



# A coordinate ascent approach to tomographic reconstruction of label images from a few projections<sup>☆</sup>

Hstau Y. Liao, Gabor T. Herman

*Department of Computer Science, The Graduate Center, City University of New York, NY, USA*

Received 1 October 2003; received in revised form 27 April 2004; accepted 10 February 2005  
Available online 12 July 2005

---

## Abstract

Our aim is to produce a tessellation of space into small voxels and, based on only a few tomographic projections of an object, assign to each voxel a label from a small predetermined set that indicates one of the components of interest constituting the object. Traditional methods are not reliable due to, among other reasons, the low number of projections. We postulate a low level prior knowledge regarding the underlying distribution of label images, and then directly estimate a label image based on the prior and the projections. We use a coordinate ascent approach for the estimation.  
© 2005 Elsevier B.V. All rights reserved.

*Keywords:* Label images; Discrete tomography; Few projections; Coordinate ascent

---

## 1. Introduction

Computerized tomography (CT) is the process of obtaining the density distribution within an object from its projections. (Here *density* refers to the physical property being imaged; e.g., attenuation coefficient, Coulomb potential, radioactivity, etc.) A *projection* consists of measured integrals of the density distribution along a set of lines that are either

---

<sup>☆</sup> This work was supported by the NIH Grant HL70472 and NSF Grant DMS0306125.

*E-mail addresses:* [h.liao@alumni.upenn.edu](mailto:h.liao@alumni.upenn.edu) (H.Y. Liao), [gabortherman@yahoo.com](mailto:gabortherman@yahoo.com) (G.T. Herman).

all parallel to each other or all sharing a common point. Since an enormous variety of possible density values may occur, a large number of projections are necessary to ensure an accurate reconstruction of the distribution; see e.g. [7,14] for discussions of methods for CT reconstruction.

In many situations, however, the ultimate aim is not the density distribution itself but rather a distribution of labels from a small predetermined set. Each of these labels corresponds to one of the components (such as a tissue type in the human body) constituting the object. For example, a large fraction of objects scanned in industrial CT (for nondestructive testing or reverse engineering) are made up of a single component material and therefore the ideal reconstruction should contain only two labels: one that indicates air and the other that indicates the material composing the object. Similar assumptions can be made in other specific application areas, such as cardiac angiography of heart chambers (presence or absence of injected dye), electron microscopy of biological macromolecules embedded in ice (presence of either protein or ice), etc. In many of these applications there are strong technical reasons why only a few projections of the object can be collected. In nondestructive testing the cost of the projection taking has to be passed on to the purchasers of the product; in cardiac angiography projections of the moving heart are best taken simultaneously and there is only room for so many X-ray sources; in electron microscopy the taking of such projections inevitably damages the macromolecules. Making use of the knowledge that the reconstruction should contain only a few values (labels) to make up for the lack of the availability of the number of projections typically required in CT is the central theme of *discrete tomography* (DT); see [9] for the foundations, algorithms, and selected applications of DT and also [3,5,6,11,15,17,19] for additional applications in medicine. Note that DT as discussed here produces directly a labeling of the components of the object based on the projections. To achieve the same in CT, the reconstructed density distribution needs to be segmented into components by an additional process.

Our main application is in electron microscopy of macromolecules. Electron microscopy is a powerful technique for imaging complex biological macromolecules in order to further the understanding of their functions. It is applicable to a broad range of situations, such as specimens arranged on a 2D crystal, on a helix, on an icosahedron, etc. The structural information provided is usually in the resolution region between 6 and 20 Å. Our proposed methodology should be applicable to the problem of even higher resolution reconstruction from only a few projections (micrographs).

Our long-term aim is to produce, based on the micrographs, a tessellation of space into small voxels, each labeled as containing ice, protein, or RNA. Traditional approaches using methods of CT would first assign, based on the projections (micrographs), to each voxel a *gray value* (which, in this case, is related to the electron density of that voxel) and then would segment this gray value image to obtain the label image. Common tools used for the segmentation step are, e.g., thresholding [18], human-interactive morphological functions [10], and fuzzy connectedness [8]. However, since typically the number of projections required in CT is much larger than that in DT, a reconstructed gray value image from only a few projections is likely to be very inaccurate and so lead to an incorrect segmentation. A particular difficulty arises if thresholding is used for segmentation (which is a common current practice) because at a resolution of 2.5 Å

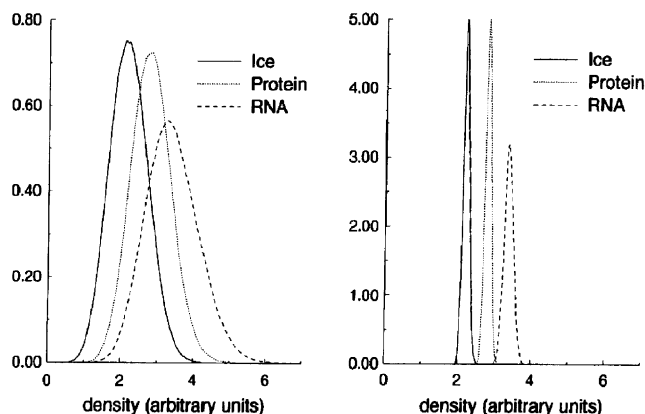


Fig. 1. Histograms of the densities corresponding to volumes sampled using voxels of edge length equal to 2.5 Å (left) and 7.5 Å (right), obtained from volumes composed of only ice, only protein, or only RNA; from [1].

or better the density distributions corresponding to different labels greatly overlap (see Fig. 1).

