

Random Spatial Structure of Geometric Deformations and Bayesian Nonparametrics

Christof Seiler¹, Xavier Pennec², and Susan Holmes¹

¹ Department of Statistics, Stanford University, USA

² Asclepios Project, INRIA Sophia Antipolis, France
`christof.seiler@stanford.edu`

Abstract. Our work is motivated by the geometric study of lower back pain from patient CT images. In this paper, we take a first step towards that goal by introducing a data-driven way of identifying anatomical regions of interest. We propose a probabilistic model of the geometrical variability and describe individual patients as noisy deformations of a random spatial structure (modeled as regions) from a common template. The random regions are generated using the distance dependent Chinese Restaurant Process. We employ the Gibbs sampler to infer regions from a set of noisy deformation fields. Each step of the sampler involves model selection (Bayes factor) to decide about fusing regions. In the discussion, we highlight connections between image registration and Markov chain Monte Carlo methods.

1 Introduction

Our long term aim is to better understand the geometric variability of patients suffering from lower back pain (LBP). We have a designed study on a dataset of 400 CT images of the spine with LBP and control patients. In this paper, we take a first step towards analyzing this dataset by introducing a data-driven way of identifying anatomical regions of interest. In a second step, the results obtained in this paper will be used in a statistical analysis of LBP and controls to describe geometric differences between regions of interest.

We assume that patients can be described as noisy deformations from a common template. The template represents an “average” patient image and “common” spatial structure. This structure is then deformed into a specific patient image (see Figure 1). The data under consideration are realizations of these deformations. Surprisingly, little work has been done to model deformations probabilistically. Most approaches focus on finding the optimal number and spatial organization of particles [11,12,15,4] (another way to parametrize the deformations) or regions [3,14] with deterministic algorithms. In this paper, we propose a probabilistic approach using Bayesian nonparametrics which allows us to estimate full posterior distributions of the spatial structure and thus provides a measure of uncertainty of our estimates. In addition, it provides a rigorous way of incorporating prior medical domain knowledge.

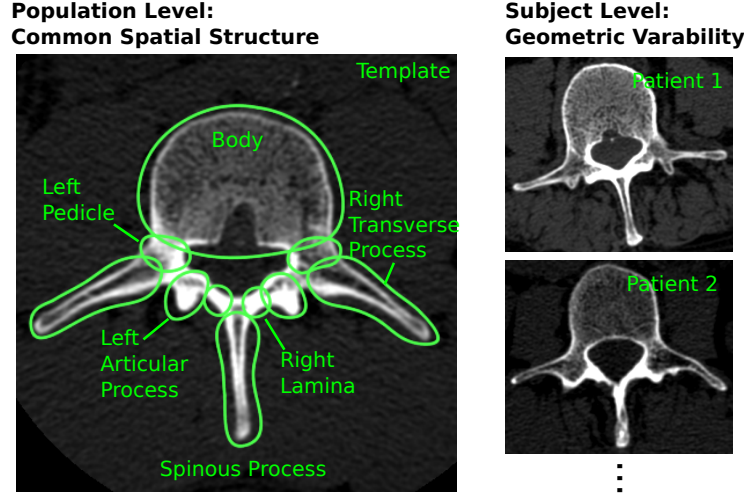


Fig. 1: Hierarchical model of the anatomy. We obtain a specific geometry of a patient (right) by deforming the common spatial structure of a template (left).

One of the earliest approaches to the modeling of shapes was introduced in [16] (first edition appeared in 1917). Thompson was interested in understanding the morphological evolution of species. He proposed using the simplest possible deformation of the underlying space to achieve a “good” match. For instance, he illustrated how to deform the skull of a chimpanzee into a human skull by using shear, affine, conformal and some other more peculiar mappings.

The majority of previous work on the subject has been focused on the most general case, the space of diffeomorphic deformations. Recently [12] studied simpler deformations, conformal mappings, and [11,15,4] reduced the complexity by discretizing deformations with “blobs” placed and optimized deterministically. For spatial structures modeled as regions, we can also find work using deterministic methods. For instance, a multiscale data-driven approach using rectangularly shaped regions aligned along image directions [3], or a multiscale tree structure motivated by the geometry of the anatomy [14].

