2010 Ninth International Conference on Machine Learning and Applications

Using an infinite von Mises-Fisher Mixture Model to Cluster Treatment Beam Directions in External Radiation Therapy

Mark Bangert
German Cancer Research Center
Heidelberg, Germany
m.bangert@dkfz.de

Philipp Hennig
Cavendish Laboratory
Cambridge, United Kingdom
ph347@cam.ac.uk

Uwe Oelfke
German Cancer Research Center
Heidelberg, Germany
u.oelfke@dkfz.de

***Abstract* —We present a method for fully automated selection of treatment beam ensembles for external radiation therapy.**

We reformulate the beam angle selection problem as a clustering problem of locally ideal beam orientations distributed on the unit sphere. For this purpose we construct an infinite mixture of von Mises-Fisher distributions, which is suited in general for density estimation from data on the D-dimensional sphere. Using a nonparametric Dirichlet process prior, our model infers probability distributions over both the number of clusters and their parameter values. We describe an efficient Markov chain Monte Carlo inference algorithm for posterior inference from experimental data in this model.

The performance of the suggested beam angle selection framework is illustrated for one intra-cranial, pancreas, and prostate case each. The infinite von Mises-Fisher mixture model (iMFMM) creates between 18 and 32 clusters, depending on the patient anatomy. This suggests to use the iMFMM directly for beam ensemble selection in robotic radiosurgery, or to generate low-dimensional input for both subsequent optimization of trajectories for arc therapy and beam ensemble selection for conventional radiation therapy.

***Keywords* - Nonparametric Bayesian Inference; Directional Statistics; Radiation Therapy; Treatment Planning; Beam Angle Optimization.**

I. Introduction

Radiation therapy aims to maximize the tumor control probability while minimizing the normal tissue complication rate. For radiation therapy treatment planning, these two conflicting clinical objectives are often translated into an objective function F through the following definition:

$$\begin{aligned} F &= \sum_i p_i \ \left\{D_i - D_i^{\text{pres}}\right\}^2 \\ &= \sum_i p_i \ \left\{\Sigma_j(w_j D_{ij}) - D_i^{\text{pres}}\right\}^2 . \end{aligned} \tag{1}$$

p_i, D_i, and D_i^{pres} denote the penalty, the actual dose, and the prescribed dose for voxel i of the discretized patient anatomy. D_i is given by a weighted linear superposition of multiple beamlets j. The dose influence matrix D_{ij} specifies the dose contribution to voxel i from beamlet j. Minimizing F with respect to the beamlet weights w_j is a convex problem. It can be solved efficiently with standard optimization techniques. The D_{ij} matrix, however, is a function of the individual treatment beams β_η constituting the treatment beam ensemble B.

$$D_{ij} = D_{ij}(B) \tag{2}$$

Hence, the overall optimization problem of finding ideal machine parameters for radiation therapy has to be defined as

$$\underset{B,\boldsymbol{w}}{\operatorname{argmin}} \left\{ \sum_i p_i \{ \ \Sigma_j(w_j D_{ij}(B)) - D_i^{pres} \ \}^2 \right\}. \tag{3}$$

The computation of the dose influence matrix D_{ij}, i.e. the simulation of the radiation transport on the patient anatomy, for a beam configuration B is computationally intensive. It complicates the optimization of beam angles, which itself is a non convex combinatoric problem with exponential complexity [3].

This is why, besides global optimization techniques [12][13], heuristic beam angle optimization (BAO) strategies receive constant scientific attention. Different scoring functions based on both geometric [9] and dosimetric considerations [6][10] have been suggested to facilitate the selection of a treatment beam ensemble. We have previously reported on a strategy to select a beam ensemble, given the number of beams K, that can be adopted to work for any scalar scoring function S [1]. It is based on the calculation of a set of locally ideal irradiation angles $\mathcal{B}^*$ which is parametrized as a set of points on the unit sphere. The spherical distribution of $\mathcal{B}^*$ is characteristic for every site and patient. We demonstrated that interpreting its centers of gravity as beam orientations for radiation therapy yields a significant improvement of the clinical dose distributions for intensity modulated radiation therapy [1].

