Shepp and Vardi [3] for positron emission tomography and by Coakley [1] for NDP.

As it is shown in Rodríguez [2] the Kullback number $I(P_\lambda : P_\theta)$ (i.e. entropy) generates the metric and the prior on the hypothesis space (see equations (1) and (9) in [2]). From 1 we have,

$$\begin{aligned} I(P_\lambda : P_\theta) &= \sum_n P_\lambda(n) \log \frac{P_\lambda(n)}{P_\theta(n)} \\ &= \sum_n P_\lambda(n) \sum_{d=1}^{D} \left\{ \theta^*(d) - \lambda^*(d) + n(d) \log \frac{\lambda^*(d)}{\theta^*(d)} \right\} \\ &= \sum_{d=1}^{D} \left\{ (\theta^*(d) - \lambda^*(d)) \left(\sum_n P_\lambda(n) \right) + \left(\sum_n n(d) P_\lambda(n) \right) \log \frac{\lambda^*(d)}{\theta^*(d)} \right\} \end{aligned}$$

the first sum inside parenthesis is one, since P_λ is a probability mass function, and the second sum in parenthesis is the mean of $n(d)$ i.e. $\lambda^*(d)$. Hence, we obtain

$$I(P_\lambda : P_\theta) = \sum_{d=1}^{D} \left\{ \theta^*(d) - \lambda^*(d) + \lambda^*(d) \log \frac{\lambda^*(d)}{\theta^*(d)} \right\} \tag{3}$$

The information metric of the hypothesis space is obtained by computing:

$$I\left(P_{\theta+\epsilon v} : P_\theta\right) = \sum_{d=1}^{D} \left\{ -\epsilon v^*(d) + (\theta^* + \epsilon v^*(d)) \log \left(1 + \epsilon \frac{v^*(d)}{\theta^*(d)} \right) \right\}$$

expanding the logarithm up to second order in ϵ and simplifying we obtain:

$$I\left(P_{\theta+\epsilon v} : P_\theta\right) = \frac{\epsilon^2}{2} \sum_{d=1}^{D} \frac{[v^*(d)]^2}{\theta^*(d)} + o(\epsilon^2)$$

from here it follows that the square of the tangent vector v attached to the point θ is

$$\|v\|_\theta^2 = {v^*}^\dagger \begin{pmatrix} \frac{1}{\theta^*(1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\theta^*(D)} \end{pmatrix} v^* \tag{4}$$

where v and v^* are assumed to be column vectors satisfying,

$$v^* = \mathbf{P}^\dagger v \tag{5}$$

with $\mathbf{P}$ the B by D matrix with (b, d) element given by $p(b, d)$. Replacing 5 into 4 we can see that the metric tensor $g(\theta)$ in matrix form is given by:

$$g(\theta) = \mathbf{P} \begin{pmatrix} \frac{1}{\theta^*(1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\theta^*(D)} \end{pmatrix} \mathbf{P}^\dagger \tag{6}$$

Therefore, by taking the square root of the determinant of equation 6 we obtain the invariant measure, $\eta(d\theta)$, defined up to a proportionality constant by,

$$\eta(d\theta) \propto \frac{d\theta}{\sqrt{\theta^*(1) \cdots \theta^*(D)}} \tag{7}$$

Using 1,3 and 7, we find that the logarithm of the posterior probability of λ given the observed counts n the value of the parameter α and an initial guess θ, is proportional to:

$$\begin{aligned} U(\lambda) &= \sum_{d=1}^{D} \{ n(d) \log \lambda^*(d) - \lambda^*(d) \\ &\quad - \alpha \left[\theta^*(d) - \lambda^*(d) + \lambda^*(d) \log \frac{\lambda^*(d)}{\theta^*(d)} \right] - \frac{1}{2} \log \lambda^*(d) \right\}. \end{aligned} \tag{8}$$

Re-arranging and simplifying the expression 8 we obtain that the most likely a posteriori value of the intensities λ is given by the solution of the following optimization problem:

$$\max_\lambda \sum_{d=1}^{D} \{ [\tilde{n}(d) - \alpha \lambda^*(d)] \log \lambda^*(d) - \beta(d) \lambda^*(d) \} \tag{9}$$

where λ^* is given by equation 2 and

$$\tilde{n}(d) = n(d) - \frac{1}{2}, \tag{10}$$

and

$$\beta(d) = 1 - \alpha \left(\log \theta^*(d) + 1 \right). \tag{11}$$

Several algorithms could be designed to solve problem 9. Of particular interest is the following extension of the algorithm used in [1].

