

# 2D PET Image Reconstruction Using Robust $L_1$ Estimation of the Gaussian Mixture Model

Azra TAFRO<sup>1,\*</sup>, Damir SERŠIĆ<sup>2</sup>, Ana SOVIĆ KRŽIĆ<sup>2</sup>

<sup>1</sup> *Institute of Processes Engineering, University of Zagreb Faculty of Forestry and Wood Technology, Svetošimunska cesta 25, 10000 Zagreb, Croatia*

<sup>2</sup> *Department of Electronic Systems and Information Processing, University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia*  
*e-mail: azra.tafr@sumfak.unizg.hr, damir.sersic@fer.hr, ana.sovic.krzic@fer.hr*

Received: February 2021; accepted: April 2022

**Abstract.** An image or volume of interest in positron emission tomography (PET) is reconstructed from gamma rays emitted from a radioactive tracer, which are then captured and used to estimate the tracer's location. The image or volume of interest is reconstructed by estimating the pixel or voxel values on a grid determined by the scanner. Such an approach is usually associated with limited resolution of the reconstruction, high computational complexity due to slow convergence and noisy results.

This paper presents a novel method of PET image reconstruction using the underlying assumption that the originals of interest can be modelled using Gaussian mixture models. Parameters are estimated from one-dimensional projections using an iterative algorithm resembling the expectation-maximization algorithm. This presents a complex computational problem which is resolved by a novel approach that utilizes  $L_1$  minimization.

**Key words:** Gaussian mixture models, positron emission tomography, Expectation-Maximization (EM) algorithm, image segmentation.

## 1. Introduction

In positron emission tomography (PET), image reconstruction implies generating an image of a radiotracer's concentration to estimate physiologic parameters for objects (volumes) of interest. Pairs of photons arise from emissions of annihilated positrons, and crystals placed in the scanner detect these pairs. When two are activated at the same time, the device records an event. Each possible pair of crystals is connected by a line or volume of response (LOR, VOR), depending on whether the scan is 2D or 3D. Assuming there are no contaminating physical effects or noise, the total number of coincidence events detected and the total amount of tracer contained in the tube are proportional. In the 2D case, we observe events along lines of response (LORs) joining two detector elements, lying within a specified imaging plane. The data are recorded as event histograms (sinograms or projected data) or as a list of recorded photon-pair events (list-mode data). Classic PET image

---

\*Corresponding author.

reconstruction methodology is explained in more detail in e.g. Tong *et al.* (2010), Alessio and Kinahan (2006) or Reader and Zaidi (2007).

Modern image reconstruction methods are mostly based on maximum-likelihood expectation-maximization (MLEM) iterations. Maximum likelihood is used as the optimization criterion, combined with the expectation-maximization algorithm for finding its solution. To overcome the computational complexity and slow convergence of the MLEM, the ordered subsets expectation-maximization (OSEM) algorithm has been introduced (Hudson and Larkin, 1994). Since MLEM or OSEM estimation of pixels or voxels is usually noisy (Tong *et al.*, 2010), use of post-filtering methods is necessary. On the one hand, image reconstruction from its projections is a mature research field with well known methods and proven results. On the other hand, known limitations made a challenge for a different approach presented in this paper.

In image segmentation, a number of algorithms based on model-based techniques utilizing prior knowledge have been proposed to model uncertainty, cf. Zhang *et al.* (2001, 1994). The Gaussian mixture model (GMM) is well-known and widely used in a variety of segmentation and classification problems (Friedman and Russell, 1997; Ralašić *et al.*, 2018), as many observed quantities exhibit behaviour congruent with the model. A good overview of the application of GMMs and their generalizations to problems in image classification, image annotation and image retrieval can be found, e.g. in Tian (2018). However, in those problems the sample is from the GMM itself, whereas in this case the observed data is lower-dimensional (i.e. lines) and the points of origin are unknown.

In this paper, we propose a new robust method for reconstructing an object from a simulated PET image. The challenge of estimating Gaussian parameters from lower-dimensional data is solved by utilizing a novel  $L_1$  minimizing algorithm. With it, our framework becomes similar to the standard Expectation-Maximization (EM) algorithm. Instead of values in individual pixels or voxels, we estimate parameters of the object's model. This enables us to model one or more objects of interest, each as a mixture of one or more components, assuming that the entire object is a source of radiation with varying intensity. We focus on the two-dimensional case for clarity, but an extension to three dimensions would follow in a similar fashion. To the best of our knowledge, this is the first attempt to use parametric models in PET image reconstruction in this way. Parameters of Gaussian mixture models have not been estimated from lower-dimensional data due to computational complexity, which is circumvented here by the algorithm that efficiently finds global minimums. This paper is organized into six sections. Section 2 gives an overview of traditional methodology in 2D PET imaging. Section 3 describes the estimation of Gaussian mixture model parameters from PET data. Section 4 shows the implementation of iterative estimates in the EM-like algorithm. Finally, experimental results in Section 5 and discussion conclude the paper.

## 2. Two-Dimensional PET Imaging

Data are collected along lines lying within a specific imaging plane. Traditionally, data are organized into sets of projections, integrals along all lines (LORs) for a fixed direction  $\phi$ . The collection of all projections for  $0 \leq \phi < 2\pi$  forms a two-dimensional function

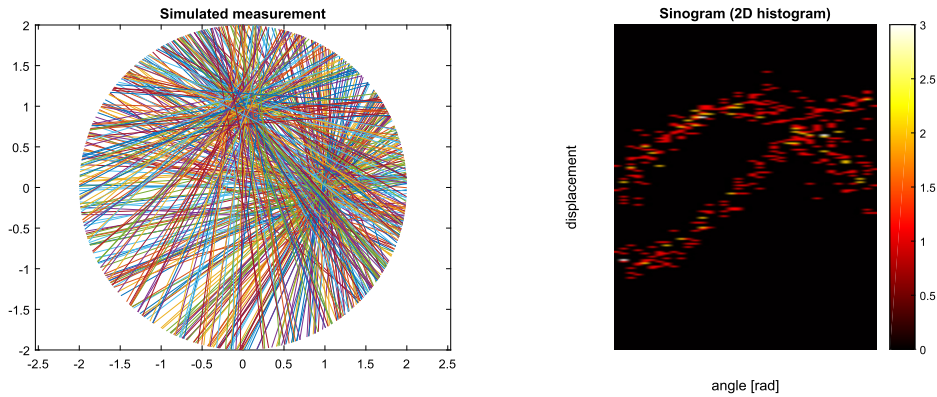


Fig. 1. (a) Simulated measurements for  $N = 400$  and  $K = 2$ . (b) The corresponding sinogram.

