Probabilistic model based methods for data fusion

Ali Mohammad–Djafari

Laboratoire des Signaux et Systèmes (CNRS–SUPELEC–UPS)
École Supérieure d'Électricité
Plateau de Moulon, 3 rue Joliot-Curie, 91192 Gif–sur–Yvette Cedex, France.
E_mail: djafari@lss.supelec.fr

Abstract

The main objective of this paper is to show how classical probabilistic methods such as Maximum Entropy (ME), maximum likelihood (ML) and/or Bayesian inference approaches can be used in data fusion problems. In the first part of this paper, these three approaches are considered separately and their basic foundations and their inter-relations are described briefly . In the second part, a few academic data fusion problem examples are given and the way these tools can be used to give solutions to them are described. The main conclusion of this part is that only the Bayesian approach can handle realistic data fusion problems. In the third part, an X ray computed tomography (CT) image reconstruction problem using two different kind of data has been considered. First, we considered the case where these data are the classical X-rays radiographic data and ultrasound echo-graphic data and analyzed different possible solutions within the Bayesian estimation approach. Finally, in the last part, we considered an easier problem which is the fusion of radiographic data and some geometrical informations coming from anatomical atlas, and proposed realistic new methods for this data fusion problem. A few preliminary simulation results show the performances of the proposed method and present the advantages of using geometrical data in CT image reconstruction.

key words: Data fusion, Maximum entropy, Maximum likelihood, Bayesian data fusion, EM algorithm, Computed tomography, Fusion of radiography and anatomic data.

1 Introduction

Data fusion is one of the active area of research in many applications such as non destructive testing (NDT), geophysical imaging, medical imaging, radio-astronomy, etc. The main objective of this paper is not to focus on any of these applications but to show how classical probabilistic methods such as Maximum Entropy (ME), maximum likelihood (ML) and/or Bayesian (BAYES) approaches can be used to do data fusion.

In the first part of the paper, we give a brief review of these methods to see explicitly that:
– ME can be used to assign a probability law to an unknown quantity when we have exact macroscopic data (expectations) on it;
– ML can be used to estimate the parameters of a probability law when we have exact microscopic data;
– Bayesian approach can be used to update a prior probability law when we have microscopic data either direct or indirect through a model based observation;
– When we have both microscopic and macroscopic data, we can use first ME to assign a prior and then the Bayesian approach to update it to the posterior law, thus doing the desired data fusion.

In the second part, first we give a few academic data fusion problem examples such as macroscopic (or low resolution) and microscopic (or high resolution) data fusion and will see how these tools can be used to give solutions to them. The main objective here is to illustrate the use of the three previous approaches for such simple data fusion problems. The main conclusion here is that only the Bayesian approach can handle realistic data fusion problems.

In the third part, an X ray computed tomography (CT) image reconstruction problem using two different kind of data has been considered. First, we considered a case of Non Destructive Testing (NDT) application where these data are the classical X-rays radiographic data and ultrasound echo-graphic data. For these kind of problems, we proposed new methods based on the Bayesian maximum *a posteriori* estimation approach. X ray data give more information on the regions volumes inside the body, while ultrasound echo-graphic data give more information about the regions borders. The main idea is how to model the object under test to be able to relate its parameters to both kind of data giving the possibility of data fusion. The main tool here is modeling regions and their borders through a compound Markov random field. Different possible reconstruction schemes ar then proposed and their implementation complexities are analyses.

Finally, in the last part, we considered a CT medical imaging problem using two different kind of data: classical radiographic (projection) data and some geometrical informations coming from anatomical atlas such as partial knowledge of the anatomic regions and/or the borders of these regions. This problem is easier than the previous case, because the region borders are given directly.

The idea of using anatomical data in medical CT imaging is not new. Many works on the subject has been done before. See for example [1, 2, 3, 4]. In [1], the authors proposed methods for using regions borders from anatomical data in medical imaging and the authors in [2, 3, 4] used the knowledge of some of regions themselves. While the application in the first reference concerns medical imaging, the application in the second references concerns industrial non destructive testing. But, combining both regions and borders informations from anatomic data is new. This work is still in development. We give here some preliminary results simulating a fan beam CT problem with limited number

of X ray data. A few results show comparisons of the results using classical back-projection or filtered back-projection methods with those obtained by the proposed method either using or not the anatomic data. These results show the advantages of using anatomic data when these data are exact and well registered with radiographic data. Some preliminary results show also the sensitivity of the proposed method to some errors in anatomical data due to un-perfect registration and other uncertainties.

2 Short description of the basic probabilistic methods

2.1 Maximum Entropy (ME)

ME can be used to assign a probability law to an unknown quantity when we have macroscopic data (expectations) on it. To see this let note by X a quantity of interest and assume that we have L sensors giving us L macroscopic data $\{\mu_l,\ l = 1, \ldots, L\}$, representing the mean values of L known functions $\{\phi_l(X),\ l = 1, \ldots, L\}$ related to the unknown quantity X:

$$\mathrm{E}\left\{\phi_l(X)\right\} = \int \phi_l(x) p(x)\, \mathrm{d}x = \mu_l, \quad l = 1, \ldots, L. \tag{1}$$

The question is then how to represent our partial knowledge of X by a probability law.

Obviously, this problem has not a unique solution. Actually these data define a class of possible solutions and we need a criterion to select one of them. The ME principle can give us this criterion and the problem then becomes:

$$\begin{aligned} &\text{maximize} && S(p) = -\int p(x)\, \ln p(x)\, \mathrm{d}x \\ &\text{subject to} && \int \phi_l(x)\, p(x)\, \mathrm{d}x = \mu_l, \quad l = 1, \ldots, L. \end{aligned}$$

The solution is given by

$$p(x) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left[-\sum_{l=1}^{L} \theta_l \phi_l(x)\right] = \frac{1}{Z(\boldsymbol{\theta})} \exp\left[-\boldsymbol{\theta}^t \boldsymbol{\phi}(x)\right], \tag{2}$$

where

$$Z(\boldsymbol{\theta}) = \int \exp\left[-\sum_{l=1}^{L} \theta_l \phi_l(x)\right] \mathrm{d}x \tag{3}$$

is the partition function and $\{\theta_1, \ldots, \theta_l\}$ are determined by the following system of equations:

$$-\frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_l} = \mu_l, \quad l = 1, \ldots, L, \tag{4}$$

See [5, 6, 7, 8, 9, 10, 11, 12, 13, 14] for more discussions on the foundation of this approach, [15] for optimization and algorithmic, and [16, 17, 18, 19, 20, 21] for some applications.

2.2 Maximum Likelihood (ML)

Let again X be the unknown quantity of interest and assume now that we can directly observe it. Assume also that, through an observation model, we have deduced a parametric form $p(x|\boldsymbol{\theta})$ for the probability law of X where $\boldsymbol{\theta}$ represents the unknown parameters. Finally assume that N samples $\boldsymbol{x} = [x_1, \ldots, x_N]$ of X have been observed (either by one or N sensors) and the question is how to determine the parameters $\boldsymbol{\theta}$. This is a classical parameter estimation problem in statistics.

Two classical methods for solving this problem are:

- Moments Method (MM): The main idea is to write a set of equations (at least L) relating the theoretical and empirical moments:

$$G_l(\boldsymbol{\theta}) = \mathrm{E}\left\{X^l\right\} = \int x^l \, p(x|\boldsymbol{\theta}) \, \mathrm{d}x = \frac{1}{N} \sum_{j=1}^{N} x_j^l, \quad l = 1, \ldots, L \tag{5}$$

  and solve them to obtain the solution.

- Maximum Likelihood (ML): Here, the main idea is to consider the data as N independent samples of X. Then, writing the expression of $p(\boldsymbol{x}|\boldsymbol{\theta})$ and considering it as a function of $\boldsymbol{\theta}$, the ML solution is defined as

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{l(\boldsymbol{\theta}|\boldsymbol{x})\} \quad \text{with} \quad l(\boldsymbol{\theta}|\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{N} p(x_j|\boldsymbol{\theta}) \tag{6}$$

It is interesting to note that, in the case of the generalized exponential families:

$$p(x|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left[-\sum_{l=1}^{L} \theta_l \, \phi_l(x)\right] = \frac{1}{Z(\boldsymbol{\theta})} \exp\left[-\boldsymbol{\theta}^t \, \boldsymbol{\phi}(x)\right] \tag{7}$$

we have

$$l(\boldsymbol{\theta}) = \prod_{j=1}^{N} p(x_j|\boldsymbol{\theta}) = \frac{1}{Z^N(\boldsymbol{\theta})} \exp\left[-\sum_{j=1}^{N} \sum_{l=1}^{L} \theta_l \, \phi_l(x_j)\right] \tag{8}$$

Then, it is easy to see that the ML solution is the solution of the following system of equations:

$$-\frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_l} = \frac{1}{N} \sum_{j=1}^{N} \phi_l(x_j), \quad l = 1, \ldots, L \tag{9}$$

Comparing equations (4) & (9), we can remark an interesting relation between these two methods. The expected values $\mu_l = \mathrm{E}\{\phi_l(x)\}$ in (4) are replaced by their respective empirical values in (9).

See [22, 23, 24] for more discussions on the foundation of this approach, [24, 25] for its relation to maximum entropy, [26] for optimization and algorithmic issues, and [27] for some applications.

2.3 ML and incomplete data: EM Algorithm

Consider the previous problem, but now assume that a sensor gives M values $\boldsymbol{y} = [y_1, \ldots, y_M]$ related to the N samples $\boldsymbol{x} = [x_1, \ldots, x_N]$ of X through a non invertible but exact deterministic relation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with $M < N$. How to determine $\boldsymbol{\theta}$?