To overcome such difficulties, we postulate a low-level Gibbs prior [21] on the underlying distribution of label images, and then directly estimate, using an approximation to the *maximum a posteriori probability* (MAP) approach, an optimum label image based on the prior and the measured projections. We choose Gibbs priors as it has been experimentally demonstrated [2,12,13,20] that for certain types of Gibbs distributions there are algorithms that recover an unknown image (that is a typical sample from the distribution) when provided with only a few projections of the image and with the values of the parameters of the Gibbs distribution. What is new in this paper compared to all the previous related works is the extra difficulty in the problem presented: unlike in [2,12,13,20], in which a gray value of a voxel is uniquely determined by the labeling of that voxel, here we allow the gray value to take any real value according to some probability distribution. Therefore, the earlier works can be thought of, after appropriate mathematical formalism adjustments, as dealing with a special case of our problem. For some of the types of Gibbs distributions in the published works, the typical samples correspond to images that have relatively large uniform regions over a background, as occur in almost all the applications that use the methods of tomography. For example, in images of biological macromolecules, the background would correspond to ice and the foreground would be the protein and the RNA. Here we report on our early experiments aimed at evaluating this kind of approach.

The paper is organized as follows. In Section 2 we formulate the problem through a probabilistic model and adopt the MAP criterion for image reconstruction. However, due to the difficulty of dealing with the posterior probability, in Section 3 we discuss an alternative criterion that we expect to deliver reasonably good reconstructions. The experimental details and results are, respectively explained in Section 4 and shown in Section 5. Finally, in Section 6 we give the conclusions and a discussion.

## 2. The proposed model

Let  $\mathbf{X}$  be a set of *label images* each one of which is an  $I$ -dimensional column vector  $\mathbf{x} = (x_1, \dots, x_I)^t$ , where  $x_i \in X$  (the finite set of *labels*), for  $1 \leq i \leq I$ . We assume that there is a prior distribution that assigns to every label image  $\mathbf{x}$  a probability  $\pi(\mathbf{x})$ . Typically, this prior distribution is a *Gibbs distribution*. This means that

$$\pi(\mathbf{x}) = Z^{-1} \exp[-H(\mathbf{x})], \quad (1)$$

where  $\pi(\mathbf{x})$  is the probability of occurrence of the image  $\mathbf{x}$ ,  $Z$  is the normalizing factor, and  $H(\mathbf{x})$  is referred to as the *energy* of  $\mathbf{x}$  (see, e.g., [21]). The energy is the sum of *local potentials*, each of which is a real number assigned to a *clique* (in the experiments reported below the cliques are  $3 \times 3$  subregions of the image). As a result,  $\pi(\mathbf{x})$  is a product of terms, each of which depends on only one clique. Typically, the local potentials depend on the labels in the cliques in a way that is invariant under various transformations (such as translations, rotations, and reflections); this results in having only a few possible values for the local potentials, these are the *parameters* of the Gibbs distribution (see the examples in [12] with five parameters).

Let  $\mathbf{Y}$  be a set of *gray value images* each one of which is a  $J$ -dimensional column vector  $\mathbf{y} = (y_1, \dots, y_J)^t$ , where  $y_j \in Y$  (the set of *gray values*,  $Y \subseteq \mathbb{R}$ ), for  $1 \leq j \leq J$ . We assume that there is a conditional distribution that, given a label image  $\mathbf{x}$ , assigns a probability  $\phi(\mathbf{y}|\mathbf{x})$  to every gray value image  $\mathbf{y}$ . (Since  $\mathbf{Y}$  is not necessarily finite, it would be more precise to say that  $\phi(\mathbf{y}|\mathbf{x})$  is the probability density function defining the conditional distribution of the gray value image  $\mathbf{y}$  given the label image  $\mathbf{x}$ . For the sake of brevity, we will continue to refer to a notation such as  $\phi(\mathbf{y}|\mathbf{x})$  as a “probability” rather than a “probability density function.”) Initially we work with the special case in which  $J = I$  and, for every label  $x$ , there is a distribution which assigns (independently) a probability  $\phi(y|x)$  to every gray value  $y$ . Consequently,  $\phi(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^J \phi(y_j|x_j)$ .

Let  $\mathbf{W}$  be a set of *measurement vectors* each one of which is a  $K$ -dimensional column vector  $\mathbf{w} = (w_1, \dots, w_K)^t$ , where  $w_k \in W \subseteq \mathbb{R}$ , for  $1 \leq k \leq K$ . We assume that there is a conditional distribution that, given a gray value image  $\mathbf{y}$ , assigns a probability  $\chi(\mathbf{w}|\mathbf{y})$  to every measurement vector  $\mathbf{w}$ . For the preliminary experiments reported below, we consider only the special case in which  $\chi(\mathbf{w}|\mathbf{y})$  has a product form; i.e., given a gray value image  $\mathbf{y}$ , there is (for  $1 \leq k \leq K$ ) a conditional probability  $\psi_k(w|\mathbf{y})$  of the  $k$ th measurement being  $w$ , and  $\chi(\mathbf{w}|\mathbf{y}) = \prod_{k=1}^K \psi_k(w_k|\mathbf{y})$ .