of the distance of the LOR from the origin, denoted by  $s$ , and angle  $\phi$ . The line-integral transform of  $f(x, y) \rightarrow p(s, \phi)$  is called the X-ray transform (Natterer, 1986), which in 2D is the same as the Radon transform (Helgason, 1999). The path for any fixed point in the projection space resembles a sinusoidal curve. This is traditionally represented by a sinogram, a graph where these paths are drawn for all points of activity simultaneously. Figure 1 illustrates one synthetic measurement and the corresponding sinogram. Classical reconstruction methods, most notably the filtered backprojection (FBP), rely on the integral transform, cf. Natterer (1986), O’Sullivan and Pawitan (1993), Alessio and Kinahan (2006). They do not include or depend on any assumptions about the data. Assumptions such as form, volume and dependence could be utilized to obtain more precise estimates.

### 3. Estimating the Gaussian Mixture Model

A Gaussian mixture model (Reynolds, 2015) is a weighted sum of  $K$  component Gaussian densities as given by the equation:

$$p(\mathbf{x}|\tau_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{k=1}^K \tau_k f(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{1}$$

where  $\mathbf{x}$  is a  $d$ -dimensional observation,  $\tau_i$  ( $i = 1, \dots, K$ ) are the weights of each Gaussian component, and  $f(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  are the Gaussian component densities. We assume that a given observation  $\mathbf{x}$  is a realization from exactly one of the  $K$  Gaussian mixture components, and each component density is a  $d$ -variate Gaussian function of the form:

$$f(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right). \tag{2}$$

The set of probabilities  $\{\tau_k\}$  such that  $\sum_{k=1}^K \tau_k = 1$  defines the probabilities that  $\mathbf{x}$  belongs to the corresponding Gaussian component. The complete Gaussian mixture model

is parameterized by the mean vectors  $\{\boldsymbol{\mu}_k\}$ , covariance matrices  $\{\boldsymbol{\Sigma}_k\}$  and mixture weights  $\{\tau_k\}$  for all Gaussian component densities.

Traditionally in this setting, the GMMs are used to model activity concentration in voxels (Layer *et al.*, 2015) or pixel values (Nguyen and Wu, 2013). Those models do not take into account the spatial correlation between observations, which can be corrected by introducing Markov random fields (Layer *et al.*, 2015; Nguyen and Wu, 2013). In other image segmentation problems, this can be addressed by modelling the mixture weights as probability vectors, thereby creating a spatially varying finite mixture model (SVFMM) (Xiong and Huang, 2015).

Here, we apply the GMM directly to the spatial data, focusing only on the locations and not the values at those locations. In other words, the location of the tracer, i.e. the particles that originate the gamma rays, is represented by the the  $K$  Gaussian components. The points  $\mathbf{x}$  that are realizations from these components are latent (unobserved). Our observations, i.e. events, are lines through these points at random angles  $\varphi$ , however, using convenient properties of Gaussian distributions we will still be able to accurately estimate the parameters.

As mentioned earlier, Gaussian mixture models have naturally appeared in many signal processing (Yu and Sapiro, 2011; Léger *et al.*, 2011) and biometrics problems (Reynolds, 2015; Reynolds *et al.*, 2000). Due to the flexibility of the mixtures, GMMs are convenient and effective for modelling various types of signals, as well as image inverse and missing/noisy data problems.

Note that in our simulations we assume that  $K$ , the number of mixture components, is known. In practical applications it should also be estimated, which is a further (and not insignificant) problem. For mixture components that are sufficiently separated in space, such as in Fig. 1, the sinogram itself could give insight into the most adequate number of components, otherwise ideas suggested in, e.g. Leroux (1992), Chen (1995), Cheng (2005), should be explored.

For clarity, in this section we focus only on one Gaussian component. Assuming we know a set of  $N$  events whose points of origin come from a Gaussian distribution with parameters  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , we can estimate those parameters.

### 3.1. Estimation of $\boldsymbol{\mu}$ Using Minimal Distance

Each event in two dimensions is uniquely given by its slope  $k = \tan \varphi$  and an intercept  $l$  (or by any two equivalent parameters) and we can write the event equations as

$$\mathbf{a}_i^T \mathbf{x} + l_i = 0, \quad \mathbf{a}_i = [\tan \varphi_i - 1]^T, \quad i = 1, \dots, N. \quad (3)$$

The mean vector  $\boldsymbol{\mu} = [\mu_x \ \mu_y]^T$  can be estimated in multiple ways. One method is to find the point in space whose total squared distance from all events is minimal. Clearly, the solution depends on our definition of distance. Formally, one would prefer to find the point whose total distance (not squared) is minimal, but that presents a nonlinear problem which is significantly more difficult to solve, and we will show that the linear problem provides a satisfying solution.

In general, the squared distance between  $d$ -dimensional vectors is defined as

$$d^2(\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1 - \mathbf{v}_2)^T \mathbf{W}(\mathbf{v}_1 - \mathbf{v}_2), \tag{4}$$

for some  $d \times d$  weight matrix  $\mathbf{W}$ . In particular, when  $\mathbf{W} = \mathbf{I}$  the distance in (4) is Euclidian, and  $\mathbf{W} = \mathbf{\Sigma}^{-1}$  gives the Mahalanobis distance. In our case,  $d = 2$  and a planar line is given by one equation, but similar results would follow for higher dimensions.

