

Probabilistic shortest path tractography in DTI using Gaussian Process ODE solvers

Michael Schober¹, Niklas Kasenburg^{1,2}, Aasa Feragen^{1,2}, Philipp Hennig¹,
Søren Hauberg³

¹ Max Planck Institute for Intelligent Systems, Tübingen, Germany

² Department of Computer Science, University of Copenhagen, Denmark

³ Technical University of Denmark

Abstract. Tractography in diffusion tensor imaging estimates connectivity in the brain through observations of local diffusivity. These observations are noisy and of low resolution and, as a consequence, connections cannot be found with high precision. We use probabilistic numerics to estimate connectivity between regions of interest and contribute a Gaussian Process tractography algorithm which allows for both quantification and visualization of its posterior uncertainty. We use the uncertainty both in visualization of individual tracts as well as in heat maps of tract locations. Finally, we provide a quantitative evaluation of different metrics and algorithms showing that the adjoint metric [8] combined with our algorithm produces paths which agree most often with experts.

1 Introduction

Diffusion tensor imaging (DTI) is a non-invasive imaging technology generating a global mapping of local brain diffusivity. DTI estimates, in each voxel, the local directional probability of water molecule displacement, represented as a tensor field. Assuming high diffusion along brain fibers, global connections can be estimated by integrating the tensor field, a process referred to as *tractography*.

Shortest path algorithms for tractography are useful for quantifying structural brain connectivity between regions of interest (ROIs), as needed for structural brain connectivity graphs. Due to low resolution and a low signal-to-noise ratio, one cannot find connections with high precision. Quantification and interpretation of the uncertainty of solutions is thus an important problem [18]. We use probabilistic numerics [12] both to estimate connectivity in the form of tracts connecting ROIs, and to quantify and visualize the tracts and their uncertainty.

We apply a recent algorithm [12] to infer a distribution of possible solutions to shortest path problems in tractography. This distribution is represented as a Gaussian Process (GP) over the solution to an ordinary differential equation (ODE). GPs offer novel ways to quantify and visualize uncertainty arising from the numerical computation, and allow marginalization over a space of feasible solutions. This has two strong advantages: 1) the inclusion of quantified numerical uncertainty into the final tractography result gives a more honest view of the accuracy of the algorithm; and 2) visualizing the uncertainty emphasizes the

fact that the shortest paths solutions found are *not* real fibers in the brain, but mathematical abstractions of probable connectivity.

Contribution and organization. In Sec. 1.1 we provide a brief overview of tractography. We then present the suggested algorithm for tractography based on a probabilistic numerical solution to the ODEs arising when a Riemannian metric is estimated from the diffusion tensors (Sec. 2). This gives a distribution of possible shortest paths in the form of a GP. We then show new visualization techniques emerging from the availability of the uncertainty of the GP solution (Sec. 3). Finally, the same section provides results from a quantitative analysis demonstrating that the adjoint metric [8] combined with GP ODE solvers produces shortest paths matching best to expert annotations compared to state-of-the-art in shortest path tractography.

1.1 Background and Related Work

Current tractography methods are either *fiber tracking* or *shortest path* methods.

Fiber tracking methods greedily trace the most probable path in any direction from a chosen seed until a stopping criterion is met. *Deterministic* fiber tracking algorithms [2] trace paths by following a non-stochastic rule, while *probabilistic* algorithms [15] randomly sample different directions in each step and continue forward from the best one(s).

While popular, these methods suffer from two problems. 1) Exploratory tracking methods continue tracing out a path until the algorithm no longer knows where to go. Because of this, many tracking algorithms either get lost in low-connected areas or in areas with crossing fibers. 2) Voxels near the starting region are explored more thoroughly than voxels far away. Some parts of the brain can, thus, be under-explored. This not only introduces a bias towards paths close to the starting region, but may also have the effect that the optimal path is never explored. This is particularly important when tractography is used to generate structural connectivity networks between ROIs for population studies of brain networks. Weak connections will be far less reliably estimated than strong ones, meaning that any graph analysis performed on the resulting network is biased towards finding differences in strong connections.