This problem is also a classical one in statistics where $\boldsymbol{x}$ is called *complete data* and $\boldsymbol{y}$ *incomplete data*. The solution here is still based on the ML. The only difference is the way to calculate the solution. The main difficulty is that, the incomplete likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$ cannot be computed easily while we know the complete data likelihood expression $p(\boldsymbol{x}|\boldsymbol{\theta})$. A well known algorithm has been developed to handle this problem which is called *Expectation-Maximization*. The main reference to this algorithm is [28] and the following summarizes the main steps to obtain it:
– Write the fundamental relation between $p(\boldsymbol{x}|\boldsymbol{\theta})$ and $p(\boldsymbol{y}|\boldsymbol{\theta})$:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta})\, p(\boldsymbol{y}|\boldsymbol{\theta}), \qquad \forall \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}. \tag{10}$$

or written differently

$$\ln p(\boldsymbol{y}|\boldsymbol{\theta}) = \ln p(\boldsymbol{x}|\boldsymbol{\theta}) - \ln p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}). \tag{11}$$

– Take the conditional expectation of both sides for given value of $\boldsymbol{y}$ and for $\boldsymbol{\theta} = \boldsymbol{\theta}'$:

$$\ln p(\boldsymbol{y}|\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}'}\left\{\ln p(\boldsymbol{x}|\boldsymbol{\theta})\right\} - \mathrm{E}_{\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}'}\left\{\ln p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta})\right\} \tag{12}$$

or written differently as functions of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$:

$$L(\boldsymbol{\theta}) = Q(\boldsymbol{\theta}, \boldsymbol{\theta}') - V(\boldsymbol{\theta}, \boldsymbol{\theta}'). \tag{13}$$

– Note that for a given $\boldsymbol{\theta}'$ and for all $\boldsymbol{\theta}$ we have

$$L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}') = [Q(\boldsymbol{\theta}, \boldsymbol{\theta}') - Q(\boldsymbol{\theta}', \boldsymbol{\theta}')] + [V(\boldsymbol{\theta}, \boldsymbol{\theta}') - V(\boldsymbol{\theta}', \boldsymbol{\theta}')]. \tag{14}$$

– Now, using the Jensen's inequality [28]

$$V(\boldsymbol{\theta}, \boldsymbol{\theta}') \leq V(\boldsymbol{\theta}', \boldsymbol{\theta}') \tag{15}$$

an iterative algorithm, known as *Expectation-Maximization (EM)*, is derived:

$$\left\{ \begin{array}{lll} \text{E:} & Q\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)}\right) & = \mathrm{E}_{\boldsymbol{x}|\boldsymbol{y};\widehat{\boldsymbol{\theta}}^{(k)}}\left\{\ln p(\boldsymbol{x}|\boldsymbol{\theta})\right\} \\ \text{M:} & \widehat{\boldsymbol{\theta}}^{(k+1)} & = \arg\max_{\boldsymbol{\theta}} \left\{ Q\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)}\right)\right\} \end{array} \right. \tag{16}$$

This algorithm insures to converge to a local maximum of the incomplete likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$.

It is interesting to see that in the case of the generalized exponential families (7), the algorithm becomes:

$$\left\{ \begin{array}{lll} \text{E:} & Q(\boldsymbol{\theta}|\boldsymbol{\theta}') & = -N \ln Z(\boldsymbol{\theta}) - \sum_{j=1}^{N} \boldsymbol{\theta}^t \mathrm{E}_{\boldsymbol{x}|\boldsymbol{y}|\boldsymbol{\theta}'}\left\{\boldsymbol{\phi}(x_j)\right\} \\ \text{M:} & -\frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_l} & = \frac{1}{N}\sum_{j=1}^{N} \mathrm{E}_{x_j|\boldsymbol{y}|\boldsymbol{\theta}^{(k)}}\left\{\phi_l(x_j)\right\}, \quad l = 1, \ldots, L \end{array} \right. \tag{17}$$

Compare the M step of this last equation with those of (4) and (9) to see still some relations between ME, ML and the EM algorithms for the particular case of exponential families.

Consider now the same problem where we want to estimate not only $\boldsymbol{\theta}$ but also $\boldsymbol{x}$. We can still use the EM algorithm with the following modification:

$$\left\{\begin{array}{lll} \text{E:} & Q\left(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}\right) & = \mathrm{E}\left\{\ln p(\boldsymbol{x}|\boldsymbol{\theta})|\boldsymbol{y};\widehat{\boldsymbol{\theta}}^{(k)}\right\} \\ & \widehat{\boldsymbol{x}}^{(k)} & = \mathrm{E}\left\{\boldsymbol{x}|\boldsymbol{y};\widehat{\boldsymbol{\theta}}^{(k)}\right\} \\ \text{M:} & \widehat{\boldsymbol{\theta}}^{(k+1)} & = \arg\max_{\boldsymbol{\theta}}\left\{Q\left(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}\right)\right\} \end{array}\right. \tag{18}$$

See [22, 23, 24, 25] for more discussions on the foundation of this approach, [29, 30, 31, 32] for algorithmic and optimization, and [33, 34] for some applications.

2.4 Bayesian Approach

Consider again one of the two last problems, but now assume that the observations $\boldsymbol{y}$ are corrupted by some errors: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ represents any uncertainties in the data, both the observation modeling errors and the measurement errors.

The main tool here is the Bayesian approach where, we use the relation between the data $\boldsymbol{y}$ and the unknowns $\boldsymbol{x}$ and the noise probability law $p_\epsilon(\boldsymbol{\epsilon}|\boldsymbol{\theta}_1)$ to define the likelihood $p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}_1) = p_\epsilon(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}|\boldsymbol{\theta}_1)$ and combine it with the prior $p(\boldsymbol{x}|\boldsymbol{\theta}_2)$ through the Bayes' rule to obtain the posterior law

$$p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \frac{p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2)}{p(\boldsymbol{y}|\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}, \tag{19}$$

where

$$p(\boldsymbol{y}|\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \int p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2)\, \mathrm{d}\boldsymbol{x}. \tag{20}$$

$\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, which are called hyperparameters, are the parameters of the prior laws $p(\boldsymbol{\epsilon}|\boldsymbol{\theta}_1)$ and $p(\boldsymbol{x}|\boldsymbol{\theta}_2)$.

The posterior law $p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ contains all the information available on $\boldsymbol{x}$. We can then use it to make any inference on $\boldsymbol{x}$ [35]. We can for example define the following point estimators:

- Maximum *a posteriori* (MAP):

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}}\left\{p(\boldsymbol{x}|\boldsymbol{y}|\boldsymbol{\theta})\right\} \tag{21}$$

- Posterior Mean (PM):

$$\widehat{\boldsymbol{x}} = \mathrm{E}\left\{\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}\right\} = \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x} \tag{22}$$

- Marginal Posterior Modes (MPM):

$$\widehat{x}_j = \arg\max_{x_i}\left\{p(x_i|\boldsymbol{y}|\boldsymbol{\theta})\right\}, \tag{23}$$

where

$$p(x_i|\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta})\, \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \ldots \mathrm{d}x_{i+1} \ldots \mathrm{d}x_n \tag{24}$$

Here we focused on the estimation of $\boldsymbol{x}$, but in practical applications we need to determine $\boldsymbol{\theta}_2$ or $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ or still to infer on both $(\boldsymbol{\theta}, \boldsymbol{x})$?

The Bayesian approach gives us the necessary tools which are all based on the joint posterior probability law

$$p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{x}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \propto p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \propto p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

which contains all the information about $(\boldsymbol{x}, \boldsymbol{\theta})$ given the data $\boldsymbol{y}$. But, unfortunately, it is often difficult to use this multi variable function directly as it is in practical applications. Very often, we need to summarize it by just one value (the best in some sense) as it was the case when we were only interested on $\boldsymbol{x}$ assuming $\boldsymbol{\theta}$ to be known. The following are some known schemes:

• Joint Maximum *a posteriori* (JMAP):

$$\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{x}}\right) = \arg\max_{(\boldsymbol{\theta}, \boldsymbol{x})} \{p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})\}$$

• Generalized Maximum Likelihood (GML):

$$\begin{cases} \widehat{\boldsymbol{x}}^{(k)} = \arg\max_{\boldsymbol{x}} \left\{p(\boldsymbol{x}|\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})\right\} \\ \widehat{\boldsymbol{\theta}}^{(k)} = \arg\max_{\boldsymbol{\theta}} \left\{p(\widehat{\boldsymbol{x}}^{(k)}|\boldsymbol{y}, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\right\} \end{cases}$$

which can be considered as an iterative alternate optimization algorithm to find the JMAP solution. Note however that, this algorithm may not always converge to JMAP solution.

• Marginalized Maximum Likelihood (MML):

$$\begin{cases} \widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})\, p(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}\right\} \\ \widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \left\{p(\boldsymbol{x}|\boldsymbol{y}; \widehat{\boldsymbol{\theta}})\right\} \end{cases}$$

which separates the estimation of $\boldsymbol{\theta}$ from the estimation of $\boldsymbol{x}$.

• MML-EM:

As we have seen it previously, an analytic expression for $p(\boldsymbol{y}|\boldsymbol{\theta})$ is rarely possible. Consequently, considering $[\boldsymbol{y}, \boldsymbol{x}]$ as the complete data and $\boldsymbol{y}$ as the incomplete data, we can use the EM algorithm to obtain the following scheme:

$$\begin{cases} \text{E:} \quad Q\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)}\right) = \mathrm{E}_{\boldsymbol{x}|\boldsymbol{y}|\widehat{\boldsymbol{\theta}}^{(k)}} \{\ln p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})\} \\ \text{M:} \quad \widehat{\boldsymbol{\theta}}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} \left\{Q\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)}\right)\right\} \end{cases}$$

and finally, we can include the estimation of $\widehat{\boldsymbol{x}}$ in this iterative scheme to obtain

$$\begin{cases} \text{E:} & Q\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)}\right) & = \mathrm{E}\left\{\ln p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{y}; \widehat{\boldsymbol{\theta}}^{(k)}\right\} \\ & \widehat{\boldsymbol{x}}^{(k)} & = \mathrm{E}\left\{\boldsymbol{x}|\boldsymbol{y}; \widehat{\boldsymbol{\theta}}^{(k)}\right\} \\ \text{M:} & \widehat{\boldsymbol{\theta}}^{(k+1)} & = \arg\max_{\boldsymbol{\theta}} \left\{Q\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)}\right)\right\} \end{cases} \tag{25}$$

which, when compared to (18), we see that $\ln p(\boldsymbol{x}|\boldsymbol{\theta})$ in (18) is replaced by $\ln p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})$ in (25) to account for indirect observations $\boldsymbol{y}$.

3 Simple academic data fusion problem examples

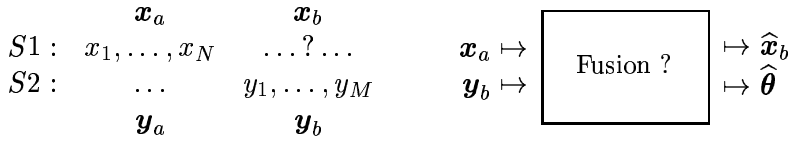
In this section, we give a few academic data fusion problems and will see how the previous tools can be used to give solutions to them. Then, we see that only the Bayesian approach can handle realistic data fusion problems.

3.1 Fusion of macroscopic and microscopic data

Observing an unknown quantity X, L sensors give L macroscopic data $\{\mu_l = \mathrm{E}_{\phi_l(X)}\{,\}\ l = 1,\ldots,L\}$ and another sensor observes X through a linear measurement instrument giving either the exact data $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ or the noisy data $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$. We are asked to infer on $\boldsymbol{x}$. Here, we can first use the ME method with macroscopic data to obtain $p(\boldsymbol{x}|\boldsymbol{\theta})$ and then, depending on the case, either use the ML or the Bayesian approach of the previous sections to infer on $\boldsymbol{x}$ or even to update both estimates $\boldsymbol{\theta}$ and $\boldsymbol{x}$.

3.2 Fusion of a direct and exact data set with an indirect and noisy one

The sensor S1 gives N samples $\boldsymbol{x}_a = \{x_1,\ldots,x_N\}$ of X and stops. The sensor S2 gives M samples $\boldsymbol{y}_b = \{y_1,\ldots,y_M\}$ related to $\boldsymbol{x}$ by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$.
We are asked to predict the unobserved samples $\boldsymbol{x}_b = \{x_{N+1},\ldots,x_{N+M}\}$ of X.