First, note that for a fixed  $\boldsymbol{\mu}$  the solution to

$$\min d^2(\mathbf{x}, \boldsymbol{\mu}) \quad \text{such that } \mathbf{a}_i \mathbf{x} + l_i = 0,$$

is (Golub and Van Loan, 2012) the solution to

$$\begin{bmatrix} 2\mathbf{W} & \mathbf{a}_i \\ \mathbf{a}_i^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}_i^\mu \\ \lambda \end{bmatrix} = \begin{bmatrix} 2\mathbf{W}\boldsymbol{\mu} \\ -l_i \end{bmatrix}. \tag{5}$$

Now the optimal  $\boldsymbol{\mu}$  is the one that minimizes the expression

$$\min \sum_{i=1}^N (\mathbf{x}_i^\mu - \boldsymbol{\mu})^T \mathbf{W}(\mathbf{x}_i^\mu - \boldsymbol{\mu}), \tag{6}$$

where  $\mathbf{x}_i^\mu$  is determined by (5). To solve this, first note that the solution to (5) is

$$\begin{bmatrix} \mathbf{x}_i^\mu \\ \lambda \end{bmatrix} = \begin{bmatrix} 2\mathbf{W} & \mathbf{a}_i \\ \mathbf{a}_i^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 2\mathbf{W}\boldsymbol{\mu} \\ -l_i \end{bmatrix}.$$

To accommodate this, we will denote the expression in (6) by  $d^2$  and expand it:

$$\begin{aligned} d^2 &= \sum_{i=1}^N (\mathbf{x}_i^\mu - \boldsymbol{\mu})^T \mathbf{W}(\mathbf{x}_i^\mu - \boldsymbol{\mu}) \\ &= \sum_{i=1}^N \left( \begin{bmatrix} \mathbf{x}_i^\mu \\ \lambda \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu} \\ \lambda \end{bmatrix} \right)^T \cdot \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \cdot \left( \begin{bmatrix} \mathbf{x}_i^\mu \\ \lambda \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu} \\ \lambda \end{bmatrix} \right). \end{aligned}$$

For simplicity, denote

$$\mathbf{x}_{i\lambda} = \begin{bmatrix} \mathbf{x}_i^\mu \\ \lambda \end{bmatrix}, \quad \boldsymbol{\mu}_\lambda = \begin{bmatrix} \boldsymbol{\mu} \\ \lambda \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}.$$

Now  $d^2$  equals

$$\sum_{i=1}^N (\mathbf{x}_{i\lambda}^T \tilde{\mathbf{W}} \mathbf{x}_{i\lambda} - \mathbf{x}_{i\lambda}^T \tilde{\mathbf{W}} \boldsymbol{\mu}_\lambda - \boldsymbol{\mu}_\lambda^T \tilde{\mathbf{W}} \mathbf{x}_{i\lambda} + \boldsymbol{\mu}_\lambda^T \tilde{\mathbf{W}} \boldsymbol{\mu}_\lambda),$$

which we will differentiate piecewise to find the minimum. We have:

1)

$$\frac{d}{d\boldsymbol{\mu}} \mathbf{x}_{i\lambda}^T \tilde{\mathbf{W}} \mathbf{x}_{i\lambda} = 2 \begin{bmatrix} 2\mathbf{W}\boldsymbol{\mu} \\ -l_i \end{bmatrix}^T \mathbf{B}_i \begin{bmatrix} 2\mathbf{W} & \mathbf{a}_i \\ \mathbf{a}_i^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2\mathbf{W} \\ \mathbf{0} \end{bmatrix},$$

where

$$\mathbf{B}_i = \left( \begin{bmatrix} 2\mathbf{W} & \mathbf{a}_i \\ \mathbf{a}_i^T & 0 \end{bmatrix}^{-1} \right)^T \tilde{\mathbf{W}}.$$

2)

$$\frac{d}{d\boldsymbol{\mu}} \boldsymbol{\mu}_\lambda^T \tilde{\mathbf{W}} \mathbf{x}_{i\lambda} = \boldsymbol{\mu}_\lambda^T \mathbf{B}_i^T \begin{bmatrix} 2\mathbf{W} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} 2\mathbf{W}\boldsymbol{\mu} \\ -l_i \end{bmatrix}^T \mathbf{B}_i \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}.$$

3)

$$\frac{d}{d\boldsymbol{\mu}} \boldsymbol{\mu}_\lambda^T \tilde{\mathbf{W}} \mathbf{x}_{i\lambda} = \boldsymbol{\mu}_\lambda^T \mathbf{B}_i^T \begin{bmatrix} 2\mathbf{W} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} 2\mathbf{W}\boldsymbol{\mu} \\ -l_i \end{bmatrix}^T \mathbf{B}_i \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}.$$

4)

$$\frac{d}{d\boldsymbol{\mu}} \boldsymbol{\mu}_\lambda^T \tilde{\mathbf{W}} \boldsymbol{\mu}_\lambda = 2\boldsymbol{\mu}_\lambda^T \tilde{\mathbf{W}} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}.$$

Note that in all calculations we use the fact that  $\mathbf{W}$  and, by extension,  $\tilde{\mathbf{W}}$  are symmetric. By plugging these equations into  $\frac{d}{d\boldsymbol{\mu}}$  and equating that with 0 to obtain the minimum, we get:

$$\begin{aligned} \frac{d}{d\boldsymbol{\mu}} d^2 &= \sum_{i=1}^N \left( \begin{bmatrix} 2\mathbf{W}\boldsymbol{\mu} \\ -l_i \end{bmatrix}^T \cdot \mathbf{B}_i \left( \begin{bmatrix} 2\mathbf{W} & \mathbf{a}_i \\ \mathbf{a}_i^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2\mathbf{W} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right) - \right. \\ &\quad \left. - \boldsymbol{\mu}_\lambda^T \left( \mathbf{B}_i^T \begin{bmatrix} 2\mathbf{W} \\ \mathbf{0} \end{bmatrix} - \tilde{\mathbf{W}} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right) \right) = 0. \end{aligned}$$

Define

$$\begin{bmatrix} \mathbf{M}_i \\ \mathbf{m}_i \end{bmatrix} = \mathbf{B}_i \left( \begin{bmatrix} 2\mathbf{W} & \mathbf{a}_i \\ \mathbf{a}_i^T & 0 \end{bmatrix} \begin{bmatrix} 2\mathbf{W} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \right),$$

$$\begin{bmatrix} \mathbf{N}_i \\ \mathbf{0} \end{bmatrix} = \mathbf{B}_i^T \begin{bmatrix} 2\mathbf{W} \\ \mathbf{0} \end{bmatrix} - \tilde{\mathbf{W}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