The input data to our approach are stationary velocity fields (SVF’s) that can be obtained by different registration algorithms [1,8,17,10]. The SVF v is the unique solution to the ODE $\partial\phi(x,t)/\partial t = v(\phi(x,t))$ with initial condition $\phi(x,0) = \text{identity}$. The reason that ODE’s are useful for image registration is that we can generate a diffeomorphic mapping of a patient image I_p to a template image I_t with $I_t(x) = I_p(\phi(x))$ with spatial position $x \in \mathbb{R}^3$, intensity image $I : \mathbb{R}^3 \mapsto \mathbb{R}$, diffeomorphic mapping $\phi : \mathbb{R}^3 \mapsto \mathbb{R}^3$. This assumption makes sense for spines in the absence of fractures and collapse of tissue.

1.1 Our Contributions

Recent developments in Bayesian nonparametrics address clustering of data into an unknown number of partitions with the distance dependent Chinese Restaurant Process (ddCRP) [2]. Here, we build upon these recent developments and introduce a probabilistic model of the spatial structure of geometric deformations at the population level. We translate recent results on segmentation of natural images [6] and motion-based segmentation of 3D objects [5] to dense deformations of medical images. Related work in the medical context are Dirichlet processes to detect spatial activation patterns in fMRI [9] or tractography segmentation [18].

Finally, we highlight promising connections between image registration and Markov chain Monte Carlo (MCMC) methods, in particular Hamiltonian (Hybrid) Monte Carlo (MC) methods, that have not yet been exploited.

2 Random Image Partitions

Our goal is to find self-similar partitions of adjacent voxels in deformations fields. To qualify as a self-similar deformation, vectors at voxel positions within one partition need to follow similar transformation laws for a set of patients. We can illustrate this idea with a hierarchical model: On the first level of the hierarchy, we model the spatial structure common to all patients in the population (Figure 1 left). The spatial structure is represented as image partitions. On the second level of the hierarchy, we model the geometrical variability of each of the partitions (Figure 1 right).

The clustering of voxels differs from general clustering of data in that we are interested in contiguous cluster regions. We want to avoid regions that are scattered across the image, which would lead to unrealistic non-local deformations. A second requirement is that the number of partitions should be data dependent and not be fixed a priori. This becomes crucial for small scale regions in anatomical structures where the different anatomical parts are unknown. The ddCRP provides both, contiguous regions and data dependent clusters from observed deformation fields.

2.1 Distance Dependent Chinese Restaurant Process

In this section, we introduce the ddCRP [2] which we will later use to sample random image partitions. The process generates links between voxels, these links can then be used in a second step to deterministically define partitions. Two voxels belong to the same partition if there is a link between them. The probability of linking voxel x_i to voxel x_j is given by³:

$$p(\mathbf{c}_i = j | D, f, \alpha) \propto \begin{cases} f(d_{ij}) & j \neq i \\ \alpha & j = i \end{cases} \quad (1)$$

³ We denote random variables in **bold** face to distinguish between deterministic and random parameters.

Here, D is the set of all distances between voxels, f is a decay function, we use the window decay $f(d) = 1[d \leq a]$ that considers voxels at most distance a from the current voxel. Setting $a = 1$ forces us to consider adjacent voxels only and ensure partitions without gaps. The parameter α describes the probability of a voxel linking to itself, which results in the creation of a new image partition, thus the larger α , the higher the probability of creating new partitions.

Using the ddCRP allows us to generate a random sample of our hierarchical model depicted in Fig. 1:

1. For each voxel x_i , sample a link from $\mathbf{c}_i \sim \text{ddCRP}(D, f, \alpha)$. Partition $z(\mathbf{c})$ are now given deterministically according to the link structure $\mathbf{c} = [\mathbf{c}_1, \dots, \mathbf{c}_n]$.
2. For each partition, sample a template-subject geometric deformation.

In this section, we described how to sample random partitions (first step of generative process), in the the next section, we will describe a way to sample geometric deformations for each partitions (second step of generative process). Finally, we will show how to do inference given a set of patient specific deformation fields (Section 2.3).

The ddCRP is a generalization of the CRP to non-exchangeable distributions on partitions. This extension is necessary to consider the spatial ordering of voxels in terms of distances between voxels.

2.2 Regional Bayesian Linear Regression

In this section, we describe a Bayesian linear regression model for predicting one affine transformation for one partition of a SVF. In a second step, we give the marginal likelihood of that model, which will be used to perform inference using the ddCRP.