The *likelihood*  $\rho(\mathbf{w}|\mathbf{x})$  of measuring  $\mathbf{w}$  given a label image  $\mathbf{x}$ , assuming (as reasonable) that  $\mathbf{w}$  does not depend directly on  $\mathbf{x}$ , is the integral (or sum if  $Y$  is finite)

$$\rho(\mathbf{w}|\mathbf{x}) = \int_{\mathbf{Y}} \chi(\mathbf{w}|\mathbf{y}) \phi(\mathbf{y}|\mathbf{x}) \, d\mathbf{y}. \quad (2)$$

By invoking Bayes' rule, the posterior probability of  $\mathbf{x}$  given the measurement  $\mathbf{w}$  is proportional to

$$\pi(\mathbf{x}) \rho(\mathbf{w}|\mathbf{x}). \quad (3)$$

### 3. The estimation approach

According to the MAP criterion, the estimate of the label image is  $\mathbf{x}_{\text{MAP}}$ , which is the  $\mathbf{x}$  in  $\mathbf{X}$  that maximizes the product of  $\pi(\mathbf{x})$  and  $\rho(\mathbf{w}|\mathbf{x})$ ; i.e.,

$$\mathbf{x}_{\text{MAP}} = \arg \max_{\mathbf{x}} [\pi(\mathbf{x})\rho(\mathbf{w}|\mathbf{x})]. \quad (4)$$

Due to the non-linearity and the non-convexity of the optimizing function of (3) (in general, a Gibbs distribution is non-convex), the MAP estimator(s) cannot be obtained using local search techniques. Furthermore, the posterior is defined on labels, rather than on continuous numerical variables; therefore, we are forced to use some kind of combinatorial optimization technique.

In order to proceed toward our stated goal, it seems desirable to perform the integration on the right-hand side of (2) and find a formula for the likelihood  $\rho(\mathbf{w}|\mathbf{x})$  in which  $\mathbf{y}$  does not appear. (The alternative way of evaluating the integral is by using some kind of Monte Carlo technique, but these are inherently slow.) Except for the case when  $\chi(\mathbf{w}|\mathbf{y})$  and  $\phi(\mathbf{y}|\mathbf{x})$  are both Gaussian, we have so far been unable to do the integration analytically. Even in the Gaussian case, the formula for the likelihood contains complicated dependencies with respect to the unknown  $\mathbf{x}$ , e.g., in the normalizing factor of the resulting distribution  $\rho(\mathbf{w}|\mathbf{x})$ . As a consequence, we decided to investigate the existence of alternative approaches that are based on criteria different from MAP, but that can be efficiently implemented and at the same time deliver good reconstructions.

In the alternative approach on which we report in this paper, we aim at finding the label image  $\hat{\mathbf{x}}$  that maximizes the product of  $\pi(\mathbf{x})$  and the term inside the likelihood summation in (2); i.e.,

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} [\pi(\mathbf{x}) \max_{\mathbf{y}} [\chi(\mathbf{w}|\mathbf{y})\phi(\mathbf{y}|\mathbf{x})]]. \quad (5)$$

In other words, the likelihood in (2), which can be viewed as an average over  $\chi(\mathbf{w}|\mathbf{y})$  with respect to  $\phi(\mathbf{y}|\mathbf{x})$ , is replaced by the value  $\max_{\mathbf{y}} [\chi(\mathbf{w}|\mathbf{y})\phi(\mathbf{y}|\mathbf{x})]$ . For the maximization, we propose a coordinate ascent approach; i.e., starting from a gray value image  $\mathbf{y}^{(0)}$ , we alternately maximize

$$F(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{x})\chi(\mathbf{w}|\mathbf{y})\phi(\mathbf{y}|\mathbf{x}) \quad (6)$$

with respect to the label image  $\mathbf{x}$  ( $x$ -step) and with respect to the gray value image  $\mathbf{y}$  ( $y$ -step). We call this maximization the  $M_{xy}$  algorithm. Clearly,  $F(\mathbf{x}, \mathbf{y})$  cannot decrease in any step, so the stationary points are local maxima. They are not likely to be global maxima due to the non-convexity of Gibbs distributions. Fig. 2 shows the corresponding pseudo-code. The termination test checks whether it is the case that  $\mathbf{x}^{(n)} = \mathbf{x}^{(n-1)}$ . In the examples in this paper, we consider the case in which there is a unique  $\arg \max_{\mathbf{y}} [\chi(\mathbf{w}|\mathbf{y})\phi(\mathbf{y}|\mathbf{x})]$  given a measurement  $\mathbf{w}$  and a label image  $\mathbf{x}$ ; therefore the condition  $\mathbf{x}^{(n)} = \mathbf{x}^{(n-1)}$  implies that  $\mathbf{y}^{(n)} = \mathbf{y}^{(n-1)}$ . (In practice, our algorithm only estimates  $\mathbf{y}^{(n)}$ . Nevertheless, since it is a deterministic algorithm, given an  $\mathbf{x}^{(n)}$ , it always yields the same estimate of  $\mathbf{y}^{(n)}$ .)

```

Initialization
 $y^{(0)}$ ,  $n=1$ 

Repeat
  x-step  $x^{(n)} = \arg \max_x [\pi(x) \phi(y^{(n-1)} | x)]$ 

  y-step  $y^{(n)} = \arg \max_y [\chi(w | y) \phi(y | x^{(n)})]$ 

   $n = n + 1$ 

Until
  Termination test == true

```

Fig. 2. Pseudo-code for the  $M_{xy}$  algorithm.

#### 4. Experimental details

In order to test this general approach in the context of tomographic reconstruction of label images from a few projections, we performed some initial experiments. We picked some computer-generated phantom images (which are label images) and their corresponding gray value images (one gray value image for each phantom image). We reconstructed the label images based on the projections by a traditional method and our method. An indicator (defined in Section 5) that tells us how good these reconstructions are was used to evaluate the two approaches. In order to compare our approach with the best one can expect from a traditional approach (recall that in the latter approach, a gray value image is first reconstructed and then segmented to result in a label image), the inputs to the two methods were different. For our approach, we used computer-simulated projection data based on the generated gray value images. For the traditional approach, we took the gray value images themselves; i.e., we assumed a perfect reconstruction of the gray value image. A segmented label image from a perfectly reconstructed gray value image should be the best one can expect from a traditional approach. In particular, we made the following choices.