We can propose the following solutions:

- Use $\boldsymbol{x}_a$ to estimate $\boldsymbol{\theta}$, the parameters of $p(\boldsymbol{x}|\boldsymbol{\theta})$ and use it then to estimate $\boldsymbol{x}_b$ from $\boldsymbol{y}_b$:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{L_a(\boldsymbol{\theta}) = \ln p(\boldsymbol{x}_a|\boldsymbol{\theta})\}$$
$$\widehat{\boldsymbol{x}}_b = \arg\max_{\boldsymbol{x}_b} \left\{p\left(\boldsymbol{x}_b|\boldsymbol{y}_b;\widehat{\boldsymbol{\theta}}\right)\right\}$$

$$\boldsymbol{x}_a \mapsto \boxed{\text{ML}} \mapsto \widehat{\boldsymbol{\theta}} \mapsto \boxed{\text{MAP}} \mapsto \widehat{\boldsymbol{x}}_b \quad (\uparrow \boldsymbol{y}_b)$$

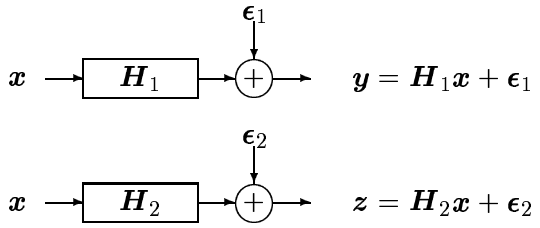
- Use both $\boldsymbol{x}_a$ and $\boldsymbol{y}_b$ to estimate directly $\boldsymbol{x}_b$ and $\boldsymbol{\theta}$ through the joint posterior probability law $p(\boldsymbol{x}_b, \boldsymbol{\theta}|\boldsymbol{x}_a, \boldsymbol{y}_b)$, for example the JMAP:

$$(\widehat{\boldsymbol{x}}_b, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{x}_b, \widehat{\boldsymbol{\theta}})} \{p(\boldsymbol{x}_b, \boldsymbol{\theta}|\boldsymbol{x}_a, \boldsymbol{y}_b)\}$$

  or any other criteria based on it as described in the previous section.

3.3 Fusion of two indirect and noisy but homogeneous data

We have two types of data on the same unknown $\boldsymbol{x}$, both related to it through linear models:



For example, consider an X ray tomography problem where $\boldsymbol{x}$ represents the mass density of the object and where $\boldsymbol{y}$ and $\boldsymbol{z}$ represent respectively a high resolution projection and a low resolution projection.

We can use directly the Bayesian approach to solve this problem:

$$p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{z}) = \frac{p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{x})\,p(\boldsymbol{x})}{p(\boldsymbol{y},\boldsymbol{z})}$$

Actually, the main difficulty here is to assign $p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{x})$ and $p(\boldsymbol{x})$. If we assume that the errors associated to the two sets of data are independent then we have $p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{z}|\boldsymbol{x})$, and if we assume to be able to assign $p(\boldsymbol{y}|\boldsymbol{x})$, $p(\boldsymbol{z}|\boldsymbol{x})$ and $p(\boldsymbol{x})$. then we can compute the posterior $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{z})$. and the calculation can be done more easily. For the purpose of illustration assume the following:

$$\begin{aligned}
\boldsymbol{\epsilon}_1 \simeq \mathcal{N}(\boldsymbol{0},\sigma_1^2\boldsymbol{I}) \longrightarrow p(\boldsymbol{y}|\boldsymbol{x};\sigma_1^2) &\propto \exp\left[-\frac{1}{2\sigma_1^2}|\boldsymbol{y}-\boldsymbol{H}_1\boldsymbol{x}|^2\right] \\
\boldsymbol{\epsilon}_2 \simeq \mathcal{N}(\boldsymbol{0},\sigma_2^2\boldsymbol{I}) \longrightarrow p(\boldsymbol{z}|\boldsymbol{x};\sigma_2^2) &\propto \exp\left[-\frac{1}{2\sigma_2^2}|\boldsymbol{z}-\boldsymbol{H}_2\boldsymbol{x}|^2\right] \\
\boldsymbol{x} \simeq \mathcal{N}(\boldsymbol{m},\Sigma) \longrightarrow p(\boldsymbol{x};\boldsymbol{m},\boldsymbol{\Sigma}) &\propto \exp\left[-\frac{1}{2}[\boldsymbol{x}-\boldsymbol{m}]^t\boldsymbol{\Sigma}^{-1}[\boldsymbol{x}-\boldsymbol{m}]\right]
\end{aligned}$$

Indeed, assume that the hyper-parameters $(\sigma_1^2,\sigma_2^2,\boldsymbol{m},\boldsymbol{\Sigma})$ are given. Then we can use, for example, the MAP estimate, given by:

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}}\{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{z})\} = \arg\min_{\boldsymbol{x}}\{J(\boldsymbol{x}) = J_1(\boldsymbol{x}) + J_2(\boldsymbol{x}) + J_3(\boldsymbol{x})\}$$

with

$$J_1(\boldsymbol{x}) = \frac{1}{2\sigma_1^2}||\boldsymbol{y}-\boldsymbol{H}_1\boldsymbol{x}||^2,$$

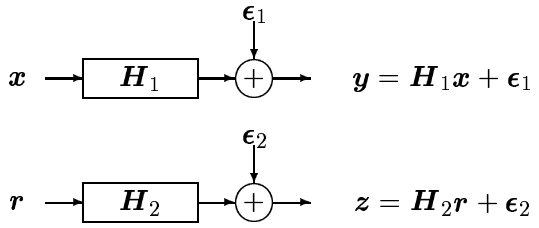
$$J_2(\boldsymbol{x}) = \frac{1}{2\sigma_2^2}||\boldsymbol{z}-\boldsymbol{H}_2\boldsymbol{x}||^2,$$

$$J_3(\boldsymbol{x}) = \frac{1}{2}[\boldsymbol{x}-\boldsymbol{m}]^t\boldsymbol{\Sigma}^{-1}[\boldsymbol{x}-\boldsymbol{m}]$$

However, in practical applications, the data $\boldsymbol{y}$ and $\boldsymbol{z}$ come from different processes and it may not be easy to relate them through two linear systems $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ to the same quantity $\boldsymbol{x}$ as it will be seen in the next section.

4 Real data fusion problems: X rays radiography and ultrasound echography

Consider a more realistic data fusion problem, where we have two different kinds of data. As an example assume a tomographic image reconstruction problem where we have a set of data $\boldsymbol{y}$ obtained by an X ray and a set of data $\boldsymbol{z}$ obtained by an ultrasound probing system (For an introduction to X ray tomography refer to [36]). The X ray data are related to the mass density $\boldsymbol{x}$ of the matter while the ultrasound data are related to the acoustic reflectivity $\boldsymbol{r}$ of the matter which is more related to the changes of material mass density inside the object. Indeed, assume that, we have linear relations, both between $\boldsymbol{y}$ and $\boldsymbol{x}$ and between $\boldsymbol{z}$ and $\boldsymbol{r}$. Then we have:



Assuming that the two sets of data are independent, we can again use the Bayes rule which now becomes

$$p(\boldsymbol{x}, \boldsymbol{r}|\boldsymbol{y}, \boldsymbol{z}) = \frac{p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{r})\, p(\boldsymbol{x}, \boldsymbol{r})}{p(\boldsymbol{y}, \boldsymbol{z})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{z}|\boldsymbol{r})\, p(\boldsymbol{x}, \boldsymbol{r})}{p(\boldsymbol{y}, \boldsymbol{z})}$$

with

$$p(\boldsymbol{y}, \boldsymbol{z}) = \int\int p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{z}|\boldsymbol{r})\, p(\boldsymbol{x}, \boldsymbol{r})\, \mathrm{d}\boldsymbol{r}\, \mathrm{d}\boldsymbol{x}.$$

Here also the main difficulty is the assignment of the probability laws $p(\boldsymbol{y}|\boldsymbol{x})$, $p(\boldsymbol{z}|\boldsymbol{r})$, and more specifically $p(\boldsymbol{x}, \boldsymbol{r})$.

Actually if we could find a mathematical relation between $\boldsymbol{r}$ and $\boldsymbol{x}$, then the problem would become the same as in the previous case. To see this, assume that we can find a relation such as $r_j = g(x_{j+1} - x_j)$ with g a monotonic increasing function. For example, assume that the body under the test is composed of horizontal layers of homogeneous materials. Then, x_j are proportional to the mass density of the material in layer j and r_j is related to the difference of material densities in layer j and $j + 1$. This relation between x_j and $(x_{j+1} - x_j)$ is not in general a linear relation and is too complicated to describe. However, just to see the relation with the previous example assume g to be a linear function then we have

$$\left\{ \begin{array}{l} \boldsymbol{y} = \boldsymbol{H}_1\boldsymbol{x} + \boldsymbol{\epsilon}_1 \\ \boldsymbol{z} = \boldsymbol{H}_2\boldsymbol{r} + \boldsymbol{\epsilon}_2 \\ \boldsymbol{r} = \boldsymbol{G}\boldsymbol{x} \end{array} \right. \longrightarrow \left\{ \begin{array}{l} \boldsymbol{y} = \boldsymbol{H}_1\boldsymbol{x} + \boldsymbol{\epsilon}_1 \\ \boldsymbol{z} = \boldsymbol{G}\boldsymbol{H}_2\boldsymbol{r} + \boldsymbol{\epsilon}_2 \end{array} \right.$$

which becomes equivalent of the previous example. But, as we mentioned, this is an unrealistic hypothesis.

For more realistic cases we need a method which does not use a physically based explicit expression of g. One approach proposed and used by the author [37] and by other collaborators *Gautier et al.* [38, 39, 40, 41] is based on a compound Markovian model where the body object $\boldsymbol{o}$ is assumed to be composed of three related quantities:

$$\boldsymbol{o} = \{\boldsymbol{r}, \boldsymbol{x}\} = \{\boldsymbol{q}, \boldsymbol{a}, \boldsymbol{x}\}$$

where $\boldsymbol{q}$ is a binary vector representing the positions of the discontinuities (edges) in the body, $\boldsymbol{a}$ a vector containing the reflectivity values such that

$$\left\{\begin{array}{l} q_j = 0 \longrightarrow r_j = 0, \\ q_j = 1 \longrightarrow r_j = a_j \end{array}\right. \quad \text{and} \quad r_j = \left\{\begin{array}{ll} g(x_{j+1} - x_j) & \text{if } |x_{j+1} - x_j| > \alpha \\ 0 & \text{otherwise} \end{array}\right.$$

where g can be any monotonic increasing function which is not needed to be known.

