Gaussian Mixture Model Estimation in 2D Pet Image Reconstruction

Azra Tafro [1,2] Damir Seršić[2]

[1]University of Zagreb Faculty of Forestry and Wood Technology
[2]University of Zagreb Faculty of Electrical Engineering and Computing
azra.tafro@sumfak.unizg.hr

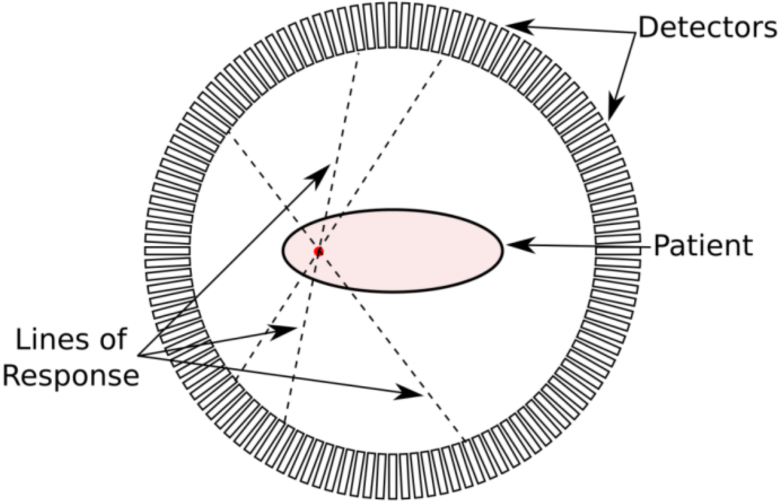
October 16th, 2021

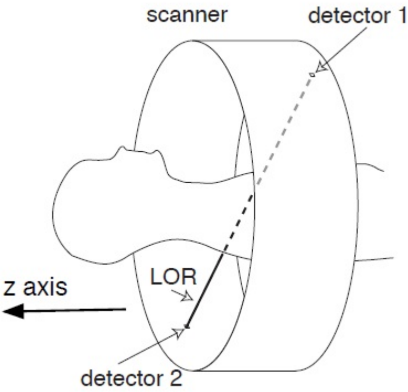
Outline

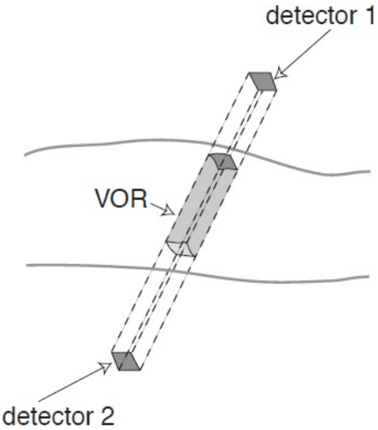
1. Introduction
   - PET Images and Reconstruction
   - Gaussian Mixture Models (GMMs)
2. Estimation of Gaussian Parameters
   - One Component Parameter Estimation
   - Iterative Algorithms for Mixtures
3. Experiments on Simulated Measurements
4. Conclusion

Original Image

- Inject living tissue with radioactive substance (tracer)
- Decaying tracer produces pairs of annihilation photons traveling in opposite directions
- detector elements (crystals) placed around the object detect lines - coincidence events
- The scanner can be
  - 3D: a tube joining two detector elements is a *volume of response* (VOR)
  - 2D: the line connecting the detector elements is a *line of response* (LOR)
- Data are recorded as event histograms (sinograms or projected data) or as a list of recorded photon-pair events (list-mode data)


Detectors
Patient
Lines of
Response

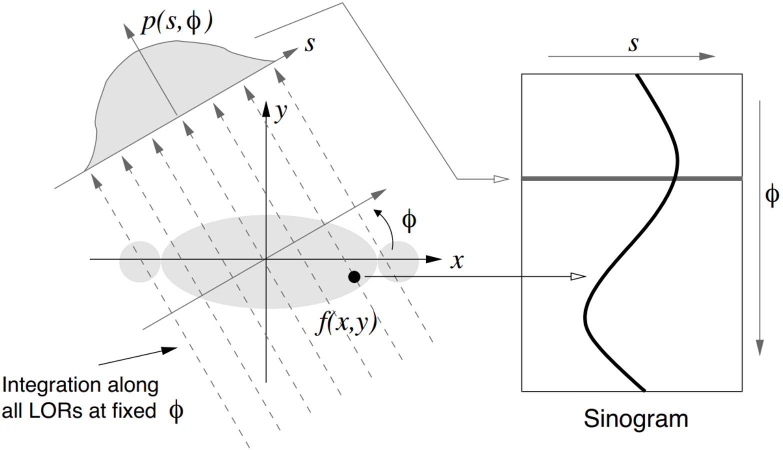

scanner
detector 1
LOR
z axis
detector 2


detector 1
VOR
detector 2

Sinogram (2D)