In this paper we introduce a more sophisticated clustering algorithm to analyze the spherical data. It sidesteps the delicate question of finding the right number of clusters by replacing the point estimate of one particular number with a probabilistic belief over the total number of clusters. By adapting previous work on infinite Gaussian Mixture

978-0-7695-4300-0/10 $26.00 © 2010 IEEE
DOI 10.1109/ICMLA.2010.114

Models (iGMM) [11] in Cartesian spaces to the D-Sphere, we develop the infinite von Mises-Fisher mixture model (iMFMM). The iMFMM is suited for general density estimation and clustering of D-dimensional spherical data. We illustrate the potential application of the iMFMM for radiation therapy treatment planning with three clinical cases.

II. Method

A. Reformulation of beam angle optimization as a spherical clustering problem

The presented methodolgy is flexible enough to work with any score function projecting the beam angle optimization problem onto the 3-sphere. Here, we apply the heuristic scalar score $S_{\boldsymbol{\beta}v}$ of voxel v for irradiation from direction $\boldsymbol{\beta}$ [1], which is defined as

$$S_{\boldsymbol{\beta}v} = \frac{d_{\text{NT}} + 100 \cdot d_{\text{OAR}}}{d_{\text{Target}}}. \tag{4}$$

d_{Target}, d_{NT}, and d_{OAR} denote the doses delivered to the target volume, to the normal tissue and to potential organs at risk (OAR). When calculating $S_{\boldsymbol{\beta}v}$ for a set of candidate directions, it is assumed that voxel v is irradiated only with a narrow pencil beam from direction $\boldsymbol{\beta}$ - not the entire target volume with a broad beam. By weighting dose contributions to potential OARs hundredfold, $S_{\boldsymbol{\beta}v}$ implies a twofold nature of the beam selection problem: In absence of OARs, the ideal irradiation angle $\boldsymbol{\beta}^*$ minimizes the ratio $d_{\text{NT}}/d^{\text{Target}}$; in presence of OARs, $\boldsymbol{\beta}^*$ will try to avoid traversing OARs and minimize $d_{\text{OAR}}/d_{\text{Target}}$.

By identifying the ideal beam direction $\boldsymbol{\beta}_v^*$ for all voxels within the target, we obtain a set of locally ideal beam directions $\mathcal{B}^*$, which is parametrized as a set of points $\{\boldsymbol{x}_i\}$ on the 3-sphere. Figure 1 shows a Mollweide projection of the three-dimensional data set $\mathcal{B}^*$ to two dimensions for an intra-cranial lesion. The central idea of spherical cluster analysis for BAO is to identify the cluster centers of the patient specific distributions $\mathcal{B}^*$ and interpret those as beam orientations for external radiation therapy.

B. The finite von Mises-Fisher mixture model

The von Mises-Fisher distribution is a spherical analogon to an uncorrelated multivariate Gaussian distribution in Cartesian space. On the D-sphere, it is defined as

$$\mathcal{F}(\boldsymbol{x}; \boldsymbol{\mu}, \tau) = \frac{\tau^{D/2-1}}{(2\pi)^{D/2} I_{D/2-1}(\tau)} \exp(\tau \boldsymbol{\mu}^T \boldsymbol{x}) \tag{5}$$

with the scalar precision parameter τ and the mean direction $\boldsymbol{\mu}$. I_ν is the modified Bessel function of the first kind and order ν. For $D = 3$, we obtain the special form

$$\mathcal{F}(\boldsymbol{x}; \boldsymbol{\mu}, \tau) = \frac{\tau}{4\pi \sinh(\tau)} \exp(\tau \boldsymbol{\mu}^T \boldsymbol{x}) \tag{6}$$

In the following, we concentrate on the 3-dimensional case, but all derivations can be easily extended to D dimensions.