From the previous equation we now have

$$\begin{aligned} \sum_{i=1}^N \left( [2\boldsymbol{\mu}^T \mathbf{W}^T, -l_i] \cdot \begin{bmatrix} \mathbf{M}_i \\ \mathbf{m}_i \end{bmatrix} - [\boldsymbol{\mu}^T, \lambda_i] \cdot \begin{bmatrix} N_i \\ \mathbf{0} \end{bmatrix} \right) &= 0, \\ \sum_{i=1}^N (2\boldsymbol{\mu}^T \mathbf{W}^T \mathbf{M}_i - l_i \mathbf{m}_i - \boldsymbol{\mu}^T N_i) &= 0, \\ \boldsymbol{\mu}^T \sum_{i=1}^N (2\mathbf{W}^T \mathbf{M}_i - N_i) &= \sum_{i=1}^N l_i \mathbf{m}_i, \end{aligned}$$

from which it follows that

$$\boldsymbol{\mu}^T = \left( \sum_{i=1}^N l_i \mathbf{m}_i \right) \cdot \left( \sum_{i=1}^N (2\mathbf{W}^T \mathbf{M}_i - N_i) \right)^{-1}. \tag{7}$$

It remains to verify that this stationary point  $\boldsymbol{\mu}$  is also a turning point. However, since a point whose sum of squared distances from all lines is minimal must exist from a geometrical perspective, the solution in (7) is indeed the (global) minimum.

### 3.2. Estimation of $\Sigma$ Using 1D Projections

In the two-dimensional setting, since  $\Sigma$  is a symmetric matrix,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix},$$

we need to estimate three parameters:  $\Sigma_{11}$ ,  $\Sigma_{12}$  and  $\Sigma_{22}$ .

Figure 2 depicts a single event and its corresponding LOR. Recall that the line is determined by two parameters, one of which is the slope  $\tan \varphi$ , where  $\varphi$  is the angle between the line and the  $x$ -axis. If we rotate the coordinate system by  $\varphi - \frac{\pi}{2}$ , in the new coordinate system the  $y$ -axis is parallel to the event. If we represent the Gaussian distribution as in the image, in this new setup it rotated by  $\psi = \frac{\pi}{2} - \varphi$ . Since a rotated Gaussian distribution is again Gaussian, in this new coordinate system the event is parallel to the  $y$ -axis, and the Gaussian distribution has parameters

$$(T\boldsymbol{\mu}, T\Sigma T^T), \quad \text{where } T = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \text{ is the rotation matrix.} \tag{8}$$

Given the original event parameters, the coordinates of the intersection of the event and the new  $x$ -axis are  $(-l \sin \psi, 0)$ . The projection of the Gaussian distribution onto the new  $x$ -axis remains Gaussian, and the parameters of this one-dimensional distribution are

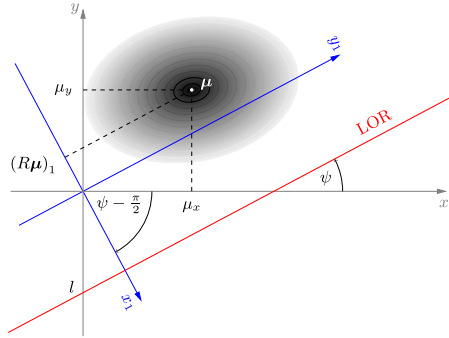


Fig. 2. Illustration of the rotation. The event (LOR) is parallel to the new y-axis.

components of the original two-dimensional Gaussian. From (8) we see that the expectation is given by

$$(T\boldsymbol{\mu})_1 = \cos \psi \mu_x - \sin \psi \mu_y,$$

and variance is

$$(T\Sigma T^T)_{11} = \cos^2 \psi \Sigma_{11} - 2 \cos \psi \sin \psi \Sigma_{12} + \sin^2 \psi \Sigma_{22}. \tag{9}$$

We can use these one-dimensional projections and their squared (Euclidian) distance from the mean, to estimate the variance in (9). This squared distance is equal to  $(\cos \psi \mu_x - \sin \psi \mu_y + l \sin \psi)^2$ , which gives us a system of equations:

$$\mathbf{A}\mathbf{s} = \mathbf{b}, \quad \text{where } \mathbf{s} = \begin{bmatrix} \Sigma_{11} \\ \Sigma_{12} \\ \Sigma_{22} \end{bmatrix}, \tag{10}$$

$$\mathbf{A} = \begin{bmatrix} \cos^2 \psi_1 & -2 \sin \psi_1 \cos \psi_1 & \sin^2 \psi_1 \\ \cos^2 \psi_2 & -2 \sin \psi_2 \cos \psi_2 & \sin^2 \psi_2 \\ \vdots & \vdots & \vdots \\ \cos^2 \psi_N & -2 \sin \psi_N \cos \psi_N & \sin^2 \psi_N \end{bmatrix},$$

$$\text{and } \mathbf{b} = \begin{bmatrix} (\cos \psi_1 \mu_x - \sin \psi_1 \mu_y + l_1 \sin \psi_1)^2 \\ \vdots \\ (\cos \psi_N \mu_x - \sin \psi_N \mu_y + l_N \sin \psi_N)^2 \end{bmatrix}.$$

This problem is clearly overdetermined, since there are (dozens of) thousands of measurements. There is no exact solution and the best approximation is found by solving

$$\min_s \|\mathbf{A}\mathbf{s} - \mathbf{b}\|. \tag{11}$$



Note that the solution will depend on the type of norm used in (11). Following classical methodology, the ordinary least squares (OLS) method gives the solution that minimizes the  $L_2$  norm. This solution can be found by solving the equivalent problem

$$A^T A s = A^T \cdot b, \quad \text{i.e. } s = (A^T A)^{-1} A^T \cdot b.$$

As we will show in Section 5, OLS performs poorly when we introduce more than one component, so we will need to use alternative methods.  $L_1$  minimization is a popular approach in engineering problems because it gives sparse solutions, but it had also been shown that it is less sensitive than the more traditional  $L_2$  minimization (Bektaş and Şişman, 2010). The main argument against  $L_1$  minimization would be computational complexity, which can be alleviated by using iterative methods. In this paper, the  $L_1$  minimization algorithm proposed in Sović Kržić and Seršić (2018) is used to solve a modified problem

$$A s = k \cdot b, \tag{12}$$

where  $k$  is a constant corrective scale factor. In a sufficiently large sample, one would have enough data points in each direction  $\phi$  to obtain a one-dimensional variance estimate. In the absence of that, we are able to obtain a robust estimator for the parameters of the covariance matrix following the median absolute deviation (MAD) estimator of deviation  $\sigma$  (Rousseeuw and Croux, 1993), i.e.

$$k = (\Phi(0.75))^2 \approx 1.4826^2,$$

where  $\Phi$  denotes the distribution function of the standard normal random variable.