With this model we can write

$$p(\boldsymbol{o}, \boldsymbol{r}) = p(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q}) = p(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{q})\, p(\boldsymbol{a}|\boldsymbol{q})\, p(\boldsymbol{q})$$

and using the Bayes rule, we have

$$p(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q}|\boldsymbol{y}, \boldsymbol{z}) \propto p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q})\, p(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q}) = p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q})\, p(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{q})\, p(\boldsymbol{a}|\boldsymbol{q})\, p(\boldsymbol{q})$$

We illustrate this approach by making the following assumptions:

- Conditional independence of $\boldsymbol{y}$ and $\boldsymbol{z}$:

$$p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q}) = p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{z}|\boldsymbol{a}, \boldsymbol{q}) = p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{z}|\boldsymbol{r}).$$

- Gaussian laws for $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ which results to:

$$p(\boldsymbol{y}|\boldsymbol{x}; \sigma_1^2) \propto \exp\left[-\frac{1}{2\sigma_1^2}||\boldsymbol{y} - \boldsymbol{H}_1\boldsymbol{x}||^2\right]; \quad p(\boldsymbol{z}|\boldsymbol{r}; \sigma_2^2) \propto \exp\left[-\frac{1}{2\sigma_2^2}||\boldsymbol{z} - \boldsymbol{H}_2\boldsymbol{r}||^2\right]$$

- Bernoulli law for $\boldsymbol{q}$: $\quad p(\boldsymbol{q}) \propto \sum_{i=1}^{n} q_i^{\lambda}(1 - q_i)^{1-\lambda}$

- Gaussian law for $\boldsymbol{r}|\boldsymbol{q}$ or equivalently for $\boldsymbol{a}|\boldsymbol{q}$:

$$p(\boldsymbol{a}|\boldsymbol{q}) \propto \exp\left[-\frac{1}{2\sigma_a^2}\boldsymbol{a}^t\boldsymbol{Q}\boldsymbol{a}\right], \quad \boldsymbol{Q} = \text{diag}[q_1, \ldots, q_n]$$

- Markovian model for $\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{q}$: $\quad p(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{q}) = p(\boldsymbol{x}|\boldsymbol{q}) \propto \exp\left[-U(\boldsymbol{x}|\boldsymbol{q})\right]$

Then, based on

$$p(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q}|\boldsymbol{y}, \boldsymbol{z}) \propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{z}|\boldsymbol{a})\, p(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{q})\, p(\boldsymbol{a}|\boldsymbol{q})\, p(\boldsymbol{q})$$

we can propose the following schemes to estimate either $\boldsymbol{x}$ or both $(\boldsymbol{x}, \boldsymbol{r})$ or equivalently $(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q})$:

- Simultaneous estimation of all the unknowns with the joint MAP estimation (JMAP):

$$(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{a}}, \widehat{\boldsymbol{q}}) = \arg\max_{(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q})} \{p(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{q}|\boldsymbol{y}, \boldsymbol{z})\} \qquad \begin{array}{l} \boldsymbol{y} \mapsto \\ \boldsymbol{z} \mapsto \end{array} \boxed{\text{JMAP}} \begin{array}{l} \mapsto \widehat{\boldsymbol{x}} \\ \mapsto \widehat{\boldsymbol{a}} \\ \mapsto \widehat{\boldsymbol{q}} \end{array}$$

- First estimate the positions of the discontinuities $\boldsymbol{q}$ and then use them to estimate $\boldsymbol{x}$ and $\boldsymbol{a}$:

$$\left\{\begin{array}{l} \widehat{\boldsymbol{q}} = \arg\max_{\boldsymbol{q}} \{p(\boldsymbol{q}|\boldsymbol{y}, \boldsymbol{z})\} \\ (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{a}}) = \arg\max_{(\boldsymbol{x}, \boldsymbol{a})} \{p(\boldsymbol{x}, \boldsymbol{a}|\boldsymbol{y}, \boldsymbol{z}, \widehat{\boldsymbol{q}})\} \end{array}\right. \qquad \begin{array}{l} \boldsymbol{y} \mapsto \\ \boldsymbol{z} \mapsto \end{array} \boxed{\text{Det.}} \mapsto \begin{array}{l} \boldsymbol{y} \mapsto \\ \widehat{\boldsymbol{q}} \mapsto \\ \boldsymbol{z} \mapsto \end{array} \boxed{\text{Est.}} \begin{array}{l} \mapsto \widehat{\boldsymbol{x}} \\ \mapsto \widehat{\boldsymbol{a}} \end{array}$$

- First estimate the positions of the discontinuities $\boldsymbol{q}$ using only $\boldsymbol{z}$ and then use them to estimate $\boldsymbol{x}$ and $\boldsymbol{a}$:

$$\begin{cases} \widehat{\boldsymbol{q}} = \arg\max_{\boldsymbol{q}} \{p(\boldsymbol{q}|\boldsymbol{z})\} \\ (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{a}}) = \arg\max_{(\boldsymbol{x},\boldsymbol{a})} \{p(\boldsymbol{x}, \boldsymbol{a}|\boldsymbol{y}, \boldsymbol{z}, \widehat{\boldsymbol{q}})\} \end{cases} \qquad \boldsymbol{z} \mapsto \boxed{\text{Det.}} \mapsto \begin{array}{l} \boldsymbol{y} \mapsto \\ \widehat{\boldsymbol{q}} \mapsto \\ \boldsymbol{z} \mapsto \end{array} \boxed{\text{Est.}} \begin{array}{l} \mapsto \widehat{\boldsymbol{x}} \\ \mapsto \widehat{\boldsymbol{a}} \end{array}$$

- First estimate $\boldsymbol{q}$ and $\boldsymbol{a}$ using only $\boldsymbol{z}$ and then use them to estimate $\boldsymbol{x}$:

$$\begin{cases} (\widehat{\boldsymbol{q}}, \widehat{\boldsymbol{a}}) = \arg\max_{\boldsymbol{q},\boldsymbol{a}} \{p(\boldsymbol{q}, \boldsymbol{a}|\boldsymbol{z})\} \\ \widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{a}}, \widehat{\boldsymbol{q}})\} \end{cases} \qquad \boldsymbol{z} \mapsto \boxed{\begin{array}{c} \text{Det.} \\ \& \\ \text{Est.} \end{array}} \begin{array}{l} \mapsto \widehat{\boldsymbol{q}} \mapsto \\ \mapsto \widehat{\boldsymbol{a}} \mapsto \\ \boldsymbol{y} \mapsto \end{array} \boxed{\text{Est.}} \mapsto \widehat{\boldsymbol{x}}$$

- First estimate only $\boldsymbol{q}$ using $\boldsymbol{z}$, then estimate $\boldsymbol{a}$ using $\widehat{\boldsymbol{q}}$ and $\boldsymbol{z}$, and finally, estimate $\boldsymbol{x}$ using $\widehat{\boldsymbol{q}}$, $\widehat{\boldsymbol{a}}$ and $\boldsymbol{y}$:

$$\begin{cases} \widehat{\boldsymbol{q}} = \arg\max_{\boldsymbol{q}} \{p(\boldsymbol{q}|\boldsymbol{z})\} \\ \widehat{\boldsymbol{a}} = \arg\max_{\boldsymbol{a}} \{p(\boldsymbol{a}|\boldsymbol{z}, \widehat{\boldsymbol{q}})\} \\ \widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{a}}, \widehat{\boldsymbol{q}})\} \end{cases}$$

$$\boldsymbol{z} \mapsto \boxed{\text{Det.}} \mapsto \widehat{\boldsymbol{q}} \qquad \begin{array}{l} \widehat{\boldsymbol{q}} \mapsto \\ \boldsymbol{z} \mapsto \end{array} \boxed{\text{Est.}} \mapsto \widehat{\boldsymbol{a}} \qquad \begin{array}{l} \widehat{\boldsymbol{q}} \mapsto \\ \widehat{\boldsymbol{a}} \mapsto \\ \boldsymbol{y} \mapsto \end{array} \boxed{\text{Est.}} \mapsto \widehat{\boldsymbol{x}}$$

- First estimate only $\boldsymbol{q}$ using $\boldsymbol{z}$ and then estimate $\boldsymbol{x}$ using $\widehat{\boldsymbol{q}}$ and the data $\boldsymbol{y}$:

$$\begin{cases} \widehat{\boldsymbol{q}} = \arg\max_{\boldsymbol{q}} \{p(\boldsymbol{q}|\boldsymbol{z})\} \\ \widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{q}})\} \end{cases} \qquad \boldsymbol{z} \mapsto \boxed{\text{Det.}} \mapsto \widehat{\boldsymbol{q}} \qquad \begin{array}{l} \widehat{\boldsymbol{q}} \mapsto \\ \boldsymbol{y} \mapsto \end{array} \boxed{\text{Est.}} \mapsto \widehat{\boldsymbol{x}}$$

In all these schemes, the detection steps (estimation of $\boldsymbol{q}$) is very difficult and computationally demanding due to the need of marginalization and a combinatorial optimization algorithm. One way to take over this difficulty is to model the object only by $(\boldsymbol{r}, \boldsymbol{x})$ without decomposing $\boldsymbol{r}$ in $(\boldsymbol{q}, \boldsymbol{a})$ anymore. However, we must catch the information on $\boldsymbol{r}$ which is almost always equal to zero (in homogeneous regions) and can take any real value in the borders of these regions. This can be down through a better choice for $p(\boldsymbol{r})$. For example, a generalized Gaussian law for $p(\boldsymbol{r})$:

$$p(\boldsymbol{r}) \propto \exp\left[-\alpha \sum_j |r_j|^\beta\right], \quad \text{with} \;\; 1 \leq \beta \leq 2. \tag{26}$$

(in place of Gaussian law which is the special case for $\beta = 2$), can be used to translate this concentration around zero while giving the possibility to have large values, thanks to long-tailed character of this distribution for the values of β near to one.

Based on this remark, a more realistic solutions has been proposed in previous works [42, 38, 40, 37, 43] which is described briefly:

Proposed method:
Use the ultrasound data $\boldsymbol{z}$ to detect the locations of some of the boundaries and use X ray data to make an intensity image preserving the positions of these discontinuities:

$$\boldsymbol{z} \mapsto \boxed{\text{Est.}} \mapsto \widehat{\boldsymbol{r}} \mapsto \boxed{q_j = \frac{|r_j|}{\sum_j |r_j|}} \mapsto \widehat{\boldsymbol{q}} \qquad \begin{matrix} \widehat{\boldsymbol{q}} \mapsto \\ \boldsymbol{y} \mapsto \end{matrix} \boxed{\text{Est.}} \mapsto \widehat{\boldsymbol{x}}$$

For the first part, with the assumptions made, we have

$$\widehat{\boldsymbol{r}} = \arg\max_{\boldsymbol{r}} \{p(\boldsymbol{r}|\boldsymbol{z})\} = \arg\min_{\boldsymbol{r}} \{J_1(\boldsymbol{r}|\boldsymbol{z})\}$$

with

$$J_1(\boldsymbol{r}|\boldsymbol{z}) = ||\boldsymbol{z} - \boldsymbol{H}_2\boldsymbol{r}||^2 + \lambda \sum_j |r_j|^\beta.$$

When $\boldsymbol{r}$ estimated, we can determine either a binary valued $\boldsymbol{q}$ from it by $q_j = 1$, if $|r_j| < s$ and $q_j = 0$ elsewhere, or a real valued $\boldsymbol{q}$ with $q_j \in [0, 1]$ by $q_j = |r_j| / \max(|r_j|)$ or any other practical solution. This real valued $\boldsymbol{q}$ can be considered as a blurred (or fuzzy) image of borders and edges.