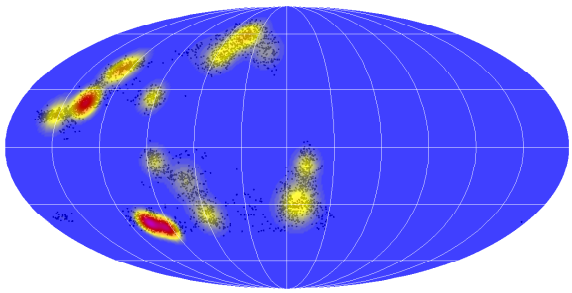


Figure 1. Mollweide projection of a set of ideal beam angles $\mathcal{B}^*$ for an intra cranial lesion. The longitude corresponds to the angle around the patient axis, the latitude to the angle towards the transversal plane. The data, indicated by black dots, is overlaid by a density estimate of the iMFMM averaged over 1000 Monte Carlo iterations.

To construct our model, we assume that the N samples $\boldsymbol{x}_i$ are each generated from a mixture of an unknown and unbounded number K of independent von Mises-Fisher distributions with unknown parameters $\boldsymbol{\mu}_k, \tau_k$.

$$p(\{\boldsymbol{x}_i\}; \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K, \tau_1, ..., \tau_K, \pi_1, ..., \pi_K) = \sum_{k=1}^{K} \pi_k \prod_{i=1}^{N} \mathcal{F}(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \tau_k) \tag{7}$$

$\pi_k = p(c_i = k)$ is the probability of sample i stemming from cluster k (and c_i indicates the assignment of sample i to a cluster k).

In this section, we will assume K to be fixed and finite; the limit $K \to \infty$ is introduced in section II-C. The derivations are closely modeled on work by Rasmussen [11] and Neal [8] with regard to Gaussian Mixture Models. Our contribution is the transformation of their work from Cartesian space to the sphere (i.e. from mixtures of Gaussians to mixtures of von Mises-Fisher distributions).

Given a prior $p(\boldsymbol{\mu}_{1...K}, \tau_{1...K}, c_{1...N})$ on the unknown parameter values, the goal of an inference algorithm on such a model is to track a posterior belief $p(\boldsymbol{\mu}_{1...K}, \tau_{1...K}, c_{1...N} | \boldsymbol{x}_{1...N})$ over the parameters of the mixture model given the observed data. In our implementation, inference is performed using Gibbs sampling [4], a widely used Markov chain Monte Carlo scheme. It consists of iteratively sampling values of all parameters of the model individually, conditioned on the current samples from all other parameters. Gibbs sampling is guaranteed to produce samples from the exact posterior in the limit of large numbers of sampling steps.

To keep the computational cost of the sampling scheme manageable, we use conjugate priors for the parameters of the mixture components. A parametric distribution

$$p(z|\boldsymbol{a}) = f(z; \boldsymbol{a}) \tag{8}$$

on the variable z with parameters $\boldsymbol{a}$ is called a conjugate prior to a likelihood $p(d|z)$ of z under the data d if the

posterior can be formulated (using Bayes' rule) in the exact parametric form of the prior.

$$p(z|d,\boldsymbol{a}) = \frac{p(d|z)p(z|a)}{\int p(d|z)p(z|a)dz} = f(z;\boldsymbol{a}') \tag{9}$$

The von Mises-Fisher distribution forms an exponential family [2]. All distributions forming exponential families have conjugate priors for their parameters, and a general construction for these priors exists.

1) Conjugate prior for $\boldsymbol{\mu}_k$ given τ_k:
For the mean parameter $\boldsymbol{\mu}_k$, the conjugate prior is itself a von Mises-Fisher distribution

$$g(\boldsymbol{\mu}_k;\tau_k,\boldsymbol{x}_0) = \mathcal{F}(\boldsymbol{\mu}_k;\boldsymbol{m}_0,t_0) \tag{10}$$

with two parameters $\boldsymbol{m_0}$ and t_0. In our experiments, we set $t_0 = 0.1$ which leads to a distribution so broadly covering the sphere that the precise value of $\boldsymbol{m_0}$ becomes irrelevant (for the lack of a better option, it was set to the mean of the entire dataset). The update rule for the posterior given data is

$$\begin{aligned} p(\boldsymbol{\mu}_k|\left\{\boldsymbol{x}_{i\in k}\right\},\tau_k) &= \mathcal{F}(\boldsymbol{\mu}_k;\boldsymbol{m}_0,t_0)\cdot\prod_{i\in k}\mathcal{F}(\boldsymbol{x}_i;\boldsymbol{\mu}_k,\tau_k) \\ &= \mathcal{F}(\boldsymbol{\mu}_k;\boldsymbol{\xi}/|\boldsymbol{\xi}|,|\boldsymbol{\xi}|). \end{aligned} \tag{11}$$

with $\boldsymbol{\xi} = t_0\boldsymbol{m}_0 + \tau_k\sum_{i\in k}\boldsymbol{x}_i$. The notation $i \in k$ confines an operation to samples that are associated with cluster j.