We considered 2D two-label images of size  $I = 63 \times 63$  and the set of labels  $X = \{0, 1\}$ . We define a *pixel* to be a component of an image (that was defined as a vector in Section 2). All the images were typical samples from the Gibbs distribution  $\pi(\mathbf{x})$  used in Experiment 1 of [12], which has five parameters. The parameters were chosen in such a way that the prior assigns higher probabilities to images that have relatively large uniform regions (which, in this case, are labeled 1) over a background (which, in this case, are labeled 0); such type of images seems to be quite common in many discrete tomography problems [3,5,6,9,12,15]. Although we could have considered a more realistic experiment in which we use images that are representative in some application area (e.g., those corresponding to biological macromolecules in electron microscopy) and then fit them by a model of Gibbs distribution with the five parameters, we did not do this at this stage because we wanted to avoid that the error resulting from the fitting precludes the evaluation of our reconstruction approach. However, we point out that [12] gives examples in which such a model was used to fit some mathematically described phantom images representing sections of a heart chamber, which resulted in an impressive reconstruction quality.

We then randomly picked 50 label images from  $\pi(\mathbf{x})$  using the *Metropolis algorithm* (see, e.g., [21]). This is a stochastic algorithm by which one can obtain samples from an arbitrary Gibbs distribution; see [12] for details of our sampling technique. Next, for each one of these label images, we randomly picked a gray value image (with  $J = I = 63 \times 63$  and  $Y = \mathbb{R}$ ). We assumed a product structure for  $\phi(\mathbf{y}|\mathbf{x})$ , implying that for a label image, a gray value image can be sampled by sampling at each pixel according to the probability  $\phi(y|x)$ . The  $\phi(y|x)$  was Gaussian distributed with mean and variance equal to  $\mu_x$ , where  $\mu_0 = 4$  and  $\mu_1 = 9$ , so that

$$\phi(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^J \phi(y_j|x_j) = \exp \left\{ - \sum_{j=1}^J \left[ \ln Z_j + \frac{(y_j - \mu_{x_j})^2}{2\mu_{x_j}} \right] \right\}, \quad (7)$$

where  $Z_j = (2\pi\mu_{x_j})^{1/2}$ . There is no particular reason for choosing these values of mean and variance, except for the fact that they clearly reflect the idea of overlapping gray values in higher-resolution electron microscopy, as shown by the left histogram in Fig. 1. For sampling the individual Gaussian distributions we used the method explained in [16].

We made two different choices for the conditional probability  $\chi(\mathbf{w}|\mathbf{y})$  of having a measurement vector  $\mathbf{w}$  given a gray value image  $\mathbf{y}$ .

The first choice was based on a traditional approach of first reconstructing the gray value image and then segmenting it to obtain a label image. As mentioned earlier, for the comparison between our method and a traditional approach, we consider an “ideal” case, in which the reconstruction of the gray value image is exactly the gray value image itself. We can bring this ideal into our framework by setting  $K = J$  and  $\mathbf{W} = \mathbf{Y}$  and choosing  $\chi(\mathbf{w}|\mathbf{y}) = \delta(\mathbf{w} - \mathbf{y})$ , where  $\delta$  is the impulse at the origin (the Dirac function). Using (2), we see that in this case  $\rho(\mathbf{w}|\mathbf{x}) = \phi(\mathbf{w}|\mathbf{x})$ .

The second choice was a more realistic one; this was the one to be used for evaluating our approach. The projections forming the measurement vector  $\mathbf{w}$  were simulated as follows. There were either three, four, or eight projections using parallel lines in each projection, such that the tangent of the angle, denoted by  $\alpha$ , between these lines and the “positive horizontal direction” was 0, infinity, and  $-1$  in the case of three projections; it was 0, infinity,  $-1$ , and 1 in the case of four projections; and it was 0, infinity,  $-1$ , 1,  $-0.5$ ,  $0.5$ ,  $-2$ , and 2 in the case of eight projections. The distance between two consecutive parallel lines in all the cases was  $l \max(\cos \alpha, \sin \alpha)$ , where  $l$  is the length of a side of a pixel. When the tangent of  $\alpha$  was 0, infinity,  $-1$ , or 1, the parallel lines intersect the centers of the pixels; and when the tangent of  $\alpha$  was  $-0.5$ ,  $0.5$ ,  $-2$ , and 2, the parallel lines intersect (if at all) the upper vertexes of the upper border pixels and the right vertexes of the right border pixels. (Note that, for any fixed direction, the length of intersection of a line with a pixel is the same for all lines and pixels.) This choice of the location of the lines is not a necessary condition for the validity of the discussed approach, but it simplifies the implementation of the proposed method. For each experiment we formed a  $K$ -dimensional column vector  $\mathbf{z} = (z_1, \dots, z_K)^t$  from the sums of the pixel values in the gray value image along the  $K$  lines in all the (three, four, or eight) projections. We chose  $\psi_k(w_k|\mathbf{y})$ , for  $1 \leq k \leq K$ , to be Gaussian distributed with mean  $z_k$  and variance equal to  $0.01 z_k$  (our way of handling the case  $z_k = 0$  is discussed

in Section 4.1); i.e.,

$$\chi(\mathbf{w}|\mathbf{y}) = \prod_{k=1}^K \psi_k(w_k|\mathbf{y}) = \exp \left\{ - \sum_{k=1}^K \left[ \ln Z'_k + \frac{(w_k - z_k)^2}{0.02z_k} \right] \right\}, \tag{8}$$

where  $Z'_k = (0.02\pi z_k)^{1/2}$ .

