Auto-calibrating Spherical Deconvolution Based on ODF Sparsity

Thomas Schultz[1] and Samuel Groeschel[2]

[1] University of Bonn, Germany
[2] Experimental Pediatric Neuroimaging and Department of Pediatric Neurology & Developmental Medicine, University Children's Hospital Tübingen, Germany*

Abstract. Spherical deconvolution models the diffusion MRI signal as the convolution of a fiber orientation density function (fODF) with a single fiber response. We propose a novel calibration procedure that automatically determines this fiber response. This has three advantages: First, the user no longer needs to provide an estimate of the response. Second, we estimate a per-voxel fiber response, which is more adequate for the analysis of patient data with focal white matter degeneration. Third, parameters of the estimated response reflect diffusion properties of the white matter tissue, and can be used for quantitative analysis.

Our method works by finding a tradeoff between a low fitting error and a sparse fODF. Results on simulated data demonstrate that auto-calibration successfully avoids erroneous fODF peaks that can occur with standard deconvolution, and that it resolves fiber crossings with better angular resolution than FORECAST, an alternative method. Parameter maps and tractography results corroborate applicability to clinical data.

1 Introduction

Constrained Spherical Deconvolution (CSD) [1] is one of the most widely used methods for inferring fiber distributions from High Angular Resolution Diffusion Imaging (HARDI). It obtains a fiber Orientation Density Function (fODF) by deconvolving the measured HARDI signal with a kernel that describes the response of a single fiber compartment. This single fiber response is a central parameter of the method. It is often assumed to be constant throughout the brain and estimated from voxels that are assumed to contain a single fiber bundle [2].

Recently, it has been pointed out that a mismatch between the true fiber response and the kernel used in CSD can lead to spurious peaks in the fODF [3]. Therefore, a more careful choice of the deconvolution kernel has the potential to improve fODF estimates, especially in cases where white matter degradation is studied, since degraded fibers can have a significantly different MR response.

* We would like to thank Bernhard Schölkopf (MPI for Intelligent Systems, Tübingen) for discussions about this work. S.G. gratefully acknowledges support through grant IZKF 2103-0-0 of the Medical Faculty of the University of Tübingen.

K. Mori et al. (Eds.): MICCAI 2013, Part I, LNCS 8149, pp. 663–670, 2013.

To the best of our knowledge, the only available method to estimate a per-voxel fiber response is FORECAST (Fiber ORientation Estimated using Continuous Axially Symmetric Tensors) [4]. FORECAST offers a fast estimate of the kernel, but we will demonstrate that it drastically underestimates the anisotropy of the single fiber response in case of crossing fiber compartments, which limits the angular resolution at which they can be resolved.

We propose an alternative way of selecting a kernel, based on balancing fitting error and fODF sparsity. This amounts to the idea that spurious fODF peaks are a result of overfitting to the measurement noise, and that they can be avoided by regularizing with an appropriate sparsity prior. The assumption that fODFs in white matter are sparse has previously been exploited to achieve super-resolved deconvolution [1]. In this work, we use it to determine the local fiber response.

2 Related Work

FORECAST models the HARDI signal as arising from a continuous mixture of fiber compartments that are each described by an axially symmetric diffusion tensor [4]. The parameters of those tensors, mean diffusivity $\bar{\lambda}$ and perpendicular diffusivity $\lambda_\perp$, are assumed to be constant within each voxel, but not throughout the brain. This is equivalent to performing spherical deconvolution with a per-voxel response function that corresponds to the axially symmetric tensor model.

In numerical experiments, FORECAST has achieved excellent results [4]. However, these experiments have used the true value of $\bar{\lambda}$, which was known from simulating the data, during analysis. When analyzing in vivo data, that true value is unknown. In this case, FORECAST uses the mean diffusivity *measured in the voxel* as an estimate of $\bar{\lambda}$, the mean diffusivity *of the single fiber response.*

Unfortunately, this heuristic underestimates $\bar{\lambda}$ significantly when two fiber compartments cross. This is why FORECAST exhibits a strong bias in estimated FA and a reduced ability to resolve crossings in our own experiments in Section 4, where we analyze simulated data with the same heuristic that FORECAST would apply to in vivo data. The key advantage of our novel method is that it estimates a per-voxel kernel without assuming prior knowledge of $\bar{\lambda}$.

