

EXPECTATION MAXIMIZATION FOR EMISSION TOMOGRAPHY

MORITZ BLUME

MATHEMATICAL PRELIMINARIES

Notation. Vectors like \mathbf{v} will be denoted by roman (= serif) bold italic letters. v_i denotes the i -th component of \mathbf{v} . Matrices will be written in upper-case roman bold italic letters, e. g. \mathbf{M} , M_{ij} being an entry at row i and column j .

Images will be stored in vectors by lexicographic ordering. For a two dimensional image with dimension $m \times n$ the order of pixels stored in a vector is $(0, 0)$, $(1, 0)$, \dots , $(m, 0)$, $(0, 1)$, \dots , $(m, 1)$, \dots , $(0, n)$, \dots , (m, n) . For three dimensional images this works completely analog.

Probability Theory. Since the EM framework involves probabilistic modeling, the reader should be familiar with some basic concepts from probability theory. The text should at least cover topics like random variables, expectation values, conditional probabilities, conditional expectations and Poisson distributions and its properties. We recommend our own tutorial [2] which is still a draft but at least covers all topics necessary for the understanding of this tutorial.

1. INTRODUCTION

The basic problem of image reconstruction in Emission Tomography (ET) - be it SPECT or PET - is to estimate an original image vector \mathbf{f} from the base of measured count numbers stored in a vector \mathbf{g} . An entry f_i stores the number of events that took place inside voxel i during a specific measurement period. g_j stores the number of events that occurred in tube j . A tube in PET is defined by two detector elements, in SPECT by one detector element.

The estimation is only possible if some information about the imaging process is available. A very essential information is the probability of detecting an event originated from box j in detector tube i :

$$(1) \quad P(\text{event detected in tube } i \mid \text{event occurred in box } j) = H_{ij} \quad ,$$

where \mathbf{h} is called the system matrix.

The average/expected number of events detected in tube i is then $\mathbb{E}[g_i] = \sum_j H_{ij} f_j$. In matrix-vector notation:

$$(2) \quad \mathbb{E}[\mathbf{g}] = \mathbf{H} \mathbf{f} \quad .$$

In essence, image reconstruction algorithms try to invert this equation in order to solve for the image \mathbf{f} . One of the best known and mostly applied algorithms

is the Expectation Maximization algorithm. The general scheme was proposed by Dempster *et al.* in 1977 [3]. It is used in numerous scientific disciplines. Since the algorithm is so general, it has to be adopted to each specific case. For ET, this has been done in 1982 by Shepp and Vardi [4].

2. EXPECTATION MAXIMIZATION FOR EMISSION TOMOGRAPHY

If you are not familiar with the EM algorithm you will have to refer to additional sources. We especially recommend our own tutorial [1] and the cited sources there.

Remember that the EM algorithm is a general iterative scheme that helps to solve maximum likelihood problems by introducing a so-called hidden random variable.

The predominant questions of an application of an EM scheme to ET are: Why do we want to apply the EM algorithm to ET? How do we apply it?

The answer to the first question is not that simple. In fact, exactly the same iterative formula that results from the EM scheme can be derived by simply calculating the maximum likelihood estimate of the incomplete data (as will be shown in the next section). The main reason why it is very convenient to know that we are using an EM algorithm is that we do not have to study things like convergence etc. anymore, since this is already backed up by the numerous theoretical results about the EM algorithm.

The answer to the second question is more difficult and we start by asking: What is the *complete* and what the *incomplete* data in the case of ET?

As mentioned, the *complete* data term as used in the paper by Dempster *et al.* is an artificial description for all the data that is necessary in order to estimate the sought distribution parameters. In case of ET, \mathbf{g} is the measured/observed data and will be considered as *incomplete*, since we only know *that* something happened on a certain tube but not *from where* exactly it originates. Accordingly, the complete data will be defined as a matrix \mathbf{G} , where G_{ij} represents the number of detected events in tube i originated from box j . Of course, if we could observe \mathbf{G} directly, the solution to the whole problem would be simple:

$$(3) \quad \hat{f}_j = \sum_i G_{ij} \ .$$

We continue with developing the complete data likelihood function:

$$(4) \quad L(\mathbf{f}) = P(\mathbf{G} \mid \mathbf{f})$$

$$(5) \quad = \prod_i \prod_j P(G_{ij} \mid \mathbf{f})$$

$$(6) \quad = \prod_i \prod_j e^{-\mathbb{E}[G_{ij}]} \frac{\mathbb{E}[G_{ij}]^{G_{ij}}}{G_{ij}!} \ .$$

We get from (4) to (5) by taking advantage of the independency of the random variables G_{ij} , and from (5) to (6) by inserting the definition of the Poisson distribution.