#### 4.1. Estimation methods for the first choice for $\chi(\mathbf{w}|\mathbf{y})$

We estimated the label images from the gray value images using both the MAP estimator of (4) and the *maximum likelihood* (ML) estimator

$$\mathbf{x}_{\text{ML}} = \arg \max_{\mathbf{x}} \phi(\mathbf{w}|\mathbf{x}) = \arg \max_{\mathbf{x}} \left\{ \prod_{k=1}^K \frac{1}{Z_k} \exp \left[ - \frac{(w_k - \mu_{x_k})^2}{2\mu_{x_k}} \right] \right\}. \tag{9}$$

We used a *simulated annealing* [4,21] via the Metropolis algorithm (see [12] for details of the choice of an annealing schedule) for finding the MAP estimator. The ML approach corresponds to thresholding the gray value image. This is because the maximization of product form of the right-hand side of (9) can be done factor by factor, and the maximization of each factor corresponds to choosing the label 0 or 1, for which the given measurement  $w_k$ , for  $1 \leq k \leq K$ , has a higher probability (density) of occurring. The two Gaussian distributions (with both the mean and the variance equal to  $\mu_0 = 4$  in one and  $\mu_1 = 9$  in the other) assign equal probability to the value 6.4683 . . . , which is therefore the level to be used for thresholding.

#### 4.2. Estimation method for the second choice for $\chi(\mathbf{w}|\mathbf{y})$

We estimated the label images using our  $M_{xy}$  algorithm based on the projections. By substituting (1), (8), and (7) into (6), we get that

$$F(\mathbf{x}, \mathbf{y}) = \left( \frac{1}{Z} \prod_{k=1}^K \frac{1}{Z'_k} \prod_{j=1}^J \frac{1}{Z_j} \right) \exp \left[ -H(\mathbf{x}) - \sum_{k=1}^K \frac{(w_k - z_k)^2}{0.02z_k} - \sum_{j=1}^J \frac{(y_j - \mu_{x_j})^2}{2\mu_{x_j}} \right]. \tag{10}$$

Since the  $1/Z$  factor in this expression does not depend on either  $\mathbf{x}$  or  $\mathbf{y}$ , it can be ignored during the optimization process. The same cannot be said for the  $1/Z'_k$  factors. We have found an efficient algorithm for the  $y$ -step of the  $M_{xy}$  algorithm that depends on a further approximation: the replacement of the variances  $0.01z_k$  in (8) by  $0.01 \max(\mu_0, w_k)$ , for  $k = 1, \dots, K$ . With this approximation, the task of maximizing  $F(\mathbf{x}, \mathbf{y})$  is replaced by that

of maximizing

$$\left( \prod_{k=1}^K \frac{1}{\tilde{Z}_k} \prod_{j=1}^J \frac{1}{Z_j} \right) \exp \left[ -H(\mathbf{x}) - \sum_{k=1}^K \frac{(w_k - z_k)^2}{0.02\tilde{w}_k} - \sum_{j=1}^J \frac{(y_j - \mu_{x_j})^2}{2\mu_{x_j}} \right], \quad (11)$$

where  $\tilde{w}_k = \max(\mu_0, w_k)$  and  $\tilde{Z}_k = (0.02\pi\tilde{w}_k)^{1/2}$ .

For the initial  $\mathbf{y}^{(0)}$  we took the gray value image produced by a reconstruction based on the (noisy) projections using, in this case, the *algebraic reconstruction technique* (ART) described in [7] (Chapter 11, Eqs. (1.1) and (1.2); see also the Appendix) with 256 cycles through the data, initial image vector with all its components set to zero and relaxation parameter equal to 0.5. The data were accessed so that all the lines in one projection were processed before going to the next projection, and the sequence of projections (with no repetition in one cycle) were chosen so that the angle between any pair of consecutive projection directions was as large as possible. To be precise, the tangents of the projection directions were ordered as 0, infinity, and  $-1$  in the case of three projection; they were ordered as 0, infinity,  $-1$ , and 1 in the case of four projections; and they were ordered as 0, infinity,  $-1$ , 1,  $-0.5$ , 0.5,  $-2$ , and 2, in the case of eight projections. The  $M_{xy}$  algorithm then alternately performed the  $x$ -step and the  $y$ -step as described in the pseudo-code of Fig. 2.

#### 4.2.1. The $x$ -step

For the maximization with respect to  $\mathbf{x}$ , we computed the MAP estimator from the current gray value image  $\mathbf{y}^{(n-1)}$ :

$$\mathbf{x}^{(n)} = \arg \max_{\mathbf{x}} \left\{ \left( \prod_{j=1}^J \frac{1}{Z_j} \right) \exp \left[ -H(\mathbf{x}) - \sum_{j=1}^J \frac{(y_j^{(n-1)} - \mu_{x_j})^2}{2\mu_{x_j}} \right] \right\}, \quad (12)$$

using, again, simulated annealing via the Metropolis algorithm. The product structure of both the prior  $\pi(\mathbf{x})$  and  $\phi(\mathbf{y}|\mathbf{x})$  is what allows a relatively efficient implementation [20] of the  $x$ -step.

#### 4.2.2. The $y$ -step

Note that since each of the  $z_k$ 's is a linear combination of the  $y_j$ 's, the last two terms inside the exponential in (11) form a quadratic in  $\mathbf{y}$ . Hence, maximizing (11) with respect to  $\mathbf{y}$  is equivalent to minimizing the quadratic

$$q(\mathbf{y}) = (\mathbf{w} - R\mathbf{y})^t \Gamma^{-1} (\mathbf{w} - R\mathbf{y}) + (\mathbf{y} - \mu_{\mathbf{x}^{(n)}})^t \Sigma^{-1} (\mathbf{y} - \mu_{\mathbf{x}^{(n)}}), \quad (13)$$

where  $R$  is the *projection matrix* that relates  $\mathbf{y}$  with  $\mathbf{z}$ :  $R\mathbf{y} = \mathbf{z}$  ( $R$  is a sparse matrix), the vector  $\mu_{\mathbf{x}^{(n)}}$  is equal to  $(\mu_{x_1^{(n)}}, \dots, \mu_{x_J^{(n)}})^t$  and  $\Gamma$  and  $\Sigma$  are nonsingular diagonal matrices with entries equal to, respectively,  $0.01\tilde{w}_k$ , for  $k = 1, \dots, K$ , and  $\mu_{x_j^{(n)}}$ , for  $j = 1, \dots, J$ .