#### 4. EM-Like Algorithm

In many statistical models, maximum likelihood parameter estimation can not be performed directly since most equations do not have an explicit closed solution. Several iterative methods have been developed to combat this, most notably the expectation-maximization (EM) algorithm. It had been proposed and used in many different circumstances before its proper introduction in Dempster *et al.* (1977). The name comes from its two-step setup, namely the expectation (E) and maximization (M) step which interchange until an acceptable solution is found. In the context of GMMs, and generally in emission tomography (Shepp and Vardi, 1982), the issue is that the true mixture membership is unknown (latent). The E step for given mixture parameters estimates probabilities  $\{\tau_k\}$  from (2) for each data point. The M step then (re)estimates the mixture parameters from the points assigned to that mixture with their corresponding probabilities. As already mentioned, our observations are lines and the true data points are latent, however, we can replicate the iterative steps using estimates from Section 3. Alternatively,

this could also be considered a Lloyd-like algorithm, where the difference from the conventional Lloyd's algorithm (Lloyd, 1982) is that we allow different distance functions of the form (4). The algorithm initializes parameters arbitrarily, and then alternates between the following steps:

1. Compute class membership probabilities. For each event, compute the squared distance from each component mean. We distinguish between a *hard* classification where we assign the event to its nearest component, and a *soft* classification where membership probability is inversely proportional to the squared distance.
2. Estimate component parameters. Either from a hard or soft classification, where each event participates with its proportional share, parameters of each component are estimated using methodology from Section 3.

The  $L_1$  minimization algorithm recursively reduces and increases dimensionality of the observed subspace and uses weighted median to efficiently find the global minimum, and has shown to overperform state-of-the-art competitive methods when there are relatively few parameters to be estimated from a very high number of equations. For details, see Appendix and Sović Kržić and Seršić (2018).

This approach allows for different variants, depending on the type of distance used (Tafro and Seršić, 2019).

In this paper, initial steps of the iterative algorithm use Euclidian distance in (4). Since later iterations improve the estimates, the distance gradually transforms into the Mahalanobis distance, i.e.

$$W = (1 - \alpha)I + \alpha \hat{\Sigma}^{-1}, \quad (13)$$

where  $\alpha$  increases from 0 to 1. It is known that, for  $\Sigma$  given, the estimate obtained using the Mahalanobis distance is also the maximum likelihood estimate for  $\mu$ . Therefore, in later iterations we obtain an MLE-like estimate of  $\mu$ .

## 5. Results

To evaluate the methodology, we experimented in the two-dimensional setting with  $K = 1$  and  $K = 2$  components.

First, for proof of concept we show that the method in Section 3 provides good estimates with both  $L_1$  and  $L_2$  minimization, for several covariance matrices with varying corresponding correlation coefficients:

$$\Sigma_1 = 0.05I, \quad \Sigma_2 = \begin{bmatrix} 0.02 & -0.01 \\ -0.01 & 0.05 \end{bmatrix}, \quad \Sigma_3 = \begin{bmatrix} 0.01 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}.$$

These matrices represent different shapes, i.e. different types of dependence, from independent ( $\Sigma_1$ ) to strongly dependent ( $\Sigma_3$ ) variables. Since the 2D covariance matrix is of

the form

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix},$$

it is determined by three parameters –  $\sigma_x$ ,  $\sigma_y$  and  $\rho(= \rho_{xy})$ . This corresponds to the three-dimensional vector

$$s = [\Sigma_{11} \ \Sigma_{12} \ \Sigma_{22}]^T = [\sigma_x^2 \ \rho\sigma_x\sigma_y \ \sigma_y^2]^T$$

in (10). Instead of observing true and estimated  $\Sigma$  matrices we will calculate the error in estimation of  $s$  for each  $s$  that corresponds to matrices above:

$$s_1 = \begin{bmatrix} 0.05 \\ 0 \\ 0.05 \end{bmatrix}, \quad s_2 = \begin{bmatrix} 0.02 \\ -0.01 \\ 0.05 \end{bmatrix}, \quad s_3 = \begin{bmatrix} 0.01 \\ 0.02 \\ 0.05 \end{bmatrix}.$$

Note that the corresponding correlation coefficients can be calculated from these vectors, and they are  $\rho_1 = 0$ ,  $\rho_2 \approx -0.3$ ,  $\rho_3 \approx 0.9$ .

### 5.1. One-Dimensional Estimation

Accuracy of an estimate  $\hat{v}$  of vector  $v$  can be assessed in many ways, for illustrative purposes we chose relative error, i.e.

$$\|e\| = \frac{\|\hat{v} - v\|}{\|v\|},$$

where  $\|\cdot\|$  denotes the standard Euclidian norm in both expressions.

Initial value of  $\alpha$  in the distance weight matrix was set to zero, and proportionally increased to 1 in the final iteration. The left side of Fig. 3 illustrates how the errors in

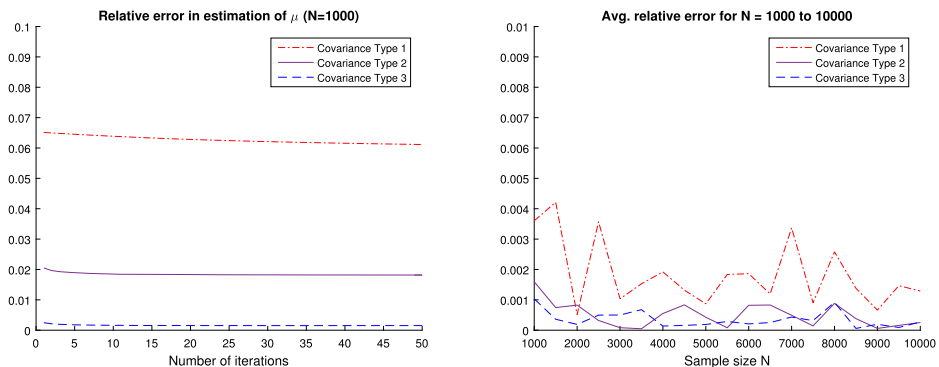


Fig. 3. Left: Relative error in estimation of  $\mu$  from a single measurement ( $N = 1000$  events). Right: Average relative error for increasing sample size, for three types of covariance.

Table 1  
Mean estimation error,  $K = 1$ .