- Events along lines of response
- Integrate object activity distribution $f(x, y)$ along all parallel LORs at angle ϕ for $0 \leq \phi < 2\pi$
- $f(x, y) \mapsto p(s, \phi)$ where s is the distance from the center of the field of view
- A fixed point traces a sinusoidal path in the projection space
- Sinogram = the superposition of all sine waves for each point of activity

Sinogram Illustrated



A. Alessio, P. Kinahan, “PET Image Reconstruction,” Nuclear medicine, vol. 1, pp. 1–22, 2006.

Image Reconstruction

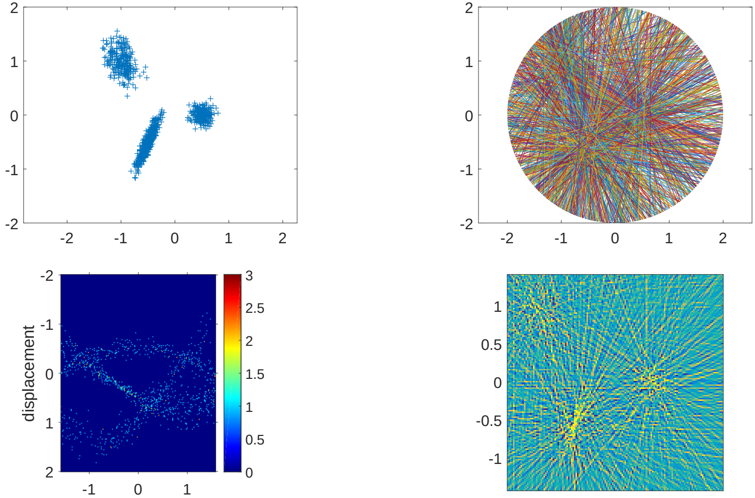
Filtered backprojection

- Based on the Fourier-slice (projection-slice) theorem
- Projection data are filtered (pre-corrected for the oversampling of the Fourier transform), then backprojected, and then inverse Fourier transformed
- Analytic and fast, but sensitive to noise and errors

Iterative algorithms

- Image is discretized into distinct pixels (voxels) which are then modeled
- Expectation-maximization algorithms: MLEM, OSEM
- More precise, computationally complex, also require some denoising

- $N = 1000$ events and $K = 3$ components: original (unknown) points, measurements, sinogram, FBP (simplified)



Gaussian Mixture Models (GMMs)

Used in a wide variety of image classification and reconstruction problems.

- Each component density is an d-variate Gaussian function:

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

- Observation $\boldsymbol{x}$ is a realization from exactly one of the K Gaussian mixture components

$$p(\boldsymbol{x}|\tau_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{k=1}^{K} \tau_k \; g(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (2)$$

- $\{\tau_k\}$ = mixture weights, i.e. the probabilities that $\boldsymbol{x}$ belongs to corresponding Gaussian components; $\sum_{k=1}^{K} \tau_k = 1$.

Estimating Mixture Parameters

- Existing algorithms (MLEM, OSEM) do not consider spatial dependence of pixels/voxels, or introduce it later (e.g. by using Markov random fields)
- "Holistic" approach to modelling the source of emissions - intensity is proportional to normal distribution density (or their mixtures)
- Main issues:
  - emission sources are unknown (latent), and the observations (lines) are lower-dimensional than the source
  - simplifying and accelerating estimation algorithms - faster scanning, less time in the scanner and less exposure to radiation

$K = 1$, Mean Vector Estimate

$\hat{\boldsymbol{\mu}} =$ the point "nearest" to all events

- We define distance using a weight matrix $\boldsymbol{W}$:

$$d^2(\boldsymbol{v}_1, \boldsymbol{v}_2) = (\boldsymbol{v}_1 - \boldsymbol{v}_2)^T \boldsymbol{W} (\boldsymbol{v}_1 - \boldsymbol{v}_2).$$