3 Balancing Fitting Error and Sparsity

Given a kernel $\mathbf{R}$, spherical deconvolution [2] obtains the spherical harmonics coefficients $\hat{\mathbf{f}}_\mathbf{R}$ of the corresponding fiber ODF $\hat{F}_\mathbf{R}(\theta, \phi)$ by minimizing

$$\hat{\mathbf{f}}_\mathbf{R} = \arg\min \|\mathbf{QRf}_\mathbf{R} - \mathbf{s}\|_2, \tag{1}$$

where $\mathbf{s}$ is a vector of measured signal attenuations in n gradient directions, $\mathbf{R}$ convolves $\mathbf{f}_\mathbf{R}$ with a kernel representing the single fiber response, and $\mathbf{Q}$ evaluates the spherical harmonics in the n gradient directions.

Even though $\mathbf{f}_\mathbf{R}$ is commonly constrained to represent a non-negative function $F_\mathbf{R}(\theta, \phi)$ [1], it has been pointed out that this does not reliably prevent erroneous

peaks in the fODF [5], especially when the kernel $\mathbf{R}$ overestimates the anisotropy of the true single fiber response [3]. We propose to reduce spurious peaks via an additional sparsity constraint on $F_{\mathbf{R}}$, by selecting a kernel $\hat{\mathbf{R}}$ such that

$$\hat{\mathbf{R}} = \arg\min \frac{1}{\sqrt{n}} \|\mathbf{Q}\mathbf{R}\hat{\mathbf{f}}_{\mathbf{R}} - \mathbf{s}\|_2 + \nu \|\hat{F}_{\mathbf{R}}(\theta, \phi)\|_{0.5}, \tag{2}$$

where ν controls the tradeoff between fitting error and sparsity. In our experiments, this parameter did not require any tuning; it has been fixed at $\nu = 0.02$.

An ambiguity in Eq. (2) arises from the fact that scaling $\mathbf{R}$ by factor c can be offset by scaling $\hat{\mathbf{f}}_{\mathbf{R}}$ by c^{-1}. We avoid this problem by fixing the scale of all candidate kernels $\mathbf{R}$ such that all $\hat{F}_{\mathbf{R}}$ integrate to unity. Since this constrains the ℓ_1 norm of $\hat{F}_{\mathbf{R}}$ to a constant, we instead use the $\ell_{0.5}$ norm to induce sparsity, which we define as $\|F(\theta, \phi)\|_{0.5} := \iint \sqrt{|F(\theta, \phi)|}$.

3.1 The Fiber Response as a Function of cFA

We simplify the problem of finding an optimum of Eq. (2) by only considering kernels that correspond to axially symmetric tensors. Their spherical harmonics coefficients are derived in [4] as a function of b value and parameters $\bar{\lambda}$ and $\lambda_\perp$.

For our purposes, a re-parametrization in terms of parallel diffusivity $\lambda_\parallel$ and Fractional Anisotropy (FA) is more convenient, since FA has a standardized range over which we can optimize. From FA and $\lambda_\parallel$, we compute $\lambda_\perp$ by observing that the difference $\Delta := \lambda_\parallel - \lambda_\perp$ is the unique root of the quadratic function

$$(2\,\mathrm{FA}^2 - 1)\Delta^2 - 4\,\mathrm{FA}^2\lambda_\parallel\Delta + 3\,\mathrm{FA}^2\lambda_\parallel^2 = 0 \tag{3}$$

that falls within $[0, \lambda_\parallel]$. Since $\bar{\lambda} := (\lambda_\parallel + 2\lambda_\perp)/3$, this easily reduces our preferred parameters to those used in [4].

During optimization, we use FA of the single fiber response as the independent variable that we modify to achieve an optimum of Eq. (2). The second parameter of $\mathbf{R}$, $\lambda_\parallel$, is always chosen such that, for the given value of FA, the fODF $\hat{F}_{\mathbf{R}}$ integrates to unity. As shown in [4], this requires that $\lambda_\parallel$ and Δ correctly predict the average signal attenuation

$$\frac{\bar{S}}{S_0} = \frac{\sqrt{\pi}}{2} \frac{\mathrm{erf}(\sqrt{b\Delta})}{\sqrt{b\Delta}} \mathrm{e}^{-b(\lambda_\parallel - \Delta)}. \tag{4}$$

For a fixed choice of FA, the attenuation predicted by Eq. (4) is a monotonic function of $\lambda_\parallel$. Therefore, we can use a simple bracketing technique to numerically solve for the value of $\lambda_\parallel$ that predicts the empirical signal attenuation.

Unlike the standard Fractional Anisotropy computed from diffusion tensor MRI, which quantifies the anisotropy of *diffusion within the voxel,* the FA parameter inferred during calibration models the anisotropy *of the individual fiber compartments.* Thus, it is affected far less by fiber crossings or fanning. To emphasize this difference, we will refer to this new measure as *calibration FA* (cFA).