2) Conjugate prior for τ_k given $\boldsymbol{\mu}_k$:
Up to normalization, the conjugate prior on the precision parameter τ_k is

$$f(\tau_k;a,b) \propto \left\{\frac{\tau_k}{4\pi\sinh(\tau_k)}\right\}^a \exp\left(\tau_k b\right) \tag{12}$$

with scalar parameters $a > b > 0$. We set $a = 5.0$ and $b = 4.7$ yielding a realistic initial distribution of the precision parameters τ_k for our data. The update rule for the posterior given data and $\boldsymbol{\mu}_k$ is

$$\begin{aligned} p(\tau_k;\left\{\boldsymbol{x}_{i\in k}\right\},\boldsymbol{\mu}_k) &\propto f(\tau_k;a,b)\prod_{i\in k}\mathcal{F}(\boldsymbol{x}_i;\boldsymbol{\mu}_k,\tau_k) \\ &\propto f(\tau_k;a+N_K,b+\sum_{i\in k}\boldsymbol{\mu}_k^T\boldsymbol{x}_i) \end{aligned} \tag{13}$$

where N_k is the number of members of cluster k. We are not aware of an efficient method to analytically obtain samples from this distribution, but any one-dimensional Markov chain Monte Carlo method can be used to produce samples from this marginal. In our implementation, we use the slice sampling algorithm [7], which is a particularly efficient Markov chain Monte Carlo method for one-dimensional distributions.

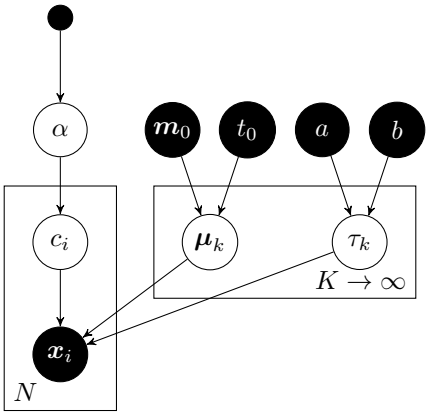


Figure 2. Directed graphical model (Bayesian network) of the iMFMM representating both the generative process used to model the data, and the factorization properties of the joint distribution of all variables in the model. Any node in the graph is conditionally independent of the rest of the graph given values of its parents, its children and parents of its children. The deterministic quantities, such as the data $\boldsymbol{x}_i$ and the hyperparameters a, b, $\boldsymbol{m}_0$ and t_0 are depicted by filled circles, while probabilistic (latent) parameters are shown as hollow circles. The inverse Gamma prior on α is shown as a small black circle. Rectangles with label N and K are so-called "plates" representing N and K copies of their contents.

3) Prior on the mixing proportions π_k:
The joint probability $p(c_1,c_2,...c_N)$ of the class memberships of the samples $\boldsymbol{x}_i$ is a multinomial distribution parameterized by the unknown mixture parameters π_k:

$$p(\boldsymbol{c}|\pi_k) = \prod_{k=1}^{K}\pi_k^{N_k} \tag{14}$$

The multinomial distribution is also a member of the exponential family, and the conjugate prior for its parameter vector $\boldsymbol{\pi}$ is the Dirichlet distribution with a K-dimensional parameter vector $\boldsymbol{\alpha}$. If we set all elements $\alpha_k = \alpha/K$ with a scalar constant α, the Dirichlet distribution puts uniform probability mass on all possible values of π_k and has the form

$$\mathcal{D}(\pi;\alpha) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/K)^K}\prod_{k=1}^{K}\pi_k^{\alpha/K-1}. \tag{15}$$