2 Proposed Solution

Here we follow closely the derivation of the EM algorithm for positron emission tomography in Vardi *et al.* [4] section (2.1). Using the objective function $U(\lambda)$ (see 9) we can write the following expressions for the first and second derivatives:

$$\frac{\partial U}{\partial \lambda(b_0)} = \sum_{d=1}^{D} \left\{ \frac{\tilde{n}(d)}{\lambda^*(d)} - \alpha - \alpha \log \lambda^*(d) - \beta(d) \right\} p(b_0, d), \tag{12}$$

$$\frac{\partial^2 U(\lambda)}{\partial \lambda(b_0) \partial \lambda(b_1)} = -\sum_{d=1}^{D} \left\{ \frac{\tilde{n}(d)}{\lambda^*(d)} + \alpha \right\} \frac{p(b_0, d) p(b_1, d)}{\lambda^*(d)}. \tag{13}$$

If all the $\tilde{n}$'s are positive the matrix of second derivatives is negative semidefinite, as it can be easily verified from equation 13. Hence, it follows that $U(\lambda)$ is concave and that sufficient conditions for $\hat{\lambda}$ to be a maximizer of U are the Kuhn-Tucker conditions as in [4]. Therefore, from 12 it follows that it is sufficient for $\hat{\lambda}$ to satisfy

$$\hat{\lambda}(b) \sum_{d=1}^{D} \beta(d) p(b, d) = \hat{\lambda}(b) \sum_{d=1}^{D} \left\{ \frac{\tilde{n}(d)}{\hat{\lambda}^*(d)} - \alpha(1 + \log \hat{\lambda}^*(d)) \right\} p(b, d) \tag{14}$$

The natural (fix point like) iterative algorithm suggested by the previous equation is:

$$\hat{\lambda}^{\text{new}}(b) = \hat{\lambda}^{\text{old}}(b) \frac{\sum_{d=1}^{D} \left\{ \frac{\tilde{n}(d)}{\hat{\lambda}^*(d)} - \alpha(1 + \log \hat{\lambda}^*(d)) \right\} p(b, d)}{\sum_{d=1}^{D} \beta(d) p(b, d)} \tag{15}$$

where

$$\hat{\lambda}^*(d) = \sum_{b=1}^{B} \hat{\lambda}^{\text{old}}(b) p(b, d) \tag{16}$$

Notice that when $\alpha = 0$ is replaced in 15 (see also 11) we recover the original EM algorithm in Coakley [1] on the data $\tilde{n}$ instead of n. This has intuitive appealing; if all the observed counts are large the two solutions agree, but if some energy bins contain only a few recorded particles the new estimator subtracts the effect of these rare events and it should produce better solutions. This needs to be tried by simulation and on real data.

3 Proposed Research to be Conducted at NIST

In order to evaluate the operating characteristics of the algorithm proposed in 15 experiments should be conducted in close contact with scentists at NIST.

First of all, the algorithm needs to be implemented in an efficient computer language like C. Special attention should be paid to minimize the required memory space and running time for the program. Simulations with relatively small values for B and D might be feasible to run on a RISC workstation. However with realistic sizes super computer computational muscle would be required.

The proposed algorithm depends on the following parameters:

- The information scalar parameter, α.
- The initial guess for the vector of intensities, θ.
- The matrix of transition probabilities, $p(b, d)$.
- The stopping criterion.

The required positive scalar parameter, α, indicates the quality of the guess, θ. Good guesses should be given larger values of α and poor or unreliable guesses should go with α close to zero. The extreme case is $\alpha = 0$ that makes θ disappear from the algorithm. A quick and dirty estimation of the intensities (perhaps using the convolution backprojection algorithm) may provide a good value for θ. Material scientists should be consulted for possible educated guesses based on the physics of the situation. Once θ is given, cross-validation or empirical bayes methods could be used for choosing α.

The matrix of probabilities could (in principle) be obtained from the physics and geometry of the experimental setting. A good starting point is the information given in Coakley's [1] paper and it should be possible to extend the results to distributions that depend not only on depth but also on the angle of view. Again input from experimental physicists would be useful here.

Finally, an adequate stopping criterion needs to be specified. If the initial guess is sufficiently reliable then, traditional stopping methods for iterative algorithms may work. However, it may be necessary to use cross-validation on the posterior distribution in a manner similar to the technique proposed by Coakley.

Conclusive answers can be given only after experiments on simulated and real data.

References

[1] Kevin J. Coakley. A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profining spectra. *IEEE Trans. Nucl. Sci.*, 38:9–15, 1991.

[2] Carlos C. Rodríguez. Entropic priors. Attached manuscript, Oct. 1991.

[3] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, M1-1:113–122, 1982.

[4] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *J. Amer. Stat. Assoc.*, 80:8–20, 1985.