Shortest path methods tackle these problems by specifically computing connections between ROIs rather than connections from a seed. To derive efficient algorithms, optimal connections are often defined as *shortest paths* under a Riemannian metric, which is inversely proportional to the local diffusion tensor [13,14]. This way, local steps of short length correspond to local steps of large diffusion. While multiple shortest paths might exist between two given voxels, only one solution is obtained. Hence, many voxels are considered (cf. Fig. 3). In more detail, from the diffusion tensor D_i at voxel i , a local distance measure $\text{dist}^2(a, b) = \alpha_i(a-b)^\top D_i^{-1}(a-b)$ is defined. Here α_i is a per-voxel scaling factor, where most commonly $\alpha_i = 1$ [13,14]. Hao et al. [11] argue that a per-voxel scaling is needed and provide a numerical scheme for computing α_i . Fuster et al. [8] show that shortest paths only correspond to free Brownian motion under the

implied Riemannian metric when $\alpha_i = \det D_i$. This is called the *adjoint metric*. We investigate this choice further in the empirical evaluation (Sec. 3).

Discrete shortest path methods construct a graph with vertices corresponding to voxel positions, edges between neighboring voxels, and edge length defined according to the local distance measure. Shortest paths can then be found either using *Dijkstra’s algorithm* [14] or *fast marching methods* [16]. This is computationally efficient, but suffers from discretization errors as the shortest paths are restricted to go through the voxels.

Continuous shortest path methods smoothly interpolate the per-voxel metric tensors $M_i = \alpha_i D_i^{-1}$ to form a continuous metric space. The shortest path from a to b can then be found by solving the following system of ODEs [13]

$$\begin{aligned} c''(t) &= -\frac{1}{2}M^{-1}(c(t)) \left[\frac{\partial \text{vec } M(c(t))}{\partial c(t)} \right]^\top [c'(t) \otimes c'(t)] \\ &=: f(c(t), c'(t)), \quad \text{with } c(0) = a, \quad c(1) = b, \end{aligned} \tag{1}$$

where $\text{vec } M \in \mathbb{R}^9$ stacks the columns of M and \otimes is the Kronecker product. The solution is a continuous curve $c : [0, 1] \rightarrow \mathbb{R}^3$, so discretization issues are avoided. Unlike discrete methods, the solution found by the numerical solver might only provide a locally optimal path. This concern, along with their ease of implementation, has made discrete methods the most common solution in shortest path tractography [8, 11, 14, 16].

Non-trivial ODEs such as (1) cannot be solved analytically. Traditional numerical algorithms return a point estimate in the solution space of all curves agreeing with the *finitely many* evaluation points of the ODE. This might lead to the impression that this is *the* true solution rather than an *uncertain* approximation. GP ODE solvers return a more accurate picture in this regard.

2 Methodology

We propose a probabilistic continuous shortest path tractography algorithm. Similarly to Lenglet et al. [13] we form a continuous Riemannian metric through trilinear interpolation of the metric tensors given at voxel locations. We then provide a method for solving Eq. (1) numerically which gives explicit estimates of the uncertainty of the solution.

2.1 Probabilistic ODE Solvers

The central idea in recent work on probabilistic solvers for ODEs is to assign a prior probability distribution over possible solutions to the ODE, then iteratively refine it to a posterior distribution by repeatedly evaluating the ODE at test inputs. For algebraic as well as conceptual reasons [4], GPs are a preferable choice for the prior distribution. We first review GPs and then discuss their application in ODE solvers.