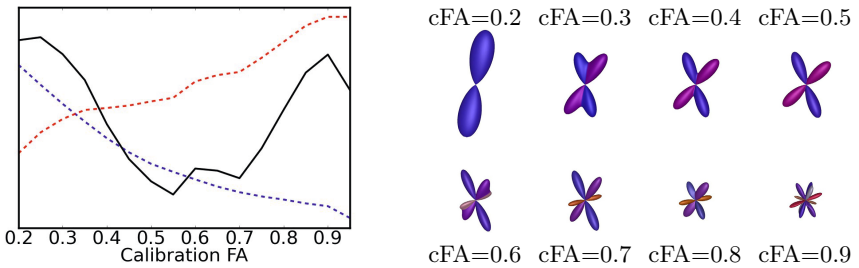


Fig. 1. Calibrating deconvolution to higher cFA reduces fitting error (blue dashes), but introduces erroneous fODF peaks. Our method avoids this via an objective function (black line) that combines fitting error with a sparsity term (red dashes).

Selecting a lower cFA amounts to using a smoother convolution kernel. Therefore, fitting residuals increase as calibration FA decreases (cf. blue dashes in Fig. 1). At the same time, fODFs become sparser, eliminating erroneous peaks that are caused by overfitting to the measurement noise when using an overly large cFA. Our method automatically determines a suitable tradeoff.

3.2 Finding the Optimal Kernel

For a given candidate $\mathbf{R}$, we evaluate Eq. (2) by first finding the corresponding $\hat{\mathbf{f}}_{\mathbf{R}}$ via a non-negative constrained version of Eq. (1). We then evaluate the fitting error and estimate the $\ell_{0.5}$ norm by taking m samples F_i of $F_{\mathbf{R}}(\theta, \phi)$, distributed uniformly on the hemisphere, and computing

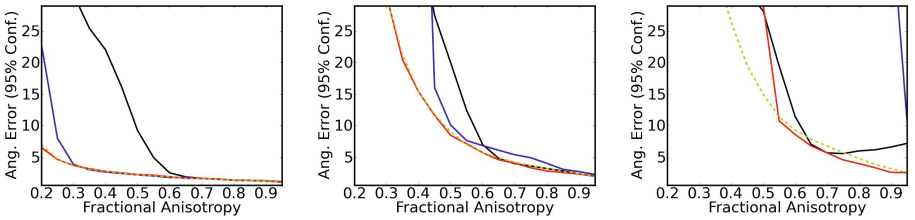
$$\|F(\theta,\phi)\|_{0.5} \approx \frac{4\pi}{m} \sum_{i=1}^{m} \sqrt{F_i}. \tag{5}$$

In practice, we simply use the same $m = 300$ directions that enforce non-negativity in the iterative implementation of CSD by Tournier et al. [1].

Since the dependence of $\hat{\mathbf{f}}_{\mathbf{R}}$ on $\mathbf{R}$ cannot be expressed in closed form, we cannot compute an analytical derivative of Eq. (2) for optimization. Instead, we simply continue to increase or decrease the value of cFA by some stepsize, as long as the steps improve the objective value, and remain within cFA $\in [0.2, 0.95]$. Once no more improvements are possible, we continue with a reduced stepsize, until it falls below 0.01. This easy-to-implement procedure permits auto-calibrated deconvolution of a full brain scan ($> 160,000$ voxels) in around 22 minutes on a standard quad-core 2.67 GHz workstation.

4 Results on Simulated Data

To confirm that sparsity-based kernel selection produces useful results, we perform numerical experiments, assuming the same acquisition parameters as the clinical sequence we routinely use for deconvolution-based tractography: 64 uniformly distributed gradient directions, $b = 2000\,\mathrm{s/mm^2}$, plus one image at $b = 0$.



(a) Single Fiber (b) 90° Crossing (c) 60° Crossing

Fig. 2. In simulated low-noise data, auto-calibration (red) produces more accurate results than deconvolution with a fixed kernel (black) or FORECAST (blue)

4.1 Studying Model Assumptions in Low-Noise Data

We focus on the effects of model assumptions by simulating diffusion weighted data with weak Rician noise ($SNR_0 = 50$) from a mixture of axially symmetric diffusion tensors with normally distributed $\bar{\lambda}$ (statistics taken from a white matter ROI in a healthy subject). We report the 95% confidence intervals, over 500 trials, of the angular distance between largest fODF peaks and ground truth fiber directions. If an fODF has too few significant peaks (using 0.1 as a threshold), we assign the maximum error (90°) to the missing ground truth direction.