For the second part which is the estimation of $\boldsymbol{x}$ given $\boldsymbol{y}$ and $\widehat{\boldsymbol{q}}$ the following criterion has been used:

$$\begin{aligned} \widehat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{q}})\} = \arg\min_{\boldsymbol{x}} \{J_2(\boldsymbol{x}|\boldsymbol{y}; \widehat{\boldsymbol{q}})\} \\ \text{with} \qquad J_2(\boldsymbol{x}|\boldsymbol{y}, \widehat{\boldsymbol{q}}) &= ||\boldsymbol{y} - \boldsymbol{H}_1\boldsymbol{x}||^2 + \lambda_2 \sum_j (1 - q_j)|x_{j+1} - x_j|^\beta, \quad 1 \geq \beta \leq 2. \end{aligned}$$

The aim of this paper is not to go through more details on these methods. The interested reader should refer to [41, 43].

5 Fusion of geometric information and radiographic data

Here, we illustrate an application of the proposed methods for the special case of CT medical imaging where we want to include some geometric information such as partial knowledge of borders of different regions and/or partial knowledge of materials in specified regions of the body in the reconstruction method. The idea of using anatomical information in computed tomography is not new. Many works on the subject has been done before. See for example [1, 41, 43, 2]. Using some partial knowledge of some regions borders can be considered as a special case of the previous example. Actually, in the second step of the proposed method in previous section, we were using the combination of the radiographic data $\boldsymbol{y}$ and $\boldsymbol{q}$ which can be considered as a geometrical (regions borders) data. It has also been used in [1] in medical imaging. Using some partial knowledge about the exact values of pixels in some specified regions has also been used recently [2]. But, combining both region and border informations from anatomic data is new.

In the following, we assume to have sinogram data $\boldsymbol{y}$ and a binary map $\boldsymbol{q}$ containing the borders of some of the regions in the body and $\boldsymbol{s}$ an image containing the attenuation constant values of some of the regions in the body (not forcibly the same regions for which we know the borders) and $\boldsymbol{\mu}$ an image indicating our degree of confidence about the knowledge of values in those regions (0 when no knowledge and 1 when high confidence). Based on the MAP Bayesian approach and the discussions in previous sections, we propose the following criterion to optimize to find the an image $\widehat{\boldsymbol{x}}$ which will be the result of fusion of these data:

$$J(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{s}, \boldsymbol{\mu}) = ||\boldsymbol{y} - \boldsymbol{H}_1\boldsymbol{x}||^2 + \lambda_1 \sum_j (1 - q_j)|x_{j+1} - x_j|^{\beta_1} + \lambda_2 \sum_j \mu_j |x_j - s_j|^{\beta_2}$$

Note that, when the hyperparameters $\lambda_1, \lambda_2 > 0$ and $1 \leq \beta_1, \beta_2 \leq 2$ and the data $\boldsymbol{y}$ and $\boldsymbol{q}, \boldsymbol{s}$ and $\boldsymbol{\mu}$ are given, this criterion is a convex function of $\boldsymbol{x}$. Then, its optimization can be done by any gradient based algorithm. In the following, we show the results obtained by this criterion for the following situations:
– when we have only the data $\boldsymbol{y}$ and no anatomical data ($\boldsymbol{q} = \boldsymbol{\mu} = \boldsymbol{0}$);
– when we have the data $\boldsymbol{y}$ and the map of borders $\boldsymbol{q}$ but no region data ($\boldsymbol{\mu} = \boldsymbol{0}$);
– when we have the data $\boldsymbol{y}$ and the map of regions (characterized by both $\boldsymbol{s}$ and $\boldsymbol{\mu}$) but no other borders data ($\boldsymbol{q} = \boldsymbol{0}$);
– when we have the data $\boldsymbol{y}$ and both the borders $\boldsymbol{q}$ and regions maps ($\boldsymbol{s}, \boldsymbol{\mu}$).

In the following, the hyperparameters β_1 and β_2 are fixed either to 2 or to 1.1 and the two regularization parameters λ_1 and λ_2 are adjusted empirically.

The object is a known numerical phantom in CT which has been proposed by Shepp and Logan [44, 34, 45]. This is a (256×256) image. The sinogram data is obtained by simulating a fan beam tomography with 64 detectors and 128 angular positions over 0 and 360 degrees for the source. The opening angle for the source is 30.4 degrees. The distance between the source and the center of the object is 600 mm and the dimensions of the reconstructed image is (400 mm $\times$ 400 mm).

Figure 1 shows this geometric configuration, the original object and the associated sinogram data.

Figure 2 shows the reconstruction results by classical back-projection or filtered back-projection methods used in commercial scanners. As it is seen on this figure, these results are not satisfactory for the data gathering configuration we proposed where we are looking for a high resolution image (256×256) from a sinogram data which has only (128×64) data points (64 detectors and 128 source positions uniformly distributed in 0, 2π). We also give here two other results obtained by optimizing the criterion

$$J(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{s}) = ||\boldsymbol{y} - \boldsymbol{H}_1\boldsymbol{x}||^2 + \lambda_1 \sum_j (x_{j+1} - x_j)^2$$

once over $\mathbf{R}^n$ and the second over $\mathbb{R}^n_+$. In both cases, we used a simple gradient algorithm, but in the second case, we imposed the positivity constraint at each iteration. These results are significantly better than the classical back-projection methods thanks to regularization terms, but they need more computations (approximately two times more computations than a simple back-projection in each iteration).

Figure 3 shows the reconstruction results when we assume to know the values of image in three regions (those in white color in the left figure) of the object. This result is obtained by optimizing the following criterion.

$$J(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{s}) = ||\boldsymbol{y} - \boldsymbol{H}_1\boldsymbol{x}||^2 + \lambda_1 \sum_j (x_{j+1} - x_j)^2 + \lambda_2 \sum_j \mu_j (x_j - s_j)^2$$

Here $\mu_j = 0$ everywhere excepted the pixels in those three regions.

Figure 4 shows the reconstruction results when we assume to know the borders of some of the regions in the image (those given in the left figure). This result is obtained by optimizing the following criterion.

$$J(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{s}) = ||\boldsymbol{y} - \boldsymbol{H}_1\boldsymbol{x}||^2 + \lambda_1 \sum_j (1 - q_j)(x_{j+1} - x_j)^2$$

Here $q_j = 0$ everywhere excepted in the pixels in borders of those three regions where $q_j = 1$.

Figure 5 shows the reconstruction results when we assume to know both the borders of some of the regions and the values of some others (the same as those used in Figures 4 and 5). This result is obtained by optimizing the following criterion.

$$J(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{q},\boldsymbol{s}) = ||\boldsymbol{y} - \boldsymbol{H}_1\boldsymbol{x}||^2 + \lambda_1 \sum_j (1-q_j)(x_{j+1}-x_j)^2 + \lambda_2 \sum_j \mu_j(x_j - s_j)$$

Obviously, fusion of more geometrical information results in more accurate results when the geometrical results are exact. Unfortunately, in practical applications, we need a first step of registration to bring the geometrical informations in the same frames of X ray radiographic data. In previous simulations, we assumed that this has been done, before starting the reconstruction.

In the following, we give some results to show the sensitivity of the results on some errors of registration. Here we simulated the cases where the geometrical atlas data are obtained with some errors on the orientation of some of the known regions (5 degree).

Figure 6 shows the the errors in regions and those in borders. and the figures 7, 8 and 9 show the results obtained with these errors in the geometrical data with the same conditions which are obtained the results of the figures 3, 4 and 5. Comparing these results, we see that the degradations due to these errors are not so crucial if the regularization parameters λ_1 and λ_2 are not too high.

There is however the possibilities to reduce these effects, by feedback iterating, i.e. by reestimating these geometrical data from the final reconstructions of the previous results and restarting the data fusion again from these new informations. But, nothing can be given on the convergence of such an iterative procedure.

6 Conclusions

In this paper, in the first and second parts, we presented the basics of three main probabilistic approaches which can be used for data fusions and gave a few simple examples of data fusion problems and show the way we can use these tools to propose solutions for those problems. The following points are the main conclusions of this part:

- ME can be used when we want to assign a probability law $p(\boldsymbol{x})$ to an unknown quantity X from the knowledge of some macroscopic data (expected values).
- ML can be used when we have a parametric form of the probability law $p(\boldsymbol{x}|\boldsymbol{\theta})$ and we have access to exact direct observations $\boldsymbol{x}$ of X, and we want to estimate the parameters $\boldsymbol{\theta}$.
- ML-EM extends the ML to the case of incomplete but exact observations.
- When the observed data are uncertain (noisy), the Bayesian approach is the most appropriate to update the prior $p(\boldsymbol{x}|\boldsymbol{\theta})$ to the posterior $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ using the likelihood $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$.
- For practical data fusion problems the Bayesian approach seems to give all the necessary tools we need.

In the third part, we illustrated the use of the Bayesian approach for a real data fusion of X ray radiographic data and ultrasound echo-graphic data in a CT image reconstruction

problem in a NDT application. Through this example, we showed that, even the Bayesian approach is coherent and easy to understand and conceptually easy to use, in real applications, we have still much to do to implement it. The following are the main difficult steps of the approach which need careful attentions: assignment or choice of the prior laws; efficient optimization of the obtained criteria; and the estimation of the hyper-parameters. We also showed that compound Markov models are convenient models to represent signals and images in a Bayesian approach of data fusion when one set of data gives informations on the borders and the other set of data gives information on the region volumes. We illustrated the the effects of prior modeling on the complexity of the obtained criteria and proposed a practically feasible method for this data fusion problem.

Finally, we illustrated the feasibility of such Bayesian estimation approach for a medical computed tomography where we used some anatomical information to obtain better reconstruction results. We used two kind of anatomical information: partial knowledge of values in some regions and partial knowledge of the borders of some other regions. We showed the advantages of using such informations on increasing the quality of reconstructions. We also showed some results to analyze the effects of some errors in anatomic data on the reconstructed results.

References

[1] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Medical Imaging*, vol. MI-12, no. 4, pp. 670–680, 1993.

[2] M. Fiani, J. Idier, and S. Gautier, "Algorithmes ART semi-quadratiques pour la reconstruction à partir de radiographies," in *Actes du 18^{e} colloque GRETSI*, (Toulouse), sep. 2001.

[3] J. Boyd and J. Little, "Complementary data fusion for limited-angle tomography," in *Proceeding of Computer Vision and Pattern Recognition 1994*, (Seattle, WA, USA), pp. 288–294, 1994.