Let $z(\mathbf{c})_k = [x_1, \dots, x_{|z(\mathbf{c})_k|}]$ denote the k th image partition and $|z(\mathbf{c})_k|$ the number of voxels in that partition. Then, we stack the corresponding velocity vectors into $\mathbf{v} = [\mathbf{v}_1, \dots, \mathbf{v}_{|z(\mathbf{c})_k|}]^\top$. We can generate a SVF v from the Lie algebraic representation of affine transformations M with $v(x) = M\tilde{x}$ and homogenous coordinate $\tilde{x} = [x, 1]^\top$. The Lie algebraic representation relates to the affine group through the Lie group exponential which is in this special case the standard matrix exponential $\exp(M) = \sum_{k=0}^{\infty} \frac{1}{k!} M^k$. We vectorize $\mathbf{m} = \text{Vect}(M_{3 \times 4})$ and put a normal prior on \mathbf{m} ($M_{3 \times 4}$ is M after removing the 4th row of zeros):

$$\mathbf{m} \sim \mathcal{N}(\mu, \Gamma), \quad (2)$$

with the mean set to $\mu = 0$ and the concentration matrix (inverse of covariance) set to $\Gamma = \text{diag}(\sigma_{r+s}^{-2}, \sigma_{r+s}^{-2}, \sigma_t^{-2}) \otimes \text{diag}(1_{3 \times 3})$.

To obtain an intuition about the hyperparameters σ_{r+s}, σ_t and to justify the independence assumption between parameters, we use a first order approximation of the Baker-Campbell-Hausdorff (BCH) formula $Z = \log(\exp(X) \exp(Y))$ and the Jordan/Schur decomposition $M = \frac{1}{2}(M - M^\top) + \frac{1}{2}(M + M^\top)$:

$$M_{3 \times 4} = T_{3 \times 4} + R_{3 \times 4} + S_{3 \times 4} = \begin{bmatrix} 0_{3 \times 3} & t \end{bmatrix} + \theta \begin{bmatrix} 0 & -r_3 & r_2 & 0 \\ r_3 & 0 & -r_1 & 0 \\ -r_2 & r_1 & 0 & 0 \end{bmatrix} + [\text{diag}(s) \ 0_{3 \times 1}]$$

for translation T , rotation R and scaling S . The well known representation of rotations in \mathbb{R}^3 using an axis and rotation around it, is the natural description in the Lie Algebra of the group of rotations and yields a decoupling of rotation axis $r = [r_1, r_2, r_3]^\top$, $|r| = 1$ and angle θ .

The hyperparameters σ_{t+r}^2 and σ_s^2 are in the log-domain but can be related to the parameters of regular affine transformations so that we can set them according to medical knowledge. The scaling factor can be obtained through $\exp(S)_{i,i} = \sum_{k=0}^{\infty} s_i^k/k! = \exp(s_i)$. The rotation is given by θ times $-1 \leq r_i \leq 1$. The translation factor is more difficult because affine groups are semidirect products $\mathbb{R}^3 \rtimes \text{GL}(3, \mathbb{R})$, meaning that the general linear part acts on the real vector space of the translational part. A first order approximation yields $\exp(T)_{i,4} = \sum_{k=0}^{\infty} s_i^k t_i \approx t_i$.

The Bayesian linear regression model for a SVF partition is given by:

$$\mathbf{v} = \Phi \mathbf{m} + \epsilon, \text{ with } \epsilon \sim \mathcal{N}(0, \frac{1}{\sigma_v^2} I_{3|z(c)_k|}), \quad (3)$$

as for the observations, let us stack the matrices $\phi(x) = \tilde{x} \otimes I_3$ at all voxels into a big matrix $\Phi = [\phi(x_1)^\top, \dots, \phi(x_{|z(c)_k|})^\top]^\top$, with dimensions $3|z(c)_k| \times 12$. The likelihood of observing a deformation follows a normal distribution,

$$\mathbf{v} | \Phi \mathbf{m}, \sigma_v^2 I_{3|z(c)_k|} \sim \mathcal{N}(\Phi \mathbf{m}, \frac{1}{\sigma_v^2} I_{3|z(c)_k|}), \quad (4)$$

and the posterior follows a normal distribution,

$$\mathbf{m} | \mathbf{v}, \mu, \Gamma \sim \mathcal{N}(\check{\mu}, \check{\Gamma}). \quad (5)$$

To compute the Bayes factor within the Gibbs sampling step of the ddCRP (next section), we analytically calculate the marginal likelihood up to a constant K that partially cancels out in the Bayes factor $B(x, z, \mu, \Gamma, \sigma_v^2)$:

$$p(\mathbf{v} | \mu, \Gamma, \sigma_v^2) = \int p(\mathbf{v} | \Phi \mathbf{m}, \sigma_v^2) p(\mathbf{m} | \mu, \Gamma) d\mathbf{m} = K \exp\left(\frac{1}{2\sigma_v^2} \mathbf{v}^\top \Phi \check{\Gamma}^{-1} \Phi^\top \mathbf{v}\right), \quad (6)$$

with the posterior concentration matrix $\check{\Gamma} = \Phi^\top \Phi + \sigma_v^2 \Gamma$.