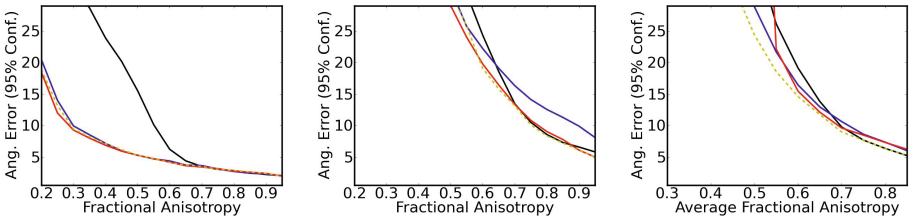
Fig. 2 compares the performance of constrained spherical deconvolution with a fixed calibration to cFA = 0.75 (black lines) to FORECAST (blue) and our proposed auto-calibration (red). We also provide the accuracy of an "oracle"-based deconvolution that uses the true kernel that was used to simulate the data (dashed yellow line). This is not a practically useful method, since the kernel is unknown for in vivo data. However, it provides a reference which confirms that in many cases, auto-calibration performs close to optimal.

Results confirm a previous finding that even when noise is very low, deconvolution is unreliable when the kernel overestimates the true anisotropy [3]. Auto-calibration effectively ameliorates this problem, especially in the single-fiber case. For a 90° fiber crossing, it performs only marginally better than FORECAST; however, it is able to resolve smaller angles, such as the 60° shown in Fig. 2 (c), which FORECAST only resolves at the highest anisotropies.

Our comparison does not include the damped Richardson-Lucy algorithm, which was proposed as an alternative way of reducing spurious peaks in deconvolution [5]. However, Parker et al. [3] report that this method is not able to deal with low target FA ≤ 0.5. Our method remains useful beyond that point.

4.2 Stability at Lower SNR

Under more realistic noise ($SNR_0 = 20$), auto-calibration continues to provide more reliable estimates than non-adaptive deconvolution, especially in the single fiber case (Fig. 3 (a)). For crossing fibers (in (b): 75° with volume fractions 0.6 : 0.4), it provides better accuracy than FORECAST, and a higher angular resolution: Unlike FORECAST, it continues to reliably resolve 60° crossings.



(a) Single Fiber (b) 75° Crossing, 0.6 : 0.4 (c) 90°, Unequal FA

Fig. 3. At $\mathrm{SNR}_0 = 20$, auto-calibration (red) continues to provide a clear advantage over standard deconvolution (black), especially for single-fiber voxels (a)

Fig. 3 (c) demonstrates that auto-calibration still performs well when its model assumptions are violated: Even though the crossing fibers do not have the same FAs in this case (simulated FAs are 0.1 larger and smaller than the average value specified in the plot), performance is on par with standard CSD.

Even though a cFA value can also be derived from the FORECAST kernel, we found that it exhibits a strong bias. Given 60- to 90-degree two-fiber crossings with FA in [0.5,0.95], the mean difference between the cFA from FORECAST and the ground truth was -0.253 ± 0.064. This bias is caused by the biased estimate of $\bar{\lambda}$, which is explained in Section 2, and is removed in our method, whose estimated cFA matches the true FA with accuracy 0.030 ± 0.061.

4.3 Required Number of $b = 0$ Measurements and b Value

Our HARDI acquisition scheme maximizes angular resolution by taking as many different gradient directions as possible within the clinically feasible time budget, leaving just a single image with $b = 0$. In contrast, Anderson takes a much larger number of $b = 0$ images for FORECAST, arguing that via Eq. (4), S_0 plays an important role in estimating the kernel [4]. We repeated our simulations with 60 uniform directions plus 5 $b = 0$ images, rather than our default 64+1, as well as with 55+10. For both single and crossing fibers, we found that taking more images at $b = 0$ reduced variance in the estimates of cFA and $\lambda_{\|}$ during calibration, and therefore leads to a more reliable estimation of the deconvolution kernel. Unfortunately, the loss of angular resolution that results from having fewer directions still led to a slight increase in angular reconstruction errors.

We also tried several b values in our numerical experiments and found that reconstructing a 75° crossing of fibers with FA $\in [0.5, 0.95]$ with a 95% confidence in orientation of less than 30° at $\mathrm{SNR}_0 = 20$, required $b \geq 1400\,\mathrm{s/mm^2}$.

