MCMC, FFT, Point Data Structures (20 pts)

This homework is due on Dec.16. Please submit your homework in pdf form, which can be either a scanned copy of your hand-written answers, or, computer generated documents (e.g., via word/latex etc). In either case, please give a clear description about your method and include computer calculated results and figures.

Please use Random Number Generator to generate three sets of bivariate Gaussian distributions on the $X - Y$ plane. The three components have standard deviations of 0.2, 0.3 and 0.5, and realized with 1000, 800 and 500 particles, respectively. The centres of these bivariate Gaussians are randomly located within a square region of 1×1.

(1) Introduce a $2D$ uniform mesh (of a reasonable size and spatial resolution) to cover the majority of the particle region, assign the particles to this mesh using the Cloud-In-Cell method. Make sure all three Gaussian components can be well resolved by your mesh grid. Through FFT calculate the clustering of your data points on scales larger than the mesh cell size and express your results as the correlation function of the density field. Plot both the correlation function $\xi(r)$ (in real space) and the power spectrum $P(k)$ (in frequencey space).

(2) Use the Pair-Estimation method with the Landy & Szalay estimator (1993) to calculate the two-point correlation function of the density field. Compare the differences between this result and the FFT correlation from (1), plot them in the same figure. Can you explain what you see?

(3) Now choose any two Gaussian components out of the three and project their $2D$ distributions on to the same X-axis. This data set will then be fitted by a Gaussian mixture model with two components. Write out the associated Bayesian formular for the prior, the joint data likelihood and the posterier probability distribution function. Apply Markov Chain Monte Carlo to sample this posterior distribution and plot the projections of the 5-dimensional posterior pdf (on to the planes, each one spanned by two model parameters) as well as the marginalized distribution for each parameter. In the same plot, please also indicate the prefixed parameters of the Gaussian components.

(4) Now suppose you do not know the generative model for the $2D$ distribution. Apply the Gaussian mixture model method to estimate the underlying density field. Use Information Criteria or any cross-validation method, estimate how many Gaussian components best describe the particle distribution? Explain how you reach this number and label out the different GMM components with circles (ellipses) on top of the particle distribution.