2.3 Inference with the Gibbs Sampler

The computation of the posterior is intractable due to the ddCRP regardless of the other terms. We approximate the posterior using the Gibbs sampler. The Gibbs sampler changes one link at a time while all others are fixed [2] proportional to the following distribution:

$$p(\mathbf{c}_i | \mathbf{c}_{i-1}, \mathbf{v}, \mu, \Gamma, \sigma_v^2, D, f, \alpha) \propto \begin{cases} p(\mathbf{c}_i | D, f, \alpha) B(\mathbf{v}, z(\mathbf{c}), \mu, \Gamma, \sigma_v^2) & \text{if } \mathbf{c}_i \text{ joins } k \text{ and } l \\ p(\mathbf{c}_i | D, f, \alpha) & \text{otherwise,} \end{cases}$$

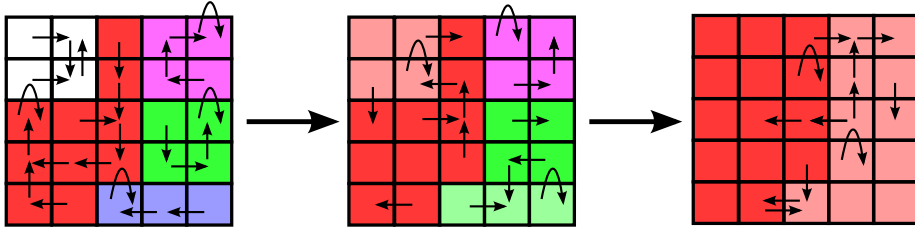


Fig. 2: Evolution of random partitions. Each color represents one partition. The links (arrows) define the partitions. Curved arrows are self references and straight arrows are links between neighboring voxels.

where B is the Bayes factor comparing two models: (1) creation of a new joint partitions k and l , (2) keeping the two partitions separated. The factor is computed over a set of i.i.d. patients t :

$$B(\mathbf{v}, z(\mathbf{c}), \mu, \Gamma, \sigma_v^2) = \frac{\prod_t p(\mathbf{v}_{z(\mathbf{c})_{k \cup l}} | \mu, \Gamma, \sigma_v^2)}{\prod_t p(\mathbf{v}_{z(\mathbf{c})_k} | \mu, \Gamma, \sigma_v^2) \prod_t p(\mathbf{v}_{z(\mathbf{c})_l} | \mu, \Gamma, \sigma_v^2)} \quad (7)$$

Figure 2 gives an example of the evolution of three Gibbs sampling steps. Fusing regions together until there is only one left.

2.4 Experiments with Synthetic and Spine CT Images

One step of the Gibbs sampler yields one partition image, which makes it difficult to visualize the posterior distributions with conventional histogram plots. Therefore, we visualize our results using movie animations of two-dimensional slices extracted from the three-dimensional images. We conducted experiments on synthetic deformations with known partitions and real spine deformations with unknown partitions that are driven by intersubject anatomical variability. The video animations are available on the corresponding author's website⁴.

3 Discussion: Image Registration and MCMC

In this paper, we illustrated how one can use the Gibbs sampler in an image registration context. We showed how to infer the random partitioning structure from a set of deformations. The estimation of deformations was modeled with Bayesian linear regression yielding analytical solutions for the marginal likelihood need in the ddCRP. For more complicated deformations other MCMC methods are need in registration problems. Here we list a few interesting developments with potential impact on image registration.

⁴ <http://stanford.edu/~cseiler/gsi2013/>

Hamiltonian MC. Hamiltonian Monte Carlo (HMC) [13] shows promising results in high dimensional problems in different scientific fields. One way to apply HMC to image registration is to use a particle-based registration methods (e.g. [11]). There are two steps:

1. Draw new momentum variable from a symmetric proposal distribution $\mathbf{p} \sim \mathcal{N}(0, M)$ (this corresponds to the momentum variable in the particle-based registration formulation).
2. Integrate L steps of the Hamiltonian dynamics and accept the end state with a biased coin toss (corresponds to a Metropolis update).

There is one major difference for HMC in registration. We have three main properties for standard HMC: (i) reversibility, (ii) conservation of Hamiltonians (energy preservation), (iii) volume preservation. In the case of registration, volume preservation only holds for global rotations and scaling in one direction while compensating in another direction. In contrast, in registration problems, we are also interested in local rotation, general anisotropic scaling and skewing. We are also concerned with the speed, since potential thousands of computations need to be performed.