Review of GP regression. A Gaussian Process $\mathcal{GP}(c; \mu, k)$ [17] is a probability distribution over real-valued functions $c : \mathbb{R} \mapsto \mathbb{R}$ such that any finite restriction to function values $\{c(t_1), \dots, c(t_N)\}$ has a Gaussian distribution. GPs are parameterized by a mean function $\mu : \mathbb{R} \mapsto \mathbb{R}$ and a covariance function $k : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$. GPs are closed under linear transformations; a linear operation Φ induces a distribution over Φc as $\mathcal{GP}(\Phi c; \Phi \mu, \Phi k \Phi^\top)$, where Φ^\top denotes application of Φ to the left. Given observations $(T, Y) = [(t_1, y_1), \dots, (t_N, y_N)]^\top$ of likelihood $\mathcal{N}(y; \Phi c, \sigma^2 \mathbf{I})$, the posterior over c is a $\mathcal{GP}(c; \tilde{\mu}, \tilde{k})$ with

$$\tilde{\mu}(t) := \mu(t) + k(t, T) \Phi^\top (\Phi k_{TT} \Phi^\top + \sigma^2 \mathbf{I})^{-1} (Y - \Phi \mu(T)) \quad (2)$$

$$\tilde{k}(t, u) := k(t, u) - k(t, T) \Phi^\top (\Phi k_{TT} \Phi^\top + \sigma^2 \mathbf{I})^{-1} \Phi k(T, u), \quad (3)$$

where $(k_{TT})_{ij} := k(t_i, t_j)$ is the covariance matrix of input locations [17, §2.2].

Fig. 1 illustrates the concept with a GP posterior belief $\mathcal{GP}(c; \tilde{\mu}, \tilde{k})$ (green) and its derivative $\mathcal{GP}(\partial c; \partial \tilde{\mu}, \partial \tilde{k} \partial^\top)$ (orange). Bold lines show posterior mean and filled areas show point-wise two times the standard deviation. The posterior was generated from 5 observations (black). Beliefs over multi-output functions $c = [c_i(t)]$, $i = 1, \dots, D$ can be constructed through vectorization. If the covariance structure is assumed to factorize between inputs and outputs

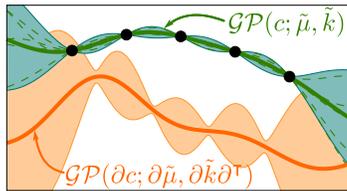


Fig. 1. GP example.

$$\text{cov}(c_i(t), c_j(u)) = V_{ij} \cdot k(t, u), \quad (4)$$

then the belief over c can be written as $p(c(t)) = \mathcal{GP}(c; \mu_c, V \otimes k)$. (The important special case $V = \mathbf{I}$ amounts to independent GP priors for each dimension of c). The covariance structure determines the regularity and uncertainty of the curve.

GP ODE solvers. In [12], Hennig & Hauberg studied a framework, originally envisioned by Skilling [19], for obtaining a posterior probability distribution $p(c)$ over the solution c to an ODE like Eq. (1). The general concept was also recently analyzed by Chkrebtii et al. [4]. It works as follows.

Since differentiation is a linear operation, a GP distribution on c induces a GP belief on its derivatives, if the mean and kernel functions are sufficiently smooth. One can now repeatedly construct an estimate $\tilde{c}(t_i)$ for the true solution $c(t_i)$ to the ODE and use it to construct approximate observations $y_i = f(\tilde{c}(t_i), \partial \tilde{c}(t_i))$ of $\partial^2 c(t_i)$. As $\tilde{c}(t_i)$ is only an approximation to the true solution $c(t_i)$, the observation y_i is imprecise up to some error that, in this framework, is estimated by propagating the current uncertainty over c through the ODE:

At step i , assume a current posterior $p(c) = \mathcal{GP}(c; \tilde{\mu}^i, \tilde{k}^i)$. We construct the estimates \tilde{c} and $\partial \tilde{c}$ as the current “best guess”, the mean: $[\tilde{c}, \partial \tilde{c}] = [\tilde{\mu}(t_i), \partial \tilde{\mu}(t_i)]$. The estimate for the error of this guess is the current marginal variance

$$\text{cov} \begin{pmatrix} \tilde{c}^i(t_i) \\ \partial \tilde{c}^i(t_i) \end{pmatrix} = \begin{pmatrix} \tilde{k}^i(t_i, t_i) & \left. \frac{\partial \tilde{k}^i(t_i, t)}{\partial t} \right|_{t=t_i} \\ \left. \frac{\partial \tilde{k}^i(t, t_i)}{\partial t} \right|_{t=t_i} & \left. \frac{\partial^2 \tilde{k}^i(t, u)}{\partial t \partial u} \right|_{t, u=t_i} \end{pmatrix} =: \Sigma^i. \quad (5)$$