$N = 1000$	$s_1$	$s_2$	$s_3$	$N = 10000$	$s_1$	$s_2$	$s_3$
$L_1$ method	13.78%	11.88%	9.28%	$L_1$ method	4.28%	3.66%	2.93%
$L_2$ method	8.27%	7.61%	7.6%	$L_2$ method	2.61%	2.38%	2.37%

the estimation of  $\boldsymbol{\mu}$  change with the number of iterations in a single experiment with  $N = 1000$  points, for all three types of covariance matrices. Depending on the type of covariance matrix, the accuracy increases as  $|\rho|$  increases, but in all three scenarios the estimation is stable within less than 10 iterations, and sufficiently accurate. In further experiments, the number of iterations in the algorithm was set to 10, with the distance metric varying according to (13). The right side of Fig. 3 shows how the accuracy in estimation of  $\boldsymbol{\mu}$  increases with the sample size, with the number of experiments set to 100 for each sample size. The relative errors decreases as the number of points increases, as is expected, but we can also notice that the error is fairly similar even for sample sizes that are very small (compared to the usual sample sizes in PET imagery).

For each of these types of covariances we then simulated a measurement from  $N = 1000$  and  $N = 10000$  points and calculated the average estimate of vector  $s$  using  $L_1$  and  $L_2$  minimization separately. We repeated the experiment 1000 times in each case. Results are given in Table 1.

Given that the variance of the traditional standard deviation estimator (from points) equals  $\frac{\sigma^4}{n-1}$ , estimations within 10% from  $N = 1000$  events are acceptable, which justifies the methodology described in Section 3. Significantly more accurate estimation from  $N = 10000$  events further confirms this. It is also notable that in general the accuracy of the estimate increases as  $|\rho|$  increases from 0 to 1, i.e. when the set of points of origin is more elongated in shape.

## 5.2. Two-Dimensional Estimation

For  $K = 2$  components we repeated the experiment as described in Section 4 for various combinations of types of vector  $s$ . We used hard classification, where we assign each line to at most one component. We experimented with various constraints, from simply assigning events to more likely components to assignments only when the probability of belonging is above a certain threshold.

For synthetic measurements from a variety of original (real) GMMs the algorithm proved robust regardless of the values of initial parameters, with estimation using  $L_1$  minimization and the scaling factor  $k$  correcting the bias. The  $L_2$  minimization method proved inefficient, since wrongly assigned events would cause unstable estimations and “breaks” in the algorithm.

Table 2 shows the percentage of correctly classified events for several combinations of covariance matrices, where a total of  $N = 4000$  events were simulated from two mixtures ( $N_1 = 2500$ ,  $N_2 = 1500$ ) for each combination. Events were classified with no probability thresholds, i.e. each event was assigned to the more likely component. We can conclude

Table 2  
Correct classification,  $K = 2$ .

Cov. 1	Cov. 2	Comp. 1	Comp. 2	Total
$\Sigma_1$	$\Sigma_2$	87.63%	93.49%	89.83%
$\Sigma_2$	$\Sigma_3$	93.52%	92.13%	93.00%
$\Sigma_3$	$\Sigma_1$	91.26%	95.83%	92.98%

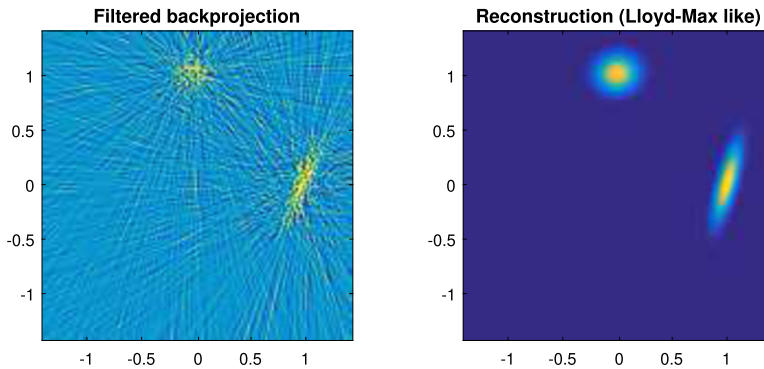


Fig. 4. (a) Classical FBP reconstruction. (b) Proposed reconstruction using  $L_1$  minimization.

overall that the algorithm has a very good classification rate which, combined with the results for individual components, gives reason to expect accurate estimations for multiple components.

An illustration of the results for  $K = 2$ ,  $N = 4000$  is shown in Fig. 4, along with the corresponding classical FBP method.

## 6. Conclusion

These results show proof of concept that it is possible to reconstruct data from latent mixture models using lower-dimensional observations and state-of-the-art computational techniques. Estimation from lower-dimensional measurements had not been thoroughly developed for Gaussian mixture models in general, and especially in this context. This paper shows that the main obstacle, slow computational speed, can be evaded by an alternative approach to minimization, which opens the door to various new procedures. Further work includes other metrics and membership calculation and utilizing traditional methods to obtain the number of mixing components and optimal initial values.

The fact that the reconstructed image is given by its parametric model (mean vectors  $\{\mu_k\}$ , covariance matrices  $\{\Sigma_k\}$  and mixture weights  $\{\tau_k\}$ ) has many benefits. This approach would allow for accurate reconstruction from significantly fewer measurements, thus decreasing patients' exposure to radiation. It is virtually of infinite resolution, since the Gaussian components can be evaluated at each spatial point. The model is sparse: it consists from only a few parameters needed for successful object representation. Hence,

future research will be oriented to compressed sensing approach (Rani *et al.*, 2018; Ralašić *et al.*, 2018, 2019) for reducing the number of projections, in this case reduced radio-tracer's concentration. Due to robustness of the proposed reconstruction method, a post-filtering step is not needed. This novel point of view aims to match the accuracy of iterative parametric models with the computational speed of the analytic approach by combining advantageous statistical models with state-of-the-art techniques from other image reconstruction problems.