It is a crucial characteristic of the Dirichlet distribution that it is possible to integrate out the values of π_k under the posterior [11], leading to a joint distribution for the c_i which is only a function of α, K, and the cluster sizes N_k. It does not depend on the individual values of c_i:

$$p(c_1,...,c_N;\alpha) = \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)}\prod_{k=1}^{K}\frac{\Gamma(N_k+\alpha/K)}{\Gamma(\alpha/K)} \tag{16}$$

During Gibbs sampling, we condition on all but one particular sample. In this case the Gamma functions cancel and we arrive at the simple discrete conditional probability:

$$p(c_i = k|c_{\setminus i},\alpha) = \begin{cases} \frac{n_{\setminus i,k}}{N_k-1+\alpha}\ , & n_{\setminus i,k} > 0 \\ \frac{\alpha/K}{N_k-1+\alpha}\ , & \text{else} \end{cases} \tag{17}$$

where $c_{\setminus i}$ indicates all indices except i and $n_{\setminus i,j}$ is the number of observations, excluding $\boldsymbol{x}_i$, that are associated with cluster k.

Also conditioning on the value of sample x_i yields the Gibbs sampling probability for c_i of:

$$p(c_i = k; \boldsymbol{c}_{\setminus i}, \alpha)\ p(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \tau_k, \boldsymbol{c}_{\setminus i}) = \frac{n_{\setminus i,k}}{N_k - 1 + \alpha}\ \mathcal{F}(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \tau_k) \tag{18}$$

The likelihood for α itself can be derived from equation 16. Together with a prior of inverse Gamma shape [11], the posterior update for α is

$$p(\alpha; n, N) = \frac{\alpha^{n-3/2} \exp(-\frac{1}{2\alpha})\Gamma(\alpha)}{\Gamma(N + \alpha)} \tag{19}$$

As there is no efficient analytical sampling scheme for this distribution, we apply the slice sampling algorithm to obtain updates for α [7].

C. The infinite limit

So far, we have assumed a constant number of clusters. But, sidestepping a few technicalities [5][8][11], it is intuitively easy to take the limit of equation 18 for $K \to \infty$. All clusters containing more than one sample, i.e. $n_{\setminus i,k} > 0$, retain a finite probability $n_{\setminus i,j}/(N_k - 1 + \alpha) \cdot p(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \tau_k)$ of being chosen. And because the overall probability of choosing any cluster has to be 1 and all clusters have parameters with the same prior distribution $p(\boldsymbol{\mu}, \tau)$, all infinitely many remaining clusters together have the finite probability

$$\frac{\alpha}{N_k - 1 + \alpha} \int p(\boldsymbol{x}_i|\boldsymbol{\mu}, \tau) \cdot p(\boldsymbol{\mu}, \tau)\ d\boldsymbol{\mu}\ d\tau \tag{20}$$

of being chosen. The integral in this equation can be approximated by Monte Carlo integration, i.e. generating l samples from the prior for $\boldsymbol{\mu}$ and τ, summing their likelihood terms $p(\boldsymbol{x}_i; \boldsymbol{\mu}, \tau)$, and dividing by l. We found that often even $l = 1$ is sufficient for convergence of the Gibbs sampler.

The resulting probability measure over probability measures, widely known as the "Dirichlet process", controls the number of mixture components. The limit-construction used here is known as the "Chinese Restaurant Process" [5].

III. Results

For inference on our spherical beam angle data set, we initialize the iMFMM with K_0 mixture components. K_0 is sampled uniformly from the interval $[10, 20]$. The corresponding starting parameters for $\boldsymbol{\mu}_k$ and τ_k are drawn from the prior and posterior updates on all model parameters are performed during a large number of Gibbs sampling iterations. The initialization of K_0 reflects our expectation regarding the number of classes underlying the beam angle data set. It reduces the number of iterations needed for burn-in and leaves the long term behavior of the iMFMM unaltered. Here, "burn-in" denotes the drift from the initial Markov chain state to regions of high probability mass.