  $\boldsymbol{W} = \boldsymbol{I}$ gives Euclidian, $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}$ gives Mahalanobis distance.
- For a given $\boldsymbol{\mu}$, denote by $\boldsymbol{x}_i^{\mu}$ the point on ith line nearest to it
- $\hat{\boldsymbol{\mu}}$ is the solution of

$$\min_{\boldsymbol{\mu}} \sum_{i=1}^{N} (\boldsymbol{x}_i^{\mu} - \boldsymbol{\mu})^T \boldsymbol{W} (\boldsymbol{x}_i^{\mu} - \boldsymbol{\mu}).$$

Note: d vs. d^2 and $\boldsymbol{I}$ vs. $\boldsymbol{\Sigma}^{-1}$ yield very similar (good) results!

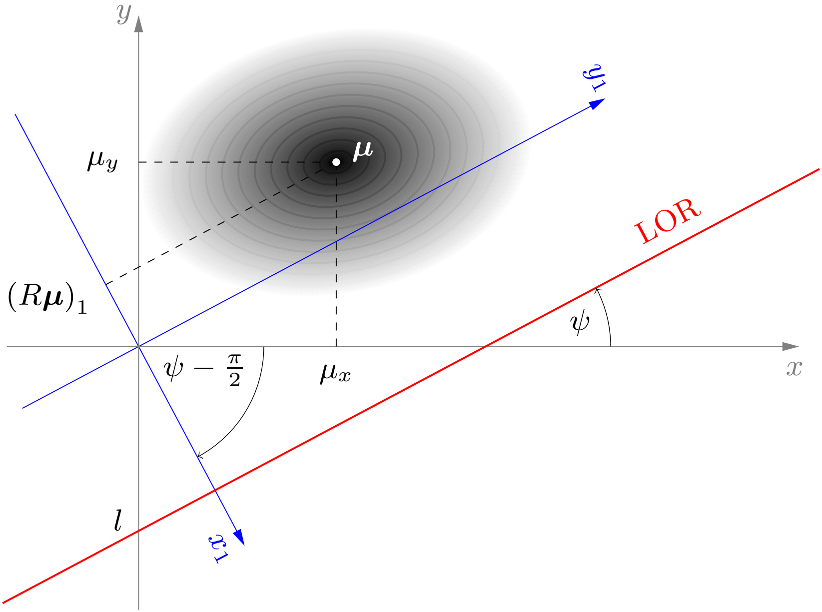
$K = 1$, Covariance Matrix Estimate

Suppose $d = 2$.

$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} \Rightarrow \text{ estimate } \Sigma_{11}, \Sigma_{12}, \Sigma_{22}.$$

- A Gaussian distribution retains properties when rotated
- Marginal distributions of a Gaussian are again Gaussian
- If an event is at angle ψ, rotating the coordinate system by $\varphi = \frac{\pi}{2} - \psi$ makes it parallel to the y-axis
- Integral along line = 1D projection onto the new x-axis

$$(\boldsymbol{\mu}, \mathbf{\Sigma}) \mapsto (R\boldsymbol{\mu}, R\mathbf{\Sigma}R^T) \mapsto ((R\boldsymbol{\mu})_1, (R\mathbf{\Sigma}R^T)_{11})$$



y
y_1
$\boldsymbol{\mu}$
μ_y
LOR
$(R\boldsymbol{\mu})_1$
ψ
$\psi - \frac{\pi}{2}$
μ_x
x
l
x_1

$K = 1$, Covariance Matrix Estimate

- Lines are given by $\boldsymbol{a}_i^T \boldsymbol{x} + l_i = 0\,,\ \boldsymbol{a}_i = [\tan\psi_i\ \ -1]^T$, $i = 1, \ldots, N$.
- New x-coordinate of each line is $-l_i \sin\varphi_i$.