[4] J. Boyd, "Limited-angle computed tomography for sandwich structures using data fusion," *Journal of Nondestructive Evaluation*, vol. 14, no. 2, pp. 61–76, 1995.

[5] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.

[6] L. A. Verdugo and P. N. Rathie, "On the entropy of continuous probability distributions," *IEEE Trans. Inf. Theory*, vol. 24, pp. 120–122, jan. 1978.

[7] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theory*, vol. 26, pp. 26–37, jan. 1980.

[8] J. M. Van Campenhout and T. M. Cover, "Maximum entropy and conditional probability," *IEEE Trans. Inf. Theory*, vol. 27, pp. 483–489, juil. 1981.

[9] E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proc. IEEE*, vol. 70, pp. 939–952, sep. 1982.

[10] R. Balian, *Du microscopique au macroscopique, cours de physique statistique de l'École Polytechnique*, vol. 1. Paris: Ellipses, 1982.

[11] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*. New York: Springer-Verlag, 1985.

[12] J. M. Borwein and A. S. Lewis, "Duality relationships for entropy-like minimization problems," *SIAM J. Control and Optimization*, vol. 29, pp. 325–338, mars 1991.

[13] A. Mohammad-Djafari, *Maximum Entropy and Linear Inverse Problems; A Short Review*, pp. 253–264. Paris: Kluwer Academic Publ., ali mohammad–djafari and guy demoment ed., 1992.

[14] A. Mohammad-Djafari, "Maximum d'entropie et problèmes inverses en imagerie," *Traitement du Signal*, pp. 87–116, 1994.

[15] S. Erlander, "Entropy in linear programs," *Mathematical Programming*, vol. 21, pp. 137–151, 1981.

[16] J. Skilling and R. K. Bryan, "Maximum entropy image reconstruction: General algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 211, pp. 111–124, 1984.

[17] J. Navaza, "On the maximum–entropy estimate of the electron density function," *Acta Crystallographica*, vol. A–41, pp. 232–244, 1985.

[18] R. Narayan and R. Nityananda, "Maximum entropy image restoration in astronomy," *Ann. Rev. Astron. Astrophys.*, vol. 24, pp. 127–170, 1986.

[19] L. K. Jones and C. L. Byrne, "General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis," *IEEE Trans. Inf. Theory*, vol. 36, pp. 23–30, jan. 1990.

[20] A. Decarreau, D. Hilhorst, C. Lemaréchal, and J. Navaza, "Dual methods in entropy maximization. Application to some problems in crystallography," *SIAM J. Optimization*, vol. 2, pp. 173–197, mai 1992.

[21] N. J. Dusaussoy and I. E. Abdou, "The extended ment algorithm: A maximum entropy type algorithm using prior knowledge for computerized tomography," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1164–1180, 1991.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[23] B. Efron, "The geometry of exponential families," *The Annals of Statistics*, vol. 6, no. 2, pp. 362–376, 1978.

[24] Frieden, "Dice, entropy and likelihood," *J. Opt. Soc. Amer.*, vol. 73, no. 12, pp. 1764–1770, 1985.

[25] A. Mohammad-Djafari and J. Idier, *Maximum Likelihood Estimation of the Lagrange Parameters of the Maximum Entropy Distributions*, pp. 131–140. Seattle, WA: Kluwer Academic Publ., C.R. Smith, G.J. Erikson and P.O. Neudorfer ed., 1991.

[26] C. Y. Chi, J. M. Mendel, and D. Hampson, "A computationally fast approach to maximum-likelihood deconvolution," *Geophysics*, vol. 49, pp. 550–565, 1984.

[27] B. R. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.*, vol. 62, pp. 511–518, avr. 1972.

[28] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and toeplitz constrained covariances," *Proc. IEEE*, vol. 75, pp. 892–906, juil. 1987.

[29] P. J. Green, "On the use of the EM algorithm for penalized likelihood estimation," *J. R. Statist. Soc. B*, vol. 52, pp. 443–452, 1990.

[30] R. J. Hathaway, "A constrained formulation of maximum-likelihood estimation for normal mixture distributions," *The Annals of Statistics*, vol. 13, pp. 795–800, 1985.

[31] B. G. Leroux and M. L. Puterman, "Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models," *Biometrics*, vol. 48, pp. 545–558, juin 1992.

[32] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, pp. 195–239, avr. 1984.

[33] Holmes, "Blind deconvolution of quantum-limited incoherent imagery: Maximum-likelihood approach," *J. Opt. Soc. Amer.*, vol. 9, juil. 1993.

[34] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Medical Imaging*, vol. MI-1, pp. 113–122, 1982.

[35] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Medical Imaging*, vol. 8, pp. 194–202, juin 1989.

[36] A. Mohammad-Djafari and J.-M. Dinten, *Reconstruction tomographique partir d'un nombre faible de projections*, ch. 12, pp. 297–320. Paris: Herms, 2001.

[37] A. Mohammad-Djafari, "Probabilistic methods for data fusion," in *Maximum Entropy and Bayesian Methods*, (Boise, ID), aot 1997.

[38] S. Gautier, G. Le Besnerais, A. Mohammad-Djafari, and B. Lavayssière, *Data fusion in the field of non destructive testing.* Maximum Entropy and Bayesian Methods, Santa Fe, NM: Kluwer Academic Publ., K. Hanson ed., 1995.

[39] S. Gautier, G. Le Besnerais, A. Mohammad-Djafari, and B. Lavayssière, "Fusion de données radiographiques et ultrasonores, en vue d'une applicaion en contrôle non destructif," (Clermont-Ferrand), Second international workshop on inverse problems in electromagnetism and acoustic, mai 1995.

[40] S. Gautier, *Fusion de données gammagraphiques et ultrasonores. Application au contrôle non destructif.* Thèse de doctorat, Université de Paris-Sud, Orsay, dc. 1996.

[41] S. Gautier, J. Idier, A. Mohammad-Djafari, and B. Lavayssière, "Traitement d'échogrammes ultrasonores par déconvolution aveugle," in *Actes du 16e colloque GRETSI*, (Grenoble), pp. 1431–1434, sep. 1997.

[42] S. Gautier, G. Le Besnerais, A. Mohammad-Djafari, and B. Lavayssière, "Vers la fusion de données gammagraphiques et ultrasonores," in *Actes du 15^{e} colloque GRETSI*, (Juan-les-Pins), pp. 869–872, sep. 1995.

[43] S. Gautier, J. Idier, A. Mohammad-Djafari, and B. Lavayssière, "X-ray and ultrasound data fusion," in *Proc. IEEE ICIP*, (Chicago, IL), pp. 366–369, oct. 1998.

[44] B. F. Logan and L. A. Shepp, "Optimal reconstruction of a function from its projections," *Duke Math. J.*, vol. 42, pp. 645–659, 1975.

[45] G. T. Herman, *Image reconstruction from projections. The fundamentals of computerized tomography.* New York: Academic Press, 1980.

List of Figures

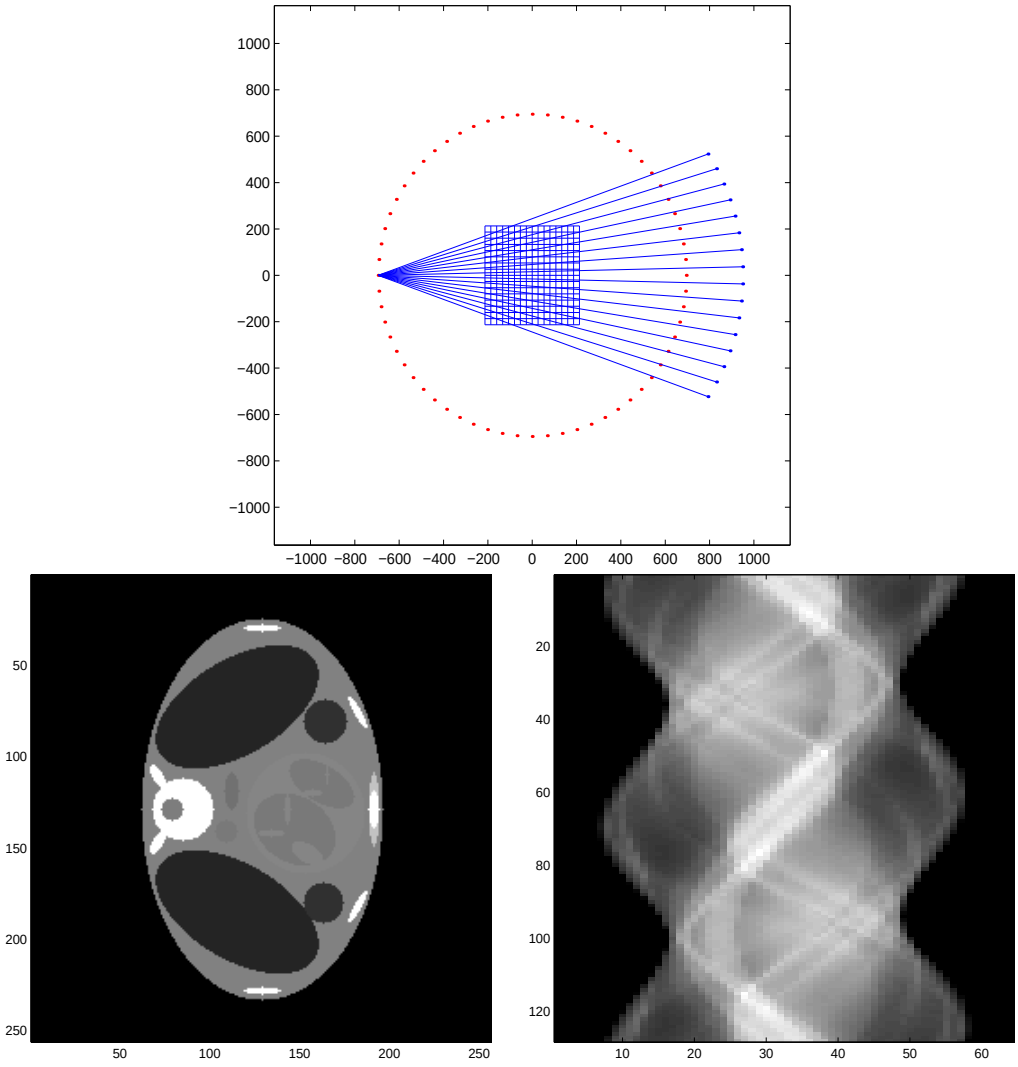


Figure 1: Geometrical configuration (top), original object (left) and its sinogram data (right). The geometry is a fan beam CT, the object is (256×256) and the sinogram data is (128×64).