## Appendix

The  $L_1$  minimization algorithm recursively reduces and increases dimensionality of the observed subspace and uses weighted median to efficiently find the global minimum. Reduction of dimensionality is achieved by extracting of parameters and inserting them into remaining equations in (12). If  $[A_{i1} \ A_{i2} \ A_{i3}]$  is the  $i$ -th row of matrix  $A$ , and  $b_i$  the  $i$ -th element of vector  $k \cdot \mathbf{b}$ , the  $i$ -th equation of the system is

$$b_i = \Sigma_{11}A_{i1} + \Sigma_{12}A_{i2} + \Sigma_{13}A_{i3}.$$

We choose equation  $j_1$  from the set  $i = 1, \dots, N$  and extract one of its parameters, e.g.  $\Sigma_{11}$ :

$$\Sigma_{11} = -\frac{A_{j_1 2}}{A_{j_1 1}}\Sigma_{12} - \frac{A_{j_1 3}}{A_{j_1 1}}\Sigma_{22} + b_{j_1}. \quad (14)$$

We insert it into all other equations and get a new system with only two unknown parameters:

$$b_i - b_{j_1}A_{i1} = \left(A_{i2} - \frac{A_{j_1 2}}{A_{j_1 1}}A_{i1}\right)\Sigma_{12} + \left(A_{i3} - \frac{A_{j_1 3}}{A_{j_1 1}}A_{i1}\right)\Sigma_{22},$$

for  $i \neq j_1$ . Now, we choose some other equation  $j_2$ ,  $j_2 \neq j_1$  and extract one of the remaining parameters, e.g.  $\Sigma_{12}$ :

$$\Sigma_{12} = -\frac{A_{j_2 3}A_{j_1 1} - A_{j_1 3}A_{j_2 1}}{A_{j_2 2}A_{j_1 1} - A_{j_1 2}A_{j_2 1}}\Sigma_{22} + \frac{(b_{j_2} - b_{j_1}A_{j_2 1})A_{j_1 1}}{A_{j_2 2}A_{j_1 1} - A_{j_1 2}A_{j_2 1}}. \quad (15)$$

We insert  $\Sigma_{12}$  into all other equations and get the system of  $N - 2$  equations with only one unknown parameter  $\Sigma_{22}$ :

$$b_i^{(1)} = A_i^{(1)}\Sigma_{22}, \quad (16)$$

where  $i \neq j_1, j_2$ ,

$$b_i^{(1)} = b_i - b_{j_1} A_{i1} - \left( A_{i2} - \frac{A_{j_1 2} A_{i1}}{A_{j_1 1}} \right) \frac{(b_{j_2} - b_{j_1} A_{j_2 1}) A_{j_1 1}}{A_{j_2 2} A_{j_1 1} - A_{j_1 2} A_{j_2 1}},$$

$$A_i^{(1)} = A_{i3} - \frac{A_{j_1 3}}{A_{j_1 1}} A_{i1} - \left( A_{i2} - \frac{A_{j_1 2}}{A_{j_1 1}} A_{i1} \right) \frac{A_{j_2 3} A_{j_1 1} - A_{j_1 3} A_{j_2 1}}{A_{j_2 2} A_{j_1 1} - A_{j_1 2} A_{j_2 1}}.$$

Parameter  $\Sigma_{22}$  is calculated by minimizing  $L_1$  norm

$$\min \sum_{\substack{i=1 \\ i \neq j_1, j_2}}^N |b_i^{(1)} - A_i^{(1)} \Sigma_{22}|$$

which equals

$$\min \sum_{\substack{i=1 \\ i \neq j_1, j_2}}^N |A_i^{(1)}| \left| \frac{b_i^{(1)}}{A_i^{(1)}} - \Sigma_{22} \right|.$$

The value of parameter  $\Sigma_{22}$  is given by the weighted median (MED):

$$(\Sigma_{22}, j_3) = \text{MED} \left( \left| A_i^{(1)} \right| \diamond \frac{b_i^{(1)}}{A_i^{(1)}} \Bigg|_{i=1, i \neq j_1, j_2}^N \right), \tag{17}$$

where  $j_3$  is an ordinal number of the concomitant equation and  $\diamond$  is the replication operator. The weighted median can be obtained using the algorithm given in Sović Kržić and Seršić (2018). The value of parameter  $\Sigma_{22}$  is an element of set  $\{b_i^{(1)}/A_i^{(1)}\}$ . Chosen equations  $\{j_1, j_2, j_3\}$  define a local minimum in 1D, a vertex.

Return to a higher dimension is achieved by putting calculated parameter value  $\Sigma_{22}$  in (15). To further descending in  $L_1$  cost surface, we fix equations  $j_1$  and  $j_3$ , and try to find new  $j_4$  using (16)–(17). If  $j_4 \neq j_2$ , new vertex is defined by  $\{j_1, j_2, j_3\} = \{j_1, j_3, j_4\}$  and we repeat the same procedure. If  $j_4 = j_2$ , we conclude that the vertex  $\{j_1, j_2, j_3\}$  is a local minimum in observed 2D subspace, thus we return to 3D by fixing  $j_2$  and  $j_3$  and try to find new  $j_4$  instead of  $j_1$ . We repeat the previous procedure until we cannot find new equation. Since the  $L_1$  cost surface is convex, the global minimum is reached. Parameters  $\Sigma_{11}$ ,  $\Sigma_{12}$  and  $\Sigma_{22}$  are given by (14), (15) and (17).

### Acknowledgments

This research has been supported by the Croatian Science Foundation under the project IP-2019-04-6703.