Assuming we have access to upper bounds or other estimates $U > \partial f / \partial c$ and $U' > \partial f / \partial c'$ on the gradients of f , we can use Σ^i to construct an estimate for the error on y_i as $[U, U']^\top \Sigma^i [U, U'] =: \Lambda^i$, which gives a likelihood function $p(y_i | \partial^2 c(t_i)) = \mathcal{N}(y_i; f(\tilde{c}^i, \partial \tilde{c}^i), \Lambda^i)$. Using (2) and (3), the belief can be updated to obtain $\tilde{\mu}^{i+1}, \tilde{k}^{i+1}$, and the process repeats.

GP ODEs for tractography. For the DTI shortest path problem (1), we make the following specific choices: For the prior covariance we use the Kronecker form of (4), with a Gaussian kernel $k(t_i, t_j) = \exp[-(t_i - t_j)^2 / (2\lambda^2)]$, which has one free model parameter, the length scale λ . In contrast to [12], we estimate both V and λ with a marginal moment matching method instead of evidence maximization [17, §2.3], which allows for more adaptive solutions. Like [12], we include boundary conditions analytically in the prior belief as direct observations of c with vanishing noise, using a regular grid with $2N$ observations spread over the domain $t = [0, 1]$. We initialize the prior mean for the solver from a discrete shortest path $v_0 = a, v_1, \dots, v_N = b$, which we pre-process with a GP least-squares smoother (with a square exponential kernel function, whose parameters are optimized using evidence maximization [17, §5.3] separately for each output dimension). This gives a smooth prior mean function for the GP solver.

GP ODE solutions. The output of the ODE solver is a posterior GP belief $\mathcal{GP}(c; \tilde{\mu}, \tilde{k})$. The repeated linear extrapolation to construct the $\tilde{c}(t_i)$ in the GP solver is structurally very similar to Runge-Kutta methods [10], but does not afford the strong analytic guarantees of those classic solvers. However, this algorithm, at any point during its runtime, retains a joint, consistent probability distribution over one locally distinct solution c and all its derivatives (cf. Fig. 1), from which *joint* samples (candidate solutions for the ODE) can be drawn. In contrast, classic solvers only return a single joint point estimate for c , with *local* error estimates that do not allow joint sampling.

3 Experiments

Data. Tractography was performed on pre-processed diffusion data of 40 subjects provided as a subsample from the Q3 release of the Human Connectome Project (HCP) [6, 7, 9, 20]. The pre-processed HCP diffusion data contains 270 diffusion directions distributed equally over 3 shells with b -values = 1000, 2000 and 3000 s/mm² [20]. DTI tensors for each voxel were computed with dtifit [1]. Segmentation was performed with FAST [22]. The cortico-spinal tract (CST) used for experiments was obtained from the expert annotated Catani tract atlas [3]. ROI atlases were constructed in “template space” by overlapping the tract atlas with regions defined by the Harvard-Oxford cortical and sub-cortical atlas [5] using the overlap with the brainstem, the hippocampus and the amygdala for one region, and the overlap with the superior frontal gyrus, the precentral gyrus and the postcentral gyrus for the second region. The warp fields provided by HCP were used to warp the ROIs from “template space” to “subject spaces”.