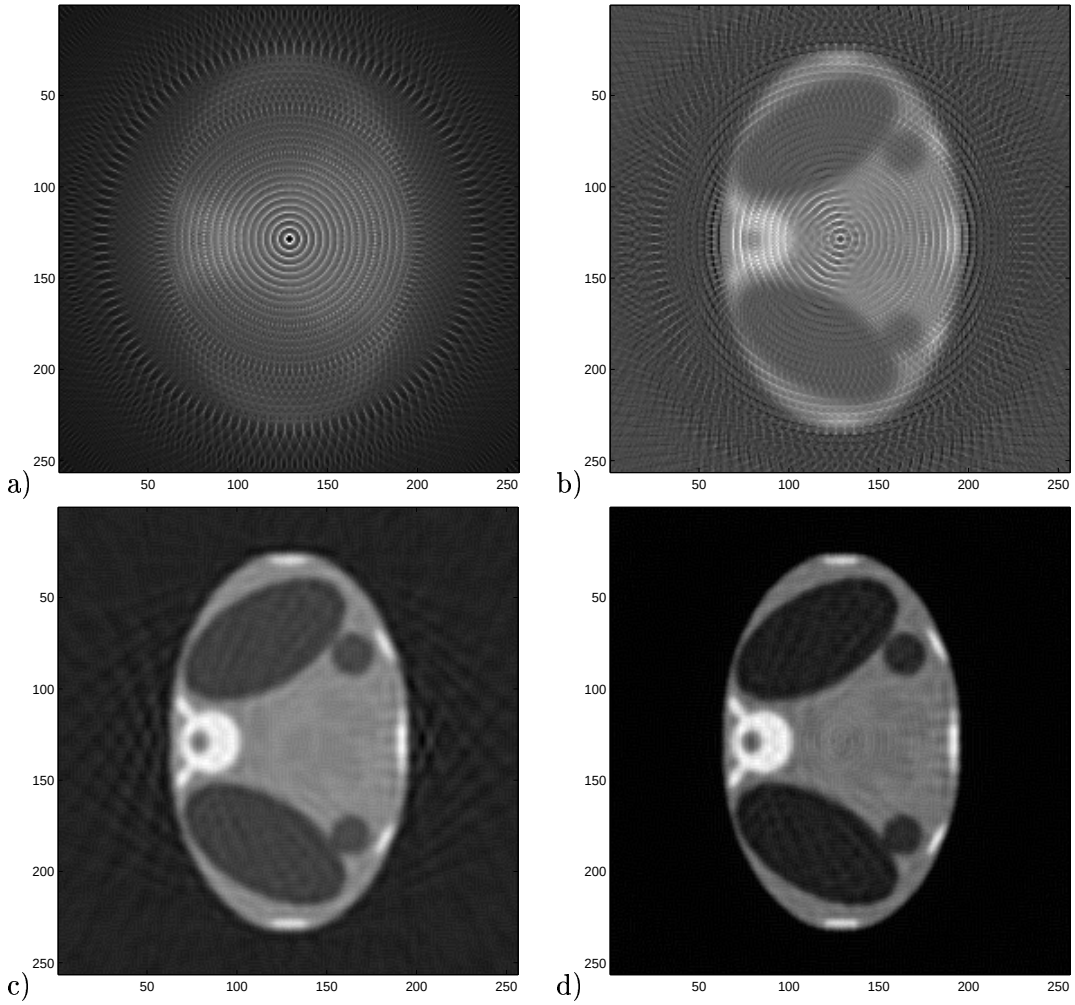


Figure 2: Reconstructions by classical methods: a) back-projection b) filtered back-projection c) quadratic regularization and d) quadratic regularization with positivity constraint.

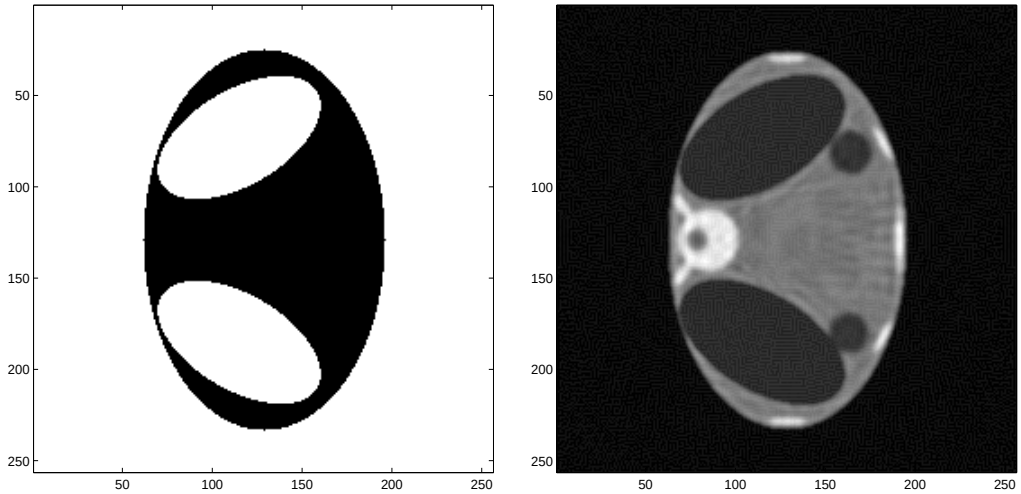


Figure 3: Reconstructions by data fusion (incomplete known regions): The result in the right is obtained by assuming to know the exact values of densities in the white regions of the figure in the left.

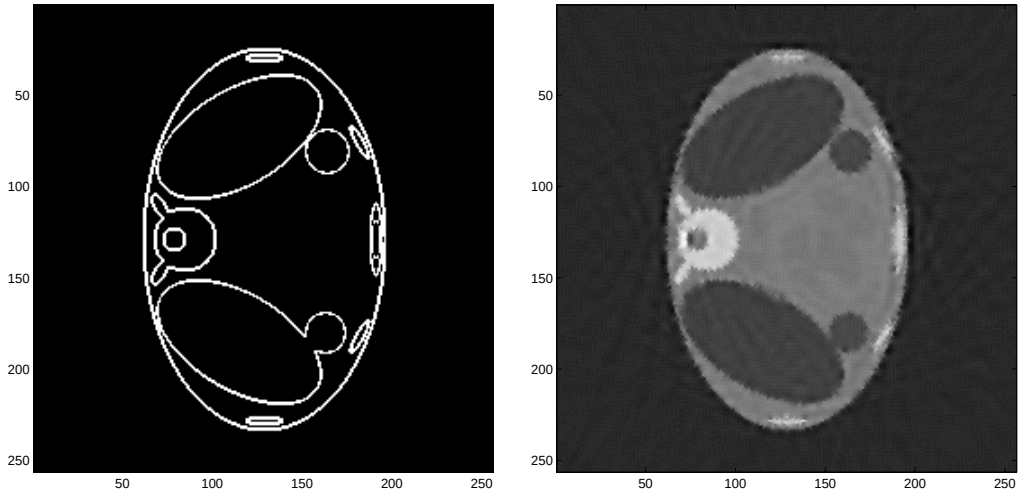


Figure 4: Reconstructions by data fusion (incomplete regions borders): The result in the right is obtained by assuming to know the borders positions of some of the regions of the object (those in the left figure).

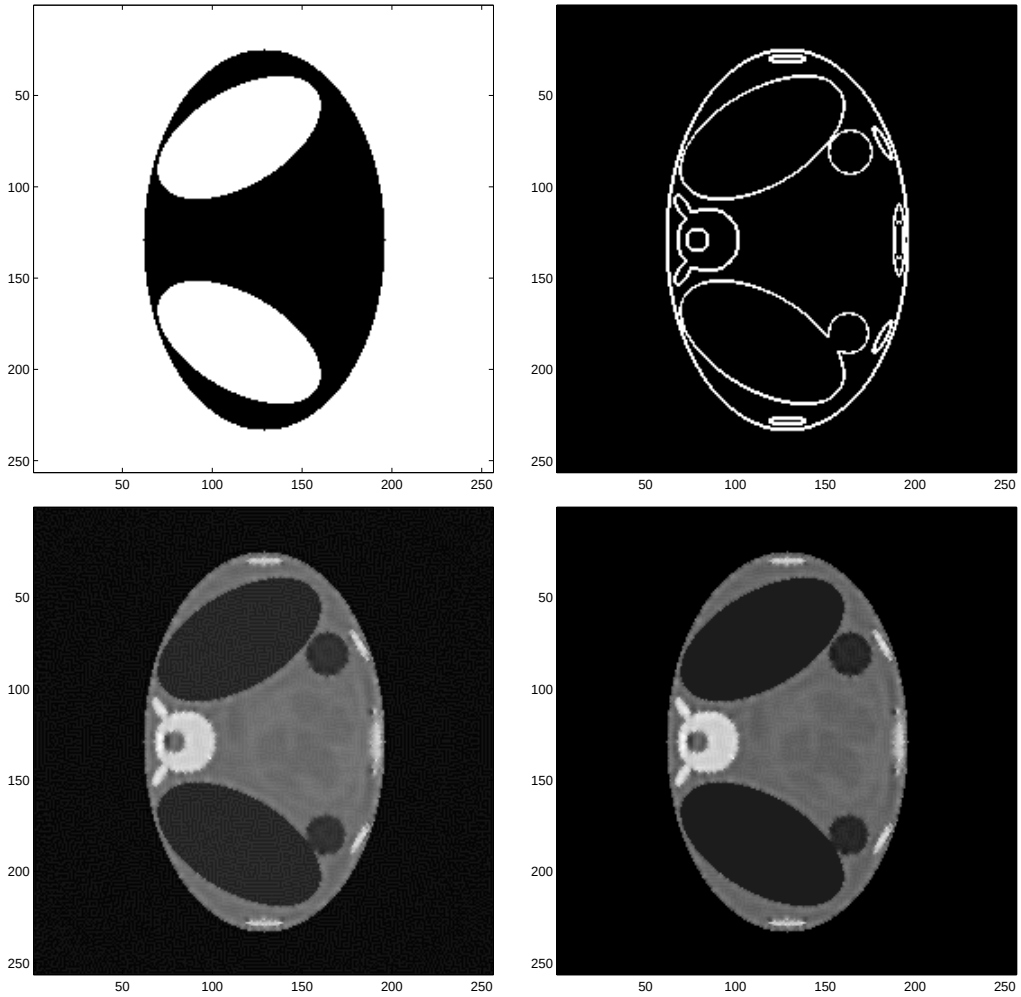


Figure 5: Reconstructions by data fusion (incomplete regions values, same as in Figure 3, and incomplete regions borders, same as in Figure 4): The results in the bottom are obtained for two different values of confidence for the regions values (different values of λ_2 which is too low at left but it seems to have good value at right.