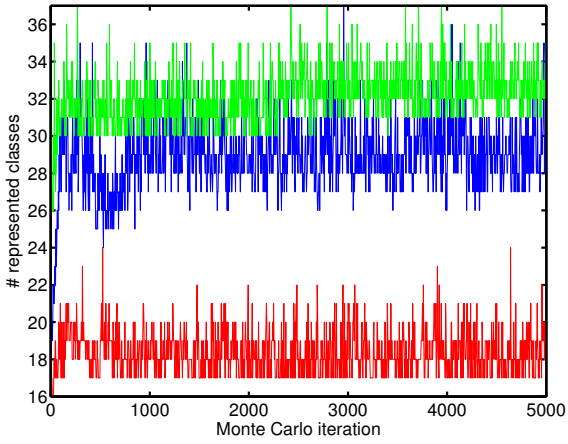
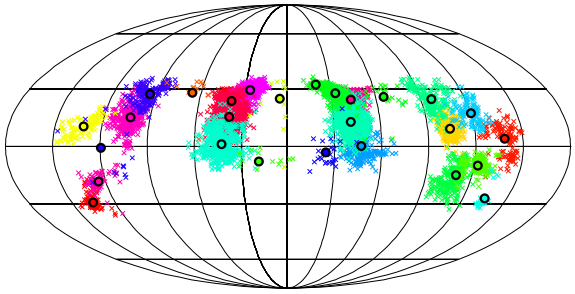


Figure 3. Number of represented classes for the initial 5000 iterations for an intra-cranial (red), pancreas (green), and prostate case (blue).

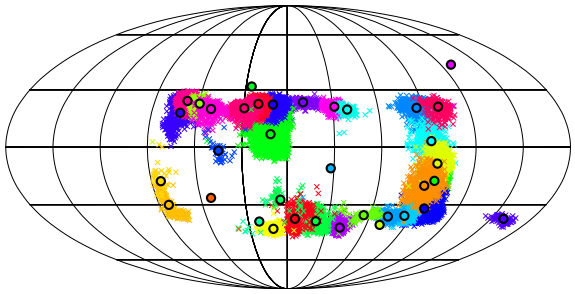
Figure 3 shows the number of represented classes during the first 5000 iterations for three patient data sets. The number of classes is rapidly adjusted by the iMFMM within the first 100 iterations and subsequently undergoes slight fluctuations for the intra-cranial and prostate data set. We decided to discard the first 2000 iterations for burn-in. The state of the iMFMM in a single iteration after burn-in (i.e. one iMFMM sample) is shown in figure 4 for every data set investigated. The full posterior distribution is formed by the set of all samples from the Gibbs scheme. Figure 1 shows an estimate for the mean of this distribution, obtained by averaging over a 1000 samples.

The autocorrelation of the number of represented classes after 2000 iterations is shown in figure 5. We do not observe a significant correlation for any of the three data sets under investigation. The effective autocorrelation length, computed as the sum of the autocorrelation between an iteration lag of -1000 and 1000 [11], does not exceed 10 iterations for the three data sets investigated.

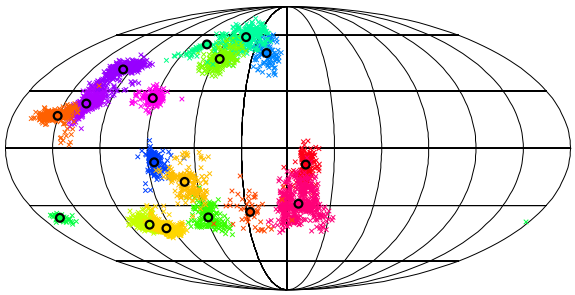
For a concrete statement regarding the number of represented classes (which in our case will correspond to the number of treatment beams) we now draw 100 independent iMFMM samples. More precisely, we evaluate 100 samples after burn-in which are each separated by one autocorrelation length. Figure 6 shows the spectra for K for the three data sets under investigation. As the data does not stem from a mixture of von Mises-Fisher distributions but from our formulation of the beam angle optimization problem, there is not a distinct number of components that explain our data. The iMFMM yields a distribution over the probability of the number of represented components. It found the highest probability for the intra-cranial data set to be generated by a mixture of $K^*_{\text{intra}} = 18$ von Mises-Fisher distributions. The analysis for the pancreas and prostate data sets yields highest