These  $\Gamma$  and  $\Sigma$  are positive definite and, consequently, the minimizer of  $q(\mathbf{y})$  is unique (see, e.g., Section 12.1 of [7]). This assures that our termination test of  $\mathbf{x}^{(n)} = \mathbf{x}^{(n-1)}$ , for  $n = 1, 2, \dots$ , implies that  $\mathbf{y}^{(n)} = \mathbf{y}^{(n-1)}$ . Among the various methods for the minimization of (13) (e.g., conjugate gradient methods, the steepest descent method, etc.) we resorted to an

easy-to-implement row-action iterative algorithm (an ART) with relatively fast convergence properties. The main idea is to construct from the original system  $\mathbf{R}\mathbf{y} = \mathbf{z}$  of equations (that is most likely to be inconsistent) a consistent system of linear equations, in such a way that there is a one-to-one onto continuous mapping between the solutions of this system and  $J$ -dimensional vectors  $\mathbf{y}$ . Furthermore, the squared (Euclidean) norm of a solution of the new system is equal to the value of the quadratic (13) for the corresponding  $\mathbf{y}$ . An ART algorithm that yields iterates converging to the (unique) minimum norm solution of the new system can then be adapted to converge to the minimizer of the quadratic. In [7] an algorithm is described that optimizes (13) for matrices  $\Gamma$  and  $\Sigma$  that are proportional to the identity matrices of appropriate sizes. Generalized to our case, the algorithm for finding  $\mathbf{y}$  that minimizes  $q(\mathbf{y})$  is the following (see the Appendix for the derivations).

During the  $n$ th  $y$ -step (see the pseudo-code in Fig. 2)

$$\mathbf{u}^{(0)} \text{ is the } K\text{-dimensional column vector of zeros,} \tag{14}$$

$$\mathbf{y}^{(n,0)} = \mu_{\mathbf{x}^{(n)}}, \tag{15}$$

$$\mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + c^{(m)}\gamma_{k_m}, \tag{16}$$

$$\mathbf{y}^{(n, m+1)} = \mathbf{y}^{(n, m)} + c^{(m)}\Sigma\mathbf{r}_{k_m}, \tag{17}$$

with

$$c^{(m)} = \frac{w_{k_m} - \mathbf{r}_{k_m}^t \mathbf{y}^{(n, m)} - \gamma_{k_m}^t \mathbf{u}^{(m)}}{\gamma_{k_m}^t \gamma_{k_m} + \mathbf{r}_{k_m}^t \Sigma \mathbf{r}_{k_m}}, \tag{18}$$

where  $k_m$  denotes  $m \pmod K + 1$ ,  $\gamma_{k_m}$  and  $\mathbf{r}_{k_m}$  are the respective transposes of the  $k_m$ th row of  $\Gamma^{1/2}$  and of  $R$  ( $\mathbf{u}^{(m)}$ ,  $\mathbf{y}^{(n, m)}$ , and  $\mu_{\mathbf{x}^{(n)}}$  are all column vectors, for  $m = 0, 1, \dots$ ). Specifically, for the experiments for this paper, we chose 256 cycles through the data (i.e., the output of the  $n$ th  $y$ -step is  $\mathbf{y}^{(n)} = \mathbf{y}^{(n, 256K)}$ ), and data were accessed in the same way as was done for obtaining  $\mathbf{y}^{(0)}$ . We note that the  $y$ -step takes up much less CPU time than the  $x$ -step (a consequence of our approximation in (8) of  $z_k$  by  $\tilde{w}_k$ ). In a 1.7 Ghz Pentium III, an  $x$ -step takes up to 15 min, while a  $y$ -step consumes about 5 s (for the case of eight projections).

### 5. Experimental results

In Table 1 we indicate the quality of the MAP and ML estimators based on the gray value image (Section 4.1) as well as of our proposed estimator (Section 4.2) based on three, four, and eight projections. The reconstruction quality is measured by the average (over the 50 images of the experiment) percentage of correctly classified labels in the estimates. Fig. 3 shows actual reconstructions by the three estimators. The algorithm  $M_{xy}$  converges quite fast to a (local) optimum: at most two iterations were necessary (for each of the 50 label images, using either three, four, or eight projections).

The significance of the difference of the reconstruction qualities between our proposed method and thresholding of the *exact* gray value image was measured using the pairwise

Table 1  
Quality of reconstruction

Estimator	Measurements	Percentage of correct classifications
MAP	Gray values	99.4±0.2
ML (thresholding)	Gray values	86.2±0.6
By our method	8 projections	94.6±2.1
By our method	4 projections	88.8±3.7
By our method	3 projections	86.2±4.0