## References

- Alessio, A., Kinahan, P. (2006). PET image reconstruction. *Nuclear Medicine*, 1, 1–22. <https://doi.org/10.1118/1.2358198>.
- Bektaş, S., Şişman, Y. (2010). The comparison of L1 and L22-norm minimization methods. *International Journal of Physical Sciences*, 5(11), 1721–1727.
- Chen, J. (1995). Optimal rate of convergence for finite mixture models. *The Annals of Statistics*, 23(1), 221–233. <https://doi.org/10.1214/aos/1176324464>.
- Cheng, Y.-m. (2005). Maximum weighted likelihood via rival penalized EM for density mixture clustering with automatic model selection. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 750–761. <https://doi.org/10.1109/TKDE.2005.97>.
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39, 1–38. <https://doi.org/10.2307/2984875>.
- Friedman, N., Russell, S. (1997). Image segmentation in video sequences: a probabilistic approach. In: *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence, UAI'97*, pp. 175–181. <http://dl.acm.org/citation.cfm?id=2074226.2074247>.
- Golub, G.H., Van Loan, C.F. (2012). *Matrix Computations*, Vol. 3. JHU Press. <https://doi.org/10.1137/1028073>.
- Helgason, S. (1999). *The Radon Transform*, Vol. 2. Springer.
- Hudson, H.M., Larkin, R.S. (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4), 601–609. <https://doi.org/10.1109/42.363108>.
- Layer, T., Blaukner, M., Knäusel, B., Georg, D., Neuwirth, J., Baum, R.P., Schuchardt, C., Wiessalla, S., Matz, G. (2015). PET image segmentation using a Gaussian mixture model and Markov random fields. *EJNMMI Physics*, 2(1), 9. <https://doi.org/10.1186/s40658-015-0110-7>.
- Léger, F., Yu, G., Sapiro, G. (2011). Efficient matrix completion with Gaussian models. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 1113–1116.
- Leroux, B.G. (1992). Consistent estimation of a mixing distribution. *The Annals of Statistics*, 1350–1360. <https://doi.org/10.1214/aos/1176348772>.
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>.
- Natterer, F. (1986). *The Mathematics of Computerized Tomography*, Vol. 32. Siam. <https://doi.org/10.1137/1.9780898719284>.
- Nguyen, T.M., Wu, Q.J. (2013). Fast and robust spatially constrained Gaussian mixture model for image segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(4), 621–635. <https://doi.org/10.1109/TCSVT.2012.2211176>.
- O'Sullivan, F., Pawitan, Y. (1993). Multidimensional density estimation by tomography. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2), 509–521. <https://doi.org/10.1111/j.2517-6161.1993.tb01919.x>.
- Ralašić, I., Seršić, D., Petrinović, D. (2019). Off-the-shelf measurement setup for compressive imaging. *IEEE Transactions on Instrumentation and Measurement*, 68(2), 502–511. <https://doi.org/10.1109/TIM.2018.2847018>.
- Ralašić, I., Tafro, A., Seršić, D. (2018). Statistical compressive sensing for efficient signal reconstruction and classification. In: *2018 4th International Conference on Frontiers of Signal Processing (ICFSP)*, pp. 44–49. <https://doi.org/10.1109/ICFSP.2018.8552059>.
- Rani, M., Dhok, S., Deshmukh, R. (2018). A systematic review of compressive sensing: concepts, implementations and applications. *IEEE Access*, 6, 4875–4894. <https://doi.org/10.1109/ACCESS.2018.2793851>.
- Reader, A.J., Zaidi, H. (2007). Advances in PET image reconstruction. *PET Clinics*, 2(2), 173–190. PET Instrumentation and Quantification. <https://doi.org/10.1016/j.cpet.2007.08.001>.
- Reynolds, D. (2015). Gaussian mixture models. In: Li, S.Z., Jain, A.K. (Eds.), *Encyclopedia of Biometrics*. Springer, Boston, MA. [https://doi.org/10.1007/978-1-4899-7488-4\\_196](https://doi.org/10.1007/978-1-4899-7488-4_196).
- Reynolds, D.A., Quatieri, T.F., Dunn, R.B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10(1–3), 19–41. <https://doi.org/10.1006/dspr.1999.0361>.
- Rousseeuw, P.J., Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424), 1273–1283. <https://doi.org/10.1080/01621459.1993.10476408>.
- Shepp, L.A., Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2), 113–122. <https://doi.org/10.1109/TMI.1982.4307558>.



- Sović Kržić, A., Seršić, D. (2018).  $L_1$  minimization using recursive reduction of dimensionality. *Signal Processing*, 151, 119–129. <https://doi.org/10.1016/j.sigpro.2018.05.002>.
- Tafro, A., Seršić, D. (2019). Iterative algorithms for Gaussian mixture model estimation in 2D PET Imaging. In: *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 93–98. <https://doi.org/10.1109/ISPA.2019.8868570>.
- Tian, D. (2018). Gaussian mixture model and its applications in semantic image analysis. *Journal of Information Hiding and Multimedia Signal Processing*, 9(3), 703–715.
- Tong, S., Alessio, A.M., Kinahan, P.E. (2010). Image reconstruction for PET/CT scanners: past achievements and future challenges. *Imaging in Medicine*, 2.5, 529–545. <https://doi.org/10.2217/iim.10.49>.
- Xiong, T., Huang, Y. (2015). Robust Gaussian mixture modelling based on spatially constraints for image segmentation. *Journal of Information Hiding and Multimedia Signal Processing*, 6(5), 857–868.
- Yu, G., Sapiro, G. (2011). Statistical compressed sensing of Gaussian mixture models. *IEEE Transactions on Signal Processing*, 59(12), 5842–5858. <https://doi.org/10.1109/TSP.2011.2168521>.
- Zhang, J., Modestino, J.W., Langan, D.A. (1994). Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. *IEEE Transactions on Image Processing*, 3(4), 404–420. <https://doi.org/10.1109/83.298395>.
- Zhang, Y., Brady, M., Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57. <https://doi.org/10.1109/42.906424>.

**A. Tafro** is an assistant professor at University of Zagreb Faculty of Forestry and Wood Technology. She worked previously as a postdoctoral researcher at the Faculty of Electrical Engineering and Computing, Department of Electronic Systems and Information Processing, and at the Faculty of Science. Her research interests include statistical analysis and signal processing. Her publications can be found on <https://orcid.org/0000-0002-1083-7934>.

**D. Seršić** is a full professor at the University of Zagreb Faculty of Electrical Engineering and Computing. He received the Diploma, Master's, and a PhD degree in technical sciences, electrical engineering from the University of Zagreb, Zagreb, Croatia, in 1986, 1993, and 1999, respectively. Professor Seršić was a visiting researcher at the Technical University Vienna, Republic of Austria, in 1995, as well as at the Colorado State University, Fort Collins, Colorado, USA, in 2012. His current research interests include the theory and applications of wavelets, advanced signal and image processing, adaptive systems, blind source separation, and compressive sensing. Dr. Seršić is a member of the European Association for Signal Processing. From 2006 to 2008, he served as the chair of the Croatian IEEE Signal Processing Chapter. His publications can be found on <https://www.fer.unizg.hr/en/damir.sersic>.

**A. Sović Kržić** is an assistant professor at the University of Zagreb Faculty of Electrical Engineering and Computing, Croatia. Her research interests include signal processing, educational robotics, e-learning, and data science. She is the president of Croatian robotics association. She has received several awards, including two national awards of Republic of Croatia: Faust Vrancic National award for the technical culture in 2013 and National science award in 2019. Her publications can be found on [https://www.fer.unizg.hr/en/ana.sovic\\_krzic](https://www.fer.unizg.hr/en/ana.sovic_krzic).