(a) Prostate lesion



(b) Pancreas lesion



(c) Intra-cranial lesion

Figure 4. One iMFMM sample for the spherical data sets of locally ideal beam angles of a prostate (a), pancreas (b), and intra-cranial lesion (c). The longitude corresponds to the angle around the patient axis, the latitude to the angle towards the transversal plane. The centroids of the mixtures $\boldsymbol{\mu}_k$ are indicated by black circles, the color coding corresponds to the current assignment of the data to the mixtures c_i. We did not attempt to visualize the precisions τ_k. Note the relatively large number of clusters required in 4(b), as the von Mises-Fisher distribution cannot model directional correlation. Considering 4(a) and 4(b) it might be straight forward to identify a smooth path on the sphere passing by the cluster centroids and respecting potential physical limitaions of the irradiation device for improved arc therpay.

probabilities for mixtures of $K^*_{\text{pancreas}} = 32$ and $K^*_{\text{prostate}} = 29$ von Mises-Fisher distributions, respectively. The number of samples used by the model might be a possible measure of the structural complexity of the treatment problem. A thorough investigation of the clinical significance of such a measure based on the iMFMM is left for future research.

Given the number of represented components K^*, it is straight forward to infer exact orientations for the treatment beams: Using a finite von Mises Fisher mixture model with K^* components, we maximize the posterior probability of the parameters $\boldsymbol{\mu}_k$, τ_k, and c_i. For the purpose of localizing

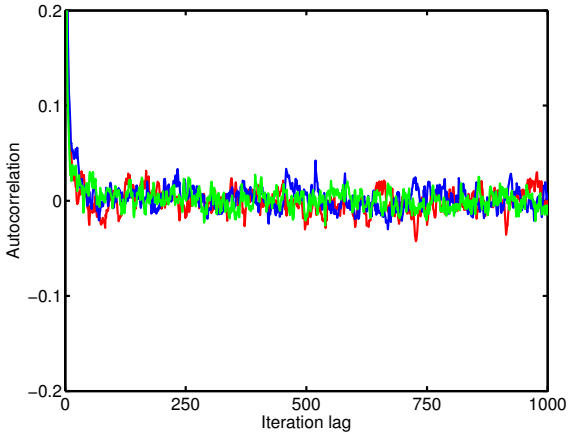


Figure 5. Autocorrelation of the number of represented classes after 2000 iterations for an intra-cranial (red), pancreas (green), and prostate case (blue).

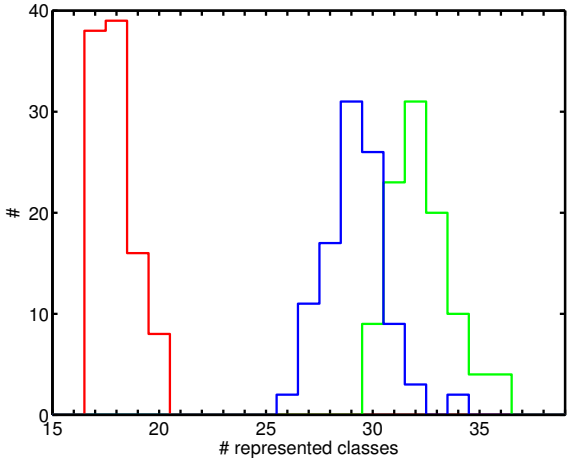


Figure 6. Histograms of the number of represented classes for an intra-cranial (red), pancreas (green), and prostate case (blue) displaying the frequency of the number of represented classes.

the treatment beams, only the centroid positions $\boldsymbol{\mu}_k$ will be of interest.

IV. Discussion and Conclusion

In extension of previous work on density estimation with Dirichlet process mixture models in Cartesian spaces [11][8], this paper introduces the infinite von Mises-Fisher mixture model as a general framework for density estimation on the D-sphere. We constructed conjugate priors for its parameters $\boldsymbol{\mu}$ and τ, and derived a Gibbs sampling scheme for posterior inference from data.