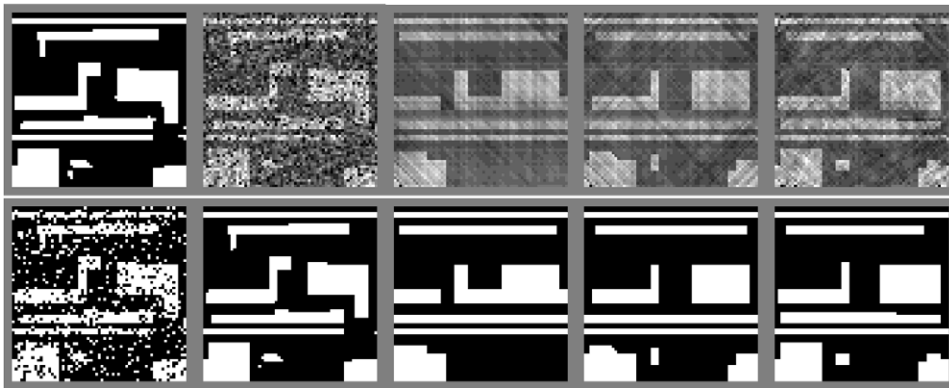


Fig. 3. One of the 50 phantom label images (top left) together with its ML estimate (bottom left) and MAP estimate (bottom of the second column) based on the gray value image itself (top of the second column). The remaining three columns are the reconstructions produced by our proposed  $M_{xy}$  algorithm based on three (third column), four (fourth column), and eight projections (right-most column): the images on the top and the bottom are, respectively, the final gray value images and the final label images. For the reconstructed label images on the bottom, the number of misclassified labels are, respectively, for all the columns, 537, 25, 581, 399, 280. Black and white correspond to, respectively, label 0 (with  $\mu_0 = 4$ ) and label 1 (with  $\mu_1 = 9$ ). The gray value images are displayed in the range 0–13.

*t*-test. We found that our method based on four or eight projections is statistically significantly superior and our method based on three projections performs as well as thresholding. However, we remind the reader that the gray value image being thresholded is assumed to be perfectly reconstructed (see Section 4).

## 6. Conclusions and discussion

Our approach to reconstructing label images from at least four projections outperforms the common practice of thresholding reconstructions. When three projections are used, our method yields reconstructions as good as thresholding the exact gray value image. However, there is hope for further improvement, since MAP estimation from gray values provides even better results.

Our algorithm is greedy in the usual sense that the objective function  $F(\mathbf{x}, \mathbf{y})$  cannot decrease in any step of an iteration. Thus, the stationary points of the  $x$ -step and the  $y$ -step correspond to local optima. Searching for the global optimum instead is an important task, since it may deliver a superior reconstruction. However, due to the presence of the Gibbs distribution  $\pi(\mathbf{x})$  in the  $F(\mathbf{x}, \mathbf{y})$  of (6), we are not aware of an efficient method for global optimization of  $F(\mathbf{x}, \mathbf{y})$ . The Gibbs distribution is defined on labels, rather than on continuous numerical variables; therefore we are forced to use some kind of combinatorial optimization technique that is further hampered by the presence of very many local optima. Another reason why we did not insist on finding the global optimum is (recall Section 3) that our original goal was to find the optimizer of the product  $\pi(\mathbf{x})\rho(\mathbf{w}|\mathbf{x})$  of (3), which is unlikely to be the same as that provided by optimizing  $F(\mathbf{x}, \mathbf{y})$ . Indeed, the discovery of an efficient algorithm for finding the  $\mathbf{x}_{\text{MAP}}$  of (4) (given a measurement vector  $\mathbf{w}$ ) is an important open problem whose difficulty lies not only in the already mentioned problems due to the Gibbs prior in the function to be optimized, but also in the complicated nature of the other term (the likelihood function of (2) that involves a summation over all the gray value images).

### Appendix

We present an ART approach (see Chapter 11 of [7]) for finding the minimizer  $\mathbf{y}^*$  of the quadratic (13).

Consider a consistent system of linear equations

$$A\mathbf{g} = \mathbf{b} \tag{19}$$

with  $S$  unknowns and  $L$  equations. For  $1 \leq l \leq L$ , let  $A_l$  be the transpose of the  $l$ th row of  $A$  and  $b_l$  be the  $l$ th component of  $\mathbf{b}$ . The following algorithm is a relaxation method for solving system (19) of equalities (Section 12.2 of [7]).

$$\mathbf{g}^{(0)} \text{ is the } S\text{-dimensional columnvector of zeros,} \tag{20}$$

$$\mathbf{g}^{(m+1)} = \mathbf{g}^{(m)} + c^{(m)} A_{k_m}, \tag{21}$$

with

$$c^{(m)} = \frac{(b_{k_m} - A_{k_m}^t \mathbf{g}^{(m)})}{A_{k_m}^t A_{k_m}}, \tag{22}$$

where  $k_m = m \pmod K + 1$ . The sequence of the  $\mathbf{g}^{(m)}$  converges to the (unique) minimum (Euclidean) norm solution of system (19).

Using now the notation in (13), consider the system

$$[\Gamma^{1/2} \quad R\Sigma^{1/2}] \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = (\mathbf{w} - R\mu_{\mathbf{x}^{(n)}}), \tag{23}$$

where the system matrix  $[\Gamma^{1/2} \ R\Sigma^{1/2}]$  is of size  $K \times (K + J)$ , the unknown  $\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$  is a column vector of size  $K + J$ , and the right-hand side  $(\mathbf{w} - R\mu_{\mathbf{x}(n)})$  is a column vector of size  $K$ . The following result is easy to prove.