5 Results on Measured Data

5.1 Mapping Kernel Parameters

Fig. 4 confirms a theoretical advantage of our novel cFA measure: In regions where decrease of traditional FA is known to be due to fiber crossings and spread

(a) FA from DT-MRI (b) cFA from Calibration (c) Neurite Density [6]

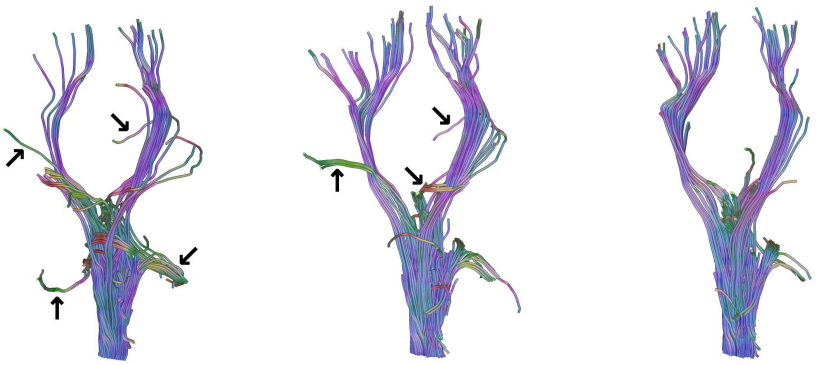
Fig. 4. Compared to Fractional Anisotropy from the diffusion tensor model (a), our novel cFA parameter (b) depends less on orientation dispersion within the voxel

(marked by arrows in (a)), cFA remains high (b). In this respect, it is similar to the neurite density parameter in the recently proposed NODDI framework [6]. However, unlike NODDI, it does not require diffusion acquisition on a second b shell. We believe that maps of kernel parameters such as cFA could complement recent efforts to derive new quantitative measures from fODFs [7,8].

ODF sparsity is reduced in regions of fanning fibers. Even though Fig. 4 (b) suggests that our method can deal with this violation of its model assumption to some degree, the impact of fanning merits closer investigation in the future.

5.2 Reduction of Erroneous Fibers in Tractography

Finally, we have used auto-calibrated deconvolution as a basis of deterministic streamline tractography in a patient with metachromatic leukodystrophy, a demyelinating white matter disorder. Fig. 5 shows that auto-calibration produced a successful tractography without requiring a user-specified single fiber response. Compared to diffusion tensor tractography and CSD with the standard global calibration procedure [2], it reduces the number of streamlines that deviate into



(a) DT-MRI (b) CSD (c) Auto-Calibration

Fig. 5. Compared to diffusion tensor (a) or constrained deconvolution tractography (b), auto-calibration (c) reduces the number of false positive streamlines

adjacent bundles (arrows in (a) and (b)). All methods start from the same seeds in the brain stem and use the same curvature-based termination criteria.

6 Conclusion

We have extended the widely used spherical deconvolution technique with a method for auto-calibration. We have successfully applied it to simulated data with different b values and noise levels, and to measured data from healthy subjects, as well as from patients with white matter degeneration, without having to tune any parameters or providing a global single fiber response.

In numerical experiments, our per-voxel kernel estimates outperform constrained deconvolution with a fixed kernel, and FORECAST, the only other method that estimates a per-voxel deconvolution kernel. Results on measured data indicate a potential for using kernel parameters for quantitative analysis, and for improving the specificity of streamline tractography.

References

1. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. NeuroImage 35, 1459–1472 (2007)
2. Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage 23, 1176–1185 (2004)
3. Parker, G.D., Marshall, D., Rosin, P.L., Drage, N., Richmond, S., Jones, D.K.: A pitfall in the reconstruction of fibre ODFs using spherical deconvolution of diffusion MRI data. NeuroImage 65, 433–448 (2013)
4. Anderson, A.W.: Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magnetic Resonance in Medicine 54(5), 1194–1206 (2005)
5. Dell'Acqua, F., Scifo, P., Rizzo, G., Catani, M., Simmons, A., Scotti, G., Fazio, F.: A modified damped richardson-lucy algorithm to reduce isotropic background effects in spherical deconvolution. NeuroImage 49, 1446–1458 (2010)
6. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C.: NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage 61(4), 1000–1016 (2012)
7. Raffelt, D., Tournier, J.D., Rose, S., Ridgway, G.R., Henderson, R., Crozier, S., Salvado, O., Connelly, A.: Apparent fibre density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. NeuroImage 59(4), 3976–3994 (2012)
8. Dell'Acqua, F., Simmons, A., Williams, S.C.R., Catani, M.: Can spherical deconvolution provide more information than fiber orientations? hindrance modulated orientational anisotropy, a true-tract specific index to characterize white matter diffusion. Human Brain Mapping (2012) Early view, doi:10.1002/hbm.22080