We apply the iMFMM to infer a treatment beam ensemble for external radiation therapy based on a set of locally ideal beam angles distributed on the unit sphere. For the data sets studied, the iMFMM returns mixtures of 18-32 beam orientations. This represents a considerable dimensionality

reduction from the original data sets containing $\sim 10^4$ locally ideal beam directions. For robotic radiosurgery treatments, where $\sim 50-100$ beam orientations are accessed, the iMFMM might be used without modification for beam selection. The iMFMM beliefs could also be used to improve trajectories for arc therapy, where the treatment beam is rotated around the patient during irradiation. In conventional linear accelerator treatments, which typically include $5-11$ beam orientations, the clusters found by the iMFMM can be used as a starting point for a global optimization algorithm. In either case, the strong dimensionality reduction achieved through the use of the nonparametric density estimation lowers the computational complexity. Alternatively, heuristic merging of clusters could be used in the search of optimal beam directions in order to arrive at a number of beams that is acceptable for conventional irradiation with a linear accelerator. Extending the infinite mixture model to integrate directional correlation on the sphere, which implies the non trivial transition from von Mises-Fisher distributions to Kent distributions, is another potential means to enhance the dimensionality reduction (consider figure 4(b)).

The focus of this paper is on the derivation of the iMFMM. A detailed assessment of the clinical impact of the suggested framework for beam angle selection is left for future research. However, previous work [1] has already provided strong evidence for the clinical value of the approximation to formulate the search for beneficial beam directions as a clustering problem of locally ideal beam angles on the 3-sphere. As the incorporation of the iMFMM into this framework implies only the replacement of the spherical K-means algorithm with a more sophisticated method, a negative impact regarding clinical performance is not expected.

It is clear that our model does not provide an optimal answer to the global optimization problem defined in equation 3, as it is not explicitly constructed to minimize this function. Instead, our model provides a probabilistically motivated projection of the locally ideal beam directions (as measured with a scalar scoring function) to a lower-dimensional space. In contrast to clustering methods based on point estimates, such as K-means clustering or factor analysis, the iMFMM, being a fully probabilistic method, does not suffer from the problem of overfitting, and can automatically determine the range of clusters to use.

ACKNOWLEDGEMENTS

The authors would like to thank Carl Scheffler for helpful discussions.

REFERENCES

[1] M. Bangert & U. Oelfke: *Spherical cluster analysis for beam angle optimization in intensity-modulated radiation therapy treatment planning*. Physics in Medicine and Biology 55, 2010.

[2] C. Bishop: *Pattern Recognition and Machine Learning* (Section 2.4). Springer, 2006.

[3] M. Ehrgott et al.: *Beam selection in radiotherapy design*. Linear Algebra and its Applications 428 (5-6), 2008.

[4] S. Geman et al.: *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*. Journal of Applied Statistics 20, 1993.

[5] T. Griffiths & Z. Ghahramani: *Infinite latent feature models and the Indian buffet process*. Advances in Neural Information Processing Systems 18, 2006.

[6] J. Meyer et al.: *Automatic selection of non-coplanar beam directions for three-dimensional conformal radiotherapy*. British Journal of Radiology 78, 2005.

[7] R.M. Neal: *Slice Sampling*. Annals of Statistics 31 (3), 2003.

[8] R.M. Neal: *Markov Chain Sampling Methods for Dirichlet Process mixture models* Technical Report 4915, Department of Statistics, University of Toronto, 1998.

[9] P. Potrebko et al.: *Improving intensity-modulated radiation therapy using the anatomic beam orientation optimization algorithm*. Medical Physics 5, 2008.

[10] A. Pugachev & L. Xing: *Computer-assisted selection of coplanar beam orientations in intensity-modulated radiation therapy*. Physics in Medicine and Biology 9, 2001.

[11] C. E. Rasmussen: *The Infinite Gaussian Mixture Model*. Advances in Neural Information Processing Systems 12, 2000.

[12] J. Stein et al.: *Number and orientations of beams in intensity-modulated radiation treatments*, Medical Physics 2,1997.

[13] X. Wu et al.: *Selection and determination of beam weights based on genetic algorithms for conformal radiotherapy treatment planning*. Physics in Medicine and Biology 9, 2000.