One-dimensional mean and variance are

$$(R\boldsymbol{\mu})_1 = \cos\varphi\mu_x - \sin\varphi\mu_y,$$
$$(R\boldsymbol{\Sigma}R^T)_{11} = \cos^2\varphi\Sigma_{11} - 2\cos\varphi\sin\varphi\Sigma_{12} + \sin^2\varphi\Sigma_{22}.$$

Each line gives a 1D projection whose squared (Euclidian) distance from the mean, $(\cos\varphi\mu_x - \sin\varphi\mu_y + l\sin\varphi)^2$ is used to estimate the variance.

Solve $\boldsymbol{A}\boldsymbol{s} = \boldsymbol{b}$, where $\boldsymbol{s} = [\Sigma_{11},\ \Sigma_{12}, \Sigma_{22}]^T$.

$$A = \begin{bmatrix} \cos^2\varphi_1 & -2\sin\varphi_1\cos\varphi_1 & \sin^2\varphi_1 \\ \vdots & \vdots & \vdots \\ \cos^2\varphi_N & -2\sin\varphi_1\cos\varphi_N & \sin^2\varphi_N \end{bmatrix},$$

$$\text{and } \boldsymbol{b} = \begin{bmatrix} (\cos\varphi_1\mu_x - \sin\varphi_1\mu_y + l_1\sin\varphi_1)^2 \\ \vdots \\ (\cos\varphi_N\mu_x - \sin\varphi_N\mu_y + l_N\sin\varphi_N)^2 \end{bmatrix}.$$

Covariance Matrix Estimate, $K = 1$

- Overdetermined system: solve $\min_{\boldsymbol{s}} \|\boldsymbol{A}\boldsymbol{s} - \boldsymbol{b}\|$.
- L_1 minimization preferred to L_2 minimization (more robust and resistant to gross and systematic errors)[1]
- One-dimensional variance estimates are from single (or at most several) points for each φ - a corrective factor is needed.

$$\min_{\boldsymbol{s}} \|\boldsymbol{A}\boldsymbol{s} - \lambda\boldsymbol{b}\|_1,$$

where $\lambda = (\Phi(0.75))^2 \approx 1.4826^2$.

[1] L_1 minimization algorithm proposed by A. Sović Kržić and D.S. (2018).

Expectation - Maximization Algorithm

- E step:
  - Given the current estimate of parameters, create a function for the expectation of the log-likelihood function
  - GMM: assign each data point its membership probabilities
- M step:
  - Compute parameters that maximize the function from the E step
  - GMM: estimate parameters of each component using points "belonging" to that component.
- Start from initial parameters (E step) or initial weights (M step)

Observations are lines, there are no (proper) max-likelihood estimators!

- STEP 1 (expectation - like)
  - Probabilistic approach: weights are calculated using Gaussian densities from previous iterations.
  - Geometric approach: weights are calculated using only geometric properties of lines (inversely proportional to distance from previously estimated mean).
- STEP 2 (maximization - like)
  - Soft classification: all lines participate in estimation of all components, proportional to weight.
  - Hard classification: lines assigned to the most likely component and participate only there.

Simulated Measurements

- $K = 2$ $N = 4000$ ($n_1 = 2500$, $n_2 = 1500$), 1000 iterations.

$$\boldsymbol{\mu}_1 = \begin{bmatrix} -0.05 \\ -0.05 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.01 & 0.02 \\ 0.02 & 0.05 \end{bmatrix},$$
$$\boldsymbol{\mu}_2 = \begin{bmatrix} 0.05 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.02 & -0.01 \\ -0.01 & 0.05 \end{bmatrix}.$$

- Distributions are presented as images (color intensity corresponds to density)
- Images are compared using the Structural Similarity Index (http://www.cns.nyu.edu/˜lcv/ssim/)