Riemannian MCMC. The authors [7] propose a MCMC method that generalizes the standard Hamilton MC case. In contrast to the standard method, it replaces the proposal probability covariance M (from the previous section) with a local covariance G , which is a function of the parameters and the data. This G takes into account the curvature of the parameter space that forms a Riemannian manifold. In particular, the authors promote the use of the expected Fisher information matrix plus the negative Hessian of the log-prior to define G .

The task of computing Riemannian geodesics in the LDDMM framework provides a direct link to Riemannian MCMC methods. One of the key challenges for translating these methods to image registration are the violation of volume preservation and the computational complexity.

4 Conclusions

In this paper, we showed how to learn the unknown spatial structure from observed geometric image deformations in the context of Bayesian nonparametrics. This direction of research is motivated by three main requirements: (i) Hierarchical description of spatial structure at population level and geometric variability at patient level. (ii) Compact discrete description of anatomical variability. (iii) Probabilistic formulation instead of point estimates to measure uncertainty in parameters and incorporate prior medical domain knowledge. All of these points are derived from clinical requirements, where the main motivation is to enable medical doctors to understand our models (which can be hard in case of infinite-dimensional diffeomorphic mappings). This forces us to rethink our modeling approach from scratch.

Acknowledgements

This work was funded by a fellowship from the Swiss National Science Foundation and a grant from the France-Stanford Center for Interdisciplinary Studies.

References

1. Ashburner, J.: A fast diffeomorphic image registration algorithm. *NeuroImage* 38(1), 95–113 (2007)
2. Blei, D.M., Frazier, P.I.: Distance dependent chinese restaurant processes. *The Journal of Machine Learning Research* 12, 2461–2488 (2011)
3. Buerger, C., Schaeffter, T., King, A.P.: Hierarchical adaptive local affine registration for fast and robust respiratory motion estimation. *Medical Image Analysis* 15(4), 551–564 (Aug 2011)
4. Durrleman, S., Allasonnière, S., Joshi, S.: Sparse adaptive parameterization of variability in image ensembles. *International Journal of Computer Vision* 101(1), 161–183 (Aug 2013)
5. Ghosh, S., Sudderth, E., Loper, M., Black, M.: From deformations to parts: Motion-based segmentation of 3D objects. In: *NIPS*. pp. 2006–2014 (2012)
6. Ghosh, S., Ungureanu, A.B., Sudderth, E.B., Blei, D.M.: Spatial distance dependent chinese restaurant processes for image segmentation. In: *NIPS* (2011)
7. Girolami, M., Calderhead, B.: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(2), 123–214 (Mar 2011)
8. Hernandez, M., Bossa, M.N., Olmos, S.: Registration of anatomical images using geodesic paths of diffeomorphisms parameterized with stationary vector fields. In: *ICCV 2007*. pp. 1–8. IEEE (Oct 2007)
9. Kim, S., Smyth, P., Stern, H.: A nonparametric Bayesian approach to detecting spatial activation patterns in fMRI data. In: *MICCAI 2006, Part II. LNCS*, vol. pp. 217–224. Springer Heidelberg (2006)
10. Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X.: LCC-demons: A robust and accurate diffeomorphic registration algorithm. *NeuroImage* (May 2013)
11. Marsland, S., McLachlan, R.: A Hamiltonian particle method for diffeomorphic image registration. *IPMI* 20, 396–407 (2007)
12. Marsland, S., McLachlan, R., Modin, K., Perlmutter, M.: Geodesic warps by conformal mappings. *International Journal of Computer Vision* pp. 1–11 (Oct 2012)
13. Neal, R.M.: MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* 54, 113–162 (2010)
14. Seiler, C., Pennec, X., Reyes, M.: Capturing the multiscale anatomical shape variability with polyaffine transformation trees. *Medical Image Analysis* (2012)
15. Sommer, S., Lauze, F., Nielsen, M., Pennec, X.: Sparse Multi-Scale diffeomorphic registration: The kernel bundle framework. *Journal of Mathematical Imaging and Vision* pp. 1–17 (2012)
16. Thompson, D.W.: *On Growth and Form*. Cambridge University Press, abridged edn. (Jul 1992)
17. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient Non-Parametric image registration. *NeuroImage* 45(1 Suppl), S61–S72 (Mar 2009)
18. Wang, X., Grimson, E.E., Westin, C.F.F.: Tractography segmentation using a hierarchical Dirichlet processes mixture model. *NeuroImage* 54(1), 290–302 (Jan 2011)