Evaluation. We subsample 250 pairs of points at random from the ROIs for all 40 subjects and compute the discrete and continuous shortest paths. To

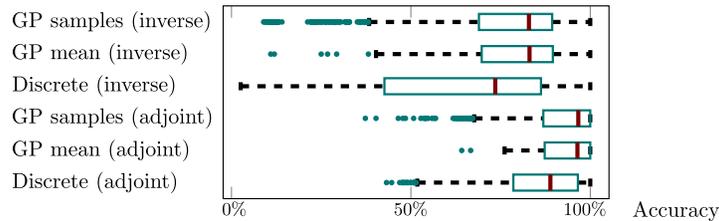


Fig. 2. Box plot of the evaluation score of different methods under different metrics.

compare the quality of the solutions we compute the set of voxels each path passes through. Taking the Catani atlas as a reference we measure *accuracy* defined as the percentage of voxels which are classified as belonging to the CST by at least one expert. We evaluate both the standard inverse metric [13,14] and the adjoint metric [8]. Fig. 2 shows the results for all 40 subjects.

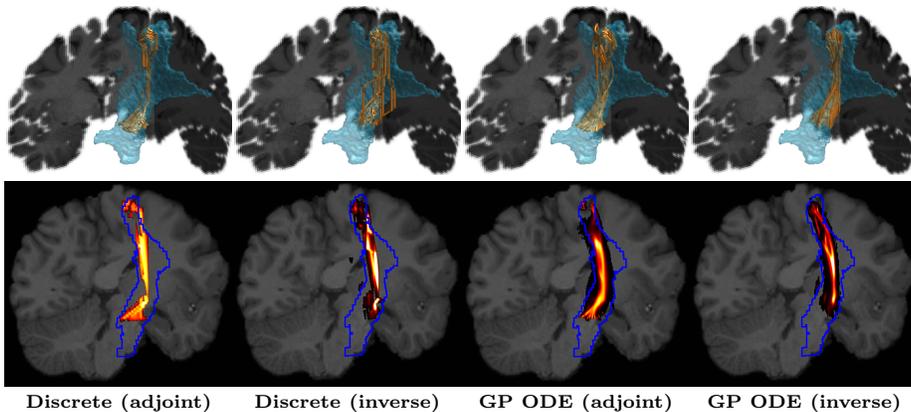


Fig. 3. Top: Geodesics under the inverse [13,14] and the adjoint metrics [8] in the right CST. Blue area shows voxels in the Catani atlas which at least one expert considered to be part of the tract. By considering different endpoints, bifurcating tracts can be discovered (1st and 3rd figure). **Bottom:** Density of discrete (left) and continuous (right) paths using two different metrics. CST of the Catani atlas as defined by at least one expert as reference (blue). **Also see the supplementary material**⁴.

Fig. 3 shows example paths from both algorithms and metrics computed on a single subject. The supplementary material⁴ further contain an animation of samples from the GP solutions. Additionally, by counting the number of discrete paths going through each voxel, we produce a 3D density of paths. For each GP posterior we sample several curves, and proceed as in the discrete case. Fig. 3

⁴ <http://probabilistic-numerics.org/ODEs.html>

shows a 2D heat map slice of the 3D density, chosen via principal component analysis to contain maximal variance.

4 Discussion and Conclusion

We use probabilistic numerics to compute shortest paths for tractography. This captures the uncertainty inherent in the shortest path computation which can be used when visualizing the results. Sampling and averaging from the GP posterior allows us to marginalize over its uncertainty. We utilize this to generate heat maps of path densities. Compared to discrete methods, these heat maps appear smoother due to marginalization, but also sharper due to the continuous description of the sampled solutions.

Fig. 3 shows the classic “spaghetti plot” used for visualizing tractography results. The discrete solutions show tendencies to straight line segments connected by rather sharp turns which cannot solely be explained by the discretization error. In general, the continuous solutions bend less drastically, as we would expect from actual fibers. The figure only shows the mean function of the GP estimates of the geodesics; **the supplements contain an animation of the uncertainty**⁴. This provides a visualization of the solutions which makes it very clear that individual paths cannot, and *should not*, be interpreted as individual fibers in the brain — a common misinterpretation of “spaghetti plots”.

We introduce a quality measure to compare different methods. First, we compare the standard inverse metric [13, 14] to the recently suggested adjoint metric [8]; empirical results show that the theoretically strong adjoint metric is consistently better. Secondly, we find that the GP posteriors agree with experts slightly more often than the discrete solutions. Generally, samples from the posterior score lower than the mean prediction, but this is to be expected since the Gaussian distribution puts non-zero probability mass everywhere.