Since maximizing the log-likelihood gives the same result but is easier to do we define

$$\begin{aligned}
 (7) \quad l(\mathbf{f}) &= \ln L(\mathbf{f}) \\
 (8) \quad &= \sum_i \sum_j -\mathbb{E}[G_{ij}] + G_{ij} \ln \mathbb{E}[G_{ij}] - \ln G_{ij}!
 \end{aligned}$$

Since $\mathbb{E}[G_{ij}]$ is the expected number of emissions from voxel j measured in tube i , it is directly related to the probability H_{ij} of an emission from voxel j being measured in tube i :

$$(9) \quad \mathbb{E}[G_{ij}] = f_j H_{ij}$$

So, the log-likelihood is

$$(10) \quad l(\mathbf{f}) = \sum_i \sum_j -f_j H_{ij} + G_{ij} \ln f_j H_{ij} - \ln G_{ij}! .$$

Since l depends on G_{ij} , which are unknown measurements, we cannot directly calculate the ML estimate for \mathbf{f} . That's where the EM algorithm comes into the game: we treat l as if it was a random variable, since in fact it *is* a random variable because a function that depends on a random variable (G_{ij} in this case!) is a random variable by itself! So, instead of maximizing l directly, we maximize its expected value. The EM algorithm can be written down in one line:

$$(11) \quad \hat{\mathbf{f}}^{(n+1)} = \underset{\mathbf{f}}{\operatorname{argmax}} \mathbb{E} \left[l(\mathbf{f}) | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right]$$

2.1. Expectation-Step. The expectation step consists in calculating

$$(12) \quad \mathbb{E} \left[l(\mathbf{f}) | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right] .$$

We insert the log-likelihood function from (10):

$$(13) \quad \mathbb{E} \left[\sum_i \sum_j -f_j H_{ij} + G_{ij} \ln f_j H_{ij} - \ln G_{ij}! \mid \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right] .$$

Due to linearity of expectation we get

$$(14) \quad \sum_i \sum_j \left(-f_j H_{ij} + \mathbb{E} \left[G_{ij} | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right] \ln f_j H_{ij} - \mathbb{E} \left[\ln G_{ij}! | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right] \right) .$$

Now, in order to proceed, we have to calculate the expectation $\mathbb{E} \left[G_{ij} | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right]$. This is a bit tricky. Since $g_i = \sum_j G_{ij}$, G_{ij} is conditioned to the fact that we already know about the sum $\sum_j G_{ij}$. Probability theory states that if we have independent Poisson random variables, the conditional probability distribution given the sum of its values is a Binomial distribution with parameters $\left(\sum_j G_{ij}, \frac{\mathbb{E}[G_{ij}]}{\sum_j \mathbb{E}[G_{ij}]} \right)$. With the expectation of a Binomial distribution (a, b) being ab and with $\mathbb{E}[G_{ij}] = f_j H_{ij}$, we get

$$(15) \quad \mathbb{E} \left[G_{ij} | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right] = g_i \frac{\hat{f}_j^{(n)} H_{ij}}{\sum_k \hat{f}_k^{(n)} H_{ik}} .$$

Intuitively the expectation of G_{ij} is corrected by the real number of measured events compared to the number of measured events according to the image estimate (note that $\frac{g_i}{\sum_k \hat{f}_k^{(n)} H_{ik}}$ is a value close or equal to one...).

We are not interested in the second expectation since it will fall away anyway in the maximization step...

2.2. Maximization-Step. Now, since the expectation is developed, we can go on in the EM algorithm (refer (11)) and maximize this expectation. We will do this by setting the derivative equal to zero:

$$(16) \quad \frac{\partial}{\partial f_l} \mathbb{E} \left[l(\mathbf{f}) | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right] = 0$$

$$(17) \quad \Leftrightarrow \sum_i -H_{il} + \sum_i \mathbb{E} \left[G_{il} | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right] \frac{1}{f_l} = 0$$

$$(18) \quad \Leftrightarrow f_l = \frac{\sum_i \mathbb{E} \left[G_{il} | \mathbf{g}, \hat{\mathbf{f}}^{(n)} \right]}{\sum_i H_{il}}$$

$$(19) \quad \Leftrightarrow f_l = \frac{\hat{f}_l^{(n)}}{\sum_i H_{il}} \sum_i \frac{H_{il} g_i}{\sum_k \hat{f}_k^{(n)} H_{ik}} .$$

In fact, this f_l is the l -th component of our new estimate, and so we denote the EM algorithm finally as

$$(20) \quad \hat{f}_l^{(n+1)} = \frac{\hat{f}_l^{(n)}}{\sum_i H_{il}} \sum_i \frac{H_{il} g_i}{\sum_k \hat{f}_k^{(n)} H_{ik}} .$$

REFERENCES

1. Moritz Blume, *Expectation maximization: A gentle introduction*, 2008.
2. ———, *An introduction to probability theory*, 2008.
3. A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society. Series B (Methodological) **39** (1977), no. 1, 1–38.
4. L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE TRANSACTIONS ON MEDICAL IMAGING **1** (1982), no. 2, 113–122.

MORITZ BLUME, TECHNISCHE UNIVERSITÄT MÜNCHEN, INSTITUT FÜR INFORMATIK / I16, BOLTZMANNSTRASSE 3, 85748 GARCHING B. MÜNCHEN, GERMANY

E-mail address: `blume@cs.tum.edu`

URL: `http://www.navab.cs.tum.edu/Main/MoritzBlume`