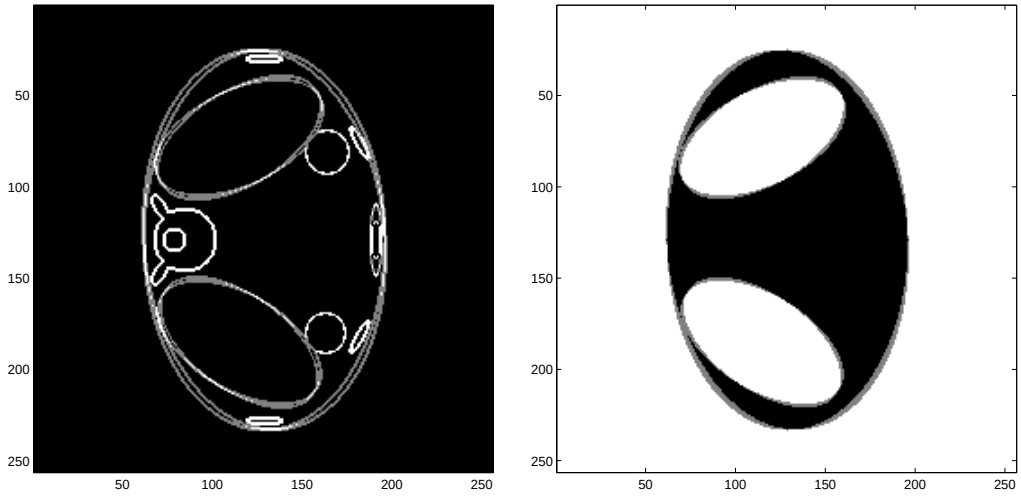


Figure 6: Geometrical knowledge with errors: Regions and borders errors.

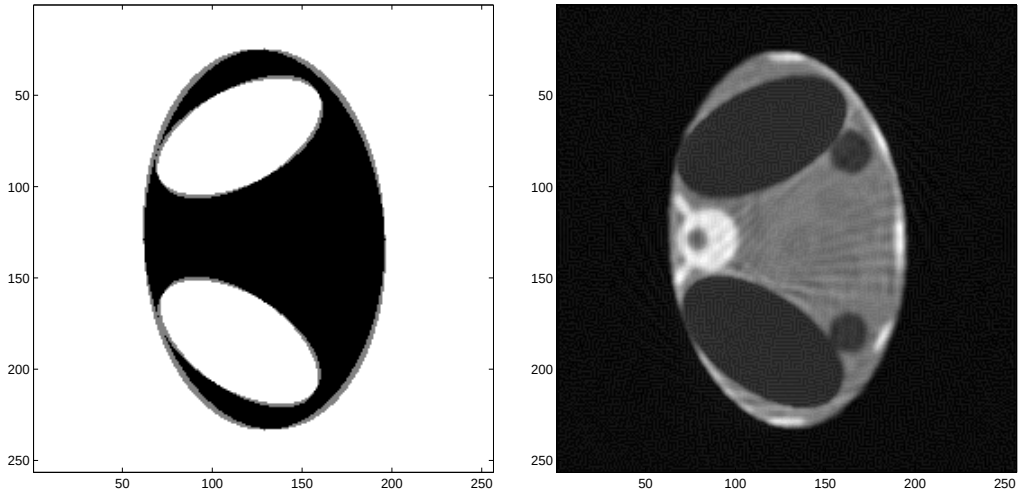


Figure 7: Reconstructions by data fusion with some errors in the support of the known regions: The result in the right is obtained by assuming to know the values of densities in black regions of of figure in the left. This result has to be compared with the result on Figure 3. Here we used the region information which is not correct in about 5% of pixel positions.

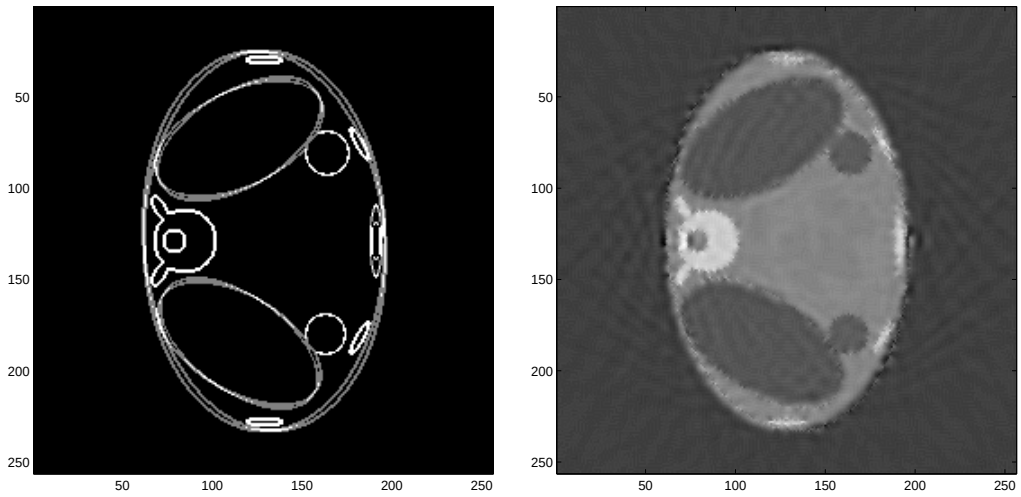


Figure 8: Reconstructions by data fusion with some errors in the support of the regions' borders: The result in the right is obtained by assuming to know the exact values of densities in black regions of of figure in the left. This result has to be compared with the results of Figure 4. Here we used the regions' borders information which is not correct in about 5% of pixel positions.

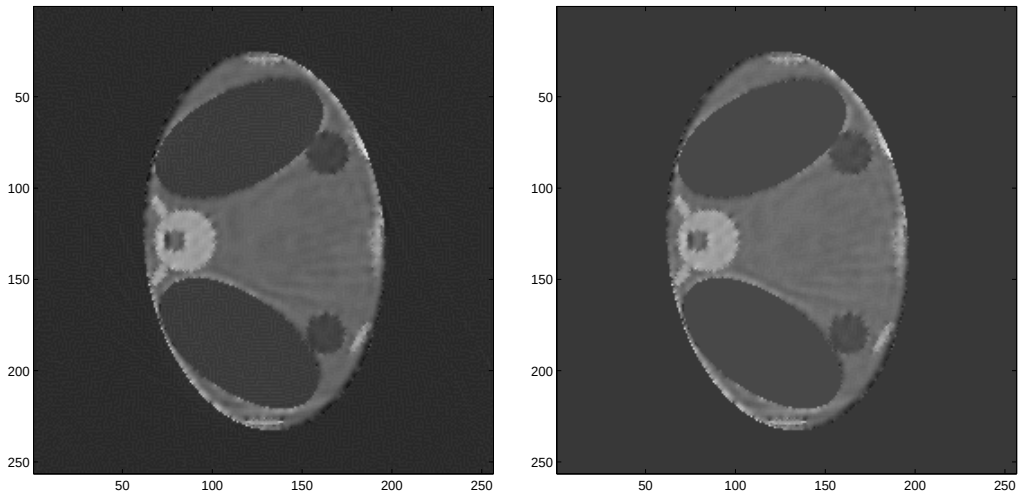


Figure 9: Reconstructions by data fusion with some errors in the geometrical data of regions and borders: These results are obtained for two different values of confidence for the regions values. These results are to be compared with those of figure 5 which were obtained with exact borders and regions informations.

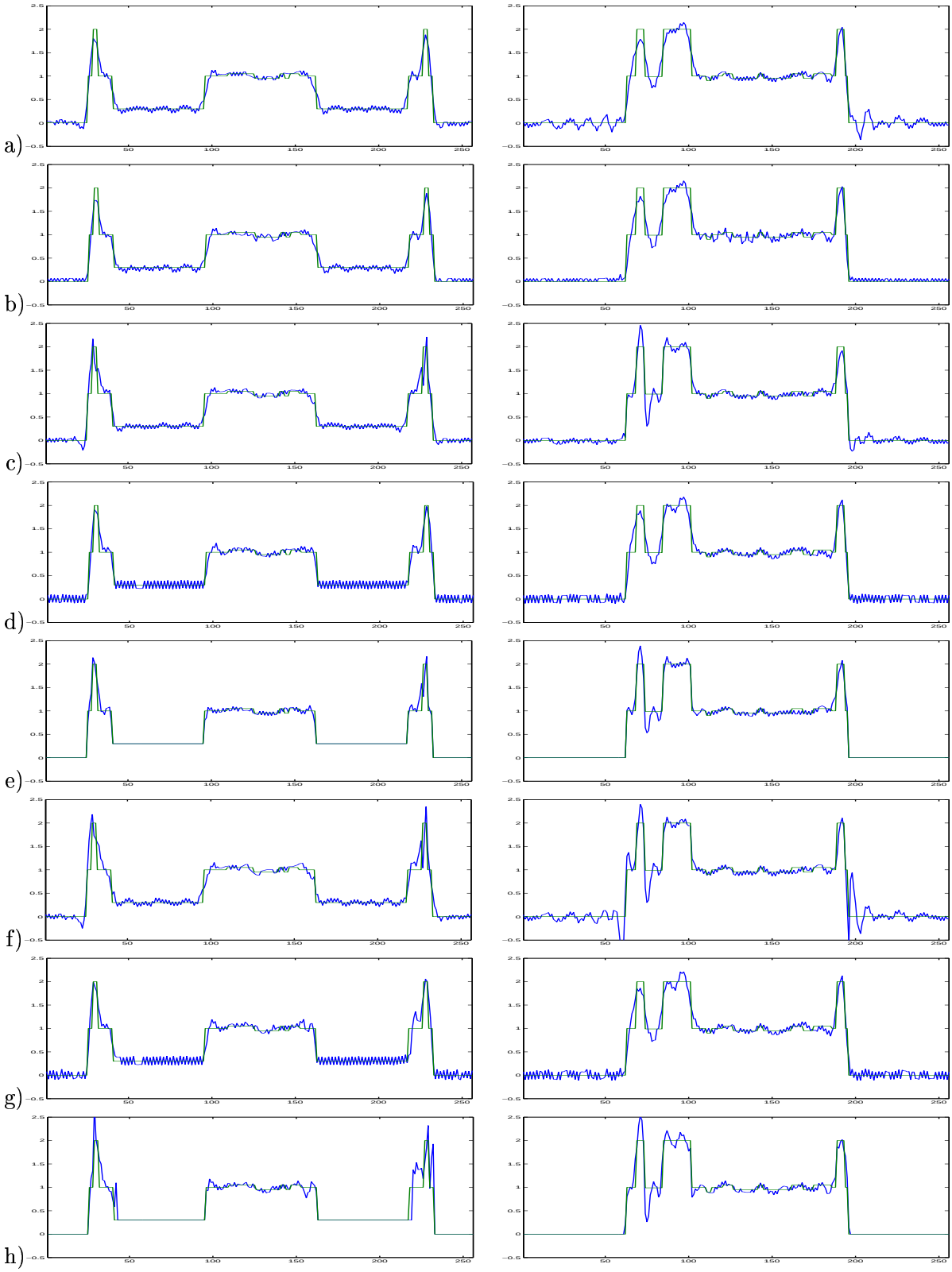


Figure 10: Comparison of the reconstructions results– horizontal (left) and vertical sections (right) of: a) quadratic regularization, b) quadratic regularization with positivity constraint, c) fusion with exact regions data, d) fusion with exact borders data, e) fusion with both exact regions and exact borders data, f) fusion with some errors in regions data, g) fusion with some errors in borders data, h) fusion with some errors in both regions and borders data.