**Lemma.** For any  $K$ -dimensional vector  $\mathbf{u}$  and  $J$ -dimensional vector  $\mathbf{v}$ ,  $\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$  is a solution of (23) if, and only if, there exists a  $J$ -dimensional vector  $\mathbf{y}$  such that

$$\mathbf{u} = \mathbf{u}(\mathbf{y}) = \Gamma^{-1/2}(\mathbf{w} - R\mathbf{y}) \quad (24)$$

and

$$\mathbf{v} = \mathbf{v}(\mathbf{y}) = \Sigma^{-1/2}(\mathbf{y} - \mu_{\mathbf{x}(n)}). \quad (25)$$

This lemma implies in particular that (23) is a consistent system of linear equations. The squared norm of any of its solutions can be written, in view of (24) and (25), as

$$(\mathbf{w} - R\mathbf{y})^t \Gamma^{-1}(\mathbf{w} - R\mathbf{y}) + (\mathbf{y} - \mu_{\mathbf{x}(n)})^t \Sigma^{-1}(\mathbf{y} - \mu_{\mathbf{x}(n)}), \quad (26)$$

which is precisely the quadratic  $q(\mathbf{y})$  of (13). It follows that if  $\mathbf{u}^{(m)}$  and  $\mathbf{v}^{(m)}$  are sequences of  $K$ -dimensional, respectively,  $J$ -dimensional, vectors such that  $\begin{bmatrix} \mathbf{u}^{(m)} \\ \mathbf{v}^{(m)} \end{bmatrix}$  converges to the minimum norm solution of (23) and if, for all  $m$ , we define

$$\mathbf{y}^{(n,m)} = \Sigma^{1/2}\mathbf{v}^{(m)} + \mu_{\mathbf{x}(n)}, \quad (27)$$

then  $\mathbf{y}^{(n,m)}$  converges to the minimizer of  $q(\mathbf{y})$  in (13).

In order to apply the general algorithm described in (20)–(22) to obtain a sequence  $\begin{bmatrix} \mathbf{u}^{(m)} \\ \mathbf{v}^{(m)} \end{bmatrix}$  that converges to the minimum norm solution of (23), we observe that for the special case

$$A_{k_m}^t = [\gamma_{k_m}^t \ \mathbf{r}_{k_m}^t \Sigma^{1/2}] \quad (28)$$

and

$$b_{k_m} = w_{k_m} - \mathbf{r}_{k_m}^t \mu_{\mathbf{x}(n)}^{(m)}, \quad (29)$$

where  $\gamma_{k_m}$  and  $\mathbf{r}_{k_m}$  are as defined after (18). This combined with (27) yields the algorithm described in (14)–(18).

As noted just above, the sequence  $\mathbf{y}^{(n,m)}$  converges to a minimizer of  $q(\mathbf{y})$  as  $m$  goes to infinity.

## References

- [1] J.M. Carazo, C.O. Sorzano, E. Rietzel, R. Schröder, R. Marabini, Discrete tomography in electron microscopy, in: G.T. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999, pp. 405–416.

- [2] B.M. Carvalho, G.T. Herman, S. Matej, C. Salzberg, E. Vardi, Binary tomography for triplane cardiography, in: A. Kuba, M. Samal, A. Todd-Pokropek (Eds.), *Information Processing in Medical Imaging*, Springer, Berlin, 1999, pp. 29–41.
- [3] G.S. Cunningham, K. Hanson, X.L. Battle, Three-dimensional reconstructions from low-count SPECT data using deformable models, *Opt. Express* 2 (1998) 227–236.
- [4] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, Wiley, New York, 2001.
- [5] J.A. Fessler, Segmented attenuation correction for PET, in: *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference*, 1992, pp. 1182–1184.
- [6] J.J. Gerbrands, C.H. Slumps, 3D reconstruction of homogeneous objects from two Poisson-distributed projections, *Pattern Recognition Lett.* 3 (1985) 137–145.
- [7] G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, 1980.
- [8] G.T. Herman, B.M. Carvalho, Multiseeded segmentation using fuzzy connectedness, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2001) 460–474.
- [9] G.T. Herman, A. Kuba, *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999.
- [10] K.H. Höhne, W.A. Hanson, Interactive 3D segmentation of MRI and CT volumes using morphological operations, *J. Comput. Assist. Tomo.* 16 (1992) 285–294.
- [11] H. Kudo, H. Nakamura, Segmented attenuation map reconstruction from incomplete transmission data, *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2000, pp. 1358–1362.
- [12] H.Y. Liao, G.T. Herman, Automated estimation of the parameters of Gibbs priors to be used in binary tomography, *Discrete Appl. Math.* 139 (2004) 149–170.
- [13] S. Matej, G.T. Herman, A. Vardi, Binary tomography on the hexagonal grid using Gibbs priors, *Internat J. Imag. System Tech.* 9 (1998) 126–131.
- [14] F. Natterer, F. Wuebbeling, *Mathematical Methods in Image Reconstructions*, SIAM, Philadelphia, 2001.
- [15] C. Pellot, A. Herment, M. Sigelle, P. Horain, H. Maitre, P. Peronneau, A 3D reconstruction of vascular structures from two X-ray angiograms using an adapted simulated annealing algorithm, *IEEE Trans. Medical Imaging* 13 (1994) 48–60.
- [16] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, New York, 1993.
- [17] N. Robert, F. Peyrin, M.J. Yaffe, Binary vascular reconstruction from a limited number of cone beam reconstructions, *Med. Phys.* 21 (1994) 1839–1851.
- [18] P.K. Sahoo, S. Soltani, A.K.C. Wong, Y.C. Chen, A survey of thresholding techniques, *Comput. Vision Graphics and Image Processing* 41 (1996) 233–260.
- [19] M. Senasli, L. Garnero, A. Herment, E. Mousseaux, 3D reconstruction of vessel lumen from very few angiograms by dynamic contours using a stochastic approach, *Graphical Models* 62 (2000) 105–127.
- [20] E. Vardi, G.T. Herman, T.Y. Kong, Speeding up stochastic reconstructions of binary images from limited projection directions, *Linear Algebra Appl.* 339 (2001) 75–89.
- [21] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer, Berlin, 1995.