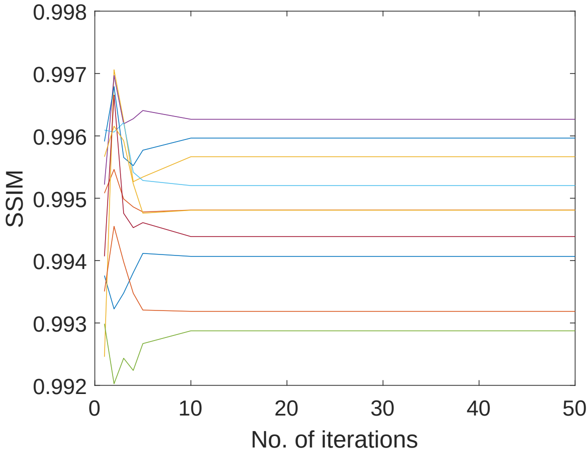
SSIM Results

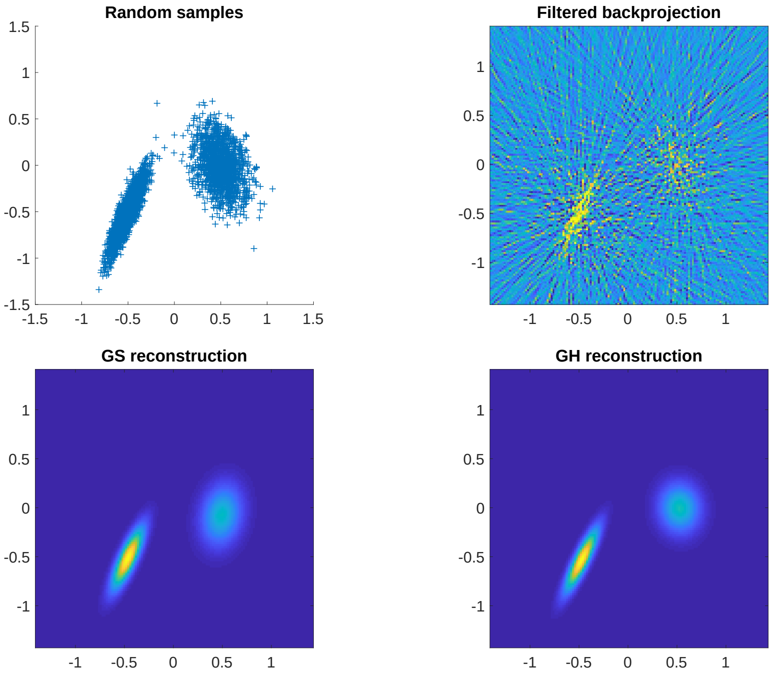
	min.	average	max.
GS	99.03%	99.33%	99.59%
GH	99.01%	99.46%	99.74%
PS (inc.)	93.15%	95.27%	96.44%
PH (inc.)	94.65%	95.87%	98.82%
FBP	95.81%	95.82%	95.82%

- Algorithm iterated $k = 10$ times, with distance d varying from Euclidian to Mahalanobis.
- Probabilistic algorithms appeared unstable (near null covariance matrices), most likely due to poor initial values.

GH Algorithm, Additional Experiments

- Geometric form of the algorithm is robust regardless of initial parameters.
- Accuracy stabilizes after approximately 10 iterations.




Random samples
Filtered backprojection
GS reconstruction
GH reconstruction
1.5
1
0.5
0
-0.5
-1
-1.5

Alternative models

Observe a sample of specific related measurements:

- line intersections, or
- points (on the lines) nearest to the center(s)

Progress:

- Both sets of points provide (using corrective factors) unbiased estimators for the single component covariance
- Line intersections present a computationally complex problem - N sources induce $\binom{N}{2}$ intersections
- Nearest points give less precise estimates, i.e. more measurements are needed
- Classification problems in multiple components scenarios (as of now)

Conclusion

Work in progress:

- Extension to 3D
- Optimal initial values
- Application to real data - detection of K, attenuation, random events etc.
- Other distributions with suitable properties

Advantages:

- Parametric model (sparse representation)
- Resistance to noise (no need for post-processing)
- Reconstruction from fewer measurements (less exposure)

References

- A. Tafro, D. Seršić, A. Sović Kržić: *2D PET Image Reconstruction Using Robust L_1 Estimation of the Gaussian Mixture Model*, preprint arXiv:1906.06961.
- A. Tafro, D. Seršić: *Iterative algorithms for Gaussian Mixture Model Estimation in 2D PET Imaging*, 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019.
- T. Matulić, R. Bagarić, D. Seršić: *Enhanced reconstruction for PET scanner with a narrow field of view by using backprojection method*, 44th International Convention on Information, Communication and Electronic Technology (MIPRO), 2021.

Thank you for your attention!

Joint work with Damir Seršić, Tomislav Matulić (UNIZG-FER) and Robert Bagarić (IRB).
Research is fully supported by the Croatian Science Foundation under the project "Sampling Renaissance" IP-2019-04-6703.