A general disadvantage of shortest path tractography is that there will always be *some* path connecting any two points, whether it is anatomically there or not. This can be alleviated to some extent by discarding improbable paths.

Recent work by Wassermann et al. [21] illustrates that GPs form a particularly useful representation of shortest paths for population studies of brain connectivity. Our GP solution to the tractography problem lends itself particularly well to this type of population analysis as it avoids the post-hoc GP fitting used by Wassermann et al. This will be further investigated in future work.

Acknowledgements. Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. A.F. is funded in part by the Danish Council for Independent Research (DFR), Technology and Production Sciences. S.H. is funded in part by the Villum Foundation; the DFR, Natural Sciences; and an `amazon.com` machine learning in education award.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B* 103(3), 247–254 (1994)
2. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine* 44(4), 625–632 (2000)
3. Catani, M., de Schotten, M.T.: A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex* 44(8), 1105–1132 (2008)
4. Chkrebti, O., Campbell, D., Girolami, M., Calderhead, B.: Bayesian uncertainty quantification for differential equations. *arXiv stat.ME 1306.2365* (Jun 2013)
5. Desikan, R.S., et al.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31(3), 968–980 (2006)
6. van Essen, D., et al.: The Human Connectome Project: a data acquisition perspective. *NeuroImage* 62(4), 2222–2231 (2012)
7. van Essen, D., et al.: The WU-Minn Human Connectome Project: an overview. *NeuroImage* 80, 62–79 (2013)
8. Fuster, A., et al.: A novel Riemannian metric for geodesic tractography in DTI. In: *Computational Diffusion MRI and Brain Connectivity*, pp. 97–104. *Mathematics and Visualization* (2014)
9. Glasser, M., et al.: The minimal preprocessing pipelines for the human connectome project. *NeuroImage* 80, 105–124 (2013)
10. Hairer, E., Nørsett, S., Wanner, G.: *Solving Ordinary Differential Equations I – Nonstiff Problems*. Springer (1987)
11. Hao, X., Whitaker, R., Fletcher, P.: Adaptive Riemannian metrics for improved geodesic tracking of white matter. In: *IPMI*, pp. 13–24. Springer (2011)
12. Hennig, P., Hauberg, S.: Probabilistic solutions to differential equations and their application to Riemannian statistics. In: *AISTATS*. vol. 33. *JMLR, W&CP* (2014)
13. Lenglet, C., et al.: Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping. In: *ECCV*. pp. 127–140 (2004)
14. O’Donnell, L., Haker, S., Westin, C.F.: New approaches to estimation of white matter connectivity in diffusion tensor MRI: Elliptic PDEs and geodesics in a tensor-warped space. In: *MICCAI* (1). pp. 459–466 (2002)
15. Parker, G., et al.: A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *J. Magn. Reson. Imaging* 18(2), 242–254 (2003)
16. Prados, E., et al.: Control theory and fast marching techniques for brain connectivity mapping. In: *CVPR*. pp. 1076–1083 (2006)
17. Rasmussen, C., Williams, C.: *GPs for Machine Learning*. MIT Press (2006)
18. Schultz, T., et al.: Fuzzy fibers: Uncertainty in dMRI tractography. In: *arXiv cs.CV 1307.3271* (2013)
19. Skilling, J.: *Bayesian solution of ordinary differential equations. Maximum Entropy and Bayesian Methods*, Seattle (1991)
20. Sotiropoulos, S.N., et al.: Effects of image reconstruction on fiber orientation mapping from multichannel diffusion MRI: reducing the noise floor using SENSE. *Magn Reson Med* 70(6), 1682–89 (2013)
21. Wassermann, D., et al.: White matter bundle registration and population analysis based on gaussian processes. In: *IPMI*. pp. 320–332 (2011)
22. Zhang, Y., et al.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE TMI* 20(1), 45–57 (2001)